ABSTRACT

Synchronous Rain streaks and Raindrops Removal (SR3) is a hard and challenging task, since rain streaks and raindrops are two wildly divergent real-world phenomena with different optical properties and mathematical distributions. As such, most existing data-driven deep Image Deraining (SID) methods only focus on one of them. Although there are only a few existing SR3 methods, they still suffer from blur textures and unknown noise in reality due to weak robustness and generalization ability. In this paper, we will propose a new and universal SID model with novel modules, termed Robust Attention Deraining Network (RadNet), with strong robustness and generalization ability that are reflected in two main aspects. (1) RadNet can restore different rain degenerations, including raindrops, rain streaks, or both; (2) RadNet can adapt to different data strategies, including single-type, superimposed-type, and blended-type. The generalization ability is also demonstrated by the performance of dealing with real rain images. Specifically, we first design a lightweight and robust attention module (RAM) with a universal attention mechanism for coarse rain removal, and then present a new deep refining module (DRM) with multi-scale blocks for precise rain removal. To solve the inconsistent labels of real scenario data, we also introduce a flow & warp module (FWM) into the network, which can greatly improve the performance of real scenario data via optical flow prediction and alignment. The whole process is unified in a network to ensure sufficient robustness and strong generalization ability. We evaluated the performance of our method under a variety of data strategies, and extensive experiments demonstrated that our RadNet could outperform other state-of-the-art SID methods.

CCS CONCEPTS

• Computing methodologies → Computer vision tasks; Image representation; Image deraining; Deep neural networks.

KEYWORDS

Synchronous rain streak and raindrop removal; robust attention deraining; deep refining module; flow & warp module.

ACM Reference Format:

Yanyan Wei, Zhao Zhang, Mingliang Xu, Richang Hong, Jicong Fan, and Shuicheng Yan. 2022. Robust Attention Deraining Network for Synchronous Rain Streaks and Raindrops Removal. In Proceedings of the 30th ACM International Conference on Multimedia (MM ’22), October 10–14, 2022, Lisboa, Portugal. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3503161.3547932
1 INTRODUCTION
Single image deraining (SID) is an important task in image restoration field [11, 19, 26], since it can affect outdoor computer vision and multimedia computing tasks. In reality, rain captured by cameras, surveillance cameras, and mobile devices has two major forms, i.e., rain streaks and raindrops. Due to different optical properties and mathematical distributions, rain streak removal and raindrop removal are usually treated as two different tasks. Since rain streaks appear more frequently in reality, more deep SID methods are proposed for rain streak removal [2, 28, 39]. However, the rain streak removal models usually cannot be well generalized to handle raindrop removal, and vice versa [16, 41]. To reduce the gap between the two tasks, researchers have recently proposed a new task called synchronous rain streaks and raindrops removal (SR\textsuperscript{3}) [21], which aims to remove both via a unified convolutional neural network (CNN). Next, we will describe the development of these three tasks.

1.1 Rain Streak Removal
Rain streak images create a rain mask before the true image. For heavy rains, they may cause a haze atmosphere due to light scattering, hence making the images blurring and haziness [29]. Separating the rain mask from the true image is an intuitive idea to solve this task. The rain streak removal problem can be modeled as

\[ O = B + S, \tag{1} \]

where \( O \) denotes a rain image that is decomposed into a rain streak component \( S \) and a clean background \( B \). This model is widely used in current SID methods [6, 15, 23, 27, 31, 35].

1.2 Raindrop Removal
Raindrop degradation occurs raindrop region is formed by rays of reflected light from a wider environment, which contains different imageries from those without raindrops. In most cases, the focus of camera is on the background scene, making the appearance of raindrops blur. The raindrop removal process is modeled as

\[ O = D + (1 - M) \odot B, \tag{2} \]

where \( M \) is a binary mask, and \( M(x) = 1 \) means the pixel \( x \) is a part of the rain region, and otherwise, it is a part of the background region. \( D \) is the effect brought by the raindrops, representing the complex mixture of the background information and the light reflected by the environment and passing through the raindrops adhered to a lens or windscreen. This model is also frequently used in current raindrop removal methods, such as [20, 22].

1.3 Synchronous Rain Streak and Raindrop Removal (SR\textsuperscript{3})
When raindrops and rain streaks appear synchronously in the same image, they will affect each other and complicate the restoration task due to the different optical properties of the rain streaks and raindrops. Moreover, it is difficult for deep learning to treat the mixed data with different distributions in a unified network, making the network hard to converge or even to perform (i.e., better on either rain streaks or raindrops). The SR\textsuperscript{3} task is modeled as

\[ O = \alpha D + (1 - M) \odot (B + S), \tag{3} \]

where \( \alpha \) is a global atmospheric lighting coefficient. Eqn. (3) is used in the recent SR\textsuperscript{3} method called complementary cascaded network (CCN) [21]. CCN tackled this hybrid task in a uniform network via a two-branch and two-stage strategy.

1.4 Remarks and Existed Problems
Current SID methods still suffer from unpleasant results over both synthetic and real-world data (see Figures 1 and 9). Therefore, we have to ask: what exactly degrades the quality of derained images based on different datasets and deraining tasks? We attempt to answer this question from two respects:

1) **Limited robustness.** Few existing SID methods consider the robustness based on different deraining tasks by employing a “divide-and-conquer” strategy. However, the degradations caused by rain streaks and raindrops are highly entangled, artificially forcing them apart will result in artifacts and blur textures. In other words, the widely-used divide-and-conquer strategy will reduce the robustness properties of deraining models.

2) **Weak generalization.** Rain streaks, raindrops, and synchronous rain streaks and raindrops often appear in reality. However, most of current deraining methods usually focus on one of them, making them suffer from limited generalization ability across various types of rain degradations in real world. That is, existing methods still cannot well handle complex patterns of rain in real scenario.

These two problems lead to a solution that we need a unified network that can synchronously remove rain streaks and raindrops to enhance the robustness and generalization in real scenario.

1.5 Our Contributions
In this paper, we mainly propose novel strategies to improve the robustness and generalization ability of deraining model over different data paradigms. Overall, the main contributions of this paper are summarized as follows:

- To improve the robustness of the SR\textsuperscript{3} task against different degenerations and data strategies, we propose a hybrid SID network, termed Robust Attention Deraining Network (RadNet). Specifically, we first design a lightweight network with a universal attention mechanism for coarse rain representation and removal, and then propose a deep neural network with multi-scale blocks for precise rain representation and removal.

- We propose a robust attention module (RAM), which can not only restore different degenerations, including raindrops, rain streaks, or both but also can adapt to different data strategies, including single-type, superimposed-type, and blended-type. Another new Deep Refining Module (DRM) with multi-stream CNN is also presented to perform refined rain removal.

- We use optical flow estimation and alignment in SID task, apparently due to the nature of RainDS\_real dataset itself. Since single image lacks frames compared with video deraining, it is groundbreaking to align the background properly, which enables our model to deliver promising results on real scenario data.

- Extensive experiments on synthetic and real-world data demonstrate the effectiveness of our RadNet. Up to 1dB-4dB and 0.1dB-0.2dB reconstruction advantage in terms of PSNR and SSIM is obtained, compared with the current SOTA SR\textsuperscript{3} methods.
Figure 2: The framework of our RadNet has three parts: (1) Flow & Warp Module (FWM) conducts optical estimation and alignment for RainDS_real; (2) Robust Attention Module (RAM) pays attention to various types of rain degradation; (3) Deep Refining Module (DRM) uses the coarse results obtained by RAM to refine the results, and then return to FWM again to obtain the final results.

Figure 3: Illustration of the intermediate outputs of our RadNet. From (a) to (d) are input, attention map by RAM, heat map of attention, and the refined result by DRM, respectively.

2 RELATED WORK

2.1 Single Image Deraining

We review some representative deep SID methods on rain streak removal, raindrop removal, and SR3, which are related to our method.

Rain Steak Removal Methods. Most of the existing SID methods mainly handle the task of rain streak removal. For example, JORDER [35] develops a multi-task deep learning architecture that learns the binary rain streak map, the appearance of rain streaks, and the clean background. PReNet [23] provides a better and simpler baseline deraining network by considering network architecture, input, output, and loss functions. Recently, RCDNet [25] utilizes the proximal gradient descent technique to design an iterative algorithm only containing simple operators for solving the model.

Raindrop Removal Methods. To restore an image from a corrupted input, the model in [5] predicts a clean output by a specialized form of CNN, with three layers and 512 neurons in each layer. AttentGAN [20] conducts a new raindrop dataset and proposes a GAN-based model to remove them by the injection of attention map into both generative and discriminative networks. A double attention mechanism is also introduced [22], which concurrently guides the CNN by shape-driven attention and channel re-calibration.

Synchronous Rain Steak and Raindrop Removal Methods. Recently, few researchers proposed approaches to solve this issue [21, 40]. For example, CCN [21] uses the NAS to find an optimal architecture adaptively. They also construct a new rain dataset, RainDS, which includes the rain images in different types and their corresponding rain-free ground-truth, including rain streak only, raindrop only, and both. DAiIAM [40] removes both rain streaks and raindrops via DAM, which respectively attends to the heavy and light rain regions. Further more, DAiIAM uses a differential-driven dual attention-in-attention model with a “heavy-to-light” scheme to remove rain via addressing the unsatisfying deraining regions. Unlike the two-stage and two-branch design in CCN and the two attention mechanisms design in DAiIAM, we propose a robust model to remove the rain streaks and raindrops synchronously in a truly unified CNN framework for better generalization ability.

2.2 Video Deraining via Optical Flow

Video deraining (VD) methods [8, 18, 34, 37, 38] focus on removing rain from multi frames that contain additional temporal information. In the video deraining task, the different image frames have mismatch due to the dynamic characteristics of the video. Therefore, many existing video rain removal methods, e.g., [34, 38], utilize optical flow to perform alignment among different frames. Specifically, a chosen frame is viewed as the reference frame, and a pre-trained FlowNet [4, 13] predicts the optical flow between adjacent frames and the reference frame, and then the adjacent frames are aligned with the reference frame according to the optical flow information. In this way, different rain image frames with the same background can be theoretically obtained. The parameters of the pre-trained FlowNet are fine-tuned through joint training with the rain removal
network to make it better adapt to the domain of rain videos. For example, [34] addresses the problem of rain removal from videos by a two-stage recurrent network with dual flow constraints. To better capture the motion patterns and retain inter-frame consistency, researchers employ two flow-based representations to regularize the learning of the proposed video deraining network.

### 3 PROPOSED SR^3 METHOD

In this section, we introduce RadNet in detail. As can be seen in Figure 2, RadNet contains three primary parts, i.e., Flow & Warp Module (FWM), Robust Attention Module (RAM), and Deep Refining Module (DRM). FWM can conduct optical estimation and alignment for unaligned data in RainDS_real. RAM can pay attention to different rain degradation phenomena with a robust attentional mechanism and perform coarse rain removal results. Finally, DRM uses a multi-stream CNN to perform refined rain removal. The overall network processing pipeline is as follows: (1) A rain image \( I_{\text{input}} \) and a clean image \( I_{\text{target}} \) are sent to FWM for optical flow estimation and alignment, and the output will be an aligned rain image \( \tilde{I}_{\text{input}} \); (2) \( \tilde{I}_{\text{input}} \) is sent to RAM to extract the attention map \( \tilde{I}_{\text{atten}} \) and predict a coarse deraining result \( \tilde{I}_{\text{coarse}} \); (3) \( \tilde{I}_{\text{coarse}} \) will be sent to DRM to obtain the derained image \( \tilde{I}_{\text{derain}} \); (4) \( \tilde{I}_{\text{derain}} \) will be sent to FWM again to obtain the final deraining result \( I_{\text{derain}} \). We will introduce FWM, RAM, and DRM, respectively.

#### 3.1 Flow & Warp Module (FWM)

Note that there exists a mismatch between rain image and clean image in the sub-datasets of RainDS dataset, i.e., RS_real, RD_real, and RDS_real. The FWM module is mainly designed to improve the generalization ability of our method for handling these real scenario rain images. Unlike video deraining [38] with multiple consecutive frames, SID task only use three discontinuous images, i.e., rain image \( I_{\text{input}} \), clean image \( I_{\text{target}} \), and derained image \( I_{\text{derain}} \). The lacked consecutive frames will make the optical flow prediction and image alignment more difficult, since the more frames involved, the easier it is for FlowNet to fine-tune the model. To solve this problem, we conducted optical flow prediction and alignment operation in two stages before and after rain removal. Specifically, taking the clean image \( I_{\text{target}} \) as the reference frame, we first predict the optical flow information from the image pair \( (I_{\text{target}}, I_{\text{input}}) \), and then align the rain image \( I_{\text{input}} \) to the clean image \( I_{\text{target}} \) according to the optical flow information as follows:

\[
\hat{f}_{i \rightarrow t} = \mathcal{F}_\text{flow}(I_{\text{target}}, I_{\text{input}}),
\]

where \( \mathcal{F}_\text{flow}(\cdot) \) is the pre-trained FlowNet2, while \( f_{i \rightarrow t} \) denotes the optical flow estimated from the rain image \( I_{\text{input}} \) to clean image \( I_{\text{target}} \). Then, we can warp the rain image \( I_{\text{input}} \) to the clean image \( I_{\text{target}} \) based on the estimated optical flow information:

\[
\tilde{I}_{\text{input}} = \text{wrap}(I_{\text{input}}, f_{i \rightarrow t}),
\]

### Table 1: Evaluation results in terms of PSNR and SSIM metrics under single-type data strategy. Red and blue colors indicate the best and second best records in each group of results, respectively.

| Method             | Rain200H | Rain200L | RS_syn | RS_real | RainDrop | RD_syn | RD_real | Average  |
|--------------------|----------|----------|--------|---------|----------|--------|---------|----------|
| AttentGAN [20]     | 23.04/0.800 | 28.07/0.935 | 27.70/0.888 | 24.47/0.765 | 30.55/0.902 | 27.25/0.910 | 22.02/0.709 | 26.17/0.844 |
| DetailNet [7]      | 26.17/0.805 | 34.43/0.959 | 31.03/0.916 | 26.08/0.860 | 25.18/0.874 | 28.25/0.912 | 22.03/0.751 | 27.60/0.868 |
| RESCAN [17]        | 26.60/0.904 | 37.07/0.983 | 38.92/0.982 | 26.41/0.878 | 25.50/0.883 | 34.49/0.975 | 22.97/0.804 | 30.28/0.916 |
| PRNet [23]         | 28.08/0.889 | 36.76/0.981 | 39.12/0.989 | 26.17/0.878 | 25.87/0.896 | 34.28/0.980 | 23.60/0.827 | 31.64/0.922 |
| JORDER-E [36]      | 29.37/0.891 | 39.25/0.985 | 40.45/0.987 | 26.66/0.883 | 26.15/0.899 | 35.54/0.982 | 24.08/0.828 | 31.64/0.922 |
| RCDNet [25]        | 30.18/0.905 | 39.75/0.986 | 44.12/0.993 | 26.89/0.891 | 26.20/0.897 | 35.51/0.981 | 24.16/0.827 | 32.40/0.926 |
| RLNet [2]          | 29.38/0.897 | 38.25/0.982 | 37.33/0.976 | 26.75/0.885 | 26.67/0.908 | 33.01/0.971 | 24.24/0.836 | 30.80/0.922 |
| Pcs2pix [14]       | 23.96/0.736 | 30.24/0.893 | 28.22/0.820 | 25.19/0.717 | 25.50/0.825 | 24.71/0.803 | 20.73/0.619 | 25.51/0.773 |
| CCN [21]           | 29.12/0.922 | 37.94/0.984 | 35.12/0.970 | 26.83/0.821 | 31.34/0.951 | 33.29/0.975 | 24.81/0.758 | 31.21/0.912 |
| RadNet             | 30.08/0.915 | 38.68/0.985 | 39.58/0.991 | 26.64/0.893 | 24.54/0.885 | 35.49/0.982 | 23.78/0.832 | 32.02/0.946 |

**Figure 4:** Deraining comparison under single-type data strategy (i.e., RS_real) with state-of-art SID methods.

**Figure 5:** Deraining comparison under single-type data strategy (i.e., RD_real) with state-of-art SID methods.
where $\text{wrap}()$ is alignment operation, and $\tilde{I}_{\text{input}}$ is aligned rain images that we expect to have the same background as clean images. However, the optical flow prediction and alignment operation may be affected by the complex rain streaks in images, making FlowNet fail in learning parameters well. We carry out optical flow prediction and alignment operation after deraining as

$$ f_{d\rightarrow t} = \text{Flow}(I_{\text{target}}, I_{\text{derain}}), \quad (6) $$

$$ \tilde{I}_{\text{derain}} = \text{wrap}(I_{\text{derain}} - f_{d\rightarrow t}). \quad (7) $$

where $f_{d\rightarrow t}$ is the optical flow estimated from the derained image $I_{\text{derain}}$ to the clean image $I_{\text{target}}$. Similarly, we warp the derained image $I_{\text{derain}}$ to the clean image $I_{\text{target}}$ based on the estimated optical flow. By predicting optical flow and aligning the rain image $I_{\text{input}}$ and derained image $I_{\text{derain}}$, we can constrain the aligned results with the reference clean image $I_{\text{target}}$ via an L1 loss:

$$ L_{\text{flow}} = \|I_{\text{input}} - I_{\text{target}}\|_1 + \|I_{\text{derain}} - I_{\text{target}}\|_1. \quad (8) $$

We use FWM twice for optical flow prediction and alignment of real scenario images. The first time is to get rain images aligned with clean images, and the network could obtain the derained image without deviation to the clean images, while the second time can make the derained images more aligned with the clean images.

### 3.2 Robust Attention Module (RAM)

LSTM [3] has been demonstrated to be useful in many computer vision tasks, e.g., image deraining and dehazing. Following [20, 23, 42], we also use the LSTM unit, and residual block [9] to construct our robust attention module RAM. To be specific, RAM can pay attention to different rain degradation phenomena due to the multi-stage processing, so that the proposed RadNet can be applied to deal with different data strategies. An example of the attention maps extracted by RAM is shown in Figure 3. We see that RAM can focus on most regions with rain streaks, raindrops, and both. By subtracting the extracted attention map from the original rain images, RadNet can obtain coarse rain removal results, which are used as DRM input.

It is noted that RAM can reduce the terrible influence caused by the divergent distribution of different degradation phenomena, suggesting the strong robustness of different types of deraining tasks and facilitating the subsequent DRM deraining.

The structure of RAM is shown in the bottom middle of Figure 2, where each stage includes a $\text{Conv} + \text{ReLU}$ operation followed by an LSTM unit, and then three $\text{Conv} + \text{ReLU}$ in the end. The LSTM unit has an input gate $i_t$, a forget gate $f_t$, an output gate $o_t$, and a cell state $c_t$. The interaction in each LSTM unit is defined as follows:

$$ i_t = \sigma(W_f [X_t, H_{t-1}] + b_i), f_t = \sigma(W_f [X_t, H_{t-1}] + b_f), $$

$$ g_t = \sigma(W_g [X_t, H_{t-1}] + b_g), o_t = \sigma(W_o [X_t, H_{t-1}] + b_o), $$

$$ c_t = f_t \odot c_{t-1} + i_t \odot g_t, H_t = o_t \odot \tanh(c_t). \quad (9) $$

where $X_t$ denotes the stack maps obtained by preoperative $t$-stage $\text{Conv} + \text{ReLU}$ unit, $c_t$ denotes the cell state that will be fed to the LSTM unit in the next stage, $H_t$ is the output of current LSTM unit and will be sent to the next $\text{Conv} + \text{ReLU}$ unit, $\{,\}$ is concatenate operation, $\sigma$ and $\tanh$ are the $\text{sigmoid}$ and $\tanh$ functions, respectively.

After obtaining the predicted image $I_{\text{input}}$, we use it as input and send it to RAM. Firstly, the attention map $I_{\text{atten}}$ will be predicted, then added with the original input image $I_{\text{input}}$ to obtain the coarse deraining result $I_{\text{coarse}}$:

$$ I_{\text{atten}} = \text{RAM}(I_{\text{input}}), \quad (10) $$

$$ I_{\text{coarse}} = I_{\text{atten}} + I_{\text{input}}. \quad (11) $$

### 3.3 Deep Refining Module (DRM)

RAM can be regarded as a lightweight network in our proposed framework to obtain the coarse rain removal results. However, note that there still exist two issues needing to be explored: 1) The task of SR$^3$ is more complex, so simply deleting the rain attention maps obtained by the proposed RAM will make the recovered images lack some details. 2) The structure of RAM is plain and shallow with a limited receptive field, which is bad for capturing the global context of the rain image, and further makes the proposed deraining model generate low-quality images.

As such, we need a more powerful network (i.e., DRM) with a stronger learning ability to solve the above issues. To be specific, DRM consists of the following two main components: 1) Dual Path Residual Dense Block (DPRDB). Traditional deep networks use the simple residual blocks [9] or dense blocks [10] to extract features, but for the complex SR$^3$ task, these blocks cannot extract appropriate features from complex data strategies. Recently, a new block called DPRDB has been proposed in [33], which can jointly discover new features and reuse features, and has been proved to be very effective in the image restoration task with complex distributions. As such, DPRDBs will be employed in the design of DRM.
for representation. 2) Multi-Scale Architecture. Since multi-scale features extracted from different kernel sizes can capture different perceptive filed information hidden in the rain images, and we therefore use DPRDB as our basic block to construct a three-branch network for in-depth deraining. As a result, DRM can extract more contextual information to guide the proposed model complete further deraining.

The structure of our DRM is illustrated in Figure 2, where each branch has five DPRDBs. The kernel sizes of each branch are set to 3, 5 and 7, respectively. We use feature shortcuts between blocks to avoid information loss. After that, we use a Conv + ReLU operation to obtain the final refined rain removal results. After obtaining the coarse deraining results $I_{\text{coarse}}$, we use it as input and send it to DRM. This process can be formulated as follows:

$$I_{\text{derain}} = DRM(I_{\text{coarse}}).$$  

The derained image $I_{\text{derain}}$ will be then sent to FWM for further optical flow estimation and alignment to obtain the final refined deraining result $I_{\text{target}}$.

3.4 Loss Function

The total loss function of RadNet is formulated in Eqn. (13):

$$L_{\text{total}} = L_{\text{flow}} + L_{\text{per}} + L_{\text{ssim}},$$  

where $L_{\text{flow}}$ is the proposed loss for optimal flow estimation and alignment in Eqn. (8). $L_{\text{per}}$ is perceptual loss to minimize the difference between perceptual features of output and target, $V_{gg_i}$ ($\cdot$) is a pre-trained VGG-16 model [24]. In our method, we use layer ReLU2_2 and ReLU2_3 as $V_{gg_i}$ ($\cdot$) and $V_{gg_2}$ ($\cdot$). $L_{\text{ssim}}$ denotes the SSIM loss for measuring the similarity between output and target.

4 EXPERIMENTS

4.1 Experiment Setup

Baselines. We divided SID methods into three categories and evaluated them based on different data strategies: 1) Raindrop removal method, i.e., AttentGAN [20]; 2) Rain streak removal method, i.e., DetailNet [7], RESCAN [17], PReNet [23], JORDER-E [36], RCDNet [25], and RLNet [2]; 3) SR$^3$ method, i.e., Pix2pix [14], CCN [21] and the proposed RadNet. In addition, we also tested a variant of the proposed network without FWM, denoted as RadNet-.

Robustness evaluation. We choose four datasets (i.e., Rain200H, Rain200L [35], RainDrop [20], and RainDS [21]) as the benchmark data. RainDS has six subsets with synthetic and real-world data, i.e., RS_syn, RD_syn, RDS_syn, RS_real, RD_real, and RDS_real. Based on the above datasets, we design three data strategies to examine the robustness of each SID method: 1) Single-type data contain only one dataset with one type of rain degeneration, i.e., rain streak (Rain200H, Rain200L, RS_syn, and RS_real) and raindrop (RainDrop, RD_syn, and RD_real). 2) Superimposed-type data has only one dataset with two types of rain degeneration in a single image, i.e., RDS_syn and RDS_real. 3) Blended-type data contain more than two datasets for training and two types of rain degeneration are

\begin{table}[h]
\centering
\caption{Evaluation results in terms of PSNR and SSIM metrics under blended-type data strategy.}
\begin{tabular}{|l|c|c|c|c|}
\hline
Method & {RD_syn + RS_syn + RDS_syn} & {RD_real + RS_real + RDS_real} & {Rain200H + Rain200L + RainDrop} & Average \\
\hline
AttentGAN [20] (cvpr'18) & 26.23/0.864 & 22.51/0.706 & 23.92/0.836 & 24.22/0.802 \\
DetailNet [7] (cvpr'17) & 27.06/0.843 & 23.54/0.785 & 23.87/0.823 & 24.82/0.817 \\
RESCAN [17] (eccv'18) & 34.13/0.969 & 23.96/0.822 & 28.42/0.950 & 30.08/0.911 \\
Pix2pix [14] (cvpr'17) & 24.77/0.760 & 22.52/0.651 & 30.10/0.921 & 30.81/0.916 \\
PReNet [23] (cvpr'19) & 34.13/0.969 & 23.96/0.822 & 30.10/0.921 & 30.40/0.904 \\
RCDNet [25] (cvpr'20) & 35.98/0.975 & 24.42/0.822 & 29.85/0.911 & 30.08/0.903 \\
RLNet [2] (cvpr'21) & 35.87/0.973 & 25.12/0.838 & 30.73/0.924 & 30.41/0.912 \\
JORDER-E [36] (tpami'19) & 34.73/0.968 & 23.96/0.822 & 28.42/0.950 & 30.08/0.911 \\
\hline
\end{tabular}
\end{table}

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{Figure7.png}
\caption{Deraining comparison under blended-type data strategy (i.e., {RD_real + RS_real + RDS_real}) with state-of-art SID methods.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{Figure8.png}
\caption{Comparison of deraining performance (in terms of PSNR and SSIM) of different methods under different data strategies.}
\end{figure}
in different images, i.e., [RD_syn + RS_syn + RDS_syn], [RD_real + RS_real + RDS_real], and [Rain200H + Rain200L + RainDrop].

Generalization evaluation. We collected real scenario data from Internet and previous works [30, 32] to build a real dataset as benchmark to examine the generalization ability of each method.

Implementation Details. We use the PyTorch platform in a Python environment on two NVIDIA GeForce GTX 3080 GPU with 24GB memory each. The whole network is optimized jointly by the Adam optimizer. The initial learning rate is set to 1e-3 for the Robust Attention Module (RAM) and Deep Refining Module (DRM), while 1e-6 for the Flow & Warp Module (FWM). The learning rate is decayed by multiplying 0.2 every 30 epochs. The batch size is 16, and each image will be randomly cropped to 128×128 pixels. We train 100 epochs to make the network convergence. Two metrics are used for evaluations, i.e., Peak Signal to Noise Ratio (PSNR) [12] and Structural Similarity Index (SSIM) [1].

4.2 Robustness Evaluation Results

Results on single-type data. The results under the single-type strategy are described in Table 1. We see that: (1) our method generally obtains highly-competitive results on two real datasets, i.e., RS_real and RD_real, while our results far exceed those of other methods. (2) Compared with the closely-related SR3 method CCN, our method performed better on the six datasets and only except for the RainDrop dataset, particularly advantageous on RS_syn, RS_real, and RD_real. Specifically, RadNet is up to 4dB PSNR advantage in RS_syn and 3dB PSNR advantage in RD_real over CCN. (3) The performance of our method on the RainDrop dataset is worse than CCN. This may be because CCN has two modules to process rain streaks and raindrops so that it can process RainDrop data containing large blur areas. (4) On average, our RadNet has the best SSIM and slightly lower PSNR performance than RCDNet.

Finally, we also illustrate some results from RS_real and RD_real data in Figures 4 and 5, from which we see our method can remove more degradation, recover clean background, and obtain significant improvement. Specifically, we see from Figure 4 that RadNet can remove most of the rain streaks. While in Figure 5, RadNet can restore the obscured area by the halo of the raindrop effectively.

Results on superimposed-type data. The restoration task under this data strategy is more difficult, which mainly focuses on processing the rain streaks and raindrops synchronously. From the numerical results in Table 2, we see that: (1) Our model obtains the best performance on both synthetic data (RDS_syn) and real data (RDS_real). Specifically, our RadNet has up to 2dB PSNR advantage over CCN on RDS_syn, while up to nearly 4dB PSNR and 0.24dB SSIM advantage over CCN on RDS_real. (2) All methods perform poorly on RDS_real data, which is mainly because the pairs of images of the RDS dataset collected from real scenarios do not correspond at the pixel level, which brings problems to all supervised networks. But in this case, our approach can still work effectively due to the proposed FWM.

We also visualize an image from RDS_real in Figure 6. The mixture of rain streaks and raindrops will cause two degradation phenomena, i.e., the loss of texture caused by rain streaks and the loss of large areas of structure caused by raindrops. These two different optical phenomena make SID methods extremely difficult to process. The figure shows the compared methods cannot completely remove the rain streaks, and cannot restore the large area obscured by raindrops. In contrast, our method can predict the areas obscured by raindrops, while removing more rain streaks.

Results on blended-type data. This is the most difficult data strategy in experiments, since the simultaneous mixing and stacking of data are disastrous for networks with weak robustness capabilities. While the amount of data of blended-type is sufficient, and if can be fully utilized, a very robust network can be trained. From the numerical results in Table 3, we see that: (1) due to the designed new network structure and modules, our method can better handle this situation and obtain the best performance. Specifically, our model can obtain up to 1dB PSNR advantage over RCDNet on the first blended-type data, while obtaining a significant near 5dB PSNR advantage and 0.15dB SSIM advantage over RLNet in the real scenario; (2) since the authors did not provide the source codes of the CCN method, we cannot provide the corresponding result in this study. However, judging from our results on blended-type data, it is basically equal to the average of the result of every single inclusive dataset, e.g., the performance of blended-type data: [RD_syn + RS_syn + RDS_syn] ≈ performance of single-type data: RD_syn + RS_syn + RDS_syn, so CCN should be unable to surpass our results.

We also visualize some results from [RD_real + RS_real + RDS_real] in Figure 7. We see that most methods remain lots of rain degradation and blur the image, while our method obtains the clearest background and makes the restored image close to sharp image.
4.3 Generalization Evaluation Results

The ability of handling real-world scenario images is important to measure the generalization power of each deraining model. Since there are no ground-truth images for comparison, we list some examples in Figure 9. As can be seen, our method has the best rain removal effect, which can not only remove more rain streaks, but also avoid the ambiguity caused by excessive rain removal and artificial noise. All methods are pre-trained on RS_real data.

4.4 Ablation Study

In this study, we mainly explore the impact of three factors, i.e., Network Module, Basic Block and Loss Function. The experiments are conducted on the RDS_syn. Specifically:

1) We show the influence of network module on the performance in Table 4. Our method with modules RAM + DRM obtains the best performance, and the setting with only RAM module obtains the worst performance since the lightweight network focuses on coarse deraining and cannot learn a good mapping from complex data.

2) We show the influence of basic block on the performance in Table 5. Our method with DPRDBs obtained the best performance. By utilizing advantages of both residual block and dense block, DPRDB can reuse previous features and explore new features, leading to stronger representation learning ability.

3) We show the influence of loss function on the performance in Table 6. The best performance is obtained by using both $\mathcal{L}_p$ and $\mathcal{L}_{sim}$. This form of the loss function can preserve the per-pixel similarity and the global structures.

4) We further explore the effectiveness of FWM based on the RainDS_real (i.e., RS_real, RD_real, and RDS_real). We describe the comparison results of our RadNet without and with FWM in Table 7. We see that the performance without FWM is much worse than with FWM. Figure 10 shows the improvement of deraining results via twice FWM operations. As can be seen, optical flow prediction and alignment are carried twice over predicting $I_{\text{input}}$ and $I_{\text{derain}}$, respectively. After the first operation, the PSNR/SSIM increased from 19.44/0.588 ($I_{\text{input}}$) to 21.34/0.775 ($I_{\text{derain}}$), and after the second operation, the deraining result increased from 27.49/0.928 ($I_{\text{derain}}$) to 28.66/0.946 ($I_{\text{derain}}$). We only use FWM for real-scenario data in RainDS, i.e., RS_real, RD_real, and RDS_real, due to the un-corresponding at the pixel level of the rain image and clean image.

5 CONCLUSION

We have explored the robustness and generalization ability in the task of synchronous rain streaks and raindrops removal. Technically, we proposed a universal robust attention deraining network for processing synthetic and real rain images. Compared with the current methods that usually attempt to solve single-type data, we first investigate the optical flow estimation and alignment for the task of rain removal, so that real scenario data can be well handled. Extensive comparisons are conducted on synthetic and real rain images to evaluate the deraining performance and generalization ability of our method. The investigated results show that our model can outperform recent related methods, especially in real-world scenario. In future, we will investigate how to reduce the parameters of our method so that it can be deployed on lightweight devices.

ACKNOWLEDGMENTS

This work is partially supported by the National Natural Science Foundation of China (62072151, 61732007, 61932009, 2020106007), and the Anhui Provincial Natural Science Fund for Distinguished Young Scholars (2008085J30). Zhao Zhang and Jicong Fan are the co-corresponding authors of this paper.
REFERENCES

[1] Alan C Brook, Xiaonian Zhao, and Thrasyvoulos N Pappas. 2008. Structural similarity quality metrics in a coding context: exploring the space of realistic distortions. *IEEE Transactions on Image Processing* 17, 8 (2008), 1261–1273.

[2] Chenghao Chen and Hao Li. 2021. Robust Representation Learning with Feedback for Single Image Deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7742–7751.

[3] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. 2015. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2625–2634.

[4] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philipp Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers, and Thomas Brox. 2015. FlowNet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE international conference on computer vision*. 2758–2766.

[5] David Eigen, Dilip Krishnan, and Rob Fergus. 2013. Restoring an image taken through a window covered with dirt or rain. In *Proceedings of the IEEE international conference on computer vision*. 633–640.

[6] Zhiwen Fan, Huafeng Wu, Xueyang Fu, Yue Huang, and Xinghao Ding. 2018. Residual-guide network for single image deraining. In *Proceedings of the 26th ACM international conference on Multimedia*. 1751–1759.

[7] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. 2017. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3585–3593.

[8] Kshitiz Garg and Shree K Nayar. 2004. Detection and removal of rain from videos. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2004. CVPR 2004, Vol. 1. IEEE, 1–1.

[9] Kaining He, Xiangyu Zhang, Shaqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[10] Guo Huang, Zihang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.

[11] Huahuo Huang, Aijing Yu, and Ran He. 2021. Memory Oriented Transfer Learning for Semi-Supervised Image Deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7732–7741.

[12] Quan Huynh-Thu and Mohammed Ghanbari. 2008. Scope of validity of PSNR in image/video quality assessment. *Electronics letters* 44, 13 (2008). 800–801.

[13] Eddy Ilg, Nikolau Mayer, Tommy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. 2017. FlowNet 2.0: Evolution of optical flow estimation with deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2462–2470.

[14] Philipp Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1125–1134.

[15] Guanbin Li, Xiang He, Wei Zhang, Huiyou Chang, and Liang Lin. 2018. Non-locally enhanced encoder-decoder network for single image de-raining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1056–1064.

[16] Siyuan Li, Iago Breno Araujo, Wenqi Ren, Zhangyang Wang, Eric K Tokuda, Xiaochun Cao, and Subhashini Venugopalan. 2020. Controlling the Rain: From rain streak removal: When cyclic consistency meets temporal correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1661–1670.

[17] Yanyan Wei, Zhao Zhang, Richang Hong, and Meng Wang. 2019. A Course-to-Fine Multi-stream Hybrid Deraining Network for Single Image Deraining. In *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 628–637.

[18] Wenhan Yang, Jiashi Feng, Shiqi Wang, Bin Cheng, and Jiaying Liu. 2021. Progressive image deraining networks: a better and simpler baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3103–3112.

[19] Yanyan Wei, Zhao Zhang, Meng Wang, and Zongming Guo. 2018. D3r-net: Network and Quasi-Sparsity Based Training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14791–14801.

[20] Siqi Ni, Xueyun Cao, Tao Yue, and Xuemei Hu. 2021. Recurrent multi-frame deraining: Combining physics guidance and adversarial learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).

[21] Sungho Yoon, Joo-Hwee Lim. 2019. Singe Image Rain Removal with Unpaired Data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1720–1729.

[22] Fang Chen, Shijie Deng, Yixin Chen, and Hui Ji. 2019. Deep learning for seeing through window with raindrops. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2463–2471.

[23] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zha, and Deyu Meng. 2019. Progressive image deraining networks: a better and simpler baseline. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3937–3946.

[24] Karen Simonyan and Andrew Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. In *Proceedings of the 3rd International Conference on Learning Representations*. 4700–4708.

[25] Yinglong Wang, Chao Ma, and Bing Zeng. 2021. Multi-Decoding Deraining Network and Quasi-Sparsity Based Training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13375–13384.

[26] Zheng Wang, Jianwu Li, and Ge Song. 2019. DTDN: Dual-task De-raining Network. In *Proceedings of the 27th ACM International Conference on Multimedia*. 1833–1841.

[27] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. 2019. Semi-supervised transfer learning for image rain removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3877–3886.

[28] Yanyan Wei, Zhao Zhang, Yinglong Wang, Zongming Guo, and Meng Wang. 2021. DerainCycleGAN: Rain attentive CycleGAN for single image deraining and rainmaking. *IEEE Transactions on Image Processing* (2021).

[29] Yanyan Wei, Zhao Zhang, Yang Wang, Huiyou Chang, Meng Zhao, and Meng Wang. 2019. Semi-deraining: A new semi-supervised single image deraining network. In *Proceedings of the IEEE International Conference on Multimedia and Expo*. 1–6.

[30] Yanyan Wei, Zhao Zhang, Huiyou Chang, Richang Hong, and Meng Wang. 2019. A Course-to-Fine Multi-stream Hybrid Deraining Network for Single Image Deraining. In *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 628–637.

[31] Wenhan Yang, Jiashi Feng, Subhashini Venugopalan, and Meng Wang. 2019. Complementary recurrent video deraining with dual-level flow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1661–1670.

[32] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiayi Liu, Zongming Guo, and Shuicheng Yan. 2017. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1357–1366.

[33] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiayi Liu, Shuicheng Yan, and Zongming Guo. 2019. Deep joint rain detection and removal from a single image with contextualized deep networks. *IEEE transactions on pattern analysis and machine intelligence* (2019).

[34] Wenhan Yang, Robby T Tan, Jiashi Feng, Siqi Ni, and Jiayi Liu. 2021. Recurrent multi-frame deraining: Combining physics guidance and adversarial learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).

[35] Wenhan Yang, Robby T Tan, Shiqi Wang, and Jiayi Liu. 2020. Self-learning video rain streak removal: When cyclic consistency meets temporal correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4854–4863.

[36] Yuntong Ye, Yi Chang, Hanyu Zhou, and Luxin Yan. 2021. Closing the Loop: Joint Rain Generation and Removal via Disentangled Image Translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2053–2062.

[37] Kaihao Zhang, Dongxu Li, Wenhan Luo, and Wenqi Ren. 2021. Multi-attractive and attentive model for joint rain streak and raindrop removal. *IEEE Transactions on Image Processing* 30 (2021), 7408–7419.

[38] Zhao Zhang, Yanyan Wei, Huiyou Chang, Yi Yang, Shuicheng Yan, and Meng Wang. 2021. Data-driven single image deraining: a comprehensive review and new perspectives. *TechRxiv* (2021). 1–20.

[39] Hongyan Zhu, Xi Feng, Joey Tianyi Zhou, Songfan Yang, Vijay Chandersekhar, Liyan Li, and Joo-Hwee Lim. 2019. Singe Image Rain Removal with Unpaired Information: A Differentiable Programming Perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 9332–9339.