



Multi-resolution Parallel Aggregation Network for Single Image Deraining

Man Qi and Yufeng Huang^(✉)

College of Electronic and Information Engineering, Shenyang Aerospace University,
Shenyang 110136, China
651595954@qq.com

Abstract. Single image deraining is an important preprocessing task, as rain streaks awfully reduce the image quality and hinder the subsequent outdoor multimedia issues. In this paper, we explore the multi-resolution representation for rain streaks through parallel hierarchical structure and multi-scale feature extraction and fusion, termed Multi-resolution Parallel Aggregation Network (MPA-Net) in end-to-end manner. Specially, considering the significant role of multi-resolution, we employ the first stage to capture the high-resolution features, progressively introduce high-to-low resolution streams to produce more stages, and then connect all stages in parallel. In each stage, Densely Connected Residual (DCR) block is involved to guide the feature extraction. Besides, Cross-Scale Feature Fusion (CSFF) is first introduced to receive and consolidate the correlated features from different scales followed with Squeeze-and-Excitation (SE) blocks, leading to rich the resolution representations. Extensive experiments demonstrate that our method outperforms the recent comparing approaches on the frequent-use synthetic and real-world datasets.

Keywords: Single image deraining · Multi-resolution parallel · Cross-Scale Feature Fusion

1 Introduction

Images taken from the rainy conditions significantly suffer from degradation, which surely subject to blurring, color distortion and content obstruction. The visibility poor quality severely effects the performance of subsequent multimedia applications. Image deraining thus has become a vital component in the vision tasks and attracts increasing attention in the multimedia area.

In general, the deraining purpose is to recover the clear background B from the obtained rainy image $O = B + R$ with the rain layer R . Since the background and rain layer are usually unknown, the deraining can be considered as a highly ill-posed problem theoretically. To make the problem be well solved, various algorithms have been designed for single image deraining, and previous researches can be mainly classified into two

This work was supported by Innovation project of College students (S202110143005, X202110143128).

categories, including model-based and data-driven methods. Along the model-based line, prior-based model approaches treat rain removal as an optimization problem and typical methods contain Gaussian Mixture Model (GMM) [1], Discriminative Sparse Coding (DSC) [2, 3], low-rank representation [4], image decomposition [5] and filter-based deraining approaches [6–9]. The model-based ways try to make prior and subjective hypotheses on rain streaks, while those methods perform well only in some specific conditions.

Another appealing solution is a data related driven method that considers the deraining question as a non-linear function and searches the suitable parameters to apart the rainy part from the background image [10]. Inspired by the deep learning, numerous data-driven learning methods have emerged for deraining and verified remarkable restoration performance. Fu et al. [11] first employ the related network with multi-layer convolutional neural network to get and remove the rain layer, and then introduce Deep Detail Network [12] that straightly wipe off the rain streaks by decreasing the mapping range. The RESCAN [13] presents a recurrent neural network and convolutional way to apply the contextual information for single image deraining. In [14], Progressive Resnet Network (PReNet) carries out the recursive computer to effectively produce the derained images progressively. Based on the recurrent network, the work of Spatial attentive network (SPANet) [15] is able to get the spatial contextual details and obtain the spatial related information in a local-to-global manner. Jiang et al. [16] explore the multi-scale collaborative to represent the rain streaks and hierarchical deep features. In [17], the convolution dictionary is employed to represent the rain streaks and a proximal gradient descent technique is utilized to simplify the deraining model. The above deep learning-based strategies, however, have evident deficiencies in utilizing the comprehensive rain information and representations. Few efforts have been used to preserve the desired fine spatial details and strong contextual information.

In this paper, we proposed a novel parallel architecture namely Multi-resolution Parallel Aggregation Network (MPA-Net), that maintains the multi-resolution feature representations and minimizes the detailed loss for single image deraining. Our main contributions are summarized as follows.

We conduct an end-to-end MPA-Net to handle the single image deraining problem, which can generate a spatially-precise and detailed output by using a novel multi-resolution parallel feature extraction structure, while receiving and consolidating rich contextual information from different scales.

To better illustrate the rain features from different scales, Cross-Scale Feature Fusion (CSFF) is first constructed to effectively exchange and combine the cross scales information, so that the rain streaks distribution can be integrated to characterize in a collaborative manner.

Comprehensive experiments are performed on six challenging datasets (4 synthetic and 2 real-world datasets) and the deraining results demonstrate that our designed method outperforms existing state-of-the-art approaches.

2 Proposed Method

We briefly introduce the proposed MPA-Net, which can properly remove the rain streaks and maximally restore the details in the rainy images. The details of MPA-Net can be

described include the overall architecture, the feature extraction and fusion modules, as well as the loss function.

2.1 Overall Architecture of MPA-Net

We design an end-to-end deraining network to restore rainy image using the multi-resolution parallel framework, which is composed of Densely Connected Residual (DCR) blocks and CSFF with Squeeze-and-Excitation (SE) blocks, as illustrated in Fig. 1. In detail, MPA-Net has a parallel multi-resolution structure that the first stage deals with the original scale, and the other two use strided convolutions to down-sample the original input image into the changed scales as 1/2 and 1/4. Next the DCR blocks are involved in each parallel stage to extract and transport the image features. Using the features from different scales, the CSFF then performs the deep feature extraction and fusion after concatenating the multi-resolution feature information. Following the CSFF, SE block is added to adaptively rescale channel-wise features and strengthen the feature hierarchy encoding quality.

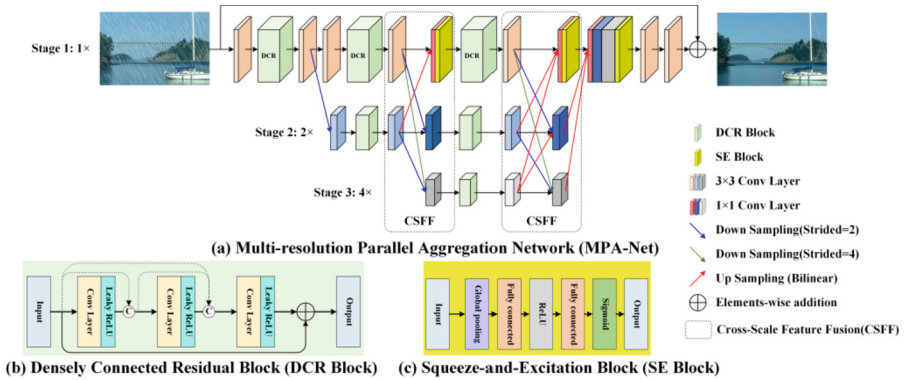


Fig. 1. The overview structure of the proposed MPA-Net framework.

Parallel Multi-resolution Pipeline is the main architecture of MPA-Net, that starts the original resolution as the first stage, progressively involves high-to-low resolution stages to produce more streams and connects all stages in parallel. Hence the later stage contains part of features from the previous stage and an extra low resolution one. As a novel feature extraction model, this structure is effectively used for extracting the fine-to-coarse features with semantically-rich as well as the coarse-to-fine one with spatially-precise feature representations. Let S_{sr} is the stage in the s th stage and r is the resolution index and the latter resolution is $1/2^{r-1}$ of the resolution of the first stage, and the 3 parallel example stages are given as follows:

$$\begin{array}{l}
 S_{11} \rightarrow S_{21} \rightarrow S_{31} \\
 \searrow S_{22} \rightarrow S_{32} \\
 \qquad \searrow S_{33}
 \end{array} \quad (1)$$

2.2 Feature Extraction and Aggregation Module

Our proposed MPA-Net contains two basic backbone blocks: (a) Densely Connected Residual (DCR) block is employed to lead the rain feature extraction and representation, (b) Squeeze-and-Excitation (SE) block is used to aggregate different scale characteristics after CSFF.

Densely Connected Residual Block [18, 19] applies the DenseNet [20] to directly receive and transport features through all the preceding layers and utilizes the Residual Net [21] to ensure the features can transport to the deeper layers in a lower computer cost. Based on the advantages of DenseNet and Residual Net, DCR block can obtain a precise negative rain feature to map the corresponding rainy image. Specifically, each DCR block consists of three convolution layers followed by leaky-ReLU with $\alpha = 0.2$ as the activation function, shown in Fig. 1(b).

The Squeeze-and-Excitation Block [22] explores the spatial and channel components, seeking to improve the feature representation capability. As depicted in Fig. 1(c), the SE block is involved to effectively aggregate different scale characteristics. In the squeeze step, a global embedding process carries out to exploit feature contextual information. Making full use of aggregated information, the excitation operation is applied to capture feature dependencies efficiently.

2.3 Cross-Scale Feature Fusion

As shown in the deraining process, it is an efficient way to get various rain streak components by combining features from different scales. Existing deraining methods usually process each scale separately or just exchange information only in an adjacent manner. Apart from the mentioned methods, the proposed method explores a novel CSFF mechanism to fuse the comprehensive feature information. Here is a scheme showing the example of CSFF unit, which sets the third stage into 3 exchange blocks. Each CSFF unit has 3 parallel convolution layers with an exchange unit across the parallel units, which can be expressed as:

$$\begin{array}{ccccc}
 S_{31}^1 \searrow & \nearrow S_{31}^2 \searrow & \nearrow S_{31}^3 \searrow & & \\
 S_{32}^1 \rightarrow C_3^1 \rightarrow S_{32}^2 \rightarrow C_3^2 \rightarrow S_{32}^3 \rightarrow C_3^3 & & & & (2) \\
 S_{33}^1 \nearrow & \searrow S_{33}^2 \nearrow & \searrow S_{33}^3 \nearrow & &
 \end{array}$$

where S_{sr}^b denotes the convolution layer in the b th block and the s th stage for the r th resolution, and the corresponding exchange unit is C_s^b .

The operation generated reliable feature representations by fusing the multi-scale features produced in the parallel stages. Mathematically, each output is a feature aggregation arriving multiple parallel streams, can be defined as:

$$\begin{aligned}
 s_1^3 &= f_{1 \times 1}^1(C(s_{1,2,0}, s_{2,2,2}, s_{3,2,2}); \eta) \\
 s_2^3 &= f_{1 \times 1}^2(C(s_{2,2,0}, s_{1,2,1}, s_{3,2,1}); \mu) \\
 s_3^3 &= f_{1 \times 1}^3(C(s_{1,2,2}, s_{2,2,1}, s_{3,2,0}); \tau)
 \end{aligned} \tag{3}$$

where η , μ , τ are the hyperparameters of the network respectively, and C is the feature map.

According to the final fused features, we can formulate the derained result as:

$$F_{out} = f_{3 \times 3} \left(f_{1 \times 1} \left(C \left(s_1^3, f \left(s_2^3 \right)_{\uparrow 2}, f \left(s_3^3 \right)_{\uparrow 4}, se \right); \theta_1 \right); \theta_2 \right) + F_0 \quad (4)$$

where se represents the SE block, \uparrow means the up sampling. $f_{n \times n}(\cdot)$ denotes a convolution of size $n \times n$ and $\{\theta_1, \theta_2\}$ indicates the hyperparameters of the network.

2.4 Loss Function

Generally, the derained output of proposed MPA-Net should be equal to the clean image in certain level. Thus, we adopt two classical loss function provides an efficient way to refine the difference between the derained image and the corresponding clean ground-truth image in the per-pixel level. Besides using per-pixel loss, L_{SSIM} is used to value the structural similarity for the derained process. Finally, the total loss function can be formulated as:

$$L = L_1 + \lambda L_{SSIM} \quad (5)$$

where λ is the weight parameter.

3 Experimental Results

In this section, we describe the experiment datasets and implementation message in details. Then comprehensive deraining researches are employed to demonstrate the effectiveness of the designed MPA-Net against the current deraining approaches. In addition, ablation studies are conducted to validate the efficiency of our designed model.

3.1 Experiment Settings

Datasets. We carry out rain removal experiments on four updated synthetic datasets: Rain200L/H [23], Rain800 [24], and Rain1400 [12], with numerous rain streaks of diverse sizes, shapes and directions. Besides, some real-world data are collected to assess the presentation of deraining and two related datasets are involved: the first one (called SPA-Data) that the rainy image is real and its ground truth is obtained by human labeling and multi-frame fusion [15], and the other with 167 rainy images collected by Internet. The detailed descriptions are tabulated in Table 1, together with the synthetic and real-world datasets.

Table 1. Datasets description. Values indicate the number of clean/rainy image pairs.

Datasets	Training set	Testing set	Type
Rain200L	1,800	200	Synthetic
Rain200H	1,800	200	Synthetic
Rain800	700	100	Synthetic
Rain1400	12,600	1400	Synthetic
SPA-Data	638,492	1000	Real-world
Internet-Data	0	167	Real-world

Setting. The detailed architecture and parameter settings of the proposed MPA-Net are depicted in Fig. 1 using Pytorch framework. In MPA-Net, the parallel stage number is 3 with the channel dimensions of 32, 64 and 128 at the corresponding resolutions as 1, 1/2 and 1/4, respectively. In the training process, we randomly select 64×64 patch pairs from the training datasets as input, and the loss function weight λ is 0.2. To accelerate the training process, Adam optimization is applied with a batch size of 16, as well as the initial learning rate is 1×10^{-3} , and then multiplied by 0.1 after every 25 epochs. Our model is trained with 200 epochs for the Rain200H, Rain200L, and Rain800 datasets, 100 epochs for Rain1400 datasets and 25 epochs for SPA-Data datasets. All the comparing testing experiments perform with the same datasets and hardware environment on the NVIDIA Tesla V100 GPU (16G).

3.2 Results on Synthetic Datasets

We compare our method with other five state-of-the-art image deraining methods, including RESCAN [13], PReNet [14], SPANet [15], MSPFN [16], and RCD-Net [17]. According to the ground truth in synthetic datasets, we perform the quantitative comparisons using Peak Signal to Noise Ratio (PSNR) and Structural SIMilarity index (SSIM). As shown in the Table 2, our proposed method gets the highest values both in PSNR and SSIM, which reflect the excellent performance and robustness of MPA-Net. The notable increasing scores in Rain200H and Rain800 reveal that our model can properly restore the rainy images especially in the heavy rain with various rainy conditions.

Besides the quantitative results, we further present several challenging examples for visual observation comparisons in Fig. 2. As displayed, the RESCAN leaves too many rain streaks in the derained images, particularly in the heavy rain condition. Clearly, PReNet, SPANet and MSPFN can remove the rain streaks in most of rain cases, while there are still some rain left in the distant or complex sceneries. By observing zoomed parts of image, the main drawbacks of RCDNet are that it tends to blur the contents and fails to reconstruct the scene detail information, and these defects can also be found in the above deraining methods. In contrast, our proposed MPA-Net can deal with majority of rain streaks in diverse rain distribution with complex background. In addition, another benefit can be found is being good at restoring the detailed structure information.

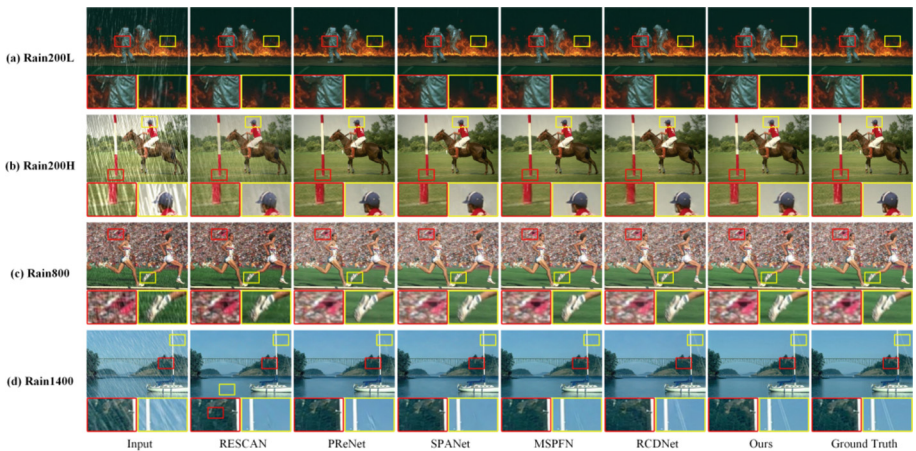
Table 2. Quantitative results evaluate of average PSNR and SSIM metrics on five benchmark datasets.

Datasets	Rain200L	Rain200H	Rain800	Rain1400	SPA-Data
Metrics	PSNR/SSIM				
Input	26.70/0.8438	13.07/0.3733	22.55/0.6850	25.24/0.8097	34.15/0.9269
RESCAN	36.94/0.9812	26.62/0.8415	24.09/0.8412	32.03/0.9314	38.11/0.9707
PReNet	36.28/0.9793	27.64/0.8846	22.83/0.7905	32.55/0.9459	40.16/0.9816
SPANet	35.60/0.9744	26.32/0.8581	24.37/0.8618	29.85/0.9148	40.24/0.9811
MSPFN	36.43/0.9810	27.78/0.8993	25.35/0.8567	32.80/0.9321	39.95/0.9768
RCD-Net	35.28/0.9710	26.18/0.8357	24.59/0.8216	33.04/0.9472	41.47/0.9834
Ours	38.65/0.9845	28.16/0.9042	26.33/0.8760	33.68/0.9579	43.02/0.9885

3.3 Results on Real-World Datasets

For practical use, we conduct additional comparisons against other deraining related algorithms on the mentioned two real-world rainy datasets. Table 2 in the last column and Fig. 3(a) compare the results on SPA-Data of all competing methods visually and quantitatively. As the natural image are more complex, all the competing methods leave some rain streaks even in the less rain streaks condition. As expected in the SPA-Data datasets, our method still exhibits remarkable performance with the better quantitative values and less rain streaks left.

Furthermore, we choose other two challenging samples from Internet-Data. For fair comparison, all the methods employ the pre-trained model trained on the Rain200H dataset to evaluate. As shown in Fig. 3(b), the rainy picture has complicated spatial space and content with heavy rainy condition. t, all the competing methods fail to remove the rain streaks far from the camera in the complex real rainy scenarios. Zooming the color

**Fig. 2.** Visual comparison of four synthetic examples, including (a) Rain200L, (b) Rain200H, (c) Rain800, and (d) Rain1400.

boxes, the other methods loss the details and blur the scene to certain extent. As the various light environment in Fig. 3(c), the above deraining algorithms fail to figure out the rain streaks from the complex surrounding background. It can be observed that the proposed method significantly competes others in removing the majority of rain streaks while preserving image details even in the dark surrounding and light effecting.

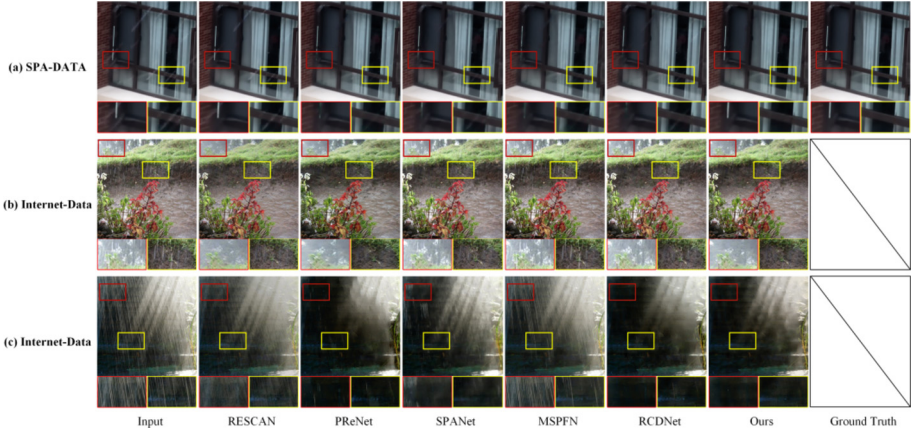


Fig. 3. Visual comparison of all the competing methods on two real-world datasets, including SPA-Data (a) and Internet-Data (b, c).

3.4 Ablation Studies

We study the main component impacts and parameter choices on the final performance. All the following ablation studies are completed in the same situation using the Rain200H dataset.

Parallel Multi-resolution Stages Number. To investigate the different number influences, we implement experiments on different numbers of parallel multi-resolution stages. From the Table 3, the increased parallel stages can lead to higher SSIM and PSNR, which bring a total gain of 2.15 dB and 0.0569 over the one stage that means better deraining performances. To balance the model performances and memory, we choose stages = 3 for our MPA-Net.

Table 3. Ablation study on different number of parallel multi-resolution stages.

Metrics	Stage = 1	Stage = 2	Stage = 3
PSNR/SSIM	26.05/0.8473	27.69/0.8850	28.16/0.9042

Feature Extraction and Aggregation Module. We analysis the effect of feature extraction and aggregation module that consist of DCR and SE blocks. The baseline module is constructed by using convolution layers in series. As displayed in Table 4, the greatest performance can be realized by employing DCR and SE blocks both, which can verify its effectiveness in the rain removal tasks.

Table 4. Ablation study on feature extraction and aggregation module.

Modules	M_1	M_2	M_3
Basic block	✓		
DCR block		✓	✓
SE block			✓
PSNR/SSIM	27.74/0.8991	27.95/0.9013	28.16/0.9042

Feature Fusion Strategies. We further perform an ablation analysis on feature fusion mechanism and three feature fusion methods as follows: (a) **W/O scale exchange** is no exchange between multi-resolution stages. (b) **adjacent-scale exchange** is only exchange between two adjacent stages. (c) **cross-scale exchange** is our designed method that fuses the features from all the stages. In general, Fig. 4 provides the visual and quantitative deraining results of three stated feature fusion strategies. As shown in Fig. 4(c), the zoomed color boxes perform better in rain removal, texture restoration and less artifact. The evaluation criteria can also reflect the CSFF get the better results under the following pictures.

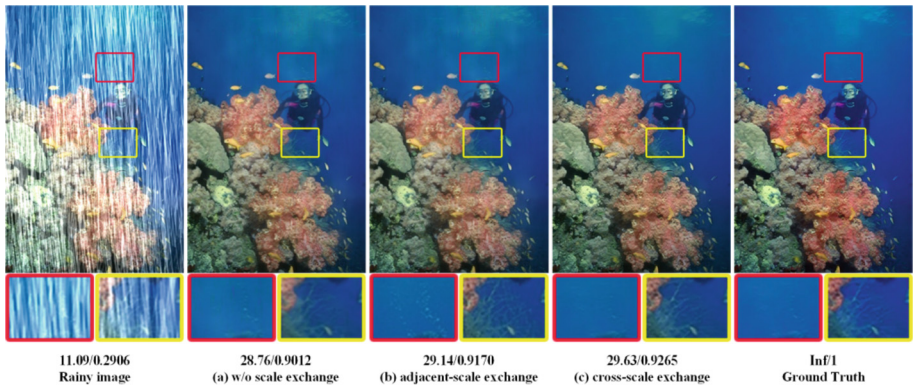


Fig. 4. Quantitative and quantitative comparison of different feature fusion strategies, with the explanation of PSNR/SSIM in the bottom of derained images.

4 Conclusion

In this paper, we propose a multi-resolution parallel aggregation network (MPA-Net) to handle the single image deraining. An original multi-resolution parallel architecture is first utilized to extract and aggregate the multi-scale features, so that the complementary parallel streams are dedicated to spatially-precise generating and provide better contextualized features. In MPA-Net, DCR block is involved to explore the feature extraction and fully propagation. In addition, an innovative CSFF mechanism is introduced to realize comprehensive information exchange, so that the features across multi-resolution stages are progressively fused together for improved representation learning. Experimental results on synthetic and real-world rainy images both demonstrate that our method outperforms other state-of-the-art approaches considerably.

References

1. Li, Y., et al.: Rain streak removal using layer priors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2736–2744 (2016)
2. Lei, J., et al.: An Image Rain Removal algorithm based on the depth of field and sparse coding. In: 24th International Conference on Pattern Recognition (ICPR), pp. 2368–2373 (2018)
3. Zhang, H., Patel, V.M.: Convolutional sparse and low-rank coding-based rain streak removal. In: IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1259–1267 (2017)
4. Zhang, L., Zuo, W.: Image restoration: from sparse and low-rank priors to deep priors [lecture notes]. *IEEE Sig. Process. Mag.* **34**(5), 172–179 (2017)
5. Lian, Q., et al.: Single image rain removal using image decomposition and a dense network. *IEEE/CAA J. Automatica Sin.* **6**(6), 1428–1437 (2019)
6. Chen, X., Huang, Y., Xu, L.: Multi-scale attentive residual dense network for single image rain removal. In: Ishikawa, H., Liu, C.-L., Pajdla, T., Shi, J. (eds.) ACCV 2020. LNCS, vol. 12623, pp. 286–300. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-69532-3_18
7. Wang, F.L., et al.: When a conventional filter meets deep learning: basis composition learning on image filters. *ArXiv, abs/2203.00258* (2022)
8. Zheng, X., Liao, Y., Guo, W., Fu, X., Ding, X.: Single-image-based rain and snow removal using multi-guided filter. In: Lee, M., Hirose, A., Hou, Z.-G., Kil, R.M. (eds.) ICONIP 2013. LNCS, vol. 8228, pp. 258–265. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-42051-1_33
9. Zhang, X., et al.: Rain removal in video by combining temporal and chromatic properties. In: 2006 IEEE International Conference on Multimedia and Expo, pp. 461–464 (2006)
10. Chen, X., Huang, Y., Xu, L.: Multi-scale hourglass hierarchical fusion network for single image deraining. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 872–879 (2021)
11. Fu, X., et al.: Clearing the skies: a deep network architecture for single-image rain removal. *IEEE Trans. Image Process.* **26**(6), 2944–2956 (2017)
12. Fu, X., et al.: Removing rain from single images via a deep detail network, pp. 1715–1723 (2017)
13. Li, X., Wu, J., Lin, Z., Liu, H., Zha, H.: Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 262–277. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_16

14. Ren, D., et al.: Progressive image deraining networks: a better and simpler baseline. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3937–3946 (2019)
15. Wang, T., et al.: Spatial attentive single-image deraining with a high quality real rain dataset. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 12270–12279 (2019)
16. Jiang, K., et al.: Multi-scale progressive fusion network for single image deraining, pp. 8346–8355 (2020)
17. Wang, H., et al.: A model-driven deep neural network for single image rain removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3103–3112 (2020)
18. Park, Y., et al.: MARA-Net: single image deraining network with multi-level connection and adaptive regional attention. arXiv preprint [arXiv:2009.13990](https://arxiv.org/abs/2009.13990) (2020)
19. Wei, Y., et al.: A coarse-to-fine multi-stream hybrid deraining network for single image deraining. In: 2019 IEEE International Conference on Data Mining (ICDM), pp. 628–637 (2019)
20. Huang, G., et al.: Densely connected convolutional networks, pp. 2261–2269 (2017)
21. He, K., et al.: Deep residual learning for image recognition, pp. 770–778 (2016)
22. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks, pp. 7132–7141 (2018)
23. Yang, W., et al.: Deep joint rain detection and removal from a single image, pp. 1685–1694 (2017)
24. Zhang, H., Sindagi, V., Patel, V.M.: Image de-raining using a conditional generative adversarial network. *IEEE Trans. Circ. Syst. Video Technol.* **30**(11), 3943–3956 (2020)