



ARTICLE

Single Image Deraining Using Dual Branch Network Based on Attention Mechanism for IoT

Di Wang, Bingcai Wei and Liye Zhang*

School of Computer Science and Technology, Shandong University of Technology, Zibo, 255000, China

*Corresponding Author: Liye Zhang. Email: zhangliye@sdut.edu.cn

Received: 23 December 2022 Accepted: 13 February 2023 Published: 28 June 2023

ABSTRACT

Extracting useful details from images is essential for the Internet of Things project. However, in real life, various external environments, such as bad weather conditions, will cause the occlusion of key target information and image distortion, resulting in difficulties and obstacles to the extraction of key information, affecting the judgment of the real situation in the process of the Internet of Things, and causing system decision-making errors and accidents. In this paper, we mainly solve the problem of rain on the image occlusion, remove the rain grain in the image, and get a clear image without rain. Therefore, the single image deraining algorithm is studied, and a dual-branch network structure based on the attention module and convolutional neural network (CNN) module is proposed to accomplish the task of rain removal. In order to complete the rain removal of a single image with high quality, we apply the spatial attention module, channel attention module and CNN module to the network structure, and build the network using the coder-decoder structure. In the experiment, with the structural similarity (SSIM) and the peak signal-to-noise ratio (PSNR) as evaluation indexes, the training and testing results on the rain removal dataset show that the proposed structure has a good effect on the single image deraining task.

KEYWORDS

Internet of Things; image deraining; dual-branch network structure; attention module; convolutional neural network

1 Introduction

In order to make traffic management more efficient and road traffic safer, the Internet of Things (IoT) and intelligent transportation technology have attracted the attention of researchers. In order to ensure the safety of users, one of the key technologies is to realize smart cities, intelligent transportation, and the interconnection between things. In the actual application scenario of the Internet of Things, the collection, transmission and analysis of various perceptual data and the control of object information data are one of the main problems that need to be tackled in the IoT technology [1,2]. There are many different technologies for data collection. For example, in vehicle networking technology, vehicle, road and personnel information is collected through sensors such as radar and camera to achieve real-time monitoring of road traffic condition, and information sharing with other vehicles using various communication technologies. In the process of development, the



Internet of Vehicles technology can apply a variety of communication technologies to conduct data interconnection, establish the relationship between vehicles, people and vehicles, and ensure the normal operation of healthy transportation in smart cities.

Among all data acquisition technologies, computer vision technology is superior to other technologies in terms of cost, interaction and security. Especially in recent years, due to the ability of deep learning algorithm to learn and adapt to different conditions, the development of deep learning algorithms makes computer vision more brilliant. The use of computer vision can process image information more efficiently and quickly, which is greatly changing the Internet of Things industry. For example, in the Internet of Vehicles (IoV) system, vision based traffic information extraction has become one of the indispensable functions of vehicles. Using computer vision technology, key traffic information is extracted from the acquired images, such as pedestrian and vehicle recognition and detection, street obstacle recognition and route planning. In the monitoring system, computer vision can be used to achieve face recognition, vehicle license plate information and other functions. For car cameras and outdoor surveillance cameras, when affected by bad weather, such as rain, the image quality taken by the camera will be seriously degraded, which is very unfavorable to feature learning and may lead to computer vision system failure. Therefore, it has important to process the degraded image and realize image rain removal.

In order to reduce the influence of rain grain on images, researchers began to study the deraining technology and have contributed a lot of ideas. Single image rain removal algorithm [3] was first proposed in 2012. In reference [4], Sun et al. proposed a convolution neural network with rainy images as input, which can directly recover clear images in the case of atmospheric veiling effects caused by distant rain-streak accumulation. Compared with video rain removal, a single image does not have the help of context time domain information, which is more challenging in the rain removal task. At present, there are two categories of related rain removal algorithms, one is the image rain removal algorithm based on the traditional way, and the other is the image rain removal algorithm based on the deep learning way. As a traditional way of image rain removal, reference [5] removed rain lines by sparse coding and dictionary learning. Reference [6] proposed a Gaussian mixture model, which uses prior knowledge based on blocks to calculate the different directions and scales of rain patterns. It can effectively solve the problem of excessive smoothing and rain pattern residue in sparse coding algorithms. In reference [7], kernel regression and non-local average filtering were used to detect and remove rain streaks. In 2018, Deng et al. [8] established a global sparse model considering the directionality and structure knowledge of rain grain. These traditional methods are based on low-level image features to separate rain grain, and the effect is not ideal.

In recent years, with the development of hardware technology and deep learning, deep learning has been widely used in computer vision tasks of rain removal. In the deep learning algorithm, the image is directly taken as the input, and the hierarchical features of the image are learned through convolution or other nonlinear extraction stages, so deep learning has a strong learning ability. Yang et al. [9] proposed the JORDER model, which used a context expansion network to obtain more image features, and this model had a better removal effect in the case of heavy rain. Fu et al. [10] proposed the DetailNet network model, which used guided filtering to divide the image into high-frequency and low-frequency layers. Then used a convolutional network to learn the features of high-frequency layers to achieve the effect of rain-grain removal. Zhang et al. [11] applied a conditional generative adversarial network [12] to image rain removal and used constraints to make the generated image more realistic. In addition to the convolutional neural network model, the application of the attention mechanism [13] in computer vision has also greatly improved the results of many tasks.

Attention mechanism is the development trend in the field of image processing. Reference [14] proposed a multi-scale feature fusion image rain removal method based on attention mechanism. The feature extraction stage is composed of multiple residual groups containing two multi-scale attention residual blocks. The multi-scale attention residual block uses the multi-scale feature extraction module to extract and aggregate the feature information of different scales, and further improves the feature extraction capability of the network through coordinate attention. Reference [15] established a rain-removal model based on space-frequency domain, and designs a multi-layer channel attention module to map the weight information of the rain line layer, enhance important features, mine the brightness difference between the rain line layers, and improve the rain line detection performance. Xu et al. [16] proposed a single image rain removal algorithm based on the joint attention mechanism, which can fully explore the interaction of attention mechanisms in different dimensions.

As mentioned above, the self-attention module, convolutional neural network module, and generative adversarial network can all carry out deep learning tasks. In this paper, a double branch parallel network structure of attention mechanism and residual network is used to remove rain grain. The attention mechanism adopts two modules: spatial attention mechanism and channel attention mechanism. Different attention modules will capture different feature information, which makes the final information feature more abundant.

2 Related Work

2.1 Rain Removal Model

The early additive composite model [17] (ACM) can be expressed as follows:

$$O = B + \tilde{S} \quad (1)$$

where O represents the composite image degraded by the rain grain, B represents the clean background layer, and \tilde{S} represents the rain grain layer. This model simply superimposes the background layer and the rain-grain layer without considering the situation of rain accumulation, which will cause excessive smoothing. Later, the screen blend model [15] (SBM) was proposed:

$$O = 1 - (1 - B) \circ (1 - S) = B + S - B \circ S \quad (2)$$

where \circ represents a point-wise multiplication operation. According to the reference [15], the background layer and the rain layer in the image are not linear superposition, and the mutual influence between the background layer and the rain layer is considered, which is more consistent with the real visual characteristics. Later, people made improvements and proposed the following model:

$$O = \alpha \left(B + \sum_{t=1}^s \tilde{S}_t R \right) + (1 - \alpha) A \quad (3)$$

where \tilde{S}_t is the same direction and shape of the rain. s is the number of superposition, α is the atmospheric transmission. The model treats areas with and without rain differently, with R representing some values of 0 or 1, 0 representing areas without rain and 1 representing areas with rain. This preserves more detail and gives a sharper image while removing rain lines.

2.2 Attention Mechanism

The attention mechanism was first applied in the field of natural language processing. In recent years, it has been shown to be able to improve the performance of network models. The attention mechanism learns a specific mask for the feature map and applies the dot product to the feature map

to get the image features. In this paper, we use two attention models: spatial attention mechanism and channel attention mechanism.

2.2.1 Spatial Attention Model

The spatial attention model will look for the relatively important information areas in the network to process. Focus more on the target area and less on the background area. In the spatial attention mechanism, the global average pooling operation (GAP) [17] and the global maximum pooling operation (GMP) are used to generate two kinds of image features. After the two features are concatenated and merged, the feature fusion is completed by a convolution operation. Finally, the weighted graph is generated by the Sigmoid activation function, and the final feature is calculated by the dot product of the original feature. The spatial attention model is shown in Fig. 1. The spatial attention mechanism compresses the channel, and the global average pooling and maximum pooling are carried out on the channel dimension. The Maximum pooling (MaxPool) operation extracts the maximum value on the channel, and the extraction times are height * width. Average pooling (AvgPool) takes the average value on the channel, and the extraction times are also height * width, and then combines the previously mentioned features to get a 2-channel feature map. The spatial attention mechanism can be expressed follows:

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \quad (4)$$

where σ represents the Sigmoid operation, and $7 * 7$ represents the size of the convolution kernel.

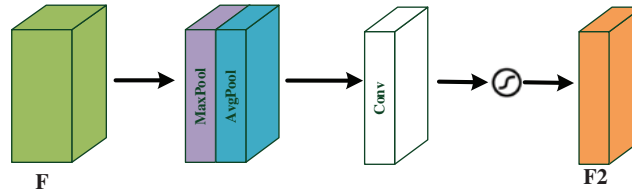


Figure 1: The spatial attention model

2.2.2 Channel Attention Model

The channel attention mechanism acts on the channel domain and pays different attention to different channels of the feature map. Because the spatial attention mechanism focuses on local features and cannot make good use of the overall features of the image, the channel attention is used to consider the interdependence between channels and adaptively repartition channel features. The running process of the channel attention module is shown in Fig. 2. Input feature map F, after global average pooling and global maximum pooling operations based on width and height, respectively, and then through Multilayer Perceptron (MLP), respectively, the output of the features by MLP are added based on elements, and then through the sigmoid activation function, finally generate channel attention weight. Therefore, the channel attention mechanism can be expressed as follows:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1(W_0(F_{avg}^c) + W_0(F_{max}^c))) \quad (5)$$

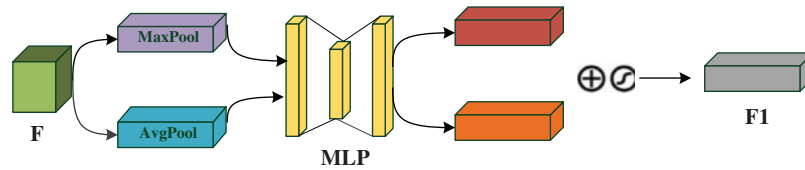


Figure 2: The channel attention model

The channel attention mechanism compresses the feature map on the spatial dimension, obtain the one-dimensional vector, and then operate.

2.3 Loss Function

A loss function that measures the difference between the model’s predicted and true values. The choice of the loss function is related to the model’s performance. Therefore, different models using different loss functions will lead to different results. In this paper, we use a combination of multiple loss functions to calculate the loss of the model.

MSE loss is widely used in deep learning, one of the most basic loss functions based on regression. This function can be expressed as:

$$L_{mse}(S, Y) = \frac{1}{N} \sum_{t=1}^N (S_t - Y_t)^2 \tag{6}$$

where S is the predicted value and Y is the ground truth. The function curve of MSE loss is very smooth and continuously derivable, and the gradient descent algorithm can be used. The continuous decrease of the difference, the gradient will also decrease, which is conducive to convergence. However, the MSE loss is greatly affected by outliers, which leads to the degradation of the performance of the model. To reduce this effect, we use Perceptual Loss. In the perceptual loss, the real value and the predicted value are transmitted through the pre-trained neural network, and the MSE of its middle layer is calculated as the loss. Perceptual loss is a kind of L2 loss based on the difference in the CNN feature map between the generated image and the target image. The perceptual loss is defined by the output characteristics of a certain layer of the pre-trained network:

$$L_x = \frac{1}{W_{ij}H_{ij}} \sum_{x=1}^{W_{ij}} \sum_{y=1}^{H_{ij}} \left\{ \begin{matrix} \Phi_{ij}(I^S)_{x,y} \\ -\Phi_{ij}(G_{\theta G}(I^B))_{x,y} \end{matrix} \right\}^2 \tag{7}$$

where Φ_{ij} represents the feature map extracted through the pre-trained convolutional neural network, and the pre-trained network model used in this paper is VGG16. W_{ij} and H_{ij} represent the size of the feature map.

The perceptual loss can extract the features of the image and try to minimize the differentiation between the images. The combination of sensing loss and MSE loss can improve the model’s overall performance.

In addition, SSIM loss [18] is also used in this paper. Structural similarity is an index to measure the similarity of two images. It can be expressed as Eq. (8):

$$SSIM(X, Y) = \frac{2\mu(X, Y) + c}{(\mu(X) + \mu(Y))^2 + c} \frac{2\sigma(X, Y) + s}{(\sigma(X) + \sigma(Y))^2 + s} \tag{8}$$

$l(X, Y)$ stands for brightness, $c(X, Y)$ stands for contrast, and $s(X, Y)$ stands for structure, as shown in the following Eq. (9):

$$\begin{aligned}
 l(X, Y) &= \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \\
 c(X, Y) &= \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \\
 s(X, Y) &= \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}
 \end{aligned} \tag{9}$$

where μ is the average value, σ is variance, σ_{xy} is the covariance.

In this experiment, the best performance of the model is achieved when MSE loss, perception loss, and structural similarity loss are used simultaneously. The ablation experiments will be described in detail in the following sections.

3 The Proposed Method

3.1 Branch Attention Module

This paper proposes a Branch Attention Module (BAM), as shown in Fig. 3. The module consists of convolution layer, activation function, spatial attention module and channel attention module. In network structure, the last stage of extracted features are fed into the branch of attention after the module. They first pass through the convolution with a convolution core size of $3 * 3$ and use the PReLU activation function. And then pass through a layer of a convolution again. They enter different branches according to the source of the features. If the incoming features come from the spatial attention module, they pass through the spatial attention mechanism again; If the incoming feature comes from the channel attention module, it will enter the channel attention mechanism. Finally, it will be added with the features of the previous stage element by element to enter the next module. After a series of ablation experiments, it is proved that the module is beneficial in improving the model's performance. After a series of ablation experiments, it is proved that the module is beneficial in improving the performance of the model.

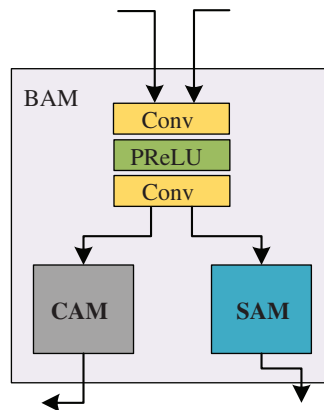


Figure 3: The structure diagram of BAM

3.2 Skip Connect

Fig. 4 is a schematic from ResNet [19] block. A nonlinear variation function can be used to describe the input and output of the network. The input is x , and the outcome is $F(x)$. F usually includes operations such as convolution and activation. The idea of skip connection is to express the output as a superposition of an input and a nonlinear transformation of the input. To a certain extent, the more layers of the network, the stronger the ability of the network to extract features and the better the performance. However, with the deepening of network depth, there will be many problems, such as gradient dissipation, gradient explosion and so on. Before using residual join, there are many ways to optimize these problems, such as better initialization strategy, various activation functions, etc., but they still need to be more, and the ability to improve the situation is limited until the residual join is widely used. Skip Connect ameliorates the gradient dissipation problem during back propagation, thus making it easy to train deep networks. In addition, the residual connection also breaks the network's symmetry and improves the network's representation ability.

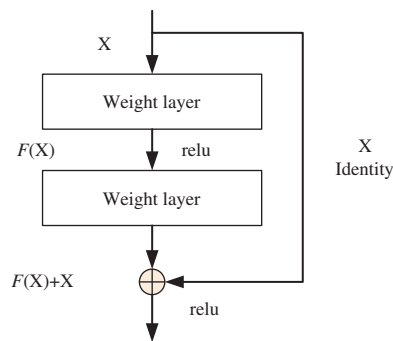


Figure 4: The structure of residual block

Based on the powerful representation ability of residual connection, we also introduce residual connection in the design of the network structure, which connects the features of each layer of the attention module with each layer of the final decoder. And the last layer directly connects the input features with the global features. See the overall structure diagram for its structure.

3.3 The Network Structure

Input a degraded image, shallow features in an image with convolution first after extraction, attention will feature respectively into space module and channel attention module for feature extraction respectively, and the extracted features into BAM module, in BAM module, according to the different way of attention mechanism, selection of different feature extraction methods. The features of BAM module are added to the features of the upper layer and passed to the next module as input. Cycle until the module of the last layer (the three layers), the last layer obtained from BAM in do and features are no longer as before, but for three layers of convolution mapping, the portion of the sampling scale factor of 2, model for bilinear model, and through a convolution kernel size is 3×3 layers of convolution, repeated three times, scale the feature size to equal the original feature size. Finally, the rain removal image after deep learning feature extraction is obtained.

The whole of neural network uses U-Net architecture, which was first applied in the field of medical image processing and was proposed at the MICCAI Conference in 2015. The adopted coder (down sampling)-decoder (up sampling) structure and jump connection are a very classic design method. The core of network structure design in this paper still continues the core idea of U-Net.

The model can extract different spatial scale features when processing image features. At the same time, low-resolution image features and up-sampling high-resolution features are fused and stacked through skip connection, and finally rich multi-scale features are obtained.

The structure of the Dual-branch deraining network based on the attention mechanism and CNN is shown in Fig. 5.

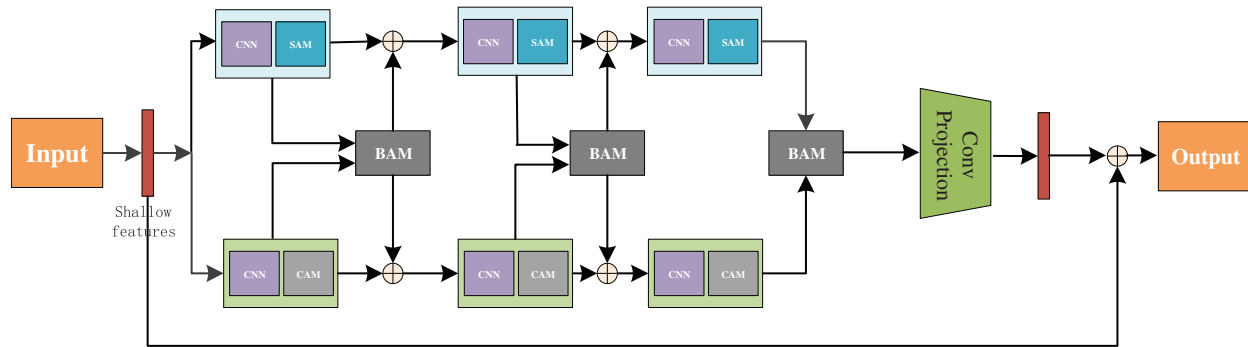


Figure 5: The network structure

4 Experimental Process and Analysis

4.1 The Experimental Details

The comparison experiment and ablation experiment in this paper were performed on the server NVIDIA RTX3060. The server system is windows 10. The programming language used in this experiment is Python, and the pytorch framework is used for training and testing. Adam [18] optimizer was used to optimize the network model in both the comparison and the ablation experiments. In the ablation experiment, the initial learning rate was set as $2e-4$, the training period was set as 100 rounds, and the batch size was set as 3. Warm-up was used to prevent network over fitting. Cosine annealing is used to decay the learning rate. After several iterations, the learning rate finally decays to $1e-6$. At the beginning of training, the image is preprocessed to increase the performance and generalization ability of the network model. After several iterations, the final training data is obtained.

4.2 Evaluation Index and Data Set Information

We selected peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) as evaluation indicators in the experiments. The higher the values of PSNR and SSIM, the better the image's rain-removal effect. In selecting of data sets, because it is difficult to obtain the corresponding rain images and rain-free images in the real world, synthetic public data sets are used for testing and training, and Rain1800 [17] dataset is used for training Test100 [18], Rain100H, Rain100L, Test1200 [18] datasets were used for testing. The dataset Rain100L has 200 pairs of images for training, 100 pairs for testing, and only one type of rain marking. The dataset Rain100H has 1800 pairs of images for training, 100 pairs for testing, and five different types of rain patterns. The dataset Rain800 has 700 pairs of training images and 100 pairs of test images, and the rain removal effect of real-world images and rainy images was compared with other algorithms.

4.3 Ablation Experiments

In this paper, we replace the network structure with different network modules, channel attention model and spatial attention model and compare them with other data sets. There are only convolutional neural network module, convolutional neural network and BAM, channel attention module and convolutional neural network and BAM, spatial attention module plus convolutional neural network and BAM, channel attention and spatial attention, and BAM, whose performance is shown in [Table 1](#).

Table 1: Performance of different combination modules in data sets

	CNN + CNN PSNR/SSIM	CNN + CNN + BAM PSNR/SSIM	CAM + CNN + BAM PSNR/SSIM	SAM + CNN + BAM PSNR/SSIM	SAM + CAM + BAM PSNR/SSIM
Rain100H	26.13/0.783	26.07/0.783	26.16/0.783	25.92/0.777	26.14/0.786
Rain100L	28.69/0.887	29.40/0.896	29.28/0.890	28.82/0.886	29.70/0.899
Test100	21.88/0.762	22.02/0.764	22.04/0.763	21.97/0.763	22.16/0.770

Taking the equivalent signal-to-noise ratio and structural similarity as indicators, we can see each network module's performance from [Table 1](#). When the spatial attention module, channel attention module and BAM module are applied at the same time, the performance effect is the best. In [Fig. 6](#), pictures are used to show the rain removal effect of each module more intuitively.

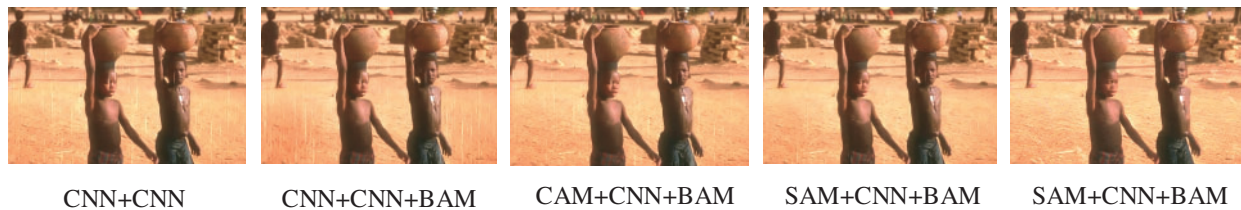


Figure 6: Comparison of rain removal effects of different combination modules

In addition to the ablation experiment of the network structure, we also compared a variety of loss functions for the selection of loss functions. They include the MSE loss function, perception loss function and structural similarity loss function and are compared in different data sets. The results are shown in [Table 2](#).

Table 2: Performance of each loss function on the rain removal data set

	MSE Loss PSNR/SSIM	MSE + Prep. Loss PSNR/SSIM	MSE + SSIM Loss PSNR/SSIM	MSE + Prep. + SSIM Loss PSNR/SSIM
Rain100H	26.14/0.786	26.15/0.805	26.29/0.809	26.25/0.812
Rain100L	29.70/0.899	29.83/0.902	29.83/0.903	29.95/0.906
Test100	22.16/0.770	22.06/0.774	22.07/0.777	22.17/0.777

In the table, we can find that when only MSE loss is used, the peak signal-to-noise ratio and structural similarity are low, the final experimental results show that the network model has the best performance when MSE loss, perception loss and structural similarity loss are used simultaneously. [Fig. 7](#) shows the effect of rain removal after different loss functions are combined.

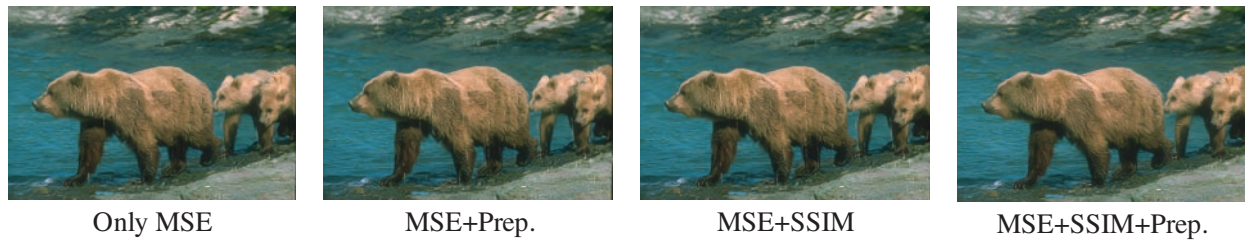


Figure 7: Comparison of rain removal effects under different loss functions

4.4 Comparative Experiment

In the comparative experiment, we compared with a variety of experimental algorithms, including DerainNet [20], DDC [10], DIDMDN [21], SEMI [22], RESCAN [23], UMRL [24], PreNet [25], MSPFN [26] and other rain removal algorithms. The comparison chart of rain removal between the proposed algorithm and other algorithms is shown in Fig. 6, and the specific experimental data are shown in Table 3.

Table 3: Comparison diagram of the proposed algorithm and other algorithms

	Test100 PSNR/SSIM	Rain100H PSNR/SSIM	RAain100L PSNR/SSIM	Test1200 PSNR/SSIM
DerainNet [20]	21.90/0.837	13.67/0.573	26.36/0.873	22.24/0.848
DDC [10]	22.63/0.825	14.51/0.499	26.75/0.858	27.59/0.882
DIDMDN [21]	21.56/0.811	16.31/0.556	23.71/0.804	27.00/0.883
SEMI [22]	21.39/0.781	15.50/0.519	24.05/0.820	24.95/0.841
RESCAN [23]	23.09/0.830	24.86/0.783	27.46/0.864	27.14/0.869
UMRL [24]	23.92/0.883	24.85/0.835	27.73/0.929	29.59/0.922
PreNet [25]	24.03/0.872	25.75/0.861	31.64/0.949	30.86/0.926
MSPFN [26]	26.97/0.898	27.42/0.864	31.66/0.921	31.59/0.928
Ours	27.40/0.905	27.64/0.875	31.82/0.947	31.34/0.925

Through comparative experiments, it is found that our method is superior to the previous algorithms in the two indexes of peak signal-to-noise ratio and structural similarity. Our proposed two-branch structure performs better in rain removal tasks.

5 Conclusion

In the Internet of Things, image feature extraction based on computer vision is one of the important data sources. In order to reduce the fuzzy effect of rainy weather on images, this paper proposed a dual branch deraining model based on attention mechanism and neural network, and proposed BAM fusion branch module in the network model. The paper also introduced the related works and models involved, such as the development of the rain removal model, the composition of spatial attention and channel attention, the used loss function, and the overall architecture of the network model. A series of ablation experiments and comparison experiments with other algorithms prove the validity and rationality of this model. The attention mechanism-based network will become

the trend of future rain removal tasks, and I will continue to study this task, hoping to get a network structure with good rain removal results on a lightweight basis.

Funding Statement: This work is supported by the National Natural Science Foundation of China (No. 62001272).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Liu, X., Zhang, X. (2019). NOMA-Based resource allocation for cluster-based cognitive industrial Internet of Things. *IEEE Transactions on Industrial Informatics*, 16(8), 5379–5388. <https://doi.org/10.1109/TII.9424>
2. Liu, X., Zhang, X. (2018). Rate and energy efficiency improvements for 5G-based IoT with simultaneous transfer. *IEEE Internet of Things Journal*, 6(4), 5971–5980. <https://doi.org/10.1109/JIoT.6488907>
3. 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/TIP.2011.2179057>
4. Sun, H., Ang, M. H., Rus, D. (2019). A convolutional network for joint deraining and dehazing from a single image for autonomous driving in rain. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 962–969. Macau, China.
5. Luo, Y., Xu, Y., Ji, H. (2015). Removing rain from a single image via discriminative sparse coding. *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 3397–3405. Santiago, Chile.
6. 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 (CVPR)*, pp. 2736–2744. Las Vegas, NV, USA.
7. Kim, J. H., Lee, C., Sim, J. Y., Kim, C. S. (2013). Single-image deraining using an adaptive nonlocal means filter. *2013 IEEE International Conference on Image Processing*, pp. 914–917. Melbourne, VIC, Australia.
8. Deng, L. J., Huang, T. Z., Zhao, X. L., Jiang, T. X. (2018). A directional global sparse model for single image rain removal. *Applied Mathematical Modelling*, 59(7), 662–679. <https://doi.org/10.1016/j.apm.2018.03.001>
9. Yang, W., Tan, R. T., Feng, J., Liu, J., Guo, Z. et al. (2017). Deep joint rain detection and removal from a single image. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1685–1694. Honolulu, HI, USA.
10. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X. et al. (2017). Removing rain from single images via a deep detail network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1715–1723. Washington DC, USA.
11. Zhang, H., Sindagi, V., Patel, V. M. (2020). Image de-raining using a conditional generative adversarial network. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(11), 3943–3956. <https://doi.org/10.1109/TCSVT.76>
12. Mirza, M., Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
13. Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L. et al. (2017). Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 6000–6010. Long Beach, California, USA.
14. Liu, Z. Y., Zhou, J., Lu, J. X., Miao, Z. L., Shao, G. F. et al. (2022). Multi-scale feature fusion image rain removal method based on attention mechanism. *Journal of Nanjing University of Information Engineering (Natural Science Edition)*, 1–11 (in Press).
15. Yang, Q., Yu, M., Fu, Q., Yan, G. (2022). Image rain removal based on frequency doubling convolution and attention mechanism. *Control and Decision*, 1–9 (in Press).

16. Xu, C., Yan, Q., Li, T. (2022). Single image rain removal algorithm based on joint attention mechanism. *Computer Applications*, 42(8), 2578–2585.
17. Lin, M., Chen, Q., Yan, S. (2013). Network in network. arXiv preprint arXiv:1312.4400.
18. 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>
19. He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778. Las Vegas, NV, USA.
20. Zhang, H., Patel, V. M. (2018). Density-aware single image de-raining using a multi-stream dense network. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 695–704. Salt Lake City, UT, USA.
21. 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>
22. Wei, W., Meng, D., Zhao, Q., Xu, Z., Wu, Y. (2019). Semi-supervised transfer learning for image rain removal. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3877–3886. Long Beach, CA, USA.
23. 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 (ECCV)*, pp. 254–269. Munich, Germany.
24. Yasarla, R., Patel, V. M. (2019). Uncertainty guided multi-scale residual learning-using a cycle spinning CNN for single image de-raining. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8405–8414. Long Beach, CA, USA.
25. Ren, D., Zuo, W., Hu, Q., Zhu, P., Meng, D. (2019). Progressive image deraining networks: A better and simpler baseline. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3932–3941. Long Beach, CA, USA.
26. Jiang, K., Wang, Z., Yi, P., Chen, C., Huang, B. et al. (2020). Multi-scale progressive fusion network for single image deraining. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8346–8355. Warangal, India.