ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)

https://doi.org/10.17559/TV-20240420001475
Original scientific paper

RSPNet: Rain Streak Prior Based Rain Removal Network

Aoran ZHAO

Abstract: Single image rain removal remains challenging due to the complex nature of rain streaks and their interaction with background textures. This study proposes RSPNet, a novel rain removal network that leverages high-frequency information as a rain streak prior. By incorporating a Butterworth filter-derived guidance map, RSPNet enhances rain streak features while preserving background details. The network architecture includes specialized modules for high-frequency feature enhancement, channel feature enhancement, multi-scale feature extraction, and feature refinement. Experiments on synthetic datasets (Rain100H, Rain100L, Rain200H, Rain200L, Rain12) and real-world rainy images demonstrate RSPNet's superior performance over state-of-the-art methods. Notably, RSPNet achieves the highest PSNR and SSIM scores across all tested datasets, with significant improvements in visual quality and detail preservation. The proposed method also shows enhanced performance in downstream tasks such as object detection and image segmentation on rain-affected images.

Keywords: feature enhancement; high-frequency guidance; rain removal; residual block

1 INTRODUCTION

Image rain removal aims to restore clean images from rainy scenes, which is crucial for many outdoor computer vision tasks [1, 3], such as autonomous driving, video surveillance, and target detection. The rain streak information in rainy images can seriously reduce the quality of the image, making it difficult to meet the requirements of high-level computer vision tasks. Hence, in recent years, the rain removal task, as an important preprocessing in high-level computer vision tasks, has received widespread attention. However, since the rain removal task is an ill-posed problem, it is still a challenge. Image rain removal can be expressed as a layer decomposition problem [4], which decomposes the rainy image into a rain streaks layer and a background layer containing the real scene content. It can be expressed by the following formula:

$$O = R + B \tag{1}$$

where O denotes the rainy image, R is the rain layer, and B represents the background layer. The rain removal method based on the above formula aims to learn the rain layer R, and then subtract it from the rainy image O to obtain the background layer, that is, the rainless image.

In the past few decades, many rain removal methods based on traditional theories have been studied. For example, Kang et al. [5] proposed a decomposition-based rain removal method based on the consideration of high-frequency components including rain streaks and background textures. This method decomposes the rainy image into high-frequency and low-frequency components, and then uses dictionary learning and sparse encoding to further extract rain components from the high-frequency components. After that, image rain removal has received more and more attention. Luo et al. [6] proposed an image rain removal method based on sparse coding. This method uses sparse coding to separate the rain layer and background layer from the rainy image. In addition, Li et al. [7] proposed a prior rain removal method based on the Gaussian mixture model, which utilizes simple prior information to process background layer and rain layer，and can be used to the removal of rain streaks with different directions and scales. Although traditional methods have achieved good rain removal results, their performance heavily depends on handcrafted prior knowledge.

In recent years, due to the powerful feature extraction ability of Convolutional Neural Networks (CNNs), deep learning-based removal methods have been widely studied and made great progress. Yang et al. [8] proposed a joint network for rain streak detection and removal. Since then, more and more researchers have applied deep learning to the field of image rain removal [9-13], and have conducted in-depth research on this basis. Fan et al. [14] proposed a cascade network, which guides the learning of deep feature extraction blocks through residual information obtained from shallow feature extraction blocks, and obtains residual estimates from coarse to fine. Ren et al. [15] proposed a progressive rain removal network by introducing circular layers, which utilizes the cross-stage dependency of depth characteristics to gradually remove rain streaks. Wang et al. [16] proposed a JDNet containing the Scale-Aggregation and Self-Attention modules to solve the single image rain removal problem. Wang et al. [17] developed intrinsic loss functions by exploiting the intrinsic priors of rainy images to facilitate training rain removal networks, decomposing a rainy image into a rain-free background layer and a rainy layer. However, due to the difficulty in obtaining pairs of rain images and real rain free images for network training, most networks are trained using synthesized rain-rainless image pairs as training data. Due to the diversity and complexity of rain streak information, most networks currently have poor generalization ability, making it difficult to adapt to rainy images in real environments. To this end, semi-supervised methods of rain removal have been proposed. While using synthetic rainy images and corresponding clean background images for network training, real rainy images are further used for unsupervised training of the network, enabling it to adapt to diverse rainy types and improve the generalization ability of network [18]. In response to this work, the network [19] for generating rainy images are proposed to expand and enrich existing datasets, which enables the proposed model to extract more complex rainfall distributions, thereby improving the performance of rain removal. However, due to the complex degradation factors in rainy images, it is difficult to obtain a simulated dataset with the same distribution as real rainy images. Although deep learning has achieved significant progress in image rain removal, there are still many challenges. For example, there is a high degree of overlap between rain streaks and background features, which makes it difficult for most methods to separate rain streaks with sufficient accuracy. Therefore, the rain removal results obtained by most methods have the problem of residual rain streaks or loss of background information.

Currently, most CNN based methods adopt an end-to-end learning approach, which often suffer from the loss of texture information or residual rain streaks due to the lack of consideration for the confusion between rain streaks and other high-frequency features. Regarding this issue, we analysed the rain streak information and high-frequency information in a rainy image, as shown in Fig. 1.

According to Eq. (1), a rain streak image in Fig. 1c can be obtained by subtracting the background image from the synthesized rainy image. Fig. 1d is the high-frequency image obtained by employing a Butterworth high pass filter on the rainy image. Fig. 1e is the high-frequency residual image by subtracting the high-frequency image from the rain streak image. From Fig. 1d and e, we can observe that the high-frequency image contains a large amount of rain streak information and a small amount of background texture information.

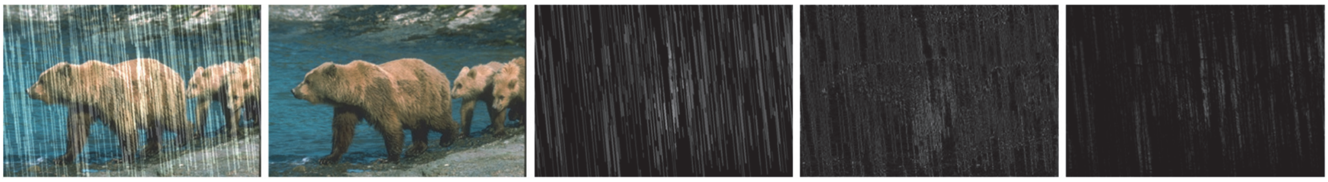


Figure 1 Rainy image and high frequency images

In the rain removal task, if the rain streak features are directly removed as high-frequency features, some background information may be lost while removing the rain streak features. If background texture information is to be retained, some rain streaks may remain in the rain removal results. Therefore, considering the presence of rain streak information and other texture information in high-frequency components, we construct a high-frequency guidance image in the network to guide the network to learn more accurate rain streak information while supplementing lost background texture information, which helps to reduce residual rain streaks and suppress edge blurring in the rain removal results.

Based on the above analysis, this paper proposes a rain streak prior based image rain removal network (RSPNet), which uses high-frequency images obtained from rainy images as rain steak prior information to help reconstruct more accurate rainless images. In RSPNet, first, a high-frequency feature enhancement module (HFEM) is designed to enhance the rain streak features by using the high-frequency component, which makes the subsequent network pay more attention to the rain streak areas. Then, a deep feature extraction module (DFEM) containing a channel feature enhancement module (CFEM), multi-scale feature extraction module (MFEM) and a feature refinement module (FRM) is constructed to further learn and refine rain streak features. CFEM can capture the channel context information, which is inspired by the image reflection removal task [20], and the channel attention module [21] is introduced into the residual structure to enhance important features in the channel dimension. In this paper, we construct the CFEM by cascading multiple residual structure with the channel attention. In addition, considering the appearance model of the raindrops, the raindrops also have a reflection effect [22]. In order to extract the spatial features of rain streaks from different scales, we construct a MFEM by using the pyramid pooling operation [23]. In MFEM, the feature maps are downsampled to different scales, and then the rain streak features at different scales are extracted by using different convolution receptive fields. Finally, a FRM is designed to further extract accurate rain streaks. The main contributions of this work over current methods can be summarized as follows:

1) A RSPNet is presented for image rain removal tasks, which utilizes high-frequency images from rainy images as guiding features to reduce residual rain streaks and loss of edge textures in the rain removal results.

2) To enhance the rain streak features, we propose an HFEM, which utilizes the high-frequency image from the rainy image as a prior map to guide the extraction of rain streak features.

3) To obtain accurate rain streak features, a DFEM including CFEM, MFEM, and FRM is constructed to achieve fine-grained learning of rain streak features at the channel dimension and multiple spatial scales.

The rest of this paper is organized as follows. In Section 2, we review the related work of image rain removal. Section 3 introduces the structure of RSPNet in detail. In Section 4, we present extensive experiments and corresponding analysis. The conclusion is summarized in Section 5.

2 RELATED WORK

In this section, we briefly review the basic modules used in the proposed network. The architecture of the traditional network is introduced first, followed by the multiscale structure, and finally, the attention mechanism is introduced.

2.1 Priori-Based Approach

Traditional methods generate prior information in the rain and background layers and optimize it by constructing cost functions. Many researchers utilize prior information from clean images and rainy images to develop rain removal algorithms [6]. Li et al. [7] proposed simple prior information to restore rainy images by constructing a Gaussian mixture model. Zhu et al. [45] defined a priori information based on a sparse representation and an estimation of the direction of rain in the background layer and constructed a priori similarity between the rain streaks and the extracted ones in the rain layer. The rain layer and background layer in the rainy image are separated by an optimization function. However, these methods rely too much on handcrafted prior information, and cannot achieve good rain removal effects in complex rainy environments.

2.2 Multiscale Structure

The pooling of pyramids proposed by He et al. [22] enables the network to capture information from different perception fields at different scales, thus improving the algorithm performance. Fu et al. [24] constructed a lightweight rain removal network by utilizing the Gaussian Laplace operator for multi-scale decomposition. Jiang et al. [25] proposed a multi-scale progressive fusion rain removal network by considering that the rain streaks affect the image with different resolutions and fuzzy degree. This network uses multiscale cooperation to represent rain streak features of different scales and fuses them. Chen et al. [26] designed a multi-scale hourglass extraction block in order to make full use of the multiscale rainfall characteristics. Yang et al. [27] also used multiscale aggregation to extract multiscale rain streak features. A large number of previous studies have shown that due to the complexity and diversity of rain streak density and direction, the multiscale structure is very effective in image rain removal task. Through different perception fields, the network can better detect and extract rain streak features.

2.3 Attention Mechanism

Recently, attention mechanisms have been widely used in deep learning. Channel attention and spatial attention are common. Channel attention focuses on "what" is meaningful in each input image. Spatial attention focuses on "where" is an informative part [28]. In the method proposed by Wang et al. [29], a self-attention mechanism is introduced to enhance the adaptability of rain removal models. In the paper of Shao et al. [30], the image is fused by using channel attention, to reduce the redundancy of fusion features and make the network pay more attention to channel features. To learn the global dependencies of features, Transformer models with multi head attention have been widely applied in the field of computer vision. Swin Transformer [31] has made significant progress in image classification, object detection, and semantic segmentation by integrating reasonable prior knowledge. In response to the issue that Swin Transformer may not capture global relationships when processing high-resolution images, Xiao et al. [32] proposed a dual Transformer based on window and space to effectively extract local and global information. Chen et al. [33] proposed a DRSformer, which utilize an expert system and Top-k sparse attention to optimize the Transformer structure. The Transformer based method still results in residual rain streaks due to neglecting the learning of local features. Recently, Chen et al. [34] proposed a hybrid CNN-Transformer feature fusion for single image deraining by combining CNN and Transformer. In summary, it can be concluded that the attention mechanism is also very meaningful for image rain removal, which can make the network pay more attention to the important features in the image.

2.4 High Pass Filtering

Image high pass filtering technology [35] can be divided into high pass filtering in the spatial domain and high pass filtering in the frequency domain. High pass filtering in the spatial domain can highlight details and edges in images, but is sensitive to noise. High pass filtering in the frequency domain mainly includes ideal high pass filtering, Butterworth high pass filtering, and Gaussian high pass filtering. The Butterworth high pass filter preserves the high-frequency components of the image to make the details and edges clearer, thereby achieving image sharpening processing [36]. It is more suitable for situations where image details and edges need to be emphasized. Gaussian high pass filter can effectively suppress noise while preserving image details, providing smoother image processing results. In this paper, the purpose of constructing a high-frequency guidance map is to emphasize the rain streak features, so a Butterworth filter was selected.

3 PROPOSED METHOD

In this section, a RSPNet is proposed, as shown in Fig. 2. Based on Eq. (1), this network is constructed to obtain a background layer (rainless image) by learning the rain streak layer and subtracting it from the rainy image. Firstly, considering that the rain streak features belong to high-frequency features, a high-frequency map is constructed using a Butterworth high pass filter on the input rainy image and used as a guide map to enhance rain streak features. Then, the constructed HFEM uses the guide map to enhance high-frequency features and make the network pay more attention to rain streak information. Next, A DFEM is constructed to extract accurate rain streak features and reconstruct the rain layer. Below, we will provide a detailed introduction to the HFEM and DFEM.

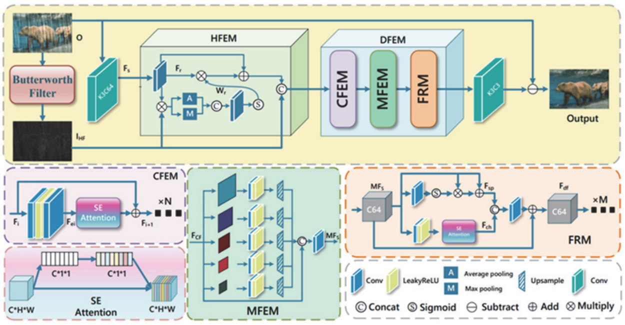


Figure 2 The structure of RSPNet
HFEM: High-frequency Feature Enhancement Module; DFEM: Deep Feature Extraction Module, consisting of CFEM, MFEM, and FRM. CFEM: Channel Feature Enhancement Module; MFEM: Multi-Scale Feature Extraction Module; FRM: Feature Refinement Module. SE: Squeeze-and-Excitation

3.1 High-Frequency Feature Enhancement Module (HFEM)

Although the high-frequency map obtained by the Butterworth high pass filter contains a large amount of rain streak information, it also contains some high-frequency information from the background layer. Therefore, the high-frequency map is not directly used as a rain layer, but as prior information of rain layer to guide the network to focus on learning rain streak features and separating the more accurate rain layer.

In HFEM, first, the shallow feature maps F_s with a channel number of 64 are dimensionally reduced to obtain the feature maps F_r with a channel number of 32 through a convolutional layer. Then, a high-frequency guidance map obtained by the Butterworth high pass filter is used to weight shallow features F_r to enhance rain streak features. And the enhanced feature maps are processed through the average pooling and maximum pooling operations to obtain a high-frequency spatial attention map W_r, which enables make the network pay attention to rain streak information. Next, the spatial attention map W_r are multiplied with the features maps F_r to enhance the rain streak features. Finally, the weighted result is added to the feature maps F_s to obtain the feature maps with rich streak information, and the high-frequency map is still used as guidance information for subsequent learning of image features. The above process can be represented by the following formula.

$$F_s = C_{3\times 3}(O),\ F_r = C_{3\times 3}(F_s) \tag{2}$$

$$W_r = Sig\left(C_{3\times3}\left(Concat\left(A\left(F_r \times I_{HF}\right), M\left(F_r \times I_{HF}\right)\right)\right)\right) \tag{3}$$

$$F_{en} = Concat\left(\left(F_r * W_r + F_r\right), I_{HF}\right) \tag{4}$$

where O represents the input rainy image, $C_{3\times3}$ represents the convolutional layer with filters of size 3 × 3, A and M represent average pooling and maximum pooling operations, respectively. *Concat* represents the concatenation operation and *Sig* is the Sigmoid activation function. I_{HF} represents the high-frequency map obtained by the Butterworth filter, W_r represents the high-frequency spatial attention map, and F_{en} represents the enhanced feature maps.

3.2 Deep Feature Extraction Module (DFEM)

A DFEM is constructed to learn and refine rain streak features from the enhanced features. This module mainly consists of CFEM, MFEM, and FRM, and their specific structure is described below.

3.2.1 Channel Feature Enhancement Module (CFEM)

CFEM is constructed by cascading N feature enhancement residual blocks to further extract the rain streak features. This residual block is composed of a convolutional layer and a Squeeze-and-Excitation (SE) attention module, which can supplement features through jump connections and avoid information loss caused by the increase in network depth. In addition, within the residual block, the 3 × 3 convolution layer is used to extract more abundant features in the local receptive field, and the SE attention module [20] is introduced to enable the residual block to selectively enhance the features $F_i\left(F_0 = F_{en}\right)$ in channels containing rich rain streak information. The operation of the i-th feature enhancement residual block can be expressed by the following formula.

$$Fe_i = C_{3\times3}\left(LRelu\left(C_{3\times3}\left(F_i\right)\right)\right) \tag{5}$$

$$CA = SE\left(Fe_i\right) = \sigma\left(W_2\delta\left(W_1\left(A\left(Fe_i\right)\right)\right)\right) \tag{6}$$

$$F_{i+1} = Fe_i * CA + F_i \tag{7}$$

where F_i represents the input of the i-th feature enhancement residual block or the output of the i-1-th residual block. Fe_i is the output of the convolutional layer in the i-th residual block, CA represents the weight matrix of channel attention, and $LRelu(\cdot)$ refers to the Leaky ReLU function. W_1 and W_2 are the weight coefficients of two fully connected layers, namely the dimensionality reducing layer and the dimensionality increasing layer. σ is a sigmoid function.

3.2.2 Multi-Scale Feature Extraction Module (MFEM)

Considering the diversity of size, shape, and direction of rain streaks, an MFEM containing six scale extraction branches is constructed to further extract the rain streak features. Here, the size of the feature maps in the minimum scale branch is set to 8 × 8, and the specific number of scales is defined as $s = \log_2 \min\left(H, W\right) - 3$. First, the inputs of these branches are obtained by down sampling the feature maps from CFEM at ratios of 1/2, 1/4, 1/8, 1/16, and 1/32. In each branch, a 1×1 convolution layer and a Leaky ReLU function are used to integrate rain streak features of different scales. The integrated feature maps are restored to the original size through up sampling operations. Then, the feature maps of the original scale are directly reused and concatenated with the output of all scale branches. Finally, a 1×1 convolution layer is used to reduce the channel dimension of the feature maps and achieve feature fusion. The operation of MFEM can be expressed by the following formula.

$$MF_i = Up\left(LRelu\left(C_{1\times1}\left(Down\left(F_{CF}\right)\right)\right)\right)\quad i = 1,2,3,4,5 \tag{8}$$

$$MF_s = C_{1\times1}\left(Concat\left(F_{CF}, MF_{i=1\cdots5}\right)\right) \tag{9}$$

where $C_{1\times1}$ represents the convolutional layer with filters of size 1×1. F_{CF} is the output of CFEM, MF_i is the output of the i-th scale branch, and MF_s represents the output of MFEM. Up and $Down$ represent the up sampling and down sampling operations, respectively.

3.2.3 Feature Refinement Module (FRM)

To obtain accurate rain streak information, a FRM composed of M feature refinement blocks is designed to enhance and refine the extracted rain streak features. The feature refinement block is designed as a dual attention residual block, which achieves feature enhancement in both spatial and channel dimensions.

In the spatial dimension, the input feature maps from MFEM first pass through a 3×3 convolution layer and a sigmoid function to generate a spatial weight map. Then, the weight map is multiplied with the input feature maps to filter features to extract more accurate rain streak information. Finally, the filtered rain streak features are added to the input feature maps to enhance the rain streak features. In the channel dimension, a SE attention module is used to achieve the feature enhancement. The two enhanced features are concatenated and integrated through a 3×3 convolutional layer. The process of a feature refinement block can be represented by the following formula:

$$F_{sp} = MF_s * \sigma\left(C_{3\times3}\left(MF_s\right)\right) + MF_s \tag{10}$$

$$F_{ch} = SE\left(LRelu\left(C_{3\times3}(MF_s)\right)\right) \tag{11}$$

$$F_{df} = C_{3\times3}\left(Concat\left(F_{sp}, F_{ch}\right)\right) + MF_s \tag{12}$$

where F_{sp} represents the enhanced feature maps in the spatial dimension, F_{ch} represents the enhanced feature maps in the channel dimension, and F_{df} is the output feature maps of a feature refinement block.

At the end of the network, the output of FRM goes through a convolution block containing a 3×3 convolution layer and a ReLU function to reconstruct the rain layer. The rain layer is subtracted from the input rain image to obtain the final rain removal image.

3.3 Loss Function

To better guide network training, a joint loss function L_{total} including reconstruction loss L_{rec}, gradient loss L_{grad} and structural similarity (SSIM) loss L_{SSIM} is defined as follows.

$$L_{total} = L_{rec} + \lambda_1 * L_{grad} + \lambda_2 * L_{SSIM} \tag{13}$$

where λ_1 and λ_2 are the trade-off parameters.

Reconstruction loss L_{rec}. The reconstruction loss is defined as L1 loss, which mainly ensures that the content of the rain removal image is closer to that of the ground truth (GT) image. Its definition can be expressed with the following formula:

$$L_{rec} = \left|O' - I_{GT}\right|_1 \tag{14}$$

where O' is the rain removal result, and I_{GT} is the GT image.

Gradient loss. To reduce the loss of texture information in the background layer, a gradient loss is defined to constrain the gradient information of the rain removal result, in order to make it consistent with the gradient information of GT image.

$$L_{grad} = \left|\nabla I_{GT}^{x} - \nabla O'^{x}\right|_1 + \left|\nabla I_{GT}^{y} - \nabla O'^{y}\right|_1 \tag{15}$$

where ∇I_{GT}^{x} and ∇I_{GT}^{y} represent the gradients of the GT image in the *x* and *y* directions, respectively. $\nabla O'^{x}$ and $\nabla O'^{y}$ represent the gradients of the rain removal result in the *x* and *y* directions, respectively.

SSIM loss. To ensure consistency in brightness, contrast, and structure between two images, as well as better visual effects, SSIM loss is defined. It can be represented by the following formula:

$$L_{SSIM} = 1 - SSIM\left(O', I_{GT}\right) \tag{16}$$

where $SSIM(\cdot)$ represents structural similarity index of two images.

4 EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we verify the validity of our proposed method through a large number of experiments, and compare it with the some state-of-the-art (SOTA) methods, including GMM [7], LPNet [24], DetailsNet [11], RESCAN [12], PRENET [15], BRNNET [30], ERRN-PRDS [37] and ECnet [38] and HCT [34]. To objectively evaluate the performance of all comparison methods, two widely used image quality evaluation metrics, such as Peak Signal to Noise Ratio (PSNR) [39] and Structural Similarity Index (SSIM) [40], are used.

To demonstrate the effectiveness of various components in the network, a series of ablation experiments are conducted, including HFEM, MFEM, and FRM, as well as the selection of loss functions and Butterworth parameters.

4.1 Implementation Details

We use a sliding window with a size of 100 × 100 to obtain image patches for training, where the sliding step size is 80. The initial learning rate is set to 0.0005. At the following 30, 50, and 80 epochs, it is attenuated by a multiple of 1/5. The weight factors λ_1 and λ_2 of the loss function are set to 0.5 by trial and error. The number of residual channel blocks N in CFEM is set to 24, and the number of double branch enhancement blocks M in DFEM is set to 6. The cutoff frequency is set to 30 and the order is set to 1 in Butterworth high-pass filter to get high-frequency information. Our network is implemented in Pytorch and trained on NVIDIA GTX 3060 GPU.

4.2 Evaluation on Synthetic Datasets

The datasets Rain200H, Rain100H, and Rain12 are tested on models trained using the heavy rain dataset RainTrainH [8]. The datasets Rain200L and Rain100L are tested on models trained using the light rain dataset RainTrainL [8]. Among all comparison methods, LPNet, BRNNET, ERRN-PRDS, ECNet, and HCT are tested using the pre-training model provided by the authors, while other comparison methods are retrained on heavy rain datasets (RainTrainH) and light rain datasets (RainTrainL).

Tab. 1 shows the average PSNR and SSIM values obtained by all comparison methods on different datasets. In Tab. 1, the best objective result is presented in bold, and while the second-best results are highlighted with an underline. From the table we can see that our proposed RSPNet has the highest PSNR and SSM values.

Table 1 Average PSNR (dB) and SSIM results of compared methods on synthetic datasets. Bold and underline represent the optimal and suboptimal values, respectively

Evaluation	RESCAN (ECCV2018)	LPNet (TNNLS2019)	PRENET (CVPR2019)	BRNNET (TIP2020)	ERRN-PRDS (TMM2021)	ECNet (WACV2022)	HCT (AAAI2023)	Proposed
Rain100H	28.82/0.863	23.40/0.822	29.46/0.899	30.47/0.913	31.67/0.927	31.43/0.921	31.51/0.913	**32.89/0.935**
Rain100L	39.22/0.983	33.42/0.959	37.48/0.979	38.16/0.982	41.02/0.989	39.66/0.986	39.70/**0.995**	**42.63**/0.991
Rain200H	28.03/0.859	23.28/0.815	29.04/0.899	29.96/0.913	30.60/0.923	30.22/0.912	30.48/0.910	**31.23/0.928**
Rain200L	38.38/0.982	33.56/0.963	37.87/0.982	38.68/0.984	40.01/0.987	39.72/0.987	39.08/0.983	**40.78/0.989**
Rain12	34.79/0.958	24.48/0.903	36.15/0.969	36.75/0.972	36.99/0.971	37.27/0.966	37.54/0.963	**38.03/0.974**

Figs. 3 and 4 show the results of rain removal on the Rain200H and Rain200L datasets, respectively. In the two figures, (a) shows the rainy image from the dataset, (j) shows the GT image, and (i) is the result of our method. As can be seen from the figures, our results contain more texture details, closest to the GT images. RESCAN and PRENET suffer from significant loss of background texture information, while LPNet results show a significant amount of residual rain streaks, especially in the sky region. The results of other methods also suffer from the loss of background information, resulting in blurred edges. Therefore, the proposed RSPNet is superior to other comparison methods.

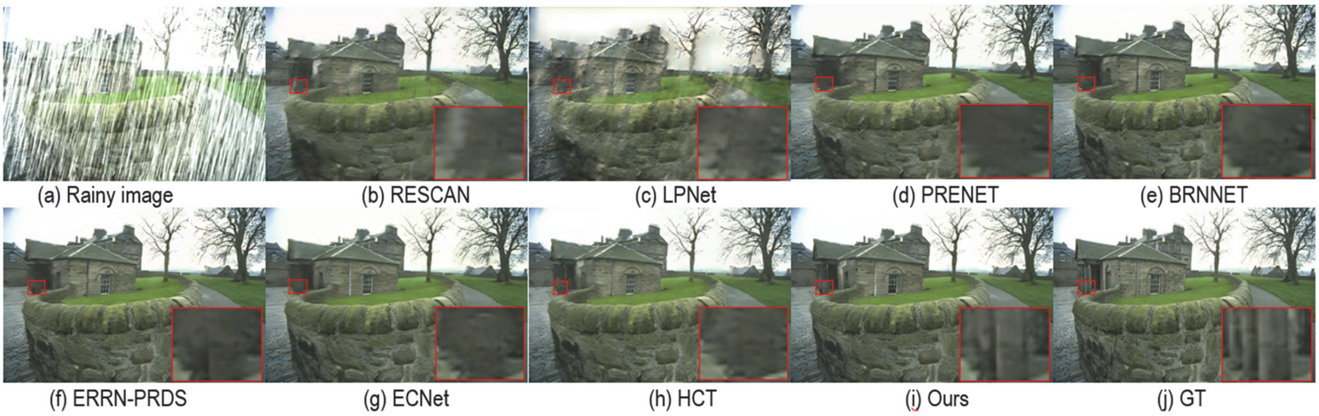

(a) Rainy image (b) RESCAN (c) LPNet (d) PRENET (e) BRNNET
(f) ERRN-PRDS (g) ECNet (h) HCT (i) Ours (j) GT

Figure 3 Visual comparison of the results obtained by comparison methods on the Rain200H dataset

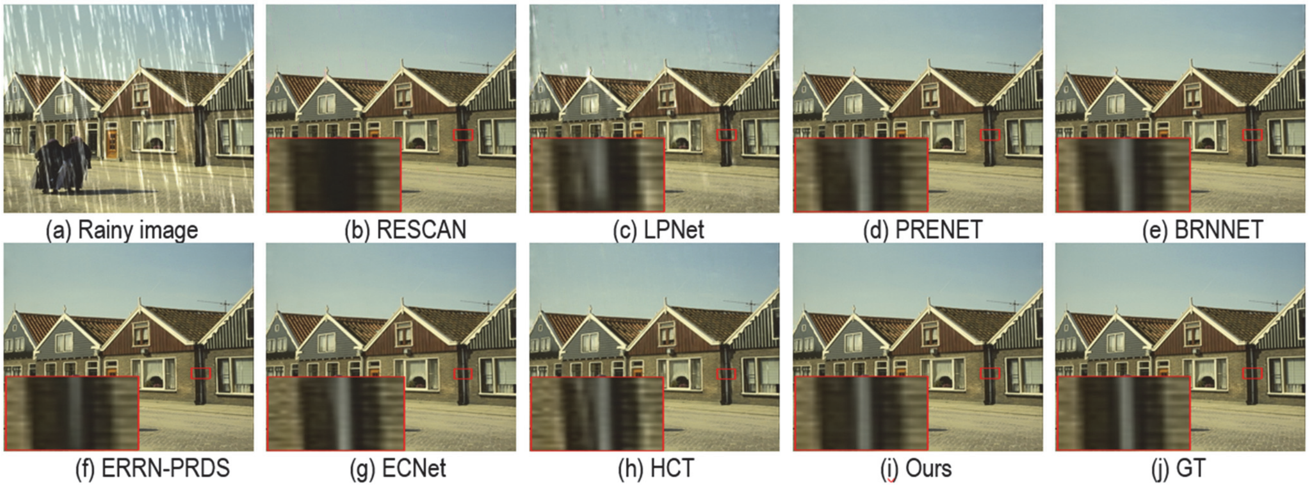

(a) Rainy image (b) RESCAN (c) LPNet (d) PRENET (e) BRNNET
(f) ERRN-PRDS (g) ECNet (h) HCT (i) Ours (j) GT

Figure 4 Visual comparison of the results obtained by comparison methods on the Rain200L dataset

4.3 Evaluation on Real Dataset

Although there are differences in the distribution between synthesized images and real rainy images, the proposed method also achieves good results on real rainy images. Figs. 5 to 7 shows the results of rain removal for rainy images in real scenes. Here we test RESCAN, PRENET, BRNNET, ERRN-PRDS, ECNet, HCT, and our methods on the real rainy images. As shown in Figures, although other methods can remove rain streaks from rainy images, they also lose more background information, and our method obtains clearer edge information. To facilitate the observation of differences between results, some local areas are selected and enlarged. From the enlarged areas, we can observe that our method can restore richer texture information, while other methods suffer from edge blurring and residual rain streaks. For example, in the results of RESCAN, PRENET, BRNNET, and ERRN-PRDS in Fig. 5, the textures in the green tile areas are relatively messy, while in the results of ECNet and HCT, the edges in the green tile areas appear blurry. In Fig. 6, the results of RESCAN, PRENET, BRNNET, and ERRN-PRDS show texture loss in the umbrella areas, while the results of ECNet and HCT show texture distortion in the umbrella areas. In Fig. 7, due to the presence of more rain streaks in the rainy image, there are still residual rain streaks in the results of ECnet and HCT.

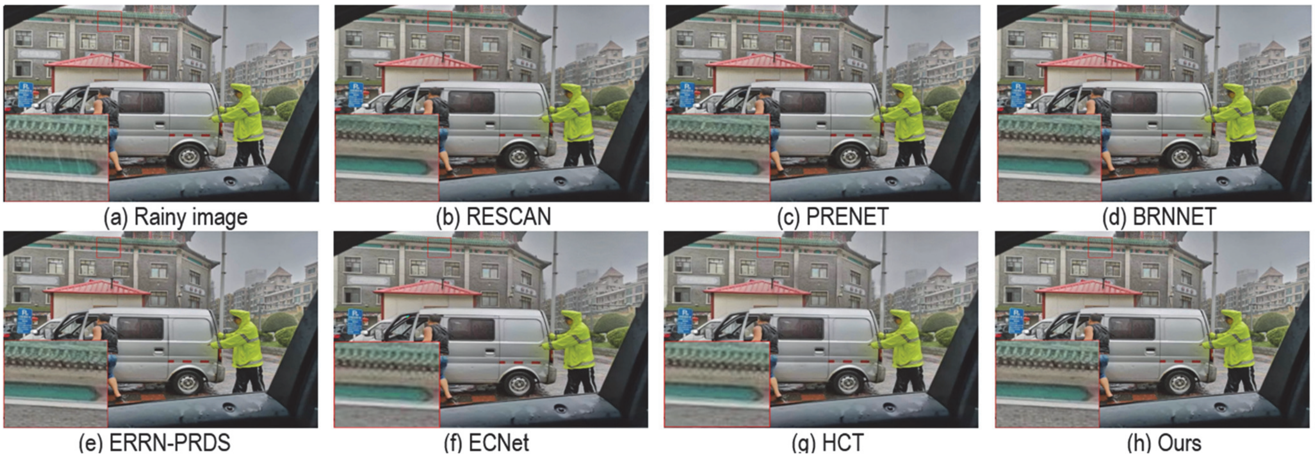

(a) Rainy image (b) RESCAN (c) PRENET (d) BRNNET
(e) ERRN-PRDS (f) ECNet (g) HCT (h) Ours

Figure 5 Visual comparison of the results obtained by comparison methods on the real dataset

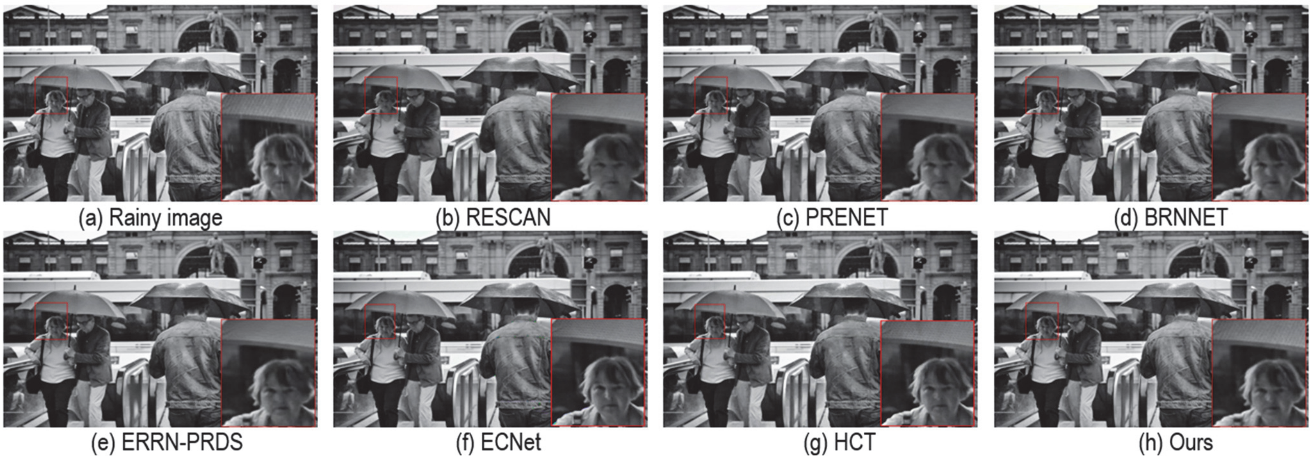

(a) Rainy image (b) RESCAN (c) PRENET (d) BRNNET
(e) ERRN-PRDS (f) ECNet (g) HCT (h) Ours

Figure 6 Visual comparison of the results obtained by comparison methods on the real dataset

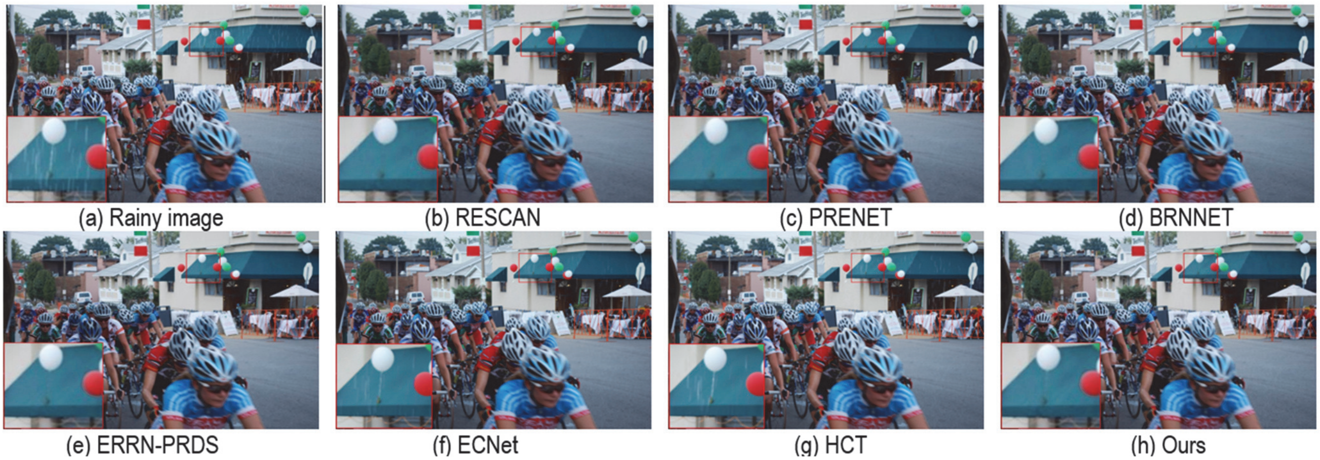

(a) Rainy image (b) RESCAN (c) PRENET (d) BRNNET
(e) ERRN-PRDS (f) ECNet (g) HCT (h) Ours

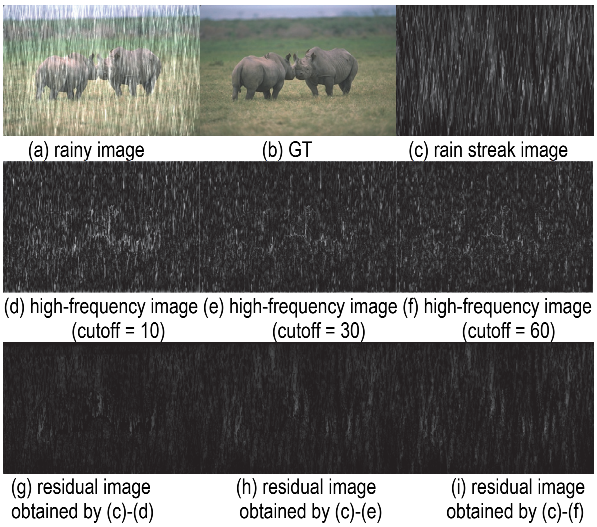
Figure 7 Visual comparison of the results obtained by comparison methods on the real dataset

4.4 Ablation Studies

To verify the effectiveness of the proposed network, we conduct several ablation experiments on the main components of the network, such as Butterworth cutoff frequency, HFEM, MFEM, and FRM, as well as loss function. All ablation experiments are performed on Rain200H [8] dataset, including 1800 training images and 200 test images. And the average PSNR and SSIM on 200 test images are used as evaluation criteria.

4.4.1 Butterworth Cutoff Frequency

The cutoff frequencies of Butterworth are different, and the extracted high-frequency information will also be different. To extract richer rain streak information, Butterworth filters with different cutoff frequencies are tested and used to obtain the high-frequency components of rainy images in the dataset, as shown in Fig. 8a and b show the rainy image and its corresponding GT image. (c) is obtained by subtracting (b) from (a), which is the rain streak image. (d), (e), and (d) is the high-frequency maps obtained by applying a Butterworth high pass filters with cutoff frequencies of 10, 30, and 60 on (a). To observe the rain streak information present in the high-frequency maps, we obtain the residual images (g), (h), and (i) by subtracting the high-frequency maps (d), (e), and (f) from the rain streak image (c).


(a) rainy image (b) GT (c) rain streak image

(d) high-frequency image (cutoff = 10) (e) high-frequency image (cutoff = 30) (f) high-frequency image (cutoff = 60)

(g) residual image obtained by (c)-(d) (h) residual image obtained by (c)-(e) (i) residual image obtained by (c)-(f)

Figure 8 Experiment on the selection of Butterworth cutoff frequency

From Figs. 8 g, h, and i, it can be seen that (g) contains less residual information compared to the other two images, but there is more background information compared to (h). And (i) has more residual information. This indicates that the larger the cutoff frequency, the less high-frequency components are extracted, and the smaller the cutoff frequency, the more background information is extracted. Therefore, this paper adopts a Butterworth high pass filter with a cutoff frequency of 30 to extract the rain streak prior map.

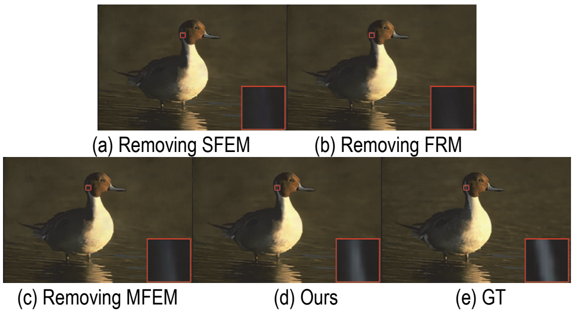
4.4.2 Effectiveness of HFEM, MFEM, and FRM

The models after removing HFEM, FRM, and MFEM from our model are named Model 1, Model 2, and Model 3, respectively, and are tested on Rain200H. Our model is Model 4. The average PSNR and SSIM values are shown in Tab. 2, and the visual comparison is shown in Fig. 9. To further demonstrate the network's ability to extract more accurate rain streak features, we compare the synthesized rain layer with the rain layer learned by the network. The visual comparison is shown in Fig. 10.

Table 2 Ablation study about HFEM, FRM, and MFEM on Rain200H dataset

Model	HFEM	FRM	MFEM	PSNR/SSIM
1	×	✓	✓	30.98/0.925
2	✓	×	✓	30.55/0.919
3	✓	✓	×	31.19/0.926
4	✓	✓	✓	**31.23/0.928**

HFEM: In the network, HFEM is used to enhance the rain streak features spatially. Here, the high-frequency guide map can make the network pay more attention to the extraction of rain streak features, and therefore, it can be used to enhance rain streak features. To verify the effectiveness of the proposed HFEM, we remove HFEM from the network and directly concatenate the high-frequency map with the extracted shallow features as inputs to the subsequence network. The modified network is retrained on the heavy rain dataset. From Tab. 2 and Fig. 9a, we can see that Mode 1 without HFEM obtains lower PSNR values than our model, i.e., Model 4, and some texture information is also lost. This also indicates that HFEM is effective.


(a) Removing SFEM (b) Removing FRM

(c) Removing MFEM (d) Ours (e) GT

Figure 9 Subjective results of ablation experiments on dataset Rain200H for HFEM, FRM and MFEM

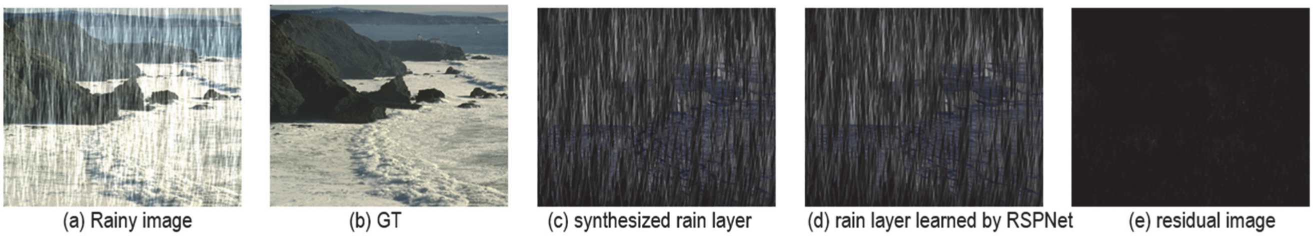

(a) Rainy image (b) GT (c) synthesized rain layer (d) rain layer learned by RSPNet (e) residual image

Figure 10 Visual comparison between synthetic rain layer and rain layer learned by RSPNet

MFEM: The adoption of multi-scale structure significantly improves the effectiveness of rain removal task in extracting complex and diverse rain streak features. To verify the function of the structure, we directly remove MFEM from the network and retrain the model for testing. From Tab. 2 and Fig. 9c, we can know that the objective indicators obtained by Model 3 have slightly decreased compared to Model 4, and the loss of texture information can also be clearly seen in the subjective results. Therefore, MFEM is effective.

FRM: FRM is used to further refine the rain streaks. To verify the validity of this module, we remove the module from the network and tested it. From Tab. 2 and Fig. 9b, we can observe that the objective indicators obtained by Model 2 without FRM have significantly decreased, and there is also a significant loss of texture information in the subjective results. Therefore, FRM is effective.

Fig. 10a and b show a rainy image and the corresponding rainless image, and (c) shows the synthesized rain layer obtained by subtracting (b) from (a). (d) is the rain layer learned by RSPNet, and (d) shows the residual image obtained by subtracting (d) from (c). We can observe from the figure that the residual image contains less information, indicating that the learned rain layer is relatively close to the artificially synthesized rain layer. Therefore, the proposed RSPNet can extract relatively accurate rain streak features.

4.4.3 Effectiveness of Loss Function

To verify the effectiveness of each loss term, such as L1 loss, gradient loss, and SSIM loss, we conducted ablation experiments on the Rain200H dataset, and the results are presented in Tab. 3. From the table, we can observe that the average PSNR and SSIM values obtained by using L1 loss alone are the lowest. When L1 loss is combined with gradient loss or SSIM loss respectively, network performance can be improved. However, the combination of the three loss terms achieves the highest measurement value. Therefore, all three loss terms are effective in improving model performance.

Table 3 Ablation experiments of different loss terms on Rain200H dataset

L1 loss	Gradient loss	1-SSIM	PSNR/SSIM
✓	✓		30.98/0.921
✓		✓	31.09/0.926
✓			30.66/0.915
✓	✓	✓	**31.23/0.928**

Discussion:

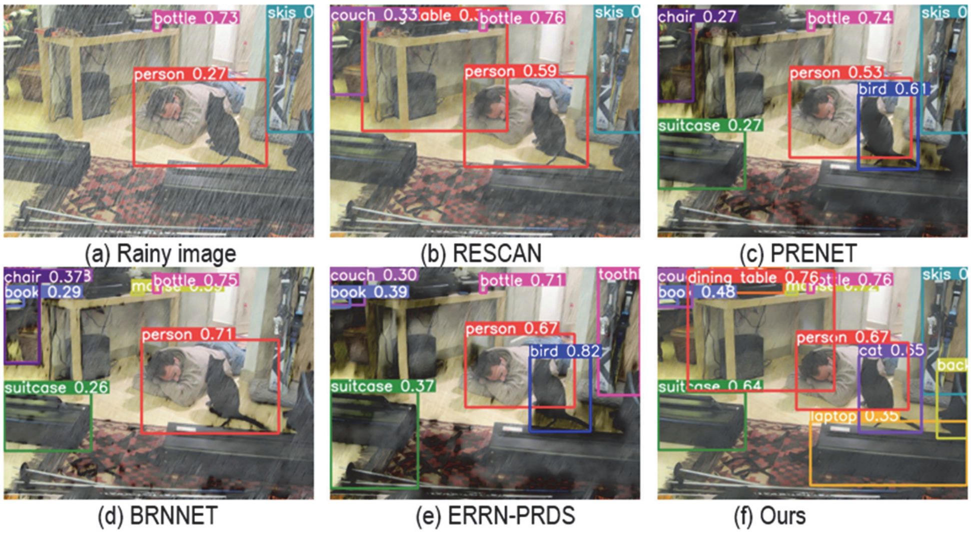
Although the proposed method has achieved considerable results, there are still certain limitations and challenges, as described below.

(1) Due to the lack of real training datasets, the proposed model trained on synthetic datasets have weak generalization on real rainy images. Due to the limitations of the imaging environment in rainy weather, real rainy images not only have linear rain features, but also features of raindrops and rain fog. At present, in rain removal tasks, most synthetic datasets only contain single rain streak features, which makes it difficult for the model to effectively remove rain information from real rainy images.

(2) How to construct a synthetic dataset that is similar to the distribution of real rainy data is currently a challenge for rain removal tasks. Currently, the synthesis of rain images is obtained by directly linearly superimposing the rain layer and the image background layer according to Eq. (1), while the degradation process of real rainy images cannot be represented by a simple linear relationship. Therefore, a reasonable degradation model needs to be defined to simulate the degradation process of rainy images, which will be beneficial for generating more realistic synthetic datasets.

4.5 Application on High-level Visual Tasks

The image rain removal task is important for improving the performance of high-level visual tasks such as image segmentation and object detection. Therefore, in order to verify that the proposed method can better assist high-level visual tasks, we conduct the comparison experiments of target recognition and image segmentation on the rain removal results. In the experiment, Photoshop is used on the COCO [41] and BDD [42] datasets to generate test datasets containing rainy images, namely COCO350 and BDD150 [25]. The target detection method proposed by Redmon et al. [43] is applied to the rain removal results from the COCO350 dataset, and the object detection results are shown in Fig.11. The rain removal results from the BDD150 dataset are used to test image segmentation, as shown in Fig. 12. The image segmentation method proposed by Mehta et al. [44] is applied to the rain removal results from the BDD150 dataset, and the image segmentation results are shown in Fig. 12. The target detection and image segmentation models are tested using the pre-training models provided by the authors.



(a) Rainy image (b) RESCAN (c) PRENET

(d) BRNNET (e) ERRN-PRDS (f) Ours

Figure 11 Target detection results on COCO350 [25] dataset using yolo3 [43]

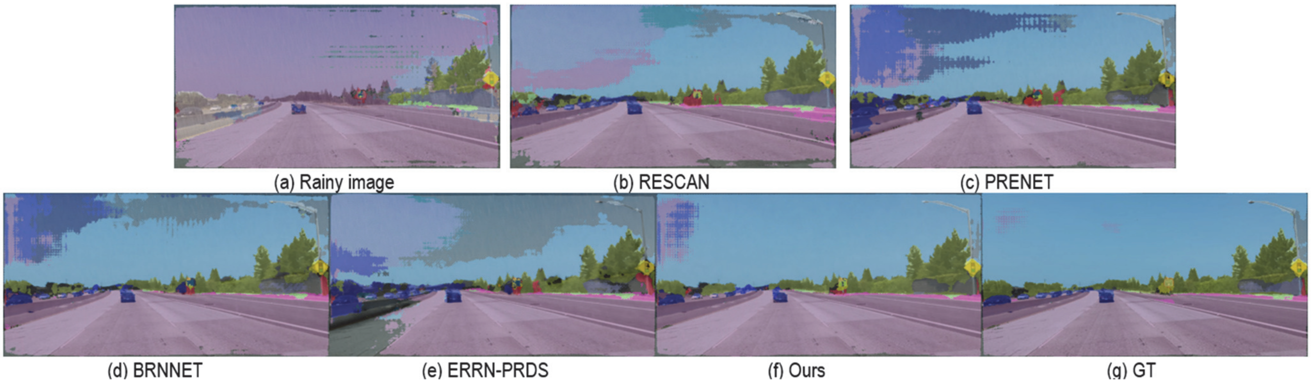


Figure 12 Segmentation results on BDD150 [25] dataset using ESPNet [44]

From Fig. 11, we can see that due to the interference of rain streaks, a small number of targets are detected in the rainy image, and the detection confidence is also relatively low. In the rain removal results, more targets can be detected, but due to the presence of residual rain marks in the results of some comparative methods, there are also problems of missed and false detections of targets. In our results, more targets are detected and higher detection confidence are achieved. Therefore, the proposed method obtains rain removal results with richer information and clearer edges.

From Fig. 12, we can observe that in the segmentation result of the rainy image, the sky and road areas are mistakenly classified into one category. In the results of other comparison methods, there are still some misclassified sky regions, and our segmentation result only has a small number of misclassified regions, which is closest to the segmentation result of the GT image. This also indicates that the rain removal results obtained by the proposed method are closest to the rainless images.

5 CONCLUSIONS

This study presents RSPNet, a novel approach to single image rain removal that leverages high-frequency information as a rain streak prior. By incorporating a Butterworth filter-derived guidance map and specialized network modules, RSPNet demonstrates superior performance in rain streak removal while preserving background details. Extensive experiments on synthetic and real-world datasets show that RSPNet consistently outperforms state-of-the-art methods in terms of PSNR and SSIM metrics. The proposed method also enhances the performance of downstream tasks such as object detection and image segmentation on rain-affected images. However, limitations exist, such as the potential over-reliance on the high-frequency prior for certain rain patterns. Future work should focus on adapting the method to a wider range of rain types and exploring the integration of temporal information for video deraining. The findings of this study offer valuable insights for advancing rain removal techniques and improving the robustness of outdoor computer vision systems.

6 REFERENCES

[1] Cai, L., Fu, Y., Zhu, T., Xiang, Y., Zhang, Y., & Zeng, H. (2021). Joint depth and density guided single image de-raining. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(7), 4108-4121. https://doi.org/10.1109/TCSVT.2021.3121012

[2] Cai, L., Fu, Y., Huo, W., Xiang, Y., Zhu, T., Zhang, Y., ... & Zeng, D. (2022). Multiscale attentive image de-raining networks via neural architecture search. *IEEE Transactions on Circuits and Systems for Video Technology*, *33*(2), 618-633. https://doi.org/10.1109/TCSVT.2022.3207516

[3] Cui, X., Wang, C., Ren, D., Chen, Y., & Zhu, P. (2022). Semi-supervised image deraining using knowledge distillation. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(12), 8327-8341. https://doi.org/ 10.1109/TCSVT.2022.3190516

[4] Ren, D., Zuo, W., Zhang, D., Zhang, L., & Yang, M. H. (2019). Simultaneous fidelity and regularization learning for image restoration. *IEEE transactions on pattern analysis and machine intelligence*, *43*(1), 284-299. https://doi.org/10.1109/TCSVT.2022.3190516

[5] Kang, L. W., Lin, C. W., & Fu, Y. H. (2011). Automatic single-image-based rain streaks removal via image decomposition. *IEEE transactions on image processing*, *21*(4), 1742-1755. https://doi.org/10.1109/TPAMI.2019.2926357

[6] Luo, Y., Xu, Y., & Ji, H. (2015). Removing rain from a single image via discriminative sparse coding. *Proceedings of the IEEE international conference on computer vision*, 3397-3405. https://doi.org/10.1109/ICCV.2015.388

[7] Li, Y., Tan, R. T., Guo, X., Lu, J., & Brown, M. S. (2016). Rain streak removal using layer priors. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2736-2744. https://doi.org/10.1109/CVPR.2016.299

[8] Yang, W., Tan, R. T., Feng, J., Liu, J., Guo, Z., & Yan, S. (2017). Deep joint rain detection and removal from a single image. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1357-1366. https://doi.org/10.1109/CVPR.2017.183

[9] Pan, J., Liu, S., Sun, D., Zhang, J., Liu, Y., Ren, J., ... & Yang, M. H. (2018). Learning dual convolutional neural networks for low-level vision. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3070-3079. https://doi.org/10.1109/CVPR.2018.00324

[10] Fu, X., Huang, J., Ding, X., Liao, Y., & Paisley, J. (2017). Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, *26*(6), 2944-2956. https://doi.org/10.1109/TIP.2017.2691802

[11] Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., & Paisley, J. (2017). Removing rain from single images via a deep detail network. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3855-3863. https://doi.org/10.1109/CVPR.2017.186

[12] Li, X., Wu, J., Lin, Z., Liu, H., & Zha, H. (2018). Recurrent squeeze-and-excitation context aggregation net for single image deraining. *Proceedings of the European conference on computer vision*, 254-269.

https://doi.org/10.1007/978-3-030-01234-2_16

[13] Zhang, H., & Patel, V. M. (2018). Density-aware single image de-raining using a multi-stream dense network. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 695-704. https://doi.org/10.1109/CVPR.2018.00079

[14] Fan, Z., Wu, H., Fu, X., Huang, Y., & Ding, X. (2018, October). Residual-guide network for single image deraining. *Proceedings of the 26th ACM international conference on Multimedia*, 1751-1759. https://doi.org/10.1145/3240508.3240694

[15] Ren, D., Zuo, W., Hu, Q., Zhu, P., & Meng, D. (2019). Progressive image deraining networks: A better and simpler baseline. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3937-3946. https://doi.org/10.1109/CVPR.2019.00406

[16] Wang, C., Wu, Y., Su, Z., & Chen, J. (2020, October). Joint self-attention and scale-aggregation for self-calibrated deraining network. *Proceedings of the 28th ACM International Conference on Multimedia*, 2517-2525. https://doi.org/10.1145/3394171.3413559

[17] Wang, Y., Ma, C., & Zeng, B. (2021). Multi-decoding deraining network and quasi-sparsity based training. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 13375-13384. https://doi.org/10.1109/CVPR46437.2021.01317

[18] Wei, W., Meng, D., Zhao, Q., Xu, Z., & Wu, Y. (2019). Semi-supervised transfer learning for image rain removal. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3877-3886. https://doi.org/10.1109/CVPR.2019.00400

[19] Wang, H., Yue, Z., Xie, Q., Zhao, Q., Zheng, Y., & Meng, D. (2021). From rain generation to rain removal. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 14791-14801. https://doi.org/10.1109/CVPR46437.2021.01455

[20] Wei, K., Yang, J., Fu, Y., Wipf, D., & Huang, H. (2019). Single image reflection removal exploiting misaligned training data and network enhancements. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8178-8187. https://doi.org/10.1109/CVPR.2019.00837

[21] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132-7141. https://doi.org/10.1109/TPAMI.2019.2913372

[22] Garg, K., & Nayar, S. K. (2007). Vision and rain. *International Journal of Computer Vision*, *75*, 3-27. https://doi.org/10.1007/s11263-006-0028-6

[23] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, *37*(9), 1904-1916. https://doi.org/10.1109/TPAMI.2015.2389824

[24] Fu, X., Liang, B., Huang, Y., Ding, X., & Paisley, J. (2019). Lightweight pyramid networks for image deraining. IEEE transactions on neural networks and learning systems, *31*(6), 1794-1807. https://doi.org/10.1109/TNNLS.2019.2926481

[25] Jiang, K., Wang, Z., Yi, P., Chen, C., Huang, B., Luo, Y., ... & Jiang, J. (2020). Multi-scale progressive fusion network for single image deraining. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8346-8355. https://doi.org/10.1109/CVPR42600.2020.00837

[26] Chen, X., Huang, Y., & Xu, L. (2021). Multi-scale hourglass hierarchical fusion network for single image deraining. In Proceedings of the IEEE conference on computer vision and pattern recognition, 872-879. https://doi.org/10.1109/CVPRW53098.2021.00097

[27] Yang, Y. & Lu, H. (2019). Single image deraining using a recurrent multi-scale aggregation and enhancement network. *Proceedings of the 2019 IEEE International Conference on Multimedia and Expo*, 1378-1383. https://doi.org/ 10.1109/ICME.2019.00239

[28] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. *Proceedings of the European conference on computer vision*, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[29] Shao, M. W., Li, L., Meng, D. Y., & Zuo, W. M. (2021). Uncertainty guided multi-scale attention network for raindrop removal from a single image. *IEEE Transactions on Image Processing*, *30*, 4828-4839. https://doi.org/10.1109/TIP.2021.3076283

[30] Ren, D., Shang, W., Zhu, P., Hu, Q., Meng, D., & Zuo, W. (2020). Single image deraining using bilateral recurrent network. *IEEE Transactions on Image Processing*, *29*, 6852-6863. https://doi.org/10.1109/TIP.2020.2994443

[31] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE international conference on computer vision*, 10012-10022. https://doi.org/10.1109/ICCV48922.2021.00986

[32] Xiao, J., Fu, X., Liu, A., Wu, F., & Zha, Z. J. (2022). Image de-raining transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(11), 12978-12995. https://doi.org/10.1109/TPAMI.2022.3183612

[33] Chen, X., Li, H., Li, M., & Pan, J. (2023). Learning a sparse transformer network for effective image deraining. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5896-5905. https://doi.org/10.1109/CVPR52729.2023.00571

[34] Chen, X., Pan, J., Lu, J., Fan, Z., & Li, H. (2023, June). Hybrid cnn-transformer feature fusion for single image deraining. *Proceedings of the AAAI conference on artificial intelligence*, *37*(1), 378-386. https://doi.org/10.1609/aaai.v37i1.25111

[35] Makandar, A. & Halalli, B. (2015). Image enhancement techniques using highpass and lowpass filters. International Journal of Computer Applications, *109*(14), 12-15. https://doi.org/10.5120/19256-0999

[36] Chen, W., Liu, Y., Zhang, J., Duan, Z., Zhang, L., Hou, X., ... & Chou, X. (2023). Blind Super-Resolution Network with Dual-Channel Attention for Images Captured by Sub-Millimeter-Diameter Fiberscope. *Electronics*, *12*(20), 4352. https://doi.org/10.3390/electronics12204352

[37] Yang, Y., Guan, J., Huang, S., Wan, W., Xu, Y., & Liu, J. (2021). End-to-end rain removal network based on progressive residual detail supplement. *IEEE Transactions on Multimedia*, *24*, 1622-1636. https://doi.org/10.1109/TMM.2021.3068833

[38] Li, Y., Monno, Y., & Okutomi, M. (2022). Single image deraining network with rain embedding consistency and layered LSTM. *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 4060-4069. https://doi.org/10.1109/WACV51458.2022.00401

[39] Huynh-Thu, Q. & Ghanbari, M. (2008). Scope of validity of PSNR in image/video quality assessment. *Electronics letters*, *44*(13), 800-801. https://doi.org/10.1049/el:20080522

[40] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, *13*(4), 600-612. https://doi.org/10.1109/TIP.2003.819861

[41] Caesar, H., Uijlings, J., & Ferrari, V. (2018). Coco-stuff: Thing and stuff classes in context. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1209-1218. https://doi.org/10.1109/CVPR.2018.00132

[42] Yu, F., Xian, W., Chen, Y., Liu, F., Liao, M., Madhavan, V., & Darrell, T. (2018). Bdd100k: A diverse driving video database with scalable annotation tooling. *arXiv preprint*, 1805.04687, *2*(5), 6.

https://doi.org/10.48550/arXiv.1805.04687

[43] Redmon, J. & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767. https://doi.org/10.48550/arXiv.1804.02767

[44] Mehta, S., Rastegari, M., Caspi, A., Shapiro, L., & Hajishirzi, H. (2018). Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation. *Proceedings of the european conference on computer vision*, 552-568. https://doi.org/10.1007/978-3-030-01249-6_34

[45] Zhu, L., Fu, C. W., Lischinski, D., & Heng, P. A. (2017). Joint bi-layer optimization for single-image rain streak removal. *Proceedings of the IEEE international conference on computer vision*, 2526-2534. https://doi.org/10.1109/ICCV.2017.276

Contact information:

Aoran ZHAO
(Corresponding author)
Tiangong University,
School of Computer Science and Technology,
Tianjin 300387, China
E-mail: zhaoaoran2022@126.com