

Research and Implementation of Image Rain Removal Based on Deep Learning

Dong Wang

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
e-mail:queerwd@163.com

Zhongsheng Wang

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
e-mail:wzhsh1681@163.com

Abstract—The traditional rain removal algorithm needs to optimize a large number of parameters, and it is only effective for rain of a specific shape, and the model generalization ability is poor. In recent years, the performance of rain removal methods based on deep learning is better than many traditional methods, but there are problems such as incomplete or excessive rain removal, and incomplete texture reconstruction of background details. This paper proposes a rain removal network based on generative confrontation, which connects the high and low frequency parts and integrates them into the model. At the same time, the attention mechanism cyclic neural network is organically combined, which can better preserve the background texture while removing rain. Theoretical can produce better rain streak removal with better color distortion.

Keywords—*Deep Learning; Generative Confrontation; High and Low Frequency; Attention Mechanism; Background Texture*

I. INTRODUCTION

The network design motivation and method are briefly described as follows. The algorithm mainly includes three parts: frequency decomposition module, generator network [1], and discriminator network. These three parts will be introduced separately below.

A. Frequency decomposition module

When training the mapping from rain images to no rain images directly on the entire image domain, the mapping range covers all possible pixel values, and the rain trace information and background information in the image will be highly aliased, making it difficult to combine the rain trace information with the background information. The information is accurately separated, causing the background information of the reconstructed image to be blurred, and there are still some rain marks left, and with the deepening of the network depth, the phenomenon of gradient disappearance may occur. Considering that almost all the information of rain marks exists in the high-frequency part of the image, the image with rain and the image without rain are decomposed into two parts: high-frequency layer and (detail layer) and low-frequency layer (base layer) respectively through guided filters, as shown in the formula: (1) and (2). Compared with the traditional bilateral filter, the use of the guided filter to decompose the image has the advantages of higher computational efficiency and stronger edge protection characteristics, and the guided filter is also more accurate in the processing of image details in terms of filtering effect.

$$X = X_{base} + X_{detail} \quad (1)$$

$$Y = Y_{base} + Y_{detail} \quad (2)$$

Figure 1 shows an example of image frequency decomposition. It can be found by observation that the edge contour information and rain trace

information of the image are preserved in the detail layer of the image with rain [2], while the base layer of the image with rain is similar to the base layer of the image without rain. , only there is a difference in the image detail layer.

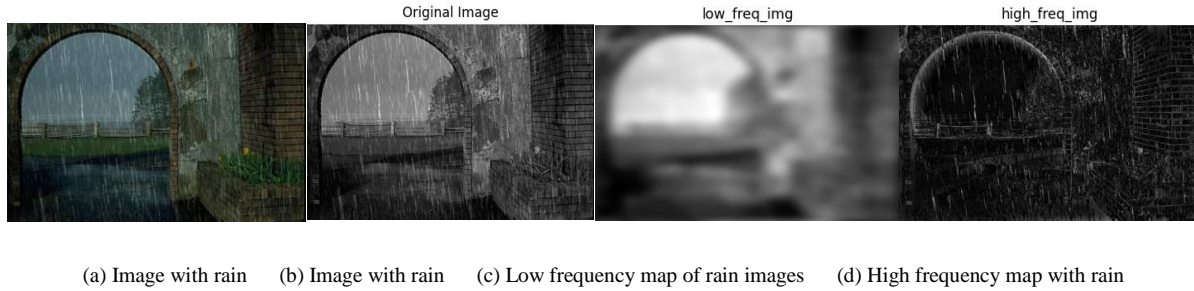


Figure 1. Example of Frequency Decomposition

It can be seen from the comparison results that the rain image is decomposed into high-frequency regions and low-frequency regions according to the frequency, and most of the rain marks exist in the high-frequency region. The low frequency map is integrated into the generator as an additional constraint for better results. The frequency decomposition module is shown in Figure 2.

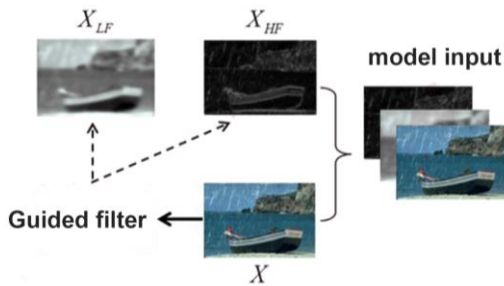


Figure 2. Frequency Decomposition Module

B. Generator network

The generator network proposed in this paper is shown in Figure 4 and is mainly divided into four

parts: (1) a convolutional layer receives the input image; (2) the combination of several residual blocks ResBlocks is used to extract deep feature information; (3)The recurrent unit adopts the long short-term memory unit LSTM, and this stage takes the output of the input and the state of the recurrent layer of the previous stage as the input; (4) A convolutional layer outputs the rain-removed result image. Represented by the following formula:

$$x^{t-0.5} = f_{in}(x^{t-1}, y) \quad (3)$$

$$s^t = f_{recurrent}(s^{t-1}, x^{t-0.5}) \quad (4)$$

$$x^t = f_{out}(f_{res}(s^t)) \quad (5)$$

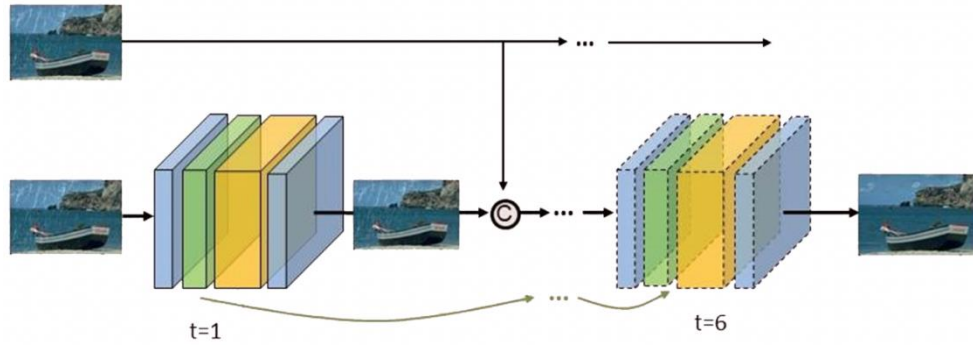


Figure 3. The network structure of the generator

- All convolution kernel size= 3×3 and padding= 1×1 , and the activation function uses ReLU. In the whole process of the network, there is no up-sampling operation, and other settings of the convolution kernel are designed to keep the resolution of the feature map unchanged, and the restored clear image will not lose the details of the image content. The first convolution kernel consists of a Conv+ReLU; the residual block consists of 5 ResBlocks, each ResBlock consists of two Conv+ReLU and an attention mechanism module; the last convolution layer consists of a Conv, no activation function.
- MSE loss: Calculates the output of the last stage (for example, the last optimization stage is T) and the MSE of the ground truth.

$$L_{MSE} = \|x^T - x^{gt}\|^2 \quad (6)$$

- Negative SSIM loss: Calculates the output of the last stage (such as the last optimization stage is T) and the ground truth SSIM.

$$L_{SSIM} = -SSIM(x^T, x^{gt}) \quad (7)$$

C. Discriminator network

In the generative adversarial network [3], the generator is responsible for generating the corresponding data, the purpose is to "fool" the discriminator, the role of the discriminator is to judge whether the input data is real or generated, and the purpose is to find out the "fake data" generated by the generator. Through continuous training, the ability of the discriminator of the generator is getting stronger and stronger, and the final generator can better realize the corresponding network function. Generative adversarial networks have great potential because they can learn and simulate the distribution of any data.

As mentioned above, although the network function is implemented by the generator, and only the generator is used in the testing process, the quality of the discriminator is crucial to the performance of the generator. If the discriminator is not set properly, after a small amount of training, the "fake data" generated by the generator is enough to make the fake data look real, and the discriminator cannot judge the real and fake data. At this time, although the generator's data is still "bad", due to the referee's standard is low, it is easy to produce "proud" mentality, and the performance cannot be further improved.

In standard generative adversarial networks, the discriminator network distinguishes whether the

input image is a real image or a fake image generated by the generator by performing binary classification, and the generator is trained to generate a fake image to convince the discriminator that it is real. Whereas in relative generative adversarial networks, the discriminator is also trained to reduce the probability that a real image is real, and the discriminator estimates the probability that a real sample is more real than a fake one.

II. EXPERIMENTAL SETUP

A. Existing Synthetic Rain Image Datasets

Since it is very difficult to obtain a large number of rain images and corresponding rain-free background images from the real world, this paper uses five commonly used synthetic benchmark rain image datasets [6], Rain100L, Rain100H, Rain800, Rain12600 and Rain12 to train and evaluate the proposed algorithm. Rain Network[4].

B. Real Rain Image Datasets

Internet-Data: Contains 149 real rainy images in total, and Figure 7 shows three typical real rainy scene images.

C. Production of scene-based depth rain image dataset

Most of the existing synthetic rain image datasets and rain removal methods are implemented based on the rain image model of Equation (8):

$$I(x) = B(x) + R(x) \quad (8)$$

Where $B(x)$ is the clean background image without rain, $R(x)$ is the additional rain streak image, and $I(x)$ is the composite rain streak

image. In heavy rain conditions, where the accumulation of rain streaks can cause attenuation and scattering, the visibility is spatially different in the image, resulting in a fog or "veil" effect in the captured image, and when the depth of the scene varies greatly in the rain image, the visibility of objects will change with the scene depth of the image, distant objects are more visually occluded by fog, and when the rain is heavy, the occlusion effect is more obvious.

This paper uses the depth information of the image to synthesize a rain data set that is more in line with the real rain scene based on the rain mark image designed in this paper and the construction site data set as the background [7]. The rain mark images provided in this paper take into account the information such as the density, direction, and scene depth of the rain marks [8]. A total of 400 images of construction sites are selected for rain image synthesis. Each image uses the depth information and rain marks in different directions to generate 12 buildings. A total of 4800 images are synthesized for the rainy day images of the construction site, of which there are 4500 images in the training set and 300 images in the test set. This method makes the rainmarks of the synthetic images more diverse and provides a more realistic rain image dataset for training the rainmark removal network.

D. Evaluation indicators

It is not objective and accurate to evaluate the quality of image restoration only through the visual perception of human eyes, because different human eyes have slightly different image resolution capabilities, which are greatly affected by subjective perception, and are also easily affected by external objective factors. Therefore, it is very important to use appropriate image evaluation indicators to accurately and objectively evaluate the effect of image rain streak removal. In

the field of image rain removal, researchers use data quantification of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) to measure the quality of image rain streak removal. PSNR and SSIM will be introduced separately below.

PSNR is mainly aimed at the absolute error between the corresponding pixels, and it does not fully consider the visual characteristics of the human eye. Therefore, when evaluating the quality of image reconstruction, the evaluation results are usually inconsistent with human subjective visual perception. Compared with PSNR, SSIM is more in line with the human eye's judgment of image quality in the measurement of image quality.

In the image rain removal task, SSIM is an evaluation index used to measure the similarity between the restored rain-free image and the real rain-free image. The value range of SSIM is [0, 1]. The more similar the two images are, the closer the value of SSIM is to 1.

E. Experiment and result analysis

Comparative Methods: The proposed rain removal method is compared with the traditional optimization method GMM, as well as the deep learning based methods JORDER and RESCAN. The GMM uses a pre-trained Gaussian mixture model as prior knowledge to decompose the image

into background scenes and rain marks. JORDER mainly proposes a multi-task joint detection rain removal network. RESCAN utilizes atrous convolution and residual learning for step-by-step rain removal.

In order to verify the effectiveness of the image rain removal method in this paper for image rain removal, this section verifies the effectiveness of the proposed method on a synthetic dataset through experiments, and compares it with GMM [9], JORDER, and RESCAN [10] for three representative images to remove rain. Algorithms were compared, and the experimental results were quantitatively analyzed using structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) evaluation indicators. It can be seen from the table 1 that the method in this paper has great advantages over other methods in both PSNR and SSIM indicators.

Figure 4 shows the different rain removal results respectively. Among them, sub-image (a) is the image with rain, sub-image (b) is the real image without rain, sub-image (c) is the GMM rain removal result, sub-image (d) is the JORDER rain-removing result, and sub-image (e) The result of draining for RESCAN, the sub-figure (f) is the draining result of this paper. They are described in detail below.

TABLE I. COMPARISON OF EXPERIMENTAL RESULTS ON SYNTHETIC DATASETS

Datasets	GMM	JORDER	RESCAN	Ours
PSNR	25.9725	26.4012	28.7863	29.1035
SSIM	0.9181	0.9246	0.9355	0.9661

In Figure 4, the amount of rain in the original rain image (a) is light rain, and the detail information covered by the rain marks is less, so more details can be added when removing the rain

mark information. Although GMM can remove most of the rain marks, the pixel details of the image after rain removal are blurred; JORDER leaves a small amount of rain marks and makes the

background image darker; RESCAN and the method in this paper are very thorough in removing the rain mark information and restoring

the background. The effect of removing rain marks is better.

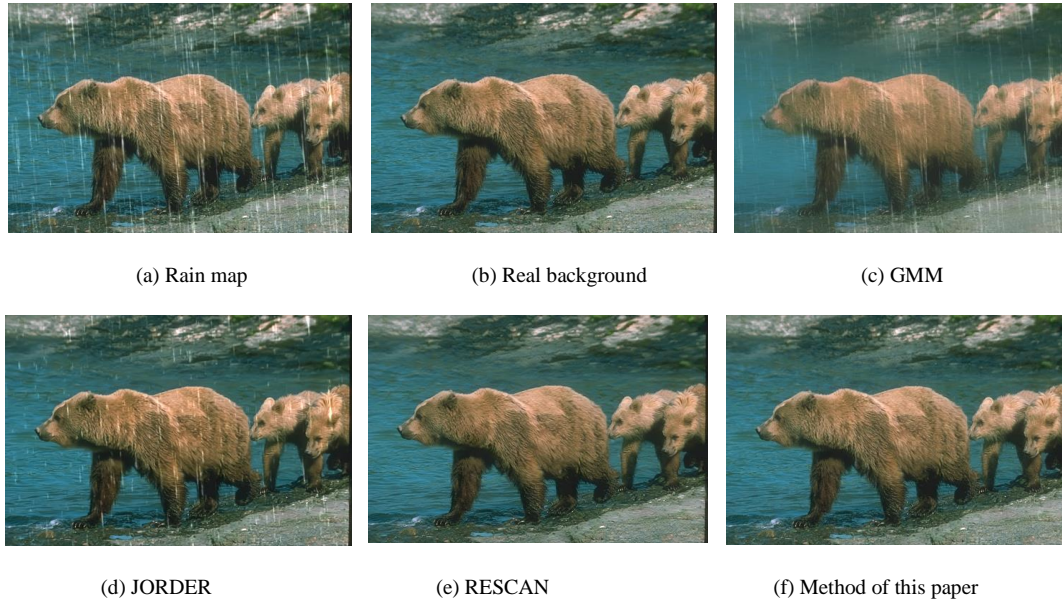
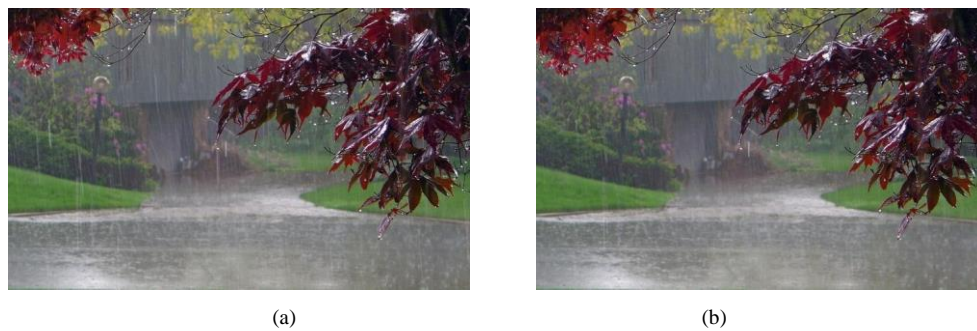


Figure 4. Visual comparison of the rain effect on the synthetic dataset Rain100H

F. Comparison on real rainy image datasets

The network model of the algorithm in this paper is trained on the synthetic rainy day data set[5]. Therefore, in the comparison of the real rainy day result images of the following methods, we will judge from various aspects according to human subjective vision, and continue to use two of the deep learning algorithms (JORDER and

RESCAN) used in the comparison of synthetic datasets. Compared with the synthetic data set, the real rainy day situation is more complex, which makes the image quality of the picture more diverse than the synthetic rainy day data set, and can better detect the generalization ability of the rain removal method.



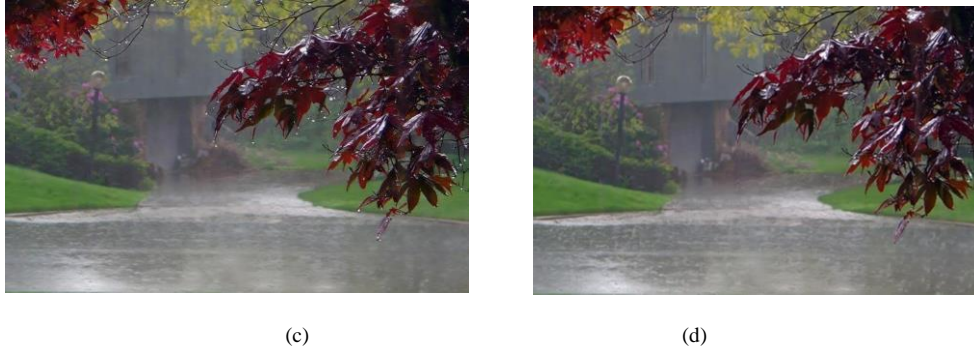


Figure 5. (a) Real rain map (b) JORDER (c) RESCAN (d) This paper method

The first image in Fig. 5 has less rainfall and residual rain marks in the results of JORDER and RESCAN. In the second image, there is a lot of rain and there is relatively large fog. It can be seen from the figure that a lot of rain marks remain in the results of JORDER, and many details in RESACAN are blurred. Relatively speaking, the method in this paper can still effectively remove the rain, and recover the rich texture background, which shows the effectiveness of the method in this paper in complex situations.

III. SUMMARY AND OUTLOOK

In this paper, some experimental attempts have been made to remove rain marks from a single image. Although some progress has been made, there are still some shortcomings, which need to be further optimized in future research work.

Since objects may move in real life scenes, and ambient lighting and camera exposure parameters may change, it is nearly impossible to take a large number of photo pairs with and without rain in the same environment. Therefore, the rain removal algorithms based on deep learning are all trained on synthetic images. Most of the existing synthetic rain image datasets add rain marks of various shapes on the background image, but as the depth of the scene increases, the intensity of the rain marks decreases, and the occlusion effect of fog increases. If a more realistic rainmark dataset can

be established, it will be of great help to improve the ability of the network to remove rainmarks.

The scale of existing deep neural network parameters is basically more than hundreds of thousands. The increase of a large number of network parameters significantly improves the network feature extraction ability and improves its task ability. However, the huge network structure limits its storage space and computing resources, Porting on the platform. In order to expand the application scenarios of deep learning network so that it can complete corresponding functions on a variety of small platforms, we should use model compression methods to optimize it. On the premise of ensuring the performance of rain removal, we should further reduce the network scale and improve its operating efficiency. It is the direction of our next research.

At present, the research ideas of model compression for deep learning networks are mainly divided into two categories: one is to compress the model through techniques such as quantization, pruning and knowledge distillation, and the other is to compact the original network by convolution operations. It can realize the functions of the original deep learning network with a small amount of calculation.

REFERENCES

- [1] Yang F, Ren J, Lu Z, et al. Rain-component-aware capsule-GAN for single image de-raining [J]. Pattern Recognition: The Journal of the Pattern Recognition Society, 2022(123-).
- [2] Wei Y, Zhang Z, Zhang H, et al. A Coarse-to-Fine Multi-stream Hybrid Deraining Network for Single Image Deraining [J]. IEEE, 2019.
- [3] Porres I, Ahmad T, Rexha H, et al. Automatic exploratory performance testing using a discriminator neural network[C]// 2020 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW). IEEE, 2020.
- [4] Yu C, Chang Y, Li Y, et al. Unsupervised Image Deraining: Optimization Model Driven Deep CNN[C]// MM '21: ACM Multimedia Conference. ACM, 2021.
- [5] Lin X, Huang Q, Huang W, et al. Single Image Deraining via detail-guided Efficient Channel Attention Network [J]. Computers & Graphics, 2021, 97(2).
- [6] Yang W, Tan R T, Feng J, et al. Deep Joint Rain Detection and Removal from a Single Image [J]. 2016.
- [7] Sakaridis, Christos, Dai, et al. Semantic Foggy Scene Understanding with Synthetic Data [J]. INTERNATIONAL JOURNAL OF COMPUTER VISION, 2018, 126(9):973-992.
- [8] Tang K, Yang J, Wang J. Investigating Haze-Relevant Features in a Learning Framework for Image Dehazing[C]// Computer Vision & Pattern Recognition. IEEE, 2014.
- [9] Mao X, Li Q, Xie H, et al. Least Squares Generative Adversarial Networks [J]. 2016.
- [10] Wang Z, Chen J, Hoi S. Deep Learning for Image Super-resolution: A Survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, PP (99):1-1.