Decomposition Makes Better Rain Removal: An Improved Attention-guided Deraining Network

Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Zhen Han, Tao Lu, Baojin Huang, Junjun Jiang

***Abstract*—Rain streaks in the air show diverse characteristics with different shapes, directions, densities, even the complex overlapped phenomenon, causing great challenges for the deraining task. Recently, deep learning based image deraining methods have been extensively investigated due to their excellent performance. However, most of the existing algorithms still have limitations in removing rain streaks while preserving rich textural details under complicated rain conditions. To this end, we propose to decompose rain streaks into multiple rain layers and individually estimate each of them along the network stages to cope with the increasing abstracts. To better characterize rain layers, an improved non-local block is designed to exploit the self-similarity of rain information by learning the holistic spatial feature correlations while reducing the calculation complexity. Moreover, a mixed attention mechanism is applied to guide the fusion of rain layers by focusing on the local and global overlaps among these rain layers. Extensive experiments on both synthetic rainy/rain-haze/raindrop datasets, real-world samples, the haze, and low-light scenarios show substantial improvements both on quantitative indicators and visual effects over the current state-of-the-art technologies. The source code is available at https://github.com/kuihua/IADN.**

***Index Terms*—Image deraining, layer-wise learning, attention mechanism, non-local network.**

I. Introduction

High-quality images in the outdoor scenes carry rich information, which is crucial for many computer vision tasks [1], [2], [3], such as object detection, recognition, and scene analysis. However, the observed image quality inevitably suffers from degradations in real-world scenarios due to bad weather conditions, such as the rain [4], [5], snow, fog [6], [7], and so on. Especially for the rainy day, rain streaks greatly affect the visibility and luminance of scenes, which in turn may lead to the failure of many computer vision systems [8], [9].

This work is supported by National Key R&D Project (2016YFE0202300) and National Natural Science Foundation of China (U1903214, 61671332, U1736206, 62071339, 62072347,62072350,61971165,61872277,61971315), and Hubei Province Technological Innovation Major Project (2019AAA049, 2019AAA045). (*Corresponding author: Zhongyuan Wang*)

K. Jiang, Z. Wang, P. Yi, Z. Han, and B. Huang are with the National Engineering Research Center for Multimedia Software, School of Computer Science, Wuhan University, Wuhan, 430072, China. E-mail: kuijiang_1994@163.com; wzy_hope@163.com; 2017202110008@whu.edu.cn; hanzhen_1980@163.com; huangbaojin@whu.edu.cn.

C. Chen is with the Department of Electrical and Computer Engineering, University of North Carolina at Charlotte. E-mail: chen.chen@uncc.edu.

T. Lu is with the Hubei Province Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan, 430205, China. E-mail: lutxyl@gmail.com.

J. Jiang is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China, and is also with the Peng Cheng Laboratory, Shenzhen, China. E-mail: jiangjunjun@hit.edu.cn.

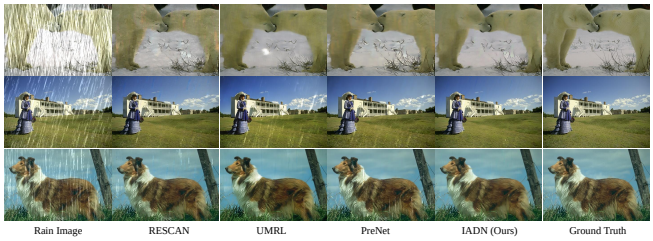

Fig. 1. Deraining examples. Although the previous stage-wise based deraining methods, such as RESCAN [20] and PreNet [21], can work well on the regions with light or sparse rain streaks, these two methods have limited ability to cope with the regions corrupted with dense or overlapped rain streaks, causing obvious color distortion and rain streaks remained. Besides, the density-estimation based method (UMRL [22]) tends to produce under- or over-deraining results. Our proposed IADN model can remove rain streaks more thoroughly and generate clearer and more credible image contents. Best viewed in color.

Therefore, inferring credible and clear contents from its rain-contaminated image is a pressing need for numerous vision tasks.

Given an observed rainy image I_{rain}, it can be mathematically considered as the superposition of a rain component I_R with the corresponding clean background image I_B as follows

$$I_{rain} = I_R + I_B. \tag{1}$$

Thus the goal of image deraining is to produce the rain-free output I_B^* from its rain-contaminated counterpart I_{rain}, approaching the clean background image I_B. Previous model-based technologies [10], [11] tend to learn the rain streak estimation using hand-crafted priors, *e.g.*, photometric appearance [12], geometrical features [13], as well as local structure correlations [14]. However, since rain streaks are of various directions, densities, and sizes, these methods are limited to meet the increasingly complex deraining tasks with hand-crafted features. More recently, numerous paired training datasets along with the deep learning frameworks [15], [16], [17] have given rise to many promising solutions on image deraining tasks [18], [19], [4]. These data-driven algorithms are capable of learning the intrinsic statistical characteristics with the powerful modeling ability of deep neural networks, which significantly promote the deraining performance and are more robust over conventional deraining techniques.

However, rain streaks in different scenarios show diverse shapes, directions, densities, and even overlaps. Most of the existing deraining methods [23], [24], [25] are particularly good enough in removing light rain, which can not fit real-world scenarios of complicated and diverse rain characteristics.

To this end, the authors in [26] use the image-level priors to learn the rain density estimation jointly and image deraining based on a density-label-based rainy dataset. However, the restoration performance overwhelmingly relies on the detection accuracy of rain density, while the dataset is only for limited rain conditions (heavy, medium, and light rain). To improve the generality, researchers [22], [27] propose to adaptively learn the density map by considering the local information of rain streaks, which in turn guides the network for deraining. Although they provide a promising solution to cope with dense rain streaks, these methods usually generate under- or over-deraining results because of the low tolerance to estimate the holistic rain streaks directly. Another effective strategy to improve deraining performance is the recurrent learning [20], [28]. These methods repeatedly learn the residual maps between the predicted deraining image and the ground truth via recurrent convolution layers. Albeit gaining the benefits, the repeated residual learning using a unified and fixed framework is unable to deal with the increasingly abstracted representation of rain layers in rainy images (especially for the overlaps of rain layers). Usually, it makes deraining results visually vulnerable and inconsistent with real contents on the contrast (Please refer to the deraining results by RESCAN [20] and PreNet [21] in Figure. 1). To sum up, there remain mainly two issues in existing methods to produce the high-quality and high-naturalness rain-free image. First, how to train a unified model to fit the diversity of rain density? Second, how to deal with the overlapping phenomenon of rain streaks located at any position and achieve the high-fidelity restoration?

To deal with these issues, in our previous work [29], we constructed an attention-guided deraining network (ADN) to learn multi-level abstractions of rain layers progressively and used the attention mechanism to guide the fusion by focusing on the overlapped regions at the channel and spatial dimensions. Albeit gaining the benefits from these effective strategies, there is the tremendous growth potential of ADN on deraining performance. First, ADN constructs a residual memory block (RMB) for rain streak representation by combining the recurrent calculation and residual learning. However, the recurrent structure (residual ConvLSTM in ADN) is specifically designed for temporal correlations rather than spatial dependence. Thus the internal correlations (self-similarity) of rain streaks at spatial dimension are underexploited. Moreover, the cascaded residual ConvLSTM in ADN trades the memory and computational cost for increased global representation to rain information, hampering the finer decomposition of rain streaks (more rain layers).

Based on these analyses above, we extend ADN [29] by both considering the holistic correlation representation of rain streaks (self-similarity), as well as layer-wise abstraction learning of rain layers more efficiently, and construct an improved attention-guided deraining network (IADN) for single image deraining. More specifically, unlike embedding the recurrent learning unit into RMB, we design an improved non-local block [30] to capture spatial feature dependencies, and further exploit the holistic self-similarity of rain information for a deep representation. It is more effective on the reduction of computation and GPU memory usage than recurrent calculation. Thus the burden reduction allows us to construct a deeper network, which can perform a more ingenious decomposition of rain streaks to cope with the finer abstractions (more rain layers), and further to release the learning difficulty. Moreover, we replace RMB in ADN with the general channel attention block (CAB) to promote the feature representation ability, and thus our model is easily extended. Following ADN, we aggregate these rain layers to regress the final residual rain image under the guidance of the content attention maps to alleviating the effects of the overlaps. Specifically, the concatenated rain layers go through a mixed attention block (MAB) to distill the final candidate components by learning the fusion weights among channel and spatial dimensions. To illustrate the aptitude of our proposed IADN method, we perform a variety of ablation studies as well as comparison experiments with state-of-the-art deraining approaches on common datasets. In summary, parts of this paper's results were published originally in this conference version [29]. However, this paper extends our earlier work in several important aspects:

- We extend ADN by considering both the feature representation and algorithm efficiency, and construct an improved ADN (IADN) for image deraining task. When compared with ADN [29], it yields better deraining performance (by 1.51dB and 0.86dB on Test100 and Test1 datasets, respectively) while enjoying 90% more efficiency both on inference time and parameters (Table. I and Figure. 7).
- We exhibit more comparison results on another four synthetic datasets with more recent deraining algorithms, including PreNet [21], LPNet [31] and ADN [29]. Moreover, more comprehensive analyses composed of the deraining performance, computational complexity, and inference time are given, showing that IADN achieves approximate or better restoration performance with leas the computational overhead (Table. III—V and Figure. 8, 9).
- Additional analytical results, as well as the visual representation of the rain layers, are presented, showing that the layer-wise scheme gradually abstracts the feature representation of rain information (Table. II and Figure. 3).
- We exhibit more comparison results on diverse rain patterns (covering rain-haze and raindrop) and other low-level vision tasks (including image dehazing and low-light enhancement) for a comprehensive evaluation (Table. VII, VIII and Figure. 11—14).

II. Related Works

In the last few years, many image deraining schemes have been proposed and achieved significant improvements. In what follows, we briefly review contemporary works closely related to ours, including the single image deraining [32], [33], non-local feature learning [30], [34] and attention mechanism [35].

A. Image Deraining

The image deraining task is to generate a high-quality rain-free image from the rain-corrupted input. The early model-based methods [12], [36], [37] use the hand-crafted priors to separate the background and rain streaks. For example,

Luo *et al.* [38] propose a dictionary learning method for single image deraining. After that, Li *et al.* [39] consider the multiple orientations and scales of the rain streaks, and propose a prior-based GMM method for rain streak estimation. However, these methods are powerful limited for the complex rain conditions, and tend to produce degraded image contents because of the dependence on the priors. Very recently, deep-learning-based approaches [40], [41], [42] have emerged for rain streak removal and demonstrated impressive restoration performance. These methods apply the powerful feature representation of deep neural networks to learn the mapping relations between the rainy image and clean background from numerous paired samples. For example, Fu *et al.* [23] pioneer a three-layer deep neural network for the rain streak estimation. They also present a rain detection and removal network by recurrently learning the physical characteristics of rain streaks. To release the learning difficulty, Li *et al.* [20] decompose the deraining task into multiple subproblems, and propose a recurrent squeeze-and-excitation context aggregation network (RESCAN) to repeatedly learning the residual rain image. To better model the rain information, Zhang *et al.* [26] consider both the density and feature representation of rain streaks, and present a multi-stream dense connection to model multi-scale features of rain streaks. Later, Zhang *et al.* [25] further incorporate quantitative, visual, and discriminative performance into the objective function, and propose a conditional generative adversarial network (GAN) for rain streak removal. More recently, Deng *et al.* [43] propose to view rain removal and detail recovery as two separate tasks, and further solve these two parts specifically. To exploit the self-similarity of rain streaks, researchers [44], [45], [46] integrate the non-local operation into the multi-scale framework to capture the spatial correlation of rain information. In [47], the authors provide a comprehensive survey of deraining methods over the last decade, dividing them into model-based and data-driven approaches, and also give insights on the historical development of deraining methods. At the same time, performance comparison and future directions are provided. In [48], the authors present a weakly-supervised technique for rain streak removal using the unpaired rainy images. More specifically, they construct a two-stage data distillation method for learning the rain map in the first "rain-to-clean" stage and the rain-free image in the second "clean-to-rain" stage. In [49], the authors propose to aggregate the advantages of the conventional model-driven prior-based and data-driven DL-based methods, and construct a novel RCDNet for single image deraining. Based on the sparsity and non-local similarity of rain streaks, they learn the specific rain kernels at different stages, and apply them to predict the rain maps while updating the background contents till the final stage. Unlike [49], we argue that the rain streaks can be decomposed into multiple rain layers due to the depth from the camera to the object, corresponding to the progressive abstractions. Thus we decompose the rain streaks into multiple rain layers, and individually learn the estimation of each rain layer, whose solutions can be combined to yield a final solution for the deraining task. To characterize the rain streaks, we employ a modified non-local block in the first step to capture the spatial similarity of rain streaks, and then abstract the representation of rain layers at different stages of the network. Finally, we aggregate these rain layers using a mixed attention block to extract and fuse the candidate components, leading to a more precise estimation of rain streaks.

B. Non-local Feature Learning

Computing the feature response between the reference position and other query pixels is a practical and effective technology to capture the long-range dependencies, which benefits more to the high-level abstraction representation of image contents. Buades *et al.* [50] propose the patch-based non-local filter algorithm for image denoising, which calculates the non-local means as the weighted average of all the pixels of the image patches. After that, a similar strategy is applied in BM3D [51], and achieves the state-of-the-art restoration performance. More recently, researchers integrate non-local operation into deep learning framework to enhance the image content representation and understanding. For example, Wang *et al.* [30] propose the non-local network to capture the long-range dependencies among spatial pixels, and showcase its efficacy for image classification. Enlightened by these, there emerge many promising non-local solutions for image restoration [52]. These methods introduce the non-local operation to explore the self-similarity of detail information at spatial dimension [44], [45], [46] or across the temporal stages [53]. Limited by the great complexity of computing the holistic response, the standard non-local operation is unsuitable to dealing with large resolution images and cope with real-time tasks. Thus, a series of efficiency-friendly strategies, such as bottleneck structure [30] and subsampling [30], [54], are introduced to save the computation and GPU memory footprint.

C. Attention Mechanism

Attention mechanism has become a popular technique to promote the discriminative learning ability by guiding the network to focus on the most valuable components. It has been successfully applied to computer vision tasks, including localization and understanding of images [55]. In previous studies [56], researchers utilize visual attention to pay close attention to the related regions concerning the classification target. For example, Fu *et al.* [57] propose to recursively learn the discriminative region attention and region-based feature representation at multiple scales in a mutually reinforcing way, and construct a recurrent attention network for fine-grained image recognition. After that, Hu *et al.* [58] present a squeeze-and-excitation (SE) block to rescale the channel responses with attention mechanism, which achieves significant performance improvements for image classification. More recently, attention mechanism has also bee applied for low-level vision tasks, such as image super-resolution [59], [60], denoising [61], and deraining [32]. These methods use the convolutions and sigmoid activations to learn the attention map through explicit supervision to capture the direction-aware features to guide the network optimization, and significantly promote the feature presentation efficiency and model performance.

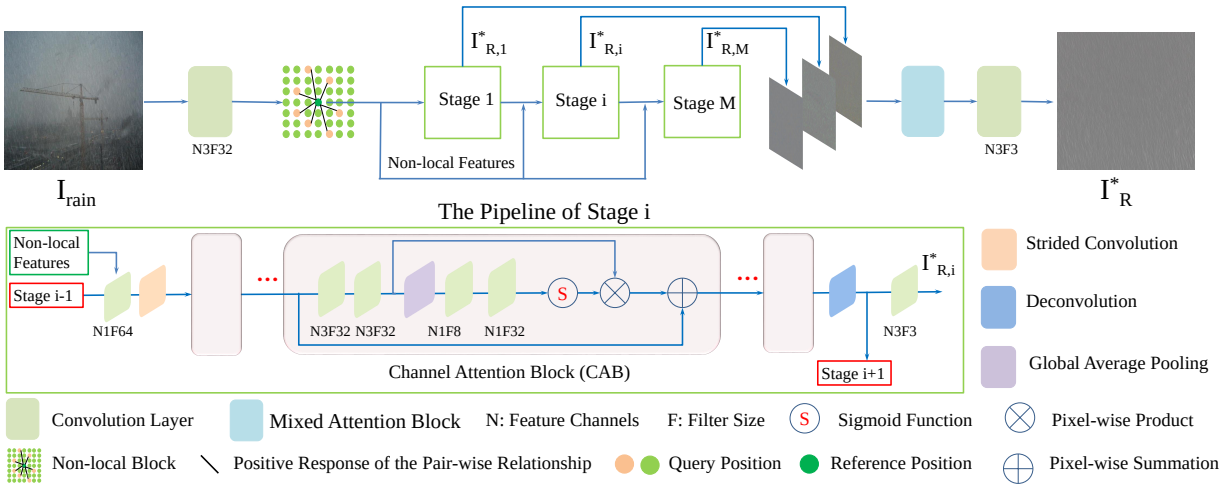

Fig. 2. Outline of the proposed improved attention-guided deraining network (IADN). IADN decomposes the learning task into N (N is set to 3 as an example.) stages, corresponding to estimate N rain layers ($I^*_{R,1}$, $I^*_{R,2}$, $I^*_{R,3}$). By fusing these rain layers through a mixed attention block (MAB), we produce the optimal approximation I^*_R of the residual rain image I_R.

III. Methods

An observed rainy image I_{rain} can be described as a linear combination of the background image I_B and its residual rain image I_R. As the diversity of rain characteristics, containing various rain shapes, directions, and densities, the distribution of I_R is complex and irregular in the air. To better model rain streaks, researchers [62], [28] regard I_R as the accumulation of multiple rain layers $I_{R,i}$ ($i \in [1, N]$). They thus neglect the discrepancy of learning tasks for different rain layers since these methods progressively remove rain streak layers through multiple recurrent stages under a unified framework. Furthermore, the simple summation among stages to regress the predicted residual rain image usually causes the loss of image details due to the overlap among rain layers. To solve these issues, we follow the decomposition strategy in [62], [28] by dividing the deraining task into multiple subproblems, corresponding to multiple rain layers, and estimate each rain layer at a specifical stage of the network to match their increasing abstraction. Moreover, we apply a mixed attention block (MAB) to rescale the feature responses to guide the fusion of these rain layers by focusing on the overlap regions among channel and spatial dimensions. The outline of our proposed IADN is shown in Figure. 2.

A. Architecture and Model Optimization

The final goal of designing IADN is to produce a high-quality rain-free image from its rainy observation I_{rain}. Motivated by [35], we utilize one initial convolutional layer $H_{ini}(\cdot)$ to extract the shallow features F_0 while projecting the input I_{rain} from the image space into feature space through the following formulation

$$F_0 = H_{ini}(I_{rain}). \tag{2}$$

Unlike the ADN [29] where F_0 is directly passed through the first stage to estimate the first rain layer, we apply a non-local operation to explore the self-similarity of rain information by calculating the spatial feature dependencies. In this way, the redundant information among similar rain patterns in a rain image (*e.g.*, similar appearance) can be aggregated to characterize the referenced rain streaks. The non-local representation of current object pixel $I(x, y)$ can be expressed as

$$I(x, y) = \sum_{\forall i} \sum_{\forall j} I(x_i, x_j) g(x_i, x_j) + I(x, y). \tag{3}$$

In Equation (3), $I(x_i, x_j)$ denotes the representation of the query positions except for $I(x, y)$, and $g(x_i, x_j)$ is the similarity matrix, denoting the pair-wise relationship between the reference position $I(x, y)$ and other position $I(x_i, x_j)$. Thus, we obtain the non-local representation F_{non} of the initial features. Then, F_{non} passes through the first stage to make a step towards deep extraction of rain information. Meanwhile, a reconstruction layer $H_{reg,1}(\cdot)$ with filter size of 3×3 is applied to regress the first rain layer, denoted as $I^*_{R,1}$. These procedures above can be defined as

$$F_1 = H_{stage,1}(F_{non}), \tag{4}$$

$$I^*_{R,1} = H_{reg,1}(F_1). \tag{5}$$

In Equation (4), $H_{stage,1}(\cdot)$ denotes the feature extraction functions, essentially several cascaded channel attention blocks (CABs) used to encode rain streak features. Following that, a reconstruction layer $H_{R,1}(\cdot)$ takes F_1 as input to predict the first rain layer. Next, F_1 is concatenated with the non-local representation F_{non} along the channel dimension, and then go through the second stage to regress the next rain layer. Similar to the first stage, the regression procedures of the i_{th} rain layer $I^*_{R,i}$ ($i \in [2, N]$) can be formulated as

$$F_i = H_{stage,i}(H_{concat}(F_{non}, F_{i-1}), \tag{6}$$

$$I^*_{R,i} = H_{reg,i}(F_i). \tag{7}$$

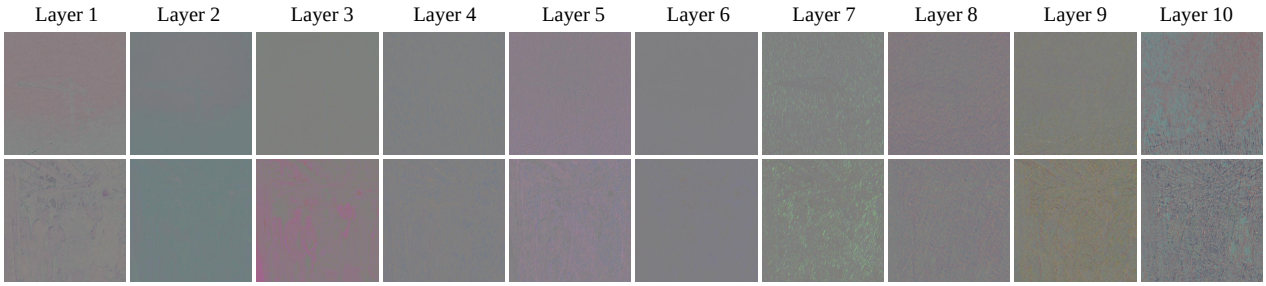

Fig. 3. Visual representation of rain layers. For better visualization, the pixel values are projected into [0, 255]. These rain layers contain rain information at different feature levels. From left to right, the network gradually abstracts the feature representation of rain information. (Best viewed in color.)

Consequently, we generate N predicted rain layers $I^*_{R,i}$ ($i \in [1, N]$) from N stages, shown in Figure. 3, which are regarded as the aggregated elements of the final residual rain image. In the previous works [62], [28], the authors tend to regress the final residual rain image via the commonly used pixel-wise summation. It is efficient and practical, but ignores the overlap among these rain streak layers, thus causing the redundancy or loss of information due to the diverse overlapped phenomenons. In particular, the overlapped strengths vary with the spatial regions and rain layers, which matter for the accurate model of rain streaks. Thus it is beneficial to make the network have the discriminative ability for these overlapped regions and layers and pay more attention to the grievously overlapped regions. Inspired by previous studies to estimate the rain density maps [22], [27], we construct a mixed attention block (MAB) using the attention mechanism [35], which can provide the network with powerful discriminative capability in focusing on the overlapped areas to learn the fusion weights adaptively. The fusion procedures are expressed as

$$I^*_R = F_{MAB}(I^*_{R,1}, I^*_{R,2}, \cdots, I^*_{R,N}). \tag{8}$$

To obtain the optimal solution, most of the existing deraining methods tend to minimize the mean squared error (l_2 loss) to generate a rain-free result approximating the ground truth. However, they usually produce blurry and oversmoothing visual effects by missing high-frequency textures since the squared penalty. In this study, we introduce the Charbonnier penalty function [63] to train the network to perform successive approximation, which is more tolerant of small errors and holds better convergence during training. The function is expressed as

$$L = \sqrt{(I_{Derain} - I_{GT})^2 + \varepsilon^2}. \tag{9}$$

In Equation (9), I_{Derain} refers to the rain-free image by subtracting the predicted residual rain maps I^*_R from the rainy input I_{Rain}. The penalty coefficient ε is empirically set to 10^{-3}.

B. Non-local Block

Context information contributes more to the accurate feature representation [64]. Especially for rain removal research, rich context information helps the network to distinguish rain streaks from the background. In ADN [29], a residual memory

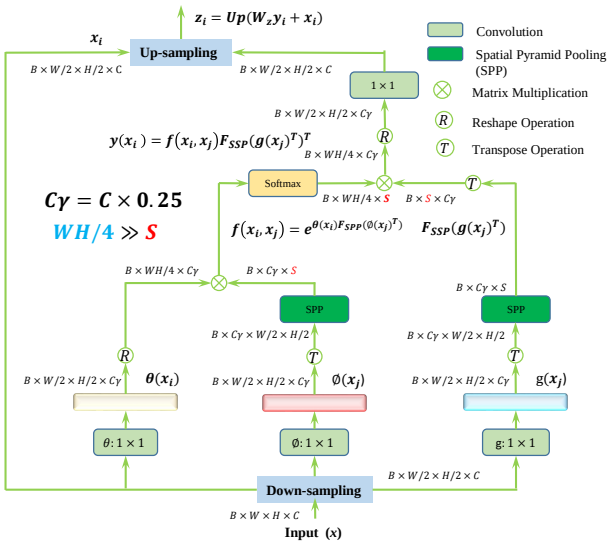

Fig. 4. Structure of the improved non-local block. When compared with the traditional non-local block [30], it introduces the spatial pyramid pooling operation to reduce the computation and GPU memory usage ($S \ll WH/4$) while maintaining the excellent modeling capability as the standard non-local module to capture the global correlated information. Meanwhile, the bottleneck design of the feature channels and the up-down architecture only require 1/16 of the original computational overhead ($W \times H \times C = W/2 \times H/2 \times C \times 0.25$,).

block is designed to capture the textural dependency across spatial series for its powerful ability in excavating strongly correlated information in spatiotemporal series. It is somewhat effective, but requires extensive calculation and memory cost. Another commonly used way to capture long-range dependence is the non-local operation [30]. More specifically, non-local block allows the network to compute the global response at a position as a weighted sum of the features at all spatial positions, thus allowing the network to exploit the complementary information from the similar and repeated patterns. However, the time and space complexities of the standard non-local operation are both quadratic to the number of positions $W \times H$ compared to normal operations since attention maps are computed for each query position via matrix multiplication.

Inspired by [65], the authors have demonstrated that the distance between the attention maps of different query positions is very small. In this work, we integrate the up-down structure, spatial pyramid pooling, and bottleneck layer into

the non-local block to save the memory footprint by sampling several channels and sparse anchor points instead of feeding all the spatial points by referring to [54]. As depicted in Figure. 4, we first sample the input with a strided convolution layer to select the query positions. We have experimentally verified that the up-down structure with strided convolution and deconvolution gains better performance by 0.1-0.2dB than that of the pooling operation or wavelet transform. Then, the embedding operations (1×1 convolution layer) in the standard non-local block is replaced with a bottleneck layer to rescale the channel response via a scaling factor. Thus for the arbitrary input features ($B \times W \times H \times C$) and the given integer factor r and sampling factor (0.5), the input is rescaled into $B \times W/2 \times H/2 \times C/r$. Moreover, before calculating the pair-wise relationship between the reference position x_i and other position x_j, we introduce the spatial pyramid pooling to sample several sparse anchor points instead of feeding all the spatial points. Thus given output pooling sizes, the outcome of the spatial pyramid pooling operation is a vector with the size of $B \times S \times C/r$, where S is much smaller than $W/2 \times H/2$. Thus the space complexity could be considerably reduced (*e.g.*, $W/2 \times H/2 \rightarrow S$ in Figure. 4.). The procedures of the improved non-local block can be defined as

$$f(x_i, x_j) = e^{\theta(down(x_i))F_{SPP}(\phi(down(x_j))^T)},$$
$$y(x_i) = \frac{1}{C(f)} \sum_{\forall j} f(x_i, x_j) F_{SPP}(g(down(x_j))^T)^T. \quad (10)$$

In Equation (10), $down(\cdot)$ is the strided convolution to sample the input with factor of 0.5. $\theta(\cdot)$ and $\phi(\cdot)$ denote two bottleneck layers with different learned parameters, W_θ and W_ϕ. $F_{SPP}(\cdot)$ refer to the spatial pyramid pooling operation. $f(x_i, x_j)$ represents the attention map for each query position. Afterward, a normalization (softmax function) is applied to $f(x_i, x_j)$ to get a unified similarity matrix. $g(\cdot)$ is the unary function that computes the representation of f while $C(f)$ is the normalization factor, defined as $C(f) = \sum_{\forall j} f(x_i, x_j)$. In this way, the feature representation is non-locally enhanced via considering all positions ($\forall j$) for each location i. Finally, a 1×1 convolution layer is implemented, acting as weighting parameter to adjust the importance of the non-local operation $w.r.t.$ the original input x and moreover, rescale the channel dimension to C, while a deconvolution layer $up(\cdot)$ is used to magnify the feature maps by the following formula

$$z_i = up(W_z\ y(x_i) + x_i). \quad (11)$$

C. Mixed Attention Block

Shown in Figure. 3, the rain layers estimated by different stages in IADN represent the diverse abstracts of rain information and contain redundancy information due to the overlapping phenomenon. Fusing these rain layers by a pixel-wise summation [62] is efficient, but inevitably causes the details missing. In this work, we design a mixed attention block (MAB), shown in Figure. 5, to conveniently cope with the overlap by performing a recalibration on feature responses. More concretely, MAB provides the network with discriminative learning capability by considering the overlapping of rain layers among channel and spatial dimensions.

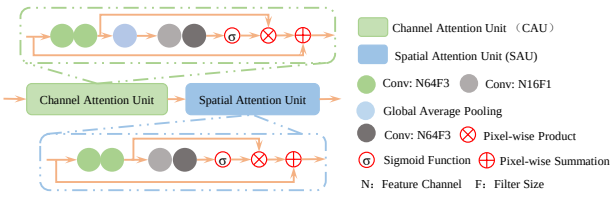

Fig. 5. Pipeline of the proposed mixed attention block (MAB).

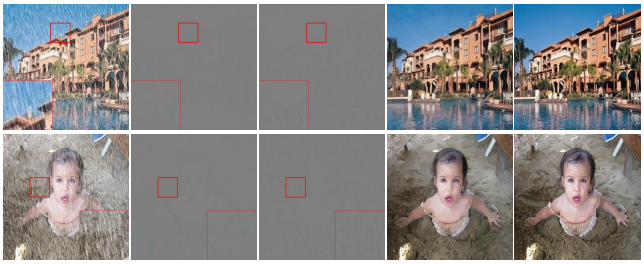

Fig. 6. Visual representation of the residual maps in the channel and spatial attention units. The residual maps of the second and third columns denote the residual components between the input and the output of the channel and spatial attention units, respectively (Please zoom in to see more details.).

Before passing the rain layers $I^*_{R,i}$ ($i \in [1, N]$) into MAB, a channel-wise concatenation is performed to obtain I^*_{cat}. After that, a channel attention unit (CAU) takes I^*_{cat} as input, and then is used to exploit the channel inter-dependence through the following function

$$F_{CAU} = H_{CAU}(I^*_{cat}). \quad (12)$$

Thus the network can adaptively distill the most informative components of rain layers by considering the contribution to the predicted residual rain image and the global overlapped strength.

Although CAU can provide a reasonable fusion scheme by rescaling feature responses at the channel dimension, the overlapped strength varies with local regions at the spatial dimension. Therefore, it is crucial to provide the network with the discriminative ability to cope with the overlap at any position and pay more attention to the grievously overlapped regions. Besides CAU, we apply a spatial attention unit (SAU) to improve the fusion quality of rain layers. This procedure can be expressed as

$$F_{SAU} = H_{SAU}(F_{CAU}), \quad (13)$$

where F_{SAU} denotes the candidate components of N rain layers. With SAU, the distilled features F_{CAU} via CAU are adaptively modulated at the spatial dimension, providing a substantive supplement for the components selection of rain layers. After that, a reconstruction layer is adopted to project F_{SAU} into image space:

$$I^*_R = H_{rec}(F_{sa}). \quad (14)$$

For a better understanding of our proposed mixed attention block, we provide the residual maps (the residual components between the input and the output) of the spatial and channel attention units in Figure. 6. For convenience, the first three

channels of the residual components are selected for the visualization. We can see that the spatial attention unit focuses on the local overlaps of rain layers. In contrast, the channel attention unit tends to cope with the global overlap of rain layers by performing a recalibration on channel feature responses.

IV. Experiments

This section evaluates our proposed deraining model I-ADN and seven other representative deraining methods on six synthetic and one real-world datasets qualitatively and quantitatively. These methods contain the typical CNN-based models (RESCAN [20] and DIDMDN [26]) and the more recent deraining algorithms UMRL [22], PreNet [21], LP-Net [31] and ADN [29]. For the sake of fairness, we retrain these aforementioned models on the unified dataset with publicly available codes provided by authors. In addition, we also evaluate IADN on rainy datasets with different rain patterns (including rain-haze and raindrop scenarios) as well as other low-level vision tasks (including image dehazing and low-light enhancement) for comprehensive verification. The widely used evaluation metrics [66], including the reference-based indexes (Peak Signal to Noise Ratio (PSNR), Feature Similarity (FSIM), and Structural Similarity (SSIM)), as well as the reference-free indicators (Naturalness Image Quality Evaluator (NIQE) [67] and Spatial-Spectral Entropy-based Quality (SSEQ) index [68]) are applied as comparison criteria to evaluate the deraining performance in this study.

A. Implementation Details

1) Data Collection: Enlightened by [25], [39], [69], we collect about 13700 clean/rain image pairs for training, varying with various streak orientations and magnitudes under distinct rain conditions. The last 100 samples in dataset Rain800 provided by [25] are used for evaluation as **Test100** in this work. Besides, the synthetic rain dataset (**Test1**) commonly used in [26], [22], composed of a total of 1,200 rain images with different orientations and scales of rain streaks, is also adopted for another comparison. **Rain100H** [70] and **Rain100L** [70] are two rainy datasets enjoying the same background images but with distinctly different rain characteristics, heavy and low rain conditions respectively, and are also used for evaluation. Moreover, the novel rainy datasets **BDD350** and **COCO350** provided by [46] are also considered for comparison, where rainy images are of diverse streak orientations and magnitudes, and at the same time have complex imaging conditions such as night scenes. Furthermore, a real-world rainy dataset [26], [71], termed as (**Real127**), is used to estimate the generalization capability of our proposed deraining model.

2) Experimental Setup: In our baseline, the rain streaks are divided into 10 rain layers ($M = 10$), corresponding to 10 stages in IADN, each of which comprises $N = 3$ channel attention blocks (CABs) and one reconstruction layer. It is expressed as that the proposed IADN simultaneously estimates 10 rain layers, and further fuses them to regress the predicted residual rain image. Before packing the training samples into IADN, the training samples are conveniently cropped into small image patches with a unified size of 96×96 pixels to obtain the sample pairs. During training, the batch size is set to 32, and the learning rate is initialized as 5×10^{-4} with the decay rate of 0.9 every 10000 steps till 1×10^{-6}. After 60 epochs on training datasets, we obtain the optimal solution with the above settings (Only one NVIDIA Titan Xp GPU and an Intel I7-8700 CPU).

B. Ablation Study

1) Ablation Study on Basic Components: Since our baseline IADN is composed of several incorporative components, including non-local feature learning (NFL), layer-wise learning scheme (LWLS), and mixed attention fusion (MAF), we conduct ablation experiments on **Test1** dataset to analyze the effects of these strategies on the final deraining performance. Using our baseline IADN, we construct the first comparison model (**Model1**) by removing the NFL to evaluate the non-local feature representation. To demonstrate the layer-wise learning strategy, we design **Model2** by directly estimating the rain streaks without adopting the rain layer decomposition. To evaluate the mixed attention fusion, we construct **Model3** by replacing MAF with the pixel-wise summation to aggregate M rain layers. Another comparison model (**Model4**) is obtained by removing the NFL and LWLS. We obtain the last comparison model (**Model5**) by removing the NFL and MAF. For the sake of fairness, we keep these models with approximately equal parameters and the same settings during training.

Quantitative evaluation results, including PSNR, SSIM, and FSIM, are tabulated in Table I. From these scores, it is obvious that the complete model IADN exhibits great superiority over its incomplete substitutes, surpassing them by a large margin. For example, compared with **Model4**, IADN achieves better scores over Model4 (removing the NFL, LWLS, and MAF from IADN) by 0.66dB and 0.05 in PSNR and SSIM, respectively. We may attribute these advantages to the effective non-local feature fusion and layer-wise learning strategy of rain layers. The former allows the network to exploit the holistic self-similarity of rain information by learning the long-range dependency of global positions. The latter provides an effective scheme to aggregate multiple rain layers by rescaling feature responses, which is more effective and practical to cope with the overlapping phenomenon of heavy rain conditions. Besides, we also compare IADN with our previous work ADN [29], showing that IADN yields better deraining performance while it is 97.5% more efficient in model efficiency and 90.0% in inference time. These improvements over ADN have provided sufficient arguments on the effectiveness and efficiency of our proposed MSHFN further. Visual comparison results are shown in Figure. 7. It is evident that combining the non-local feature learning strategy with the layer-wise learning scheme to characterize the rain streak via a mixed attention block produces better restoration results, enjoying clearer image contents, richer details, and less color distortion. It is also noted that $Model3$ can benefit from the holistic feature representation and layer-wise learning, but still fails to recover credible image textures since it is limited to cope with the overlapped phenomenon among rain layers with the simple pixel-wise summation. From these ablation experiments, we

TABLE I

INVESTIGATIONS OF NON-LOCAL FEATURE LEARNING (NFL), LAYER-WISE LEARNING SCHEME (LWLS), MIXED ATTENTION FUSION (MAF) AS WELL AS THE BASIC MODULES (CHANNEL ATTENTION BLOCK (CAB) IN IADN AND RESIDUAL MEMORY BLOCK (RMB) IN ADN [29]) ON SYNTHETIC DATASET $Test1$. WE OBTAIN THE MODEL PARAMETERS (MILLION) AND AVERAGE INFERENCE TIME (SECOND) OF DERAINING ON IMAGES WITH THE SIZE OF **512**$\times$ **512**. IT IS NOTED THAT REMOVING LWLS FROM THE BASELINE, MAF WILL ALSO BE REMOVED IN THE COMPARISON MODEL. ADN DENOTES THE DERAINING MODEL PROPOSED IN OUR PREVIOUS CONFERENCE VERSION.

Model	Rain Image	Model1	Model2	Model3	Model4	Model5	ADN [29]	IADN
NFL	×	×	✓	✓	×	×	×	✓
LWLS	×	✓	×	✓	×	✓	✓	✓
MAF	×	✓	×	×	×	×	✓	✓
RCAB	×	✓	✓	✓	✓	✓	×	✓
RMB	×	×	×	×	×	×	✓	×
PSNR	22.15	32.02	31.75	32.07	31.63	31.89	31.43	**32.29**
SSIM	0.732	0.913	0.912	0.914	0.911	0.912	0.914	**0.916**
FSIM	0.881	0.957	0.956	0.957	0.956	0.956	0.955	**0.958**
Par.(M)	–	0.978	0.953	**0.945**	0.981	0.990	31.55	0.980
Time (S)	–	0.103	0.123	0.128	**0.094**	0.101	1.314	0.132

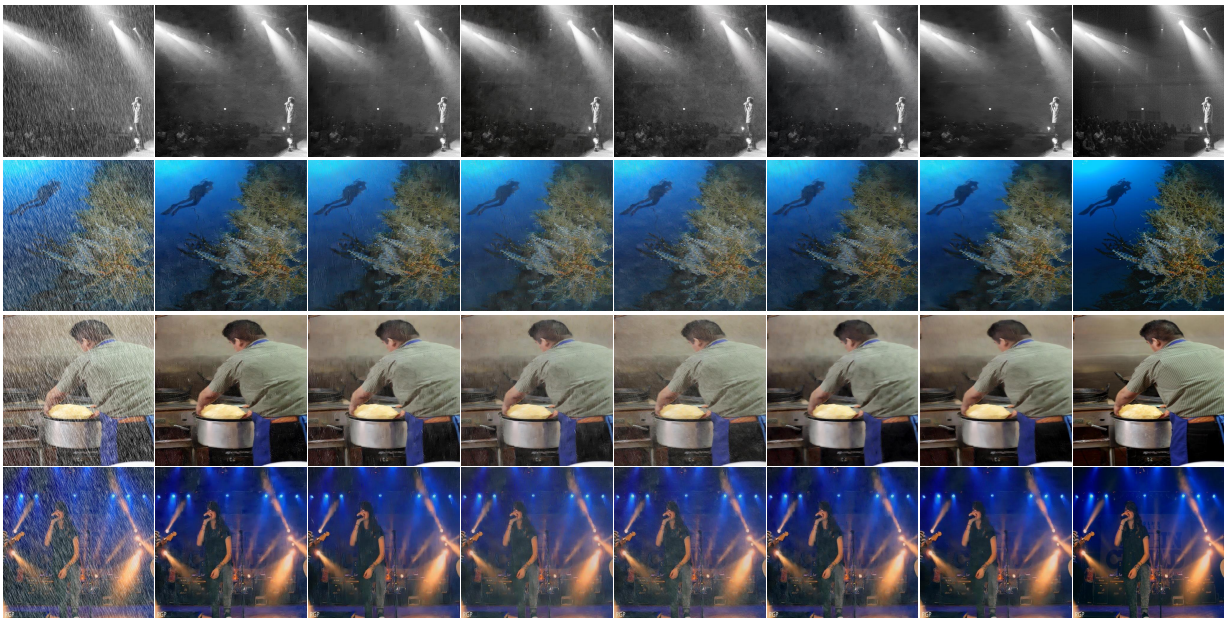

Fig. 7. Evaluation of the basic components on the **Test1** dataset.

validate the effectiveness of each component in our proposed improved attention-guided deraining network (IADN) for rain streaks removal.

2) Ablation Study on Parameters M and N: We assess the influence of the number of rain layers (**M**) and the depth of CAB (**N**) in each stage on deraining performance. Based on the baseline (**M** $=$ 10, **N** $=$ 3), we construct three comparison models, termed as IADN$_{M16N1}$, IADN$_{M12N2}$ and IADN$_{M8N4}$, while keeping approximately the equal model parameters. As shown in Table II, we obtain the best scores when **M** and **N** are set to 10 and 3 respectively. It indicates that using a reasonable decomposition strategy and network design contributes to better prediction results. Moreover, when simply increasing the number of CAB (IADN$_{M10N4}$, IADN$_{M10N3}$) or the decomposition level of rain layers (IADN$_{M10N4}$, IADN$_{M8N4}$), the deraining performance yields a slight gain (0.06dB and 0.14dB), but with additional 19% and 24% of the parameters respectively. Considering the tradeoff between efficiency and deraining performance, we set **M** and **N** to 10 and 3 respectively in the following experiments.

TABLE II

EVALUATION OF THE NUMBER OF STAGES (**N**) AND RESIDUAL ATTENTION BLOCKS (**M**), AS WELL AS MODEL PARAMETERS AND AVERAGE INFERENCE TIME (SECOND) ON **TEST1200** DATASET. IADN$_{NaMb}$ DENOTES THE MODEL WITH **N** $= a$ AND **M** $= b$.

Model	PSNR	SSIM	FSIM	Par.(M)	Time (S)
IADN$_{N10M4}$	32.35	0.917	0.958	1.170	0.148
IADN$_{N16M1}$	32.20	0.913	0.957	0.962	0.161
IADN$_{N12M2}$	32.25	0.915	0.958	0.946	0.143
IADN$_{N8M4}$	32.21	0.915	0.958	0.941	0.123
IADN$_{N10M3}$	32.29	0.916	0.958	0.980	0.132

C. Comparisons with State-of-the-arts

1) Synthesized Data: To verify the deraining performance of our IADN method, we compare IADN with six other repre-

TABLE III
THE COMPARISON RESULTS OF AVERAGE PSNR, SSIM AND FSIM ON **TEST100/TEST1** DATASETS. WE OBTAIN THE MODEL PARAMETERS (MILLION) AND AVERAGE INFERENCE TIME (SECOND) OF DERAINING ON IMAGES WITH SIZE OF **512**$\times$ **512**. * DENOTES THE RECURSIVE NETWORK USING THE PARAMETER SHARING STRATEGY.

Methods	RESCAN* [20]	DIDMDN [26]	UMRL [22]	PreNet* [21]	LPNet [31]	ADN [29]	IADN (Ours)
Dataset	**Test100/Test1**						
PSNR	25.00/30.51	22.56/29.65	24.41/30.55	24.81/31.36	23.39/25.00	25.20/31.43	**26.71/32.29**
SSIM	0.835/0.882	0.818/0.901	0.829/0.910	0.851/0.911	0.743/0.782	0.857/0.914	**0.865/0.916**
FSIM	0.909/0.944	0.899/0.950	0.910/0.955	0.916/0.955	0.861/0.899	0.920/0.955	**0.924/0.958**
Par.(M)	0.150	0.372	0.984	0.169	**0.007**	31.55	0.980
Time (S)	0.546	0.315	0.112	0.163	**0.027**	1.314	0.132

TABLE IV
THE COMPARISON RESULTS OF AVERAGE PSNR, SSIM AND FSIM ON **RAIN100H/RAIN100L** DATASETS.

Methods	RESCAN [20]	DIDMDN [26]	UMRL [22]	PreNet [21]	LPNet [31]	ADN [29]	IADN (Ours)
Dataset	**Rain100H/Rain100L**						
$PSNR$	26.36/29.80	17.35/25.23	26.01/29.18	26.77/32.44	16.00/25.57	26.86/28.32	**27.86/32.53**
$SSIM$	0.786/0.881	0.524/0.741	0.832/0.923	**0.858/0.950**	0.517/0.728	0.824/0.870	0.835/0.934
$FSIM$	0.864/0.919	0.726/0.861	0.876/0.940	**0.890/0.956**	0.699/0.832	0.872/0.889	0.875/0.942

TABLE V
THE COMPARISON RESULTS OF AVERAGE PSNR, SSIM AND FSIM ON **COCO350/BDD350** DATASETS.

Methods	RESCAN [20]	DIDMDN [26]	UMRL [22]	PreNet [21]	LPNet [31]	ADN [29]	IADN (Ours)
Dataset	**COCO350/BDD350**						
$PSNR$	17.04/16.71	17.12/16.85	17.68/17.36	17.53/16.90	15.43/14.87	17.37/17.02	**18.18/17.91**
$SSIM$	0.745/0.646	0.753/0.658	0.769/0.679	0.765/0.652	0.677/0.520	0.766/0.671	**0.790/0.719**

sentative deraining methods on six synthesized rainy datasets, including **Rain100H** [70], **Rain100L** [70], **Test100** [25], **Test1** [26], **BDD350** [46] and **COCO350** [46]. The competing methods include RESCAN [20], DIDMDN [26], UMRL [22], PreNet [21], LPNet [31] and ADN [29].

Qualitative results are tabulated in Table III, IV and V. On the **Test100** and **Test1** datasets, our proposed IADN method gains the best scores among all indicators over all competing models with acceptable model complexity. For example, the deraining performance shows significant improvements over the recurrent deraining technologies (RESCAN [20] and PreNet [21]), surpassing them by 1.78dB and 0.93dB on PSNR respectively while enjoying less inference time. In addition, when compared with more recent deraining methods, including the uncertainty guided multi-scale residual learning architecture (UMRL) [22], the light-weight deraining algorithm (LPNet) [31] as well as our previous work (ADN) [29], IADN is still very competitive by achieving better scores in all indexes. Moreover, Table IV tabulates the comparison results on **Rain100H/Rain100L** datasets, indicating that most of the deraining technologies obtain impressive performance on light rain condition, showing high consistency. However, only our IADN still performs favorably on the heavy rain condition, exhibiting great superiority over other competing methods in terms of PSNR. Moreover, in the novel rainy datasets (**BDD350** and **COCO350**), which are more challenging due to diverse rain conditions and night scenarios, our proposed IADN method still gains the best-evaluated scores. These results further demonstrate the effectiveness of our proposed layer-wise learning strategy and mixed attention fusion scheme, which are more suitable to characterize rain streaks under complex rain conditions, such as heavy and/or overlapped rain streaks.

Visual comparison results on **Test100/Test1** are reported in Figure. 8. Obviously, only our proposed IADN algorithm can recover clear and credible image textures while wiping out main rain streaks in these scenarios. For more convincing evidence, we also provide additional visual comparisons on **Rain100H/Rain100L** datasets, shown in Figure. 9. All can see that IADN can remove more rain streaks and better keep the image fidelity than these competing methods. Especially for the heavy rain condition, our proposed IADN method shows significant superiority in generating high-quality image contents. In contrast, other deraining methods fail to remove the rain streaks, and with obvious color distortion. We guess that these visible improvements on restoration quality may benefit from the elaborate design of the framework, including non-local residual learning, layer-wise learning strategy, and mixed attention fusion. These effective strategies are integrated into a unified framework, allowing the network to exploit the holistic spatial dependency to characterize rain streaks in a layer-wise manner effectively. Besides, it provides a novel and

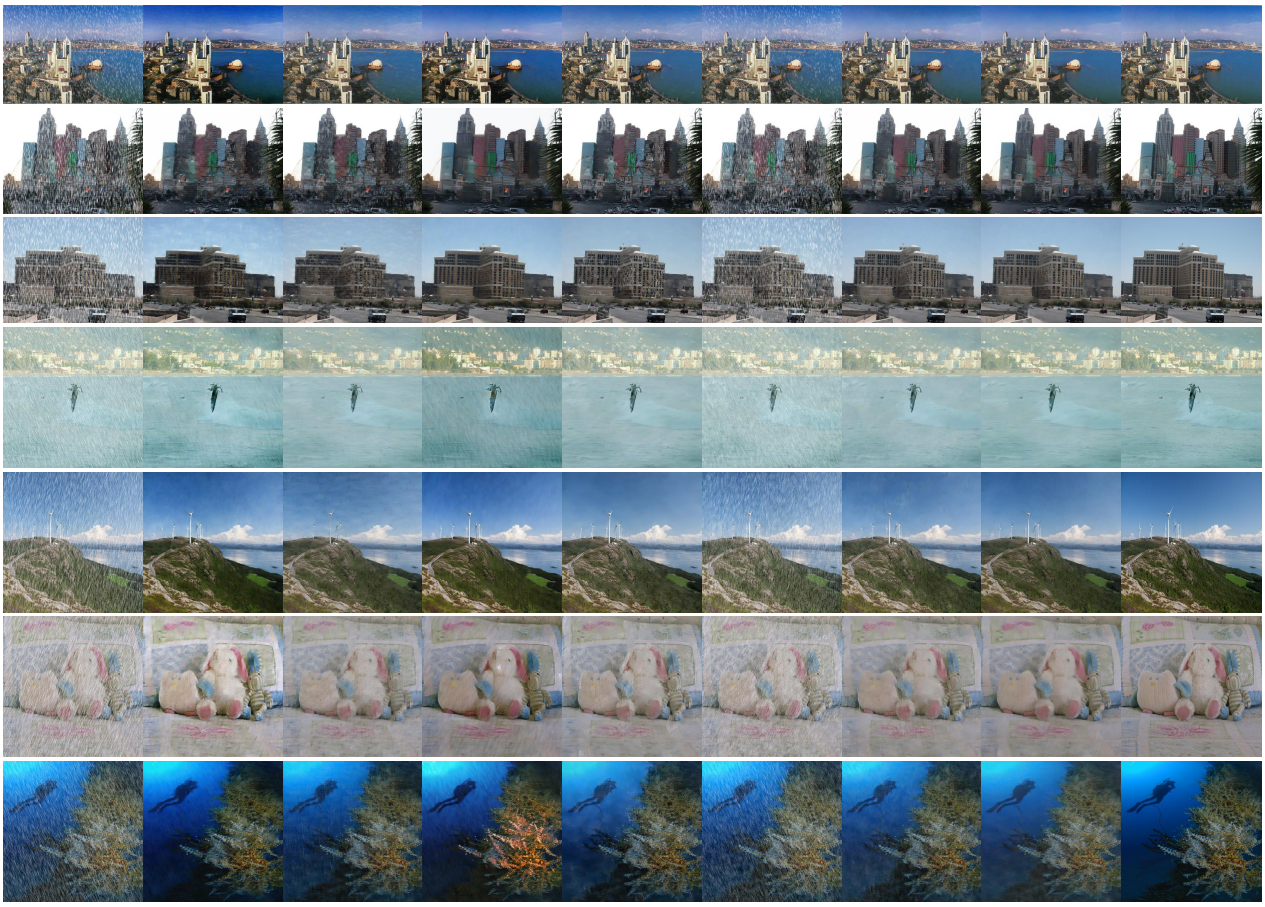

Fig. 8. Comparison results on $Test100/Test1$ datasets with six representative deraining methods, including RESCAN [20], DIDMDN [26], UMRL [22], PreNet [21], LPNet [31] and ADN [29].

TABLE VI
COMPARISON RESULTS OF AVERAGE NIQE AND SSEQ ON 127 REAL-WORLD SAMPLES. THESE TWO METRICS CAN EFFECTIVELY EVALUATE THE RESTORATION QUALITY WITHOUT REFERENCE. IN PARTICULAR, A LOWER VALUE INDICATES A HIGHER QUALITY IMAGE.

Methods	RESCAN [20]	DIDMDN [26]	UMRL [22]	LPNet [31]	PreNet [21]	ADN [29]	IADN (Ours)
NIQE	3.852	3.929	3.984	3.989	3.835	3.782	**3.769**
SSEQ	30.09	32.42	29.48	29.62	29.61	**28.82**	29.12

effective scheme for the restoration tasks with overlaps among rain streaks.

2) Real-world Data: To better estimate the robustness and generality of our proposed IADN model, additional experiments are performed on real-world scenarios. Inspired by [26], [71], we collect 127 real-world rainy samples, which are diverse in terms of contents as well as rain intensity and scales. Since the ground truth is unavailable, we thus introduce two additional quantitative indicators without reference, such as Naturalness Image Quality Evaluator (NIQE) [67] and Spatial-Spectral Entropy-based Quality (SSEQ) [68], to distinguish from pixel-based evaluation fashion. In particular, smaller scores of SSEQ and NIQE indicate better perceptual quality and clearer contents. Qualitative results are tabulated in Table VI. Still, our proposed IADN method has the lowest and second-lowest average values on NIQE and SSEQ, respectively. These results give additional evidence that our method generates an image with greater quality improvement. Moreover, the visual comparison results are reported in Figure. 10. Our proposed IADN recovers the clear and credible contents in the first two scenarios while effectively removing the main rain streak and snow. However, other competing methods generate results with visible rain streaks and snow reminded, and tend to blur the visual effect.

3) Rain-haze and Raindrop Data: To further verify the effectiveness of the proposed IADN method, we perform comparison experiments with the representative deraining methods (RESCAN [20], UMRL [22], and PreNet [21]) on multiple rain patterns, including the rain-haze and raindrop conditions. The RainCityscapes provided by [72] is a commonly used rain-haze dataset covering 30 scenarios and 1080 rainy samples with diverse rain-haze effects. The raindrop samples are col-

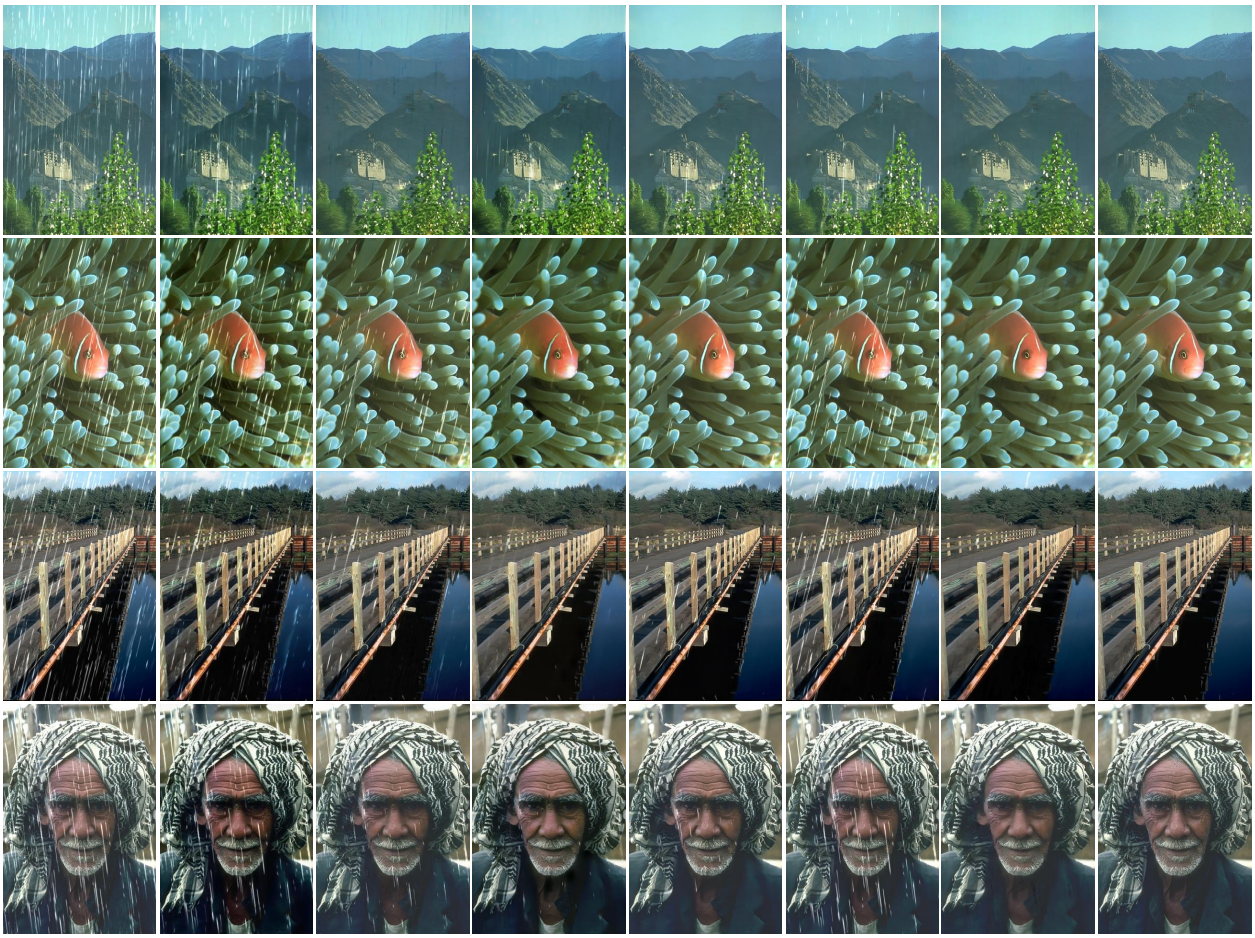

Fig. 9. The restoration results on **Rain100H/Rain100L** datasets various with different orientations and scales of rain streaks.

Fig. 10. Restoration results on real-world scenarios covering different rain or snow conditions.

Fig. 11. Restoration results on rain-haze scenarios.

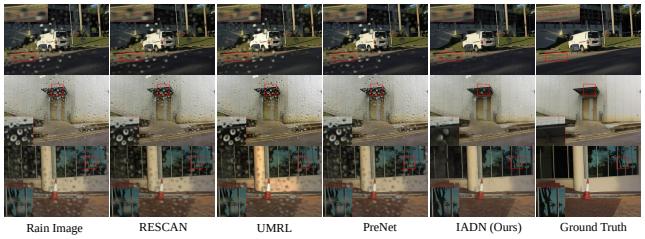

Fig. 12. Restoration results on raindrop scenarios.

lected from [32] with a total of 307 scenarios. Quantitative results in terms of the average PSNR and SSIM are tabulated in Table VII. It is obvious that our proposed IADN method is highly competitive as compared with these top-performing competitors. For example, IADN surpasses PreNet by 2.32dB in PSNR on the raindrop dataset. Visual comparisons in Figure. 11 and 12 show that the iterative algorithms (RESCAN and PreNet) aim at removing the rain streaks or raindrops that reside in the estimated image via a recurrent strategy, but tend to obtain over-smoothed results. In contrast, our proposed IADN model and the UMRL method can promote the image fidelity and discernibility while removing the main rain streaks. However, the results produced by UMRL show obvious color distortion (please refer to the "signboard" in the second scenario in Figure. 11.). Besides, our proposed IADN method is considerably more effective in removing raindrops than these competitors, producing results with fewer artifacts and clearer textural details.

TABLE VII
COMPARISON RESULTS OF AVERAGE PSNR AND SSIM ON RAIN-HAZE AND RAINDROP DATASETS.

Methods	RESCAN [20]	UMRL [22]	PreNet [21]	IADN (Ours)
Rain-haze Dataset [72]				
PSNR	17.73	**18.03**	17.42	17.90
SSIM	0.859	**0.875**	0.870	0.872
Raindrop Dataset [32]				
PSNR	23.27	21.96	23.32	**25.65**
SSIM	0.788	0.668	0.797	**0.824**

D. Evaluation via Other Low-level Vision Tasks

Due to the generality of the basic modules in feature representation, we extend our proposed improved attention-guided deraining network (IADN) to major low-level image enhancement tasks, such as image dehazing [81], [82] and low-light enhancement [78]. For the dehazing task, inspired by DADN [75], we choose 6000 synthetic paired hazy images provided by [83] to train both our proposed IADN as well as other representative dehazing methods, including AOD [73],

TABLE VIII
COMPARISON RESULTS OF IMAGE DEHAZING AND LOW-LIGHT ENHANCEMENT TASKS IN TERMS OF PSNR/SSIM. * DENOTES THAT THE RESULTS ARE OBTAINED WITH THE RELEASED TEST CODES DIRECTLY. ALTHOUGH MSBDN GAINS BETTER EVALUATION SCORES ON THE DEHAZING TASK, IT TAKES MORE THAN 31 TIMES THE PARAMETERS AND APPROXIMATELY 3 TIMES TRAINING SAMPLES THAN OUR PROPOSED IADN MODEL.

Methods	AOD [73]	EPDN [74]	DADN [75]	MSBDN* [76]	IADN (Ours)
Image Dehazing					
SOTS	20.35/0.896	24.17/0.944	23.29/0.859	**32.38/0.973**	26.48/0.951
TestSet A	18.10/0.849	20.33/0.918	20.08/0.837	**26.94/0.956**	22.91/0.927
Par. (M)	0.002	17.38	54.59	31.35	0.980
Methods	Zero-DCE [77]	RetinexNet [78]	DeepUPE* [79]	Kind [80]	IADN (Ours)
Low-light Image Enhancement					
LOL1000	15.53/0.420	18.23/0.792	16.64/0.773	**18.93/0.851**	18.71/0.839
TEST148	14.42/0.392	15.45/0.758	18.41/0.766	17.27/0.832	**18.86/0.840**
VOC144	15.42/0.348	18.08/0.749	19.43/0.781	**22.19/0.834**	19.46/0.781
Par. (M)	0.079	0.445	0.999	8.016	0.980

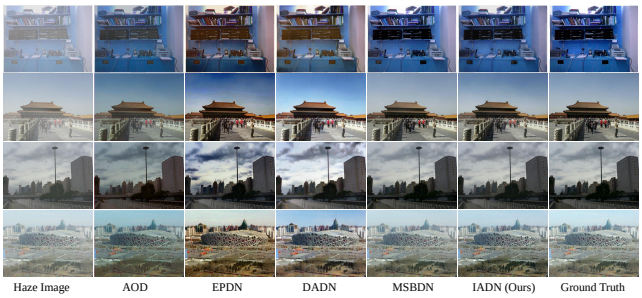

Fig. 13. Comparison results of image dehazing on synthetic datasets.

EPDN [74] and DADN [75]. Since MSBDN [76] has not released the training codes, we directly apply its pre-trained model, which is trained on 16000 image pairs for evaluation. To verify our proposed IADN method, we compare it with other competing methods on two commonly used test datasets (SOTS and TestSet A). SOTS is the test subset of the RESIDE dataset [83] which contains 500 indoor hazy images and 500 outdoor hazy images. TestSet A is the test subset of the NYU-v2 database [84], covering 3169 haze images. Quantitative results in terms of PSNR, SSIM and the model parameters are tabulated in Table VIII. Obviously, our proposed IADN method has superior performance over the competing models, surpassing the DADN method by 3.19dB and 2.83dB in PSNR on the SOTS and TestSet A datasets, respectively. Although the top-performing method MSBDN gains better evaluation scores, it is specially designed for image dehazing with more than 31 times the parameters and approximately 3 times training samples than our proposed IADN model. Visual comparisons are shown in Figure. 13. It is clear that our proposed IADN model achieves comparable results on image dehazing task against these representative dehazing methods for producing better visual effects with clearer contents and more credible textures. Other competitors fail to remove the haze effect and tend to blur the contents and produce results with obvious color distortion.

For the low-light enhancement task, inspired by [78], we use the LOL dataset, which includes 500 low/normal-light image pairs, to train our proposed IADN model as well as other representative low-light enhancement methods, including RetinexNet [78], Zero-DCE [77] and Kind [80]. Three commonly used test datasets (LOL1000 [78], TEST148 [85]

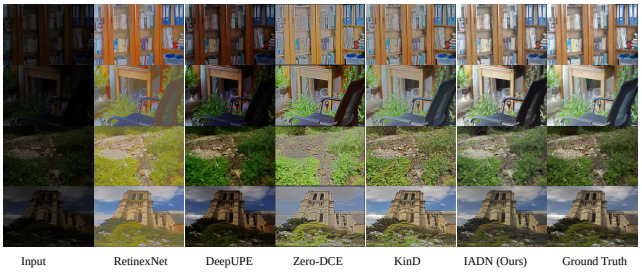

Fig. 14. Comparison results of low-light image enhancement on synthetic datasets.

and VOC144 [86]) are used for evaluation. Quantitative results in Table VIII show that our proposed IADN model exhibits great competitiveness against these representative methods specially designed for low-light image enhancement task. Visual comparisons are shown in Figure. 14. It is clear that other competitors tend to relight the dark regions but with obvious color distortion and overexposure, while our proposed IADN model can restore clear image details while adjusting the exposure and illumination in a more visually pleasing manner.

V. Conclusions

In this paper, we propose a novel improved attention-guided deraining network (IADN) for rain streak removal. IADN especially models multiple rain layers under the stages of the network, corresponding to the multi-level abstracts of rain information in one rainy image. Besides, we propose an improved non-local block to exploit the self-similarity of similar rain patterns, and further apply the channel attention blocks (CAB) to extract rich textual information from the non-local features to represent rain layers at each stage. Moreover, a mixed attention block (MAB) is constructed to guide the fusion of rain layers by focusing on the overlapped phenomenon at holistic and local dimensions. Extensive experiments on several synthetic and real-world rain datasets, as well as other low-level computer vision tasks (image dehazing and low-light enhancement) have demonstrated the superiority and generalization capability of our proposed IADN algorithm over other state-of-the-art methods.

References

[1] Q. Leng, M. Ye, and Q. Tian, "A survey of open-world person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 4, pp. 1092–1108, 2020.
[2] W. Yu, T. Yang, and C. Chen, "Towards resolving the challenge of long-tail distribution in UAV images for object detection," in *WACV*, 2020.
[3] P. Foggia, A. Saggese, and M. Vento, "Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 9, pp. 1545–1556, 2015.
[4] L. Zhu, Z. Deng, X. Hu, H. Xie, X. Xu, J. Qin, and P. Heng, "Learning gated non-local residual for single-image rain streak removal," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1–1, 2020.
[5] Q. Wu, L. Wang, K. N. Ngan, H. Li, F. Meng, and L. Xu, "Subjective and objective de-raining quality assessment towards authentic rain image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3883–3897, 2020.
[6] J. L. Yin, Y. C. Huang, B. H. Chen, and S. Z. Ye, "Color transferred convolutional neural networks for image dehazing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3957–3967, 2020.
[7] Y. Peng, Z. Lu, F. Cheng, Y. Zheng, and S. Huang, "Image haze removal using airlight white correction, local light filter, and aerial perspective prior," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 5, pp. 1385–1395, 2020.
[8] M. Teichmann, M. Weber, M. Zoellner, R. Cipolla, and R. Urtasun, "Multinet: Real-time joint semantic reasoning for autonomous driving," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 1013–1020.
[9] Z. Shao, L. Wang, Z. Wang, W. Du, and W. Wu, "Saliency-aware convolution neural network for ship detection in surveillance video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 781–794, 2020.
[10] P. C. Barnum, S. Narasimhan, and T. Kanade, "Analysis of rain and snow in frequency space," *IJCV*, vol. 86, no. 2-3, p. 256, 2010.
[11] L.-W. Kang, C.-W. Lin, and Y.-H. Fu, "Automatic single-image-based rain streaks removal via image decomposition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1742–1755, 2011.
[12] K. Garg and S. K. Nayar, "When does a camera see rain?" in *ICCV*, vol. 2, 2005, pp. 1067–1074.
[13] W. Ren, J. Tian, Z. Han, A. Chan, and Y. Tang, "Video desnowing and deraining based on matrix decomposition," in *CVPR*, 2017, pp. 4210–4219.
[14] Y.-L. Chen and C.-T. Hsu, "A generalized low-rank appearance model for spatio-temporally correlated rain streaks," in *CVPR*, 2013, pp. 1968–1975.
[15] K. Jiang, Z. Wang, P. Yi, and J. Jiang, "Hierarchical dense recursive network for image super-resolution," *attern Recognition*, vol. 107, p. 107475, 2020.
[16] T. Yang, S. Zhu, and C. Chen, "Gradaug: A new regularization method for deep neural networks," in *NIPS*, vol. 33, 2020.
[17] P. Yi, Z. Wang, K. Jiang, Z. Shao, and J. Ma, "Multi-temporal ultra dense memory network for video super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2503–2516, Aug 2020.
[18] H. Zhang and V. M. Patel, "Convolutional sparse and low-rank coding-based rain streak removal," in *WACV*, 2017, pp. 1259–1267.
[19] X. Wang, J. Chen, K. Jiang, Z. Han, W. Ruan, Z. Wang, and C. Liang, "Single image de-raining via clique recursive feedback mechanism," *Neurocomputing*, vol. 417, pp. 142–154, 2020.
[20] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha, "Recurrent squeeze-and-excitation context aggregation net for single image deraining," in *ECCV*, 2018, pp. 254–269.
[21] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng, "Progressive image deraining networks: a better and simpler baseline," in *CVPR*, 2019, pp. 3937–3946.
[22] R. Yasarla and V. M. Patel, "Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining," in *CVPR*, 2019, pp. 8405–8414.
[23] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, "Clearing the skies: A deep network architecture for single-image rain removal," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2944–2956, 2017.
[24] R. Yasarla, V. A. Sindagi, and V. M. Patel, "Syn2real transfer learning for image deraining using gaussian processes," in *CVPR*, 2020, pp. 2723–2733.
[25] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3943–3956, 2020.
[26] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," in *CVPR*, 2018, pp. 695–704.
[27] R. Yasarla and V. M. Patel, "Confidence measure guided single image de-raining," *IEEE Trans. Image Process.*, vol. 29, pp. 4544–4555, 2020.
[28] 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 *CVPR*, 2019, pp. 3838–3847.
[29] K. Jiang, Z. Wang, P. Yi, C. Chen, Y. Yang, X. Tian, and J. Jiang, "Attention-guided deraining network via stage-wise learning," in *ICASSP*, 2020, pp. 2618–2622.
[30] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *CVPR*, 2018, pp. 7794–7803.
[31] X. Fu, B. Liang, Y. Huang, X. Ding, and J. Paisley, "Lightweight pyramid networks for image deraining," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 6, pp. 1794–1807, 2020.
[32] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *CVPR*, 2018, pp. 2482–2491.

[33] R. Li, R. T. Tan, and L. F. Cheong, "All in one bad weather removal using architectural search," in *CVPR*, 2020, pp. 3172–3182.

[34] P. Yi, Z. Wang, K. Jiang, J. Jiang, T. Lu, and J. Ma, "A progressive fusion generative adversarial network for realistic and consistent video super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2020.

[35] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *ECCV*, 2018, pp. 286–301.

[36] J. Bossu, N. Hautière, and J.-P. Tarel, "Rain or snow detection in image sequences through use of a histogram of orientation of streaks," *IJCV*, vol. 93, no. 3, pp. 348–367, 2011.

[37] J. Xu, W. Zhao, P. Liu, and X. Tang, "An improved guidance image based method to remove rain and snow in a single image," *Computer and Information Science*, vol. 5, no. 3, p. 49, 2012.

[38] Y. Luo, Y. Xu, and H. Ji, "Removing rain from a single image via discriminative sparse coding," in *CVPR*, 2015, pp. 3397–3405.

[39] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, "Rain streak removal using layer priors," in *CVPR*, 2016, pp. 2736–2744.

[40] R. Li, L. F. Cheong, and R. T. Tan, "Single image deraining using scale-aware multi-stage recurrent network," 2017.

[41] X. Lin, L. Ma, B. Sheng, Z. Wang, and W. Chen, "Utilizing two-phase processing with fbls for single image deraining," *IEEE Transactions on Multimedia*, pp. 1–1, 2020.

[42] H. Wang, Z. Yue, Q. Xie, Q. Zhao, and D. Meng, "From rain removal to rain generation," 2020.

[43] S. Deng, M. Wei, J. Wang, Y. Feng, L. Liang, H. Xie, F. L. Wang, and M. Wang, "Detail-recovery image deraining via context aggregation networks," in *CVPR*, 2020, pp. 14 560–14 569.

[44] G. Li, X. He, W. Zhang, H. Chang, L. Dong, and L. Lin, "Non-locally enhanced encoder-decoder network for single image de-raining," pp. 1056–1064, 2018.

[45] P. Bo, Z. Deming, J. Junjun, and L. Xianming, "Single image deraining via scale-space invariant attention neural network," *ACMMM*, pp. 375–383, 2020.

[46] K. Jiang, Z. Wang, P. Yi, C. Chen, B. Huang, Y. Luo, J. Ma, and J. Jiang, "Multi-scale progressive fusion network for single image deraining," in *CVPR*, 2020, pp. 8343–8352.

[47] W. Yang, R. T. Tan, S. Wang, Y. Fang, and J. Liu, "Single image deraining: From model-based to data-driven and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2020.

[48] H. Lin, Y. Li, X. Fu, X. Ding, Y. Huang, and J. Paisley, "Rain oer me: Synthesizing real rain to derain with data distillation," *IEEE Trans. Image Process.*, vol. 29, pp. 7668–7680, 2020.

[49] H. Wang, Q. Xie, Q. Zhao, and D. Meng, "A model-driven deep neural network for single image rain removal," in *CVPR*, 2020, pp. 3100–3109.

[50] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *CVPR*, vol. 2, 2005, pp. 60–65.

[51] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space," in *ICIP*, vol. 1, 2007, pp. I–313.

[52] D. Liu, B. Wen, Y. Fan, C. C. Loy, and T. S. Huang, "Non-local recurrent network for image restoration," in *NIPS*, 2018, pp. 1673–1682.

[53] P. Yi, Z. Wang, K. Jiang, J. Jiang, and J. Ma, "Progressive fusion video super-resolution network via exploiting non-local spatio-temporal correlations," in *ICCV*, 2019, pp. 3106–3115.

[54] Z. Zhu, M. Xu, S. Bai, T. Huang, and X. Bai, "Asymmetric non-local neural networks for semantic segmentation," in *ICCV*, 2019, pp. 593–602.

[55] M. Jaderberg, K. Simonyan, A. Zisserman *et al.*, "Spatial transformer networks," in *NIPS*, 2015, pp. 2017–2025.

[56] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *CVPR*, 2017, pp. 3156–3164.

[57] J. Fu, H. Zheng, and T. Mei, "Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition," in *CVPR*, 2017, pp. 4438–4446.

[58] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *CVPR*, 2018, pp. 7132–7141.

[59] K. Jiang, Z. Wang, P. Yi, G. Wang, K. Gu, and J. Jiang, "Atmfn: Adaptive-threshold-based multi-model fusion network for compressed face hallucination," *IEEE Trans. Multimedia*, vol. 22, no. 10, pp. 2734–2747, Oct 2020.

[60] K. Jiang, Z. Wang, P. Yi, T. Lu, J. Jiang, and Z. Xiong, "Dual-path deep fusion network for face image hallucination," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2020.

[61] S. Anwar and N. Barnes, "Real image denoising with feature attention," in *ICCV*, 2019, pp. 3155–3164.

[62] Y. Yang and H. Lu, "Single image deraining using a recurrent multi-scale aggregation and enhancement network," in *ICME*, 2019, pp. 1378–1383.

[63] W. Lai, J. Huang, N. Ahuja, and M. Yang, "Fast and accurate image super-resolution with deep laplacian pyramid networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2599–2613, 2019.

[64] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *CVPR*, 2016, pp. 1646–1654.

[65] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "Gcnet: Non-local networks meet squeeze-excitation networks and beyond," in *ICCVW*, 2019, pp. 1971–1980.

[66] G. Wang, Z. Wang, K. Gu, L. Li, Z. Xia, and L. Wu, "Blind quality metric of dibr-synthesized images in the discrete wavelet transform domain," *IEEE Trans. Image Process.*, vol. 29, pp. 1802–1814, 2020.

[67] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2012.

[68] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Processing: Image Communication*, vol. 29, no. 8, pp. 856–863, 2014.

[69] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *CVPR*, 2017, pp. 3855–3863.

[70] 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 *CVPR*, 2017, pp. 1357–1366.

[71] W. Wei, D. Meng, Q. Zhao, Z. Xu, and Y. Wu, "Semi-supervised transfer learning for image rain removal," in *CVPR*, 2019, pp. 3877–3886.

[72] X. Hu, C.-W. Fu, L. Zhu, and P.-A. Heng, "Depth-attentional features for single-image rain removal," in *CVPR*, 2019, pp. 8022–8031.

[73] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *ICCV*, vol. 1, no. 4, 2017, p. 7.

[74] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced pix2pix dehazing network," in *CVPR*, 2019, pp. 8160–8168.

[75] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," in *CVPR*, 2020, pp. 2808–2817.

[76] D. Hang, P. Jinshan, H. Zhe, L. Xiang, W. Fei, and Y. Ming-Hsuan, "Multi-scale boosted dehazing network with dense feature fusion," in *CVPR*, 2020, pp. 2154–2164.

[77] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *CVPR*, 2020, pp. 1777–1786.

[78] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *BMVC*, 2018, p. 155.

[79] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *CVPR*, 2019, pp. 6849–6857.

[80] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020.

[81] Y. Pang, J. Xie, and X. Li, "Visual haze removal by a unified generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 11, pp. 3211–3221, 2019.

[82] C. Son and X. Zhang, "Near-infrared fusion via color regularization for haze and color distortion removals," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 11, pp. 3111–3126, 2018.

[83] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, 2019.

[84] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *ECCV*, 2012, pp. 746–760.

[85] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *arXiv preprint arXiv:1906.06972*, 2019.

[86] J. W. Feifan Lv, Feng Lu and C. Lim, "Mbllen: Low-light image/video enhancement using cnns," in *BMVC*, 2018.