

# Dual Attention-in-Attention Model for Joint Rain Streak and Raindrop Removal

Kaihao Zhang, Dongxu Li, Wenhan Luo, and Wenqi Ren

**Abstract**—Rain streaks and raindrops are two natural phenomena, which degrade image capture in different ways. Currently, most existing deep deraining networks take them as two distinct problems and individually address one, and thus cannot deal adequately with both simultaneously. To address this, we propose a Dual Attention-in-Attention Model (DAiAM) which includes two DAMs for removing both rain streaks and raindrops. Inside the DAM, there are two attentive maps - each of which attends to the heavy and light rainy regions, respectively, to guide the deraining process differently for applicable regions. In addition, to further refine the result, a Differential-driven Dual Attention-in-Attention Model (D-DAiAM) is proposed with a “heavy-to-light” scheme to remove rain via addressing the unsatisfying deraining regions. Extensive experiments on one public raindrop dataset, one public rain streak and our synthesized joint rain streak and raindrop (JRSRD) dataset have demonstrated that the proposed method not only is capable of removing rain streaks and raindrops simultaneously, but also achieves the state-of-the-art performance on both tasks.

**Index Terms**—Rain streaks, raindrops, joint deraining, dual attention, attention-in-attention, differential-driven module.



## 1 INTRODUCTION

As one of the commonest weather phenomena, rain causes visibility degradation and destroys the performance of many computer vision systems, *e.g.*, object detection [1], [2], outdoor surveillance [3], [4] and autonomous driving [5], [6]. Rain removal is to restore clean images from rainy ones, which is an important problem in computer vision field and still challenging due to its various types (*i.e.*, rain streaks and raindrops), and different intensities (*i.e.*, heavy and light rain).

In the last decade, a set of methods have been proposed for rain removal. For rain streak removal, some methods model the physical characteristics of rain and generate sharp version with various image priors [7], [8], [9], [10]. we have also witnessed significant progress of deep learning based methods [11], [12], [13], [14], [15]. Some others focus on raindrop removal via detecting and removing raindrop using multiple images or single image [16], [17], [17], [18], [19]. Despite of the achieved promising performance, there still exist major challenges in rain removal:

- Rain streaks and raindrops are two related but different types. The rain streaks lead to the occlusion of objects and scene, while raindrops can cause change of shape. In the real world, both of them often appear simultaneously. However, most deep learning based

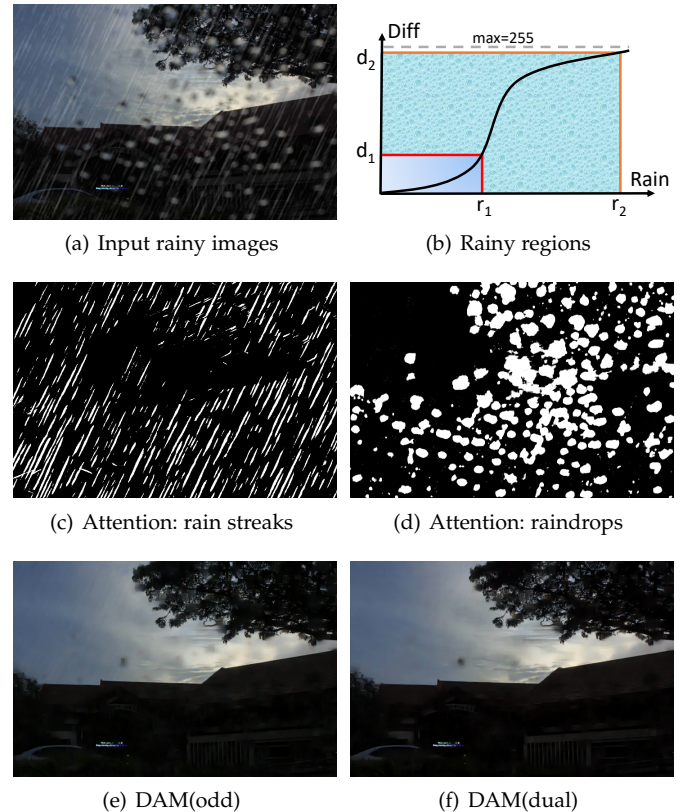


Fig. 1. Analyses and deraining results. (a) is an input rainy image. (b) describes the relationship of the rain intensity and the difference between rainy and clean images. (c) and (d) are the generated attention maps for rain streaks and raindrops, respectively. (e) and (f) are the deraining results of the proposed DAM with odd attention and dual attention, respectively.

- Kaihao Zhang and Dongxu Li are with the College of Engineering and Computer Science, Australian National University, Canberra, ACT, Australia. E-mail: {kaihao.zhang@anu.edu.au; dongxu.li@anu.edu.au}
- W. Luo is with the Tencent AI Laboratory, Shenzhen 518057, China. E-mail: whluo.china@gmail.com
- W. Ren is with State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing, 100093, China. E-mail: rwq.renwenqi@gmail.com.

Manuscript received April 19, 2005; revised August 26, 2015.

deraining methods and datasets typically focus on one of them.

- As Fig. 1(b) shows, the pixel difference between clean

and rainy images increases as the rain becomes heavier. Previous attention based methods use a fixed threshold  $d_1$  to determine whether a pixel is part of rain regions. These methods focus only on the top-right heavy rainy region and ignore the bottom-left light rainy region. In this case, the efficacy of attention mechanism will be restricted if  $d_1$  is set inappropriately large or small.

- For many cases like heavy rain, the current rain removal methods can remove rain to some extent and generate a derained image with less rain. However, it is difficult to further improve the performance by simply modifying the structure of deep networks like increasing the depth.

To address the first and second problems, a new framework which exploits the cues from different types of rain is proposed. Specially, we propose a Dual Attention-in-Attention Model, termed as **DAiAM**, to remove rain streaks and raindrops, simultaneously. It contains two branches, corresponding to two Dual Attention Model (**DAM**). Each DAM removes one type of rain via simultaneously focusing on different rain intensities. Different from previous attention-based deraining methods, which learn only the attention map of heavy rain regions (top-right regions in Fig. 1(b)), an advantage of the DAM is that it also pays attention to the light rain regions (bottom-left regions in Fig. 1(b)). One pair of heavy-rain-aware and light-rain-aware attention maps is generated to help remove rain from multiple regions. As such, the proposed method avoids the negative effects from unsuitable thresholds. Fig. 1(e) and 1(f) show the attention maps for rain streaks and raindrops, respectively.

For the third challenge, a Differential-driven Dual Attention-in-Attention Model (**D-DAiAM**), is proposed based on a “heavy-to-light” scheme. The input rainy images and output derained images from DAiAM are processed with the proposed differential-driven module, guiding the learning of the following DAiAM to further remove rain with different intensities or different types.

In order to evaluate the performance of the proposed method on rain streak and raindrop removal, a joint rain streak and raindrop dataset (**JRSRD**), is built. The rain streaks and raindrops often happen simultaneously, thus evaluating methods in this scenario is necessary to verify the performance of different methods in the wild.

The contributions of this work can be summarized as: 1) To address the problem of joint rain streak and raindrop removal, a dual attention-in-attention model (DAiAM), is proposed to remove two variations of rain. 2) Inside DAiAM, there are two well-designed DAMs, which focus on local regions with different rainy intensities. The generated intensity-aware attention maps enable better removal of rain in multiple regions. 3) D-DAiAM is proposed to alleviate the limitation of increasing depth and width of deraining methods, and thus improve the image quality. 4) A new JRSRD dataset of both rain streaks and raindrops is built. We compare the proposed method with current deraining methods. Experimental results show that the proposed method achieves not only the state-of-the-art performance on public rain streak dataset and raindrop dataset, but also

consistently better results on images with both rain streaks and raindrops.

## 2 RELATED WORKS

Our work is an attempt for jointly addressing the rain streak and raindrop removal based on attention mechanism. The following is a brief review of related works on rain streak removal, raindrop removal, as well as attention mechanism, respectively.

### 2.1 Rain Streak Removal

Traditional methods design hand-crafted priors to remove rain streaks [8], [20], [21], [22], [23], [24], [25], [26], [27], [28]. Kang *et al.* [8] use a bilateral filter to decompose an image into the low- and high-frequency parts, which are then decomposed into different components by performing dictionary learning and sparse coding. Similarly, Huang *et al.* [21] present a method to first learn an over-complete dictionary from the image high spatial frequency parts and then perform unsupervised clustering on the dictionary atoms. Zhu *et al.* [25] use a joint optimization process with three image priors to remove rain-streak details.

Recently, deep learning achieves significant success in low-level vision tasks [29], [30], [31], [32], [33], [34], which also include rain streak removal [11], [12], [13], [14], [15], [35], [36], [37]. Fu *et al.* [12] propose a deep network to remove background interference and focus on the structure of rain based on prior knowledge. Zhang *et al.* [15] introduce a DID-MDN model to jointly estimate rain density and remove rain. Li *et al.* [13] propose a deep convolutional and recurrent neural network for deraining. To make the derained images more realistic, Zhang *et al.* [35] introduce a CGAN-based model with additional regularization. Wang [38] explore the intrinsic prior structure of rain streaks and then propose a novel interpretable network to remove the rain streaks from rainy images. Li *et al.* [39] propose a comprehensive benchmark analysis of several single image deraining networks. Zhu *et al.* [26] and Hu *et al.* [27] introduce two non-local networks to improve the performance of image deraining. Wang *et al.* [28] rethink about the image deraining and reformulate rain streaks as transmission medium together with vapors to address the problem of image deraining.

In addition, there still exists some video-based rain streak removal methods [40], [41], [42], [43], [44], [45]. Specially, Chen *et al.* [43] use a super-pixel segmentation scheme to help restore clean frames via a robust deep CNN. Liu *et al.* [42] remove rain streak via classifying all pixels. More recently, Yang *et al.* [44] introduce a two-stage recurrent network to capture the motion consistency to remove rain streaks.

### 2.2 Raindrop Removal

Most methods for rain streak removal are not directly applicable for raindrop removal. Therefore, many methods are proposed like raindrop detection and removal [16], [17], [18], [46], [47], [48], [49], [50], [51], [52]. Specially, Kurihata *et al.* [46] use PCA to learn the shape of raindrops, which are then utilized to match rainy regions. Yamashita *et al.* [47]

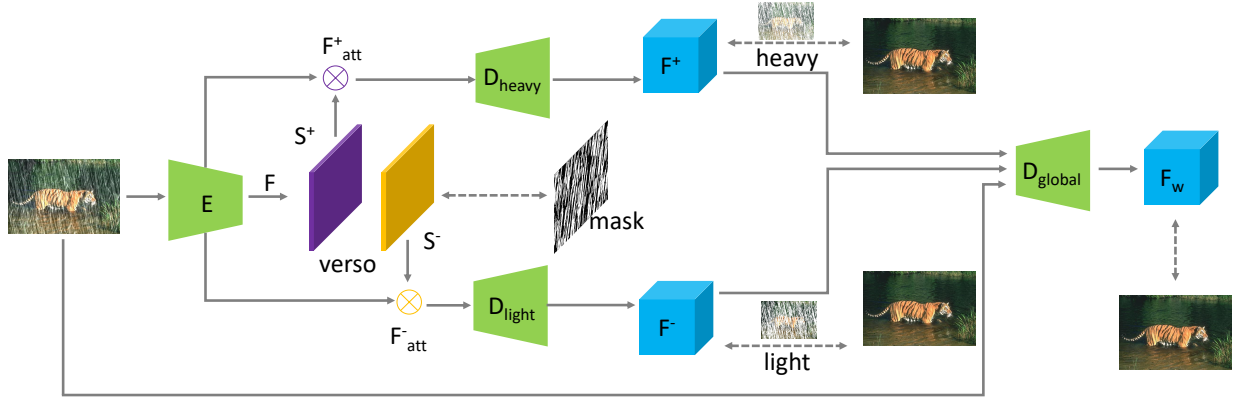


Fig. 2. The framework of DAM for image deraining. It contains three main branches, *i.e.*, heavy-rain branch, light-rain branch and full-image branch. The dual attention sub-network in the middle is utilized to generate a pair of heavy-rain-aware and light-rain-aware maps to pointedly remove rain from different regions. The original rainy image and the intermediate results are then concatenated to generate the final deraining image.

introduce a method based on the stereo measurement and disparities between stereo image pair. Position of raindrops can be detected. Finally, sharp image can be obtained by replacing raindrop regions. Roser *et al.* [16] propose a method to perform monocular raindrop detection. [49] introduces a method to exploit local spatio-temporal cue for video raindrop removal. They first model and detect adherent raindrops, then remove them and restore the images. More recently, there are many methods using CNN for single image raindrop removal [18], [19], which are trained with pairs of raindrops and corresponding sharp images. Quan [50] propose a CNN-based method to restore an image taken through glass window in rainy weather via using shape-driven attention and channel re-calibration.

Almost all existing methods disserve the two tasks and focus on either rain streaks or raindrops [53]. Meanwhile, most datasets typically contain only one kind of rain. Given that the two phenomenons usually appear simultaneously in the real world, a new dataset including raindrops and rain streaks is built in this paper.

### 2.3 Attention Mechanism

The visual attention model is effective in understanding image. It has achieved great success in tasks like object recognition [54], [55], [56], image captioning [57], [58] and saliency detection [59], [60], [61], [62]. For example, Ba *et al.* [54] use an attention mechanism to help their model decide where to focus its computation and thus propose a new method to train their object recognition model. Xu *et al.* [57] propose an attention based approach which can automatically learn to describe the content of images. Chen *et al.* [59] introduce a reverse attention module for salient object detection via guiding residual learning in a top-down manner. For deep image deraining, there are also some methods of attention mechanism [14], [63]. They utilize a threshold value to classify the regions of input rainy images into two classes like “rain” or “no-rain”, and then derive a spatial attention map to remove rain. As discussed, unsuitable thresholds cause errors, and restrict the potential of attention mechanism.

## 3 METHOD

We first take rain streak removal as an example to introduce the architecture and learning details of DAM. Then we represent DAiAM (Sec. 3.4) to jointly remove rain streaks and raindrops. Finally, a D-DAiAM framework (Sec. 3.5) is discussed to overcome the limitation of single model.

### 3.1 Overall Architecture of DAM

The overall architecture of the proposed **DAM** is shown in Fig. 2. A rainy image is fed into DAM to learn two attention maps, *i.e.*, heavy-rain-aware and light-rain-aware maps. The heavy-rain-aware map learns the attention which indicates the regions with heavy rain, and the light-rain-aware map represents the regions with light rain (Sec. 3.2).

Different from other deraining methods which directly concatenate the attention maps to generate final images, we produce two different kinds of intermediate results by two sub-networks in Sec. 3.3. The two attention maps provide not only attention to generate the final global deraining image, but also the reference to evaluate the performance of two sub-networks of DAM. Finally, the intermediate results concatenated with the input rainy image are put into a global decoder to generate the deraining image.

### 3.2 Dual Intensity-Aware Maps

In general, the DAM takes input images and produce weighting maps to focus on different spatial regions of images. By doing so, different sub-networks can exactly focus on different spatial regions that contribute most for differentiated image deraining. Specially, the proposed DAM take rainy images as input to capture the features  $F$  from the first-step encoder  $E$ . Then the feature maps are fed into two attention sub-networks to generate heavy-rain-aware and light-rain-aware maps, respectively. The heavy-rain-aware map  $S^+$  can be defined as:

$$S^+ = g(W * F + b), \quad (1)$$

where  $*$ ,  $W$  and  $b$  denote respectively convolution, convolution filters and biases.  $g$  is the sigmoid function.

Fig. 3 shows the architecture of  $E$ , and Table 1 presents the detailed configurations. As Fig. 3 shows,  $E$  is a structure

TABLE 1  
Configurations of the encoder  $E$ .

| layers    | kernel size  | output | operations | skip connection |
|-----------|--------------|--------|------------|-----------------|
| CNN       | $3 \times 3$ | 32     | -          | ResBlock1       |
| ResBlock1 | $3 \times 3$ | 32     | LReLU      | ResBlock2       |
| ResBlock2 | $3 \times 3$ | 32     | LReLU      | ResBlock3       |
| ResBlock3 | $3 \times 3$ | 32     | LReLU      | -               |
| LSTM      | $3 \times 3$ | 32     | Tanh       | -               |

of a recurrent network consisting of one CNN layer, three residual block and one LSTM layer. The output is the feature  $F$  captured from  $E$ . The attention map can be generated based on Eq. (1) in main submission.

Then we can similarly generate the light-rain-aware map based on Eq. (1). The heavy and light rain regions are a pair of complementary regions. Thus a constraint of them is set as:

$$S^+ + S^- = 1. \quad (2)$$

The two attention maps are two weighting maps which denote different region-aware attentions from the input features. Based on them, it is easy for the following sub-networks to pay attention to different regions and obtain different outputs. The operation to obtain the different features based on the two attention maps can be represented as,

$$F_{att}^+ = F \otimes S^+, \quad (3)$$

$$F_{att}^- = F \otimes S^-, \quad (4)$$

where  $\otimes$  denotes the channel-wise Hadamard matrix operation.  $F_{att}^+$  and  $F_{att}^-$  have the same size as  $F$  but are two re-weighted features by the two attention maps to focus on heavy-rain and light-rain regions, respectively. The  $S^-$  is the light-rain-aware attention map, where light-rain regions have higher weights and the heavy-rain regions have lower values. In order to guarantee that  $S^+$  learns the heavy-rain regions, we develop another constraint to make it focus on the heavy-rain regions and thus simultaneously push  $S^-$  to learn the light-rain regions. The loss function with this constraint is represented as

$$\mathcal{L}_{att} = \sum_{x=1}^X \sum_{y=1}^Y M_{(x,y)} - S_{(x,y)}^+, \quad (5)$$

where  $M$  is the rain-aware mask.  $X$  and  $Y$  are the width and height of the input features. Different from the previous methods [14], [63], which use a binary mask to represent the rain and no-rain regions, we apply a “soft” manner. Specially, we calculate the difference of images between the rainy and non-rainy versions and then normalize to the range between 0 and 1. This not only denotes whether the regions are rainy or not, but also represents the intensity of rain. In this way, we can avoid the negative effects caused by inappropriate thresholds and binary masks. Based on the above mechanism, two different attention maps are obtained with focus on heavy-rain and light-rain regions, respectively.

Namely, during the training stage, if the datasets do not provide rain streak and raindrops maps, we can generate them via calculating the difference between a rainy

TABLE 2  
Configurations of the proposed  $D_{heavy}$  and  $D_{light}$ .

| layers    | Kernel size  | output | operations | skip connection |
|-----------|--------------|--------|------------|-----------------|
| CNN1      | $3 \times 3$ | 64     | -          | ResBlock1       |
| ResBlock1 | $3 \times 3$ | 64     | LReLU      | ResBlock2       |
| ResBlock2 | $3 \times 3$ | 64     | LReLU      | ResBlock3       |
| ResBlock3 | $3 \times 3$ | 64     | LReLU      | ResBlock4       |
| ResBlock4 | $3 \times 3$ | 64     | LReLU      | ResBlock5       |
| ResBlock5 | $3 \times 3$ | 64     | LReLU      | -               |
| CNN2      | $3 \times 3$ | 3      | -          | -               |

image and its corresponding clean image. Heavier rain corresponds to greater values in the map. This works well in the practice.

### 3.3 Attentive Deraining from Regional and Global Levels

After the two attention maps are generated, we can improve the performance of deep deraining networks with them as reference. Specially, the attended features with the heavy-rain-aware attention map  $S^+$  and light-rain-aware attention map  $S^-$  are sent into two decoder networks to reconstruct two different deraining images with focus on different regions. The learning process can be defined as:

$$\mathcal{L}_{heavy} = I_c - D_{heavy}(F_{att}^+, I_i), \quad (6)$$

$$\mathcal{L}_{light} = I_c - D_{light}(F_{att}^-, I_i), \quad (7)$$

where  $I_c$  denotes the clean image and  $I_i$  is the input rainy image. The decoder networks  $D_{heavy}$  and  $D_{light}$  generate two deraining images, and the attentions of them are different.  $\mathcal{L}_{heavy}$  specially constrains the network  $D_{heavy}$  to mainly focus on the heavy-rain regions but consider less the light-rain regions due to the weighting values from  $S^+$ . The  $\mathcal{L}_{light}$  pushes the  $D_{light}$  to remove rain from light regions. Finally, both of the intermediate deraining images are concatenated with the original rainy image to generate the final deraining image via a global decoder, denoted as:

$$I_o = D_{global}(F^+, F^-, I_i), \quad (8)$$

where  $I_o$  is the derained image. We use MSE to update the model as

$$\mathcal{L}_{global} = \sum_{x=1}^X \sum_{y=1}^Y I_{c(x,y)} - I_{o(x,y)}. \quad (9)$$

The final loss function of the DAM contains  $\mathcal{L}_{att}$ ,  $\mathcal{L}_{heavy}$ ,  $\mathcal{L}_{light}$  and  $\mathcal{L}_{global}$ , which is defined as,

$$\mathcal{L}_{DAM} = \alpha \cdot \mathcal{L}_{att} + \beta_1 \cdot \mathcal{L}_{heavy} + \beta_2 \cdot \mathcal{L}_{light} + \mathcal{L}_{global}, \quad (10)$$

where  $\alpha$ ,  $\beta_1$  and  $\beta_2$  are three parameters to balance different loss functions, respectively.

$D_{heavy}$  and  $D_{light}$  share a similar recurrent structure, including one CNN layer, five residual blocks and another CNN layer, as Fig. 4 shows. Table 2 provides the detailed configurations. The  $D_{global}$  has a similar architecture, i.e., a recurrent structure of one CNN layer, two residual blocks, and one additional CNN layer. Its network configurations can refer to Table 3.

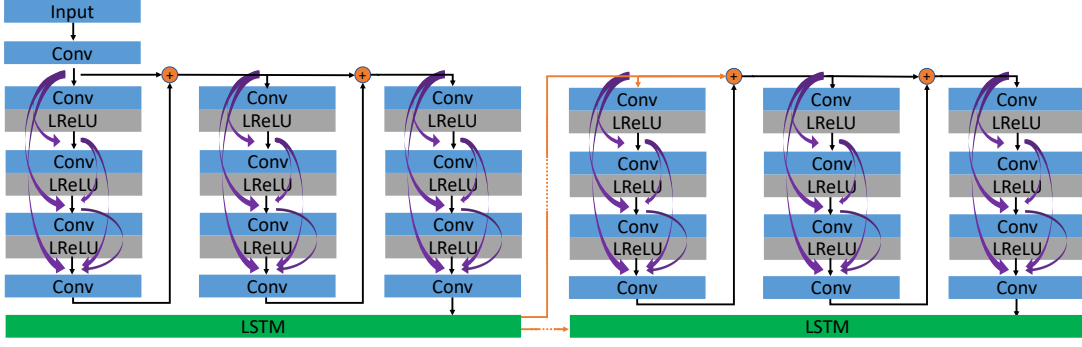


Fig. 3. The architecture of  $E$ . The detail of the architecture is shown in Table. 1. The arrows in the ResBlock mean the residual learning.

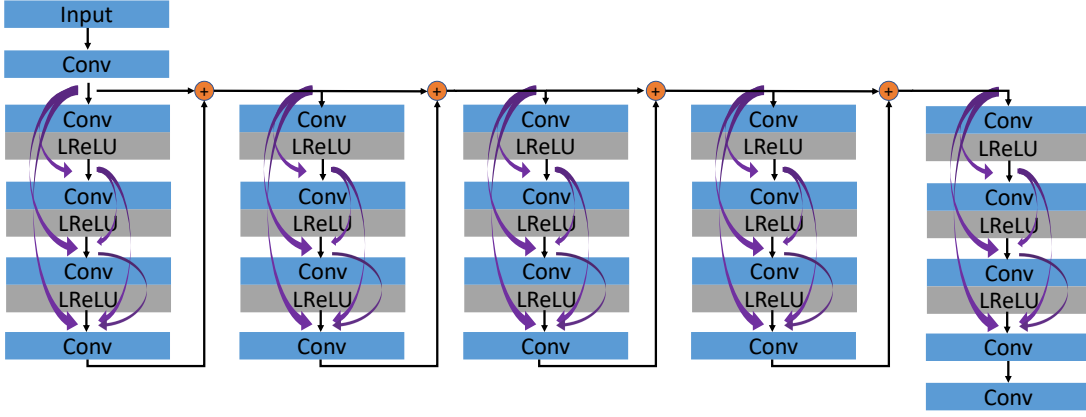


Fig. 4. The architecture of  $D_{heavy}$  and  $D_{light}$ . The detail of the architecture is shown in Table. 2. The arrows in the ResBlock mean the residual learning.

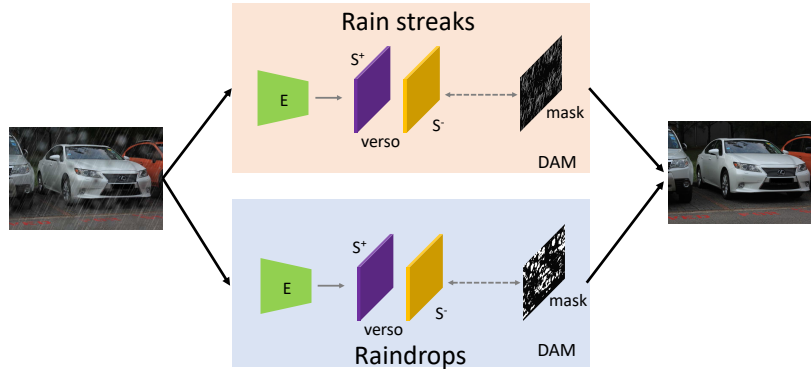


Fig. 5. The framework of DAiAM for joint rain streak and raindrop removal. DAiAM takes a rainy image as input to capture attention maps for rain streaks and raindrops via two DAMs. Then the outputs of them are concatenated to generate final deraining result.

TABLE 3  
Configurations of the proposed  $D_{global}$ .

| layers    | Kernel size  | output | operations | skip connection |
|-----------|--------------|--------|------------|-----------------|
| CNN1      | $3 \times 3$ | 64     | -          | ResBlock1       |
| ResBlock1 | $3 \times 3$ | 64     | LReLU      | ResBlock2       |
| ResBlock2 | $3 \times 3$ | 64     | LReLU      | -               |
| CNN2      | $3 \times 3$ | 3      | -          | -               |

### 3.4 Dual Attention-in-Attention Model

As discussed above, raindrops and rain streaks are two different rain types and usually appear simultaneously in the real world. In this case, rain removal becomes a more challenging problem. Previous methods [53] often focus on removing one type of rain from rainy images. To simultaneously remove both of them, a Dual Attention-in-Attention Model, **DAiAM**, is proposed. The first attention-in-attention model is responsible for heavy rain raindrops and rain streaks, while the second attention-in-attention model is responsible for light rain raindrops and rain streaks. As a

comparison, in the Dual Attention Model (DAM), the first attention model is responsible for heavy rain raindrops or rain streaks, while the second attention model is responsible for light rain raindrops or rain streaks. Among the Attention-in-Attention model, the first attention is responsible for rain types, while second attention is responsible for the rain intensities.

Fig. 5 shows the core idea of DAiAM. Image of raindrops and rain streaks is fed into our proposed DAiAM, which has two branches to pay attention to removal of raindrops and rain streaks, respectively. The branch for raindrop removal is similar to the method of removing rain streaks, which is represented in the above based on DAM. The main difference is that the attention loss function  $\mathcal{L}_{att}$  is calculated based on the mask of raindrops, rather than rain streaks. In this way, the DAiAM first pays attention to two kinds of rain variations, and then focuses on two kinds of rain intensity in different branches. The final loss function of DAiAM is defined as,

$$\mathcal{L}_{DAiAM} = \mathcal{L}_{streak} + \mathcal{L}_{drop} + \mathcal{L}_{global}, \quad (11)$$

where  $\mathcal{L}_{drop}$  and  $\mathcal{L}_{streak}$  are two loss functions to remove raindrops and streaks, respectively. The loss functions of them are

$$\mathcal{L}_{streak} = \alpha \cdot \mathcal{L}_{att}^{streak} + (\beta_1 \cdot \mathcal{L}_{heavy}^{streak} + \beta_2 \cdot \mathcal{L}_{light}^{streak}), \quad (12)$$

$$\mathcal{L}_{drop} = \alpha \cdot \mathcal{L}_{att}^{drop} + (\beta_1 \cdot \mathcal{L}_{heavy}^{drop} + \beta_2 \cdot \mathcal{L}_{light}^{drop}), \quad (13)$$

where  $\alpha$ ,  $\beta_1$  and  $\beta_2$  are parameters to balance different loss terms. The attention loss function  $\mathcal{L}_{att}^{drop}$  and  $\mathcal{L}_{att}^{streak}$  are calculated based on the masks of raindrops and rain streaks, respectively.

Finally, the proposed structure implements the fusion operation of two branches, as is achieved via  $D_{global}$ . Similar to the method for rain streak removal in Fig. 2, we use a parallel architecture to detect raindrop and then extract features (F+ and F-). All of them are concatenated and then fed into  $D_{global}$  to generate final derained images.

### 3.5 Differential-Driven DAiAM (D-DAiAM)

Rain has different intensities and various types. Images exhibiting both rain streaks and raindrops also pose increasing difficulty of deraining. Deep deraining methods can remove rain to some extent and transfer the heavy-rain images to light-rain ones [15], [64]. However, the performance of a single model is often limited. Simply increasing neural network depth is easy to exhaust the potential and difficult to further improve the performance of rain removal, even for some special heavy rain removal methods [65].

Li *et al.* [53] show that light rainy images are easier to derain. Therefore, we propose a differential-driven dual attention-in-attention model, **D-DAiAM**, to remove various kinds of rain. Different from most methods [53] which aim to directly derive final deraining images via increasing the depth or width of a single model, we aim to remove heavy rains via transferring heavy rain to light rain and then to no rain in multiple stages. In each stage, we use a DAiAM to generate better visible deraining images and attention information driven by the *differential between the*

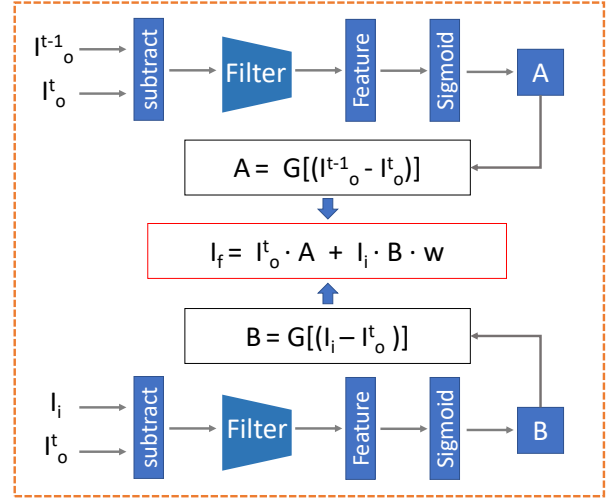


Fig. 6. The illustration of the differential-driven module. It consists of three streams, *i.e.*, two differential streams and a fusion stream. The FilterNet inside it pointedly selects key regions to help remove rain in the next stage.

TABLE 4  
Performance of different model structures on the Rain Streak dataset [14] in terms of PSNR and SSIM.

| Methods        | PSNR         | SSIM         |
|----------------|--------------|--------------|
| GMM [23]       | 15.05        | 0.425        |
| DDN [12]       | 21.92        | 0.764        |
| RGN [11]       | 25.25        | 0.841        |
| JORDER [14]    | 26.54        | 0.835        |
| RESCAN [13]    | 28.88        | 0.866        |
| PReNet [66]    | 29.46        | 0.899        |
| <b>DAM</b>     | <b>29.99</b> | <b>0.905</b> |
| <b>D-DAiAM</b> | <b>30.35</b> | <b>0.907</b> |

*current output and original input*, and the *differential between the current and previous outputs*.

Specifically, this process is conducted via a differential-driven module. As shown in Fig. 6, we calculate two types of differential. One is the difference between the current output  $I_o^t$  and the original input  $I_i$ . By comparing these two items, the differential is able to guide the following stage to focus on the remaining rainy regions in  $I_o^t$ . The other is the difference between the current and the previous outputs ( $I_o^t$  and  $I_o^{t-1}$ ). This differential leads the next stage to pay special attention to regions of the current output  $I_o^t$  which are not handled well in the current stage.

Based on these two kinds differential, we employ two *FilterNets* to generate soft maps  $A$  and  $B$  for our purpose, *i.e.*, the mark of regions needing special attention in the next stage. The FilterNet includes three convolutional layers with  $2 \times 2$  kernels to perceive local regions, rather than directly using the input differences. We apply these two soft maps to the original input  $I_i$  and the current output  $I_o^t$  and fuse them, as defined in

$$I_f = I_o^t \otimes A + I_i \otimes B \cdot w, \quad (14)$$

where  $w$  balances different types of differential.

The coarsest-level DAiAM locates in the begin of D-DAiAM. A latent deraining image is generated at the end of

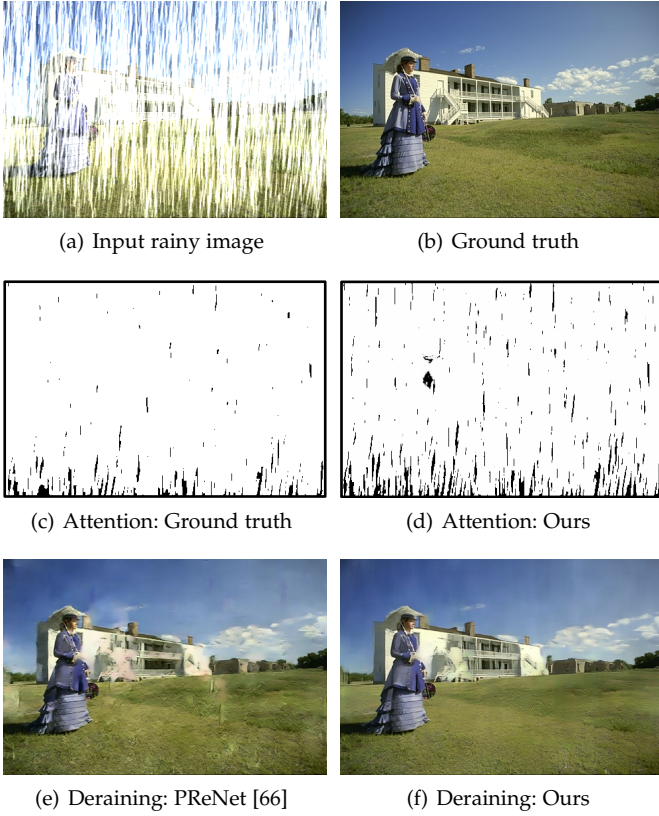


Fig. 7. Heavy rain streak removal results of sample images from Rain Streak dataset [14].

this stage. Even there still exists rain, the generated deraining image exhibits lighter rain. Then, the information from the coarsest level output is addressed by the differential-driven module, and then fed into finer-level network (which has a similar architecture as DAiAM) with deraining images. The final derained image is the output of the last DAiAM. The objective function to update the D-DAiAM is denoted as:

$$\mathcal{L} = \sum_{t=1}^N ||I_o^t - I_c||, \quad (15)$$

where  $I_o^t$  is the derained image in the  $t$ -th stage and  $I_c$  is the ground-truth image.

## 4 EXPERIMENTS

We first introduce the implementation details. Then the performance of rain streak removal and raindrop removal is compared with the state-of-the-art methods on two public datasets. We develop a new dataset of joint rain streaks and raindrops and test different deraining methods on it. Further, ablation study is carried out to verify the components of our proposal. Finally, the application of deraining in real-world scenarios is demonstrated.

### 4.1 Implementation Details

The weights of the proposed networks are initialized with Gaussian distribution with zero mean and a standard deviation of 0.01. The parameters are updated after a mini-batch

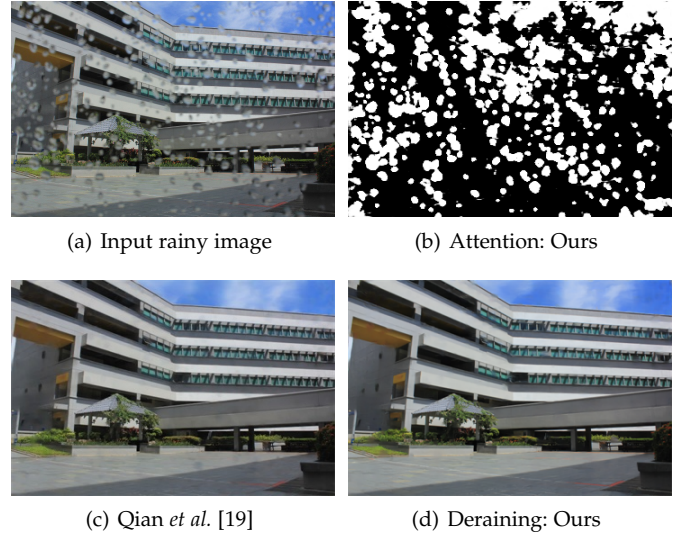


Fig. 8. Raindrop removal results on sample images from the Qian *et al.* [19] Raindrop dataset.

TABLE 5  
Performance of different model structures on the Raindrop dataset [19] in terms of PSNR and SSIM.

| Methods                 | PSNR         | SSIM          |
|-------------------------|--------------|---------------|
| DID-MDN [15]            | 24.76        | 0.7930        |
| DDN [12]                | 25.23        | 0.8366        |
| JORDER [14]             | 27.52        | 0.8239        |
| Qian <i>et al.</i> [19] | 30.55        | 0.9023        |
| Quan <i>et al.</i> [50] | 30.86        | 0.9263        |
| Hao <i>et al.</i> [52]  | 30.17        | 0.9128        |
| <b>DAM</b>              | <b>30.26</b> | <b>0.9137</b> |
| <b>D-DAM</b>            | <b>30.63</b> | <b>0.9268</b> |

of size 4 in each iteration. In the training stage,  $112 \times 112$  patches at random locations of an image are cropped to increase the number of training samples. We also randomly flip training images (horizontally) to further augment the training set. The models are trained under a learning rate which starts with a value of  $10^{-4}$  and reduces to  $10^{-6}$  after the training has converged. The hyper-parameters  $\alpha$ ,  $\beta_1$ ,  $\beta_2$  and  $w$  are set as 0.8, 1.0, 0.3 and 0.5, respectively. To reduce training time, we apply one differential-driven module in our practice. The encoder  $E$  contains three residual blocks [67] and one LSTM layer.  $D_{heavy}$  and  $D_{light}$  contain one CNN layer, five residual blocks and another CNN layer.  $D_{global}$  contains two residual blocks and one CNN layer. The size of all the kernels in this work is set to  $3 \times 3$ . ReLU function is adopted after convolution operation except the last CNN layer in each structure.

### 4.2 Results on Rain Streak Dataset

Yang *et al.* [14] build a dataset of heavy rain streaks, named as Rain100H. In order to synthesize heavy rain, they apply two different methods, including the photo-realistic rendering techniques proposed by [68] and directly adding simulated sharp line streaks to clear images. The Rain100H dataset consists of 1,800 and 100 pairs of images for training and testing, respectively. [66] removes some training images with the same background contents as testing images. Table

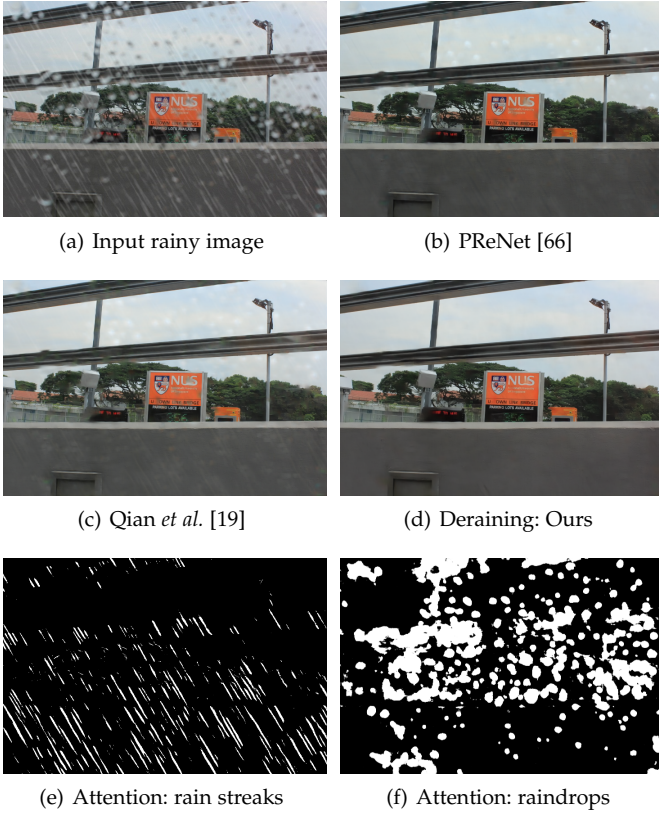


Fig. 9. Rain streak and raindrop removal results on sample images from JRSRD dataset.

4 reports the comparison results with the state-of-the-art rain streak removal methods, including GMM [23], DDN [12], RGN [11], JORDER [14], RESCAN [13] and PReNet [66]. Note that, as the rainy images contain only rain streaks, our full method D-DAiAM degrades as D-DAM in this scenery. Specially, this is only a dual-attention model which focuses on heavy and light rain streak regions. The quantitative results demonstrate the advance of our proposed method over the existing methods. Fig. 7 shows the qualitative deraining results and the associated attention maps. Our result is better than that of PReNet [66]. The latent attention map is also close to the ground truth.

### 4.3 Results on Raindrop Dataset

Qian *et al.* [19] capture 1,119 pairs of images with different background scenes and raindrops. They use two glasses to model the raindrops. One is clean to capture GT images. The other is sprayed with water to generate corresponding rainy version. The training set and testing set A include 861 and 58 pairs, respectively. In order to verify the performance of the propose method, we compare with state-of-the-art deraining methods. As mentioned before, our method becomes D-DAM in this case. Table 4 presents the results of DID-MDN [15], DDN [12], JORDER [14], Qian *et al.* [19] and ours, respectively. The deraining results and attention maps are provided in Fig. 8. Both the quantitative and the qualitative results reveal that our method is more advanced.

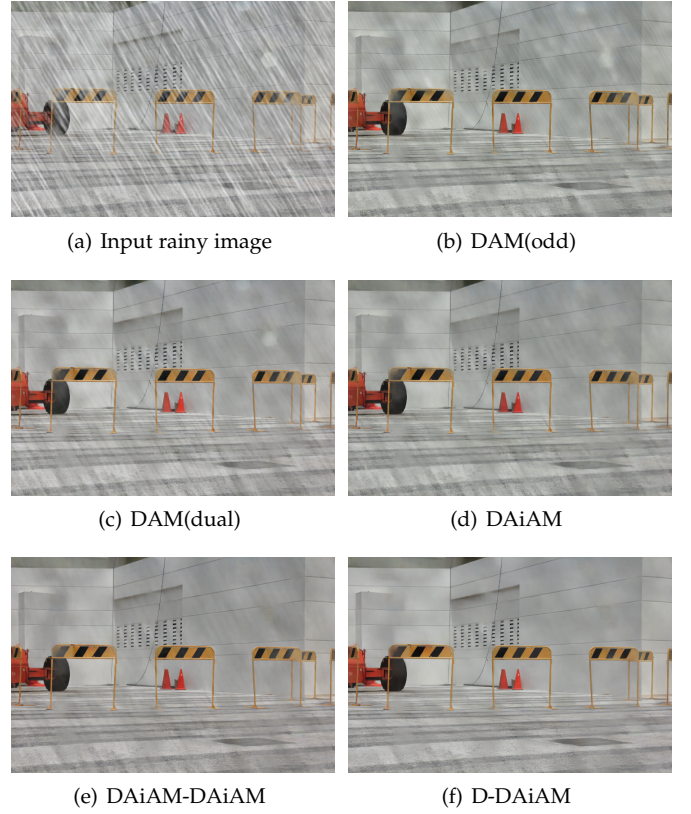


Fig. 10. Ablation study results of rain streak and raindrop removal on sample images from JRSRD dataset. Zoom-in for details.

### 4.4 Results on the Joint Rain Streak and Raindrop Dataset

There are many rain removal datasets for image deraining [14], [15], [19], [53], [63]. However, most of them focus on either rain streaks or raindrops. To this end, we synthesize a new joint rain streak and raindrop (JRSRD) dataset to evaluate the performance of different methods for removing both of them. Specially, the JRSRD training set contains 3,444 synthetic rainy images, generated using images with raindrops from [19]. We synthesize four images with different intensity levels of rain streaks for each image via Photoshop, which provides official methods to synthesize rain streak. In addition, many previous popular datasets are synthesized based on this strategy like Rain800 [35], Rain12000 [15] and Rain14000 [12]. The noise levels are set between 20% and 60% to model various intensity. The JRSRD testing set contains 232 pairs. The rainy images in our synthesized dataset contain both rain streaks and raindrops. Therefore, we apply DAiAM to remove rain. The performance compared with three current deraining methods is shown in Table 6. The “Qian *et al.* [19] + PReNet [66]” means that we first use Qian *et al.* [19] to remove raindrops from rainy images, and then use PReNet [66] to remove rain streaks. The “PReNet [66] + Qian *et al.* [19]” represents the reverse order. Params means the parameters of different deep deraining networks. Time is the inference time. FLOPs means floating point operations. All of them are re-trained on the proposed JRSRD dataset. Our proposed method beats these CNN-based methods on the task of joint

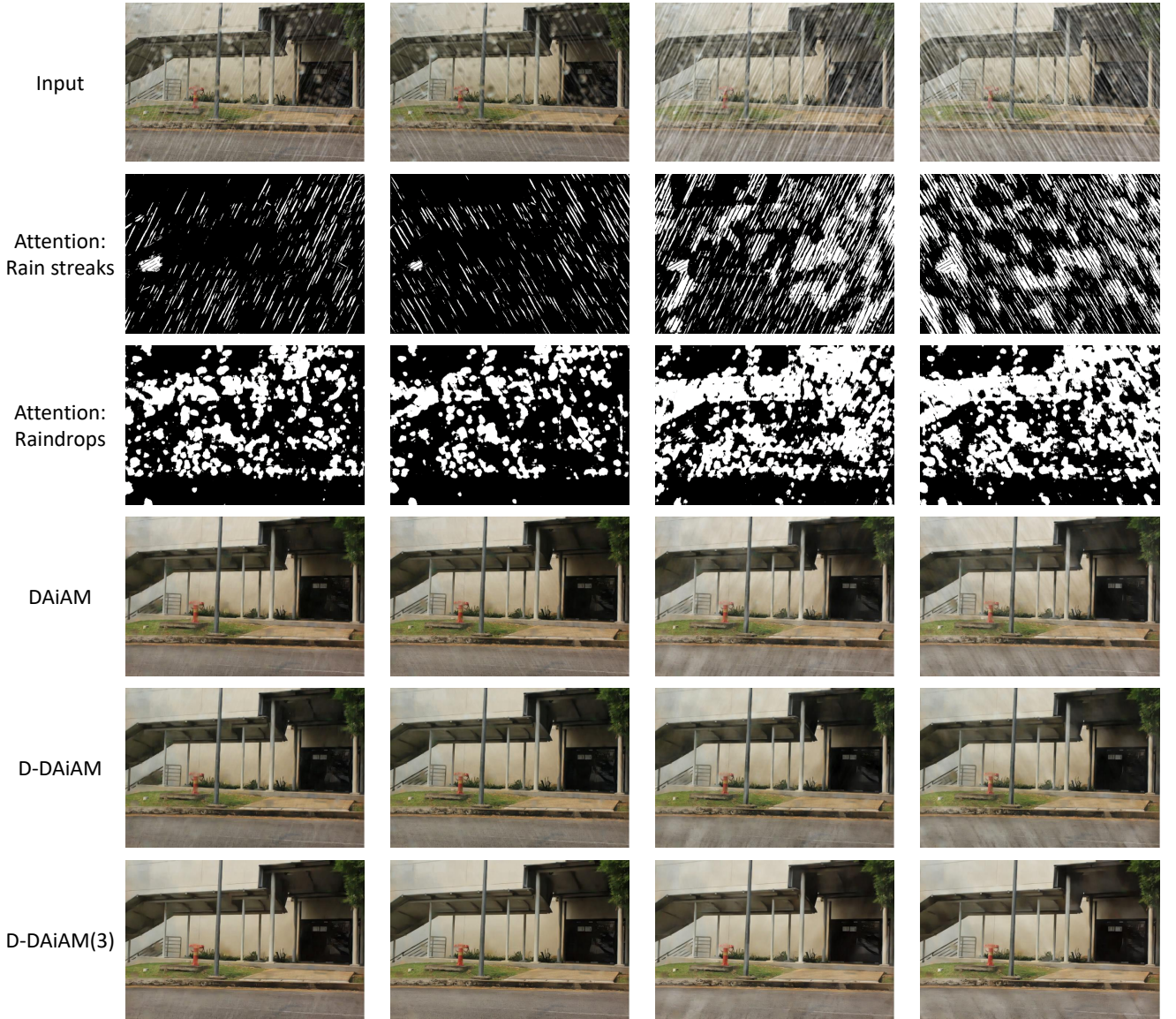


Fig. 11. Ablation study results of rain streaks and raindrop removal on different rain intensities. From top to bottom are the input, our attention maps for rain streaks and raindrops, DAiAM, D-DAiAM and D-DAiAM(3), respectively. Please zoom-in for details.

rain streak and raindrop removal. Exemplar visual results are given in Fig. 9, suggesting that the proposed method is capable of generating cleaner images.

#### 4.5 Ablation Study

To demonstrate the effectiveness of DAM, DAiAM and differential-driven module, we compare these structures with several variant structures. Different from previous methods which merely focus on heavy rain, the proposed DAM generates two feature maps paying attention to heavy rain and light rain, respectively. Thus we compare to model without attention, DAM(zero), and the models with one or two attention maps, which are named as DAM(odd) and DAM(dual), respectively. Then, we compare the performance of the proposed dual attention-in-attention model, DAiAM, which can jointly perceive rain streaks and raindrops. The D-DAiAM is the model which removes rain

using the differential-driven module. We compare it with the method directly connecting two DAiAM, termed as DAiAM-DAiAM. We also aggressively use two differential-driven modules in D-DAiAM(3). Table 7 shows the performance of them in terms of PSNR and SSIM. Apparently, the counterpart without attention performs worst. Using attention of heavy rain improves the performance, as demonstrated by DAM(odd). While dual attention mechanism further improves the results. The DAiAM outperforms these three by simultaneously removing both raindrops and rain streaks. Directly connecting two DAiAM as DAiAM-DAiAM indeed boosts the values, while the improvement is not as significant as that of the proposed D-DAiAM. Fig. 10 presents exemplar visual deraining results, which also suggest the effectiveness of the proposed method.

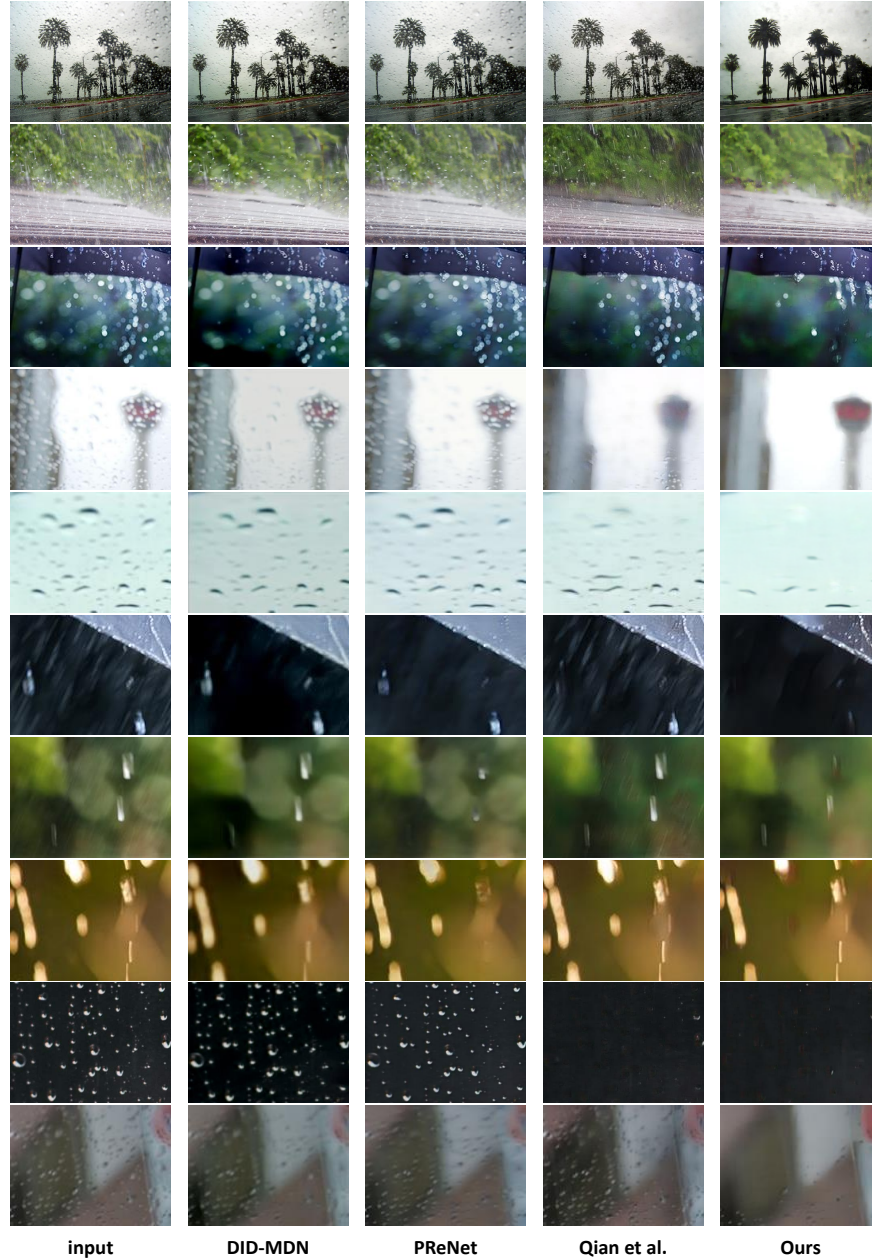


Fig. 12. The performance of different methods on real-world rainy images. From the left to right are the input, DID-MDN [15], PReNet [66], Qian *et al.* [19] and ours. DID-MDN and PReNet are two rain streak removal methods, which only work on removing rain streaks. Qian *et al.* is a raindrop removal method, which does not work on rain streak removal. Our proposed method achieves better performance by removing rain streaks and raindrops simultaneously on real-world rainy images.

TABLE 6

Performance of different model structures on the JRSRD dataset.

| Methods                               | PSNR         | SSIM         | Params/Time/FLOPs    |
|---------------------------------------|--------------|--------------|----------------------|
| RESCAN [13]                           | 21.05        | 0.768        | 0.15M/0.07s/2.46E+11 |
| PReNet [66]                           | 23.29        | 0.789        | 0.17M/0.10s/3.37E+11 |
| Qian <i>et al.</i> [19]               | 22.49        | 0.772        | 6.00M/0.13s/6.83E+11 |
| Qian <i>et al.</i> [19] + PReNet [66] | 23.89        | 0.796        | 6.17M/0.23s/10.2E+11 |
| PReNet [66] + Qian <i>et al.</i> [19] | 23.68        | 0.793        | 6.17M/0.23s/10.2E+11 |
| <b>DAiAM</b>                          | <b>24.67</b> | <b>0.819</b> | 3.60M/0.13s/8.90E+11 |
| <b>D-DAiAM</b>                        | <b>25.26</b> | <b>0.825</b> | 7.20M/0.28s/17.8E+11 |

TABLE 7

Ablation study on the JRSRD dataset in terms of PSNR and SSIM.

| Methods     | PSNR  | SSIM  |
|-------------|-------|-------|
| DAM(zero)   | 21.97 | 0.729 |
| DAM(odd)    | 23.41 | 0.791 |
| DAM(dual)   | 24.15 | 0.806 |
| DAiAM       | 24.67 | 0.819 |
| DAiAM-DAiAM | 24.84 | 0.823 |
| D-DAiAM     | 25.26 | 0.825 |
| D-DAiAM(3)  | 25.68 | 0.833 |

#### 4.6 Performance of Deraining for Different Rain Intensities

In this section, we evaluate the proposed method for removing different types of rain with different levels of intensity.

Fig. 11 present exemplar visual deraining results of our full method D-DAiAM and its two variants, DAiAM and D-DAiAM(3). Observing from the comparison, we have the following findings. 1) It is more difficult to remove heavy rain than the light rain. All the methods exhibit more artifacts when rain intensity becomes heavier. 2) Compared with DAiAM, our D-DAiAM and D-DAiAM(3) refine the performance of rain removal, especially for heavy rain, which shows the effectiveness of the differential-driven module. 3) The proposed method is capable of focusing on rain streaks and raindrops simultaneously, which shows the effectiveness of the dual attention-in-attention model. 4) Meanwhile, during the generation of attention maps, the rain streaks and raindrops can affect each other, which will cause erroneous attention maps. This shows the necessity of focusing on heavy-aware and light-aware regions simultaneously.

#### 4.7 Deployment in Real World

The proposed method is also evaluated on real-world images from the Internet. Fig. 12 shows the visual deraining results of different methods. DID-MDN [15] and PReNet [66] are two state-of-the-art methods for rain streak removal, and Qian *et al.* [19] is one of the best methods to remove raindrops [53]. The proposed method achieves better performance on removing both rain streaks and raindrops than other methods, due to the proposed dual attention-in-attention mechanism. Rain streaks and raindrops are focused simultaneously via this mechanism-based network. Inside the proposed network, there are two well-designed DAMs, which also focus on local regions with different rainy intensities. The intensity-aware attention maps enable better removal of rain in different regions. The compared methods can only remove either raindrops (*e.g.*, [19]) or rain streaks (*e.g.*, [15] and [66]). Consider that Fig. 12 has shown that the proposed method achieves significantly better results than other models, thus we do not conduct quantitative evaluation such as recognition score or user study.

## 5 CONCLUSION

In this paper, we tackle the problem of joint removal of raindrops and rain streaks. A dual attention-in-attention model, DAiAM, is presented to focus on raindrops and rain streaks simultaneously. Inside DAiAM, we propose a dual attention model, DAM. The proposed DAM learns two intensity-aware maps to remove rain from heavy and light rainy regions. We further introduce a differential-driven module to optimize the deraining process. Experimental results have demonstrated that our method performs best against the state-of-the-art methods and is capable of deraining well in real-world scenarios. In the future, we will consider generating images with purely heavy or purely light rains for supervision to help remove rains.

## ACKNOWLEDGMENT

This work is supported by Fund project of Jimei University (No. zp2020042), Xiamen Key Laboratory of Marine Intelligent Terminal R&D and Application (No. B18208).

## REFERENCES

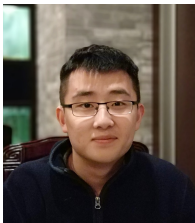
- [1] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015.
- [2] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017.
- [3] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proceedings of the IEEE international conference on computer vision*, 2015.
- [4] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 2, pp. 316–322, 2005.
- [5] G. Yang, X. Song, C. Huang, Z. Deng, J. Shi, and B. Zhou, "DrivingStereo: A large-scale dataset for stereo matching in autonomous driving scenarios," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [6] B. Li, W. Ouyang, L. Sheng, X. Zeng, and X. Wang, "Gs3d: An efficient 3d object detection framework for autonomous driving," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [7] S.-H. Sun, S.-P. Fan, and Y.-C. F. Wang, "Exploiting image structural similarity for single image rain removal," in *IEEE International Conference on Image Processing*, 2014.
- [8] L.-W. Kang, C.-W. Lin, and Y.-H. Fu, "Automatic single-image-based rain streaks removal via image decomposition," *IEEE transactions on image processing*, vol. 21, no. 4, pp. 1742–1755, 2011.
- [9] Y.-L. Chen and C.-T. Hsu, "A generalized low-rank appearance model for spatio-temporally correlated rain streaks," in *Proceedings of the IEEE international conference on computer vision*, 2013.
- [10] X. Zhang, H. Li, Y. Qi, W. K. Leow, and T. K. Ng, "Rain removal in video by combining temporal and chromatic properties," in *2006 IEEE international conference on multimedia and expo*. IEEE, 2006, pp. 461–464.
- [11] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, "Clearing the skies: A deep network architecture for single-image rain removal," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2944–2956, 2017.
- [12] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [13] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha, "Recurrent squeeze-and-excitation context aggregation net for single image deraining," in *European Conference on Computer Vision*, 2018.
- [14] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [15] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [16] M. Roser and A. Geiger, "Video-based raindrop detection for improved image registration," in *Proceedings of the IEEE international conference on computer vision Workshops*, 2009.
- [17] M. Roser, J. Kurz, and A. Geiger, "Realistic modeling of water droplets for monocular adherent raindrop recognition using bezier curves," in *Asian Conference on Computer Vision*, 2010.
- [18] D. Eigen, D. Krishnan, and R. Fergus, "Restoring an image taken through a window covered with dirt or rain," in *Proceedings of the IEEE international conference on computer vision*, 2013.
- [19] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [20] P. C. Barnum, S. Narasimhan, and T. Kanade, "Analysis of rain and snow in frequency space," *International journal of computer vision*, vol. 86, no. 2, pp. 256–274, 2010.
- [21] 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 Transactions on multimedia*, vol. 16, no. 1, pp. 83–93, 2013.
- [22] Y. Luo, Y. Xu, and H. Ji, "Removing rain from a single image via discriminative sparse coding," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3397–3405.
- [23] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, "Rain streak removal using layer priors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

- [24] Y. Chang, L. Yan, and S. Zhong, "Transformed low-rank model for line pattern noise removal," in *Proceedings of the IEEE international conference on computer vision*, 2017.
- [25] L. Zhu, C.-W. Fu, D. Lischinski, and P.-A. Heng, "Joint bi-layer optimization for single-image rain streak removal," in *Proceedings of the IEEE international conference on computer vision*, 2017.
- [26] L. Zhu, Z. Deng, X. Hu, H. Xie, X. Xu, J. Qin, and P.-A. Heng, "Learning gated non-local residual for single-image rain streak removal," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.
- [27] X. Hu, L. Zhu, T. Wang, C.-W. Fu, and P.-A. Heng, "Single-image real-time rain removal based on depth-guided non-local features," *IEEE Transactions on Image Processing*, vol. 30, pp. 1759–1770, 2021.
- [28] Y. Wang, Y. Song, C. Ma, and B. Zeng, "Rethinking image de-raining via rain streaks and vapors," in *European Conference on Computer Vision*. Springer, 2020, pp. 367–382.
- [29] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.
- [30] K. Zhang, W. Luo, Y. Zhong, L. Ma, W. Liu, and H. Li, "Adversarial spatio-temporal learning for video deblurring," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 291–301, 2018.
- [31] K. Zhang, W. Luo, Y. Zhong, L. Ma, B. Stenger, W. Liu, and H. Li, "Deblurring by realistic blurring," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2737–2746.
- [32] L. Zheng, Y. Li, K. Zhang, and W. Luo, "T-net: Deep stacked scale-iteration network for image dehazing," *arXiv preprint arXiv:2106.02809*, 2021.
- [33] K. Zhang, W. Luo, B. Stenger, W. Ren, L. Ma, and H. Li, "Every moment matters: Detail-aware networks to bring a blurry image alive," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 384–392.
- [34] K. Zhang, R. Li, Y. Yu, W. Luo, C. Li, and H. Li, "Deep dense multi-scale network for snow removal using semantic and geometric priors," *arXiv preprint arXiv:2103.11298*, 2021.
- [35] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE transactions on circuits and systems for video technology*, vol. 30, no. 11, pp. 3943–3956, 2019.
- [36] K. Zhang, W. Luo, Y. Yu, W. Ren, F. Zhao, C. Li, L. Ma, W. Liu, and H. Li, "Beyond monocular deraining: Parallel stereo deraining network via semantic prior," *arXiv preprint arXiv:2105.03830*, 2021.
- [37] K. Zhang, W. Luo, W. Ren, J. Wang, F. Zhao, L. Ma, and H. Li, "Beyond monocular deraining: Stereo image deraining via semantic understanding," in *European Conference on Computer Vision*. Springer, 2020, pp. 71–89.
- [38] H. Wang, Q. Xie, Q. Zhao, and D. Meng, "A model-driven deep neural network for single image rain removal," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.
- [39] S. Li, W. Ren, F. Wang, I. B. Araujo, E. K. Tokuda, R. H. Junior, R. M. Cesar-Jr, Z. Wang, and X. Cao, "A comprehensive benchmark analysis of single image deraining: Current challenges and future perspectives," *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1301–1322, 2021.
- [40] M. Li, Q. Xie, Q. Zhao, W. Wei, S. Gu, J. Tao, and D. Meng, "Video rain streak removal by multiscale convolutional sparse coding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [41] J. Liu, W. Yang, S. Yang, and Z. Guo, "D3r-net: Dynamic routing residue recurrent network for video rain removal," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 699–712, 2018.
- [42] —, "Erase or fill? deep joint recurrent rain removal and reconstruction in videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [43] J. Chen, C.-H. Tan, J. Hou, L.-P. Chau, and H. Li, "Robust video content alignment and compensation for rain removal in a cnn framework," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [44] W. Yang, J. Liu, and J. Feng, "Frame-consistent recurrent video deraining with dual-level flow," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [45] K. Zhang, D. Li, W. Luo, W.-Y. Lin, F. Zhao, W. Ren, W. Liu, and H. Li, "Enhanced spatio-temporal interaction learning for video deraining: A faster and better framework," *arXiv preprint arXiv:2103.12318*, 2021.
- [46] H. Kurihata, T. Takahashi, I. Ide, Y. Mekada, H. Murase, Y. Tamatsu, and T. Miyahara, "Rainy weather recognition from in-vehicle camera images for driver assistance," in *IEEE Proceedings. Intelligent Vehicles Symposium*, 2005. IEEE, 2005, pp. 205–210.
- [47] A. Yamashita, Y. Tanaka, and T. Kaneko, "Removal of adherent waterdrops from images acquired with stereo camera," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2005, pp. 400–405.
- [48] A. Yamashita, I. Fukuchi, and T. Kaneko, "Noises removal from image sequences acquired with moving camera by estimating camera motion from spatio-temporal information," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 3794–3801.
- [49] S. You, R. T. Tan, R. Kawakami, Y. Mukaigawa, and K. Ikeuchi, "Adherent raindrop modeling, detection and removal in video," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 9, pp. 1721–1733, 2015.
- [50] Y. Quan, S. Deng, Y. Chen, and H. Ji, "Deep learning for seeing through window with raindrops," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019.
- [51] S. Alletto, C. Carlin, L. Rigazio, Y. Ishii, and S. Tsukizawa, "Adherent raindrop removal with self-supervised attention maps and spatio-temporal generative adversarial networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.
- [52] Z. Hao, S. You, Y. Li, K. Li, and F. Lu, "Learning from synthetic photorealistic raindrop for single image raindrop removal," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.
- [53] S. Li, I. B. Araujo, W. Ren, Z. Wang, E. K. Tokuda, R. H. Junior, R. Cesar-Junior, J. Zhang, X. Guo, and X. Cao, "Single image deraining: A comprehensive benchmark analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [54] J. Ba, V. Mnih, and K. Kavukcuoglu, "Multiple object recognition with visual attention," *arXiv preprint arXiv:1412.7755*, 2014.
- [55] K. Gregor, I. Danihelka, A. Graves, D. J. Rezende, and D. Wierstra, "Draw: A recurrent neural network for image generation," *arXiv preprint arXiv:1502.04623*, 2015.
- [56] T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, and Z. Zhang, "The application of two-level attention models in deep convolutional neural network for fine-grained image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [57] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *International Conference on Machine Learning*, 2015.
- [58] Q. You, H. Jin, Z. Wang, C. Fang, and J. Luo, "Image captioning with semantic attention," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [59] S. Chen, X. Tan, B. Wang, and X. Hu, "Reverse attention for salient object detection," in *European Conference on Computer Vision*, 2018.
- [60] Y. Lv, J. Zhang, Y. Dai, A. Li, B. Liu, N. Barnes, and D.-P. Fan, "Simultaneously localize, segment and rank the camouflaged objects," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11 591–11 601.
- [61] A. Li, J. Zhang, Y. Lv, B. Liu, T. Zhang, and Y. Dai, "Uncertainty-aware joint salient object and camouflaged object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 071–10 081.
- [62] Y. Mao, J. Zhang, Z. Wan, Y. Dai, A. Li, Y. Lv, X. Tian, D.-P. Fan, and N. Barnes, "Transformer transforms salient object detection and camouflaged object detection," *arXiv preprint arXiv:2104.10127*, 2021.
- [63] T. Wang, X. Yang, K. Xu, S. Chen, Q. Zhang, and R. W. Lau, "Spatial attentive single-image deraining with a high quality real rain dataset," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [64] X. Hu, C.-W. Fu, L. Zhu, and P.-A. Heng, "Depth-attentional features for single-image rain removal," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [65] R. Li, L.-F. Cheong, and R. T. Tan, "Heavy rain image restoration: Integrating physics model and conditional adversarial learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

- [66] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng, "Progressive image deraining networks: a better and simpler baseline," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [67] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [68] K. Garg and S. K. Nayar, "Photorealistic rendering of rain streaks," *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3, pp. 996–1002, 2006.



**Kaihao Zhang** is currently pursuing the Ph.D. degree with the College of Engineering and Computer Science, The Australian National University, Canberra, ACT, Australia. His research interests focus on computer vision and deep learning. He has more than 20 referred publications in international conferences and journals, including CVPR, ICCV, ECCV, NeurIPS, AAAI, ACMMM, IJCV, TIP, TMM, etc.



**Dongxu Li** is a Ph.D candidate at The Australian National University. His research interests are mainly computer vision and deep learning, including visual sequence representation learning, vision-language learning and multi-modal learning. Before starting PhD, Dongxu obtained his Bachelor degree from The Australian National University with first-class honours in Computing.



**Wenhan Luo** is currently working as a senior researcher in the Tencent, China. His research interests include several topics in computer vision and machine learning, such as motion analysis (especially object tracking), image/video quality restoration, reinforcement learning. Before joining Tencent, he received the Ph.D. degree from Imperial College London, UK, 2016, M.E. degree from Institute of Automation, Chinese Academy of Sciences, China, 2012 and B.E. degree from Huazhong University of Science and Technol-

ogy, China, 2009.



**Wenqi Ren** is an Associate Professor in Institute of Information Engineering, Chinese Academy of Sciences, China. He received his Ph.D. degree from Tianjin University, Tianjin, China, in 2017. During 2015 to 2016, he was supported by China Scholarship Council and working with Prof. Ming-Husan Yang as a joint-training Ph.D. student in the Electrical Engineering and Computer Science Department, at the University of California at Merced. He received Tencent Rhino Bird Elite Graduate Program Scholarship

in 2017, MSRA Star Track Program in 2018. His research interests include image processing and related high-level vision problems.