# Multi-scale Attention Map Guided Image Rain Removal Network

**Gu Kunyuan[1], Pang Xiaoyan[1,a,*], Zhu Xiaoli[1], Zhang Peng[1]**

*[1]School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan, 454000, China*
*[a]1224019737@qq.com*
*[*]Corresponding author*

***Abstract:*** *To solve the problem of low running speed of image rain removal models, a multi scale attention map guided image rain removal network is proposed. Firstly, an original resolution main network with residual structure is constructed for image rain removal. Secondly, in order to enhance the rain removal ability of the main network, especially the recognition ability of multi-scale features, a multi-scale attention map generation network is designed as an auxiliary to the main network. The generated multi-scale attention maps are used to enhance the features of rain stripes in the image, guiding the main network to identify and remove them in turn. In addition, this paper also designed the dual attention guided convolution block, which adds the ability to pay attention to spatial features on the basis of traditional channel attention mechanisms to achieve joint attention to spatial and channel features. Experimental results show that the proposed method maintains the comparability and stability of the restored image quality while significantly reducing the amount of network parameters and effectively improving the speed.*

***Keywords:*** *Image rain removal, Deep learning, Residual structure, Multi-scale dilated convolutions, Attention module*

## 1. Introduction

Under weather conditions such as rain, snow, and fog, the quality of the captured image can be affected. In particular, rain can seriously reduce the visibility of the scene, making the captured image blurred, causing loss of image details. This will affect the reliability of image processing algorithms and the smooth implementation of tasks such as pedestrian detection[1], security monitoring[2], and object recognition[3]. Image restoration is the task of restoring clean images from degraded versions of images. Rain is a typical example of its degradation. Image rain removal is an ill posed problem because there are countless feasible solutions. In order to limit the solution space to valid images or natural images, traditional image rain removal methods use a priori image made by hand. However, designing such a priori is a challenging task, and the method is often non generalizable. To solve this problem, the latest method uses convolutional neural networks (CNN) in deep learning to learn more general feature representations by capturing natural image statistics from large-scale data.

With the advancement of deep learning technology, many network and functional modules have been developed for image applications, including recursive residual learning[12][13], dilated convolution[14], and attention mechanisms[15][16]. However, existing methods mainly have two limitations. On the one hand, spatial details and high-level context information are very useful for rainwater removal, but many methods remove raindrops based on image patches, ignoring contextual information in larger areas of the image. On the other hand, major CNN models[17][18] typically use multiple convolutional blocks (convolutional layers and activation functions) to learn target features at the original resolution of the input image, continuously stacking a large number of convolutional blocks to increase network depth, and learning context dependencies under large receptive fields. However, such models cannot accurately learn the multi-scale features of the image and the structural features of different rain layers. In order to solve these problems, some literatures[19][20][21] have studied the use of multi-scale information for image rain removal, mainly using multiple isomorphic subnets to model different scales of several resolution images. However, the multi subnet structure makes the network more complex and increases computing costs. What's more, inspired by multi-scale dilated convolution networks, some literatures[12][13][14] combines multi-scale dilated convolution to construct lightweight networks to learn multiscale image features. Since the purpose of the rain removal task is to restore all the original information of the image, these networks may have insufficient modeling capabilities at the original resolution of the input image.

To address these limitations, this paper proposes a multi-scale attention map guided image rain removal network, which is further enhanced on the basis of main network with residual structure. Firstly, the multi-scale attention map generation network is used to generate attention maps at multiple scales to guide the rain removal work of the main network in turn. Secondly, a dual attention guided convolution block that focuses on both spatial and channel features is constructed to further improve the network's ability to recognize rain stripes. The proposed network has achieved satisfactory results on data sets. Experimental results show that it not only improves the running speed of the model, but also improves the imaging quality. The contributions of this paper mainly include three aspects:

(1) This paper proposes a multi-scale attention map guided rain removal network. The network is assisted by an auxiliary network to remove rain from the main network. The main network learns detailed image features from the original resolution of the input image, and the auxiliary network uses the generated multi-scale attention map to guide the main network in rainwater removal.

(2) This paper constructs a multi-scale attention map generation network as an auxiliary to the main network, using multi-scale dilated convolution to focus on the multi-scale features of the image, and guiding the feature learning process of the main network through the generated multi-scale feature map.

(3) This paper also constructs a dual attention guided convolution block to automatically learn spatial and channel features to restore the original spatial structure and detailed texture of the image, and enhance the main network's perception of rain stripe features.

## 2. Related Work

The early methods of image rain removal mainly include filter based methods[4][5], sparse coding based methods[6][7], and Gaussian mixture model based methods[8][9]. These methods all rely on a priori image made manually, and the generalization ability of the model is weak, which cannot achieve good rain removal effects in some real-world scenarios. The recovery method based on convolutional neural network has achieved the most advanced results. This paper briefly introduces the single image rain removal models based on deep learning, including single resolution deep models and multi resolution deep models.

### *2.1. Single resolution deep models*

The image rain removal models based on single resolution achieves rain removal at the original resolution of the image, and these models attach importance to the construction of a rain removal module. In 2017, Fu et al.[10][11] first explored a simple deep convolutional neural network to learn mapping relationships between images, and further expanded the architecture by adding deep residual blocks and global skip connections to improve performance. The network removes rainfall from the high-frequency components of the decomposed rainfall image, and verifies the significant advantages of the image rain removal method based on depth learning. In the same year, Yang et al.[12]detected the location of rain by predicting a binary rain stripe mask, and gradually removed rain stripes and accumulated rain using a circular framework. In 2018, Li et al.[13] used a dilated convolutional network to expand the receptive field, learning the interdependencies between channels through a squeeze excitation module, and using a circular structure to remove rainwater layer by layer. In 2019, Wang et al.[14] built a spatial attention module to establish a spatial distribution characteristic map of rainwater through four-way cyclic convolution operations, guiding the network to identify and eliminate rain patterns from local to global perspectives. In 2020, He et al.[15] constructed a network architecture that integrates channel attention, spatial attention, and multi-layer features, using interpolation-based pyramid attention blocks to perceive spatial information at different scales, and improving the visibility of severely degraded images. In 2021, Su et al.[16] modeled two-dimensional rainfall images as vector sequences in vertical, lateral, and channel directions. Aggregating information repeatedly from all three directions enables the model to capture long-term dependencies between channels and spatial locations. In 2022, Liang et al.[17] proposed a lightweight recursive transformer, which constructed a recursive local window self-attention structure with residual connections, requiring only a small amount of computing resources to achieve rain stripe removal. All of these existing depth models operate on the original resolution of the input image to improve performance by increasing the size of the network. In order to establish a long dependency relationship, it is often necessary to overlay dozens or hundreds of convolution modules.

### *2.2. Multi resolution deep models*

The multi resolution based image rain removal models process rain images at multiple resolutions by constructing a more complex network to achieve rain removal, which are typically achieved in the gradual manner. In 2019, Ren et al.[18] constructed a residual network composed of five residual blocks as the basic module of the rain removal network, and then jointly used six of these modules to gradually remove rain streaks from the image. In 2020, Jiang et al.[19] synergistically learned the characteristics of rain stripes through a pyramid structure, using complementary and redundant information in spatial dimensions to characterize the target rain stripes. In 2022, Jiang et al.[20] decomposed the rainfall removal task into a bilateral grid learning stage and a joint feature refinement stage. The former preserves image edge details while expanding the distance between rain patterns and background information. The latter uses a dual path interaction module to dynamically and gradually decouple the rain trace content and intermediate features of clear image details. In the same year, Li et al.[21] restored images through two encoder-decoder structure subnets, using spatial filtering branches and energy-based attention branches as core components to restore images degraded by rain. These image rain removal models build complex network architectures to fully learn input data of different resolutions, but the network structure is complex and has a large amount of parameters, making it difficult to achieve faster operation speed.

In this paper, a lightweight single resolution network is used, and multi-scale feature maps are used to guide the network in learning rainwater features at the original resolution. On the one hand, the original resolution main network is designed based on the dual attention guided convolution block and residual structure to learn fine textures at the original image resolution; On the other hand, the multi-scale attention map generation network is constructed based on the multi-scale dilated convolutional structure and the supervised attention module, and the generated multi-scale feature map is used to assist the main network in removing rainwater.

## 3. Algorithm Design

### *3.1. Network structure*

In order to balance the speed and quality of image rain removal, this paper proposes a multi-scale attention map guided image rain removal network (MSAMNet). The network is mainly composed of the original resolution main network (ORMNet), the multi-scale attention map generation network (MAGNet), and the dual attention guided convolution block (DAB). The overall network architecture is shown in Figure 1. The original resolution main network learns the representative features of the image at the original image resolution without changing the image scale. The multi-scale attention map generation network learns the multi-scale context features of images and generates feature maps at various scales to guide the main network in learning. As the basic unit for processing images in the original resolution main network, the dual attention guided convolution block strengthens the network's attention to the spatial and channel characteristics of images.

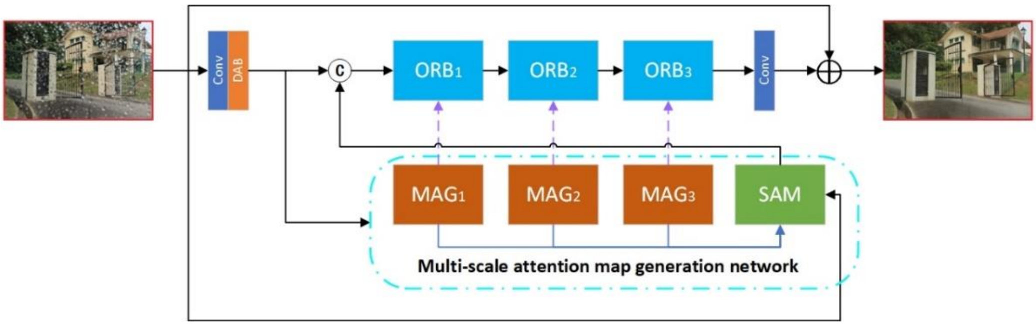


*Figure 1: Overall structure of the MSAMNet.*

Specifically, the input rain image first undergoes the convolution feature refinement module to convert the three channel features $X$ of the input RGB into features $X_S$ with $S$ channels, and then undergoes preliminary refinement of the features through the attention mechanism. After that, the converted features are imported into ORMNet and MAGNet respectively to further extract rainwater characteristics. ORMNet is composed of three original resolution blocks, and each ORB is composed of multiple cascaded DABs that introduce residual connection. It can gradually remove rainwater from the original rain map and restore a clear background image. MAGNet uses its generated feature attention map to guide ORMNet to identify the characteristics of different rain stripes. On the one hand, it uses dilation convolutions with different dilation rates to extract rainwater feature maps with different resolutions, aggregates them with ORBs in ORMNet, and sequentially guides ORBs to remove rain stripe. On the other hand, the supervised attention module (SAM) is used to aggregate the rainwater features learned by different scale dilation convolutions, further selecting and enhancing $X_S$ at the original resolution, for easy removal by ORMNet. Finally, the convolution layer reconstructs the rainwater characteristics at the end of ORMNet.

### *3.2. Original resolution main network*

Image rain removal task requires restoring the spatial details and high-level context information of the background image. In order to achieve network lightweight, this paper introduces residual connection. By nesting residual in the residual structure, a group of dual attention guided convolution blocks are stacked to form the original resolution module, and then three original resolution modules are cascaded to further form the ORMNet. Since the network always maintains the same resolution as the input during the learning process, without using down sampling or other resolution reduction operations, the network can retain spatial details from the input image to the output image.

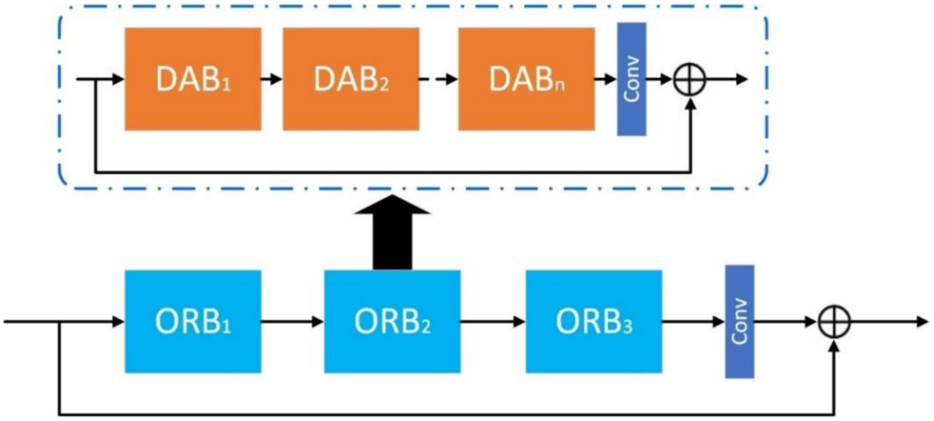


*Figure 2: Original resolution main network.*

The main structure of the original resolution main network is a residual in residual structure (RIR), which allows multiple residual connections to bypass rich low-frequency background information, allowing the network to focus on learning high-frequency rain features. As shown in Figure 2, the outer residual structure takes ORB as the basic module, and the outer residual connection helps the network learn more abstract rough level rainwater features; The inner residual structure uses DAB as the basic module, and the inner residual connection helps the network learn more specific rainwater details.

### *3.3. Multi-scale attention graph generation network*

The image rain removal task requires removing multiple rain layers with different blurring levels and resolutions. In order to further improve the recognition ability of the main network for spatial characteristics of rainwater, this paper constructs a multi-scale attention map generation network. This network generates attention maps at multiple scales by learning the features of rain stripes at different scales, and guides the ORB in the corresponding main network to remove rain stripes. Finally, a supervised attention module[22] is also introduced to finally fuse the attention maps of multiple scales generated by the network, enhance the rainwater features at the original resolution, and fuse them with the data flow in the main network. Before the data is processed by the ORB module, highlight the key features in the original data.

Specifically, as shown in Figure 3, the multis-cale attention map generation network first extracts the features of different scales using three dilated convolutions with different dilation factors (DFs), which can expand the receptive field (RF) exponentially without losing resolution. Due to the different sizes of rain stripes at different locations in the image, using dilated convolution[23] can improve the recognition ability of the network for different types of rain stripes. Specifically, in this paper, dilated convolutions with DFs of 1, 2, and 3 are used, and the rainwater image with original resolution is first converted into a rainwater feature space by the convolution layer. After entering three paths with DFs of 1, 2, and 3, each path is cascaded with two dilated convolutions with the same DF. Under different paths, the RF sizes of the first layer of dilated convolution are respectively $3\times3$, $5\times5$, and $7\times7$, and the RF sizes of the second layer of dilated convolution are respectively $5\times5$, $9\times9$, and $13\times13$. The network extracts the context information of rain stripes at different locations under multiple scales of different paths, and generates three rain feature attention maps corresponding to each other; Finally, the feature map is aggregated and enters the supervised attention module. In addition, the rainwater feature maps generated by the dilated convolution are also transmitted to the corresponding ORB modules in the main network to enhance their processing ability for rainwater, and achieve the removal of rain stripes in the image layer by layer on a scale basis.

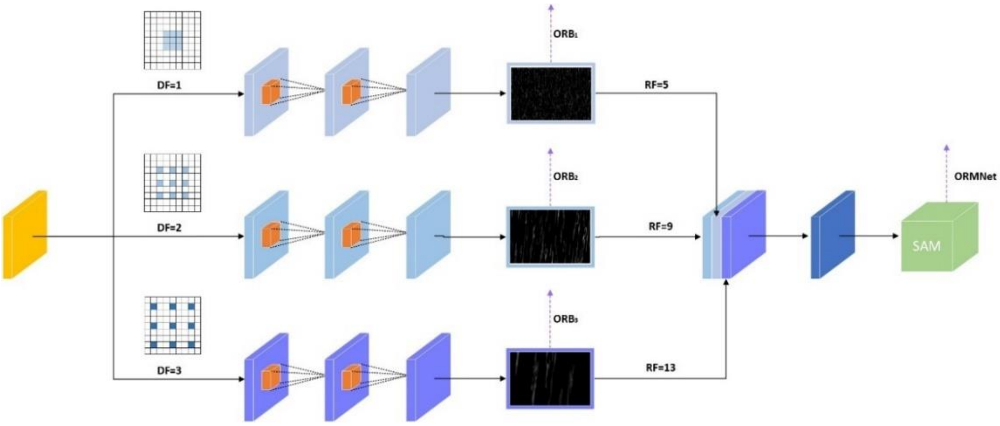


*Figure 3: Multi-scale attention graph generation network.*

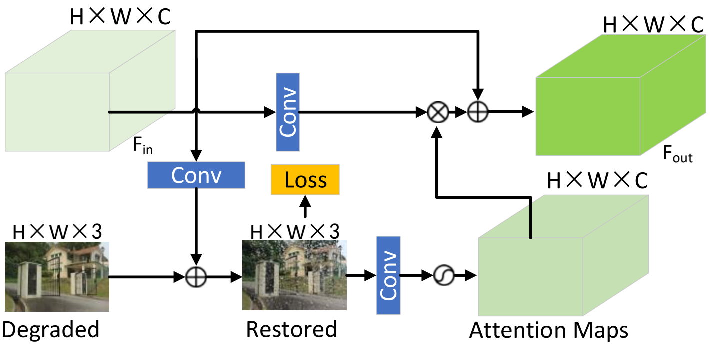


*Figure 4: Supervised attention module.*

This paper uses a supervised attention module[22] to connect the main network and the auxiliary network, as shown in Figure 4. It uses the generated attention map to suppress features with less information in the current network, allowing only useful features to be propagated. The attention network subtracts the rainwater characteristic map of the rainwater layer from the currently inputted degradation map containing rainwater to generate a clearer picture, and uses it to calculate the loss from the original clear picture. The supervision network is used to predict clear rainless images, and to calibrate the rainwater features learned by the auxiliary network by generating the attention map. The rainwater features enhanced by the attention map will be transmitted to the main network, and these features enhanced data flows will converge with the data flows of the main network, facilitating the learning of the main network.

### *3.4. Dual attention guided convolution block*

In order to improve the network's ability to focus on key areas and representative features, this paper combines spatial attention[24] and channel attention[25] with convolutional layers to construct a feature dual attention guided convolution block (DAB), which is used as the basic component of the main network ORMNet. The principle of the dual attention guided convolution block is shown in Figure 5. For any given feature map $\hat{X}$, first convert it to $X$ using two convolutional layers. The spatial and channel attention mechanisms used are mainly composed of the global average pooling function (GAP), convolutional layers, and activation functions. They are used to learn the spatial attention map $A_S$ and channel attention map $A_C$. The feature mappings guided by spatial and channel attention are $X_S$ and

$X_C$ , respectively. The specific process is as follows:

$$X_S = \hat{X} \odot A_S \tag{1}$$

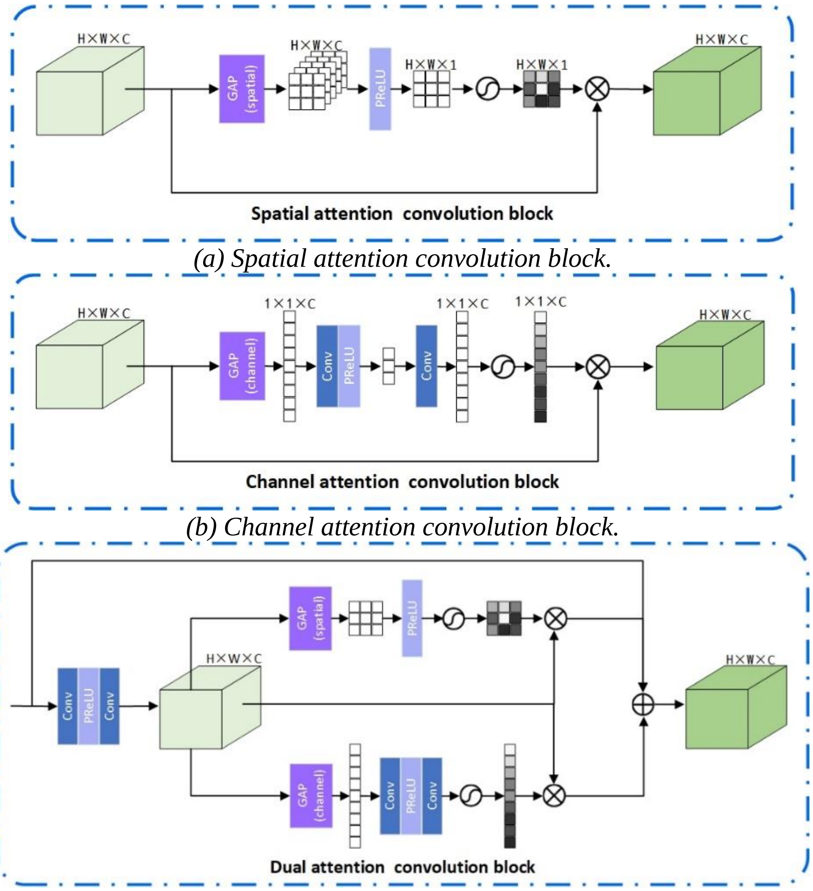
$$X_C = \hat{X} \odot A_C \tag{2}$$

Where $\odot$ represents the element-wise multiplication.

The final output is the aggregation of input and attention directed features:

$$\bar{X} = X + X_S + X_C \tag{3}$$

The original resolution module consists of the dual attention guided convolution blocks and the residual connections. After the rain image is processed by multiple cascaded original resolution modules in the main network, the final rain-free image is generated through the convolutional layer responsible for features reconstruction.



*(a) Spatial attention convolution block.*

*(b) Channel attention convolution block.*

*(c) Dual attention convolution block.*

*Figure 5: Dual attention convolution block and spatial and channel attention convolution block.*

### *3.5. Loss function*

Both the main network ORMNet and the auxiliary network MAGNet are set with a loss function to calculate the loss. The total loss $L_{total}$ is:

$$L_{total} = L_{main} + L_{auxiliary} \tag{4}$$

The network predicts the residual image $R_S$ by learning the characteristics of rainwater, and compares it with the input degraded image $I$ to produce a clear background image $X_B$ , that is $X_B = I - R_S$ . Specifically, for the loss $L_{main}$ of the main network and the loss $L_{auxiliary}$ of the auxiliary network, the following loss functions are constructed to optimize the network:

$$L=\sum_{S=1}^{3}\left[L_{char}\left(X_B,Y\right)+\lambda L_{edge}\left(X_B,Y\right)\right] \tag{5}$$

Where $Y$ represents the original clear background image, and $L_{char}$ represents Charbonnier's loss[22]:

$$L_{char}=\sqrt{\left\|X_B-Y\right\|^2+\varepsilon^2} \tag{6}$$

$L_{edge}$ represents edge loss[26][27]:

$$L_{edge}=\sqrt{\left\|\Delta\left(X_B\right)-\Delta(Y)\right\|^2+\varepsilon^2} \tag{7}$$

Where $\Delta$ represents a Laplace operator, the value of parameter $\varepsilon$ is set to $10^{-3}$, and parameter $\lambda$ controls the relative importance of the losses $L_{char}$ and $L_{edge}$, with the value set to $0.05$ [28].

## 4. Experimental Results and Analysis

In this section, the proposed multi-scale attention map guided image rain removal network is evaluated on Rain100L, Rain100H, and Rain800 datasets. The experiments include ablation experiments and comparative experiments with other methods to verify the good performance of the network proposed in this paper.

### *4.1. Experimental configuration*

The rain removal network model proposed in this paper is based on the PyTorch deep learning framework, and the hardware environment for training and testing is NVIDIA Tesla V100-SXM2. During training, the batch size is set to 16, and training is performed on image patches of size $256\times256$. In order to increase training data, horizontal and vertical flipping are randomly applied to images for data enhancement. Using the Adam optimizer[29] with an initial learning rate of $2\times10^{-4}$, use the Cosine annexing strategy[30] to slowly reduce the learning rate to $1\times10^{-6}$. In addition, for the number of DABs associated with ORB and the number of layers of dilated convolution in MAGNet, this paper analyzes the settings of these two hyperparameters to determine the optimal value. The number of DABs in the ORB can improve the network's ability to learn rainwater characteristics, but excessive numbers will greatly increase the network's parameters amount and operating speed. According to Figure 6, when the number of DABs in the ORB cascade is 4, the network's rain removal effect reaches an ideal level. Similarly, for the number of layers of dilated convolution in MAGNet, combined with Figure 7, it can be seen that when the number of layers is 2, the network achieves an ideal rain removal effect. At this time, the scale of the attention map is appropriate, which is beneficial for the main network to learn the characteristics of rainwater.

In this paper, peak signal to noise ratio (PSNR)[31] and structural similarity index (SSIM)[32] are used as objective evaluation methods for image quality, and numerical values are used to intuitively compare the performance of rain removal methods. During the experiment of removing rain, the clear images with and without rain must be in the same background, and the background corresponding to the rain area must be completely consistent. The data sets Rain100L and Rain100H proposed by Yang et al.[33] are selected, and there are 1800 training images and 100 test images in the Rain100L data set. The background image is selected from the BSD200 dataset[34]. Each image consists of light rain stripes in one direction. Rain100H is also composed of 1800 images and 100 test images. The background is also selected from the BSD200 dataset[34], but the Rain100H rain map contains five types of rain stripes. In addition, the Rain800 proposed by Zhang et al.[35] is also selected, which artificially synthesize a dataset by adding rain stripes with different intensities and directions to the image. The dataset consists of 700 training images and 100 test images. These images are selected from the UCID dataset[36] and the BSD500 dataset[37].

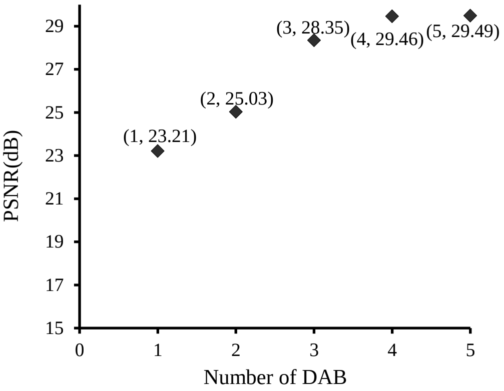


*Figure 6: Rain removal results of different number of DAB modules in the network.*

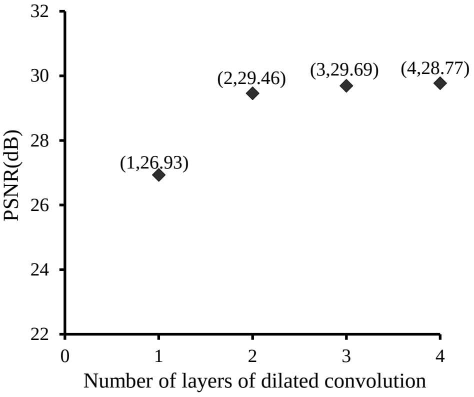


*Figure 7: Rain removal results of dilated convolution with different layers of network.*

### *4.2. Ablation experiment*

In order to analyze the impact of various components of the network on the rain removal results, this paper adds the dual attention guided convolution blocks and the multi-scale attention map generation network to the RIR structured network with traditional attention mechanisms. Based on the quantitative and qualitative evaluation of experimental results, it can be proven that the original RIR structured network with traditional attention mechanisms has a preliminary rain removal ability. The proposed dual attention guided convolution block and multi-scale attention map generation network can both improve the rain removal ability of the network to a certain extent, and when used together, the network can achieve the most ideal rain removal effect as a whole.

The quantitative results are shown in Table 1, where Single Net is the RIR structured network with traditional attention mechanisms, and DAB is the dual attention guided convolution block. Net+DAB is a single net that introduces the dual attention guided convolution blocks, MAGNet is a multi-scale attention map generation network, Net+MAGNet is a single net that introduces the multi-scale attention map generation network, and our method uses both DAB and MAGNet. Based on the analysis in Figure 8, it can be seen that the Single Net has a certain ability to remove rain. However, because it only considers processing the image at the original resolution, it is unable to remove rain stripes of all scales, and there are significant rain stripe residues in the image. Net+DAB and Net+MAGNet have further improved in quantitative evaluation indicators, verifying the effectiveness of the DAB and MAGNet constructed in this paper in terms of rain removal tasks. However, Net+MAGNet has a higher quantitative evaluation index value than Net+DAB, indicating that it contributes more to the improvement of the rain removal ability of the entire network, and that the multi-scale attention map generated by MAGNet can improve the recognition ability of the main network for rain pattern features, Effectively guide the main network to learn rain stripe features. Referring to Figure 8, it can also be found that Net+MAGNet has a stronger ability to restore spatial details and a brighter and clearer background. Due to the simultaneous use of DAB and MAGNet, our method generates clearer images, and the detailed features of the images are further refined, making it almost impossible to detect the presence of residual rain stripes.

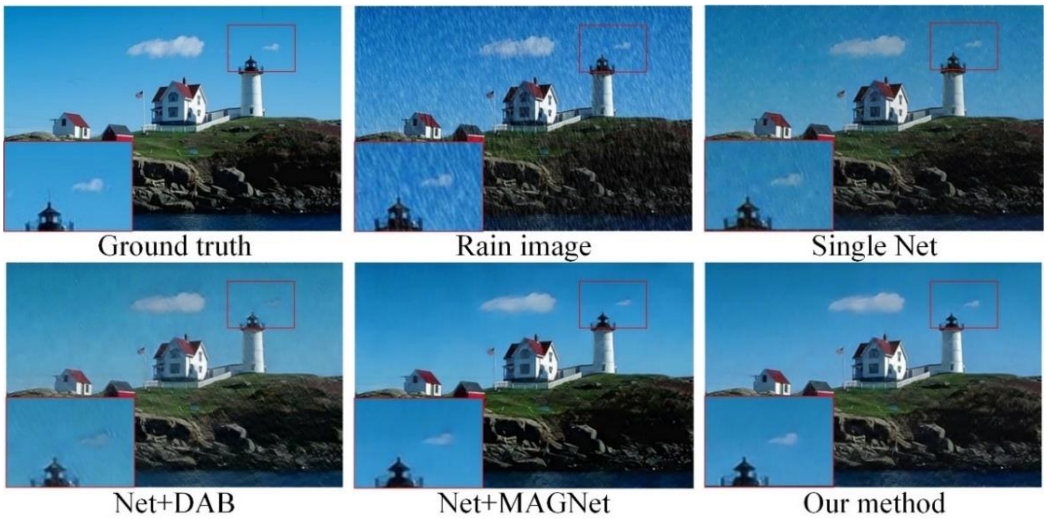


*Figure 8: Rain removal results of different modules of the network.*

*Table 1: Quantitative evaluation results of different modules of the network.*

| Method | Rain100L | | Rain100H | | Rain800 | | Average | |
|---|---|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| Single Net | 32.22 | 0.926 | 28.25 | 0.852 | 26.53 | 0.871 | 29.00 | 0.883 |
| Net+DAB | 33.90 | 0.944 | 28.68 | 0.857 | 28.27 | 0.876 | 30.28 | 0.892 |
| Net+MAGNet | 34.66 | 0.953 | 29.45 | 0.869 | 28.93 | 0.883 | 31.01 | 0.901 |
| **Our method** | **36.37** | **0.969** | **29.56** | **0.883** | **29.46** | **0.896** | **31.79** | **0.916** |

### *4.3. Comparative experiments with other methods*

Compare the GAMNet proposed in this paper with JORDER[12], RESCAN[13], PReNet[18], SPANet[14], and MSPFN[19] on the dataset Rain100L, Rain100H, and Rain800. Experiments have shown that the method proposed in this paper achieves good results on both PSNR and SSIM indicators, and the visual perception of the rain removal results is also optimal, which is superior to the above five methods.

The quantitative evaluation results are shown in Table 2, where Parameters is the parameter quantity of the corresponding network model, and Time (s) is the running time of the network model when the image with size $512 \times 512$ is used as input. As early rain removal networks, JORDER and RESCAN can well cope with rain in simple scenarios. Although the amount of parameters is relatively large, the complexity of the model structure is small, and its running time is relatively small. Based on the analysis in Figure 9, it can be seen that PReNet is a lightweight rain removal network with the smallest amount of parameters and the lowest running time. Although it enhances the communication of information within the model, it is difficult to fully learn the characteristics of rain strips of different sizes only by processing the image at the original resolution, which is only equivalent to the Single Net in the ablation experiment in this paper. SPANet emphasizes the attention to spatial characteristics, thus improving the quality of rain removal to a certain extent compared to PReNet. However, in the PSNR index of Rain800, its value is significantly lower, indicating that the network may have shortcomings in model generalization. As a multi-stage network, MSPFN has the best rain removal effect among the five comparison methods, but its model parameters are too large and the rain removal speed is not very ideal. The model proposed in this paper not only outputs clear results of rain removal, but also has fewer parameters and shorter running time.

In order to further verify the generalization ability of the method proposed in this paper, the rain removal network proposed in this paper was used to test the real rainfall images taken on rainy days. The rain removal results for real scenes are shown in Figure 10. From the visual perspective, it can be seen that the MSAMNet proposed in this paper can effectively remove the rain stripes in the images and improve the image clarity.

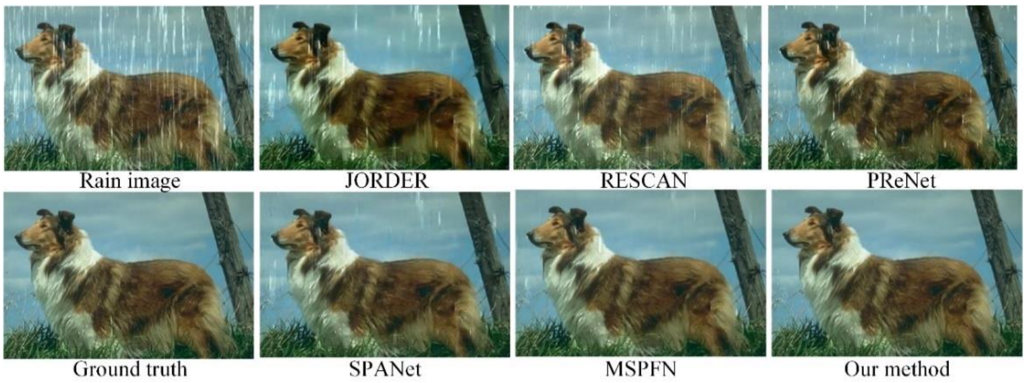


*Figure 9: Rain removal results of different methods.*

*Table 2: Quantitative evaluation results of different methods.*

| Method | Rain100L | | Rain100H | | Rain800 | | Model analysis | |
|---|---|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | Parameters | Times |
| Rain image | 26.71 | 0.843 | 13.79 | 0.367 | 22.18 | 0.663 | - | - |
| JORDER | 31.95 | 0.959 | 22.05 | 0.727 | 22.24 | 0.776 | 4169024 | 0.43 |
| RESCAN | 29.80 | 0.881 | 26.75 | 0.835 | 24.99 | 0.830 | 499668 | 0.61 |
| PReNet | 32.44 | 0.950 | 26.77 | 0.858 | 24.79 | 0.849 | 168963 | 0.156 |
| SPANet | 35.79 | 0.965 | 26.27 | 0.865 | 26.41 | 0.838 | 283716 | 1.72 |
| MSPFN | 32.64 | 0.925 | 27.39 | 0.843 | 27.01 | 0.851 | 15823424 | 1.81 |
| **Our method** | **36.37** | **0.970** | **29.56** | **0.883** | **29.46** | **0.896** | **2553105** | **0.27** |

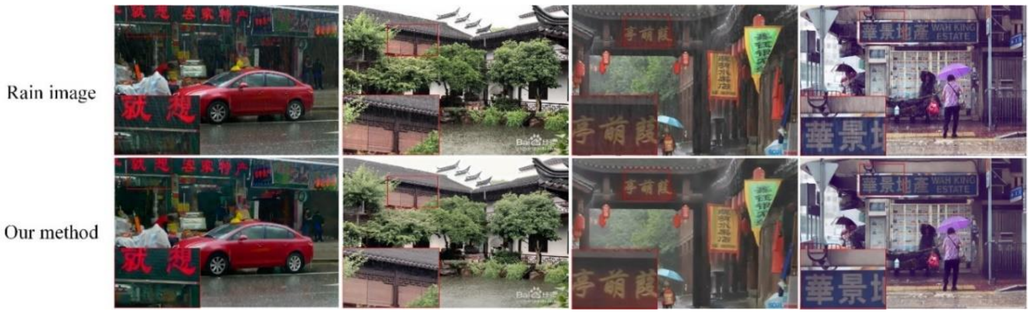


*Figure 10: The results of the method in this paper in the real scene.*

## 5. Summary

This paper proposes a multi-scale attention map guided image rain removal network (MSAMNet). The network consists of two parts, the main network is responsible for feature learning at the original resolution, and the auxiliary network is responsible for generating multi-scale feature attention maps to guide the main network in learning rain stripe features. In order to fully utilize the features learned by the auxiliary network, in addition to sequentially transferring the three scales feature maps generated by the network to the main network, these feature maps are aggregated and transferred to the supervised attention module for processing, generating the original resolution attention map, which guides the main network in learning features at the original resolution. Experimental results show that the method proposed in this paper can well balance the speed and quality of rain removal, and achieve good rain removal effect while ensuring the speed of network rain removal.

## Acknowledgements

This work was supported by the National Natural Science Foundation of China Grant No. 61602157, Henan Science and Technology Planning Program Grant No. 202102210167.

## References

*[1] Sha M, Zeng K, Tao Z, et al. Lightweight Pedestrian Detection Based on Feature Multiplexed Residual Network [J]. Electronics, 2023, 12(4): 918.*

*[2] Le V T, Kim Y G. Attention-based residual autoencoder for video anomaly detection [J]. Applied Intelligence, 2023, 53(3): 3240-3254.*

*[3] Ashiq F, Asif M, Ahmad M B, et al. CNN-based object recognition and tracking system to assist visually impaired people [J]. IEEE Access, 2022, 10: 14819-14834.*

*[4] Santhaseelan V, Asari V K . Utilizing local phase information to remove rain from video[J]. International Journal of Computer Vision, 2015, 112(1):71-89.*

*[5] Zheng X, Liao Y, Guo W, et al. Single-image-based rain and snow removal using multi-guided filter[C]//Neural Information Processing: 20th International Conference, ICONIP 2013, Daegu, Korea, November 3-7, 2013. Proceedings, Part III 20. Springer Berlin Heidelberg, 2013: 258-265.*

*[6] Luo Y, Xu Y, Ji H. Removing rain from a single image via discriminative sparse coding [C] // Proceedings of the IEEE international conference on computer vision. 2015: 3397-3405.*

*[7] Zhang H, Patel V M. Convolutional sparse and low-rank coding-based rain streak removal[C]//2017 IEEE Winter conference on applications of computer vision (WACV). IEEE, 2017: 1259-1267.*

*[8] Li Y., Tan R. T., Guo X., Lu J., & Brown M. S. Rain streak removal using layer priors. In IEEE conference on computer vision and pattern recognition (pp. 2736–2744).*

*[9] Li Y., Tan R. T., Guo X., Lu J., & Brown M. S. Single image rain streak decomposition using layer priors. IEEE Transactions on Image Processing, 26(8), 3874–3885.*

*[10] Fu X, Huang J, Ding X, et al. Clearing the skies: A deep network architecture for single-image rain removal [J]. IEEE Transactions on Image Processing, 2017, 26(6): 2944-2956.*

*[11] Fu X, Huang J, Zeng D, et al. Removing rain from single images via a deep detail network [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 3855-3863.*

*[12] Yang W, Tan R T, Feng J, et al. Deep joint rain detection and removal from a single image[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1357-1366.*

*[13] Li X, Wu J, Lin Z, et al. Recurrent squeeze-and-excitation context aggregation net for single image deraining [C]//Proceedings of the European conference on computer vision (ECCV). 2018: 254-269.*

*[14] Wang T, Yang X, Xu K, et al. Spatial attentive single-image deraining with a high quality real rain dataset [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 12270-12279.*

*[15] He D, Shang X, Luo J. Adherent mist and raindrop removal from a single image using attentive convolutional network[J]. Neurocomputing, 2022, 505: 178-187.*

*[16] Su Z, Zhang Y, Zhang X P, et al. Non-local channel aggregation network for single image rain removal [J]. Neurocomputing, 2022, 469: 261-272.*

*[17] Liang Y, Anwar S, Liu Y. Drt: A lightweight single image deraining recursive transformer [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 589-598.*

*[18] Ren D, Zuo W, Hu Q, et al. Progressive image deraining networks: A better and simpler baseline [C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 3937-3946.*

*[19] Jiang K, Wang Z, Yi P, et al. Multi-scale progressive fusion network for single image deraining [C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 8346-8355.*

*[20] Jiang R, Li Y, Chen C, et al. Two-stage learning framework for single image deraining[J]. IET Image Processing, 2022.*

*[21] Li F, Shen L, Mi Y, et al. DRCNet: Dynamic Image Restoration Contrastive Network[C]//Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX. Cham: Springer Nature Switzerland, 2022: 514-532.*

*[22] Zamir S W, Arora A, Khan S, et al. Multi-stage progressive image restoration[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 14821-14831.*

*[23] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions [J]. arXiv preprint arXiv:1511.07122, 2015.*

*[24] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.*

*[25] Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 286-301.*

*[26] Cho S J, Ji S W, Hong J P, et al. Rethinking coarse-to-fine approach in single image deblurring [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 4641-4650.*

*[27] Jiang L, Dai B, Wu W, et al. Focal frequency loss for image reconstruction and synthesis [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 13919-13929.*
*[28] Tu Z, Talebi H, Zhang H, et al. Maxim: Multi-axis mlp for image processing[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 5769-5780.*
*[29] Kingma D P, Ba J. Adam: A method for stochastic optimization [J]. arXiv preprint arXiv:1412.6980, 2014.*
*[30] Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts [J]. arXiv preprint arXiv:1608.03983, 2016.*
*[31] Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM[C]//the 20th International Conference on Pattern Recognition. IEEE, 2010: 2366-2369.*
*[32] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.*
*[33] Yang W, Tan R T, Feng J, et al. Deep joint rain detection and removal from a single image [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1357-1366.*
*[34] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]//Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. IEEE, 2001, 2: 416-423.*
*[35] Zhang H, Sindagi V, Patel V M. Image de-raining using a conditional generative adversarial network [J]. IEEE transactions on circuits and systems for video technology, 2019, 30(11): 3943-3956.*
*[36] Schaefer G, Stich M. UCID: An uncompressed color image database[C]//Storage and Retrieval Methods and Applications for Multimedia 2004. SPIE, 2003, 5307: 472-480.*
*[37] Arbelaez P, Maire M, Fowlkes C, et al. Contour detection and hierarchical image segmentation [J]. IEEE transactions on pattern analysis and machine intelligence, 2010, 33(5): 898-916.*