Multi-scale enhancement and aggregation network for single image deraining

Rui Zhang
Shandong University of Finance & Economics
Jinan, China
zhchchao@sina.com

Yuetong Liu
Shandong University of Finance & Economics
Jinan, China
henlyta@163.com

Huijian Han
Shandong University of Finance & Economics
Jinan, China
hanhuijian@sdufe.edu.cn

Yong Zheng
Shandong University of Finance & Economics
Jinan, China
1443066340@qq.com

Tao Zhang
Shandong University
Jinan, China
tao.zhang.sdu@foxmail.com

Yunfeng Zhang
Shandong University of Finance & Economics
Jinan, China
yfzhang@sdufe.edu.cn

Abstract

Rain streaks in an image will appear with different sizes and orientations, resulting in severe blurring and visual quality degradation. Previous CNN-based algorithms achieved encouraging derained results, while there are certain limitations in the description of rain streaks and the restoration of scene structure in different environments. In this paper, we propose an efficient multi-scale enhancement and aggregation network (MEAN) to solve the single image deraining problem. Specifically, considering the importance of large receptive fields and multi-scale features for depicting rain, we introduce a multi-scale enhanced unit (MEU) to capture long-range dependencies and exploit features at different scales. Simultaneously, an attentive aggregation unit (AAU) is designed to utilize the informative features on spacial and channel dimensions, and aggregate effective information to eliminate redundant features for rich scenario details. To improve the deraining performance of the encoder-decoder network, we utilize AAU to filter the information in the encoder network and concatenate useful features to the decoder network, which is conducive to predicting high-quality clean and rain-free images. Experiments results on synthetic datasets and real-world samples show that our method achieves significant deraining performance compared to state-of-the-art approaches.

***Keywords:** Single image deraining, multi-scale enhancement and aggregation, encoder-decoder network*

1. Introduction

Rain is a common weather phenomenon that hinders outdoor monitoring and human visual perception. Due to the presence of rain streaks severely degrades the visibility of objects in the images, many computer vision tasks can be interfered with, such as road surveillance, object detection and tracking, and consumer camera. Therefore, rain removal has always been a fundamental problem in computer vision research.

Removing rain from the images is a very challenging ill-posed problem, as we not only have to eliminate rain streaks completely, but also predict areas that are masked and blurred behind the rain streaks. To address the deraining problem, many methods have been designed to recover the degraded scenes from the rainy images. Traditional methods [17, 2, 13] decompose rainy images into rain streak layers and clear background layers by learning the composition pattern of rain streaks and introducing appropriate prior knowledge, and achieve satisfactory visual effects in light rain scenarios. However, rain streaks in real conditions are much more complex than existing prior assumptions, making these traditional optimization-based approaches limited in the capability to model and remove rain streaks.

Recently, deep learning-based deraining methods [26, 19, 11, 1] have achieved tremendous progress. Most of these approaches attempted to obtain negative residual maps or directly predict the clean images from original rainy inputs by stacking convolutional layers. However, networks based on convolutional layers are only able to model local information, but larger-scale contextual features are inevitably ignored. The dilated convolution is applied by

some methods [6, 8] to expand the receptive field to a certain extent. Despite the multi-scale information captured, this operation is essentially a local feature calculation process, the feature extraction is still limited by the size of the receptive field. In addition, several researchers consider introducing different modules to effectively extract deep features for obtaining high-quality deraining results. [16] and [4] enhance the channel dimensional features by employing squeeze-and-excitation block [9] to improve the feature representation ability of the network. Although these methods demonstrate impressive rain removal performance, the derained results tend to be over-smoothed and blurry details due to the lack of spatial dimension information. Moreover, as the depth of the network deepens, the effective features obtained by shallow layers may be lost. [7] utilizes skip-layer connections to transmit shallow features into subsequent modules, but meaningful shallow features can not be effectively screened and useless features are also inevitably passed.

To address the above limitations, we propose a multi-scale enhancement and aggregation network (MEAN) for single image deraining, which is an ensemble composed of an encoder network and a corresponding decoder network. The multi-scale enhanced unit (MEU) is exploited in the proposed network to obtain efficient features at different scales while capturing long-range dependencies. In addition, the MEU can also fuse the features in the current unit and utilize contextual information to better restore rain-free images with ample details. To effectively extract useful information for deep feature representation, we use an attentive aggregation unit (AAU) to enhance and deliver shallow features of the encoder part into the decoder part. The AAU can adaptively recalibrate features by integrating channel attention and spatial attention, and generate effective feature maps.

In summary, the main contributions of our method are listed below:

(1) We propose a multi-scale enhancement and aggregation network (MEAN) for image rain removal, which utilizes different scale information and effective features to completely remove rain streaks and effectively restore clear rain-free images.

(2) A multi-scale enhanced unit (MEU) is designed to capture multi-scale and global features, which can exploit information at different scales while expanding the receptive field to improve the network deraining performance.

(3) We employ an attentive aggregation unit (AAU) to obtain channel and spatial dimensions features. By applying the AAU, only the valid features of the encoder part can be transmitted to the decoder part for efficient feature representation.

(4) Extensive experimental results demonstrate that the proposed method achieves excellent deraining performance against advanced approaches on both synthetic datasets and real-world images.

2. Related work

In general, single image deraining approaches are classified into two classes: traditional methods and deep learning-based methods. Traditional approaches [17, 2, 13, 3, 14] model rain streaks by designing hand-crafted physical features, or utilize prior knowledge to constrain ill-posed problems. However, previous traditional deraining methods fail in complex rain scenarios, which can easily lead to the degradation of background content.

Recently, deep learning has been introduced for single image deraining and demonstrated impressive performance. For example, [6] designed a deep detail network for focusing on high-frequency rain streaks and removing background interference. [23] modeled the rain binary mask and atmospheric veils to jointly detect and remove rain streaks. [16] proposed a unified deraining network to gradually recover clean images, which incorporated a recurrent neural network to guide the deraining in later stages. [26] used a multi-stream dense network to determine the rain-density information and effectively remove the corresponding rain streaks guided by the density label. [20] introduced an entangled representation learning model that included a two-branched encoder to obtain better derained images. [21] presented a deraining approach that combined the temporal properties of rain with human supervision to obtain rain-free results from a series of nature rainy images. [24] constructed a novel deraining approach named recurrent hierarchy enhancement network for rain removal. [10] formulated a depth-guided attention mechanism and constructed a novel deep neural network to learn effective features for complete rain removal. [5] built a residual-guide feature fusion network that can deal with different rain scenarios and progressively predicts high-quality results. [8] presented the EfficientDeRain method to eliminate rain from images, which can utilize pixel-wise dilation filtering and the kernel prediction network to automatically predict the multi-scale kernels for each pixel. [4] introduced an end-to-end network that consisted of two sub-network with a comprehensive loss function to derain and obtain lost scene details. [12] explored the multi-scale collaborative representation for rain streaks and presented a multi-scale progressive fusion deraining network to completely eliminate rain. [19] deployed a bilateral recurrent network to model the interplay between the rain streak layer and the clean background layer for image deraining. [7] integrated both local and global features into a dual graph convolutional network and exploited multi-dimension contextual information to generate clean derained results. [1] captured features with multi-scale extraction, hierarchical distillation and information aggregation, and proposed an end-to-end frame-

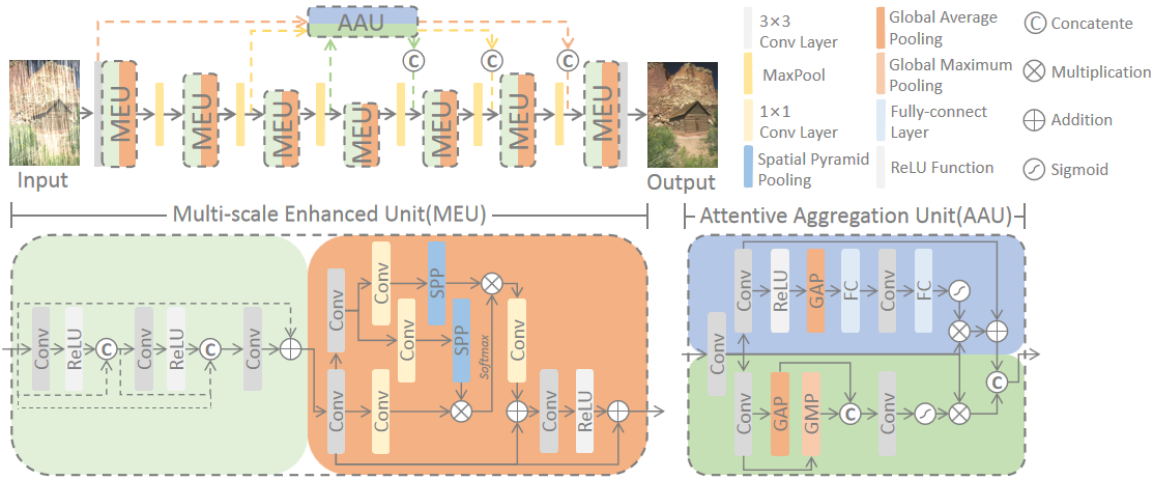


Figure 1. Overview of the proposed MEAN.

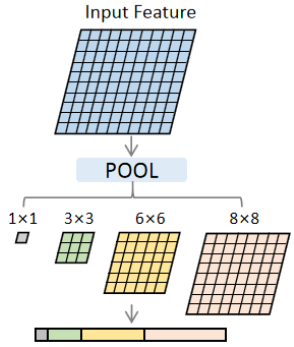


Figure 2. Overview of the spatial pyramid structure. $POOL$ represents the pooling operation. $n \times n$ denotes the output $width \times height$ of the pooling layer.

work to generate rain-free outputs. [11] analyzed the rain and formulated a rainy image model, and designed a novel end-to-end depth-guided network to produce clean images. [25] employed a multi-stage progressive network to restore degraded images, which decomposed the challenging task into multiple sub-tasks to obtain satisfactory outputs.

3. Proposed method

We show the proposed MEAN for single image deraining in Fig. 1. Our network consists of a deep convolution encoder-decoder network that delivers effective features from the encoder network to the decoder network for recovering degraded images. Specifically, the MEAN includes several multi-scale enhanced units (MEU) to capture features at different scales and expand the receptive field of the network. As the core component of the MEAN, the MEU can also integrate deep features and utilize contextual information to improve the rain removal performance of the network. In the MEAN, MEUs are reduced to half the size of the previous unit or expanded to twice the size of the previous part by exploiting $MaxPool$. To fully exploit the valuable information of the encoder network, we use the attentive aggregation unit (AAU) to recalibrate shallow features and remove redundant features, which is beneficial to the decoder network for obtaining high-quality clean outputs. We will introduce the details of the key unit and loss function in the following.

3.1. Multi-scale enhanced unit

To capture multi-scale features to improve the rain removal performance, some methods [8, 4, 7] employ dilated convolution to effectively obtain contextual information. However, rain streaks and objects in the rainy scene tend to be spatially long, and the derained results using only dilated convolutional layers computed in local regions still be limited by the receptive field. The non-local network [22] is proposed to capture long-range dependencies, and non-locally enhanced encoder-decoder network[15] and dual graph convolutional network [7] adopt non-local operation and achieve satisfactory visual effects to a certain extent. Applying a non-local algorithm in the network structure can calculate the response of a location as a weighted sum of the features of all locations to expand the receptive field from a local area to the entire rainy image.

To fully obtain multi-scale information while expanding the receptive field to predict high-quality outputs, we use a multi-scale enhanced unit (MEU) to effectively acquire and enhance features, which fuses features from the current block and integrates them into non-local operation instead of transferring inter-stage by referring to [7]. The MEU consists of two types of blocks, Fig. 1 illustrates the detail. The first block is the densely residual block that can process shallow features to guide subsequent deep feature extraction. This block employs the residual network structure for deep feature transmission and uses a dense network to share

Figure 3. Comparison results of different rain removal approaches on Rain200H. (a)Input (b)DDN [6] (c)RESCAN [16] (d)ReHEN [24] (e)EfNet [8] (f)DRD-Net [4] (g)MPRNet [25] (h)DGCN [7] (i)Our (j)Ground Truth

efficient features with subsequent operations. The second block is an improved non-local block, which integrates a spatial pyramid structure based on the non-local neural network. By applying improved non-local block, the network can fully explore multi-scale information while expanding the receptive field. The regular non-local block can expand the receptive field of the network to some extent, and this block generally involves several important steps. Firstly, the non-local block feeds the features into three 1×1 convolutions separately, which are employed to convert the input features into different embeddings θ, ϕ and g. Secondly, The result of matrix multiplication of θ and ϕ is regularized, and matrix multiplication with g. Finally, the output of the previous step is fed into a 1×1 convolution and add with the input features of the non-local block to obtain the output features of the non-local operation. The conventional non-local block in deep learning-based deraining approaches is generally explained as:

$$y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) g(x_j) \tag{1}$$

$$f(x_i, x_j) = e^{\theta(x_i)^T \phi(x_j)} \tag{2}$$

where i is the index to calculate the output location, and j is the index to enumerate all possible locations. x and y denote the input feature and the output, respectively. The

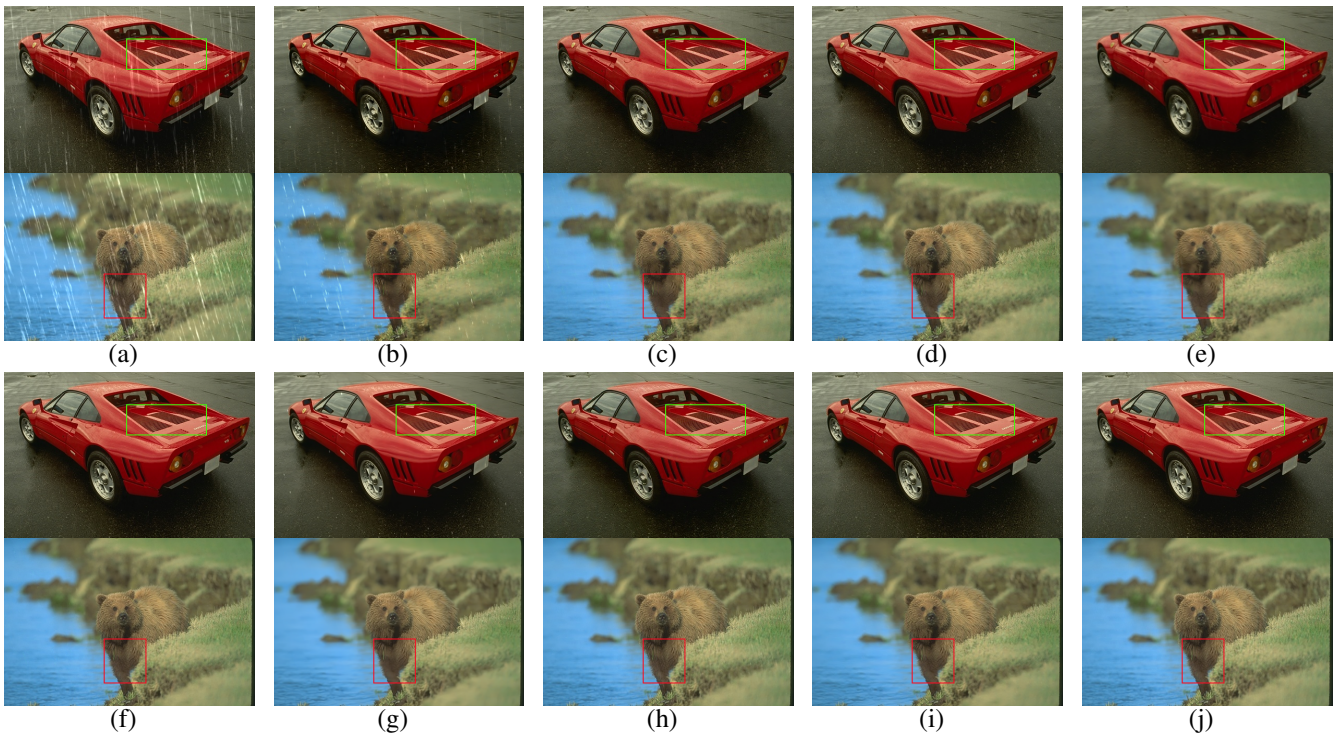


Figure 4. Comparison results of different rain removal approaches on Rain200L. (a)Input (b)DDN [6] (c)RESCAN [16] (d)ReHEN [24] (e)EfNet [8] (f)DRD-Net [4] (g)MPRNet [25] (h)DGCN [7] (i)Our (j)Ground Truth

response is normalized by the factor $C(x)$. For simplicity, non-local operation adopts g in the form of a linear embedding: $g(x_j) = W_g x_j$. $\theta(x_i) = W_\theta x_i$ and $\phi(x_j) = W_\phi x_j$. W_g, W_θ and W_ϕ are weights to be learned.

We embed a spatial pyramid structure in a non-local network to utilize multi-scale features effectively. The spatial pyramid pooling can exploit context aggregation based on different regions to capture contextual information. The specific implementation of the spatial pyramid operation is shown in Fig. 2. By applying spatial pyramid pooling, the input features of this part are transformed into the vectors, and the feature map is formed by multiplying vectors corresponding to θ and ϕ of the regular non-local block, multiplying them by improved g, and a 1×1 convolution. Compared with the traditional non-local network, MEU processed features through a pooling layer after ϕ and g, and concatenated the four results as the input for subsequent operations. By using spatial pyramid pooling in the non-local network, sufficient global information and features at different scales were integrated, which is beneficial to improve the deraining performance for satisfactory rain-free outputs.

3.2. Attentive aggregation unit

Using the information recorded of the encoder part in the decoder part proves to be beneficial for the removal of rain streaks in the degraded image [15]. Nonetheless, useless features inevitably appear in the shallow features of the encoder, and the direct employ of shallow features may lead to residual rain streaks or artifacts in the restored image. Based on the above motivations, we utilize an attentive aggregation unit (AAU) to capture meaningful features in the encoder stage and concatenate them into the decoder stage for efficient feature enhancement and redundant feature removal. We feed the input features of this unit into channel attention and spatial attention respectively, and finally aggregate the outputs. The AAU encourages the network to extract useful information for recovering abundant details in the background scene. Channel attention and spatial attention are paralleled in AAU, they can achieve feature redirection and generate effective attention maps. By integrating channel and spatial useful information, the meaningful features in the encoder network can be fully employed in the decoder network to improve the network learning ability.

3.3. Loss function

Generally, the output of the proposed network ought to be similar to the ground truth at a certain level. To obtain high-quality derained results, a hybrid loss function is used to train our method, including the structural similarity (SSIM) [15] loss and the L_1-norm loss. The SSIM is a metric that can effectively evaluate image quality. It measures the similarity between two images by including three

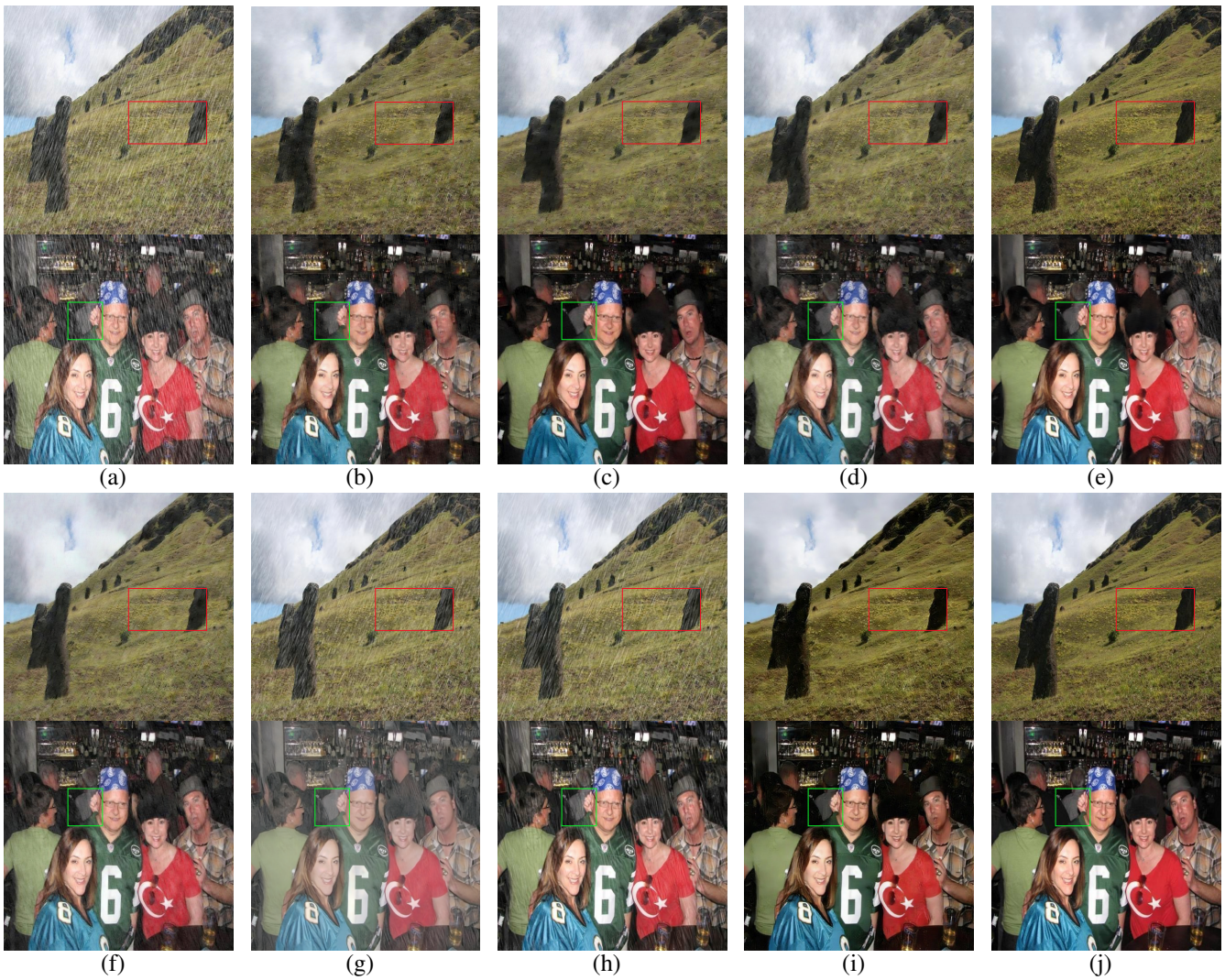


Figure 5. Comparison results of different rain removal approaches on Rain1200. (a)Input (b)DDN [6] (c)RESCAN [16] (d)ReHEN [24] (e)EfNet [8] (f)DRD-Net [4] (g)MPRNet [25] (h)DGCN [7] (i)Our (j)Ground Truth

aspects including structure, contrast and brightness. To better preserve background details of the derained outputs, we adopt the SSIM loss to supervise the proposed MEAN.

$$L_{SSIM} = -log(f_{ssim}(O, GT) + \xi_1) \tag{3}$$

where L_{SSIM} represents SSIM loss. O indicates the results of deraining network O, and GT indicates the ground truth image. f_{ssim} denotes the SSIM value between O and GT. ξ_1 indicates an extremely small number to avoid having the denominator equal to zero.

Furthermore, we also use the L_1-norm loss, which can adequately constrain the differences between brightness and color properties.

$$L_1 = \frac{1}{N}\sum_{i=1}^{N}(\parallel O - GT \parallel) \tag{4}$$

where L_1 represents L_1-norm loss. N denotes the total number of pixels in the derained output. O and GT indicate the results of deraining network and corresponding the ground truth image, respectively.

The final hybrid loss function can be formulated as:

$$L = L_{SSIM} + \lambda L_1 \tag{5}$$

where λ represents the weight parameter of L_1 loss, empirically set equal to 0.1.

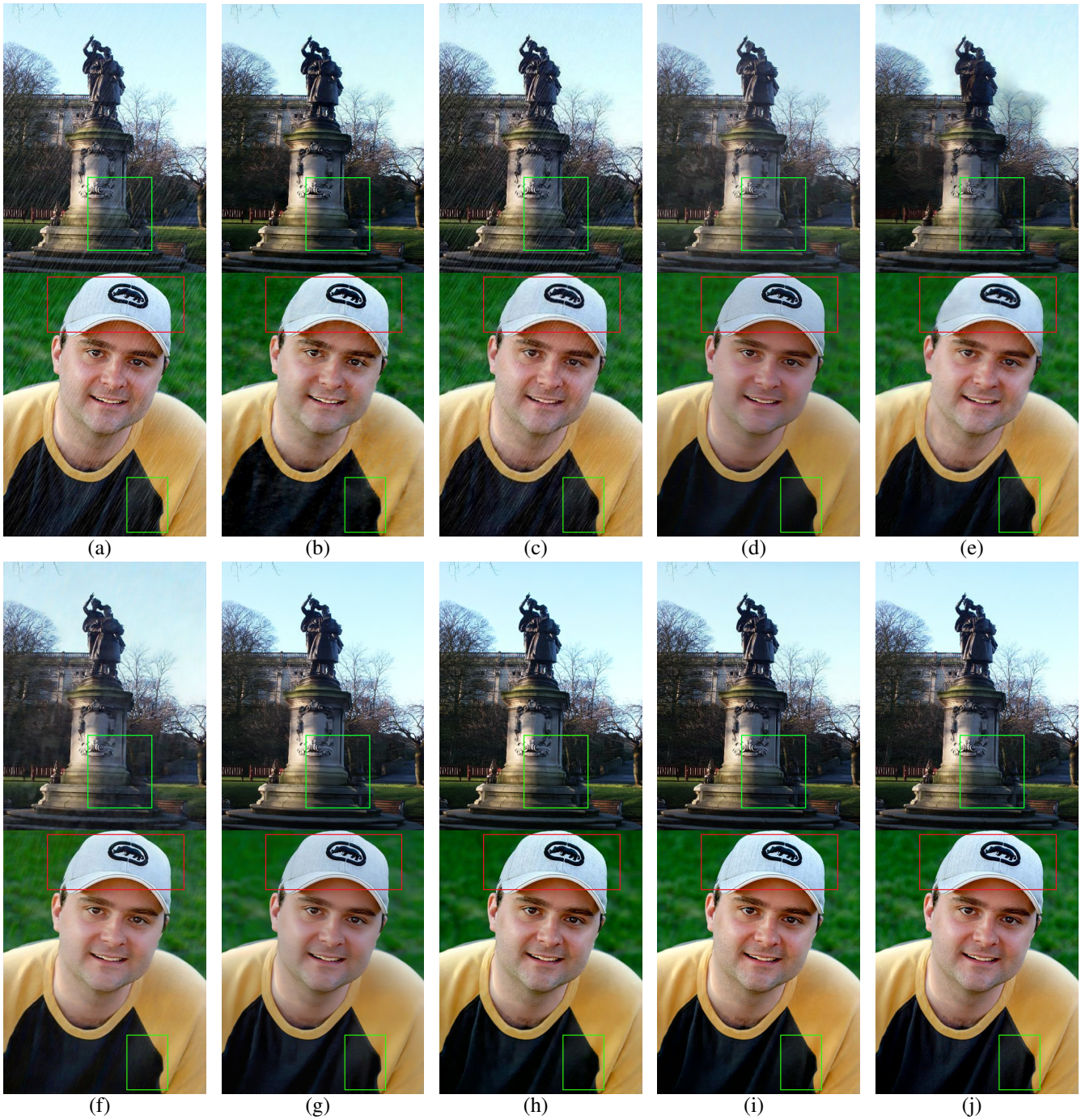


Figure 6. Comparison results of different rain removal approaches on Rain1400. (a)Input (b)DDN [6] (c)RESCAN [16] (d)ReHEN [24] (e)EfNet [8] (f)DRD-Net [4] (g)MPRNet [25] (h)DGCN [7] (i)Our (j)Ground Truth

4. Experimental results

We compare our network with seven state-of-the-art approaches: deep detail network [6] (denotes as DDN), recurrent squeeze-and-excitation context aggregation net [16] (denoted as RESCAN), recurrent hierarchy enhancement network [24] (denoted as ReHEN), EfficientDerain network [8] (denotes as EfNet), detail-recovery image deraining network [4] (denoted as DRD-Net), dual graph convolutional network [7] (denoted as DGCN), and multi-stage progressive image restoration network [25] (denoted as MPRNet).

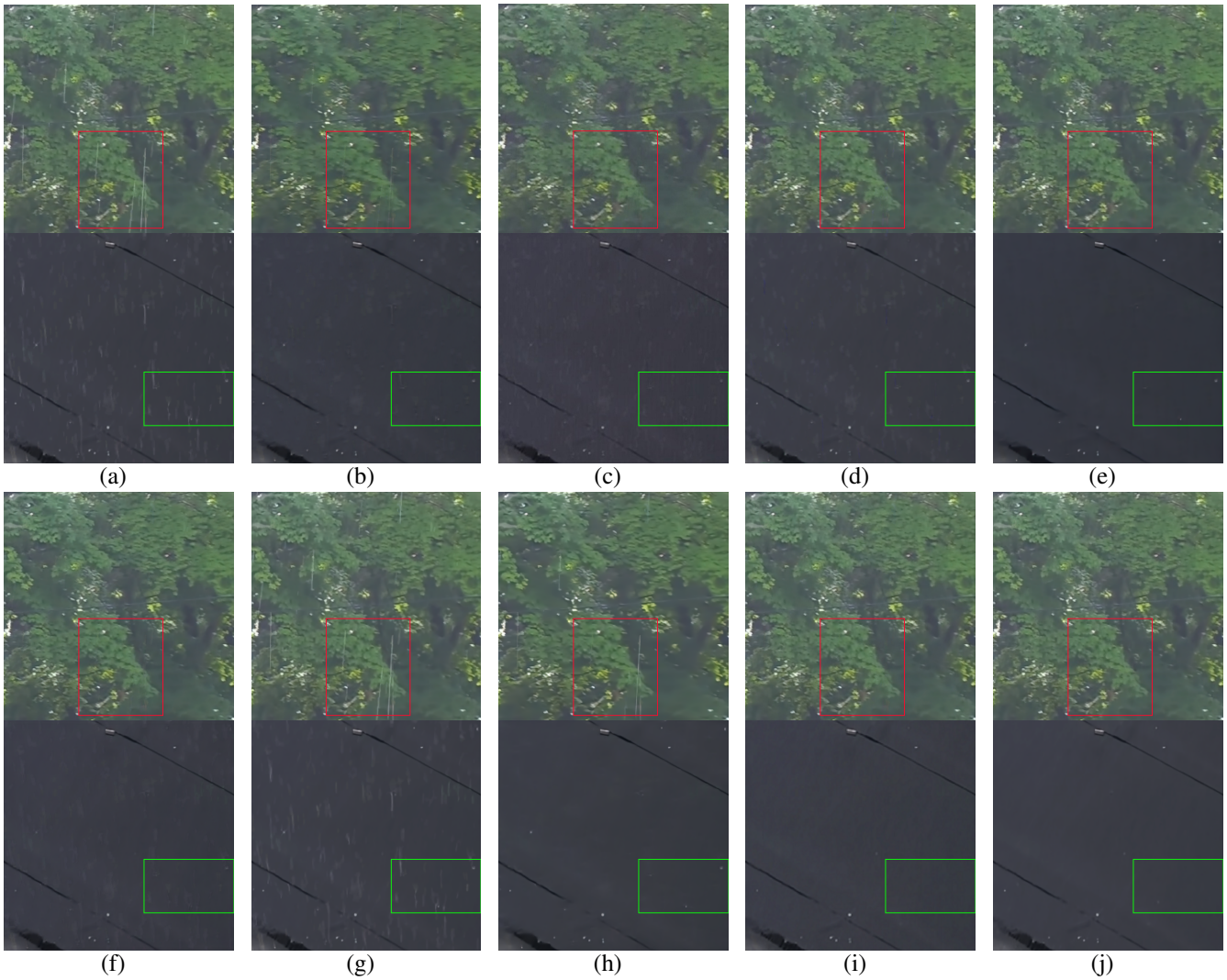


Figure 7. Comparison results of different rain removal approaches on SPA-Data. (a)Input (b)DDN [6] (c)RESCAN [16] (d)ReHEN [24] (e)EfNet [8] (f)DRD-Net [4] (g)MPRNet [25] (h)DGCN [7] (i)Our (j)Ground Truth

Datasets	Rain200H	Rain200L	Rain1200	Rain1400	SPA-Data
Method	PSNR/SSIM				
DDN[6]	24.64/0.850	33.01/0.972	30.97/0.912	30.00/0.904	36.16/0.946
RESCAN[16]	26.60/0.897	37.07/0.987	33.38/0.942	31.94/0.935	38.11/0.971
ReHEN[24]	27.88/0.850	37.26/0.972	32.64/0.914	31.33/0.918	37.99/0.966
EfNet[8]	32.34/0.908	37.10/0.986	34.79/0.970	32.55/0.910	41.27/0.983
DRD-Net[4]	28.16/0.920	37.15/0.987	29.93/0.882	32.57/0.939	38.39/0.972
DGCN[7]	31.09/0.910	37.36/0.988	34.41/0.963	33.07/0.945	41.98/0.989
MPRNet[25]	30.37/0.885	36.59/0.954	32.76/0.918	33.27/0.929	41.30/0.985
MEAN	**32.40/0.926**	**37.39/0.990**	**34.81/0.974**	**33.39/0.951**	**42.17/0.991**

Table 1. Quantitative results of different methods on synthetic datasets.

4.1. Implementation details

The proposed method is trained on the NVIDIA GTX 2080 Ti GPUs based on PyTorch. We employ the Adam

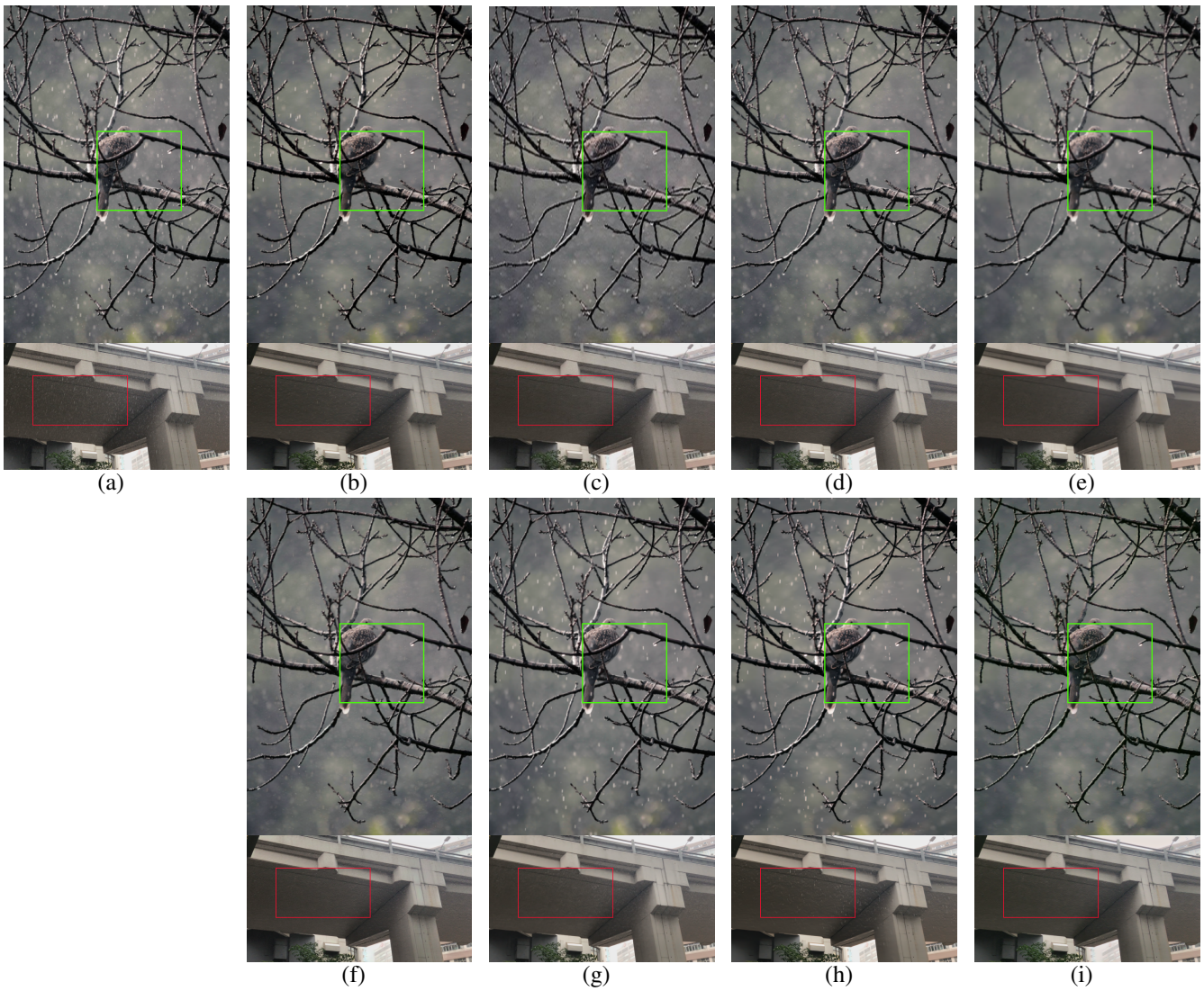


Figure 8. Comparison results of different rain removal approaches on real-world images. (a)Input (b)DDN [6] (c)RESCAN [16] (d)ReHEN [24] (e)EfNet [8] (f)DRD-Net [4] (g)MPRNet [25] (h)DGCN [7] (i)Our

algorithm to update the parameters of our network during training. The initial learning rate is 0.001, and the rate will decay by 0.2 when the epoch number reaches 20 and 40. The training of the network will terminate after 80 epochs.

4.2. Synthetic data

We conduct extensive experiments on five synthetic datasets to demonstrate the deraining performance of our method. Rain200H (with heavy rain streaks) and Rain200L (with light rain streaks) [23] contain 1800 synthetic training images and 200 testing images, respectively. Rain200H/L are synthesized from BSD200 [18] with five kinds of rain streaks. Rain1200 [26] comprises 12000 rainy/clean image pairs, which are composed of three rain density levels (light, medium and heavy rain streaks). Rain1400 [6] contains 14000 synthetic image pairs that have 14 types of different sizes and directions of rain streaks. SPA-Data [21] provides a large-scale dataset using 170 real-world rainy videos, which contains 84 different scenarios. Labels of each dataset are not employed for the network training.

Figs. 3−7 illustrates the rain removal performance of competitive methods on five synthetic rainy datasets. As shown, the derained outputs of the proposed approach are better than other methods in completely eliminating rain and efficiently restoring the rich scenario textures. By checking the part in the box, it can be seen that other methods have also removed or blurred details like rain streaks, such as grass in Fig. 5. ReHEN only uses useful infor-

mation on the channel dimension, while RESCAN, EFF, and DRD-Net only exploit the dilated convolutional layer to obtain limited receptive fields, and they all have varying degrees of rain streaks residual. MPRNet and DGNL cannot effectively combine information of different scales and global features, and they also leave rain streaks and lose some scene details in the derained image. By contrast, our method can enlarge the receptive field and capture effective features at multiple scales to completely eliminate rain from rainy inputs by applying MEUs. The AAU is utilized to achieve feature recalculation and obtain a valid attention map for generating richer background details. Comparatively, the images recovered by the proposed method are closer to the ground truth and obtain high-quality rain-free images.

We compute the peak-signal-to-noise ratio (PSNR) and structure similarity (SSIM) for the quantitative evaluation. Especially, the higher PSNR and SSIM values indicate that the network has the strong deraining ability and the outputs have clean details. Table 1 reports the quantitative results of the state-of-the-art methods and the proposed MEAN. The proposed method utilizes appropriate hybrid loss to ensure high-quality rain removal results. The SSIM loss guarantee that the rain removal outputs remain rich details, and the L_1-norm loss can confine the difference between brightness and color attributes. By using the hybrid loss function, the proposed network can obtain better derained results. Our method achieves the best quantitative results on five synthetic rain datasets, which shows that the proposed method can completely remove rain streaks while preserving more scene details. Rain streaks of different sizes and orientations are contained in these rainy datasets, and the higher PSNR and SSIM values confirm the generality and robustness of the MEAN.

4.3. Real-world data

To verify the generalization of the MEAN, we qualitatively compare our method with other advanced approaches on real-world images and show several examples in Fig. 8. For the first image, especially in the cropped box, our method can sufficiently remove rain streaks while other methods have varying degrees of rain residue. Our method is capable of aggregating information at different scales and obtaining long distance dependence to eliminate rain streaks of various sizes in the rainy image, and utilizes channel and spatial attention mechanisms to maintain tiny background structure in the image. The other advanced deraining methods can not efficiently employ spatial contextual information and capture informative features, thus they predict the output with rain steaks and artifacts. For the second image, other methods hand down a number of artifacts, while our method can generate better outputs. Based on MEUs, our method exploits AAU to integrate effective features and apply meaningful information at the encoder stage to obtain satisfactory outputs. The above examples illustrate our proposed method is an effective rain removal algorithm that can restore high-quality derained images and also can preserve plentiful details and background texture information.

Table 2. NIQE comparison on real-world samples.

Method	NIQE
DDN[6]	4.09
RESCAN[16]	3.92
ReHEN[24]	3.67
EfNet[8]	3.39
DRD-Net[4]	3.65
DGCN[7]	3.61
MPRNet[25]	3.50
MEAN	3.31

Furthermore, we calculate the naturalness image quality evaluator (NIQE) on real-world datasets and the results are shown in Table 2. We observe that our method has a lower NIQE value, which reveals that the MEAN can obtain more satisfactory and natural results. This also demonstrates that our method is an excellent derainer than others on real-world images.

4.4. Ablation study

For the proposed units, we perform ablation studies to demonstrate the effectiveness and necessity of our method. In particular, we use the densely residual block in a multi-scale enhanced unit (MEU) as the baseline. For different units, their abbreviations are as follows:

•R_1 : Only the densely residual block in the MEU is used as the baseline, and the multi-scale information and global features are not utilized, and the decoder part is not connected to the encoder part.

•R_2 : Only the full MEU is used, and the effective information from the encoder network is not concatenated and exploited by the decoder network.

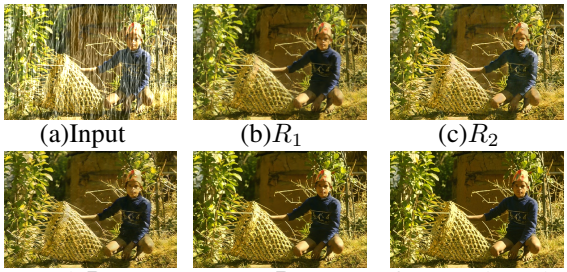
•R_3 : Use the unbroken MEU, concatenate encoder stage information in the decoder stage, but do not use the attentive aggregation unit (AAU).

•R_4 : Both MEU and AAU are applied, i.e., our proposed final network.

The quantitative results of different unit combinations are shown in Table 3. We can see that applying multi-scale information and long-range dependencies is beneficial to improve the network rain removal performance compared to the baseline. Connecting the features of the encoder network to the decoder network can upgrade the capability to learn the features to a certain extent. However, the interference of redundant features may lead to further transmission of useless information, leaving rain streaks or artifacts in

Units	MEU		Concat	AAU	PSNR/SSIM
	DRB	INB			
R_1	✓				31.80/0.907
R_2	✓	✓			31.95/0.914
R_3	✓	✓	✓		32.33/0.920
R_4	✓	✓	✓	✓	**32.40/0.926**

Table 3. The results of different units on Rain200H. MEU denotes the multi-scale enhanced unit. DRB and INB represent the densely residual block and improved non-local block, respectively. AAU indicates the attentive aggregation unit.



(a)Input (b)R_1 (c)R_2 (d)R_3 (e)R_4 (f)Ground Truth

Figure 9. Visual comparisons of key component effects in ablation studies.

the image. Fig. 9 shows the derained results of the different combinations of units and operations. It can be seen that the derained results using multi-scale features and global information can remove all rain streaks. By applying the AAU, abundant scene detail can be preserved in the outputs. Compared with several other unit combinations, the deraining performance of the network incorporating MEU and AAU is promoted to the greatest extent and can effectively predict clear images.

Moreover, we conducted the ablation study with different hybrid loss functions. Interestingly, we found that the proposed approach displays better rain removal performance in certain fixed scenarios by adding perceptual loss. The quantitative results of the hybrid loss function are exhibited in Table 4. With the addition of perceptual loss, the quantitative results of rain removal output were slightly improved. Fig. 10 displays the visual difference between the two hybrid loss functions. By observing the derained results, it can be found that after the addition of perceptual loss, some details similar to the rain streaks are retained, but some rain streaks are inevitably left in the output. In summary, the addition of perceptual loss was more helpful in improving quantitative results and promoting the retention of minute details. However, for the complete removal of rain streaks, $L = L_{SSIM} + \lambda L_1$ is more advantageous. Different hybrid loss functions can be applied to various rainy scenarios under different recovery criteria.

Table 4. Quantitative results of different hybrid loss functions on Rain200H. λ and μ are set to 0.1 and 0.05, respectively.

Loss	PSNR/SSIM
$L_a = L_{SSIM} + \lambda L_1$	32.40/0.926
$L_b = L_{SSIM} + \lambda L_1 + \mu L_{perceptual}$	32.43/0.930

(a)Input (b)L_a (c)L_b

Figure 10. Visual comparisons of different hybrid loss function.

5. Conclusion

In this paper, we have introduced a multi-scale enhancement and aggregation network (MEAN) that is based on the encoder-decoder structure for single image deraining. To effectively capture long-range dependencies and multi-scale information, we apply a multi-scale enhanced unit to improve the network rain removal performance. Moreover, we concatenate meaningful information from the two parts of the encoder network and the decoder network for efficient feature transfer. By utilizing the attentive aggregation unit (AAU), we further remove redundant features of the encoder part to better help predict clear derained outputs. Quantitative and qualitative results reveal that our network outperforms other comparative rain removal approaches on both synthetic and natural rainy datasets.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 61972227); the Natural Science Foundation of Shandong Province (No. ZR201808160102); Shandong Provincial Natural Science Foundation Key Project (No. ZR2020KF015); the Key Research and Development Project of Shandong Province (No. 2019GSF109112); the Science and Technology Plan for young talents in Colleges and Universities of Shandong Province (No. 2020KJN007); the Scientific Research Studio in Colleges and Universities of ji'nan City (No. 2021GXRC092); the Science and Technology Research Program for Colleges and Universities in Shandong Province (No. KJ2018BZN029).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author' contributions

Rui Zhang: Methodology, Investigation, Writing - review and editing. Yuetong Liu: Software, Writing - original

draft, Writing - review and editing. Huijian Han: Conceptualization, Data curating, Formal analysis. Yong Zheng: Conceptualization, Resources, Investigation. Tao Zhang: Methodology, Visualization, Writing - review and editing. Yunfeng Zhang: Supervision, Data curating, Validation.

References

[1] X. Chen, Y. Huang, and L. Xu. Multi-scale hourglass hierarchical fusion network for single image deraining. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 872–879, 2021. 1, 2

[2] Y. Chen and C. Hsu. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1968–1975, 2013. 1, 2

[3] L. Deng, T. Huang, X. Zhao, and T. Jiang. A directional global sparse model for single image rain removal. *Appl. Math. Model.*, 59:662–679, 2018. 2

[4] S. Deng, M. Wei, J. Wang, Y. Feng, L. Liang, H. Xie, F. Wang, and M. Wang. Detail-recovery image deraining via context aggregation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 14560–14569, 2020. 2, 3, 4, 5, 6, 7, 8, 9, 10

[5] Z. Fan, H. Wu, X. Fu, Y. Huang, and X. Ding. Residual-guide network for single image deraining. In *Proceedings of the ACM International Conference on Multimedia*, page 1751–1759, 2018. 2

[6] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3855–3863, 2017. 2, 4, 5, 6, 7, 8, 9, 10

[7] X. Fu, Q. Qi, Z. Zha, Y. Zhu, and X. Ding. Rain streak removal via dual graph convolutional network. In *Proceedings of the the AAAI Conference on Artificial Intelligence*, pages 1–9, 2021. 2, 3, 4, 5, 6, 7, 8, 9, 10

[8] Q. Guo, J. Sun, F. Juefei-Xu, L. Ma, X. Xie, W. Feng, and Y. Liu. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining. *arXiv preprint arXiv:2009.09238*, 2020. 2, 3, 4, 5, 6, 7, 8, 9, 10

[9] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 7132–7141, 2018. 2

[10] X. Hu, C. Fu, L. Zhu, and P. Heng. Depth-attentional features for single-image rain removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8014–8023, 2019. 2

[11] X. Hu, L. Zhu, T. Wang, C. W. Fu, and P. A. Heng. Single-image real-time rain removal based on depth-guided non-local features. *IEEE Trans. Image Process.*, 30:1759–1770, 2021. 1, 3

[12] K. Jiang, Z. Wang, P. Yi, C. Chen, B. Huang, Y. Luo, J. Ma, and J. Jiang. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8343–8352, 2020. 2

[13] L. Kang, C. Lin, and Y. Fu. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Trans. Image Process.*, 21(4):1742–1755, 2011. 1, 2

[14] J. Kim, C. Lee, J. Sim, and C. Kim. Single image deraining using an adaptive nonlocal means filter. In *Proceeding of the IEEE International Conference on Image Processing*, pages 914–917, 2013. 2

[15] G. Li, X. He, W. Zhang, H. Chang, L. Dong, and L. Lin. Non-locally enhanced encoder-decoder network for single image de-raining. In *Proceedings of the ACM International Conference on Multimedia*, pages 1056–1064, 2018. 3, 5

[16] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European conference on computer vision*, pages 254–269, 2018. 2, 4, 5, 6, 7, 8, 9, 10

[17] Y. Li, R. Tan, X. Guo, J. Lu, and M. Brown. Rain streak removal using layer priors. In *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2736–2744, 2016. 1, 2

[18] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2, pages 416–423, 2001. 9

[19] D. Ren, W. Shang, P. Zhu, Q. Hu, D. Meng, and W. Zuo. Single image deraining using bilateral recurrent network. *IEEE Trans. Image Process.*, 29:6852–6863, 2020. 1, 2

[20] G. Wang, C. Sun, and A. Sowmya. Erl-net: Entangled representation learning for single image de-raining. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5644–5652, 2019. 2

[21] T. Wang, X. Yang, K. Xu, S. Chen, Q. Zhang, and R. Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12262–12271, 2019. 2, 9

[22] X. Wang, R. Girshick, A. Gupta, and K. He. Non-local neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7794–7803, 2018. 3

[23] W. Yang, R. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, July 2017. 2, 9

[24] Y. Yang and H. Lu. Single image deraining via recurrent hierarchy enhancement network. In *Proceedings of the ACM International Conference on Multimedia*, pages 1814–1822, 2019. 2, 4, 5, 6, 7, 8, 9, 10

[25] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M. H. Yang, and L. Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 14821–14831, 2021. 3, 4, 5, 6, 7, 8, 9, 10

[26] H. Zhang and V. Patel. Density-aware single image deraining using a multi-stream dense network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 695–704, 2018. 1, 2, 9