



Multi-scale Attentive Residual Dense Network for Single Image Rain Removal

Xiang Chen¹ , Yufeng Huang¹  , and Lei Xu² 

¹ Shenyang Aerospace University, Shenyang 110136, China
cx@cvgpu.com, yufengh_sau@sina.com

² Shenyang Fire Science and Technology Research Institute of MEM,
Shenyang 110034, China
xulei@syfri.cn

Abstract. Single image deraining is an urgent yet challenging task since rain streaks severely degrade the image quality and hamper the practical application. The investigation on rain removal has thus been attracting, while the performances of existing deraining have limitations owing to over smoothing effect, poor generalization capability and rain intensity varies both in spatial locations and color channels. To address these issues, we proposed a Multi-scale Attentive Residual Dense Network called MARD-Net in end-to-end manner, to exactly extract the negative rain streaks from rainy images while precisely preserving the image details. The architecture of modified dense network can be used to exploit the rain streaks details representation through feature reuse and propagation. Further, the Multi-scale Attentive Residual Block (MARB) is involved in the dense network to guide the rain streaks feature extraction and representation capability. Since contextual information is very critical for deraining, MARB first uses different convolution kernels along with fusion to extract multi-scale rain features and employs feature attention module to identify rain streaks regions and color channels, as well as has the skip connections to aggregate features at multiple levels and accelerate convergence. The proposed method is extensively evaluated on several frequent-use synthetic and real-world datasets. The quantitative and qualitative results show that the designed framework performs better than the recent state-of-the-art deraining approaches on promoting the rain removal performance and preserving image details under various rain streaks cases.

1 Introduction

Images captured under outside conditions often affect by rain drops/streaks, which alter the image color and obstruct or distort content [1–3]. The visibility degradations and artifacts severely hinder the performance of computer vision tasks, like target detection [4], object tracking [5] and image recognition [6].

Xi. Chen and Y. Huang—Authors contributed equally to this work, listed alphabetically.

Hence, rain removal has become an important preprocessing step and attracted much attention lately in the pattern recognition and computer vision [7–11].

In the recent years, various researches have been proposed for the single image deraining, and existing methods can be roughly divided into two categories: model-based and data-driven approaches [12]. The model-based methods can be further divided into filter-based and prior-based ones. Considering the single deraining as a signal filtering task, filter-based methods employ the edge preserving and physical properties filter to recover the rain-free images [7, 9, 10, 13]. While the prior-based methods consider the deraining as an optimization question and apply handcrafted image prior to regularize the solution process, including discriminative sparse [9, 14] Gaussian mixture model (GMM) [10] and low-rank representation [15]. Different from the model-based approaches, data-driven methods formulate deraining as a procedure of learning a non-linear function and find the proper parameters to map the rainy part into the background scene [16]. Motivated by the success of deep learning, the researchers model the mapping function with the convolution neural networks (CNN) or the Generative Adversarial Networks (GANs) [17]. The CNN methods directly get the deterministic mapping function from the rainy image to the clear background [3, 18–21] and the GANs produce the deraining image inspired by their abilities on synthesizing visually appealing clean image [22].

Although effective in certain applications, the above methods suffer from several limitations. The rationality of model-based strategies refers to the subjective assumptions, and hence such learning schemes may not always adapt to diverse rainy conditions. The deep learning techniques neglect the intrinsic knowledge of rain, which make themselves easily trapped into the overfitting to training process. Most of the deraining methods generally fail to restore the structures and details, even yielding blurry background scenes. And it is difficult to get the derained image for a real-world rainy image, when the background and rain streaks merge with each other, especially in the heavy rain condition.

To address the mentioned issues, this paper conducts a Multi-scale Attentive Residual Dense Network called MARD-Net by leveraging the strong propagable capabilities of the dense network with advance residual blocks. The dense network provides a powerful capability to connect to all subsequent layers, from which the feature maps can be fully reused and smoothly transported to each layer. Multi-Scale Attention Residual Block (MARB) is introduced to better utilize multi-scale information and feature attention for improving the rain feature representation capability. Combing the features of different scales and layers, multi-scale manner is an efficient way to capture various rain streak components especially in the heavy rainy conditions. Reference to bright channel prior (BCP) [23] and the uneven distribution of rainy images, channel-wise and spatial attention mechanisms are involved in the MARB, since it helps the network to adjust the three-color channels respectively and identify the rainy region properly. We evaluate the proposed network on the public competitive benchmark synthetic and real-world datasets and the results significantly outperform the current outstanding methods on most of the deraining tasks.

In summary, our major contributions are summarized:

- We propose an end to end MARD-Net to address the single image deraining problem, which can effectively remove the rain streaks while well preserve the image details. The modified dense network is applied to boost the model performance via multi-level features reuse and maximum information flow between layers. It can reduce the loss of information transmission and vanishing-gradient, while fully utilizing the features of different layers to restore the details.
- To our knowledge, the Multi-Scale Attention Residual Block (MARB) is first constructed to improve the representation of rain streaks. The different convolution kernels along with fusion are employed to get multi-scale features for adapting the various rain cases. Then the feature attention module is applied to better extract the feature by using color channel and spatial information.
- Extensive experiments are carried out on six challenging datasets (4 synthetic and 2 real-world datasets). Our proposed network outperforms the state-of-the-art methods in visually and quantitatively comparisons. Furthermore, ablation studies have been provided to verify the rationality and necessity of important modules involved in our network.

2 Background and Related Work

An observed rainy image I can be generally modeled as a linear sum of a clean background B and a rain component R , which can be expressed by the formula as:

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

Based on the Eq. 1, deraining methods can be done either by removing R from I to get B or by directly estimating B from I . To make the deraining better be solved, various conventional methods adopted numerous prior models about rain or background to constrain the solution space [16]. Fu et al. [24] considered the image rain removal as a signal decomposition issue and performed the bilateral filter to decompose the low- frequency and high-frequency layers for getting the derained result. The discriminative sparse coding [9] was proposed to approximate the rain and background layers. To represent scales of rain and various orientations, Li et al. [10] employed the GMM based on patch priors for the rainy image to remove the rain streaks. These traditional methods usually make simple yet subjective hypotheses on the rain distribution and falling character, which work well only in some certain cases.

In recent years, numbers of deep learning based single image deraining approaches were proposed through constructing various networks [16]. Fu et al. [18] first designed the DerainNet for the image deraining, and further proposed Deep Detail Network (DDN) [3], which directly reduced the mapping range and removed the rain streak. Conditional GAN [25] was utilized to deal with the rain removal problem. Later, Qian et al. [26] introduced the attention mechanism into the GAN network, and tried to learn more about rain regions and their

surrounding features. With the help of depth-guided attention mechanism, Hu [27] developed an end to end Depth-attentional Features network (DAF-Net) to estimate the rain-free image that formulated the attention to learn the feature and regressed a residual map. Zhang et al. [28] presented a multi-stream dense network combined with residual classifier process. In [29], Gated Context Aggregation Network (GCANet) was an end-to-end network, which try to restore the rain image from the gridding artifacts by adopting the smoothed dilation technique. The work of [20] offered a recurrent squeeze and excitation context aggregation net (RESCAN) to tackle the overlap layer in the rainy image. To handle with various rain scenes, Yang et al. [16] designed a multi-stage joint rain detection and estimate network (JORDER_E) and discussed the possible aspects as architecture and loss that effected on the deraining task. Lightweight Pyramid Networks (LPNet) [30] pursued a light-weighted pyramid to remove rain, so that the obtained network became simple and contained less parameters. In [19], the PReNet performed a stage-wise operation that repeatedly unfolded several Resblocks and a LSTM layer to effectively generate the rain-free images progressively. The work of [31], Spatial attentive network (SPANet) developed a spatial attention unit based on recurrent network and utilized a branch to capture the spatial details for removing rain in a local-to-global manner. However, most existing deraining researches do not notice the underlying connection of rain streaks across different scales and few attempts have been made to exploit the feature attention of the rainy image.

3 Proposed Method

The goal of this paper is to remove the rain, while maximally keeping the original structure and color in the image. We propose a Multi-scale Attentive Residual Dense Network (MARD-Net), including the overall network architecture, multi-scale attention residual block (MARB) and loss function.

3.1 Design of MARD-Net

We propose an end-to-end trainable MARD-Net that can take diverse rainy images as input and can well represent the rain steak feature through the MARB module. The overall architecture of MARD-Net is illustrated in Fig. 1. Based on the DenseNet framework, the proposed method ensures the maximum information flow through rain feature reuse, yielding condensed models that efficiently reduce the parameter numbers and are easy to be trained. Due to the different scales and shapes of rain streaks, it is an effective way that combining features from different scales and feature attention module to capture various rain streak components. In addition, skip connections are generally used in the residual block as they can aggregate features at multiple scales and accelerate the training process. Further, the MARB can better capture the feature with different scales and rain streak structure information, as discussed in the following parts.

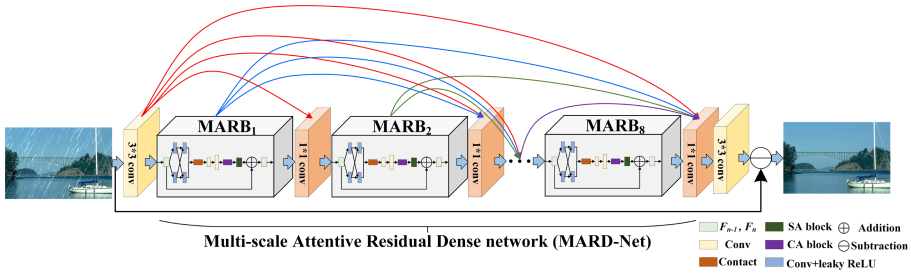


Fig. 1. The overall architecture of our proposed MARD-Net for image deraining. MARB is shown in Fig. 2. The goal of the MARD-Net is to estimate the clean image from the corresponding rainy image. The input of each layer consists of all preceding feature-maps and combines features by concatenating them. The blocks with same color share the same parameters.

3.2 Multi-scale Attention Residual Block (MARB)

Combining features at different scales effectively, multi-scale features have been widely employed to get a better information of the object and its surrounding context. In order to better solve the rain removal problem, an attention mechanism is introduced to strengthen the capability of extracting information, which is beneficial to improve the network performance and accuracy [32, 33]. Inspired by these ideas, the MARB is applied to extract multi-scale features and guide to learn rain information effectively, as shown in Fig. 2.

The MARB can be described in detail mathematically. Referring to Fig. 2, the MARB is set to have an input feature of F_0 , which first passes through the different convolutional layer which sizes are 3×3 and 5×5 respectively, and its output is expressed as:

$$F_1^{3 \times 3} = f_{3 \times 3}(F_0; \eta_0^{3 \times 3}) \tag{2}$$

$$F_1^{5 \times 5} = f_{5 \times 5}(F_0; \eta_0^{5 \times 5}) \tag{3}$$

where $F_1^{n \times n}$ denotes the output of a convolution of size $n \times n$, $f_{n \times n}(\cdot)$ presents a convolution of size $n \times n$, and $\eta_0^{n \times n}$ means the hyperparameter of a convolution of size $n \times n$. The image features can be further extracted by the convolution layer of size 3×3 or 5×5 respectively.

$$F_2^{3 \times 3} = f_{3 \times 3}((F_1^{3 \times 3} + F_1^{5 \times 5}); \eta_1^{3 \times 3}) \tag{4}$$

$$F_2^{5 \times 5} = f_{5 \times 5}((F_1^{3 \times 3} + F_1^{5 \times 5}); \eta_1^{5 \times 5}) \tag{5}$$

where $F_2^{n \times n}$ denotes the output of a convolution layer of size $n \times n$ and $\eta_1^{n \times n}$ means the hyperparameter of a convolution layer of size $n \times n$. The activation functions for these convolution layers use Leaky-ReLU with $\alpha = 0.2$ as the activation function in general.

To further improve the network representation capabilities, MARB introduces the inter-layer multi-scale information fusion, which can integrate multi-scale information with the features of different scales. This structure guarantees

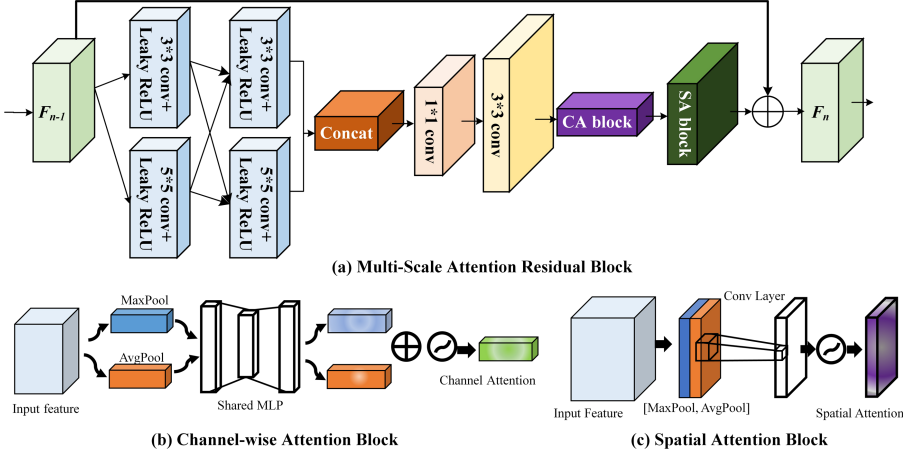


Fig. 2. (a) The architecture of our proposed Multi-Scale Attention Residual Block (MARB) consists the multi-scale residual blocks and the feature attention module. The feature module has two sequential sub-modules: channel-wise attention (CA) block (b) and spatial attention (SA) block (c).

that the input information can be propagated through all parameter layers, so that the MARB can learn the primary image features through different scales and features. To accelerate the training procedure, global skip connection is introduced among different MARB modules, which helps back-propagate gradient to update parameters. This skip connection can also propagate lossless information through the entire network directly, therefore it is useful for estimating the final derained image.

Rain density distribution patterns vary dramatically across different color channels, therefore the BCP prior [23], may be an effective way to get different weighted information for channel features. In [34], the research also finds that the channel-wise attention scheme with BCP can help the network better preserve the pixel brightness in derained images than previous methods, which treat different channels equally. Hence, the channel attention can capture the rain region and assist to extract important features. Meanwhile, the distributions of rain streaks are almost unevenly and may vary across different spatial locations. Therefore, the spatial attention may be also important to deal with rain region. Therefore, the spatial attention may be also important to deal with rain region. Multi-scale information fusion is achieved by using convolution layers of size 1×1 and 3×3 , while channel-wise and spatial attention modules can also be introduced to improve feature fusion. We can reformulate the final output as:

$$F_{out} = sa(ca((f_{3 \times 3}(f_{1 \times 1}(C(F_2^{3 \times 3}, F_2^{5 \times 5}); \eta_2); \eta_3); \delta_0); \delta_1)) + F_0 \quad (6)$$

where F_{out} denotes the output of the MARB, $sa(\cdot)$ and $ca(\cdot)$ indicate the spatial attention mechanism and channel attention mechanism, respectively, and

$\{\eta_2; \eta_3; \delta_0; \delta_1\}$ indicates the hyperparameters of the MARB output. This operation enables the network to better explore and reorganize features in different scales.

3.3 Loss Function

Mean squared error (MSE) is widely used as the loss function to evaluate the derained image and its corresponding ground truth. However, it usually leads to the blurry and over-smoothed of high-frequency textures, which do harm to remove the rain and restore the image content. To address the above drawbacks, we combine the MSE with SSIM as the proposed loss function to balance between image deraining and background structure preservation.

MSE Loss. Given an input rainy image I_i , the output rain-free image is $G(I_i)$ and the ground truth is J_i . Hence, a pixel-wise MSE loss can be defined as follows:

$$L_{MSE} = \frac{1}{HWC} \sum_{x=1}^H \sum_{y=1}^W \sum_{z=1}^C \|G(I_i) - J_i\|^2 \quad (7)$$

where H , W and C represent height, width, and number of channels respectively.

SSIM Loss. SSIM is an important indicator to measure the structural similarity between two images [35], with the equation as follows:

$$SSIM(G(I), J) = \frac{2\mu_{G(I)}\mu_J + C_1}{\mu_{G(I)}^2 + \mu_J^2 + C_1} \cdot \frac{2\sigma_{G(I)}\sigma_J + C_2}{\sigma_{G(I)}^2 + \sigma_J^2 + C_2} \quad (8)$$

where μ_x , σ_x^2 are the mean and the variance value of the image: x . The covariance of two images is σ_{xy} , C_1 and C_2 are constants value used to maintain equation stability. SSIM ranges from 0 to 1 and in the deraining issue the greater value means that the derained image are more similar to the ground truth image, so the SSIM loss can be defined as:

$$L_{SSIM} = 1 - SSIM(G(I), J) \quad (9)$$

Total Loss. The total loss is defined by combing the MSE loss and the SSIM loss as follows:

$$L = L_{MSE} + \lambda L_{SSIM} \quad (10)$$

where λ is the hyperparameter that can balance the weights between the MSE loss and SSIM loss. With the proper setting, the hybrid loss can keep the per-pixel similarity while preserving the global structures, which can help the rain removal model to obtain additional realistic derained images.

4 Experiments

In this section, we conduct comprehensive experiments to demonstrate the effectiveness of the proposed MARD-Net for image draining. Compared with the current state-of-the-art algorithms, the qualitative and quantitative analysis are carried out on the synthesized benchmark and real-world rainy datasets. In addition, ablation studies also perform to validate the effectiveness of our designed components.

4.1 Datasets and Performance Metrics

Datasets. We carry out experiments on four **synthetic benchmark datasets**: Rain100L [1], Rain100H [1], Rain12 [10] and Rain1400 [3], including rain streaks with various sizes, shapes and directions. With only one type of rain streaks, Rain100L contains 200 image pairs for training and the remaining 100 ones for evaluation. Compared with Rain100L, Rain100H is a dataset with 5 types of rain streak directions and consists of 1800 image pairs for training and 100 ones for testing. By training on Rain100L, like [19], Rain 12 is utilized to be a testing sample since it only includes 12 image pairs. With 14 types of streak orientations and magnitudes, Rain 1400 has 14000 rain synthetic images from 1000 clean images, where 12600 rainy images are selected as training data and the other 1400 ones are chosen for testing. **Real-world Datasets** are very important to evaluate the performance of deraining and two real-world datasets are involved for testing: the one with 185 real-world rainy images collected by [12], and the other with 34 images released by [36].

Performance Metrics. As the ground truths available for synthetic data, the rain removal method’s performance can be evaluated on Peak Signal-to-Noise Ratio (PSNR in dB) and the Structural Similarity Index Measure (SSIM) [35]. The higher value of PSNR indicates better performance to remove rain streaks from the rainy image. The greater SSIM score nearest to 1 means that the two different images are more similar to each other. As no ground truth exits for real-world datasets, we may present the visual comparisons and zoom local parts for the real-world images.

4.2 Training Details

The detailed structure and parameter settings of the proposed model are given in Fig. 1 using Pytorch framework and the number of MARB is set to 8 to get a better result as discussed in the Ablation Studies part. Using Adam optimization [37] in the training process, its parameters can be set as followed: the learning rate is 1×10^{-3} and batch size is 32. Considering the loss function, the weight value of SSIM is set as $\lambda = 0.2$, empirically. We train the network for 100 epochs in total and reduce the learning rate by half every 25 epochs on a workstation with a NVIDIA Tesla V100 GPU (16 G). All subsequent experiments are performed with the same environment described in implementation details. To

encourage more comparisons from the community, we will publicly release our codes on GitHub: <https://github.com/cxtalk/MARD-Net>.

4.3 Evaluation on Synthetic Datasets

In this section, we reveal the effective performance of our method by conducting a mass of experiments on frequent-used synthetic datasets: Rain100L, Rain100H, Rain1400 and Rain12. The proposed MARD-Net method is compared to five recent state deraining methods: GCANet [29], RESCAN [20], LPNet [30], SPANet [31] and PReNet [19]. All the methods use the source codes and default parameters specified published in the published literature. As the availability of ground truth in synthetic data, the results are evaluated using PSNR and SSIM. Table 1 shows the average evaluation criteria of each pairs of rain-free and deraining images with diverse and complicated rain types. From the table, the proposed method obtained the highest value of PSNR and SSIM in all synthetic datasets, which reflected the better robustness and generality of MARD-Net. The most notably increasing score in Rain100H and Rain1400 noted that our approach could properly remove the rain streaks and restore the image especially in the heavy rain as well as in the various rainy conditions.

Table 1. Quantitative results evaluate in term of average PSNR (dB) and SSIM on the synthesized benchmark datasets, including Rain100L, Rain100H, Rain1400 and Rain12. The best results are highlighted in bold. It is worth noting that the PSNR and SSIM are calculated in the RGB color space.

Datasets	Rain100L		Rain100H		Rain1400		Rain12	
Metrics	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Input	26.89	0.8382	13.55	0.3785	25.24	0.8097	30.13	0.8553
GCANet	31.65	0.9325	23.80	0.8114	27.84	0.8406	30.76	0.8819
RESCAN	36.12	0.9691	28.88	0.8660	29.88	0.9053	36.43	0.9519
LPNet	33.39	0.9579	23.39	0.8208	26.45	0.8326	34.83	0.9568
SPANet	35.33	0.9694	25.11	0.8332	28.57	0.8913	35.85	0.9572
PReNet	37.41	0.9783	29.45	0.8980	30.73	0.9184	36.66	0.9618
Ours	37.84	0.9814	30.19	0.9153	31.68	0.9215	36.88	0.9726

In addition to the results by quantitative evaluation, we also provide visual observation derained images. Some corresponding pictures directly show visual difference in rain removal images, as particularly seen in Fig. 3 with crop and zoom in two local patch regions. As displayed, the GCANet leaves many rain streaks in the recovered images, especially in the heavy rain cases (Rain100H and Rain1400). The main drawbacks of RESCAN in the comparison show color degradation with different rain patterns (Rain1400) and there are still some streaks left after deraining. LPNet fails to remove the rain-streaks completely in

the diverse rain pattern (Rain1400) and brings serious rain artifacts and blurred region to the derained image. Clearly, PReNet and SPANet, have the ability to remove most of rain streaks in different rain cases. However, by observing zoomed color boxes, we find that they lose some detailed information and lead to color degradation to a certain extent. In general, the proposed MARD-Net can successfully remove majority of rain in various rain patterns even in the heavy rain condition, and another benefit of our method is being good at preserving of color and detailed structure information similar to the ground truths.

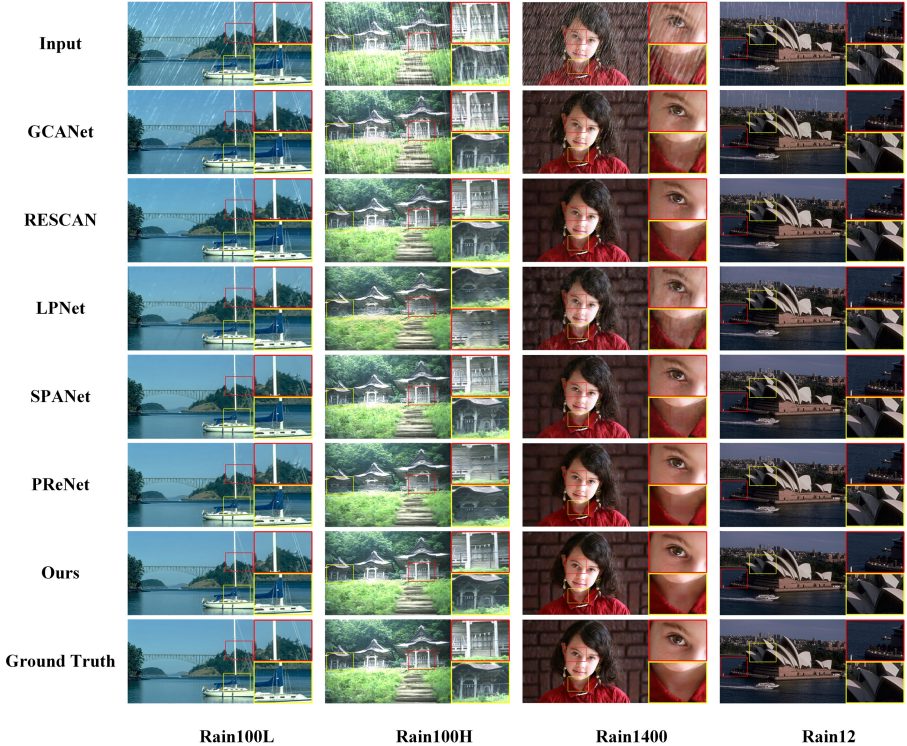


Fig. 3. Visual quality comparisons of all competing methods on synthetic datasets, including Rain100L, Rain100H, Rain1400 and Rain12. Zooming in the figure can provide a better look at the restoration quality. (Color figure online)

4.4 Evaluations on Real Rainy Images

To evaluate the effectiveness for practical use, we conduct a further comparison on the mentioned two real-world rainy datasets. Figure 4 demonstrates two real-world samples since the above one with various spatial information in the image space, while the other contains rich texture and content information. All the methods employ the pre-trained model trained on the same synthetic

rainy datasets. Even though RESCAN, PReNet and SPANet achieve remarkable rain removal performance on synthetic datasets, all competing methods leave some rain streaks to a certain extent in various spatial space and complex content condition. Since objects far or near from the camera are mainly affected differently by the rain, our method significantly removes the majority of rain streaks by introducing multi-scale features and attention information. Due to the overfitting-to training-samples process [38] and loss of the spatial information, the competing methods fail to remove the rain streak with various spatial conditions, as seen in the above picture. With complex content information and texture details in the image below, the competing methods fail to remove the rain streaks on the road. Through the zoomed color boxes, we can see the obvious detail structure and information loss for derained results. Clearly, the proposed model performs well on deraining and restoring the details and color information with feature reuse and transferring in different scales and layers.

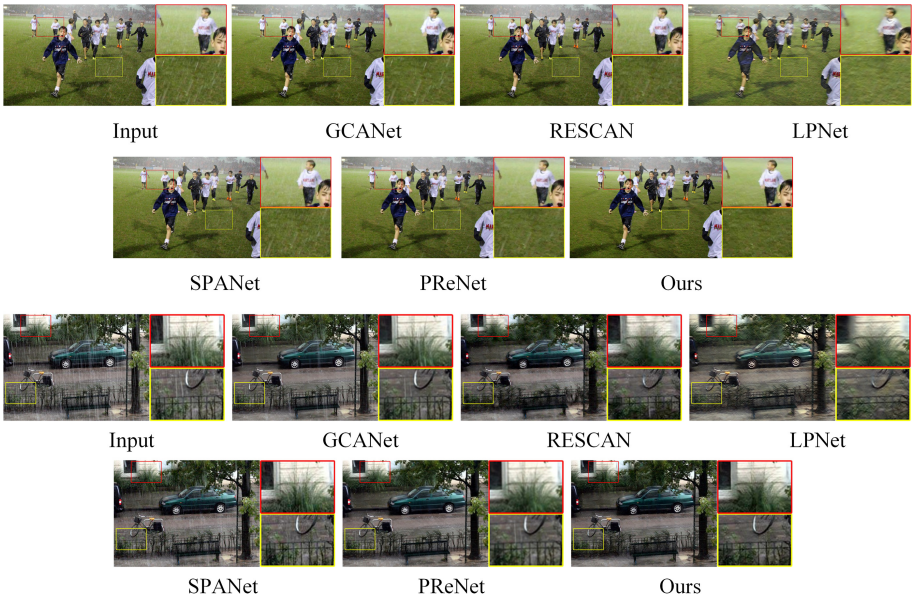


Fig. 4. Visual quality comparisons of all competing methods on real-world datasets. Zooming in the figure can provide a better look at the restoration quality.

4.5 Ablation Studies

We conduct the ablation study to explore the effectiveness of the parameters and configuration in our proposed network. All the studies are performed in the same environment by using the Rain100L dataset.

Multi-scale Attentive Residual Block Numbers

To study the influences of different numbers, we perform the experiments with different numbers of MARB to the proposed network. Specifically, MARB numbers N is set to $N \in \{2, 4, 6, 8, 10\}$ and the performances are illustrated in Table 2. As seen, increasing blocks can bring higher PSNR and SSIM values, while the value improvement is limited after $N = 8$ with extra time-consuming. Hence, 8 is chosen as our default setting to achieve the balance between effectiveness and efficiency.

Table 2. Ablation study on multi-scale attentive residual block (MARB) numbers. PSNR and SSIM results among different settings of MARD-Net on Rain100L dataset.

Block No.	$N = 2$	$N = 4$	$N = 6$	$N = 8$ (default)	$N = 10$
PSNR	34.84	35.50	36.79	37.84	37.86
SSIM	0.9635	0.9718	0.9786	0.9814	0.9810

Channel-Wise and Spatial Attention Modules

To further verify the effectiveness of feature attention modules, we conduct the studies with different variants of Multi-scale Residual Block. The baseline module is constructed by removing the channel-wise and spatial attention. As shown in Table 3, feature attention module is able to bring improvements in both PSNR and SSIM. The best performance is achieved by using the channel-wise and spatial attention both, with bringing a total gain of 1.89 dB over the baseline that verifies helpful to the task of rain removal.

Table 3. Ablation study on feature attention modules. PSNR and SSIM results among different decompositions on Rain100L dataset. The term ‘‘CA’’ and ‘‘SA’’ denote the channel-wise attention block and spatial attention block, respectively. It shows that the combination of all the designed components is the best.

	Baseline	CA	SA	SA+CA
PSNR	35.95	36.85	36.14	37.84
SSIM	0.9757	0.9785	0.9766	0.9814

Loss Function

We further investigate the impact of using the MSE and SSIM loss functions. In Table 4, the quantitative evaluations of different loss functions can be seen under the same conditions. We note that PSNR is a function of MSE, and SSIM focuses on structural similarity which is appropriate for preserving details. In this case, the quantitative performance measure of MSE and SSIM should favor the objective function that optimizes over this measure.

Table 4. Ablation study on loss functions. The results of different losses on Rain100L dataset.

	MSE	SSIM	MSE+SSIM
PSNR	36.71	35.93	37.84
SSIM	0.9671	0.9798	0.9814

5 Conclusion

In this paper, we present a novel Multi-scale Attentive Residual Dense Network named MARD-Net to handle the single image deraining problem. In MARD-Net, dense network is applied to explore the potential of network through feature reuse and fully information propagation. An innovative Multi-scale Attentive Residual Block is first utilized to identify and represent the rain streak features. Different convolution kernels along with progressive fusion are designed to explore the multi-scale rain patterns features. In addition, feature attention module is introduced to achieve the raining removal more adaptive in different color channels and spatial distribution. Extensive experiments on both frequent-use synthetic and real-world datasets demonstrate that the proposed MARD-Net achieves superior performance to the recent state deraining methods. In the future, we will further work on employing our network idea into semi/unsupervised scenarios and some other low-vision tasks.

References

1. Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.: Deep joint rain detection and removal from a single image. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1685–1694 (2017)
2. Yang, W., Tan, R.T., Feng, J., Guo, Z., Yan, S., Liu, J.: Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 1377–1393 (2020)
3. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1715–1723 (2017)
4. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149 (2017)
5. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement. *arXiv preprint arXiv:1804.02767* (2018)
6. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ICLR 2015: International Conference on Learning Representations 2015 (2015)
7. Garg, K., Nayar, S.K.: Vision and rain. *Int. J. Comput. Vis.* **75**, 3–27 (2007). <https://doi.org/10.1007/s11263-006-0028-6>
8. Zhang, X., Li, H., Qi, Y., Leow, W., Ng, T.: Rain removal in video by combining temporal and chromatic properties. In: 2006 IEEE International Conference on Multimedia and Expo, pp. 461–464 (2006)

9. Luo, Y., Xu, Y., Ji, H.: Removing rain from a single image via discriminative sparse coding. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 3397–3405 (2015)
10. Li, Y., Tan, R.T., Guo, X., Lu, J., Brown, M.S.: Rain streak removal using layer priors. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2736–2744 (2016)
11. Gu, S., Meng, D., Zuo, W., Zhang, L.: Joint convolutional analysis and synthesis sparse representation for single image layer separation. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 1717–1725 (2017)
12. Li, S., et al.: Single image deraining: a comprehensive benchmark analysis. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3838–3847 (2019)
13. Xu, J., Zhao, W., Liu, P., Tang, X.: Removing rain and snow in a single image using guided filter. In: 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), vol. 2, pp. 304–307 (2012)
14. Zhang, H., Patel, V.M.: Convolutional sparse and low-rank coding-based rain streak removal. In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1259–1267 (2017)
15. Chen, Y.L., Hsu, C.T.: A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In: 2013 IEEE International Conference on Computer Vision, pp. 1968–1975 (2013)
16. Yang, W., Tan, R.T., Wang, S., Fang, Y., Liu, J.: Single image deraining: from model-based to data-driven and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.* 1 (2020)
17. Goodfellow, I., et al.: Generative adversarial nets. In: *Advances in Neural Information Processing Systems*, vol. 27, pp. 2672–2680 (2014)
18. Fu, X., Huang, J., Ding, X., Liao, Y., Paisley, J.: Clearing the skies: a deep network architecture for single-image rain removal. *IEEE Trans. Image Process.* **26**, 2944–2956 (2017)
19. Ren, D., Zuo, W., Hu, Q., Zhu, P., Meng, D.: Progressive image deraining networks: a better and simpler baseline. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3937–3946 (2019)
20. Li, X., Wu, J., Lin, Z., Liu, H., Zha, H.: Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 262–277 (2018)
21. Yang, W., Liu, J., Yang, S., Guo, Z.: Scale-free single image deraining via visibility-enhanced recurrent wavelet learning. *IEEE Trans. Image Process.* **28**, 2948–2961 (2019)
22. Shen, Y., et al.: Mba-raingan: multi-branch attention generative adversarial network for mixture of rain removal from single images. *arXiv preprint arXiv:2005.10582* (2020)
23. Yan, Y., Ren, W., Guo, Y., Wang, R., Cao, X.: Image deblurring via extreme channels prior. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6978–6986 (2017)
24. Fu, Y.H., Kang, L.W., Lin, C.W., Hsu, C.T.: Single-frame-based rain removal via image decomposition. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1453–1456 (2011)
25. Zhang, H., Sindagi, V., Patel, V.M.: Image de-raining using a conditional generative adversarial network. *IEEE Trans. Circ. Syst. Video Technol.* 1 (2019)

26. Qian, R., Tan, R.T., Yang, W., Su, J., Liu, J.: Attentive generative adversarial network for raindrop removal from a single image. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2482–2491 (2018)
27. Hu, X., Fu, C.W., Zhu, L., Heng, P.A.: Depth-attentional features for single-image rain removal. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8022–8031 (2019)
28. Zhang, H., Patel, V.M.: Density-aware single image de-raining using a multi-stream dense network. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 695–704 (2018)
29. Chen, D., et al.: Gated context aggregation network for image dehazing and deraining. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1375–1383 (2019)
30. Fu, X., Liang, B., Huang, Y., Ding, X., Paisley, J.: Lightweight pyramid networks for image deraining. *IEEE Trans. Neural Netw.* **31**, 1794–1807 (2020)
31. Wang, T., Yang, X., Xu, K., Chen, S., Zhang, Q., Lau, R.W.: Spatial attentive single-image deraining with a high quality real rain dataset. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12270–12279 (2019)
32. Liu, J., Gong, X.: Attention mechanism enhanced LSTM with residual architecture and its application for protein-protein interaction residue pairs prediction. *BMC Bioinform.* **20**, 1–11 (2019)
33. xu qin, zhilin wang, Bai, Y., Xie, X., Jia, H.: Ffa-net: feature fusion attention network for single image dehazing. In: *AAAI 2020 : The Thirty-Fourth AAAI Conference on Artificial Intelligence*, vol. 34, pp. 11908–11915 (2020)
34. Du, Y., Xu, J., Qiu, Q., Zhen, X., Zhang, L.: Variational image deraining. In: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 2406–2415 (2020)
35. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004)
36. Li, S., Ren, W., Zhang, J., Yu, J., Guo, X.: Single image rain removal via a deep decomposition-composition network. *Comput. Vis. Image Underst.* **186**, 48–57 (2019)
37. Kingma, D.P., Ba, J.L.: Adam: a method for stochastic optimization. In: *ICLR 2015: International Conference on Learning Representations 2015* (2015)
38. Wang, H., Li, M., Wu, Y., Zhao, Q., Meng, D.: A survey on rain removal from video and single image. arXiv preprint [arXiv:1909.08326](https://arxiv.org/abs/1909.08326) (2019)