



Single image deraining via deep pyramid network with spatial contextual information aggregation

Cong Wang¹ · Yutong Wu¹ · Yu Cai¹ · Guangle Yao² · Zhixun Su^{1,3,4} · Hongyan Wang¹

Published online: 20 January 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Rain streaks usually give rise to visual degradation and cause many computer vision algorithms to fail. So it is necessary to develop an effective deraining algorithm as preprocess of high-level vision tasks. In this paper, we propose a novel deep learning based deraining method. Specifically, the multi-scale kernels and feature maps are both important for single image deraining. However, the previous works ignore the two multi-scale information or only consider the multi-scale kernels information. Instead, our method learns multi-scale information both from the perspectives of kernels and feature maps, respectively, by designing spatial contextual information aggregation module and pyramid network module. The former module can capture the rain streaks with different sizes and the latter module can extract rain streaks from different scales further. Moreover, we also employ squeeze-and-excitation and skip connections to enhance the correlation between channels and transmit the information from low-level to high-level, respectively. The experimental results show that the proposed method achieves significant improvements over the recent state-of-the-art methods in Rain100H, Rain100L, Rain1200 and Rain1400 datasets.

Keywords Single image deraining · Pyramid network · Spatial contextual information aggregation · Residual learning

1 Introduction

Recently, the task of deraining has received much attention due to the importance in the preprocessing of artificial intelligence applications, e.g. object detection, object tracking, semantic segmentation, and other high-level visions. So it is meaningful to explore an effective deraining method. In

this paper, a deep learning-based deraining algorithm is proposed.

The methods of deraining can be divided roughly into two categories: video based methods and single image based methods. The video based methods [1, 2, 7, 21, 31] can analyze the difference between frames, thus they are easier than the single image deraining methods. In this paper, we explore the study on single image deraining method.

The problem of single image deraining has been studied intensively for long years. Most of the methods can be formalized as below:

$$O = B + R \quad (1)$$

where O , B and R denote rainy image, clear image and rainy streaks, respectively. As only the rainy image is available, this is a highly ill-posed problem. To make this problem well solve, numerous algorithms use prior knowledge about rainy streaks and clear images, e.g. low-rank prior [3], sparse representation [18], Gaussian mixture model [16], to constrain the solution space. Although these algorithms achieve the promising performance, the prior knowledge used in these algorithms does not hold for some cases, as shown in Fig. 1.

In the past few decades, a large number of single image deraining methods have been proposed. Since 2012, due to

This work was supported by the Natural Science Foundation of China [grant numbers 61572099]; Major National Science and Technology Project of China [grant number 2018ZX04016001-011].

✉ Zhixun Su
zxsu@dlut.edu.cn

Cong Wang
congwang@mail.dlut.edu.cn

- ¹ Dalian University of Technology, Dalian, China
- ² Chengdu University of Technology, Chengdu, China
- ³ Guilin University of Electronic Technology, Guilin, China
- ⁴ School of Mathematical Sciences, Dalian University of Technology, Dalian, China



Fig. 1 An example in real-world datasets compared with priors based deraining methods. It can be observed that the prior based methods fail to restore rain-free image. Our method generates better deraining result

the powerful feature representation capability of the Convolutional Neural Networks (CNNs), deep learning has been widely employed in the image domain, significantly improving the performance of image classification, object detection [17], pose estimation [4], semantic segmentation [20], and also the deraining methods [5, 6, 14, 15, 22–25, 29]. Although the deep learning based single image deraining has achieved great success, they still suffer from some limitations. We summarize as follows:

- On the one hand, the spatial contextual information [9] that can be learned by multi-scale kernels can be utilized to remove the rain streaks with different sizes from rainy images. But the methods [5, 6, 30] employ either a residual network or a symmetric encoder-decoder structure with adversarial learning, neglecting the spatial contextual information. Although the methods [27, 29] design the element-wise sum between multi-stream dilation convolutions or use several convolution kernels with different sizes that consider the multi-scale kernels information into deraining, they ignore the following limitation: multi-scale feature maps.
- On the other hand, multi-scale feature maps which have been proved to be effective in many computer vision tasks [4, 17, 19, 28] are neglected in some deraining methods [5, 6, 14, 15, 27, 29, 30], leading to poor deraining performance.
- Finally, the semantic correlation between channels, which can make the channels closely related, is important in deep-learning based tasks, but it is not considered into the deraining methods [5, 6, 27, 29, 30] that lose some important semantic information.

To overcome these limitations, in this paper, we propose a novel end-to-end deep-learning based deraining method, which is based on the pyramid network with spatial contextual information aggregation. The proposed network includes two types of modules: (a) spatial contextual information aggregation module and (b) pyramid network module. The spatial contextual information aggregation

module consists of multiple dilation convolutions which can obtain more spatial contextual information and the multi-scale features from the perspective of kernels. Different from the element-wise sum among multi-stream dilation convolutions [27], we employ the 1×1 convolution to fuse the features of rain streaks with different sizes extracted by the multi-scale kernels. To obtain the multi-scale features from the perspective of feature maps, we introduce a pyramid network module in the proposed method. The pyramid network module can acquire the multi-scale feature maps and fuse these features to maintain the primary information of rain streaks. Finally, we embed squeeze-and-excitation [8] into these two types of modules to recalibrate the feature response of each feature map adaptively that can capture more semantic correlation between channels.

In summary, this paper makes the following contributions:

- A spatial contextual information aggregation module is proposed to learn the rain streaks with different sizes, which can acquire the multi-scale features from the perspective of kernels.
- We consider the pyramid network module into our network to obtain and fuse the multi-scale features of rain streaks from the perspective of the feature maps.
- As we know, this is the first paper that considers the two multi-scale information, i.e. kernels and feature maps, into deraining simultaneously and their effectiveness has been verified.
- Quantitative and qualitative experimental evaluations on both synthetic and real-world datasets have demonstrated the superiority of our proposed method, which achieves significant improvements over the state-of-the-art methods.

2 Related work

In this section, we present a review of the recent related works about deraining methods and pyramid network.

2.1 Single image deraining

As aforementioned, the single image deraining methods can be grouped into two categories, prior based methods, and deep-learning based methods.

Prior based methods The research of deraining starts from prior based method. The most widely known is image decomposition [11], non-local filter [12], sparse codes [18], low-rank model [3], and Gaussian mixture model [16]. Kang et al. [11] propose a single-image-based rain removal framework via properly formulating rain removal as an image decomposition problem based on morphological component analysis, where they decompose an image into the low- and high-frequency parts using a bilateral filter. The high-frequency part is then decomposed into a rain component and a nonrain component by performing dictionary learning and sparse coding. Kim et al. [12] first detect rain streak regions by analyzing the rotation angle and the aspect ratio of the elliptical kernel at each pixel location and then perform the nonlocal means filtering on the detected rain streak regions by selecting nonlocal neighbor pixels and their weights adaptively. Luo et al. [18] propose a dictionary learning based algorithm for single image deraining, which is to sparsely approximate the patches of two layers by very high discriminative codes over a learned dictionary with strong mutual exclusivity property and the discriminative sparse codes lead to accurate separation of two layers from their non-linear composite. Chen et al. [3] think first rain streaks usually reveal similar and repeated patterns on imaging scene, then design a low-rank model from a matrix to tensor structure to capture the correlated rain streaks and utilize the model to remove rain streaks from image in a unified way. Li et al. [16] use simple patch-based priors for both the background and rain layers and these priors are based on Gaussian mixture models and can accommodate multiple orientations and scales of the rain streaks. Prior based methods dominated the deraining task before 2017, however, the execution speed and deraining performance are unsatisfactory.

Deep-learning based methods From 2017, several deep learning based deraining methods achieve great success on speed and deraining results. They mostly either learn the residual about rain streaks or first detect rain streak then remove them or use the density of rain streaks to guide the learning of network. And these methods led to our subsequent research. Fu et al. [5, 6] firstly introduce deep-learning methods to single image deraining. They decompose rainy images O into low- and high-frequency parts and map high-frequency parts to rain streaks images R by a deep residual network, lastly utilize (1) to obtain a

clean image B . Yang et al. [27] develop a multi-task deep learning architecture that learns the binary rain streak map, the appearance of rain streaks, and the clean background and based on the designment, they propose a recurrent rain detection and removal network that removes rain streaks and clears up the rain accumulation iteratively and progressively. Li et al. [14] propose a non-locally enhanced encoder-decoder network framework, which utilizes non-locally enhanced dense blocks that are designed to not only fully exploit hierarchical features from all the convolutional layers but also well capture the long-distance dependencies and structural information. Li et al. [15] propose a novel deep network architecture based on deep convolutional and recurrent neural networks for single image deraining, where they utilize squeeze-and-excitation [8] to allot each learnable value to rain streaks with different sizes. Zhang et al. [29] present a novel density-aware multi-stream densely connected convolutional neural network-based algorithm for joint rain density estimation and deraining that they enable the network itself to automatically determine the rain-density information and then efficiently remove the corresponding rain-streaks guided by the estimated rain density label.

2.2 Pyramid network

Recently, the conventional spatial pyramid approaches have been combined successfully with neural network architectures to deal with various vision tasks. There are several networks based on the spatial pyramid. Ranjan et al. [19] develop a spatial pyramid network to estimate optical flow, which estimates large motions in a coarse-to-fine approach by warping one image of a pair at each pyramid level by the current flow estimate and computing an update to the flow. Lin et al. [17] propose a feature pyramid network for object detection, where they exploit the inherent multi-scale, pyramidal hierarchy of deep convolutional networks to construct feature pyramids with marginal extra cost. A top-down architecture with lateral connections is developed for building high-level semantic feature maps at all scales. Chen et al. [4] come up with a cascaded pyramid network for multi-person pose estimation which can successfully localize the simple keypoints like eyes and hands but may fail to precisely recognize the occluded or invisible keypoints. Zhang et al. [28] present a densely connected pyramid network for single image dehazing.

A great deal of work has been done to verify that the multi-scale information is effective for various vision tasks and it is necessary to introduce the multi-scale information to deraining problem. So in this paper, we utilize a pyramid network to obtain the multi-scale information from the perspective of feature maps.

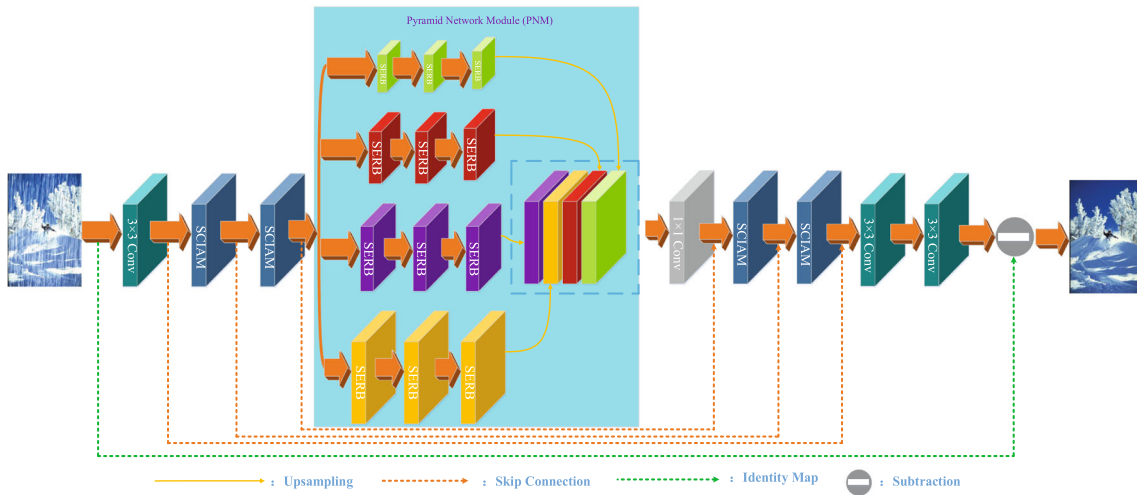


Fig. 2 Overall Network Framework. The SCIAM and SERB are shown in Figs. 3 and 4, respectively

3 Proposed method

The proposed method shown in Fig. 2 includes two types of modules: (a) spatial contextual information aggregation module (SCIAM) and (b) pyramid network module (PNM). The SCIAM shown in Fig. 3 consists of several dilation convolutions and one squeeze-and-excitation operation. The basic component of PNM, as shown in Fig. 4, is made up of the improved residual block that is a squeeze-and-excitation enhancing residual block. Skip connections are utilized to enable the computation of long-range spatial dependencies as well as efficient usage of the feature activation of proceeding layers.

3.1 Overall framework network

Compared with the complex background image information, rain streaks are easier to learn. So our network is

to learn the nonlinear mapping from rainy images to rain streaks, then obtains the final clear background image via (1). Spatial contextual information aggregation module can acquire the spatial information and the correlation between channels, where multi-scale information from the perspective of kernels is obtained using several dilation convolutions with different factors. And the pyramid network module can gain multi-scale information from the perspective of feature maps. The proposed network can be roughly divided into three stages: Shallow layer processing stage, Pyramid network module processing stage and Reconstructing rain streaks stage. The shallow layer processing stage and reconstructing rain streaks stage are based on the spatial contextual information aggregation module. And the pyramid network module processing stage is based on the pyramid network.

Mathematically, we describe the three stages of the network as follows:

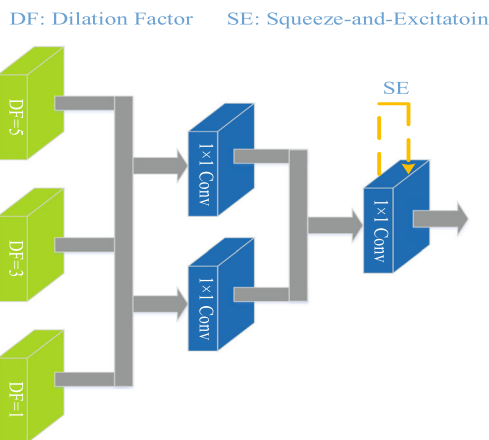


Fig. 3 The proposed spatial contextual information aggregation module (SCIAM)

Shallow layer processing stage Firstly, the rainy image is input into the shallow layer to convert image space to shallow feature space:

$$F_0 = Conv_{3 \times 3}(O) \tag{2}$$

where F_0 is the shallow feature layer. $Conv_{3 \times 3}$ is a 3×3 convolution operation.

Then two SCIAMs following the shallow feature space are designed to capture more spatial contextual information of the rain streaks with different sizes by several dilation convolutions with different factors:

$$F_i = SCIAM_i(F_{i-1}), i = 1, 2. \tag{3}$$

where $SCIAM_i$ denotes the i -th operation of SCIAM that will be introduced in Section 3.2 detailedly.

Pyramid network module (PNM) processing stage we introduce the multi-scale feature maps into our single image deraining method. Firstly, pooling operation with different sizes of kernel and stride is adopted to acquire multi-scale feature maps information:

$$P_i = Pooling_i(F_2), i = 1, 2, \dots, 2^{j-1}. \tag{4}$$

where $Pooling_i$ denotes the pooling operation with the kernel $i \times i$ and stride $i \times i$. P_i denotes the obtained original multi-scale features. j is the level of the pyramid network.

Then the original multi-scale features are fed to pyramid network module to extract the deeper rain streaks further:

$$H_i = PNM_i(P_i), i = 1, 2, \dots, 2^{j-1}. \tag{5}$$

where H_i denotes the output of pyramid network module of i -th level. PNM_i denotes the pyramid network module in i -th level, where there are three basic blocks in every level.

Lastly, we fuse all the outputs of the pyramid network module by 1×1 convolution to learn further the primary rain streaks:

$$F_3 = Conv_{1 \times 1}(Cat[H_1, H_2, \dots, H_{2^{j-1}}]), \tag{6}$$

where Cat and $Conv_{1 \times 1}$ denote concatenation at the dimension of channel and 1×1 operation, respectively. In this paper, we set $j = 4$ empirically.

Reconstructing rain streaks stage To reconstruct rain streaks, we employ two SCIAMs to process them, while to enable the computation of long-range spatial dependencies as well as efficient usage of the feature activation of proceeding layers, skip connections are utilized:

$$F_i = \begin{cases} SCIAM(Conv_{1 \times 1}(Cat[F_{6-i}, F_{i-1}])), & i = 4, 5. \\ Conv_{3 \times 3}(Conv_{1 \times 1}(Cat[F_{6-i}, F_{i-1}])) & i = 6. \end{cases} \tag{7}$$

The reconstructed rain streaks \hat{R} can be obtained:

$$\hat{R} = Conv_{3 \times 3}(F_6), \tag{8}$$

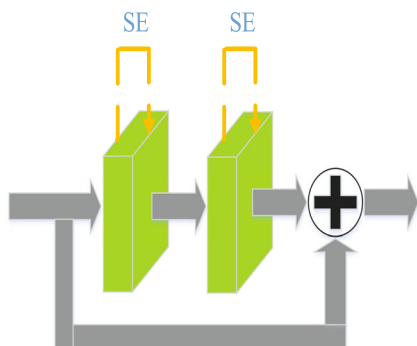


Fig. 4 The basic unit of PNM: Squeeze-and-Excitation Residual Block (SERB)

The final estimated rain-free images \hat{B} are gained via (1):

$$\hat{B} = O - \hat{R}, \tag{9}$$

3.2 Spatial contextual information aggregation module (SCIAM)

Spatial contextual information and correlation between channels are important for single image deraining [9]. Considering the above two key points, we design a spatial contextual information aggregation module, as shown in Fig. 3. It utilizes adequately the spatial contextual information and correlation between channels by using several dilation convolutions and squeeze-and-excitation operation, respectively, to learn better rain streaks representation information.

First, we utilize several dilation convolutions with different dilation factors to produce spatial contextual information:

$$y_r = Conv_r(x), r = 1, 3, 5. \tag{10}$$

where $Conv_r$ denotes dilation convolution with factor r . x and y_r denote the input signal and the corresponding output, respectively.

Then all the spatial contextual information are fused in pairs by 1×1 convolution to further learn better rain streaks structures:

$$z_{s,t} = Conv_{1 \times 1}(Cat[y_s, y_t]), s, t = 1, 3, 5. \tag{11}$$

Finally, we utilize squeeze-and-excitation (SE) to adaptively recalibrate the feature response of each feature map to acquire the correlation between channels:

$$z = SE(Conv_{1 \times 1}(Cat[z_{1,3}, z_{3,5}])), \tag{12}$$

where z is the final output of SCIAM.

3.3 Pyramid network module (PNM)

Although the SCIAM can obtain the multi-scale information from the perspective of kernels, the multi-scale features from the perspective of feature maps are ignored, which are also very important for single image deraining. To address this issue, multi pooling operations are utilized, then the information is processed by several of our proposed residual blocks with semantic correlation. The improved residual block is a normal residual block with squeeze-and-excitation enhancing, as shown in Fig. 4. We denote the improved residual block operation as $SERB$ blew. Moreover, the multi-scale feature maps are obtained via pyramid network by using several pooling operations and the overall processing procedure is expressed by (4, 5 and 6).

We describe these processes in mathematical detail as follows:

The features at different levels are forward to several *SERBs*, respectively:

$$h_i^l = SERB_l(h_i^{l-1}), i = 1, 2, \dots, 2^{j-1}. l = 1, 2, 3. \quad (13)$$

where h_i^l denotes the output of the l -th *SERB* at j -th pyramid level. The h_i^0 and h_i^3 are the P_i and H_i in the (5), respectively.

3.4 Loss function

We use *MSE* as the loss function:

$$\mathcal{L} = \frac{1}{HWC} \sum_{t=1}^H \sum_{s=1}^W \sum_{k=1}^C \|\widehat{\mathbf{B}}_{t,s,k} - \mathbf{B}_{t,s,k}\|_2^2. \quad (14)$$

where H, W, C denotes the height, width and channel number of a rain-free image, respectively. $\widehat{\mathbf{B}}$ and \mathbf{B} denote the estimated clean image and background image, respectively. Actually, this loss is equivalent to $\frac{1}{HWC} \sum_{t=1}^H \sum_{s=1}^W \sum_{k=1}^C \|\widehat{\mathbf{R}}_{t,s,k} - \mathbf{R}_{t,s,k}\|_2^2$ according to the rainy image decomposition (1). $\widehat{\mathbf{R}}$ and \mathbf{R} denote estimated rain streaks and corresponding groundtruth, respectively.

4 Experiments

In this section, we demonstrate the effectiveness of the proposed method by conducting extensive experiments on four synthetic datasets and a real-world dataset. We compare our results with six state-of-the-art methods, including two prior based methods, DSC [18] (ICCV15) and LP [16] (CVPR16), four deep-learning based methods, DDN [6] (CVPR17), JORDER [27] (CVPR17), RESCAN [15] (ECCV18) and DID [29] (CVPR18).

4.1 Datasets and measurements

Synthetic datasets We perform deraining experiments on four widely used synthetic datasets: Rain100L [27], Rain100H [27], Rain1200 [29] and Rain1400 [6]. They have various rain streaks, including different sizes, shapes and directions. Rain100H and Rain100L have 1800 images for training and 200 images for testing, respectively.

Rain1200 has 12000 images for training and 1200 images for testing. Rain1400 has 12600 images for training and 1400 images for testing. All testing datasets are assured to have different background images with training datasets. We select Rain100H as our analysis dataset.

Measurements Peak signal to noise ratio (PSNR) [10] and structure similarity index (SSIM) [26] are widely used to evaluate the quality of restored results with ground truth. PSNR is based on the error between corresponding pixels, i.e. estimated deraining result and ground truth. The higher its value, the better the restored image will be. SSIM is a measure of similarity between two images. The closer to 1, the more similar the two images are. We use them as our measurement criteria on synthetic datasets. As it is difficult to acquire the ground truth for real-world images, we only evaluate the performance on the real-world images visually.

4.2 Implementing details

We randomly crop 128×128 patch pairs from training image datasets as inputs with a mini-batch size of 12 to train our network. The ADAM [13] is used as the optimization algorithm with an initialized learning rate of 0.001, and the rate will be divided by 10 at 240K and 320K iterations, and model terminates training after 400K iterations. The number of channels is 48 and the non-linear activation is LeakyReLU with $\alpha = 0.2$ for all convolution layers. We use PyTorch to perform all experiments on an NVIDIA GTX 1080Ti GPU. As our entire model is fully convolutional, the testing process only takes 0.05 seconds when handling a test image with 512×512 pixels on a PC with a GTX 1080Ti GPU.

4.3 Results on synthetic datasets

The comparison of our method with six state-of-the-art methods on four datasets is illustrated in Table 1. It can be observed that our method is far superior to others. Especially, the proposed method achieves significant improvements on Rain100H datasets.

Table 1 Quantitative experiments evaluated on four synthetic datasets

Dataset	DSC [18]		LP [16]		DDN [6]		JORDER [27]		RESCAN [15]		DID [29]		Ours	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Rain100H	15.66	0.42	14.26	0.54	22.26	0.69	23.45	0.74	25.92	0.84	26.12	0.83	28.24	0.89
Rain100L	24.16	0.87	29.11	0.88	34.85	0.95	36.11	0.97	36.12	0.97	36.14	0.96	38.20	0.98
Rain1200	21.44	0.79	22.46	0.80	30.95	0.86	29.75	0.87	32.35	0.89	29.65	0.90	33.16	0.92
Rain1400	22.03	0.80	26.53	0.83	29.99	0.89	28.90	0.90	29.84	0.90	31.18	0.91	31.45	0.92

The best results are boldfaced

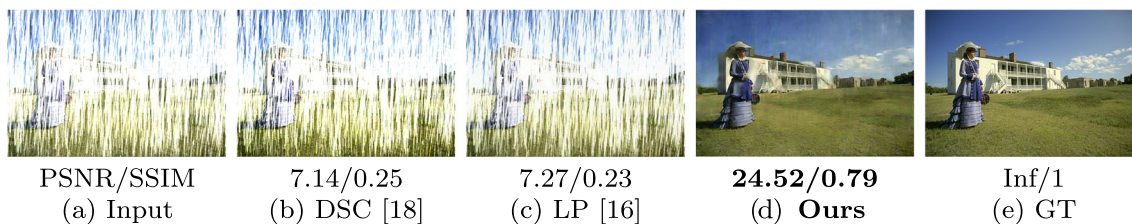


Fig. 5 An example in synthetic datasets compared with priors based deraining methods. The results of DSC [18] and LP [16] are unacceptable. The proposed method generates a much clear image

Furthermore, we give several visual examples. The performance of the prior based methods on synthetic example is shown in Fig. 5. It is obvious that the results of prior based methods are unacceptable, while our restored result is clean and clear. We also show some results on deep learning based methods in Fig. 6. We can see that other methods remain some rain streaks and have artifacts, while our restored results obtain the best performance.

4.4 Results on real-world datasets

To verify the robustness in real-world images, we also give some examples on real-world datasets. The comparison of our method with the prior based methods, DSC [18] and LP [16], is showed in Fig. 7. The performances on prior based methods are unacceptable and there are a mass of remaining rain streaks. Our result shown in Fig. 7d is better

and removes a good deal of rain streaks. We also compare our method with other deep learning based methods in Fig. 8. For the first example, our result removes all the rain streaks and the results of JORDER [27], DDN [6] and DID [29] remain some rain streaks. Furthermore, the results of JORDER [27], DDN [6] and RESCAN [15] lose the details in the masked boxes, while our result preserves better details. For the other examples, we obtain better performance, while the others have residual rain streaks. We provide more deraining results in real-world datasets in Fig. 9.

4.5 Ablation study

As our proposed method is based on several basic components, including dilation convolution, squeeze-and-excitation, pyramid network and skip connections, where

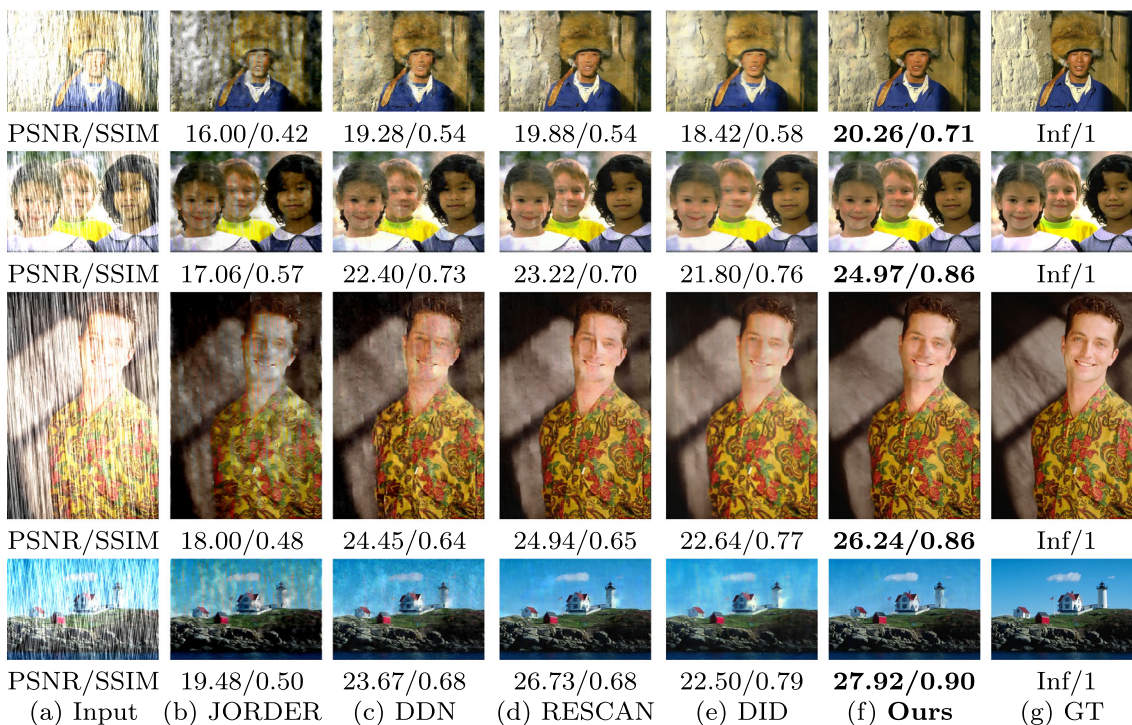


Fig. 6 Several examples in synthetic datasets compared with deep-learning based deraining methods. The other methods remain some rain streaks or have macroscopic artifacts. Our method obtains the best performance



Fig. 7 An example in real-world datasets compared with priors based deraining methods. The results of prior based methods are unacceptable, while the proposed method generates a much clear image

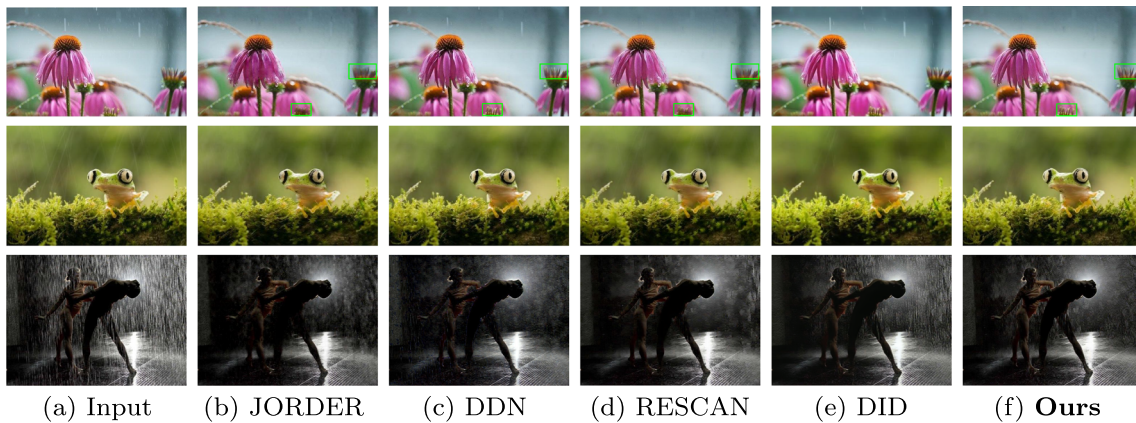


Fig. 8 There examples in real-world datasets compared with deep-learning based deraining methods. For the first example, our result remove all the rain streaks and the results of JORDER [27], DDN [6] and DID [29] remain some rain streaks. Furthermore, the results of JORDER [27], DDN [6] and RESCAN [15] lose the details in the masked boxes, while our result preserves the better details

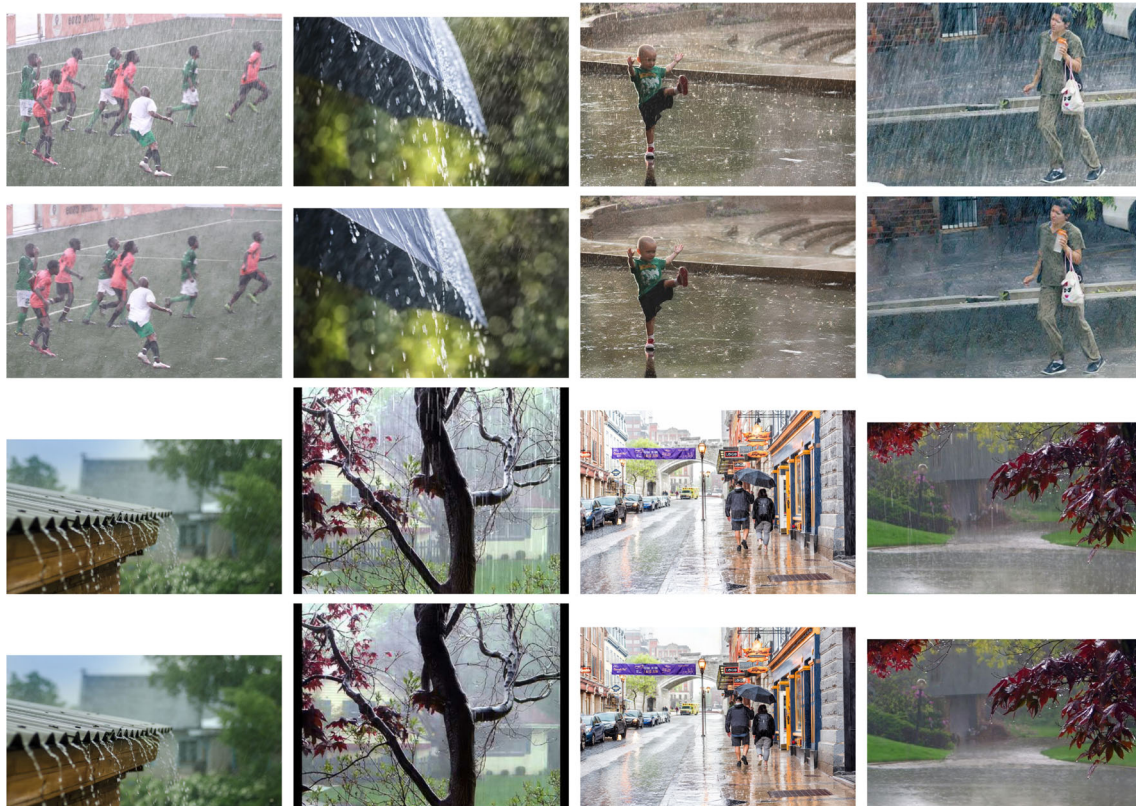


Fig. 9 More our deraining results

Table 2 Ablation study on different components of our proposed method

Experiments	R_1	R_2	R_3	R_4	R_5	R_6
DC			✓	✓	✓	✓
SE		✓		✓	✓	✓
Pyramid	✓	✓	✓		✓	✓
Skip	✓	✓	✓	✓		✓
PSNR	28.01	28.20	27.79	27.37	28.02	28.24
SSIM	0.88	0.88	0.87	0.86	0.88	0.89

The ✓ symbol denotes that the corresponding component is adopted
The best results are boldfaced

dilation convolutions denote the multi-scale information from the perspective of kernels and pyramid network represents the multi-scale information from the perspective of feature maps, it is meaningful to discuss their effects. We denote the abbreviations as below:

- R_1 : Our proposed network without dilation convolutions and squeeze-and-excitation.
- R_2 : Our proposed network without dilation convolutions.
- R_3 : Our proposed network without squeeze-and-excitation.
- R_4 : Our proposed network without pyramid network.
- R_5 : Our proposed network without skip connections.
- R_6 : Our proposed network with dilation convolutions, squeeze-and-excitation, pyramid network and skip connections. We provide more deraining results in real-world datasets in Fig. 9.

From the results in Table 2, it can be observed that our method benefit from each component. Specially, R_4 , i.e. the method without pyramid network has the worse performance, which proves the importance of multi-scale information from the perspective of feature maps playing important an role in our method. In like manner, the multi-scale information from the perspective of kernel, i.e. R_2 , also plays an important role in our method, i.e. R_6 .

Furthermore, we provide a visual example as comparison of the effectiveness on every component in Fig. 10. We can observe that our result had the highest PSNR and SSIM. We provide more deraining results in real-world datasets in Fig. 9.

5 Conclusion

In this paper, we propose a novel end-to-end deep learning based deraining method. The proposed method considers two multi-scale perspectives: multi-scale kernels and multi-scale feature maps by designing the spatial contextual information aggregation module and pyramid network module, respectively. The multi-scale kernels can acquire more spatial contextual information by several dilation convolutions with different factors and the multi-scale feature maps fuse features at different scales by pyramid network module. Furthermore, the squeeze-and-excitation and skip connections are adopted, respectively, to learn the correlation between channels and transmit the information from low-level to high-level features. The experimental results shows that our proposed method outperforms the state-of-the art methods on four widely used datasets: Rain100H, Rain100L, Rain1200 and Rain1400, while

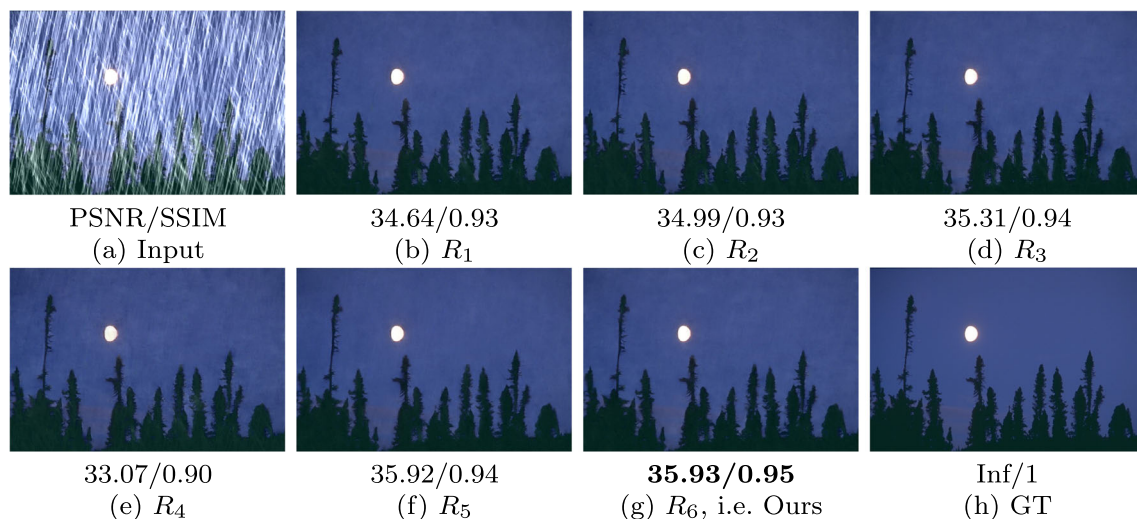


Fig. 10 An examples in synthetic datasets compared with priors based deraining methods. The proposed method generates a much clear image

having better performance on real-world dataset, and this also illustrates that the multi-scale information is important for single image deraining.

Furthermore, we plan to apply the multi-scale trick to other low-level visions, e.g. deblurring, dehazing and denoising.

References

- Brewer N, Liu N (2008) Using the shape characteristics of rain to identify and remove rain from video. In: Structural, syntactic, and statistical pattern recognition, pp 451–458. https://doi.org/10.1007/978-3-540-89689-0_49
- Chen J, Tan C, Hou J, Chau L, Li H (2018) Robust video content alignment and compensation for rain removal in a CNN framework. In: CVPR, pp 6286–6295. <https://doi.org/10.1109/CVPR.2018.00658>. http://openaccess.thecvf.com/content_cvpr_2018/html/Chen_Robust_Video_content_cvpr_2018_paper.html
- Chen Y, Hsu C (2013) A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In: ICCV, pp 1968–1975. <https://doi.org/10.1109/ICCV.2013.247>
- Chen Y, Wang Z, Peng Y, Zhang Z, Yu G, Sun J (2018) Cascaded pyramid network for multi-person pose estimation. In: CVPR, pp 7103–7112. <https://doi.org/10.1109/CVPR.2018.00742>. http://openaccess.thecvf.com/content_cvpr_2018/html/Chen_Cascaded_Pyramid_Network_CVPR_2018_paper.html
- Fu X, Huang J, Ding X, Liao Y, Paisley J (2017) Clearing the skies: a deep network architecture for single-image rain removal 26(6), 2944–2956. <https://doi.org/10.1109/TIP.2017.2691802>
- Fu X, Huang J, Zeng D, Huang Y, Ding X, Paisley J (2017) Removing rain from single images via a deep detail network. In: CVPR, pp 1715–1723. <https://doi.org/10.1109/CVPR.2017.186>
- Garg K, Nayar SK (2004) Detection and removal of rain from videos. In: CVPR, pp 528–535. <https://doi.org/10.1109/CVPR.2004.79>
- Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: CVPR, pp 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>. http://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html
- Huang D, Kang L, Yang M, Lin C, Wang Y (2012) Context-aware single image rain removal. In: ICME, pp 164–169. <https://doi.org/10.1109/ICME.2012.92>
- Huynh-Thu Q, Ghanbari M (2008) Scope of validity of psnr in image/video quality assessment. *Electron Lett* 44(13):800–801
- Kang L, Lin C, Fu Y (2012) Automatic single-image-based rain streaks removal via image decomposition 21(4), 1742–1755. <https://doi.org/10.1109/TIP.2011.2179057>
- Kim J, Lee C, Sim J, Kim C (2013) Single-image deraining using an adaptive nonlocal means filter. In: ICIP, pp 914–917. <https://doi.org/10.1109/ICIP.2013.6738189>
- Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. In: CoRR, vol. arXiv:1412.6980
- Li G, He X, Zhang W, Chang H, Dong L, Lin L (2018) Non-locally enhanced encoder-decoder network for single image deraining. In: ACM MM, pp 1056–1064. <https://doi.org/10.1145/3240508.3240636>
- Li X, Wu J, Lin Z, Liu H, Zha H (2018) Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: ECCV, pp 262–277. https://doi.org/10.1007/978-3-030-01234-2_16
- Li Y, Tan RT, Guo X, Lu J, Brown MS (2016) Rain streak removal using layer priors. In: CVPR, pp 2736–2744. <https://doi.org/10.1109/CVPR.2016.299>
- Lin T, Dollár P, Girshick RB, He K, Hariharan B, Belongie SJ (2017) Feature pyramid networks for object detection. In: CVPR, pp 936–944. <https://doi.org/10.1109/CVPR.2017.106>
- Luo Y, Xu Y, Ji H (2015) Removing rain from a single image via discriminative sparse coding. In: ICCV, pp. 3397–3405. <https://doi.org/10.1109/ICCV.2015.388>
- Ranjan A, Black MJ (2017) Optical flow estimation using a spatial pyramid network. In: CVPR, pp 2720–2729. <https://doi.org/10.1109/CVPR.2017.291>
- Shelhamer E, Long J, Darrell T (2017) Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell* 39(4):640–651. <https://doi.org/10.1109/TPAMI.2016.2572683>
- Tripathi AK, Mukhopadhyay S (2014) Removal of rain from videos: a review. *SIViP* 8(8):1421–1430. <https://doi.org/10.1007/s11760-012-0373-6>
- Wang C, Wang H, Su Z, Yang Y (2019) Embedding non-local mean in squeeze-and-excitation network for single image deraining. In: ICMEW, pp 264–269. <https://doi.org/10.1109/ICMEW.2019.00-76>
- Wang C, Zhang M, Pan J, Su Z (2019) Single image rain removal via densely connected contextual and semantic correlation net. *J Electron Imag* 28(3):033018. <https://doi.org/10.1117/1.JEI.28.3.033018>
- Wang C, Zhang M, Su Z, Wu Y, Yao G, Wang H (2019) Learning a multi-level guided residual network for single image deraining. *Signal Process Imag Commun* 78:206–215. <https://doi.org/10.1016/j.image.2019.07.003>. <http://www.sciencedirect.com/science/article/pii/S0923596519305582>
- Wang C, Zhang M, Su Z, Yao G, Wang Y, Sun X, Luo X (2019) From coarse to fine: a stage-wise deraining net. *IEEE Access* 7:84420–84428. <https://doi.org/10.1109/ACCESS.2019.2922549>
- Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Processing* 13(4):600–612. <https://doi.org/10.1109/TIP.2003.819861>
- Yang W, Tan RT, Feng J, Liu J, Guo Z, Yan S (2017) Deep joint rain detection and removal from a single image. In: CVPR, pp 1685–1694. <https://doi.org/10.1109/CVPR.2017.183>
- Zhang H, Patel VM (2018) Densely connected pyramid dehazing network. In: CVPR, pp 3194–3203. <https://doi.org/10.1109/CVPR.2018.00337>. http://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_Densely_Connected_Pyramid_CVPR_2018_paper.html
- Zhang H, Patel VM (2018) Density-aware single image de-raining using a multi-stream dense network. In: CVPR, pp 695–704. <https://doi.org/10.1109/CVPR.2018.00079>. http://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_Density-Aware_Single_Image_CVPR_2018_paper.html
- Zhang H, Sindagi V, Patel VM (2017) Image de-raining using a conditional generative adversarial network. In: CoRR, vol arXiv:1701.05957
- Zhang X, Li H, Qi Y, Leow WK, Ng TK (2006) Rain removal in video by combining temporal and chromatic properties. In: ICME, pp 461–464. <https://doi.org/10.1109/ICME.2006.262572>

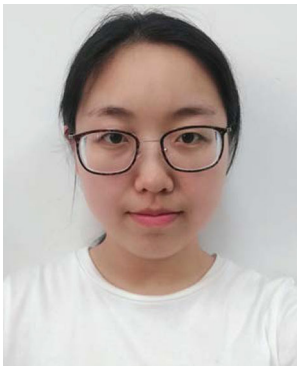
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Cong Wang received the B.S. degree from Inner Mongolia University, China, in 2017. He is currently pursuing the master's degree in computational mathematics with the School of Mathematical Sciences, Dalian University of Technology, China, under the supervision of Prof. Zhixun Su. His current research interests include deep learning, computer vision, and image processing.



Guangle Yao received the Ph.D. degree from the University of Electronic Science and Technology of China. During his Ph.D. study, he was with the Institute of Optics and Electronics, Chinese Academy of Sciences. He is currently with the Chengdu University of Technology. His research interests include action recognition, object detection, visual tracking, and deep learning.



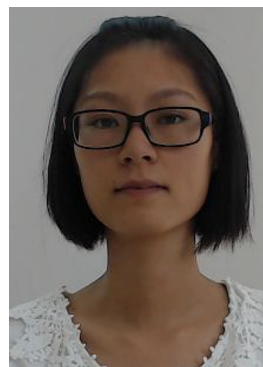
Yutong Wu received the B.S. degree from Inner Mongolia University, China, in 2019. She is currently pursuing the master's degree in computational mathematics with the School of Mathematical Sciences, Dalian University of Technology, China, under the supervision of Prof. Zhixun Su. Her current research interests include deep learning, computer vision, and image processing.



Zhixun Su received the Ph.D. degree in computational mathematics from the Dalian University of Technology, China, in 1993, where he is currently a Professor. He is also a Visiting Professor with the Guilin University of Electronic Technology. His current research interests include computer graphics, computer vision, computational geometry, and machine learning. He is currently a member of the Standing Committee of China Society of Computational Mathematics.



Yu Cai received the Ph.D. degree in computational mathematics from the Dalian University of Technology, China, in 2019. His current research interests include computer graphics, computer vision, computational geometry, and machine learning.



Hongyan Wang received the M.S. degree in computational mathematics from the Dalian University of Technology in 2013, where she is currently pursuing the Ph.D. degree with the School of Mathematical Sciences. Her research interests include image deblurring and related vision problems.