Multi-Scale Progressive Fusion Network for Single Image Deraining

Kui Jiang[1] Zhongyuan Wang[1*] Peng Yi[1] Chen Chen[2*]
Baojin Huang[1] Yimin Luo[3] Jiayi Ma[1] Junjun Jiang[4]
[1]Wuhan University [2]University of North Carolina at Charlotte
[3]King's College London [4]Harbin Institute of Technology

Abstract

Rain streaks in the air appear in various blurring degrees and resolutions due to different distances from their positions to the camera. Similar rain patterns are visible in a rain image as well as its multi-scale (or multi-resolution) versions, which makes it possible to exploit such complementary information for rain streak representation. In this work, we explore the multi-scale collaborative representation for rain streaks from the perspective of input image scales and hierarchical deep features in a unified framework, termed multi-scale progressive fusion network (MSPFN) for single image rain streak removal. For the similar rain streaks at different positions, we employ recurrent calculation to capture the global texture, thus allowing to explore the complementary and redundant information at the spatial dimension to characterize target rain streaks. Besides, we construct multi-scale pyramid structure, and further introduce the attention mechanism to guide the fine fusion of these correlated information from different scales. This multi-scale progressive fusion strategy not only promotes the cooperative representation, but also boosts the end-to-end training. Our proposed method is extensively evaluated on several benchmark datasets and achieves the state-of-the-art results. Moreover, we conduct experiments on joint deraining, detection, and segmentation tasks, and inspire a new research direction of vision task driven image deraining. The source code is available at https://github.com/kuihua/MSPFN.

1. Introduction

Due to substantial degradation of the image content in rain images and videos, traditional image enhancement algorithms [27] struggle to make desirable improvements on image quality. Therefore, developing specialized solutions for image deraining is imperative to a wide range of tasks [12], *e.g.* object detection and semantic segmentation.

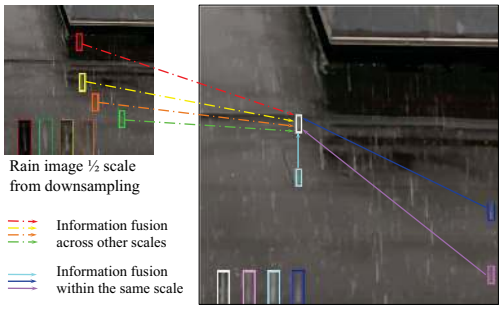


Figure 1. Demonstration of the collaborative representation of rain streaks. Specifically, similar rain patterns among rain streaks, both within the same scale (highlighted in cyan, pink and dark blue boxes) or cross different scales (highlighted in red, yellow, orange and green boxes), can help reconstruct the target rain streak (white box in the original rain image) with the complementary information (*e.g.* similar appearance, formation, etc.).

Traditional deraining methods [2, 1, 5, 9, 32] use simple linear-mapping transformations and are not robust to variations of the input [11], *e.g.*, rain streaks with various directions, densities and sizes. Recently, deep-learning based methods [6, 35, 16] which operate with convolutional and non-linear layers have witnessed remarkable advantages over traditional methods. Despite obvious improvements on feature representation brought by those methods [6, 16], their single-scale frameworks can hardly capture the inherent correlations of rain streaks across scales.

The repetitive samples of rain streaks in a rain image as well as its multi-scale versions (multi-scale pyramid images) may carry complementary information (*e.g.* similar appearance) to characterize target rain streaks. As illustrated in Fig. 1, the rain streaks (highlighted in the white box) in the original rain image share the similar rain patterns with the rain streaks (highlighted in the cyan, pink and dark blue boxes) at different positions as well as those (highlighted in the red, yellow, orange and green boxes) in the 1/2 scale rain image. Therefore, rain streaks both from the same scale (solid arrows) and across different scales (dashed ar-

*Corresponding author

rows) encode complementary or redundant information for feature representation, which would help deraining in the original image. This correlation of image contents across scales has been successfully applied to other computer vision tasks [10, 33]. Recently, authors in [8, 44] construct pyramid frameworks to exploit the multi-scale knowledge for deraining. Unfortunately, those exploitations fail to make full use of the correlations of multi-scale rain streaks (although restricted to a fixed scale-factor of 2 [10]). For example, Fu *et al*. [8] decompose the rain image into different pyramid levels based on its resolution, and then individually solve the restoration sub-problems at the specific scale space through several parallel sub-networks. Such decomposition strategy is the basic idea of many recurrent deraining frameworks [19]. Unlike [8] completing the deraining task from each individual resolution level, Zheng *et al*. [44] present a density-specific optimization for rain streak removal in a coarse-to-fine fashion, and gradually produce the rain-free image stage-by-stage [15]. However, there are no direct communications of the inter-level features across cascaded pyramid layers except for the final outputs, thus failing to take all-rounded advantages of the correlated information of rain streaks across different scales. Consequently, these methods [8, 44] are still far from producing the desirable deraining results with the limited exploitation and utilization of multi-scale rain information.

To address these limitations of the prior works, we explore the multi-scale representation from input image scales and deep neural network representations in a unified framework, and propose a multi-scale progressive fusion network (MSPFN) to exploit the correlated information of rain streaks across scales for single image deraining. Specifically, we first generate the Gaussian pyramid rain images using Gaussian kernels to down-sample the original rain image in sequence. A coarse-fusion module (CFM) (§3.1) is designed to capture the global texture information from these multi-scale rain images through recurrent calculation (Conv-LSTM), thus enabling the network to cooperatively represent the target rain streak using similar counterparts from global feature space. Meanwhile, the representation of the high-resolution pyramid layer is guided by previous outputs as well as all low-resolution pyramid layers. A fine-fusion module (FFM) (§3.2) is followed to further integrate these correlated information from different scales. By using the channel attention mechanism, the network not only discriminatively learns the scale-specific knowledge from all preceding pyramid layers, but also reduces the feature redundancy effectively. Moreover, multiple FFMs can be cascaded to form a progressive multi-scale fusion. Finally, a reconstruction module (RM) is appended to aggregate the coarse and fine rain information extracted respectively from CFM and FFM for learning the residual rain image, which is the approximation of real rain streak distribution. The overall framework is outlined in Fig. 2. The main contributions of this paper are as follows:

- We uncover the correlations of rain streaks in an image and propose a novel multi-scale progressive fusion network (MSPFN) which collaboratively represents rain streaks from multiple scales via the pyramid representation.
- To better characterize rain streaks of different scales, we devise three basic modules, coarse-fusion module (CFM), fine-fusion module (FFM) and reconstruction module (RM), to effectively extract and integrate the multi-scale information. In these modules, the complementary information of similar patterns with rain streaks, both within the same scale or across different scales (pyramid layers), is progressively fused to characterize the rain streaks distribution in a collaborative/cooperative manner.
- Apart from achieving the state-of-the-art deraining performance in terms of the conventional quantitative measurements (*e.g*. PSNR and SSIM), we build several synthetic rain datasets based on COCO [3] and BDD [38] datasets for joint image deraining, detection and segmentation tasks. To the best of our knowledge, we are the first to apply mainstream vision-oriented tasks (detection and segmentation) for comprehensively evaluating the deraining performance.

2. Related Work

In the last few years, substantial improvements [24, 18, 4, 17] have been observed on rain image restoration. In this work, we mainly focus on single image deraining because it is more challenging.

2.1. Single Image Deraining

Previous traditional methods for single image deraining [5, 14] fail under the complex rain conditions and produce degraded image contents due to the limited linear-mapping transformation. Very recently, deep-learning based approaches [24, 29, 39] have emerged for rain streak removal and demonstrated impressive restoration performance. For example, Fu *et al*. [6] introduce a three-layer convolutional neural network (CNN) to estimate and remove rain streaks from its rain-contaminated counterpart. To better represent rain streaks, Zhang *et al*. [40] take the rain density into account and present a multi-task CNN for joint rain density estimation and deraining. Later, Zhang *et al*. [41] further incorporate quantitative, visual and discriminative performance into the objective function, and propose a conditional generative adversarial network for rain streak removal. In order to alleviate the learning difficulty, recurrent frameworks [19, 36, 26] are designed to remove rain streaks in a stage-wise manner.

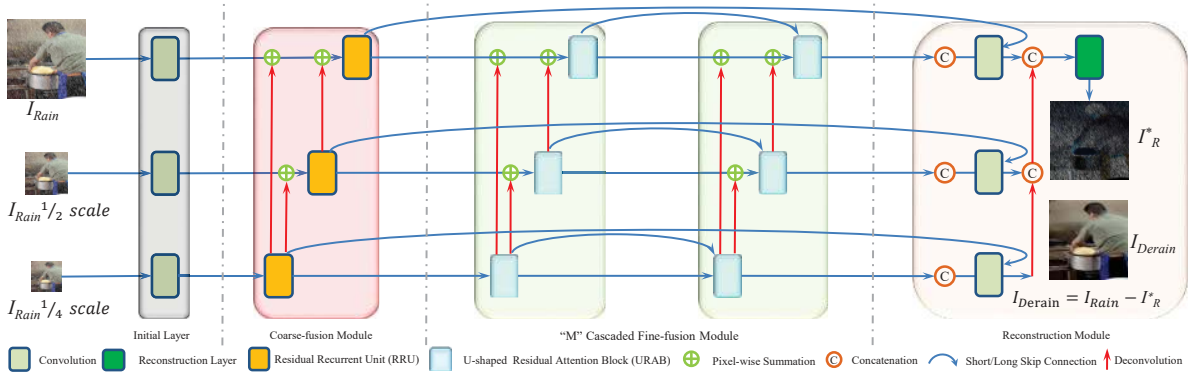


Figure 2. Outline of the proposed multi-scale progressive fusion network (MSPFN). We set the pyramid level to 3 as an example. MSPFN consists of four parts: initial feature extraction, coarse fusion, fine fusion, and rain streak reconstruction, which are combined to regress the residual rain image I^*_R. We produce the rain-free image I_{Derain} by subtracting I^*_R from the original rain image I_{Rain}. The goal is to make I_{Derain} as close as possible to the rain free image I_{Clean}.

2.2. Multi-scale Learning

Rain streaks in the air show the apparent self-similarity, both within the same scale or across different scales, which makes it possible to exploit the correlated information across scales for rain streak representation. However, most existing deraining methods [16, 40] ignore the underlying correlations of rain streaks across different scales. Only a few attempts [8, 44] have been made to exploit the multi-scale knowledge. Fu *et al*. [8] decompose the restoration task into multiple subproblems and employ a set of parallel subnetworks to individually estimate the rain information in a specific pyramid scale space. However, it does not exploit and utilize the correlated information among these pyramid layers. Different from the parallel pyramid framework in [8], Zheng *et al*. [44] propose the cascaded pyramid network, which is similar to LapSRN [15], to iteratively remove rain streaks. However, only the high-level features are used to help the adjacent pyramid representation, which results in losing some useful hierarchical and scale features in a deep cascaded network. The significance of these features produced at different stages has been verified on image reconstruction tasks [28, 43].

Different from these methods [8, 44], in this work we introduce a novel framework MSPFN to achieve the collaborative representation of rain streaks across different scales, where the rich multi-scale rain information extracted from the Gaussian pyramid images is progressively aggregated along the pyramid layers and stages of the network. As a result, our predicted rain streak distribution is more accurate via the multi-scale collaborative representation.

3. Proposed Method

Fig. 2 shows the overall pipeline of our proposed multi-scale progressive fusion network (MSPFN) for image deraining by excavating and exploiting the inherent correlations of rain streaks across different scales. We present the details of each building block and the loss function in the following.

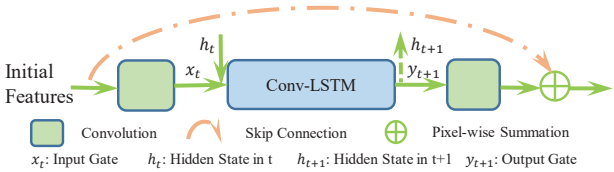


Figure 3. Pipeline of the proposed residual recurrent units (RRU).

3.1. Multi-scale Coarse Fusion

For a given rain image, our method first generates the Gaussian pyramid rain images using Gaussian kernels to down-sample the original rain image into different scales, *e.g*. 1/2 and 1/4. The network takes as input the pyramid rain images and extracts the shallow features through multiple parallel initial convolution layers (see the first block of "initial layer" in Fig. 2). Based on the initial features from each scale, the coarse-fusion module (CFM) then performs the deep extraction and fusion of multi-scale rain information through several parallel residual recurrent units (RRU), as shown in Fig. 3. The reasons for designing CFM are three folds: (a) To exploit the repetition of rain streaks under the same scale, we apply the recurrent calculation and residual learning to capture the global texture information, making it possible to cooperatively represent target rain streaks. More accurately, we introduce Conv-LSTM to model the information flow of context textures at spatial dimension with the recursive memory, where the contextual texture correlations are transformed into structured cyclic dependencies to capture the complementary or redundant rain information (*e.g*. the solid arrows in Fig. 1). (b) The multi-scale structure provides an alternative solution to greatly increase the receptive filed to cover more contents while maintaining a

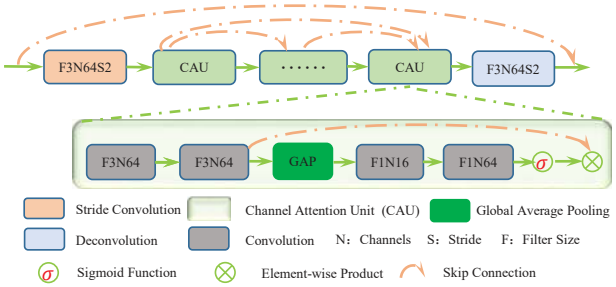


Figure 4. Pipeline of our proposed U-shaped residual attention block (URAB). URAB is composed of several cascaded channel attention units (CAUs) to promote the fusion of the multi-scale rain information and reduce the feature redundancy by focusing on the most useful channels.

shallow depth. (c) The high-resolution representations benefit from the outputs of previous stages as well as all low-resolution pyramid layers via iterative sampling and fusion.

3.2. Multi-scale Fine Fusion

The outputs of CFM go through the fine-fusion module (FFM) to refine the correlated information from different scales. As shown in Fig. 2, FFM enjoys the similar multi-scale structure with CFM for convenience. Unlike CFM, we introduce the channel attention unit (CAU) to enhance the discriminative learning ability of the network through focusing on the most informative scale-specific knowledge, making the cooperative representation more efficient. To alleviate the computation burden, we apply the strided convolution to reduce the spatial dimension of features, and finally utilize the deconvolution layer to increase the resolution to avoid losing resolution information, resulting in the U-shaped residual attention block (URAB). As depicted in Fig. 4, URAB is composed of several CAUs, along with the short skip connections to help the fine representation of multi-scale rain information. Moreover, long skip connections are used between cascaded FFMs to achieve progressive fusion of multi-scale rain information as well as to facilitate the effective backward propagation of the gradient.

3.3. Rain Streak Reconstruction

To learn the final residual rain image, we further integrate both low- and high-level multi-scale features respectively from CFM and FFM via a reconstruction module (R-M), schematically depicted in Fig. 2. Specifically, the outputs from CFM are concatenated with the outputs from the last FFM, and then a convolution layer is used to learn the channel interdependence and rescale the feature values from the two modules. Similarly, the iterative sampling and fusion of rain information across different pyramid layers are implemented to estimate the residual rain image.

3.4. Loss Function

Mean squared error (MSE) is the commonly used loss to train the network [40, 34]. However, it usually produces blurry and over-smoothed visual effect with the loss of high-frequency textures due to the squared penalty. In this work, we perform the successive approximation to the real rain streak distribution I_R with the guidance of the Charbonnier penalty function [15], which is more tolerant of small errors and holds better convergence during training. The function is expressed as

$$L_{con} = \sqrt{(I_R^* - I_R)^2 + \varepsilon^2}. \quad (1)$$

In Equation (1), I_R^* denotes the predicted residual rain image. The predicted rain-free image I_{Derain} is generated by subtracting I_R^* from its rain-contaminated counterpart I_{Rain}. The penalty coefficient ε is empirically set to 10^{-3}.

In order to further improve the fidelity and authenticity of high-frequency details while removing rain streaks, we propose the additional edge loss to constrain the high-frequency components between the ground truth I_{Clean} and the predicted rain-free image I_{Derain}. The edge loss is defined as

$$L_{edge} = \sqrt{(Lap(I_{Clean}) - Lap(I_{Derain}))^2 + \varepsilon^2}. \quad (2)$$

In Equation (2), $Lap(I_{Clean})$ and $Lap(I_{Derain})$ denote the edge maps respectively extracted from I_{Clean} and I_{Derain} via the Laplacian operator [13]. Then, the total loss function is given by

$$L = L_{con} + \lambda \times L_{edge}, \quad (3)$$

where the weight parameter λ is empirically set to 0.05 to balance the loss terms.

4. Experiments and Discussions

We conduct extensive experiments on several synthetic and real-world rain image datasets [7, 41, 29] to evaluate the restoration performance of our proposed MSPFN as well as six state-of-the-art deraining methods. These representative methods include DerainNet [6], RESCAN [19], DIDMD-N [40], UMRL [37], SEMI [31] and PreNet [26]. There is no unified training datasets for all competing methods in this paper, e.g. PreNet refers to JORDER [35] and uses 1254 pairs for training. UMRL refers to [40] and uses 12700 images for training. Therefore, directly taking the results from their papers is unfair and meaningless. To this end, we collect about 13700 clean/rain image pairs from [41, 7] for training our network as well as other competing methods for a fair comparison. In particular, these competing methods are retrained in the experiments with their publicly released codes and follow their original settings under the unified training dataset. Separately, the

Table 1. Dataset description. A total of 13712 clean/rain image pairs are used for training. There are additional 4300 labeled reference samples as well as 200 real-world scenarios for testing.

Datasets	Training Samples	Testing Samples	Name
Rain14000 [7]	11200	2800	**Test2800**
Rain1800 [35]	1800	0	**Rain1800**
Rain800 [41]	700	100	**Test100**
Rain100H [35]	0	100	**Rain100H**
Rain100L [35]	0	100	**Rain100L**
Rain1200 [40]	0	1200	**Test1200**
Rain12 [20]	12	0	**Rain12**
Real200 [29, 31]	0	200	**Real200**
RID/RIS [18]	0	2495/2348	**RID/RIS**
Total Count	13712	9343	-

Table 2. Evaluation of the basic components in our baseline MSPFN on **Test100** dataset. We obtain the average inference time of deraining on images with size of **512**$\times$ **384**.

Models	Model1	Model2	Model3	Model4	Model5	Model6	MSPFN
PSNR	26.56	27.01	23.69	26.75	26.48	26.88	**27.29**
SSIM	0.861	0.864	0.831	0.863	0.862	0.865	**0.869**
FSIM	0.921	0.923	0.905	0.923	0.921	0.923	**0.925**
Ave. inf. time (s)	0.192	0.224	**0.113**	0.238	0.141	0.180	0.308
Par. (Millions)	5.53	11.30	**2.29**	11.75	5.60	8.45	13.22

detailed descriptions of the used datasets are tabulated in Table 1. In order to quantitatively evaluate the restoration quality, we adopt the commonly used evaluation metrics, such as Peak Signal to Noise Ratio (PSNR), Feature Similarity (FSIM) [42], and Structural Similarity (SSIM) [30].

4.1. Implementation Details

In our baseline, the pyramid levels are set to 3, *i.e*. the original scale, 1/2 scale and 1/4 scale. In CFM, the filter numbers of each recurrent Conv-LSTM are respectively set to 32, 64, and 128, corresponding to the gradually increasing resolution. The depths/numbers of FFM (**M**) and CAU (**N**) are set to 10 and 3, respectively. We use Adam optimizer with batch size of 8 for training on one NVIDIA Titan Xp GPU. The learning rate is initialized to 2×10^{-4} and reduced by half at every 20000 steps till 1×10^{-6}. We train the network for 30 epochs with the above settings.

4.2. Ablation Studies

Validation on Basic Components. Using our baseline model ($\mathbf{M} = 10$, $\mathbf{N} = 3$), we design six comparison models to analyze the effects of the proposed basic modules (CFM and FFM), multi-scale pyramid framework, and multi-scale progressive fusion scheme on deraining performance. Quantitative results on **Test100** dataset are listed in Table 2. From the results, our baseline MSPFN exhibits great superiority over its incomplete versions, including Model1 (single-scale framework with only the original input), Model2 (removing CFM from MSPFN), and Model3 (removing all FFMs from MSPFN), surpassing them by 0.73dB, 0.28dB, and 3.60dB (PSNR), respectively. Moreover, we construct Model4 by applying the fusion strategy in [8] to verify the effectiveness of the proposed multi-scale progressive fusion scheme. It is evident that MSPFN gains a significant improvement over Model4 by 0.54dB with an acceptable complexity increase. Model5 ($\mathbf{M} = 5$, $\mathbf{N} = 1$) and Model6 ($\mathbf{M} = 6$, $\mathbf{N} = 3$) are the simplified variants of MSPFN with smaller depths. When compared with the single-scale framework (Model1), Model5 has the approximately equal amount of parameters but achieves faster inference speed with the multi-scale pyramid framework. Model6 has the similar computation complexity but more parameters as compared with Model1. The results show that Model5 achieves the comparable performance while it's a quarter more efficient. Model6 gains the better scores over Model1 by 0.32dB while keeping the similar computation complexity. We attribute these advantages to the effective cooperative representation of rain streaks among different pyramid layers and stages of the network.

Parameter Analysis on M and N. We assess the influence of the depth of FFM (**M**) and the number of CAU (**N**) on deraining performance. Based on our baseline ($\mathbf{M} = 10$, $\mathbf{N} = 3$), we construct three comparison models, *i.e*. MSPFN$_{M17N1}$, MSPFN$_{M13N2}$ and MSPFN$_{M8N5}$, while keeping approximately the same number of parameters. As shown in Table 3, the performance declines with the reduction of **M**. This indicates the important role of FFM for exploiting the multi-scale rain information in a progressive fashion. When increasing the number of CAU (MSPFN$_{M17N2}$), it yields a slight improvement (0.13dB), but with additional 30% of the parameters. We also add two models MSPFN$_{M30N1}$ and MSPFN$_{M5N1}$ for comparison. The former is designed to pursue a better deraining performance with more FFMs to enhance multi-scale fusion, while the latter is a lightweight model with smaller depth ($\mathbf{M} = 5$, $\mathbf{N} = 1$) and width (all filter channels $=$ 32). Meanwhile, the strided convolution and deconvolution are employed twice in our proposed U-shaped residual attention block (URAB) of MSPFN$_{M5N1}$ to further alleviate the computation burden. As we expected, MSPFN$_{M30N1}$ achieves the best scores for all the metrics. MSPFN$_{M5N1}$ still obtains the acceptable performance, although being a much lighter network. *Considering the tradeoff between efficiency and deraining performance, we set **M** and **N** to 17 and 1 respectively in the following experiments.*

4.3. Comparisons with State-of-the-arts

4.3.1 Synthesized Data

We compare our MSPFN ($M = 17, N = 1$) with other six top-performing deraining methods [6, 19, 40, 37, 31, 26] on five synthetic datasets. Quantitative results are shown in Table 4. One can see that MSPFN achieves remarkable improvements over these state-of-the-art methods. For example, MSPFN surpasses DerainNet [6] and DIDMDN [40]

Table 3. Evaluation of the depth of FFM (**M**), the number of CAU (**N**), as well as the model parameters on **Test100** dataset. MSPFN$_{MaNb}$ denotes the model with **M** $= a$ and **N** $= b$.

Models	MSPFN$_{M30N1}$	MSPFN$_{M17N1}$	MSPFN$_{M17N2}$	MSPFN$_{M13N2}$	MSPFN$_{M10N3}$	MSPFN$_{M8N5}$	MSPFN$_{M5N1}$
PSNR	**27.91**	27.50	27.63	27.42	27.29	27.13	24.99
SSIM	**0.879**	0.876	0.877	0.874	0.869	0.867	0.850
SSIM	**0.929**	0.928	0.928	0.927	0.925	0.924	0.916
Par. (Millions)	21.81	13.35	17.20	13.63	13.22	14.56	**1.65**

Figure 5. Restoration results on synthetic datasets, including **Rain100H**, **Rain100L**, **Test100**, and **Test1200**.

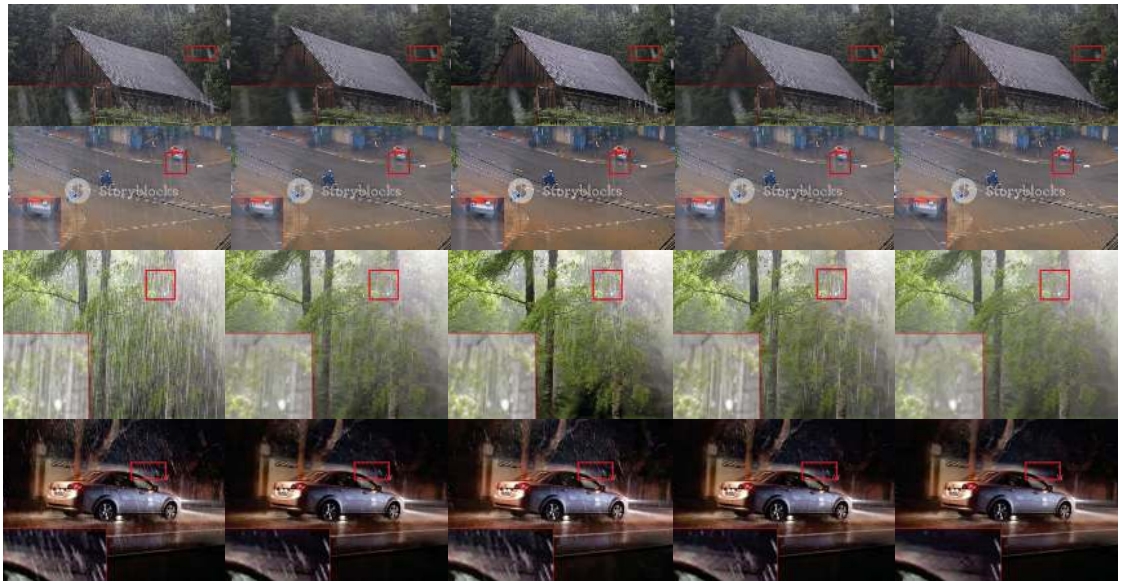
by 9.01dB and 2.74dB, respectively, in terms of PSNR on Test1200 dataset. Visual results on different rain conditions (diverse rain streak orientations and magnitudes) are presented in Fig. 5. MSPFN exhibits impressive restoration performance on all scenarios, generating results with rich and credible image textures while removing main rain streaks. For other comparison methods, they tend to blur the image contents, or still leave some visible rain streaks. For example, only our MSPFN restores the clear and credible image details in the "Giraffe" image, while the competing methods fail to remove rain streaks and their results have obvious color distortion.

4.3.2 Real-world Data

We conduct additional comparisons on three real-world datasets, including Real200 [40], Rain in Driving (RID) and Rain in Surveillance (RIS) datasets [18], to further verify the generalization capability of MSPFN. RID and RIS cover 2495 and 2348 samples, collected from car-mounted cameras and networked traffic surveillance cameras in rainy days respectively. Moreover, we use another two quantitative indicators, Naturalness Image Quality Evaluator (NIQE) [23] and Spatial-Spectral Entropy-based Quality (SSEQ) [22], to quantitatively evaluate the reference-free restoration performance. The smaller scores of SSEQ and NIQE indicate better perceptual quality and clearer contents. The results are listed in Table 5. As expected, our proposed MSPFN has the best average scores on 200 real-world samples, outperforming the state-of-the-art deraining methods [19, 37, 26] by a large margin. Moreover, we show four representative deraining examples in Fig. 6 for visual comparison. In the last image, obvious rain streaks are observed in the results of other deraining methods, but our

Table 4. Comparison results of average PSNR, SSIM and FSIM on several widely used rain datasets, including **Rain100H**, **Rain100L**, **Test100**, **Test2800**, and **Test1200**. **MSPFN$_{w/o\ Eloss}$ denotes our model without the edge constraint in the loss function**.

Methods	**Test100** PSNR/SSIM/FSIM	**Rain100H** PSNR/SSIM/FSIM	**Rain100L** PSNR/SSIM/FSIM	**Test2800** PSNR/SSIM/FSIM	**Test1200** PSNR/SSIM/FSIM	Average PSNR/SSIM/FSIM
DerainNet [6]	22.77/0.810/0.884	14.92/0.592/0.755	27.03/0.884/0.904	24.31/0.861/0.930	23.38/0.835/0.924	22.48/0.796/0.879
RESCAN [19]	25.00/0.835/0.909	26.36/0.786/0.864	29.80/0.881/0.919	31.29/0.904/0.952	30.51/0.882/0.944	28.59/0.857/0.917
DIDMDN [40]	22.56/0.818/0.899	17.35/0.524/0.726	25.23/0.741/0.861	28.13/0.867/0.943	29.65/0.901/0.950	24.58/0.770/0.876
UMRL [37]	24.41/0.829/0.910	26.01/0.832/0.876	29.18/0.923/0.940	29.97/0.905/0.955	30.55/0.910/0.955	28.02/0.880/0.927
SEMI [31]	22.35/0.788/0.887	16.56/0.486/0.692	25.03/0.842/0.893	24.43/0.782/0.897	26.05/0.822/0.917	22.88/0.744/0.857
PreNet [26]	24.81/0.851/0.916	26.77/0.858/0.890	**32.44/0.950/0.956**	31.75/0.916/0.956	31.36/0.911/0.955	29.42/0.897/0.934
MSPFN$_{w/o\ Eloss}$ (Ours)	26.93/0.865/0.924	28.33/0.842/0.883	32.18/0.928/0.939	32.70/0.928/0.964	32.22/0.914/0.958	30.51/0.895/0.934
MSPFN (Ours)	**27.50/0.876/0.928**	**28.66/0.860/0.890**	32.40/0.933/0.943	**32.82/0.930/0.966**	**32.39/0.916/0.960**	**30.75/0.903/0.937**



Rain Image RESCAN UMRL PreNet MSPFN (Ours)

Figure 6. Comparison results on four real-world scenarios with RESCAN [19], UMRL [37] and PreNet [26].

MSPFN can well preserve more realistic and credible image details while effectively removing main rain streaks.

Table 5. Comparison results of average NIQE/SSEQ on real-world datasets (**Real200, RID, and RIS**). The smaller scores indicate better perceptual quality.

Methods	RESCAN [19]	UMRL [37]	PreNet [26]	MSPFN (Ours)
Real200	4.724/30.47	4.675/29.38	4.620/29.51	**4.459/29.26**
RID	6.641/40.62	6.757/41.04	7.007/43.04	**6.518/40.47**
RIS	6.485/50.89	**5.615/43.45**	6.722/48.22	6.135/43.47

4.3.3 Other Applications

Image deraining under complex weather conditions can be considered as an effective enhancement of image content. It can potentially be incorporated into other high-level vision systems for applications such as object detection and segmentation. This motivates us to investigate the effect of restoration performance on the accuracy of object detection and segmentation based on some popular algorithms, *e.g.* YOLOv3 [25], Mask R-CNN [12], and RefineNet [21]. To this end, we randomly select a total of 850 samples from COCO [3] and BDD [38] datasets to create three new synthetic rain datasets COCO350 (for detection), BDD350 (for detection), and BDD150 (for segmentation) through Photoshop. These rain images are of diverse streak orientations and magnitudes, and at the same time have complex

Table 6. Comparison results of joint image deraining, object detection, and semantic segmentation on COCO350, BDD350, and BDD150 datasets. MSPFN* denotes the lightweight model with lighter depth and width comparing to MSPFN.

Methods	Rain input	RESCAN [19]	PreNet [26]	MSPFN* (Ours)	MSPFN (Ours)
Deraining; Dataset: **COCO350/BDD350**; Image Size: **640**$\times$ **480/1280**$\times$ **720**					
PSNR	14.79/14.13	17.04/16.71	17.53/16.90	17.74/17.38	**18.23/17.85**
SSIM	0.648/0.470	0.745/0.646	0.765/0.652	0.773/0.678	**0.782/0.761**
Ave.inf.time (s)	–/–	0.55/1.53	0.22/0.76	**0.08/0.23**	0.58/1.24
Object Detection; Algorithm: **YOLOv3** [25]; Dataset: **COCO350/BDD350**; Threshold: 0.6					
Precision (%)	23.03/36.86	28.74/40.33	31.31/38.66	30.99/39.91	**32.56/41.04**
Recall (%)	29.60/42.80	35.61/47.79	37.92/48.59	37.99/49.74	**39.31/50.40**
IoU (%)	55.50/59.85	59.81/61.98	60.75/61.08	61.06/61.90	**61.69/62.42**
Deraining; Dataset: **BDD150**; Image Size: **1280**$\times$ **720**					
PSNR	18.00	20.96	21.52	21.73	**22.48**
SSIM	0.722	0.859	0.886	0.887	**0.904**
Ave.inf.time (s)	–	1.53	0.76	**0.23**	1.24
Semantic Segmentation; Algorithm: **RefineNet** [21]; Dataset: **BDD150**					
mPA (%)	33.29	45.34	50.28	50.25	**52.96**
mIoU (%)	20.49	31.52	33.42	33.74	**35.90**

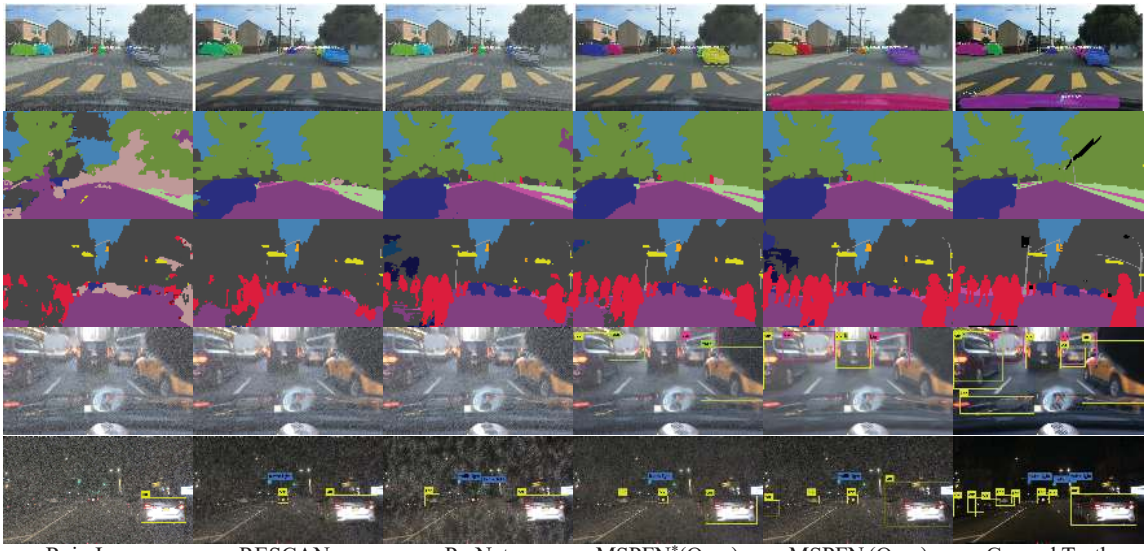


Figure 7. Examples of joint deraining, object detection and segmentation. The first row denotes the instance segmentation results of Mask R-CNN [12] on BDD150 dataset. The second and third rows are the comparison results of semantic segmentation by RefineNet [21] on BDD150 dataset. We use YOLOv3 [25] for object detection on COCO350 dataset and the results are shown in the last two rows. MSPFN* denotes the lightweight model with lighter depth and width comparing to MSPFN.

imaging conditions such as night scenes. By using our proposed deraining algorithm MSPFN as well as other top-performing deraining methods [19, 26], the restoration procedures are directly implemented on these three datasets to produce the rain-free images. And then we apply the public available pre-trained models of YOLOv3 (for detection), Mask R-CNN (for instance segmentation), and RefineNet (for semantic segmentation) to perform the the downstream tasks. Qualitative results, including the deraining performance as well as the precision of the subsequent detection and segmentation tasks, are tabulated in Table 6. In addition, visual comparisons are shown in Fig. 7.

It is obvious that rain streaks can greatly degrade the detection accuracy and segmentation precision, night scenarios in particular, *i.e.* by missing targets and producing low detection or segmentation confidence (mean pixel accuracy (mPA) and mean Intersection of Union (mIoU)). In addition, the detection precision of the produced rain-free images by MSPFN shows a notable improvement over that of original rain inputs by nearly 10%, and MSPFN achieves the best results of 52.96% mPA as well as 35.90% mIoU for semantic segmentation task on BDD150. When compared with other top-performing deraining models, the rain-free images generated by MSPFN show more credible contents with more details, which effectively promote the detection and segmentation performance. Moreover, we also evaluate our lightweight deraining model MSPFN* with lighter depth ($\mathbf{M} = 5$, $\mathbf{N} = 1$) and width (with all filter channels of 32) since computation efficiency is crucial for mobile devices and applications require real-time throughput such as autonomous driving. MSPFN* still achieves competitive performance compared with other models [19, 26] while it's a half more efficient in terms of inference time.

5. Conclusion

In this paper, we propose a novel multi-scale progressive fusion network (MSPFN) to exploit the multi-scale rain information to cooperatively represent rain streaks based on the pyramid framework. To achieve this goal, we design several basic modules (CFM, FFM and RM) along with our proposed multi-scale progressive fusion mechanism to explore the inherent correlations of the similar rain patterns among multi-scale rain streaks. Consequently, our predicted rain streak distribution is potentially more correct due to the collaborative representation of rain streaks across different scales. Experimental results on several synthetic deraining datasets and real-world scenarios, as well as several downstream vision tasks (*i.e.* object detection and segmentation) have shown great superiority of our proposed MSPFN algorithm over other top-performing methods.

6. Acknowledgement

This work is supported by National Key R&D Project (2016YFE0202300) and National Natural Science Foundation of China (U1903214, 61671332, U1736206, 41771452, 41771454, 61971165), and Hubei Province Technological Innovation Major Project (2019AAA049, 2018CFA024).

References

[1] Peter C Barnum, Srinivasa Narasimhan, and Takeo Kanade. Analysis of rain and snow in frequency space. *International journal of computer vision*, 86(2-3):256, 2010.

[2] Jérémie Bossu, Nicolas Hautière, and Jean-Philippe Tarel. Rain or snow detection in image sequences through use of a histogram of orientation of streaks. *International journal of computer vision*, 93(3):348–367, 2011.

[3] Holger Caesar, Jasper Uijlings, and Vittorio Ferrari. Cocostuff: Thing and stuff classes in context. In *IEEE Conference on CVPR*, pages 1209–1218, 2018.

[4] Jie Chen, Cheen-Hau Tan, Junhui Hou, Lap-Pui Chau, and He Li. Robust video content alignment and compensation for rain removal in a cnn framework. In *IEEE Conference on CVPR*, pages 6286–6295, 2018.

[5] Yi-Lei Chen and Chiou-Ting Hsu. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In *IEEE International Conference on ICCV*, pages 1968–1975, 2013.

[6] Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Trans. Image Process.*, 26(6):2944–2956, 2017.

[7] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *IEEE Conference on CVPR*, pages 3855–3863, 2017.

[8] Xueyang Fu, Borong Liang, Yue Huang, Xinghao Ding, and John Paisley. Lightweight pyramid networks for image deraining. *IEEE Transactions on Neural Networks and Learning Systems*, 2019.

[9] Kshitiz Garg and Shree K Nayar. When does a camera see rain? In *IEEE International Conference on ICCV*, volume 2, pages 1067–1074, 2005.

[10] Daniel Glasner, Shai Bagon, and Michal Irani. Super-resolution from a single image. In *IEEE International Conference on ICCV*, pages 349–356, 2009.

[11] Rana Hanocka, Amir Hertz, Noa Fish, Raja Giryes, Shachar Fleishman, and Daniel Cohen-Or. Meshcnn: a network with an edge. *ACM Transactions on Graphics (TOG)*, 38(4):90, 2019.

[12] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn. In *IEEE International Conference on ICCV*, Oct 2017.

[13] Behzad Kamgar-Parsi and Azriel Rosenfeld. Optimally isotropic laplacian operator. *IEEE Trans. Image Process.*, 8(10):1467–1472, 1999.

[14] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Trans. Image Process.*, 21(4):1742–1755, 2011.

[15] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *IEEE Conference on CVPR*, pages 624–632, 2017.

[16] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. Single image deraining using scale-aware multi-stage recurrent network. *arXiv preprint arXiv:1712.06830*, 2017.

[17] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *IEEE Conference on CVPR*, pages 1633–1642, 2019.

[18] Siyuan Li, Iago Breno Araujo, Wenqi Ren, Zhangyang Wang, Eric K Tokuda, Roberto Hirata Junior, Roberto Cesar-Junior, Jiawan Zhang, Xiaojie Guo, and Xiaochun Cao. Single image deraining: A comprehensive benchmark analysis. In *IEEE Conference on CVPR*, pages 3838–3847, 2019.

[19] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *European Conference on ECCV*, pages 254–269, 2018.

[20] Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. Rain streak removal using layer priors. In *IEEE Conference on CVPR*, pages 2736–2744, 2016.

[21] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *IEEE Conference on CVPR*, pages 1925–1934, 2017.

[22] Lixiong Liu, Bao Liu, Hua Huang, and Alan Conrad Bovik. No-reference image quality assessment based on spatial and spectral entropies. *Signal Processing: Image Communication*, 29(8):856–863, 2014.

[23] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a "completely blind " image quality analyzer. *IEEE Signal Process. Lett*, 20(3):209–212, 2013.

[24] Rui Qian, Robby T Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. In *IEEE Conference on CVPR*, pages 2482–2491, 2018.

[25] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[26] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: a better and simpler baseline. In *IEEE Conference on CVPR*, pages 3937–3946, 2019.

[27] Marvin Teichmann, Michael Weber, Marius Zoellner, Roberto Cipolla, and Raquel Urtasun. Multinet: Real-time joint semantic reasoning for autonomous driving. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 1013–1020, 2018.

[28] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao. Image super-resolution using dense skip connections. In *IEEE International Conference on ICCV*, pages 4799–4807, 2017.

[29] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson WH Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *IEEE Conference on CVPR*, pages 12270–12279, 2019.

[30] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004.

[31] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain re-

moval. In *IEEE Conference on CVPR*, pages 3877–3886, 2019.

[32] Jing Xu, Wei Zhao, Peng Liu, and Xianglong Tang. An improved guidance image based method to remove rain and snow in a single image. *Computer and Information Science*, 5(3):49, 2012.

[33] Chih-Yuan Yang, Jia-Bin Huang, and Ming-Hsuan Yang. Exploiting self-similarities for single frame super-resolution. In *Asian conference on ACCV*, pages 497–510. Springer, 2010.

[34] Wenhan Yang, Jiaying Liu, Shuai Yang, and Zongming Guo. Scale-free single image deraining via visibility-enhanced recurrent wavelet learning. *IEEE Trans. Image Process.*, 28(6):2948–2961, 2019.

[35] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *IEEE Conference on CVPR*, pages 1357–1366, 2017.

[36] Youzhao Yang and Hong Lu. Single image deraining using a recurrent multi-scale aggregation and enhancement network. In *IEEE Conference on ICME*, pages 1378–1383, 2019.

[37] Rajeev Yasarla and Vishal M Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *IEEE Conference on CVPR*, pages 8405–8414, 2019.

[38] Fisher Yu, Wenqi Xian, Yingying Chen, Fangchen Liu, Mike Liao, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving video database with scalable annotation tooling. *arXiv preprint arXiv:1805.04687*, 2018.

[39] He Zhang and Vishal M Patel. Convolutional sparse and low-rank coding-based rain streak removal. In *IEEE Winter Conference on WACV*, pages 1259–1267, 2017.

[40] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *IEEE Conference on CVPR*, pages 695–704, 2018.

[41] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *IEEE Trans. Circuits Syst. Video Technol.*, 2019.

[42] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.*, 20(8):2378–2386, 2011.

[43] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *IEEE Conference on CVPR*, pages 2472–2481, 2018.

[44] Yupei Zheng, Xin Yu, Miaomiao Liu, and Shunli Zhang. Residual multiscale based single image deraining. In *Conference on BMVC*, 2019.