

# Scale-Free Single Image Deraining Via Visibility-Enhanced Recurrent Wavelet Learning

Wenhan Yang<sup>1</sup>, Member, IEEE, Jiaying Liu<sup>1</sup>, Senior Member, IEEE, Shuai Yang<sup>1</sup>,  
and Zongming Guo<sup>1</sup>, Member, IEEE

**Abstract**—In this paper, we address a rain removal problem from a single image, even in the presence of large rain streaks and rain streak accumulation (where individual streaks cannot be seen and thus are visually similar to mist or fog). For rain streak removal, the mismatch problem between different streak sizes in training and testing phases leads to poor performance, especially when there are large streaks. To mitigate this problem, we embed a hierarchical representation of wavelet transform into a recurrent rain removal process: 1) rain removal on the low-frequency component and 2) recurrent detail recovery on high-frequency components under the guidance of the recovered low-frequency component. Benefiting from the recurrent multi-scale modeling of wavelet transform-like design, the proposed network trained on streaks with one size can adapt to those with larger sizes, which significantly favors real rain streak removal. The dilated residual dense network is used as the basic model of the recurrent recovery process. The network includes multiple paths with different receptive fields, thus it can make full use of multi-scale redundancy and utilize context information in large regions. Furthermore, to handle heavy rain cases where rain streak accumulation is presented, we construct a detail appearing rain accumulation removal to not only improve the visibility but also enhance the details in dark regions. The evaluation of both synthetic and real images, particularly on those containing large rain streaks and heavy accumulation, shows the effectiveness of our novel models, which significantly outperforms the state-of-the-art methods.

**Index Terms**—Single image deraining, recurrent process, wavelet transform, scale-free, residual dense network.

## I. INTRODUCTION

**B**AD weather conditions lead to a series of visibility degradations, which alter the content and color of images. The accompanying detail loss and signal distortion result in the failure of many outdoor computer vision applications, which assume that their inputs are high-quality clean video frames. Rain streaks are one of the most common degradations in rain frames. They cause severe intensity and light fluctuations in

small regions, and therefore obstruct and blur the background scene.

In the past decades, the endeavors of many researchers have been dedicated to rain image restoration. Previous methods [1]–[4] regard single image rain removal as a signal separation problem between rain streaks and background images (rain-free images), based on their texture appearance patterns. Different basic models are employed in these works, such as frequency domain representation [1], sparse representation [4], Gaussian mixture model [5], which differentiate rain streaks and background images.

Recently, there are some new approaches using deep networks to facilitate rain removal. In [6] and [7], 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 [8], a deep network designed for removing heavy rain from images is proposed. The network jointly detects and removes rain streaks, and performs an alternate rain streak and accumulation removal to enhance the visibility in rain scenes.

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

- The degradations of rain scenes in real-world scenarios are very complex. In existing rain models, the diversity of rain sizes is often neglected. Specifically, large streaks are very hard to be synthesized and modeled in the training phase, thus are hard to be removed in the practical scenario. Some previous works [9] try to address the problem by constructing a multi-path network trained with different sizes of rain streaks. However, a general scale-free architecture whose testing performance does not rely on scale patterns of rain streaks in the training set is missing.
- The multi-scale dependency of background images and rain streaks is seldom analyzed and modeled. However, this information provides abundant context clues to infer details lost in the degradation caused by rain streaks. Furthermore, the multi-scale signal analysis provides a convenient access to decomposing the original image into small scales, which facilitates the creation of scale-free rain removal architecture.
- Although some previous works [10]–[12] try to include context information, a general and easily equipped framework for that purpose is absent. Most methods have a limited receptive field, thus it is hard to obtain the context

Manuscript received May 10, 2018; revised October 20, 2018 and December 5, 2018; accepted January 8, 2019. Date of publication January 11, 2019; date of current version April 10, 2019. This work was supported in part by the National Natural Science Foundation of China under Contract 61772043, in part by the Beijing Natural Science Foundation under Contract L182002 and Contract 4192025, and in part by the NVIDIA Corporation through the GPU. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Weisheng Dong. (Corresponding author: Jiaying Liu.)

The authors are with the Institute of Computer Science and Technology, Peking University, Beijing 100080, China (e-mail: yangwenhan@pku.edu.cn; liujiaying@pku.edu.cn; pkuwilliamyang@pku.edu.cn; guozongming@pku.edu.cn).

Digital Object Identifier 10.1109/TIP.2019.2892685

information from large regions. The potential of sharply increasing the receptive field to obtain more surrounding information still needs discussion.

- In a rainy day, clouds in the sky take away most of the light, thus the background layer of the captured rain image is usually in a dim light. Previous rain streak accumulation (where individual streaks cannot be seen, and thus visually similar to mist or fog) removal or dehazing methods do not consider the low light condition, thus the recovered results maybe over-dark and many details are invisible.

Considering these limitations of existing works, we explore possible deep learning architectures that can effectively restore clean images from inputs containing very large rain streaks which do not appear in the training set and heavy rain accumulation, and are flexible to embed multi-scale context information.

Specifically, we embed a hierarchical representation of wavelet transform into a recurrent recovery process. The rain image is decomposed and reconstructed recurrently. After the decomposition, all bands are processed at small scales. Thus, the large streaks that do not appear in the training set become small and at this small scale, their distribution can be covered and simulated by the training set. Hence, the rain streaks in the decomposed bands are easier to be removed. The model trained on streaks with one size can adapt to those with larger sizes. In this way, the architecture of recurrent wavelet learning can realize ‘scale-free’ rain removal and successfully remove large streaks. Then, we use dilated residual dense networks as the basic model of the two steps of the recurrent wavelet learning – 1) rain removal on the low-frequency component; 2) recurrent detail recovery on high-frequency components. The network has a very large receptive field. Its dense connection structure makes full use of the multi-scale redundancy in the image. Thus, more context information is obtained, which facilitates rain removal. Furthermore, apart from rain streaks, we consider degradations including rain streak accumulation under low light conditions. Then, a linear composition recovery model is proposed and a deep network is utilized to learn to not only improve the visibility but also light up the details in dark regions.

Our contributions are summarized as follows,

- A recurrent wavelet learning is built to achieve a scale-free rain streak removal. Benefiting from the recurrent multi-scale modeling of wavelet transform-like design, the proposed network trained on streaks with one size can adapt to those with larger sizes. This can significantly benefit the rain removal on real rain.
- Dense networks are used as the basic model of the recurrent wavelet learning. Its network architecture equivalently has multiple paths with different receptive fields, and thus makes full use of multi-scale redundancy in the image.
- Dilated convolutions are used to construct some dense blocks. With the dilated convolutions, the dense blocks enlarge the receptive field at a faster rate and more context information is obtained.

- The degradations including rain accumulation under low light conditions are looked into, and a deep network based on a linear composition recovery model is developed to improve the visibility and light up the dark details.

The rest of this paper is organized as follows. Section II briefly reviews the related work. Section III presents the recurrent wavelet learning for scale-free rain streak removal. The specific network structures – dilated residual dense networks – to model the two steps of the recurrent wavelet learning are introduced in Section IV. Section V considers rain streak accumulation removal in low light conditions. Experimental results and concluding remarks are presented in Sections VI and VII, respectively.

## II. RELATED WORKS

### A. Single Image Rain Removal

Single image deraining is a highly ill-posed problem and is addressed by a signal separation or texture classification route. Kang *et al.* [1] attempted to separate rain streaks from the high frequency layer by sparse coding. Then, a generalized low rank model [13] was proposed, where the rain streak layer is assumed to be low rank. Kim *et al.* [14] first detected rain streaks and then removed them with the non-local mean filter. Luo *et al.* [4] proposed a discriminative sparse coding method to separate rain streaks from background images. Li *et al.* [5] exploited Gaussian mixture models to separate the rain streaks. The presence of deep learning promotes the development of single image deraining. In [6], a deep network that takes the image detail layer as its input and predicts the negative residues was constructed. It has a good capacity to keep texture details. But it cannot handle heavy rain cases where rain streaks are dense. In [8], a deep joint rain detection and removal method was proposed to recurrently remove rain streaks and accumulation, obtaining impressive results in heavy rain cases. However, rain streaks and textures of the background are intrinsically overlapped in the feature space. Thus, the remaining weak streaks or over-smoothed background textures are usually presented in the results. In [9], to treat the rain streaks differently, several parallel sub-networks are trained with different scales of rain streaks. However, it is not easy to synthesize large rain streaks, and the testing performance is still bounded by the streak sizes used in the training phase. Zhang and Patel [15] proposed a joint rain estimation and removal method. The network automatically determines the rain-density information and then efficiently removes the corresponding rain-streaks guided by the estimated rain-density label. In [16], to exploit the directional characteristic, a transformation is incorporated into the image decomposition model to facilitate mapping the image signal and imposing a low rank prior. In [17], the rain streak removal is designed as an alternating process of removing rain streaks from a single input image, and removing normal texture details from the estimated rain streak layer into a rain-free background layer. Gu *et al.* [18] proposed a joint convolution analysis and synthesis sparse representation model, where a single image is decomposed into two layers which represent large-scale image structures and fine-scale image textures, respectively. In [19],

image decomposition and dictionary learning are integrated and a 3-layer hierarchical scheme is used to remove both rains and snows. Following [6], [8], [9], and [15], our work also focuses on deep-learning based single image rain streak removal. Differently, we construct a scale-free architecture for rain streak removal. Namely, the testing performance does not rely on the streak sizes used in the training phase. Compared with [15], which removes rains with the awareness of different densities, our work deals with the problem of streak size mismatch. In [15], the density is injected into the network as an explicit predictor. In our work, the domain shift in streak sizes between training and testing phases is implicitly addressed by a recurrent network inspired by unrolling the wavelet transform into a recurrent process. Furthermore, the rain accumulation in low light conditions are looked into and a deep network is developed for the corresponding inverse restoration.

### B. Deep Learning in Image Processing

In recent years, deep learning-based image processing applications emerged with promising performance. These applications include denoising [20], super-resolution [21]–[28], deblurring [29], and style transfer [30], *etc.* There are also some recent works on bad weather restoration or image enhancement, such as dehazing [31], raindrop and dirt removal [32], light enhancement [33]–[35] and rain removal [6], [36], deblocking [37]. Besides, with the superior modeling capacity than shallow models, deep learning-based methods begin to solve harder problems, such as blind image denoising [20], image compression [38], quality assessment [39], and video coding [40]–[43].

The rise of the ResNet and DenseNet further advances the development of the related tasks. It has been observed that refining features progressively like ResNet [44] or concatenating and fusing features from different levels like DenseNet [45] leads to better representations of pixels and their contexts for low-level visions. The related beneficial tasks include super-resolution [22], [46]–[49], rain removal [8], [15], dehazing [50], inpainting [51], and compression artifacts removal [52], deblurring [53], *etc.* In [22], the connection between ResNet and traditional band filter recovery is presented, and edge information is embedded into the ResNet as the priors for a more accurate high-frequency detail inference. In [46], [51]–[53], ResNet is used as the generator of generative adversarial networks. Yang *et al.* [8] extended the baseline ResNet to a contextualized dilated network. In each block, the output features aggregate the representations of the three convolution paths with different dilated factors, which expands the receptive fields effectively and well preserve local details. In [15], a multi-stream densely connected de-raining network is proposed to efficiently leverage features from different scales. In [50], a new edge-preserving densely connected encoder-decoder structure with multi-level pyramid pooling module is exploited to estimate the transmission map for further dehazing. In [48], Tai *et al.* combined recursive units, gate units and the densely connected structure to simulate the mechanism of the human persistent memory for image restoration. Zhang *et al.* [47] combined the structure of ResNet

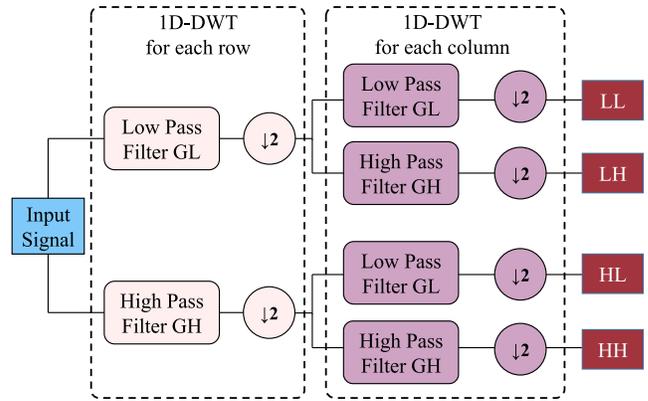


Fig. 1. Illustration of discrete wavelet transform (DWT).

and DenseNet. Dense blocks are used to obtain dense local features. All features in each dense block are connected by skip connections, and then are fused in the last layer adaptively in a holistic way. In [49], Zhang *et al.* utilized residual in residual to construct deep networks. The network consists of several residual groups, which further contains some residual blocks. In our work, following [15], [47], and [50], we also explore to construct a more effective network for low-level visions. We integrate residual networks, dense connectedly networks and dilated convolutions into a unified dilated residual dense network, where the information flow is obtained from different levels and their interdependencies are maximized to better infer the clean background image.

### C. Wavelet Transformation-Based Image Processing

Many wavelet-based methods have already been proposed for low-level image processing problems. The general flowchart of discrete wavelet transform is provided in Fig. 1. Many works focus on video super-resolution [54]. A sequence of low-resolution images is utilized to infer the information of a high-resolution image. There are also lots of works on single-image super-resolution, including interpolation-based [55] and statistic-based methods [56]. In [55], a modified version of classical wavelet-based interpolation method was proposed. In [57], a hybrid wavelet convolution network was presented. A set of sparse coding candidates is encoded by wavelet transform and a convolution network is employed for sparse coding. In [58], the wavelet transform is adopted for separating the variations of data at different scales. Some works focus on wavelet transform-based denoising [59] and deblocking [60]. There is also a work utilizing one-layer wavelet transform [61] for rain removal. Some recent works borrow from Laplacian pyramid structures and craft a progressive up-sampling method for image generation [62] and super-resolution [26], [63]. The first one cascades several convolutional networks within a Laplacian pyramid framework to generate images in a coarse-to-fine fashion, leading to a superior image generation performance. The second one reconstructs the sub-band residuals of high-frequency signals progressively and provides high-quality super-resolution results with a low computational complexity. Compared to Laplacian pyramid-based deep learning methods [26], [63], our approach includes both progressive down-sampling and up-sampling processes. After the down-sampling

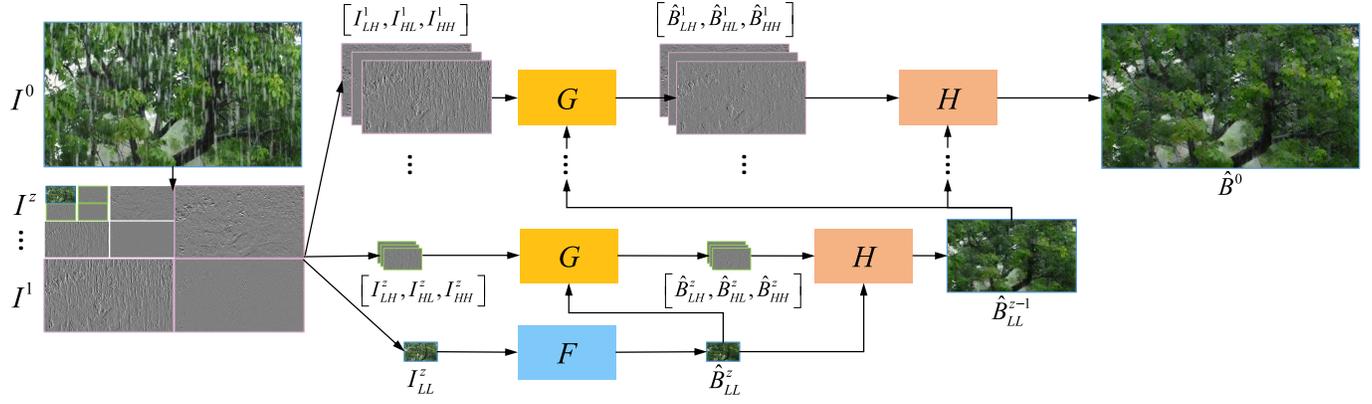


Fig. 2. The framework of our proposed recurrent wavelet learning for rain streak removal.

operation, the spatial resolution of the features is reduced, then the successive operations become more light-weighted and have a higher computational efficiency. Compared to all previous wavelet / Laplacian transform-based methods, we adopt a recurrent wavelet learning for scale-free rain streak removal, and the dilated dense network that makes use of multi-scale redundancy is constructed. Besides projecting the signal into a space where learning the mapping from input and target images more easily, we specially pay attention to unrolling the wavelet transform into an recurrent network, which decomposes the rain image into a small scale and performs the rain removal more effectively.

### III. RECURRENT WAVELET LEARNING FOR SCALE-FREE RAIN STREAK REMOVAL

In this section, we aim to address the problem of the diversity of rain streaks in sizes and densities, and develop a scale-free rain streak removal framework. Our overall network architecture is shown in Fig. 2. The rain image ( $I^0$ ) is first decomposed into different components ( $I^1 = \{I^1_{LH}, I^1_{HL}, I^1_{HH}\}$ ,  $I^2 = \{I^2_{LH}, I^2_{HL}, I^2_{HH}\}, \dots, I^z = \{I^z_{LL}, I^z_{LH}, I^z_{HL}, I^z_{HH}\}$ ) by wavelet transformations. Then, we perform rain removal (the process  $F$ ) for the low-frequency average components, *i.e.*  $I^z_{LL}$ , in which scale rain streaks can be better recognized and removed by the network. Then, guided by the low-frequency rain removal result, the residual rain streaks in high-frequency bands are removed (the process  $G$ ). After that, the estimated background low-frequency and high-frequency bands are combined by inverse wavelet transformation (the process  $H$ ). The decomposition and reconstruction processes can be performed recurrently. After the decomposition, the spatial resolutions of all bands are shrunk. Therefore, the images are equal to being processed at small scales. Large streaks in the original rain image become small and are easier to be removed. Thus, the architecture of recurrent wavelet learning can realize ‘scale-free’ rain removal. Here, we only briefly present overall architecture of recurrent wavelet learning. The specific network structures and configurations of  $F$  and  $G$  will be introduced in Section IV.

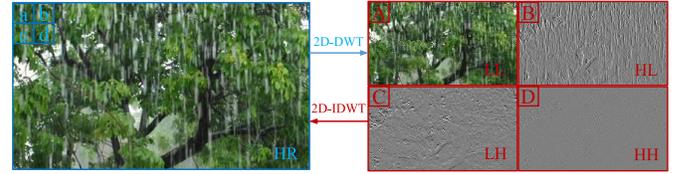


Fig. 3. Illustration for 2D-DWT and 2D-IDWT. The four pixels in each  $2 \times 2$  patch are utilized to calculate coefficient values in four band maps.

#### A. Wavelet Decomposition

Our method is built on wavelet transform. The transform decomposes an image into several sub-images of the same size containing wavelet coefficients. The Haar wavelet is employed to depict different-frequency facial information. The 2D fast wavelet transform (FWT) [64] is used to calculate Haar wavelets. The calculation process is shown in Fig. 1. Equivalently, as shown in Fig. 3, the Haar wavelet coefficients at one level are computed as follows,

$$\begin{aligned} A &= (a + b + c + d)/4, \\ B &= (a - b + c - d)/4, \\ C &= (a + b - c - d)/4, \\ D &= (a - b - c + d)/4, \end{aligned}$$

where  $a, b, c, d$  are four pixels in every  $2 \times 2$  block. The height and width of the decomposed bands are half of the original image signal. It is observed from Fig. 3 that, the main structure component of the signal is in  $A$ .  $B, C$  and  $D$  contain high-frequency details. The wavelet bands at different scales are calculated by performing the decomposition recurrently in Fig. 1. We use  $E$  to denote the wavelet decomposition process. Then, the decomposition of the original image and low-frequency band is formulated as follows,

$$\begin{aligned} [I^1_{LL}, I^1_{LH}, I^1_{HL}, I^1_{HH}] &= E(I^0), \\ [I^2_{LL}, I^2_{LH}, I^2_{HL}, I^2_{HH}] &= E(I^1_{LL}), \\ &\dots \\ [I^{j+1}_{LL}, I^{j+1}_{LH}, I^{j+1}_{HL}, I^{j+1}_{HH}] &= E(I^j_{LL}), \\ &\dots \\ [I^z_{LL}, I^z_{LH}, I^z_{HL}, I^z_{HH}] &= E(I^{z-1}_{LL}), \end{aligned} \quad (1)$$

where  $z$  is the order number of the decomposition.

### B. Recurrent Wavelet Learning

After the decomposition, a rain image turns to many wavelet bands. Then, we need to build two mappings  $F$  and  $G$ , which will be illustrated as follows, for the restoration of these bands to get rain-free bands. First, for the top low frequency rain signal, *i.e.*  $I_{LL}^2$  in the Fig. 2, a process  $F$  is adopted to map  $I_{LL}^j$  into the rain-free one  $\hat{B}_{LL}^j$ , denoted as

$$\hat{B}_{LL}^z = F(I_{LL}^z). \quad (2)$$

After performing rain removal on the top low frequency band, the predicted main structure of rain-free image  $\hat{B}_{LL}^z$  is obtained. Then, under the guidance of  $\hat{B}_{LL}^z$ , we perform rain removal on high-frequency components. We use  $G$  to signify this process,

$$\left[ \hat{B}_{LH}^z, \hat{B}_{HL}^z, \hat{B}_{HH}^z \right] = G([\hat{B}_{LL}^z, I_{LH}^z, I_{HL}^z, I_{HH}^z]). \quad (3)$$

After recovering the high-frequency components, the wavelet bands at a small scale can be combined into the low-frequency band at a larger scale. We use  $H$  to signify this process,

$$\hat{B}_{LL}^{z-1} = H([\hat{B}_{LL}^z, \hat{B}_{LH}^z, \hat{B}_{HL}^z, \hat{B}_{HH}^z]). \quad (4)$$

Specifically,  $H$  is the 2D-IDWT transform, as shown in Fig. 3, the inverse Haar wavelet transform at one level is calculated as follows,

$$\begin{aligned} a &= A + B + C + D, \\ b &= A - B + C - D, \\ c &= A + B - C - D, \\ d &= A - B - C + D. \end{aligned}$$

The height and width of the reconstructed bands are twice those of the input signals.

With Eqs. (3) and (4), we obtain the recovery of the whole rain-free image as follows,

$$\begin{aligned} \left[ \hat{B}_{LH}^z, \hat{B}_{HL}^z, \hat{B}_{HH}^z \right] &= G([\hat{B}_{LL}^z, I_{LH}^z, I_{HL}^z, I_{HH}^z]), \\ \hat{B}_{LL}^{z-1} &= H([\hat{B}_{LL}^z, \hat{B}_{LH}^z, \hat{B}_{HL}^z, \hat{B}_{HH}^z]). \\ \left[ \hat{B}_{LH}^{z-1}, \hat{B}_{HL}^{z-1}, \hat{B}_{HH}^{z-1} \right] &= G([\hat{B}_{LL}^{z-1}, I_{LH}^{z-1}, I_{HL}^{z-1}, I_{HH}^{z-1}]), \\ \hat{B}_{LL}^{z-2} &= H([\hat{B}_{LL}^{z-1}, \hat{B}_{LH}^{z-1}, \hat{B}_{HL}^{z-1}, \hat{B}_{HH}^{z-1}]). \\ &\dots \\ \hat{B}_0 &= H([\hat{B}_{LL}^1, \hat{B}_{LH}^1, \hat{B}_{HL}^1, \hat{B}_{HH}^1]). \end{aligned}$$

### IV. DILATED RESIDUAL DENSE NETWORK

In the last section, we propose a recurrent wavelet learning for single image rain removal.  $F$  and  $G$  are used to signify the process of removing rain streaks from the top low frequency component and inferring high-frequency details under the guidance of the recovered low frequency component, respectively.

Here, we construct dilated residual dense networks to model  $F$  and  $G$ . We also use this network as the basic model of the detail appearing rain accumulation removal method, which is illustrated in V-B. Our motivation and network design methodology are briefly presented as follows. First, we aim to

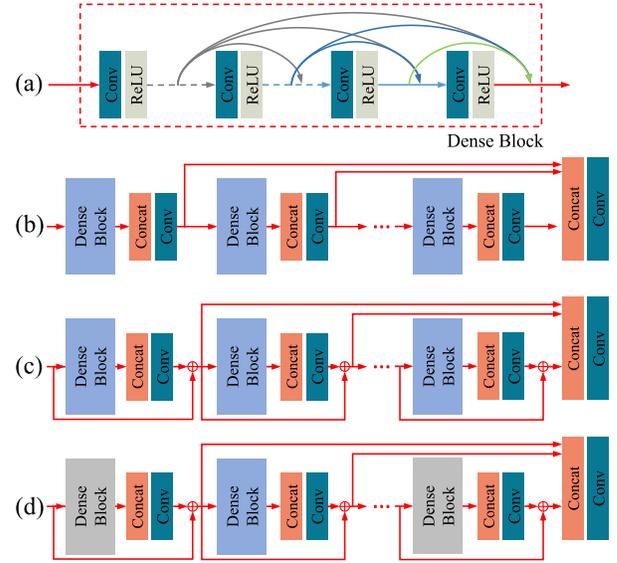


Fig. 4. Network architecture: (a) Dense network (Dense Block). (b) Dense network with progressive channel compression. (c) Dense network with progressive channel compression and residual learning. (d) Some Dense Blocks are composed of dilated convolutions (D-Dense Block), which further enlarge the receptive field of the network.

model the multi-scale dependency in the image, thus a network with dense connections, with different receptive fields, is built. Second, to avoid excessive growth of parameters, a progressive channel compression scheme is used. Third, after the channel compression, one dense block is only connected to the successive one. To make them connect with each other to facilitate the local redundancy modeling and utilize features across different levels, residual learning is also utilized in our network to further improve the network representation capacity. Fourth, to enlarge the receptive field of the network, parts of the convolutions in the dense network are dilated convolutions.

#### A. Dense Network (Dense Block)

A preliminary network architecture to model the multi-scale dependency is the dense network as shown in Fig. 4 (a). The connections of the network have different receptive fields, thus making use of the information at different scales. The output of  $c$ -th convolution layer is formulated as

$$F_c = \sigma(W_c[F_1, F_2, \dots, F_{c-1}]), \quad (5)$$

where  $\sigma$  is the ReLU activation function.  $W_c$  is the weights of the  $c$ -th convolution layer.  $[F_1, F_2, \dots, F_{c-1}]$  is the concatenation of output feature maps of preceding  $(c-1)$  convolutions, generating a  $(c-1) \times G$  channel feature maps.  $G$  is the increasing channel number of each convolution. Note that, the outputs of the preceding layers have direct connections to all subsequent layers. These outputs are from the layers having different receptive fields, thus extract useful dense features at different scales.

#### B. Progressive Channel Compression

The network in Fig. 4 (a) may face the problem of parameter explosion. The  $c$ -th convolution layer will generate  $(c-1) \times G$

channel feature maps. To make the generated feature maps more compact, we add channel compression mechanism in the network, as shown in Fig. 4 (b). The whole network is splitted into several dense blocks, each of which is fed into a concatenation layer and a  $1 \times 1$  convolution for channel compression. Assume the output feature map after the  $d$ -th dense block as well as the the following concatenation and convolution layer is  $F^d$ , then

$$F^d = \sigma(W_c^d[F^{d-1}, F_1^d, F_2^d, \dots, F_{c-1}^d]), \quad (6)$$

where  $F_j^d$  is the output feature map of the  $j$ -th convolution in the  $d$ -th dense block. Suppose this process is denoted by  $X^d$ , then the output feature map of the  $(d+1)$ -th dense block is

$$\begin{aligned} F^{(d+1)} &= X^{(d+1)}(F_d), \\ &= X^{(d+1)}(X^d(\dots(X^0(F_0))\dots)), \end{aligned} \quad (7)$$

where  $F_0$  is the feature map extracted by the first convolution from the input image.

After all  $N$  dense blocks, all output features from each dense block are concatenated and then go through a convolution layer to generate the final feature  $R$  as follow,

$$R = W_R[F^0, F^1, F^2, \dots, F^{(N-1)}, F^N]. \quad (8)$$

### C. Residual Dense Block

To alleviate the vanishing-gradient problem in training the dense network in Fig. 4 (b), we add residual connections between different dense blocks as shown in Fig. 4 (c). With the residual connections, Eq. (6) turns to

$$F^d = \sigma(W_c^d[F^{d-1}, F_1^d, F_2^d, \dots, F_{c-1}^d]) + F^{d-1}. \quad (9)$$

### D. Dilated Residual Dense Block (D-Dense Block)

To enlarge the receptive field of the network, we use dilated convolutions to construct some blocks. The dilated convolution [65] weights pixels with a step size of a dilated factor, and thus increases its receptive field without losing pixel resolution accuracy. The specific network configuration of the dilated residual dense block is shown in Table I. In each residual dense block, we use six convolution layers. In the dilated residual dense block, the dilated factors of some convolutions are larger than 1, thus the receptive field of network will expand rapidly. The receptive field of Dense Block is  $13 \times 13$  and that of D-Dense Block is  $25 \times 25$ . The enlarged receptive field makes network capable of obtaining more context information, which facilitates rain removal.

### E. Implementation Details

In the recurrent wavelet learning, the order of wavelet transform is set to 2.  $F$  and  $G$  are modeled by two dilated residual dense networks. Four blocks (including two Dense Blocks and D-Dense Blocks) are used in each network. Their arrange order is Dense Block, D-Dense Block, Dense Block and D-Dense Block sequentially. The channel of the output feature map of the first convolution and that of each compression convolution are set to 64, and the kernel size of them is  $1 \times 1$ . The kernel size of other convolutions is  $3 \times 3$ . The convolution layer

TABLE I  
THE RECEPTIVE FIELD ANALYSIS OF DENSE BLOCK AND D-DENSE BLOCK

Dense Block			
#Conv	Kernel Size	Dilated factor	Receptive field
1	3×3	1×1	3×3
2	3×3	1×1	5×5
3	3×3	1×1	7×7
4	3×3	1×1	9×9
5	3×3	1×1	11×11
6	3×3	1×1	13×13
D-Dense Block			
#Conv	Kernel Size	Dilated factor	Receptive field
1	3×3	1×1	3×3
2	3×3	2×2	7×7
3	3×3	1×1	9×9
4	3×3	2×2	13×13
5	3×3	1×1	15×15
6	3×3	4×4	25×25

number in each block is set to 6. And the output number  $G$  of these convolutions is set to 32. The specific setting of the D-Dense Block is provided in Table I.

### F. 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 decompose them using wavelet transformation and obtain the first order decomposition results  $\{y_i^{LL}, y_i^{LH}, y_i^{HL}, y_i^{HH}\}_{i=1, \dots, N}$  and  $\{x_i^{LL}, x_i^{LH}, x_i^{HL}, x_i^{HH}\}_{i=1, \dots, N}$ .  $y_i^{LL}, y_i^{LH}, y_i^{HL}, y_i^{HH}$  are sub-bands containing wavelet coefficients for average, vertical, horizontal and diagonal details of the rain image.  $x_i^{LL}, x_i^{LH}, x_i^{HL}, x_i^{HH}$  are sub-bands containing wavelet coefficients for average, vertical, horizontal and diagonal details of the rain-free image.

$F$  is directly trained based on the rain and rain-free image  $\{y_i, x_i\}_{i=1, \dots, N}$ . Let  $\Theta_F$  collect all parameters in  $F$ . We adopt mean squared error to train  $F$ :

$$L(\Theta_F) = \frac{1}{2N} \sum_{i=1}^N \|F(y_i; \Theta_F) + y_i - x_i\|^2. \quad (10)$$

After training  $F$ , we then train  $G$  based on the decomposition results. Let  $\Theta_G$  collect all parameters in  $G$ . The predicted rain-free average low frequency component by  $F$  is denoted as  $\hat{x}_i^{LL}$ . The input and output vector are represented as

$$\begin{aligned} \mathbf{y}_i &= \{\hat{x}_i^{LL}, y_i^{LH}, y_i^{HL}, y_i^{HH}\}, \\ \mathbf{x}_i &= \{x_i^{LL}, x_i^{LH}, x_i^{HL}, x_i^{HH}\}. \end{aligned}$$

We also adopt mean squared error to train  $G$ :

$$L(\Theta_G) = \frac{1}{2N} \sum_{i=1}^N \|G(\mathbf{y}_i; \Theta_G) + \mathbf{y}_i - \mathbf{x}_i\|^2. \quad (11)$$

## V. DETAIL APPEARING RAIN ACCUMULATION REMOVAL

To handle heavy rain cases including rain streak accumulation, in this section, the degradation model including rain streak accumulation under low light conditions is illustrated.

Then, a linear composition recovery model is developed, and the related training loss is presented.

#### A. Rain Streak Accumulation in Low Light Conditions

The widely used rain model [4], [5], [10] considers a rain image as a linear combination of the background image and rain streak, which is expressed as:

$$\mathbf{O} = \mathbf{B} + \mathbf{S}, \quad (12)$$

where  $\mathbf{B}$  is the background layer, and  $\mathbf{S}$  is the rain streak layer.  $\mathbf{O}$  is the input image with rain streaks.

In the real world, rain appearance is not only formed by individual rain streaks, but also by accumulation of rain streaks. When rain accumulation is dense, the individual streaks cannot be observed clearly. This rain streak accumulation, whose visual effect is similar to mist or fog, causes the atmospheric veiling effect as well as blur, especially for distant scenes. Besides, in a rainy day, clouds in the sky takes away most of the light, thus the background layer of the captured rain image is usually in a dim light.

To accommodate these two phenomena (*i.e.*, rain streak accumulation and low light degradation), a new rain model is proposed. The model comprises of a background or rain streak contaminated layer that goes through a Gamma transformation. It also includes the appearance of rain accumulation, by relying on the Koschmieder model that is approximately applicable to many turbid media, including mist, fog (*e.g.* [66]) and underwater (*e.g.* [67], [68]). The new rain model is expressed as:

$$\mathbf{R} = \alpha \odot \mathbf{I}^\gamma + (1 - \alpha) \odot \mathbf{A}, \quad (13)$$

where  $\mathbf{I}$  is the rain-free background layer  $\mathbf{B}$  or a rain streak contaminated layer  $\mathbf{O}$ .  $\mathbf{A} \in [0, 1]$  is the global atmospheric light, and  $\alpha \in [0, 1]$  is the atmospheric transmission.  $\gamma > 1$  is an decoding Gamma, and  $\mathbf{I}^\gamma$  is the gamma expansion of  $\mathbf{I}$ .

#### B. Solution: A Linear Composition Recovery Model

Based on Eq. (13), the inverse recovery is to obtain  $\mathbf{I}$  given  $\mathbf{R}$ . This process jointly considers the recovery of the degradation with rain streak accumulation and low light condition. Thus, it can not only improve the visibility but also make details in dark regions visible.

The main difficulty of the task is the complex nonlinearity in the degradation. Directly solving Eq. (13) leads to the following solution:

$$\mathbf{I} = \left[ \frac{(\mathbf{R} - (1 - \alpha) \odot \mathbf{A})}{\alpha} \right]^{\frac{1}{\gamma}}. \quad (14)$$

The division and exponential functions in Eq. (14) cause two problems: 1) the nonlinearity changes the form of solutions, thus the good estimations of  $\alpha$  and  $\gamma$  cannot guarantee a good estimator of  $\mathbf{I}$ ; 2)  $\alpha$  and  $\gamma$  in some intervals will have a great influence on  $\mathbf{I}$ , thus the inaccurate estimations of  $\alpha$  and  $\gamma$  may cause artifacts in the visual results.

To address these issues, we generalize Eq. (13) into a linear composition model as follows,

$$\begin{aligned} \mathbf{R} &= \alpha \odot \mathbf{I}^\gamma + (1 - \alpha) \odot \mathbf{A}, \\ &= \alpha \odot (\mathbf{I} - \Delta \mathbf{I}^{(\gamma)}) + (1 - \alpha) \odot \mathbf{A}, \end{aligned} \quad (15)$$

where  $\Delta \mathbf{I}^{(\gamma)}$  is the change of  $\mathbf{I}$  due to Gamma transformation. Then, Eq. (15) can be transformed into

$$\mathbf{I} = (\mathbf{R} - x) \odot y + z, \quad (16)$$

where  $x = (1 - \alpha) \odot \mathbf{A}$ ,  $y = 1/\alpha$  and  $z = \Delta \mathbf{I}^{(\gamma)}$ . Then, the estimation  $\hat{\mathbf{I}}$  can be obtained by first estimating  $x$ ,  $y$  and  $z$  based on  $\mathbf{R}$ , then followed by linear operations. This change makes estimating  $x$ ,  $y$  and  $z$  consistent with estimating  $\mathbf{I}$ , thus the above-mentioned two issues are mitigated. In our work, we use the proposed dilated residual dense network (illustrated in Section IV) to learn the mappings between  $\mathbf{R}$  and  $x$ ,  $y$ ,  $z$ .

#### C. Training Loss

Suppose we have a collection of paired patches  $\{h_i, f_i\}_{i=1, \dots, M}$  with and without rain streak accumulation, where  $M$  is the total number of training patches. These patches can be synthesized by the degradation model of Eq. (13). We follow two assumptions [31] to synthesize the training pair: 1) image content is uncorrelated with medium transmission and Gamma; 2) medium transmission and Gamma is locally constant. Thus, an arbitrary transmission  $\alpha$  and Gamma  $\gamma$  can be employ on an individual image patch. The transmission and Gamma is sampled uniformly from  $[0, 1]$  and  $[1, 1.5]$ , respectively. The atmospheric light  $\mathbf{A}$  is set to 1. Then, based on the sampled transmission, Gamma and atmospheric light, we can convert  $\{h_i, f_i\}_{i=1, \dots, M}$  to  $\{h_i, xx_i, yy_i, zz_i\}_{i=1, \dots, M}$  based on Eq. (16), where  $h_i$ ,  $f_i$ ,  $xx_i$ ,  $yy_i$  and  $zz_i$  are the training instances that correspond to  $\mathbf{R}$ ,  $\mathbf{I}$ ,  $x$ ,  $y$  and  $z$  in Eq. (16). Let  $H_x$ ,  $H_y$  and  $H_z$  denote the processes of estimating  $x$ ,  $y$  and  $z$  based on  $\mathbf{R}$ , respectively. Let  $\Theta_{H_x}$ ,  $\Theta_{H_y}$  and  $\Theta_{H_z}$  collect all parameters in  $H_x$ ,  $H_y$  and  $H_z$ , respectively. We adopt mean squared error to train them:

$$\begin{aligned} L(\Theta_H) &= \frac{1}{2M} \sum_{i=1}^M \left( \|H_x(h_i; \Theta_{H_x}) - xx_i\|^2 \right. \\ &\quad \left. + \|H_y(h_i; \Theta_{H_y}) - yy_i\|^2 \right. \\ &\quad \left. + \|H_z(h_i; \Theta_{H_z}) - zz_i\|^2 \right). \end{aligned} \quad (17)$$

## VI. EXPERIMENTS

#### A. Dataset

We compare our method with state-of-the-art methods on a few benchmark datasets: (1) *Rain12<sup>1</sup>* [5], which includes 12 synthesized rain images with only one type of rain streaks; (2) *Rain100L<sup>2</sup>* [8], which is the synthesized data set with only one type of rain streaks; (3) *Rain100H*, which is our synthesized data set with five streak directions; (4) *Rain20H*, a random subset of *Rain100H*, used for ablation analysis.

<sup>1</sup><http://yu-li.github.io/>

<sup>2</sup>[http://www.icst.pku.edu.cn/struct/Projects/joint\\_rain\\_removal.html](http://www.icst.pku.edu.cn/struct/Projects/joint_rain_removal.html)

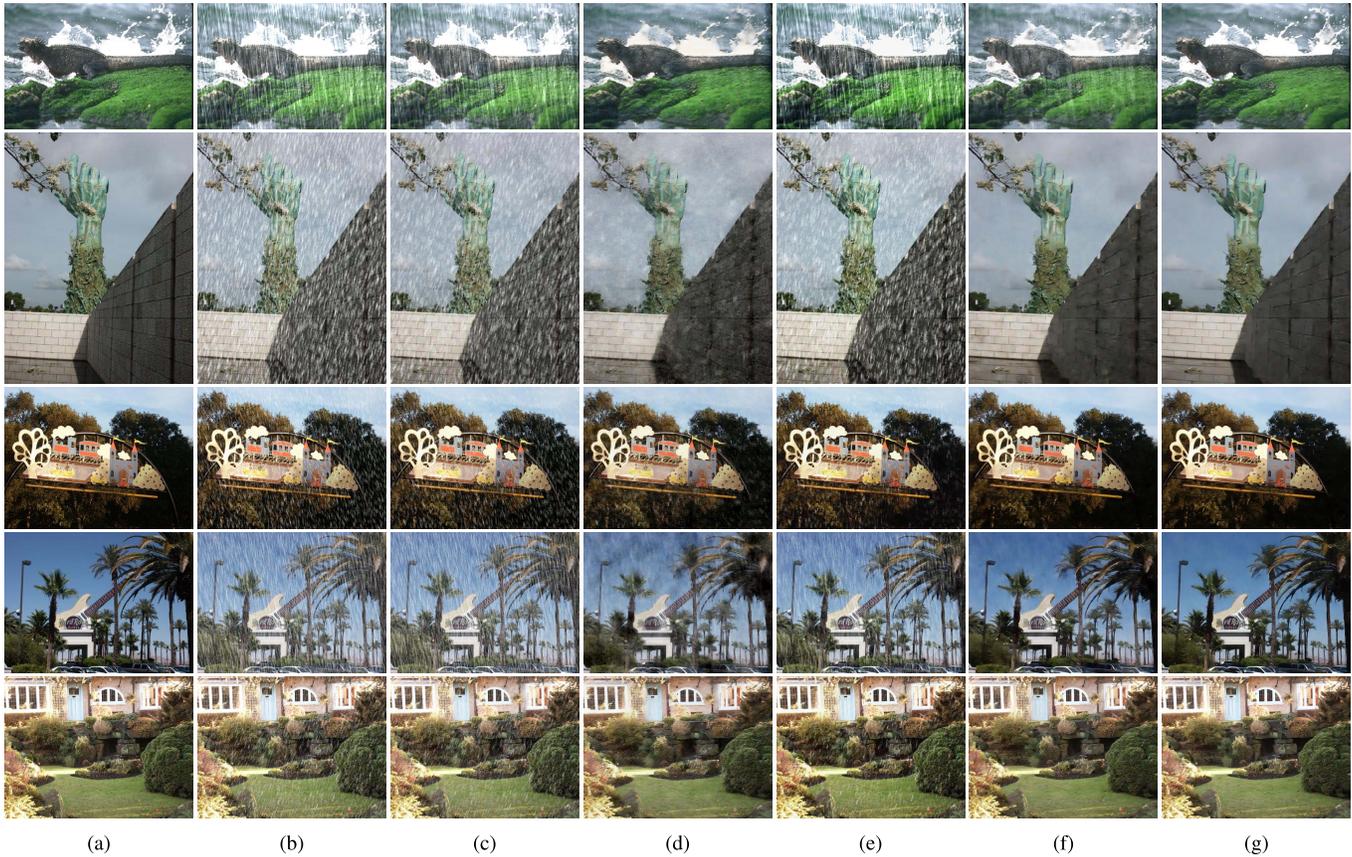


Fig. 5. Visual comparison of our RWL with state-of-the-art rain removal algorithms on synthetic rain images. It is observed that, our RWL is better at removing rain streaks and restoring details. The top panel: testing images from *Rain100H*. Others: testing images from *Rain800*. (a) Ground Truth. (b) Input. (c) UGSM. (d) DetailNet. (e) DID-MDN. (f) JORDER. (g) RWL.

(5) *Rain100H-S2* and *Rain100-S3*, 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.

### B. Baseline Methods

We compare our rain streak removal method with seven state-of-the-art methods: image decomposition (ID) [1], CNN-based rain drop removal (CNN) [32], discriminative sparse coding (DSC) [4], layer priors (LP) [5], deep detail network (DetailNet) [6], joint rain detection and removal (JORDER) [8], directional global sparse model (DGSM) [69], joint convolutional analysis and synthesis (JCAS) [18], density-aware multi-stream dense network (DID-MDN) [15], image de-raining using a conditional generative adversarial network (ID-CGAN) [70], and a common CNN baseline for image processing – SRCNN [21], trained for deraining. DetailNet and JORDER are retrained with the online available codes. Other methods are directly evaluated with the online available codes.

For rain accumulation removal, three state-of-the-art methods are compared: DehazeNet [31], nonlocal image dehazing (Nonlocal) [71], gated fusion network (GFN) [72]. The codes of the three methods are kindly provided by the authors.

TABLE II  
PSNR RESULTS AMONG DIFFERENT METHODS

Dataset	<i>Rain12</i>	<i>Rain100L</i>	<i>Rain100H</i>	<i>Rain800</i>
ID	27.21	23.13	13.78	20.54
DSC	30.02	24.16	15.66	25.57
LP	32.02	29.11	14.26	27.09
CNN	26.65	23.70	13.21	23.95
SRCNN	34.41	32.63	18.29	25.10
DetailNet	35.31	33.50	23.93	25.22
UGSM	33.30	28.83	14.06	23.12
JCAS	33.09	29.91	14.26	22.25
DID-MDN	30.14	28.27	13.85	22.55
ID-CGAN	20.78	23.39	16.86	23.81
JORDER	36.02	36.11	24.10	26.73
RWL	<b>36.32</b>	<b>36.75</b>	<b>26.89</b>	<b>27.79</b>

For the experiments on synthesized data, two metrics Peak Signal-to-Noise Ratio (PSNR) [73] and Structure Similarity Index (SSIM) [74] 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.

### C. Quantitative Evaluation

Tables II and III show the results of different methods on *Rain12*, *Rain100L* and *Rain100H*. As observed, our method considerably outperforms other methods in terms of both

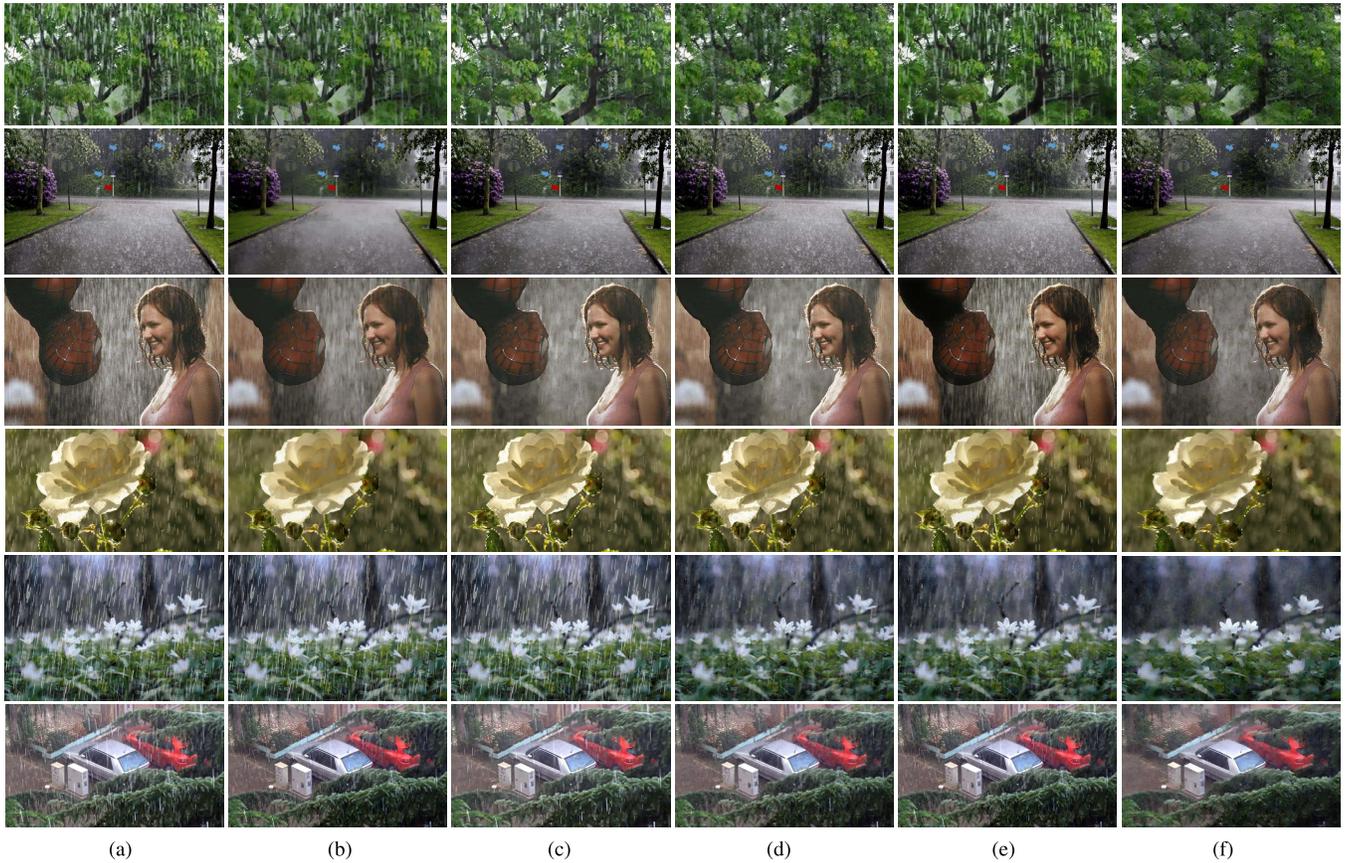


Fig. 6. Visual comparison of our RWL with state-of-the-art rain removal algorithms on real rain images including large rain streaks. It is observed that, our RWL successfully removes most large rain streaks. Zooming-in the figure will provide a better look at the restoration quality. (a) Input. (b) LP. (c) DetailNet. (d) JORDER. (e) DID-MDN. (f) RWL.

TABLE III  
SSIM RESULTS AMONG DIFFERENT METHODS

Baseline	<i>Rain12</i>	<i>Rain100L</i>	<i>Rain100H</i>	<i>Rain800</i>
ID	0.7534	0.7022	0.3968	0.6739
DSC	0.8679	0.8728	0.5444	0.6521
LP	0.9082	0.8786	0.4225	0.7801
CNN	0.7829	0.8076	0.3712	0.6589
SRCNN	0.9421	0.9357	0.6124	0.8232
DetailNet	0.9485	0.9444	0.7251	0.8228
UGSM	0.9323	0.8876	0.4454	0.7675
JCAS	0.9276	0.9053	0.4837	0.7682
DID-MDN	0.8762	0.8625	0.3748	0.7639
ID-CGAN	0.8519	0.8275	0.4921	0.8072
JORDER	0.9598	0.9705	0.7490	0.8683
RWL	<b>0.9632</b>	<b>0.9754</b>	<b>0.8406</b>	<b>0.8795</b>

PSNR and SSIM. The PSNR of recurrent wavelet learning gains over JORDER more than 1dB. Such a large gain demonstrates the effectiveness of proposed recurrent wavelet learning and dilated residual dense network on synthesized heavy rain images.

#### D. Qualitative Evaluation

Fig. 5 shows the results of synthetic images. Figs. 6 shows the results of real images including large rain streaks. Fig. 7 presents the results of real images with rain streak accumulation. As observed, our method significantly outperforms

previous state-of-the-art methods and successfully removes most of rain streaks, enhances the visibility and lights up details in dark regions.

#### E. Ablation Study for Recurrent Wavelet Learning

To evaluate the effectiveness of recurrent wavelet learning, we perform an ablation analysis in Fig. 8. It is observed that, the recurrent wavelet learning (d) outperforms both directly processing the rain input (b) and only processing the low-frequency average component (c).

#### F. Ablation Study for Dilated Residual Dense Network

We compare four versions of our network: the version without residual learning and dilated convolutions (V0), the version without dilated convolutions (V1), the version without residual learning (V2), the full version (V3). In this comparison, only one dense block is adopted in four networks. The evaluation is performed on *Rain20H*, a random sampled 20 testing images from *Rain100H*. The evaluation PSNR and SSIM result during the training is shown in Figs. 10 and 11. It is observed that, the dilated convolution significantly boosts the performance (from V0 to V2 and V1 to V3). The added residue also leads to a superior performance (from V0 to V1 and V2 to V3). We also compare the visual results of different versions of our method in Fig. 9. The DenseNet without residual learning

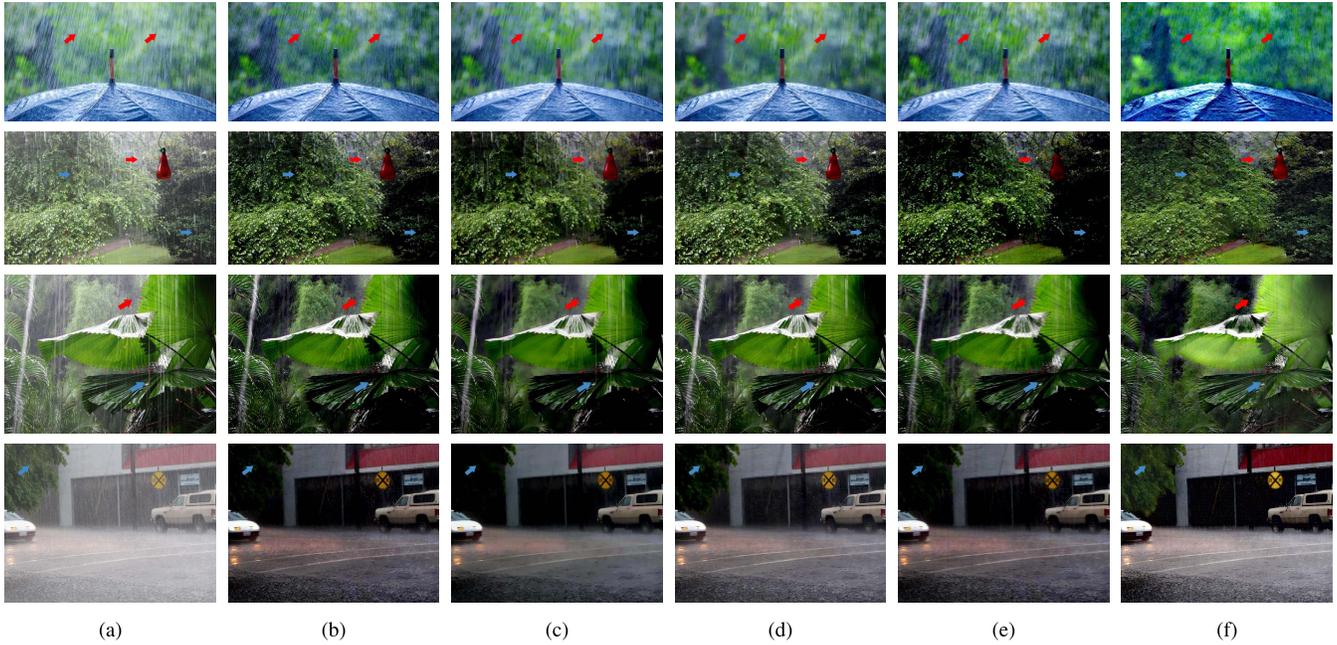


Fig. 7. Visual comparison of our method with state-of-the-art rain removal algorithms on real rain images with rain streak accumulation. Our method is the combination of RWL and our detail appearing rain accumulation removal method. Other methods are conducted with DehazeNet [31]. It is observed that, our method is better at removing rain streak accumulation (denoted by red arrows) and lighting up the details in dark regions (denoted by blue arrows). (a) Input. (b) DSC. (c) LP. (d) DetailNet. (e) JORDER. (f) Proposed.

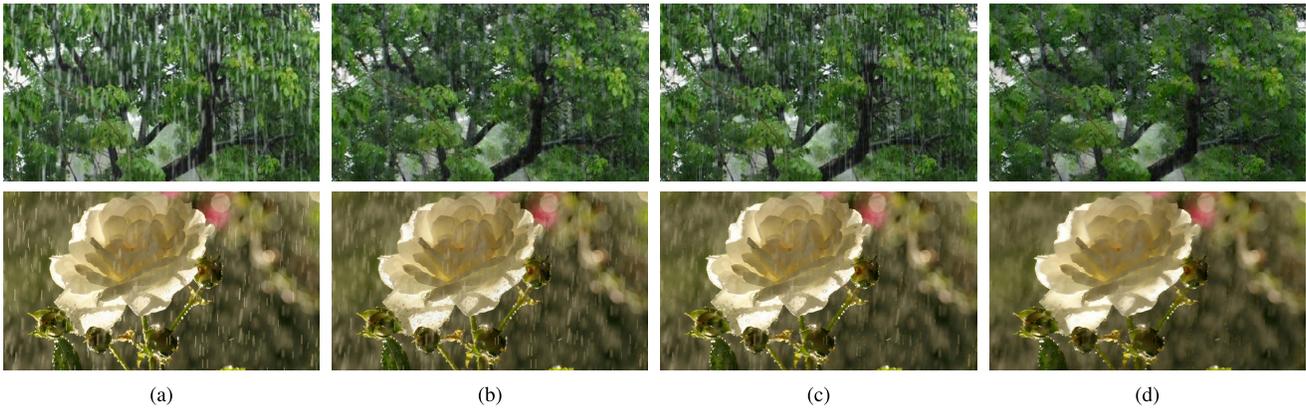


Fig. 8. Ablation study for recurrent wavelet learning.  $F_1$  directly adopts rain removal on the input image.  $F$  performs rain removal on the low-frequency component of the input image.  $F+G$  performs the whole recurrent wavelet learning. Zooming-in the figure will provide a better look at the restoration quality. (a) Input. (b)  $F_1$ . (c)  $F$ . (d)  $F + G$ .

is also compared. It is observed that, the usage of dilated convolutions and residual learning jointly leads to cleaner rain removal and finer detail recovery.

*G. Visual Results of Models With / Without Residual Learning*

We also compare the visual results of our models with and without residual learning in Fig. 12. It is observed that, residual learning slightly improves the performance in visual quality (from V0 to V1 and from V2 to V3).

*H. Evaluation on Streak Size Mismatch*

To prove the effectiveness of RWL to handle 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 IV. It is observed that,  $F + G$  achieves superior performance than  $F_1$ . The gain is almost 2 dB on *Rain100H-S2* and 1.8 dB on *Rain100H-S3*. Some visual results are provided in Fig. 13.

*I. Wavelet Reconstruction Results*

The wavelet decomposition and reconstruction results are presented in Fig. 14. It is clearly shown that, the streaks appear on average and vertical components. After deraining, most streaks in these two components are removed.

*J. Comparing With State-of-the-Art Dehazing Methods*

To demonstrate the superiority of our detail appearing rain accumulation removal method, we compare the

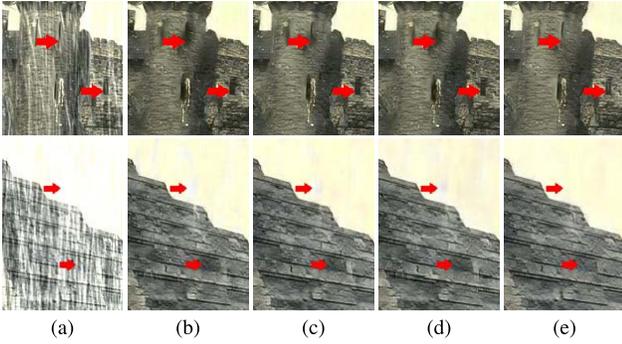


Fig. 9. Ablation study for dilated residual dense network in visual quality. (a) Input. (b) DenseNet. (c) V0. (d) V1. (e) V3.

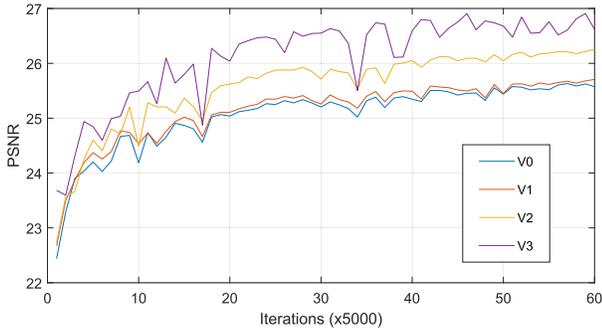


Fig. 10. Ablation study for dilated residual dense network in PSNR.

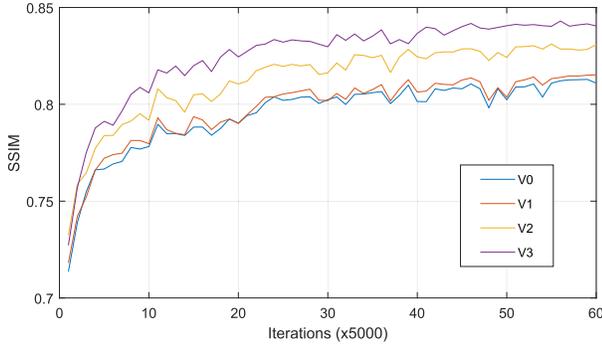


Fig. 11. Ablation study for dilated residual dense network in SSIM.

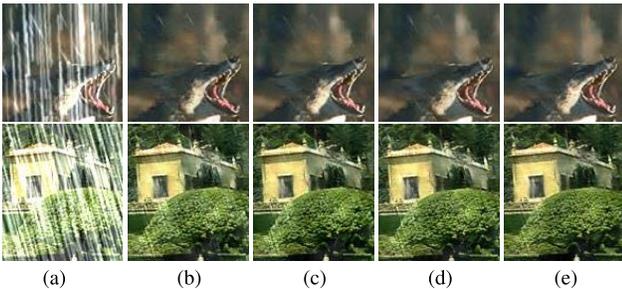


Fig. 12. Visual results of our models with and without residual learning. (a) Input. (b) V0. (c) V1. (d) V2. (e) V3.

proposed method with other state-of-the-art dehazing methods in Fig. 15. It is observed that, the result of Nonlocal has color shift. DehazeNet and GFN tend to produce darker results

TABLE IV  
PSNR AND SSIM RESULTS WHEN TRAINING AND TESTING STREAK SIZES ARE DIFFERENT

Baseline	<i>Rain100H-S2</i>		<i>Rain100H-S3</i>	
Metric	PSNR	SSIM	PSNR	SSIM
JORDER [8]	16.11	0.6464	15.21	0.6483
$F_1$	18.87	0.7624	16.48	0.7296
$F$	18.48	0.7532	16.02	0.7134
$F + G$	<b>20.83</b>	<b>0.8050</b>	<b>18.28</b>	<b>0.7824</b>

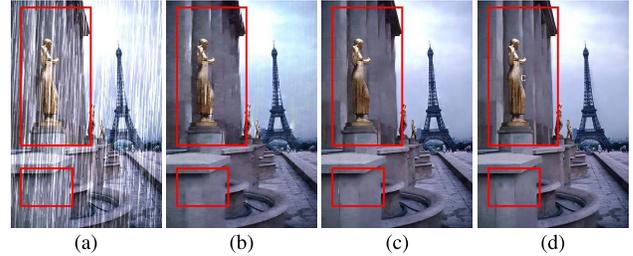


Fig. 13. Evaluation results when training and testing streak sizes are different. (a) Input. (b) JORDER. (c)  $F_1$ . (d)  $F + G$ .

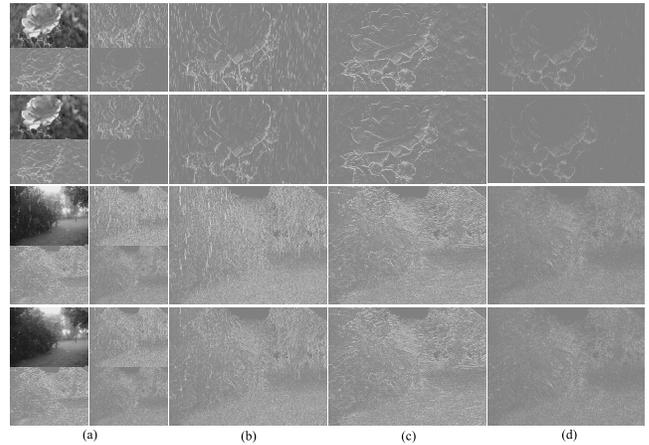


Fig. 14. Wavelet reconstruction results. The 1st and 3rd rows: rain images. The 2nd and 4th rows: rain removal images. (a): the second order wavelet. (b)-(d): the first order wavelet.

and retains some accumulation. Comparatively, our detail appearing rain accumulation removal method is successful in removing most accumulation and lighting up dark details in the images.

### K. Application in Computer Vision

We show two cases of applying our method as pre-processing for a commercial computer vision system, Clarifai,<sup>3</sup> which is an advanced image recognition system based on a deep convolutional network. The two images are shown in Fig. 16. Before rain removal, these images are categorized as ‘Nature’, which is a rough category. After rain removal by our method, they are labeled accurately as ‘Tree’ and ‘Flower’, respectively.

<sup>3</sup><https://www.clarifai.com/>.

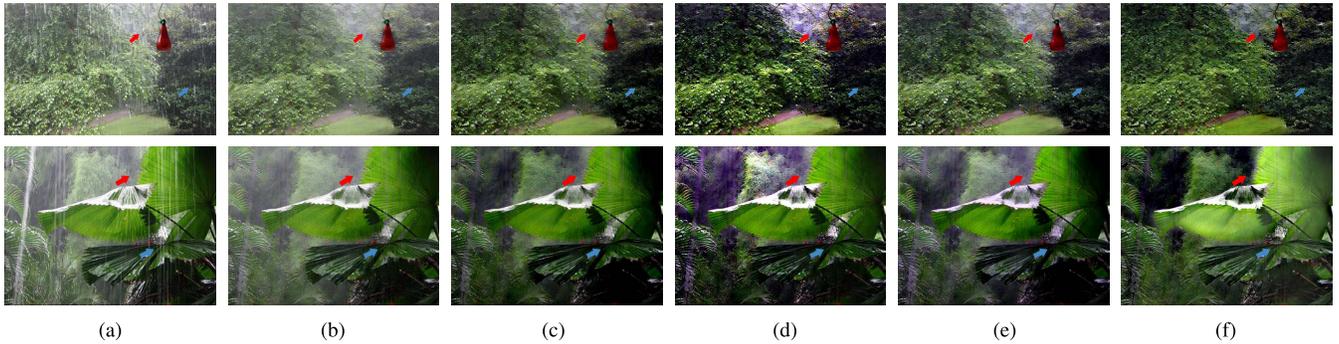


Fig. 15. Visual comparison of our detail appearing rain accumulation removal with state-of-the-art dehazing algorithms on real rain images with rain streak accumulation. All methods take the rain streak removal results produced by our RWL as their inputs. It is observed that, our detail appearing rain accumulation removal method is successful in removing most accumulation (denoted by red arrows) and lighting up details in dark regions (denoted by blue arrows). (a) Input. (b) RWL. (c) DehazeNet. (d) Nonlocal. (e) GFN. (f) Proposed.

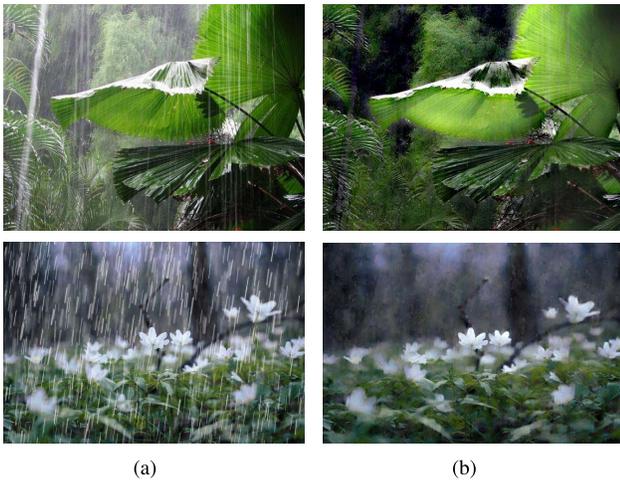


Fig. 16. Image recognition results on the images before and after rain-streak removal. Top panel: (a) Before, labeled as ‘Nature’. (b) After, labeled as ‘Tree’. Bottom panel: (a) Before, labeled as ‘Nature’. (b) After, labeled as ‘Flower’.

TABLE V  
PSNR AND SSIM RESULTS WITH AND WITHOUT RECOVERED LOW FREQUENCY GUIDANCE

Baseline	Metric	<i>Rain12</i>	<i>Rain100L</i>	<i>Rain100H</i>	<i>Rain800</i>
RWL	PSNR	36.32	36.75	26.89	27.79
RWL w/o BLL		36.21	36.52	26.47	27.54
RWL	SSIM	0.9632	0.9754	0.8406	0.8795
RWL w/o BLL		0.9618	0.9732	0.8323	0.8793

L. Analysis on Effect of Recovered Low Frequency Guidance

To verify the effectiveness of low frequency guidance  $\hat{B}_{LL}^z$  in Eq. (3), we compare with the RWL model where the recovered low frequency guidance is not included in the input of  $G$ . As shown in Table V, the recovered low frequency guidance in  $G$  slightly improves the performance, which demonstrates the positive effect of the recovered low frequency guidance on the rain removal of high-frequency components.

M. Analysis on Effect of Inverse Wavelet Transform

We compare our RWL with the one that replaces  $H$  in Eq. (4) with deconvolution layers. As shown in Table VI,

TABLE VI  
PSNR AND SSIM RESULTS WITH INVERSE DISCRETE WAVELET TRANSFORM AND DECONVOLUTION LAYER

Baseline	Metric	<i>Rain12</i>	<i>Rain100L</i>	<i>Rain100H</i>	<i>Rain800</i>
RWL	PSNR	36.32	36.75	26.89	27.79
RWL (Deconv)		36.26	36.64	26.68	27.67
RWL	SSIM	0.9632	0.9754	0.8406	0.8795
RWL (Deconv)		0.9634	0.9748	0.8357	0.8791

the original RWL with  $H$  is slightly superior to the model with deconvolution layers in PSNR and SSIM.

VII. CONCLUSION

In this paper, we construct dilated residual dense networks following the recurrent wavelet learning for rain streak removal. This architecture has two benefits: 1) the recurrent wavelet learning allows to mitigate the streak size mismatch problem between training and testing phases; 2) dilated residual dense networks model the intrinsic dependency among different scales of an image and effectively expand the receptive field. These properties facilitate the removal of large rain streaks. We also consider to model rain accumulation and low light condition. Thus, the developed approach produces the results with better visibility and brighter details. The evaluation on both synthetic and real images, particularly on those containing large rain streaks, shows the effectiveness and superiority of the proposed method.

REFERENCES

- [1] 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.
- [2] 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.
- [3] 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.*, Oct. 2014, pp. 4482–4486.
- [4] Y. Luo, Y. Xu, and H. Ji, “Removing rain from a single image via discriminative sparse coding,” in *Proc. IEEE Int. Conf. Computer Vis.*, Dec. 2015, pp. 3397–3405.
- [5] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, “Rain streak removal using layer priors,” in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2736–2744.

- [6] X. Fu, J. Huang, D. Z. 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.
- [7] 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.
- [8] 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 Int. Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1357–1366.
- [9] R. Li, L.-F. Cheong, and R. T. Tan. (Dec. 2017). "Single image deraining using scale-aware multi-stage recurrent network." [Online]. Available: <https://arxiv.org/abs/1712.06830>
- [10] D.-A. Huang, L.-W. Kang, M.-C. Yang, C.-W. Lin, and Y.-C. F. Wang, "Context-aware single image rain removal," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2012, pp. 164–169.
- [11] J. Chen and L.-P. Chau, "Rain removal from dynamic scene based on motion segmentation," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2013, pp. 2139–2142.
- [12] X. Xue, X. Jin, C. Zhang, and S. Goto, "Motion robust rain detection and removal from videos," in *Proc. IEEE 14th Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2012, pp. 170–174.
- [13] 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.
- [14] 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. (ICIP)*, Sep. 2013, pp. 914–917.
- [15] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1–10.
- [16] Y. Chang, L. Yan, and S. Zhong, "Transformed low-rank model for line pattern noise removal," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 1735–1743.
- [17] 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.*, Oct. 2017, pp. 2545–2553.
- [18] 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.*, Oct. 2017, pp. 1717–1725.
- [19] 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.
- [20] 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.
- [21] C. Dong, 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. 2015.
- [22] 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.
- [23] W. Yang, S. Xia, J. Liu, and Z. Guo, "Reference guided deep super-resolution via manifold localized external compensation," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.
- [24] W. Yang, J. Feng, G. Xie, J. Liu, Z. Guo, and S. Yan, "Video super-resolution based on spatial-temporal recurrent residual networks," *Comput. Vis. Image Understand.*, vol. 168, pp. 79–92, Mar. 2017.
- [25] S. Xia, W. Yang, J. Liu, and Z. Guo, "Dual recovery network with online compensation for image super-resolution," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2018, pp. 1–5.
- [26] W. Lai, J. Huang, N. Ahuja, and M. Yang, "Fast and accurate image super-resolution with deep Laplacian pyramid networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published.
- [27] W. Yang, S. Deng, Y. Hu, J. Xing, and J. Liu, "Real-time deep video spatial resolution upconversion system (struct++ demo)," in *Proc. ACM Int. Conf. Multimedia*, Oct. 2017, pp. 1255–1256.
- [28] Y. Hu, J. Liu, W. Yang, S. Deng, L. Zhang, and Z. Guo, "Real-time deep image super-resolution via global context aggregation and local queue jumping," in *Proc. IEEE Vis. Commun. Image Process.*, Dec. 2017, pp. 1–4.
- [29] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf. (2014). "Learning to deblur." [Online]. Available: <https://arxiv.org/abs/1406.7444>
- [30] L. A. Gatys, A. S. Ecker, and M. Bethge. (2015). "A neural algorithm of artistic style." [Online]. Available: <https://arxiv.org/abs/1508.06576>
- [31] 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.
- [32] 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.
- [33] K. G. Lore, A. Akintayo, and S. Sarkar. (2015). "LLNet: A deep autoencoder approach to natural low-light image enhancement." [Online]. Available: <https://arxiv.org/abs/1511.03995>
- [34] W. Wang, C. Wei, W. Yang, and J. Liu, "GLADNet: Low-light enhancement network with global awareness," in *Proc. IEEE Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 751–755.
- [35] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Sep. 2018, p. 1.
- [36] J. Liu, W. Yang, S. Yang, and Z. Guo, "D3R-Net: Dynamic routing residue recurrent network for video rain removal," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 699–712, Feb. 2019.
- [37] X. Zhang, W. Yang, Y. Hu, and J. Liu, "DMCNN: Dual-domain multi-scale convolutional neural network for compression artifacts removal," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2018, pp. 390–394.
- [38] 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. 2017.
- [39] Y. Fang, C. Zhang, W. Yang, J. Liu, and Z. Guo, "Blind visual quality assessment for image super-resolution by convolutional neural network," in *Multimedia Tools Appl.*, vol. 77, no. 22, pp. 29829–29846, Nov. 2018.
- [40] J. Liu, S. Xia, W. Yang, M. Li, and D. Liu, "One-for-All: Grouped variation network based fractional interpolation in video coding," *IEEE Trans. Image Process.*, to be published.
- [41] S. Xia, W. Yang, Y. Hu, S. Ma, and J. Liu, "A group variational transformation neural network for fractional interpolation of video coding," in *Proc. Data Compress. Conf.*, Mar. 2018, pp. 127–136.
- [42] Y. Hu, W. Yang, S. Xia, W.-H. Cheng, and J. Liu, "Enhanced intra prediction with recurrent neural network in video coding," in *Proc. Data Compress. Conf.*, Mar. 2018, p. 413.
- [43] Y. Hu, W. Yang, S. Xia, and J. Liu, "Optimized recurrent network for intra prediction in video coding," in *Proc. IEEE Vis. Commun. Image Process.*, Dec. 2018, pp. 1–5.
- [44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [45] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2261–2269.
- [46] 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.
- [47] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1–10.
- [48] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 4549–4557.
- [49] 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 Int. Conf. Comput. Vis.*, Jul. 2018, pp. 1–16.
- [50] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1–10.
- [51] W. Dong, M. Yuan, X. Li, and G. Shi. (2018). "Joint demosaicing and denoising with perceptual optimization on a generative adversarial network." [Online]. Available: <https://arxiv.org/abs/1802.04723>
- [52] L. Galteri, L. Seidenari, M. Bertini, and A. Del Bimbo, "Deep generative adversarial compression artifact removal," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2017, pp. 4836–4845.
- [53] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8183–8192.
- [54] H. Ji and C. FermÅller, "Robust wavelet-based super-resolution reconstruction: Theory and algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 4, pp. 649–660, Apr. 2009.
- [55] S. Naik and N. Patel, "Single image super resolution in spatial and wavelet domain," *Int. J. Multimedia Appl.*, vol. 5, no. 4, p. 23, 2013.

[56] S. Zhao, H. Han, and S. Peng, "Wavelet-domain HMT-based image super-resolution," in *Proc. IEEE Int. Conf. Image Process.*, vol. 2, Sep. 2003, pp. II-953-II-956.

[57] X. Gao and H. Xiong, "A hybrid wavelet convolution network with sparse-coding for image super-resolution," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2016, pp. 1439-1443.

[58] S. Mallat, "Understanding deep convolutional networks," *Philos. Trans. A, Math., Phys., Eng. Sci.*, vol. 374, no. 2065, p. 20150203, 2016.

[59] V. Gupta, R. Mahle, and R. S. Shriwas, "Image denoising using wavelet transform method," in *Proc. Int. Conf. Wireless Opt. Commun. Netw. (WOCN)*, Jul. 2013, pp. 1-4.

[60] T.-C. Hsung, D. P. Lun, and W.-C. Siu, "A deblocking technique for block-transform compressed image using wavelet transform modulus maxima," *IEEE Trans. Image Process.*, vol. 7, no. 10, pp. 1488-1496, Oct. 1998.

[61] L. Shen, Z. Yue, Q. Chen, F. Feng, and J. Ma. (Jan. 2018). "Deep joint rain and haze removal from single images." [Online]. Available: <https://arxiv.org/abs/1801.06769>

[62] E. Denton, S. Chintala, A. Szlam, and R. Fergus, "Deep generative image models using a Laplacian pyramid of adversarial networks," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2015, pp. 1486-1494.

[63] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. Int. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 5835-5843.

[64] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674-693, Jul. 1989.

[65] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learn. Represent.*, 2016, pp. 1-13.

[66] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *Int. J. Comput. Vis.*, vol. 48, no. 3, pp. 233-254, 2002.

[67] Y. Tian and S. G. Narasimhan, "Seeing through water: Image restoration using model-based tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2303-2310.

[68] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 81-88.

[69] 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. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0307904X18301069>

[70] H. Zhang, V. Sindagi, and V. M. Patel. (2017). "Image de-raining using a conditional generative adversarial network." [Online]. Available: <https://arxiv.org/abs/1701.05957>

[71] D. Berman *et al.*, "Non-local image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1674-1682.

[72] W. Ren *et al.*, "Gated fusion network for single image dehazing," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1-9.

[73] 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, Jun. 2008.

[74] 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.

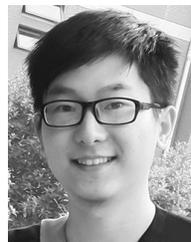


**Wenhan Yang** (S'17-M'18) received the B.S. and Ph.D. degrees (Hons.) in computer science from Peking University, Beijing, China, in 2012 and 2018, respectively. He is currently a Post-Doctoral Research Fellow with the Department of Computer Science, City University of Hong Kong. He was a Visiting Scholar with the National University of Singapore, from 2015 to 2016. His current research interests include deep-learning-based image processing, bad weather restoration, related applications and theories.



**Jiaying Liu** (S'08-M'10-SM'17) received the B.E. degree in computer science from Northwestern Polytechnic University, Xi'an, China, and the Ph.D. degree (Hons.) in computer science from Peking University, Beijing, China, in 2005 and 2010, respectively.

She is currently an Associate Professor with the Institute of Computer Science and Technology, Peking University. She has authored over 100 technical articles in refereed journals and proceedings. She holds 28 granted patents. Her current research interests include image/video 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 for Young Faculties. She has also served as a TC Member for the IEEE CAS-MSA/EOT and APSIPA IVM, and APSIPA Distinguished Lecturer from 2016 to 2017. She is a CCF/IEEE Senior Member.



**Shuai Yang** received the B.S. degree in computer science from Peking University, Beijing, China, in 2015, where he is currently pursuing the Ph.D. degree with the Institute of Computer Science and Technology.

His current research interests include image inpainting, depth map enhancement, and image stylization.



**Zongming Guo** (M'09) received the B.S. degree in mathematics, and the M.S. and Ph.D. degrees in computer science from Peking University, Beijing, China, in 1987, 1990, and 1994, respectively.

He is currently a Professor with the Institute of Computer Science and Technology, Peking University. His current research interests include video coding, processing, and communication.

Dr. Guo is an Executive Member of the China Society of Motion Picture and Television Engineers. He was a recipient of the First Prize of the State Administration of Radio Film and Television Award in 2004, the First Prize of the Ministry of Education Science and Technology Progress Award in 2006, the Second Prize of the National Science and Technology Award in 2007, and the Wang Xuan News Technology Award and the Chia Tai Teaching Award in 2008. He received the Government Allowance granted by the State Council in 2009. He received the Distinguished Doctoral Dissertation Advisor Award from Peking University in 2012 and 2013.