Removing Arbitrary-Scale Rain Streaks via Fractal Band Learning With Self-Supervision

Wenhan Yang, *Member, IEEE*, Shiqi Wang, *Member, IEEE*, and Jiaying Liu, *Senior Member, IEEE*

***Abstract*— Data-driven rain streak removal methods, most of which rely on synthesized paired data, usually come across the generalization problem when being applied in real scenarios. In this paper, we propose a novel deep-learning based rain streak removal method injected with self-supervision to obtain the capacity of removing more varied-scale rain streaks in practical applications. To this end, in this work, efforts are made from two perspectives. 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 as neural forms and effectively improves the capacity to capture discriminative features for deraining. Second, to further improve the generalization ability of FBL to remove rain streaks of varied scales, we incorporate scale-robust self-supervision to regularize the network training. The constraint forces the extracted features of an input rain image at different scales to be equivalent after rescaling operations. Therefore, our method can offer similar responses based on solely image content without the interference of scale change and is capable to remove varied-scale rain streaks. Extensive experiments in quantitative and qualitative evaluations demonstrate the superiority of our method for rain streak removal, especially for the real cases where very large rain streaks exist, and prove the effectiveness of each component.**

***Index Terms*— Rain streak removal, varied scale, frequency band learning, self-supervision, deep network.**

I. 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 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.

As many degradation factors cause information loss, it is highly ill-posed to address single image rain removal problem. Previous methods [28], [32], [47], [55] take single image rain removal as a signal separation problem that separates rain streaks and background images, namely the corresponding rain-free versions, from their mixed versions. In these works, various models are developed to extract rain streaks and background images based on their texture appearance patterns, such as frequency domain representation [32], sparse representation [47], and Gaussian mixture model [43].

Recently, the appearance of deep learning gives rise to deep-network based approaches. In [16], 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 [65], a deep network is designed to jointly detects and removes rain streaks for heavy rain cases. Successive works [42], [70] 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 complicated. Existing rain models often neglect the diversity of rain scales. A model trained with streaks of a scale is difficult to be generalized to handle streaks of a different one. Some previous works [40], [64] make preliminary attempts. However, they either use a few times parameters of a single-scale model [40] or are restricted to process the images of a certain scale whose magnitude is the order of two compared to the scale in the training phase [64]. However, a more economic and flexible framework that is easily to be generalized to process varied-scale rain images is still absent.
- Recent deep-learning based methods take pure feed-forward CNN, ResNet or DenseNet *etc.* as backbone methods. However, there is no theoretical or conceptual connection between these neural models and traditional theories. Such gap prevents us 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. However, intuitively, rain removal is a signal separation problem, where the features in the

Manuscript received May 8, 2019; revised April 2, 2020; accepted April 30, 2020. Date of publication May 19, 2020; date of current version July 6, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2018AAA0102702, in part by the Hong Kong RGC Early Career Scheme under Grant 9048122 (CityU 21211018), in part by the National Natural Science Foundation of China under Contract 61772043, and in part by the Beijing Natural Science Foundation under Contract L182002. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Marta Mrak. *(Corresponding author: Shiqi Wang.)*

Wenhan Yang and Shiqi Wang are with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: wyang34@cityu.edu.hk; shiqwang@cityu.edu.hk).

Jiaying Liu is with the Institute of Computer Science and Technology, Peking University, Beijing 100080, China (e-mail: liujiaying@pku.edu.cn).

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TIP.2020.2993406

texture or frequency band domain play an important role. Thus, the properties of traditional band theory provide 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 with two notable features. First, the developed architecture connects with frequency band recovery theory. Second, it is capable of effectively restoring clean images from inputs which contain rain streaks of different scales that may not appear in the training set. To this end, we first briefly illustrate the connections between frequency band recovery and deep learning. Subsequently, 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 of varied scales, a scale-robust self-supervision is utilized to regularize network training. FBL with self-supervision (FBL-SS) forces consistency of the extracted features of inputs at different scales after zooming. Therefore, FBL-SS can acquire scale-robust feature representation given input images at different scales. With the learned band feature representations and the power of scale-robust self-supervision, our FBL-SS is superior to previous deraining networks and is capable to remove varied-scale rain streaks. Our contributions are summarized as follows,

- We provide a novel understanding to link frequency band recovery theory 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, and prove the effectiveness of its each components.
- A scale-robust self-supervision constraint is proposed to regularize the training of a deraining network. The constraint guides the network to extract scale-robust features and deal with rain images containing varied-scale rain streaks. It can significantly benefit the rain removal on real rain images at large scales.

The rest of this paper is organized as follows. Section II briefly reviews the related work. Section III connects the traditional band recovery theory and deep learning techniques, designs new deep networks inspired by the connection, and introduces scale-robust self-supervision to obtain better generalization capacity to handle rain streaks of varied scales. Experimental results and concluding remarks are presented in Sections IV and V, respectively.

II. Related Works

A. Single Image Rain Removal

Single image rain removal is highly ill-posed as the rain streaks occlude and contaminate the background, causing a loss of information. To address the issue, the earliest works regard the single-image rain removal as a signal separation problem from the mixture of rain and background signals, based on texture patterns and the spatial redundancy. Barron *et al.* [32] utilized sparse coding to extract the rain layer from the high-frequency signal extracted by the bilateral filter. In [4], Chen *et al.* developed a generalized low-rank model that enforces the low-rank consistency among the adjacent rain frames to reveal randomly distributed rain streaks. In [34], Kim *et al.* detected and removed rain streaks via the non-local mean filter. Luo *et al.* [47] proposed a novel rain image synthesis model "screen blend model", which enables to capture nonlinear relationship between the rain streaks and backgrounds. Based on the model, the discriminative sparse coding is then applied to extract rain streaks. In [43], online learning is adopted and the Gaussian mixture model is applied to capture the rain streak distribution of the input rain images. In [3], Chang *et al.* injected the line pattern of rain streaks into the proposed model via a rotated image decomposition framework, and applied a compositional directional total variational model and low-rank prior to separate the rain streak and background. In [79], Zhu *et al.* proposed an iterative deraining process, where the background signal is removed from the rain layer and the rain signal is separated from the background layer progressively. Gu *et al.* [20] constructed a joint convolution analysis and synthesis sparse representation model. In this model, an image is decomposed into two layers, where one is approximated by the analysis sparse representation model to represent image large-scale structures while the other is modeled by the synthesis sparse representation to capture fine-scale textures. In [59], Wang *et al.* proposed a 3-layer hierarchical scheme consisting of both image decomposition and dictionary to remove rains and snows.

In 2017, the presence of deep learning promotes the development of single-image deraining and various deep-learning based deraining methods are designed. In [16], a deep network is utilized to estimate the negative residue based on the initially extracted high-frequency details. In [64], [65], a deep network is built to perform rain streak detection and removal jointly, and at the same time to handle rain streaks and accumulation recurrently. In [40], to remove the rain streaks of different scales, several parallel sub-networks are particularly trained with those streaks respectively and then the results are fused to generate the final prediction. Dong *et al.* [70] first detected rain density and then adopted a multi-scale dense network to remove rain streaks with density guidance. In [50], a progressive recurrent network is built via repeatedly unfolding a shallow network, with recurrent layers injected with gate units to model dependencies of deep features among different recurrent stages for rain streak removal. In [17], the Gaussian

Laplacian image pyramid decomposition is introduced into deep networks, which makes the learning process easier and the deraining process more efficiently. In [68], the network first extracts the rain content at different scales as well as its confidence measure. Then, guided by this content, the final derained output is obtained. In [58], a large-scale dataset with rain/rain-free image pairs is captured. After that, a spatial attentive network is built to remove rain streaks from local to global. In [25], Hu *et al.* created a new dataset with rain streaks and accumulation and designed an end-to-end deep network to extract the depth-attentional features for regressing the residual signal.

Following [16], [40], [65], [70], our work also follows the route of deep-learning based single image rain streak removal. In particular, we construct a deep network motivated by band recovery theory. The domain gap in streak scales between training and testing phases is addressed by network training with self-supervision. Compared to [65], [66], which jointly detects and removes rain streaks, our work additionally focuses on the generalization performance of rain streak removal methods especially when rain streaks of different sizes are met. The work in [64] also aims to improve the generalization capacity to handle rain streaks of different scales and it constructs a recurrent neural network via unrolling the wavelet transform. Comparatively, our work is based on self-supervised learning and introduces the scale-robust feature-level supervision to obtain more robust feature learning and achieve superior deraining performance.

B. Deep-Learning Based Image Processing and Related Architectures

In recent years, deep learning has brought fast developments to the field of image processing. The related applications include denoising [72], [78], super-resolution [9], [62], [63], deblurring [52], compression artifact removal [8], [74], and style transfer [19], *etc*. recent works begin to pay attention to bad weather restoration or image enhancement, *e.g.* dehazing [2], [6], raindrop and dirt removal [12], [48], low-light enhancement [46], [61], [67], color constancy [1], and underwater enhancement [39], *etc*. Besides, due to its distinguished capacity to model complex mappings from the degraded images/videos, deep networks have been applied to addressing more sophisticated issues, such as blind image denoising [72], image compression [27], [31], quality assessment [13], texture smoothing [80] and video coding [26], [44], [45], [76].

The commonly used network architectures for image processing include feed-forward CNN, U-Net, ResNet, DenseNet and the combination of them. The earliest works employ the simplest feed-forward CNN, such as super-resolution convolutional neural network [9] and artifacts reduction convolutional neural network [8]. U-Net [69], [73] is another simple but powerful network architecture consisting of an encoder and decoder that down-sample feature maps and then up-sample them progressively. Skip connections are used to connect the feature maps from the encoder to decoder to avoid the resolution loss.

The rise of ResNet and DenseNet has further promoted the development of the related tasks. For example, in very deep super-resolution [33] and super-resolution dense network [57], ResNet and DenseNet brings in performance leaps, respectively. From the perspective of feature learning, it leads to better representations of pixels and their contexts for the low-level image processing via refining features progressively like ResNet [24] or concatenating and fusing features from different levels like DenseNet [29]. The related beneficial tasks include super-resolution [38], [56], [63], [75], [77], dehazing [21], [77], inpainting [10], compression artifacts removal [18], and deblurring [36]. In [63], Yang *et al.* made efforts in explaining the connection between ResNet and traditional frequency band recovery theory, and injected the edge information into the ResNet as the prior to infer more accurate high-frequency details. In [10], [18], [36], [38], ResNet is adopted as the generator in the generative adversarial networks. In [77], Zhang *et al.* constructed a densely connected encoder-decoder network injected with the edge-preserving constraint to infer the transmission map for image dehazing. In [56], Tai *et al.* integrated recursive units, gate units and the densely connected structure to simulate the human persistent memory for image restoration. Zhang *et al.* [77] proposed a network that integrates ResNet and DenseNet. Dense blocks are utilized to extract dense local features. All features in each dense block are connected by residual connections, and then fused in the last layer. In [75], residual in residual structure is proposed to construct a very deep network, where several residual groups are cascaded and connected with long skip connections. In [22], by unrolling back-projection process, iterative up- and down-sampling layers are stacked into a dense network. The different layers that perform at different scales are mutually-connected for image super-resolution.

In our work, we inherit the advantages of these previous works, and provide a unified viewpoint, *i.e.* band recovery theory, to understand deep networks for low-level visions. Furthermore, along this vein, we develop a new effective backbone for rain streak removal.

III. Fractal Band Learning

A. From Band Recovery to Deep Learning

In [63], a systematic paradigm is provided to establish the relationship between ResNet [24] 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}; 2) A summation is used to combine the new band signal (residue) $F_{\mathrm{RB}}(f_{t-1})$ and the input (previous band estimators) f_t:

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

where $F_{\mathrm{RB}}(\cdot)$ 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 [11], steerable filter [14], ringlet [7], and the related processing methods based on them [53], [54], four critical properties are revealed:

- *Band refinement*. The given band features extracted from the source domain, *e.g.* a rain image, are transformed into

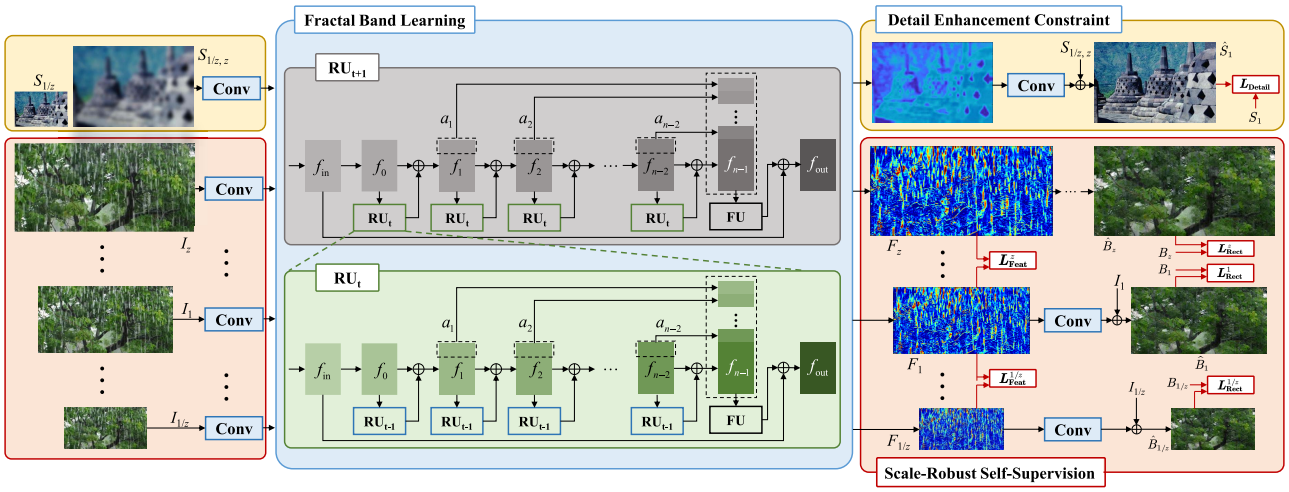


Fig. 1. The framework of our proposed fractal band learning constrained by scale-robust self-supervision (FBL-SS) for rain removal. It performs learned band feature operations progressively and is augmented by being stacked in a fractal form. The model is trained with image-level reconstruction constraint ($\mathbf{L}_{\text{Rect}}$), feature-level scale-robust self-supervision loss ($\mathbf{L}_{\text{Feat}}$), and detail enhancement constraint ($\mathbf{L}_{\text{Detail}}$). $\mathbf{L}_{\text{Rect}}$ guarantees the accuracy of rain removal results at the training scale, and the rough signal fidelities at larger or smaller scales, since some small details are poorly regularized due to rescaling operations. $\mathbf{L}_{\text{Feat}}$ enforces consistency of the feature responses of the network across different scales. $\mathbf{L}_{\text{Detail}}$ focuses on compensating for the lost details caused by the rescaling operations used in $\mathbf{L}_{\text{Rect}}$ and $\mathbf{L}_{\text{Feat}}$.

the target domain, *e.g.* a rain-free image. As the analysis above, ResNet works in this way.

- *Band expansion*. One band feature is split into several band features, and the total band feature number is increased. The popular DenseNet [29] provides 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 1×1 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.* [37], following this rule.

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

B. Overview of Fractal Band Learning

The framework of our proposed FBL for rain streak removal is illustrated in Fig. 1:

- **Fractal Band Learning** (**FBL**). We target at building 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. 1. To model the hierarchical band dependency, FBL is designed with 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 top of Fig. 1, changes of the band feature dimension in the learning process are provided. The summation operations in dash boxes signify residual additions. From left to right, the band features are refined. Band features are expanded and fused in turn at the penultimate layer.
- **Scale-Robust 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 at 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. 1. After training with this constraint, the extracted features are more scale-invariant and the model is better at removing rain streaks of varied scales.
- **Detail Enhancement Constraint**. Scale-robust self-supervision constraint enforces the consistency of features generated by the network at different scales. However, the rescaling operations that the constraint includes cause the detail loss. To preserve or even enhance high-frequency details, we introduce the detail enhancement constraint. The same network takes blurred

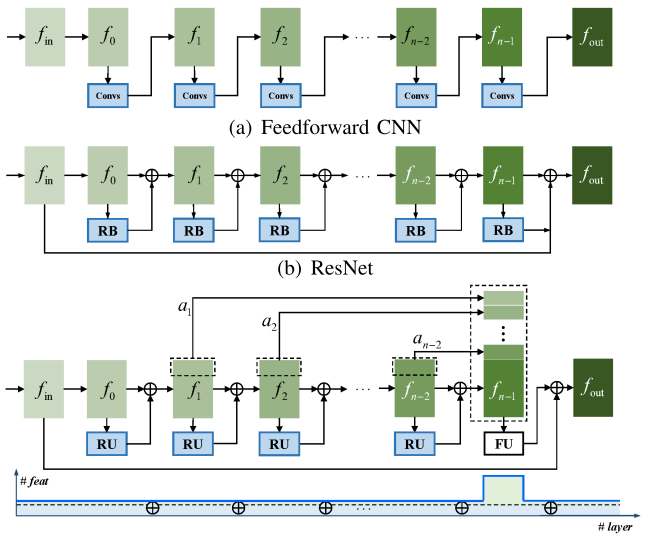


Fig. 2. The framework of our band learning. The left side presents network structures. The bottom side illustrates band feature dimension changes during the learning process.

non-rain images as its input and predicts sharp images as its output. Thus, when facing rainy images with very large rain streaks, the network is capable of removing the large streaks while preserving details well.

C. 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 following these rules step by step.

1) Feed-Forward CNN: The network structure used in SRCNN [9] is shown in Fig. 2 (a). The feature f_t is transformed through a chain of convolutions:

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

where $*$ and σ denote the convolution operation and ReLU activation function, respectively. W_t and b_t are the weights and biases of the t-th convolution layer, respectively. Even in its simplest fashion, 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 difficult 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.

2) ResNet: To characterize tiny structures in signals, residual network is employed in VDSR [33] and DEGREE [63]. As shown in Fig. 2 (b), each residual block extracts a residue and adds it with the original signal:

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

where $F_{\mathrm{RB}}(\cdot)$ denotes the feature transform of residual block. It can be set as stacked convolutions or more complicated modules. As discussed in [63], this operation equals to *progressive band refinement*, which facilitates high frequency detail reconstruction.

3) 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:

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

where $F_{\mathrm{RU}}(\cdot)$ 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 $\{a_t | t = 1, 2, \ldots, n-2\}$ are concatenated with the last output feature f_{n-1}:

$$f_{\mathrm{expansion}} = \left[a_1, a_2, a_3, \ldots, a_{n-2}, f_{n-1}\right]. \tag{5}$$

Subsequently, we conduct *band fusion* to transform $f_{\mathrm{expansion}}$ into a narrow one:

$$f_n = F_{\mathrm{FU}}\left(f_{\mathrm{expansion}}\right), \tag{6}$$

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

$$f_{\mathrm{out}} = f_{\mathrm{in}} + f_n. \tag{7}$$

As observed from the bottom side of Fig. 2 (c), the information is updated regularly, expanded and fused at the end.

4) Fractal Band Learning: In addition to the band properties – *refinement*, *expansion* and *fusion* introduced before, the potential *hierarchical structure* and *dependency* among band features are also important for signal modeling. Thus, we extend our band learning to fractal band learning. The architecture is presented in Fig. 1. 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 viewpoint of band feature dimension changes (in the bottom of Fig. 2 (c)), the information is processed and flowed to the end. In Fig. 2 (a), the information is directly to the end. In Fig. 2 (b), the information is updated regularly without dimension changes. In Fig. 2 (c), the information is updated regularly, expanded and fused at different levels many times, thus a meaningful band feature is obtained.

D. Scale-Robust 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,

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

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,

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

Subsequently, their consistency is ensured as follows,

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

or

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

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.

E. Detail Enhancement Constraint

The components mentioned above may lead to detail loss when dealing with images with large rain streaks, since the rescaling operations used to construct the reconstruction constraint and feature consistency lead to the ignorance of recovering details. Therefore, we add the detail enhancement constraint to remind the network of detail preservation or enhancement. For a rain-free input S_1, we first down-sample it into $S_{1/z}$ by a factor of z and then up-sample it back to the original scale $S_{1/z,z}$ to generate a blurred image. Then, the same deraining network takes $S_{1/z,z}$ as its input and generates the output $\hat{S}_1$, which is constrained to be close to S_1 as follows,

$$\begin{aligned} \hat{S}_1 &= F_{\text{Rect}}\left(F_{\text{FBL}}\left(S_{1/z,z}\right)\right) + S_{1/z,z}, \\ L_{\text{Detail}}\left(S_1, \hat{S}_1\right) &= \left\|S_1 - \hat{S}_1\right\|, \end{aligned} \tag{12}$$

where $F_{\text{Rect}}(\cdot)$ denotes the process to generate the negative rain layer or detail enhancement layer.

F. Network Training

Suppose we have a collection of paired rain and rain-free images $\{y^i, x^i\}_{i=1,\ldots,N}$, where N is the total number of training samples. Then, we rescale each sample using Bilinear interpolation to get $\{y_z^i, x_z^i\}_{i=1,\ldots,N}$ and $\left\{x_{1/z,z}^i, x^i\right\}_{i=1,\ldots,N}$ 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 Θ, $\hat{x}_z^i$, P_z^i to denote all the parameters in F_{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 train $F_{\text{FBL}}(\cdot)$ and $F_{\text{Rect}}(\cdot)$:

$$\begin{aligned} L(\Theta) &= \frac{1}{2N}\sum_{i=1}^{N}\sum_{z>1}\Big(\lambda_z\Big(w_1 L_{\text{Rect},z}^i(\Theta) + w_2 L_{\text{Feat},z}^i(\Theta) \\ &\quad + w_3 L_{\text{Detail},z}^i(\Theta)\Big)\Big), \\ L_{\text{Rect},z}^i(\Theta) &= \left\|F_{\text{Down}}\left(F_{\text{Streak}}(y_z^i;\Theta), z\right) + y^i - x^i\right\|, \\ L_{\text{Feat},z}^i(\Theta) &= \left\|F_{\text{Down}}\left(F_{\text{FBL}}(y_z^i;\Theta), z\right), -F_{\text{FBL}}(y^i;\Theta)\right\|, \\ L_{\text{Detail},z}^i(\Theta) &= \left\|F_{\text{Streak}}(x_{1/z,z}^i;\Theta) + x_{1/z,z}^i - x^i\right\|, \end{aligned} \tag{13}$$

where λ_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.* $z = \{1/2, 1/3, 1/4\}$, we use the following loss to train our network:

$$\begin{aligned} L(\Theta) &= \frac{1}{2N}\sum_{i=1}^{N}\sum_{z<1}\Big(\lambda_z\Big(w_1 L_{\text{Rect},z}^i(\Theta) + w_2 L_{\text{Feat},z}^i(\Theta)\Big)\Big), \\ L_{\text{Rect},z}^i(\Theta) &= \left\|F_{\text{Streak}}(y_z^i;\Theta) + F_{\text{Down}}\left(y^i - x^i, 1/z\right)\right\|, \\ L_{\text{Feat},z}^i(\Theta) &= \left\|F_{\text{Down}}\left(F_{\text{FBL}}(y_z^i;\Theta), 1/z\right), -F_{\text{FBL}}(y^i;\Theta)\right\|. \end{aligned} \tag{14}$$

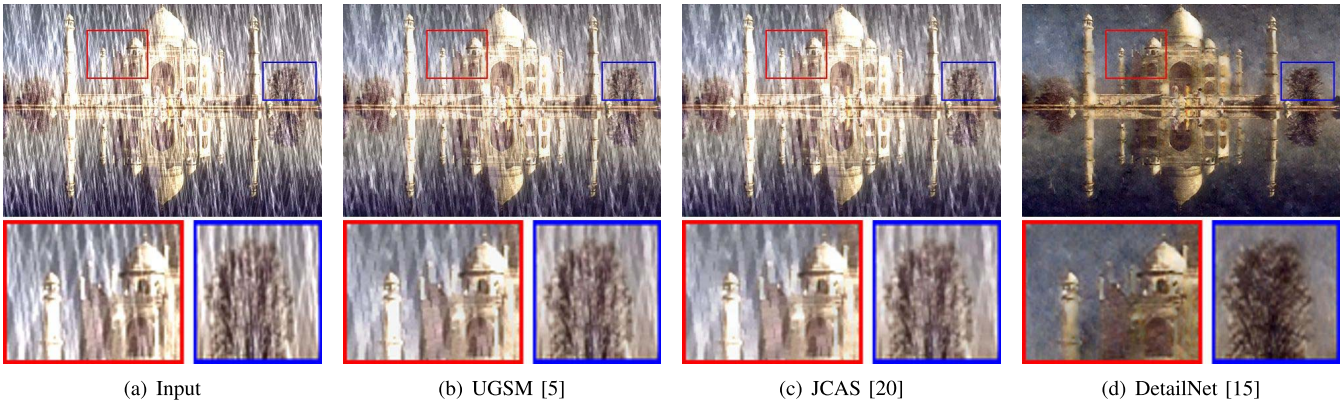
In this case, up-sampling operations are not involved. Thus, we do not introduce detail enhancement constraint.

The training images are cropped into small sub-images with a size of 64×64 pixels. We train our model on Pytorch. ADAptive Moment estimation (ADAM) [35] is used for training the model. The batch size and weight decay are set to 16 and 0.0001, respectively. The network is initialized by MSRA algorithm [23]. In particular, we set the learning rate as 0.0001 throughout the whole training procedure. We only allow at most 200 epochs.

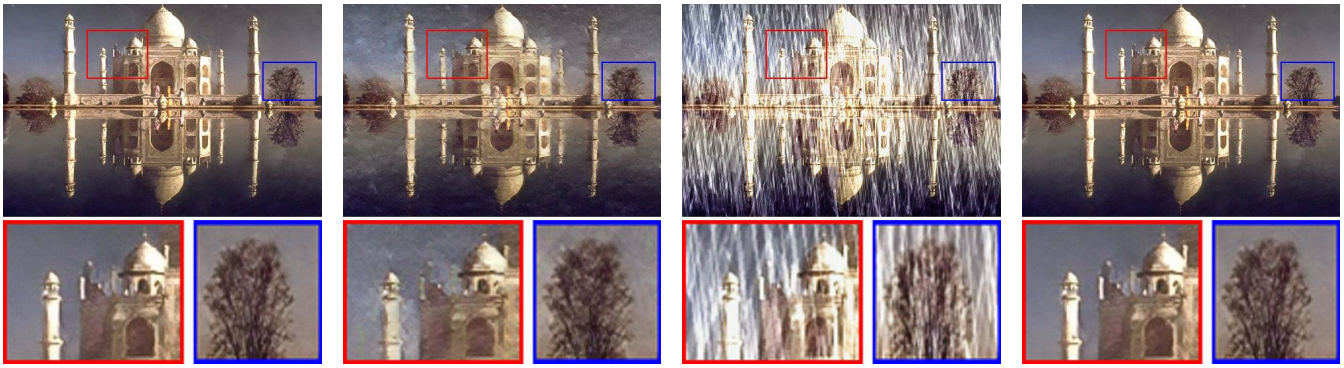
IV. EXPERIMENTS

A. Dataset

We compare our method with state-of-the-art methods on a few benchmark datasets: (1) *Rain100L* [65], which is the synthesized data set with only one type of rain streaks; (2) *Rain100H* [65], which is the synthesized data set with five streak directions; (3) *Rain100H-S2* and *Rain100-S3* proposed in [64], 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. (4) *Rain800* [71], 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. (5) *Rain1400* [16], including 12,600 paired training images and 1,400 paired testing images. (6) *Rain1200* [70], including 12,000 paired training images and 1,200 paired testing images.



(a) Input (b) UGSM [5] (c) JCAS [20] (d) DetailNet [15]



(e) JORDER [65] (f) RESCAN [42] (g) DID-MDN [70] (h) FBL

Fig. 3. Visual quality comparison of different rain removal algorithms on a synthesized rain image from *Rain100H*. It is observed that, our FBL successfully removes most rain streaks and better preserves texture details. Zooming-in the figure will provide a better look at the restoration quality. Enlarged results are provided in the supplementary material.

B. Baseline Methods

We compare our method with seven state-of-the-art methods: image decomposition (ID) [32], discriminative sparse coding (DSC) [47], layer priors (LP) [43], CNN-based rain drop removal (CNN) [12], deep detail network (DetailNet) [16], directional global sparse model (DGSM) [5], joint convolutional analysis and synthesis (JCAS) [20], density-aware multi-stream dense network (DID-MDN) [70], joint rain detection and removal (JORDER) [65], recurrent squeeze-and-excitation context aggregation net (RESCAN) [42], and recurrent wavelet learning (RWL) [64]. 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.

For the experiments on synthesized data, two metrics Peak Signal-to-Noise Ratio (PSNR) [30] and Structure Similarity Index (SSIM) [60] are used as comparison criteria. We evaluate the results only in the luminance channel, which has a significant impact on the human visual system to perceive the image quality.

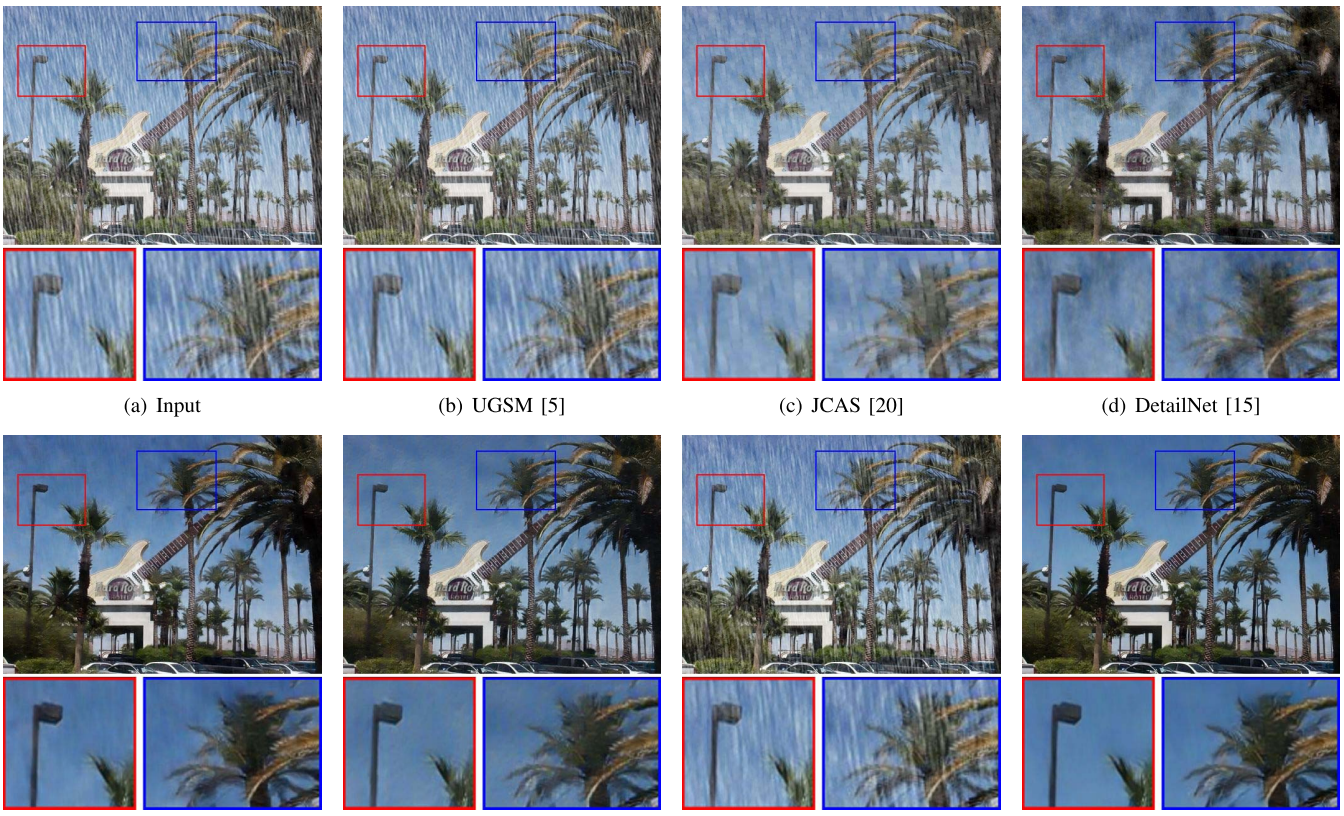
TABLE I

PSNR RESULTS AMONG DIFFERENT METHODS ON *Rain100L*, *Rain100H*, AND *Rain800*

Baseline	*Rain100L*	*Rain100H*	*Rain800*
ID [32]	23.13	13.78	20.54
DSC [47]	24.16	15.66	25.57
LP [43]	29.11	14.26	27.09
CNN [12]	23.70	13.21	23.95
SRCNN [9]	32.63	18.29	25.10
DetailNet [15]	33.50	23.93	25.22
UGSM [5]	28.83	14.06	23.12
JCAS [20]	29.91	14.26	22.25
DID-MDN [70]	28.27	13.85	22.55
ID-CGAN [71]	23.39	16.86	23.81
JORDER [65]	36.11	24.10	26.73
RESCAN [42]	38.58	26.06	28.04
RWL [64]	36.75	26.89	27.79
LPNet [17]	33.61	23.16	22.21
PReNet [50]	38.03	28.94	23.31
SPANet [58]	29.43	14.16	22.67
UMRL [68]	31.98	24.91	24.37
FBL	**40.32**	**30.47**	**28.68**

C. Quantitative Evaluation

Tables I and II show the results of different methods on *Rain100L*, *Rain100H*, and *Rain800*. As observed, our method considerably outperforms previous 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. We also compare our FBL on two popular datasets: *Rain1400* [16] and *Rain1600* [70] in Table III. Our method also achieves superior performance gains to previous methods, which provides useful evidence regarding the effectiveness of our method.



(a) Input (b) UGSM [5] (c) JCAS [20] (d) DetailNet [15]

(e) JORDER [65] (f) RESCAN [42] (g) DID-MDN [70] (h) FBL

Fig. 4. Visual quality comparison of different rain removal algorithms on a synthesized rain image from *Rain800*. It is observed that, our FBL successfully removes most rain streaks and better preserves texture details. Zooming-in the figure will provide a better look at the restoration quality. Enlarged results are provided in the supplementary material.

TABLE II

SSIM RESULTS AMONG DIFFERENT METHODS ON *Rain100L*, *Rain100H*, AND *Rain800*

Baseline	*Rain100L*	*Rain100H*	*Rain800*
ID [32]	0.7022	0.3968	0.6739
DSC [47]	0.8728	0.5444	0.6521
LP [43]	0.8786	0.4225	0.7801
CNN [12]	0.8076	0.3712	0.6589
SRCNN [9]	0.9357	0.6124	0.8232
DetailNet [15]	0.9444	0.7251	0.8228
UGSM [5]	0.8876	0.4454	0.7675
JCAS [20]	0.9053	0.4837	0.7682
DID-MDN [70]	0.8625	0.3748	0.7639
ID-CGAN [71]	0.8275	0.4921	0.8072
JORDER [65]	0.9705	0.7490	0.8683
RESCAN [42]	0.9805	0.8128	0.8748
RWL [64]	0.9754	0.8406	0.8795
LPNet [17]	0.9583	0.8011	0.7893
PReNet [50]	0.9804	0.8934	0.8011
SPANet [58]	0.9043	0.4305	0.7395
UMRL [68]	0.9552	0.8103	0.8191
FBL	**0.9863**	**0.8973**	**0.8875**

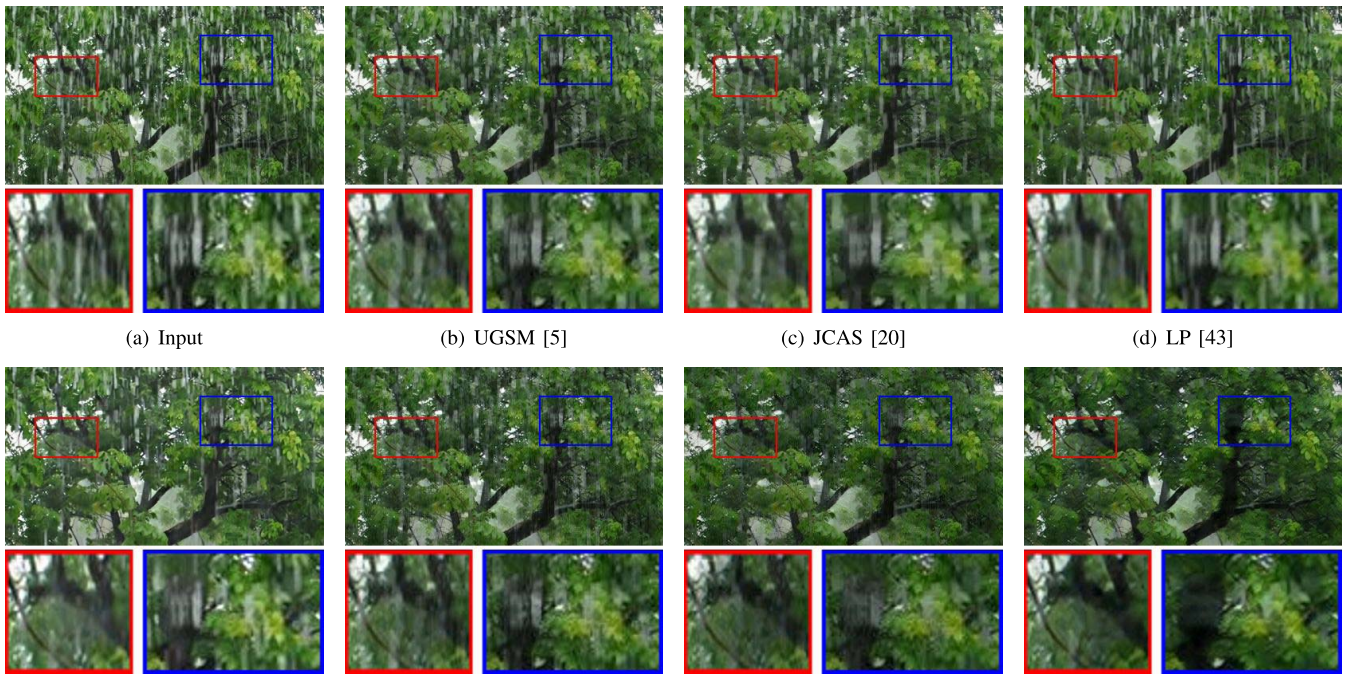
TABLE III

PSNR AND SSIM RESULTS OF DIFFERENT METHODS ON *Rain1400* AND *Rain1200*

Dataset	*Rain1400*		*Rain1200*	
Metric	PSNR	SSIM	PSNR	SSIM
LP [43]	26.53	0.83	22.46	0.80
JCAS [20]	26.80	0.85	25.16	0.81
DDN [15]	29.99	0.89	30.95	0.86
JORDER [65]	28.90	0.90	29.75	0.87
DID-MDN [70]	29.84	0.90	29.65	0.90
RESCAN [42]	31.18	0.91	32.35	0.89
RWL [64]	27.80	0.86	29.58	0.88
LPNet [17]	23.17	0.82	22.65	0.80
PReNet [50]	32.57	0.93	26.21	0.75
SPANet [58]	26.13	0.82	21.05	0.66
UMRL [68]	30.54	0.90	32.36	0.92
FBL	**33.25**	**0.93**	**35.16**	**0.93**

D. Qualitative Evaluation

Figs. 3-8 show the results of synthesized and real images with rain streaks of different scales. As observed, our method significantly outperforms previous state-of-the-art methods. In general, JORDER, RESCAN and FBL achieve better visual quality than other methods. As shown in Figs. 3-8, our FBL is better at removing rain streaks than JORDER and RESCAN, especially for large rain streaks. Besides, our FBL is also superior to other methods in detail preservation. Fig. 3 and Fig. 4 show the results of different methods on two synthetic images from *Rain100H* and *Rain800*, respectively. The input image in Fig. 3 includes long and sharp rain streaks while that in Fig. 4 includes moderate-size and blurry streaks. The results demonstrate that our method successfully removes both kinds of rain streaks. Fig. 5 shows the results of a real image



(a) Input (b) UGSM [5] (c) JCAS [20] (d) LP [43]

(e) DetailNet [15] (f) JORDER [65] (g) RWL [64] (h) FBL-SS

Fig. 5. Visual quality comparison of different rain removal algorithms on a real rain image. It is observed that, our FBL-SS successfully removes most rain streaks and better preserves texture details. Zooming-in the figure will provide a better look at the restoration quality. Enlarged results are provided in the supplementary material.

including very large and abrupt rain streaks. It is clearly showed that, without the consideration of generalization performance to handle rain streaks of different scales, most methods (Fig. 5(b)-(f)) fail to remove very large rain streaks. Benefiting from the recurrent inference, RWL can better remove large streaks. Comparatively, our method obtains the cleanest result with very clear details. Fig. 6 shows the results of a real image including moderate-size and blurry streaks. The results also demonstrate the superiority of our method to remove blurry streaks and our result owns the cleanest background. Fig. 7 includes the results of an image with sparse and small streaks, our method is capable to remove most of the streaks while generating fewer artifacts. Fig. 8 shows the results of a synthesized rain image (*Rain100H-S2*), while its scale is different from that of the training set (*Rain100H*). In this case, the superiority of our method can be observed more clearly. In our result, the background is well restored while those of other methods include obvious visual artifacts.

E. Running Time Comparison

The running time of different methods is presented in Table IV. It is observed that, our FBL is comparable to other recent deep learning methods in time complexity and runs much faster than traditional optimization-based methods.

TABLE IV

THE RUNNING TIME (IN SECONDS) OF DIFFERENT METHODS. (G) AND (C) DENOTE THE IMPLEMENTATION ON GPU AND CPU, RESPECTIVELY. THE SIZE OF THE TESTING IMAGE: 500 × 500

Baseline	Time	Baseline	Time
DSC [47] (C)	611.91	JORDER [65] (G)	1.46
LP [43] (C)	2708.20	DetailNet [15] (G)	0.58
UGSM [5] (C)	2.30	DID-MDN [70] (G)	2.94
JCAS [20] (C)	179.56	RESCAN [42] (G)	0.61
CNN [12] (G)	1529.85	FBL (G)	0.81

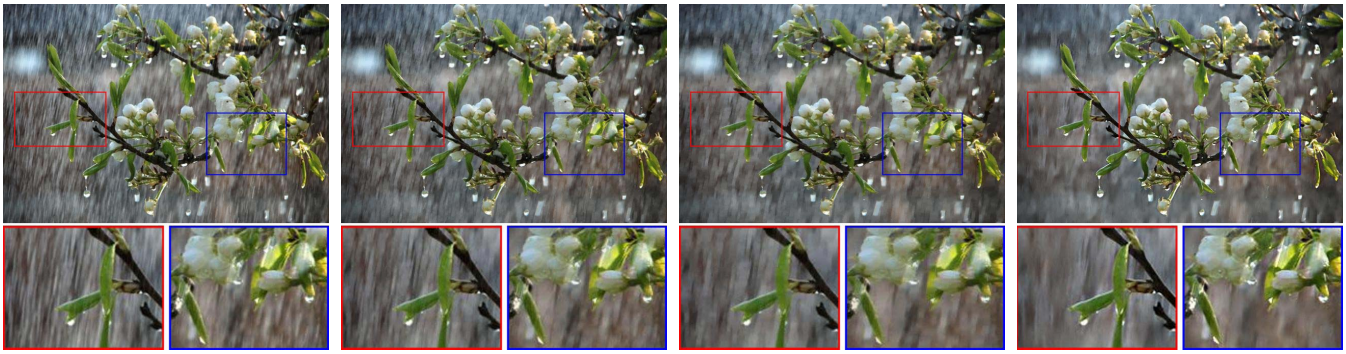
TABLE V

PSNR AND SSIM RESULTS WHEN TRAINING AND TESTING STREAK SIZES ARE DIFFERENT. FBL-SS DENOTES THE PROPOSED FBL DERAINING METHOD WITH SCALE-ROBUST SELF-SUPERVISION

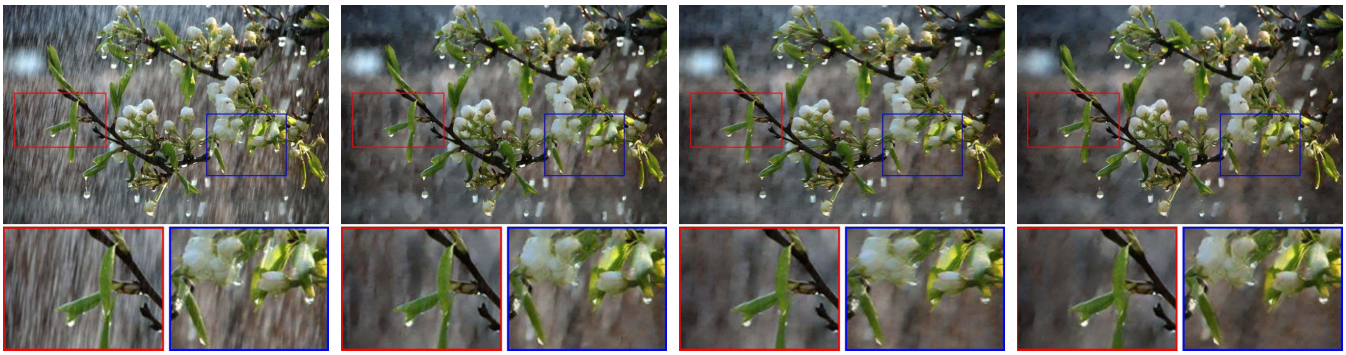
Baseline	*Rain100H-S2*		*Rain100H-S3*	
Metric	PSNR	SSIM	PSNR	SSIM
DetailNet [15]	15.98	0.6328	15.03	0.6410
JORDER [65]	16.11	0.6464	15.21	0.6483
DID-MDN [70]	14.36	0.5383	14.70	0.6142
RESCAN [42]	17.13	0.6965	16.31	0.6787
JCAS [20]	14.27	0.6350	14.07	0.5779
UGSM [5]	14.03	0.6260	13.82	0.5569
FBL	18.51	0.7355	16.32	0.6860
R-CNN	19.45	0.7668	17.32	0.7481
RWL [64]	20.83	0.8050	18.28	0.7824
FBL-SS	**22.25**	**0.8063**	**20.55**	**0.7999**

F. Evaluation on Streak Size Mismatch

To prove the effectiveness of our FBL-SS to handle the streak size mismatch problem between training and testing phases, we evaluate different methods on two testing sets, *Rain100H-S2* and *Rain100H-S3* [64]. The streak sizes of these two sets are twice and three times of those in *Rain100H*, respectively. The testing results are provided

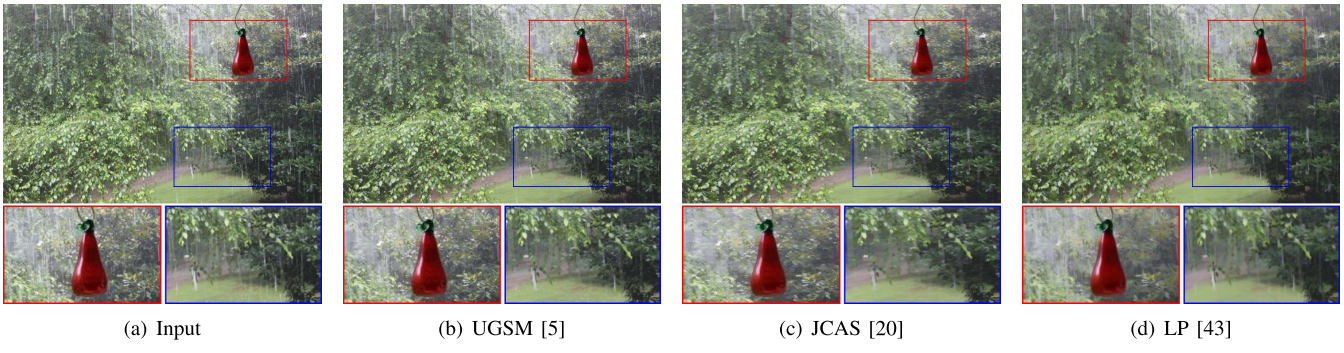


(a) Input (b) UGSM [5] (c) JCAS [20] (d) LP [43]

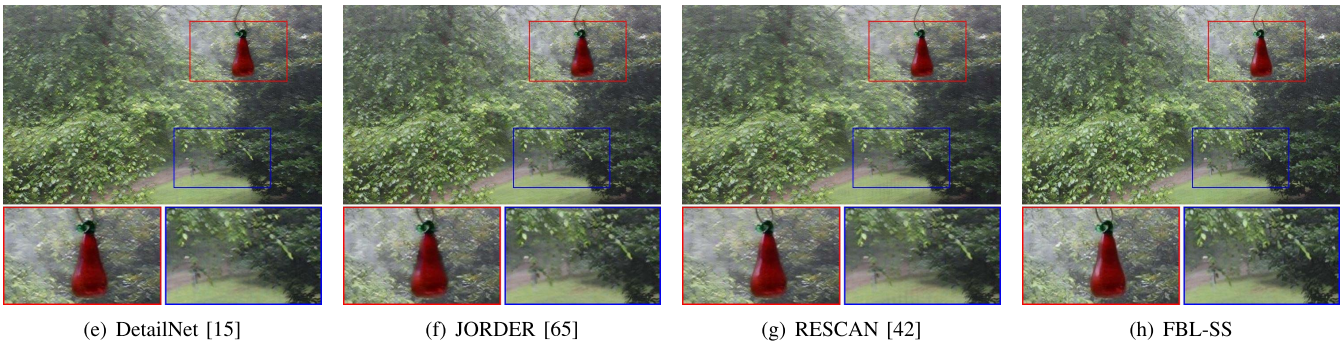


(e) DetailNet [15] (f) JORDER [65] (g) RESCAN [42] (h) FBL-SS

Fig. 6. Visual quality comparison of different rain removal algorithms on a real rain image. It is observed that, our FBL-SS successfully removes background rain streaks. Zooming-in the figure will provide a better look at the restoration quality. Enlarged results are provided in the supplementary material.



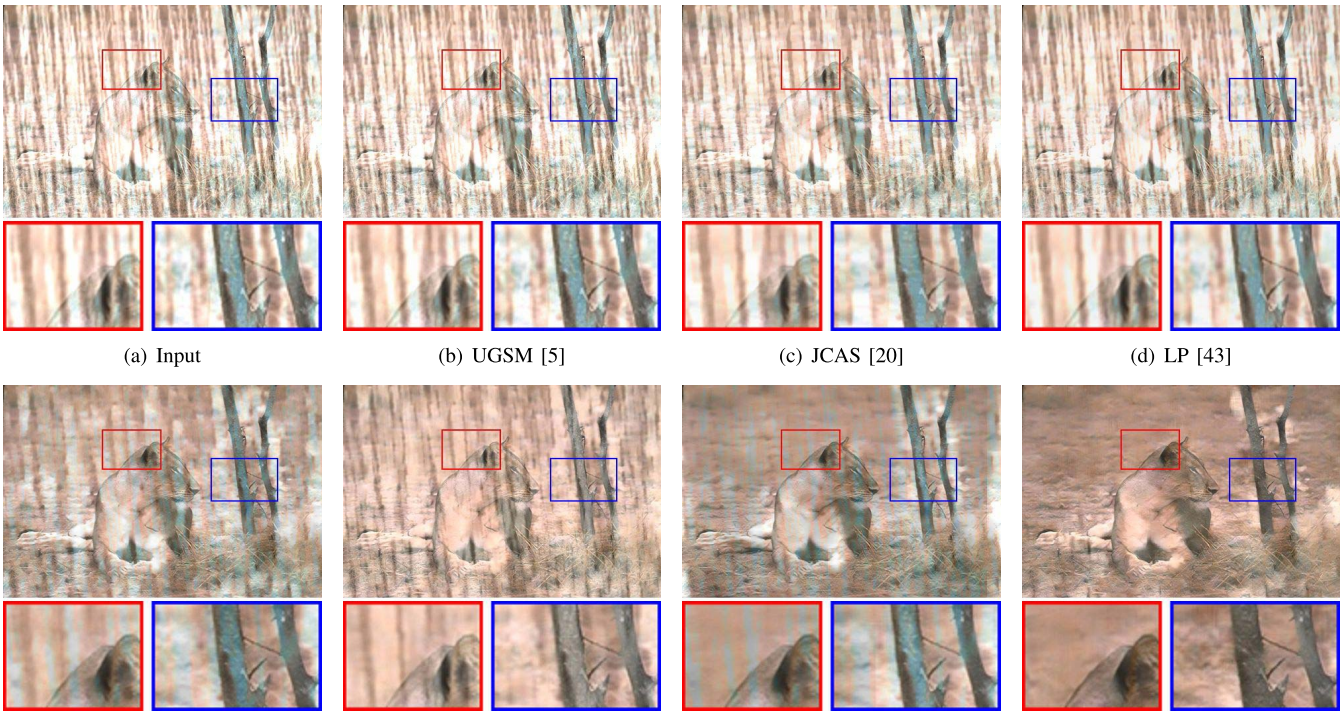
(a) Input (b) UGSM [5] (c) JCAS [20] (d) LP [43]



(e) DetailNet [15] (f) JORDER [65] (g) RESCAN [42] (h) FBL-SS

Fig. 7. Visual quality comparison of different rain removal algorithms on a real rain image. It is observed that, our FBL-SS successfully removes most rain streaks and better preserves texture details. Zooming-in the figure will provide a better look at the restoration quality. Enlarged results are provided in the supplementary material.

in Table V. It is observed that, FBL-SS achieves superior performance than RWL, another state-of-the-art method that also aims to handle varied-scale rain streaks removal. The gain is more than 1.3 dB on *Rain100H-S2* and 2 dB on *Rain100H-S3*. Some visual comparison is provided in Fig. 8.



(a) Input (b) UGSM [5] (c) JCAS [20] (d) LP [43]

(e) DetailNet [15] (f) JORDER [65] (g) RWL [64] (h) FBL-SS

Fig. 8. Visual quality comparison of different rain removal algorithms on a synthetic rain image from *Rain100H-S2*. It is observed that, our FBL-SS successfully removes most rain streaks and better preserves texture details. Zooming-in the figure will provide a better look at the restoration quality.

TABLE VI

THE ABLATION ANALYSIS FOR OUR FBL. R DENOTES REFINEMENT. E&F DENOTES BAND EXPANSION AND FUSION. F DENOTES FRACTAL STRUCTURE. SR DENOTES SCALE-ROBUST CONSTRAINT. THE EVALUATIONS ARE PERFORMED ON *Rain100H*

Methods	R	E&F	F	SR	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	✓	✓	✓	×	29.02	0.8921
FBL-SS	✓	✓	✓	✓	28.29	0.8782

TABLE VII

CROSS DATASET EVALUATION

Test (Train)	JORDER [65]	DetailNet [15]
Rain100L (*Rain100H*)	32.98	32.64
Rain100H (*Rain100L*)	18.31	19.41
Test (Train)	RESCAN [42]	FBL
Rain100L (*Rain100H*)	33.67	34.33
Rain100H (*Rain100L*)	20.13	22.35

G. Ablation Study for Network Structures

We compare five 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 VI. For a fair comparison, the number of the convolution units used in all versions is set to 36. Except for the first convolution, the last convolution, and the convolutions used for band fusion, the channel numbers of all convolutions are set to 64. Thus, these methods have only slight differences in their parameter numbers. It is observed that, each component contributes to the final performance. The comparison between ResNet and SBL demonstrates the effectiveness of band expansion and fusion. The comparisons between ResNet and F-ResNet, SBL and FBL illustrate the importance of fractal structures. Note that, the streak scales of the training and testing data have no domain shift. Therefore, the superiority of the self-robust constraint that helps our model handle more varied scale rain streaks cannot be witnessed in the comparison between FBL and FBL-SS.

H. Ablation Study for Self-Supervision

The scale-robust self-supervision provides the capacity of "scale-robust" rain removal. We perform an ablation analysis on real images in Fig. 9. It is observed that, the results generated by FBL-SS (with self-supervision) are obviously better than FBL (without self-supervision).

I. Cross Dataset Evaluation

We perform cross dataset evaluations where the model is trained on the training set of *Rain100H*/*Rain100L* and test on the testing set of *Rain100L*/*Rain100H*. The PSNR results show the superior generality capacity of FBL compared to previous methods as shown in Table VII. It is demonstrated that, our FBL is capable to not only fit the texture patterns of

TABLE VIII

DETECTION RESULTS (MAP %) ON A SUBSET OF RID DATASET [41] BY FASTER R-CNN [51] AND YOLOV3 [49]. THE BEST RESULTS ARE DENOTED IN BOLD

Method	Metric	Rainy	LPNet [17]	PReNet [50]	SPANet [58]	UMRL [68]	FBL
Faster R-CNN [51]	Bicycle AP	28.57	14.82	20.15	25.59	28.24	**29.85**
	Bus AP	12.29	3.86	2.05	8.45	**13.48**	12.95
	Car AP	26.36	9.96	9.39	25.08	**28.30**	26.58
	Motorbike AP	21.07	13.60	8.60	19.06	20.54	**21.17**
	Person AP	39.46	24.72	21.35	38.94	**41.67**	38.73
	mAP	21.29	11.16	10.26	19.52	**22.04**	21.55
YOLOv3 [49]	Bicycle AP	29.49	16.04	12.12	24.34	29.03	**29.90**
	Bus AP	29.05	5.76	4.84	31.12	28.53	**29.19**
	Car AP	32.67	13.31	16.43	31.10	**33.51**	32.48
	Motorbike AP	19.71	7.75	6.68	20.80	**20.62**	19.61
	Person AP	40.23	22.42	24.81	41.47	41.31	**42.35**
	mAP	25.19	10.88	10.81	24.80	25.50	**25.59**

(a) Input (b) FBL (c) FBL-SS

Fig. 9. Ablation study for self-supervision. FBL performs rain removal by the FBL trained without self-supervision. FBL-SS performs rain removal by the FBL trained with self-supervision.

rain streaks in the given training set but also explore intrinsic useful information for rain streak removal in the general sense.

J. CV Task Evaluation

We then study the object detection performance on the rain input images and the corresponding derained results by different methods, using the state-of-the-art object detection models: Faster R-CNN [51] and YOLOv3 [49]. The mean Average Precision (mAP) results are adopted to compare all algorithms and the pre-trained models on MS COCO are used. A subset of RID dataset [41], whose name start will 'H' and 'W' including 402 images, is used for comparison. The results are shown in Table VIII. It is observed that, LPNet [17], PReNet [50] and SPANet [58] lead to degraded performance. Comparatively, UMRL [68] and the proposed FBL can slightly boost the detection performance. The results demonstrate that, our method is not only capable to remove rain streaks to benefit human vision experience but also can restore high-level semantic information that benefits the machine vision.

V. CONCLUSION

In this paper, we design a fractal band learning network trained with self-supervision to remove rain streaks of varied scales. Our motivations originate from two aspects. First, rain streak removal is highly correlated to frequency domain analysis. Therefore, we build a fractal band learning network to perform frequency band feature operations, which outperforms previous deraining methods on all synthesized paired testing sets. Second, we enforce the learned features of an image at different scales extracted by our FBL consistent after rescaling operations to handle the rain streaks of unseen scales. This constraint effectively benefits removing rain streaks in real images. Extensive experimental results demonstrate the superiority of our method in rain streak removal to previous methods and also show the effectiveness of its each component.

REFERENCES

[1] J. T. Barron, "Convolutional color constancy," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 379–387.
[2] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.
[3] Y. Chang, L. Yan, and S. Zhong, "Transformed low-rank model for line pattern noise removal," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1735–1743.
[4] Y.-L. Chen and C.-T. Hsu, "A generalized low-rank appearance model for spatio-temporally correlated rain streaks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1968–1975.
[5] L.-J. Deng, T.-Z. Huang, X.-L. Zhao, and T.-X. Jiang, "A directional global sparse model for single image rain removal," *Appl. Math. Model.*, vol. 59, pp. 662–679, Jul. 2018.
[6] Z. Deng *et al.*, "Deep multi-model fusion for single-image dehazing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2453–2462.
[7] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.
[8] C. Dong, Y. Deng, C. C. Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 576–584.
[9] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[10] W. Dong, M. Yuan, X. Li, and G. Shi, "Joint demosaicing and denoising with perceptual optimization on a generative adversarial network," 2018, *arXiv:1802.04723*. [Online]. Available: http://arxiv.org/abs/1802.04723
[11] T. Edwards, "Discrete wavelet transforms: Theory and implementation," Stanford Univ., San Francisco, CA, USA, Tech. Rep. 59672516, 1992.

[12] D. Eigen, D. Krishnan, and R. Fergus, "Restoring an image taken through a window covered with dirt or rain," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 633–640.

[13] Y. Fang, C. Zhang, W. Yang, J. Liu, and Z. Guo, "Blind visual quality assessment for image super-resolution by convolutional neural network," *Multimedia Tools Appl.*, vol. 77, no. 22, pp. 29829–29846, Nov. 2018.

[14] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.

[15] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, "Clearing the skies: A deep network architecture for single-image rain removal," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2944–2956, Jun. 2017.

[16] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1715–1723.

[17] X. Fu, B. Liang, Y. Huang, X. Ding, and J. Paisley, "Lightweight pyramid networks for image deraining," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jul. 22, 2019, doi: 10.1109/TNNLS.2019.2926481.

[18] L. Galteri, L. Seidenari, M. Bertini, and A. D. Bimbo, "Deep generative adversarial compression artifact removal," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4836–4845.

[19] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," 2015, *arXiv:1508.06576*. [Online]. Available: http://arxiv.org/abs/1508.06576

[20] S. Gu, D. Meng, W. Zuo, and L. Zhang, "Joint convolutional analysis and synthesis sparse representation for single image layer separation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1717–1725.

[21] T. Guo, X. Li, V. Cherukuri, and V. Monga, "Dense scene information estimation network for dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1–9.

[22] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1664–1673.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[25] X. Hu, C.-W. Fu, L. Zhu, and P.-A. Heng, "Depth-attentional features for single-image rain removal," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8022–8031.

[26] Y. Hu, W. Yang, M. Li, and J. Liu, "Progressive spatial recurrent neural network for intra prediction," *IEEE Trans. Multimedia*, vol. 21, no. 12, pp. 3024–3037, Dec. 2019.

[27] Y. Hu, W. Yang, and J. Liu, "Coarse-to-fine hyper-prior modeling for learned image compression," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, Feb. 2020, pp. 1–8.

[28] D.-A. Huang, L.-W. Kang, Y.-C.-F. Wang, and C.-W. Lin, "Self-learning based image decomposition with applications to single image denoising," *IEEE Trans. Multimedia*, vol. 16, no. 1, pp. 83–93, Jan. 2014.

[29] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[30] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, 2008.

[31] F. Jiang, W. Tao, S. Liu, J. Ren, X. Guo, and D. Zhao, "An end-to-end compression framework based on convolutional neural networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 10, pp. 3007–3018, Oct. 2018.

[32] L.-W. Kang, C.-W. Lin, and Y.-H. Fu, "Automatic single-image-based rain streaks removal via image decomposition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1742–1755, Apr. 2012.

[33] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[34] J.-H. Kim, C. Lee, J.-Y. Sim, and C.-S. Kim, "Single-image deraining using an adaptive nonlocal means filter," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 914–917.

[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2014, pp. 1–15.

[36] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8183–8192.

[37] G. Larsson, M. Maire, and G. Shakhnarovich, "FractalNet: Ultra-deep neural networks without residuals," in *Proc. ICLR*, 2017, pp. 1–11.

[38] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.

[39] C. Li *et al.*, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.

[40] R. Li, L.-F. Cheong, and R. T. Tan, "Single image deraining using scale-aware multi-stage recurrent network," 2017, *arXiv:1712.06830*. [Online]. Available: http://arxiv.org/abs/1712.06830

[41] S. Li *et al.*, "Single image deraining: A comprehensive benchmark analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3838–3847.

[42] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha, "Recurrent squeeze-and-excitation context aggregation net for single image deraining," in *Proc. IEEE Eur. Conf. Comput. Vis.*, Oct. 2018, pp. 254–269.

[43] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, "Rain streak removal using layer priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2736–2744.

[44] J. Lin, D. Liu, H. Yang, H. Li, and F. Wu, "Convolutional neural network-based block up-sampling for HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 12, pp. 3701–3715, Dec. 2019.

[45] J. Liu, S. Xia, and W. Yang, "Deep reference generation with multi-domain hierarchical constraints for inter prediction," *IEEE Trans. Multimedia*, early access, Dec. 23, 2019, doi: 10.1109/TMM.2019.2961504.

[46] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," 2015, *arXiv:1511.03995*. [Online]. Available: http://arxiv.org/abs/1511.03995

[47] Y. Luo, Y. Xu, and H. Ji, "Removing rain from a single image via discriminative sparse coding," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3397–3405.

[48] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2482–2491.

[49] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: http://arxiv.org/abs/1804.02767

[50] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng, "Progressive image deraining networks: A better and simpler baseline," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3932–3941.

[51] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[52] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Learning to deblur," 2014, *arXiv:1406.7444*. [Online]. Available: http://arxiv.org/abs/1406.7444

[53] A. Singh and N. Ahuja, "Super-resolution using sub-band self-similarity," in *Proc. Asia Conf. Comput. Vis.*, D. Cremers, I. Reid, H. Saito, and M.-H. Yang, Eds. Cham, Switzerland: Springer, 2015, pp. 552–568.

[54] S. Song, Y. Li, J. Liu, and Z. Quo, "Joint sub-band based neighbor embedding for image super-resolution," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 1661–1665.

[55] S.-H. Sun, S.-P. Fan, and Y.-C.-F. Wang, "Exploiting image structural similarity for single image rain removal," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 4482–4486.

[56] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4549–4557.

[57] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4809–4817.

[58] T. Wang, X. Yang, K. Xu, S. Chen, Q. Zhang, and R. W. H. Lau, "Spatial attentive single-image deraining with a high quality real rain dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12262–12271.

[59] Y. Wang, S. Liu, C. Chen, and B. Zeng, "A hierarchical approach for rain or snow removing in a single color image," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3936–3950, Aug. 2017.

[60] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[61] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2018, pp. 1–15.

[62] W. Yang, S. Deng, Y. Hu, J. Xing, and J. Liu, "Real-time deep video spatial resolution upconversion system (STRUCT++ Demo)," in *Proc. ACM Multimedia Conf. (MM)*, New York, NY, USA, 2017, pp. 1255–1256.
[63] W. Yang *et al.*, "Deep edge guided recurrent residual learning for image super-resolution," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5895–5907, Dec. 2017.
[64] W. Yang, J. Liu, S. Yang, and Z. Guo, "Scale-free single image deraining via visibility-enhanced recurrent wavelet learning," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2948–2961, Jun. 2019.
[65] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, "Deep joint rain detection and removal from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1685–1694.
[66] W. Yang, R. T. Tan, J. Feng, Z. Guo, S. Yan, and J. Liu, "Joint rain detection and removal from a single image with contextualized deep networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 6, pp. 1377–1393, Jun. 2020.
[67] W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu, "From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 1–8.
[68] R. Yasarla and V. M. Patel, "Uncertainty guided multi-scale residual learning-using a cycle spinning CNN for single image de-raining," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8397–8406.
[69] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5505–5514.
[70] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 695–704.
[71] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Jun. 3, 2019, doi: 10.1109/TCSVT.2019.2920407.
[72] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[73] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *Proc. IEEE Eur. Conf. Computer Vis.*, Jun. 2016, pp. 649–666.
[74] X. Zhang, W. Yang, Y. Hu, and J. Liu, "DMCNN: Dual-domain multi-scale convolutional neural network for compression artifacts removal," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 390–394.
[75] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. IEEE Eur. Conf. Comput. Vis.*, Jul. 2018, pp. 286–301.
[76] Y. Zhang, T. Shen, X. Ji, Y. Zhang, R. Xiong, and Q. Dai, "Residual highway convolutional neural networks for in-loop filtering in HEVC," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3827–3841, Aug. 2018.
[77] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.
[78] L. Zhu, C.-W. Fu, M. S. Brown, and P.-A. Heng, "A non-local low-rank framework for ultrasound speckle reduction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 493–501.
[79] L. Zhu, C.-W. Fu, D. Lischinski, and P.-A. Heng, "Joint bi-layer optimization for single-image rain streak removal," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2545–2553.
[80] L. Zhu, X. Hu, C.-W. Fu, J. Qin, and P.-A. Heng, "Saliency-aware texture smoothing," *IEEE Trans. Vis. Comput. Graphics*, early access, Dec. 21, 2018, doi: 10.1109/TVCG.2018.2889055.

Wenhan Yang (Member, IEEE) received the B.S. and Ph.D. degrees (Hons.) in computer science from Peking University, Beijing, China, in 2012 and 2018, respectively. He was a Visiting Scholar with the National University of Singapore from September 2015 to September 2016 and from September 2018 to November 2018. He is currently a Postdoctoral Research Fellow of the Department of Computer Science, City University of Hong Kong. His current research interests include deep-learning-based image processing, and bad weather restoration and its related applications and theories.

Shiqi Wang (Member, IEEE) received the B.S. degree in computer science from the Harbin Institute of Technology in 2008 and the Ph.D. degree in computer application technology from Peking University in 2014. From 2014 to 2016, he was a Postdoctoral Fellow of the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. From 2016 to 2017, he was a Research Fellow of the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore. He is currently an Assistant Professor with the Department of Computer Science, City University of Hong Kong. He has proposed over 40 technical proposals to ISO/MPEG, ITU-T, and AVS standards, and authored or coauthored more than 150 refereed journal articles/conference papers. His research interests include video compression, image/video quality assessment, and image/video search and analysis. He received the Best Paper Award from PCM 2017, the IEEE Multimedia 2018, and the IEEE ICME 2019, and he has coauthored an article that received the Best Student Paper Award from the IEEE ICIP 2018.

Jiaying Liu (Senior Member, IEEE) received the Ph.D. degree (Hons.) in computer science from Peking University, Beijing, China, in 2010.

She is currently an Associate Professor with the Wangxuan Institute of Computer Technology, Peking University. She has authored over 100 technical articles in refereed journals and proceedings, and holds 42 granted patents. Her current research interests include multimedia signal processing, compression, and computer vision. She was a Visiting Scholar with the University of Southern California, Los Angeles, from 2007 to 2008. She was a Visiting Researcher with Microsoft Research Asia in 2015 supported by the Star Track Young Faculties Award. She has served as a member of the Membership Services Committee, IEEE Signal Processing Society, the Multimedia Systems and Applications Technical Committee (MSA TC), the Visual Signal Processing and Communications Technical Committee (VSPC TC), IEEE Circuits and Systems Society, the Image, Video, and Multimedia (IVM) Technical Committee, APSIPA. She is a Senior Member of the CSIG and CCF. She has also served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, and the *Journal of Visual Communication and Image Representation* (JVCI) (Elsevier), the Technical Program Chair of the IEEE VCIP-2019/ACM ICMR-2021, the Publicity Chair of the IEEE ICME-2020/ICIP-2019, and the Area Chair of CVPR-2021/ECCV-2020/ICCV-2019. She was the APSIPA Distinguished Lecturer (2016–2017).