Image Restoration Network Under Complex Meteorological Environment: GRASPP-GAN

MA JINGYI[1], ZHANG TIEJUN[2], JING GUODONG[3], YAN WENJUN[1], AND YANG BIN[1]

[1]Gansu Branch of China Meteorological Administration Training Centre, Lanzhou 730020, China
[2]China Meteorological Administration, Institute of Arid Meteorology, Beijing 10010, China
[3]China Meteorological Administration Training Center, Beijing 10010, China

Corresponding author: Ma Jingyi (mjycma@126.com)

ABSTRACT The color and contrast of objects in the image will be affected by meteorological factors, especially rain and snow will block part of the image, which will change the information contained in the image. Image restoration under bad weather conditions has practical application value. At present, most of the research focuses on the removal of fog, and the research on complex rain and snow is relatively less. Rain has more prominent features in gradient domain, and it is more distinct from non-rain image texture. In this paper, Generative Adversarial Networks is used to combine the information of image in gradient domain and spatial domain to get better performance of rain removal. Gradient aided coding is used in the generator to generate depth features that are more conducive to rain removal. In the discriminator, the gradient is used as an additional input to provide more recognizable rain and non-rain information, which enhances the discriminator's ability to distinguish the image generated by the generator and the ground truth. By modifying the network structure of the expanded spatial pyramid pooling (ASPP), the abnormal rain removal results produced by the generator are reduced. Experimental results show that the proposed method improves the performance of rain removal and the visual quality of the generated image.

INDEX TERMS Image restoration, GAN, meteorology, rain.

I. INTRODUCTION

Target detection [1], tracking [2] and identification [3] systems are widely used in modern city construction, crisis prevention and treatment, security and other fields, including unmanned driving technology which is developing continuously with the rise of artificial intelligence. In these applications, it is often hoped that the clear image or image sequence can be obtained to better realize the detection, tracking and recognition of the target. However, images or image sequences obtained by imaging equipment in outdoor environment are easily affected by various meteorological environmental factors, resulting in the decrease of visual effects, data quality and application value of images or image sequences. For example, snow will block local targets in the image, and sometimes even block the content of the image containing important information: although raindrops are translucent, their transmission to light is limited, so they will also partially block useful targets in the image, making them

The associate editor coordinating the review of this manuscript and approving it for publication was Tao Zhou.

blurred [4], [5].In the case of heavy rain, not only obvious rain lines appear in the image, but also accumulated rain forms Haze-like Effect, which reduces image contrast and leads to image color distortion [4], [6], [7].This situation has a greater impact on the information contained in the image, and even overwhelms the useful tiny features in the image, which makes the performance of some computer vision algorithms based on tiny features of the image lower. The southern region is mainly affected by rain and fog, so to make the target monitoring, tracking and identification system can work stably, eliminate the impact of these weather is crucial.

The existing rain-removal methods based on deep learning are usually designed and trained for a specific training set and test set containing similar rain-lines to remove a certain type of rain [6]–[11]. One of the biggest limitations of these methods lies in its extension (Generalization Performance) performance is inferior, that is to say, when enadversarial other types of rain grain, the performance of these methods will be markedly reduced. One challenge of these approaches is to deal with rain streaks with wide and fuzzy edges. In view of the above problems, this paper trains a Generative

Adversarial Networks that is sensitive to the shape and size of rain stripes to deal with more challenging rain stripes. Therefore, the network proposed in this paper has better expansibility and processing ability for different types of rain stripes. Considering that the complex rain pattern can damage the image background, the rain removal network is a very important aspect for the restoration ability of background texture. Gradient is a direct and effective tool to reflect image texture. It has been widely used in other computer vision tasks, such as super-resolution of images [12]. In order to improve the texture of degraded images, gradients are widely used in this paper to assist the training of Generative Adversarial Networks. In addition, extensive experiments have found that Global Features can impair performance for low-level computer vision tasks.

The main contributions of this paper are as follows.

1) A novel RASPP network structure was established to extract multi-scale features of rain images. RASPP brings the benefits of ASPP to lower-level computer vision tasks by overcoming the image black block effect introduced by global features.

2) In this paper, the residual network structure is modified to generate depth features that keep the scale constant, which makes it possible to use the superior residual structure in low-level computer vision tasks.

3) A gradient-guided GAN rain removal network is established in this paper. Gradient-assisted coding, optimization, and discrimination are introduced to help identify rain-removal results and ground truth.

II. RELATED WORK

As a dynamic weather factor, rain has complex shapes and scales. It is difficult to fully extract the characteristics of rain by a single scale convolutional layer. References [8] and [6] use dense connection blocks with multi-scale and expansive convolution (Atreus Convolution) to extra multi-scale features of rain images. Atreus spatial pyramid pooling [13] (ASPP), which is widely used in mid/high-level computer vision tasks, is used to extract the multi-scale features of images in this paper. Compared with the expansive convolution used in [6], ASPP contains a point-wise convolution layer. This convolutional layer is similar to the cut-path in the residual network [14], which is used to maintain the depth characteristics of the original scale. In addition, ASPP includes an extraction layer of global features to obtain more comprehensive depth features. Pooling and usually space pyramid [15] (Spatial Pyramid Pooling, SPP) than ASPP can get by adjusting the convolution parameters and the original images with the same depth of multi-scale feature of size.

For low-level rain removal tasks, using ASPP directly does not yield very good results. A large number of experiments have found that the depth characteristics obtained in the large receptive field have impaired the performance of the low-level computer vision algorithms sensitive to local image details. Some bad image effects will inevitably appear in the

final result of the rain. Therefore, the algorithm in this paper removed the layer of extracting global features from ASPP and the adjacent expansive convolutional layer with large Atrous rate, and searched for Atrous rate suitable for rain removal task to form a new ASPP structure, which was named RASPP for convenience. In experiments, RASPP solves the problem of singularity in the image. In addition, symmetric filling is used instead of 0 filling in the convolutional layer to obtain more accurate image edges. For the network structure, the algorithm in this paper modifies the residual network [14] (ResNet) as the encoder. In order to adapt to the image recognition task, the original residual structure ADAPTS to the low-level computer vision task by modifying the residual network to obtain the depth feature with the same size. In addition, all the convolution filling modes in the residual network are changed to symmetric filling.

Rain pattern will distort the texture and detail of the background image, so texture recovery is essential in lossless background restoration. In the spatial domain, background content will affect the image's rain-removal performance, so some work to restore image details at high frequencies of the image [16], [17]. However, extensive experiments show that the rain information is more obvious in the gradient domain, and the background texture features are also more prominent. Based on this, the algorithm in this paper uses the generator in gradient-assisted GAN network to extract depth features more conducive to rain removal. At the same time, the MSE cost of establishing a gradient domain is optimized in the process of backward propagation to optimize the generator's network parameters. In addition, gradient information promotes the convergence of network training. In the absence of gradient, the proposed network converges after 45 epochs, and the convergence of the network with gradient information requires only 13 epochs. In this paper, the gradient-assisted feature extraction structure is named GRASPP.

With GRASPP, good rain-removal performance has been obtained. However, for some rain images, a little texture may still not be recovered and some rain pattern traces still exist in the final result of rain removal. In order to further improve the performance, Adversarial Training is introduced into the network [18]. As mentioned before, the texture and rain pattern of the image show better Discrimination in the Discriminator. Therefore, the Discriminator is also trained with the assistance of the gradient, so that the network can recognize the result produced by the generator and the corresponding ground truth from both the spatial domain and the gradient domain. Higher recognition accuracy leads to better generator performance. Therefore, in this paper, a gradient guided Conditional Generative Adversarial Network (CGAN) is established to realize the rain removal task in a single image.

The content of this paper is arranged as follows: Section 3 systematically introduces the proposed algorithm and network structure as well as the corresponding training details; In section 4, the proposed algorithm is experimentally verified and compared with existing methods in
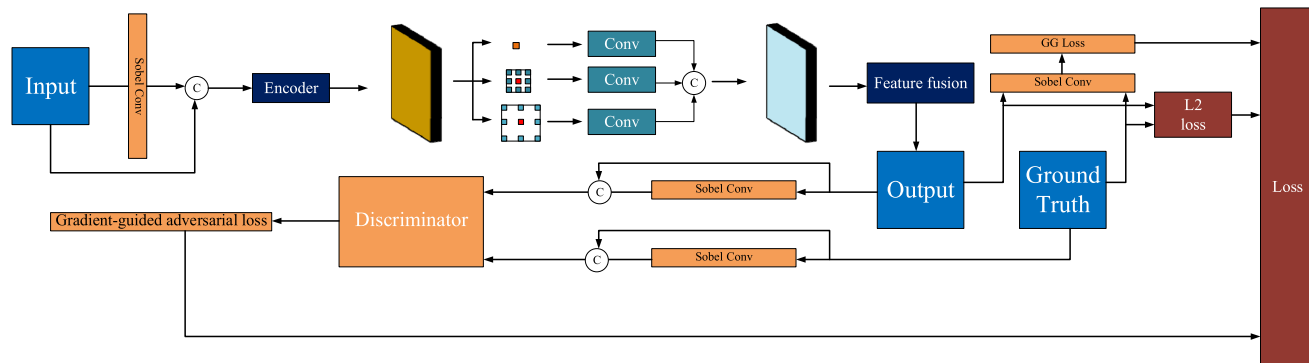
FIGURE 1. Network Architecture, the generator consists of three main parts: the encoder, RASPP, and the feature fusion block. The encoder is a modified RESnet-18 to extract scale invariant features; RASPP contains three parallel paths, each of which contains an expansive convolutional layer and a point-by-point convolutional layer; Feature fusion module contains three common convolutional layers; The loss function of the training network is composed of three parts: 2-norm MSE loss (L2loss) in the spatial domain to restore the overall image content, 2-norm MSE loss (GG loss) in the gradient domain to repair the damaged image texture, and the gradient-guided adversarial loss to identify the result of rain elimination and the corresponding ground truth.

subjective and objective aspects, and ablation experiments are conducted. Section 5 summarizes the whole paper.

III. GENERATIVE ADVERSARIAL NETWORKS BASED ON GRADIENT ASSISTANT

In order to remove rain streaking from the single image, resnet-18 [63] was used as the backbone of the rain removal network to extract depth features with the assistance of gradient. In order to keep the scale of depth feature unchanged to adapt to the task of rain removal, the stride of all down sampling in the residual network was reset to 1. In addition, the pre-trained model in the residual network is used to initialize the modified residual network to accelerate the convergence of the network and obtain more robust features. A modified ASPP network structure was introduced to diversify the extracted depth features to adapt to the changeable rain pattern. In addition, two MSE cost functions from the spatial domain and the gradient domain are combined to optimize the network parameters to restore the image background. Rain pattern is more obvious in gradient region, which is more conducive to the identification of rain and non-rain background. Therefore, a gradient-assisted discriminator is introduced to form the performance of the adversary training promotion generator. Because of the extensive auxiliary role of gradient in the whole network training, the rain elimination network in this paper is named gradient-assisted generation adversarial network based on modified ASPP.

A. GRASPP-GAN

The rain-removing frame in this paper consists of four main parts: feature encoder, modified ASPP block, multi-scale feature fusion module, and discriminator. The concrete structure is shown in Figure 1. In addition, three cost functions are used to optimize the rain-removing network in the adversarial training: 2-norm MSE loss (L2 loss) in the spatial domain is used to recover the approximate image content, such as the

intensity information of color and pixel; MSE loss (GG loss) of a gradient domain is used to repair the background detail texture. Gradient-guided adversarial loss further promotes network performance. In the training, these three loss are used to optimize the network parameters with different weight correlation.

The encoder takes a single RGB rain image I as input and outputs depth feature spectrum f_1 through training. In this process, a convolution layer of Sobel gradient is built to obtain the gradient of rain image I in both horizontal and vertical directions. This gradient is used to guide the encoder encoding depth features to help eliminate rain. Then, RASPP module applies the expansion convolution with different expansion rates to extract the multi-scale feature f_2 from the depth feature f_1 to adapt to the scale variability of rain pattern. Finally, the feature fusion module acts like a decoder to fuse multi-scale feature f_2 and generate the result d of rain removal.

The following loss function is adopted in training:

$$Loss = L_2 + \alpha L_g + \beta L_{gan} \tag{1}$$

where, α, β are the weights, $L_2 = \|d - g\|_2$ is used to measure the difference between the precipitation result d in the spatial domain and the corresponding ground truth g. L_g and L_{gan} are gradient domain losses and gradient guidance adversarial losses.

B. GRASPP
1) ENCODER
The problem of eliminating rain is different from other computer vision tasks that are processed pixel by pixel and do not require a deep network to achieve good results. It can be verified by using ResNet-18, ResNet-34 and ResNet-50 [14] as the main encoder of the algorithm in this paper. The experimental results showed that they had similar rain-removal performance, and the maximum difference between PSNR and SSIM was 0.03dB and 0.004, respectively. Therefore,

Resnet-18, which has fewer layers, is used to reduce the run time and the number of network parameters. However, ResNet is mainly used in medium/advanced computer vision tasks and requires down-sampling to learn depth features. In order to introduce ResNet's superior performance into low-level computer vision tasks that require constant feature scale, the following modifications were made to ResNet in this paper:

(1) Set all the steps of the convolutional layer of the lower sample to 1 to keep the resolution of the depth feature unchanged.

(2) In general, the zero-padding method used in convolution is replaced by symmetric padding to obtain better image boundary estimation.

(3) The modified linear units (ReLUs) are all replaced by Leaky Corrected Linear Units (Leaky relu) to relieve Dead Neuron [19].

(4) The pre-training model of the residual network is used to initialize the modified encoder to accelerate the convergence.

(5) Before the residual network is a Sobel gradient convolution layer to calculate the vertical and horizontal gradients of rain image I. This gradient is also input into the residual network together with the rain image f so that the network can also extract depth features from the gradient domain of the rain image

2) ASPP MODULE

Inspired by the excellent performance of Deeplab V3+ [20] and DORN [99], this paper also applies the module of atrous spatial pyramid pooling (ASPP) to extract the multi-scale features of rain images. ASPP extracts large-scale depth features in a large acceptance domain and also contains a global feature, but experiments have shown that large-scale features can damage low-level rain elimination tasks (detailed experiments will be shown in the following sections). Therefore, the last two layers of the original ASPP structure were removed from the algorithm in this paper, and the expansion rates of the remaining three layers were set to 1, 2, and 4. Experiments have verified that this is the most conducive to the performance of rain removal and the avoidance of image anomalous effects of the expansion rate.

3) FEATURE DECODING MODULE

As a depth feature decoder, the feature decoder module contains three convolutional layers, all of which have a convolution kernel size of 3×3 and a step size of 1. These convolutions are always filled with symmetry, and the first two convolution layers are followed by a Batch Normalization and a ReLU activation layer.

C. SIMILARITY MEASURE IN GRADIENT DOMAIN

At present, most rain elimination networks based on deep learning are trained by minimizing the difference between rain removal results and the corresponding ground truth in the space city. This restores most of the image content, such

as color and pixel intensity, but some tiny image textures are often over looked. It is found that the texture details and rain pattern are usually more distinct in the gradient domain. This can be regarded as a prior information of the rain image in the gradient domain. An example is shown in Figure 3. Therefore, depth feature extraction and network optimization in gradient domain can make up for the shortage of spatial domain operation.

Because of the prior knowledge of rain pattern and background texture in gradient domain, the encoder can extract the depth feature of rain image from gradient domain. Similarly, the parameters of the generator network are optimized by complementing the spatial domain and the gradient domain. The network is optimized by minimizing the difference between rain removal results and ground truth in the spatial domain. By minimizing the difference between rain removal results in gradient domain and the corresponding ground truth, the network was further optimized to restore the fine image texture.

Before introducing the specific cost function, the difference of image information carried by gradient domain and spatial domain is firstly analyzed. In fact, the gradient domain and the spatial domain are to some extent complementary, as shown in Figure 2. (a) and (b) are rain images and corresponding ground truth, respectively. (c) and (d) are gradient spectra of rain images and corresponding ground truth in horizontal direction. (e) is the rain layer needed to synthesize the rain image. By comparing (a)(b) with their corresponding gradient (c)(d), it is found that the spatial domain and the gradient domain respectively emphasize different information content. The spatial domain mainly reflects the overall content, color and pixel intensity of the image, while the gradient domain focuses on displaying some texture information of the image that is not obvious in the spatial domain.
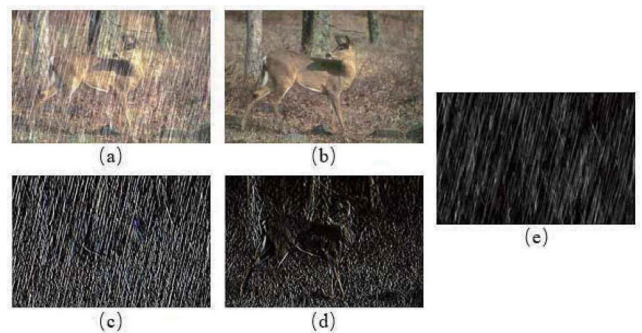

FIGURE 2. Comparison of image features between spatial domain and gradient domain;(a) and (b) are RGB images with and without rain, respectively: (c) and (d) are gradient spectra of the horizontal direction (x axis) corresponding to (a) and (b) respectively; (e) is the rain layer.

Therefore, during the training process, an MSE loss function Lg is added in the gradient domain to assist in optimizing the proposed rain removal network and recover the tiny image

details and textures lost in the spatial domain:

$$L_g = \frac{1}{n} \sum_{i=1}^{n} \left[\|\nabla_x(d_i) - \nabla_x(g_i)\|_2 + \|\nabla_y(d_i) - \nabla_y(g_i)\|_2 \right] \tag{2}$$

where, ∇x and ∇y represent the gradient operators in the horizontal and vertical directions respectively, and n is the number of training samples. The Sobel convolutional layer is designed to extract the gradient in the horizontal and vertical direction of the image. Sobel convolution layer takes Sobel operator [21] (Operator) as the convolution kernel, and the parameters are set to remain unchanged in the training process.

D. GRADIENT-ASSISTED DISCRIMINATOR

In order to further improve the network performance, an effective gradient-assisted adversarial training is introduced. The gradient provides more rain and non-rain information to help the discriminator identify rain image and ground truth. The reason for using gradients is also that rain bands are more pronounced in gradient regions as mentioned above. Specifically, after the generator GRASPP produces rain removal results, the rain removal results /ground truth and their corresponding gradient spectrum cascades are combined as the input of discriminator, and finally a score is obtained. This rating table indicates the probability that the input is the rain result or the corresponding ground truth. The discriminator is optimized by maximizing the probability of correctly assigning tags. In the generation adversarial training, the discriminator and generator are trained synchronously according to the following formula:

$$\min_G \max_D V(G, D)$$
$$= \mathbb{E}_{z \in \zeta} [\log(D(z, S(z)))] + \mathbb{E}_{x \in \chi} [\log(1 - D(G(x), S(G(x))))] \tag{3}$$

where, $x = [x1, \ldots, xn]$ is the input RGB rain image, $z = [g1 \ldots gn]$ is the corresponding ground-truth. S (•) is Sobel convolution layer. χ and ζ are the distribution of input rain image and corresponding ground truth. D and G are discriminators and generators. The discriminator is a binary classifier, so Cross Entropy Loss is used as the cost function of discriminator training.

The detailed network structure of the discriminator is shown in Figure 3. It consists of four convolution blocks, a global average pooling layer, and a full connection layer. The first convolution block contains a convolutional layer with a kernel size of 4 × 4 and a layer of Leaky ReLU activation functions. The next three convolution blocks all contain a convolution layer, a batch standardization layer, and a Leaky ReLU activation function layer. All convolutional layer steps are set to 2. The mathematical expression of Leaky ReLU is as follows:

$$\text{Leaky ReLU}(x) = \max(0, x) + \text{negativeslope} * \min(0, x) \tag{4}$$
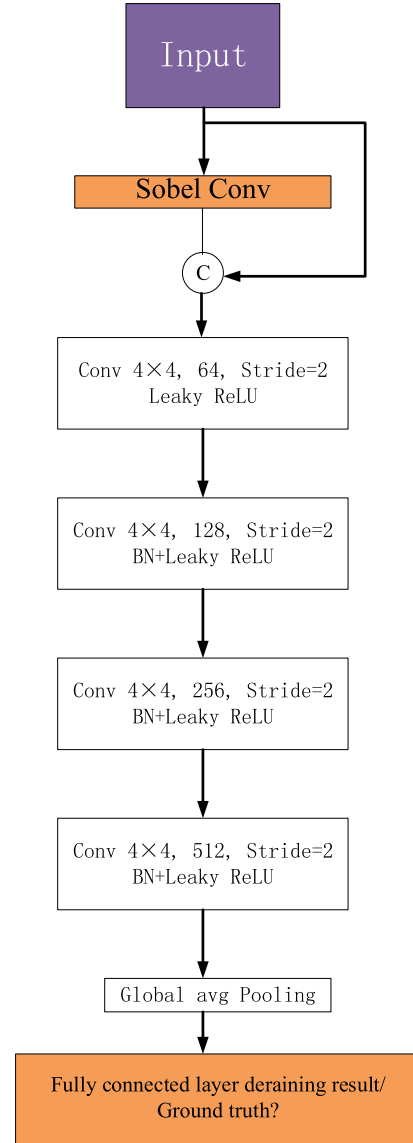

FIGURE 3. The network structure of discriminator.

where, negativeslope is set to 0.2 in this paper. The gradient guidance adversarial loss:

$$L_{gan} = \mathbb{E}_{x \in \chi} [\log(D(G(x), S(G(x))))] \tag{5}$$

IV. EXPERIMENTS AND ANALYSIS

In order to evaluate the performance of different methods, PSNR and SSIM [22] are used as objective evaluation indicators. For qualitative evaluation, visual results on synthetic rain images and real rain images are presented respectively. As there are many literatures on rain removal, this algorithm adopts a reasonable way to compare with the existing literatures on rain removal: three robust and widely used literatures [8,16,110] and three recent literatures [7], [9], [11] are used to compare with the proposed method.

In the process of training, 128 × 128 image blocks are intercepted from each training sample to form the training

sample of the input network. All image blocks are randomly intercepted from the original training pair. Adam optimizer is adopted to train the network, and the initial learning rate of generator and discriminator network is set to 0.001 and 0.1 respectively. When the network training stops convergence,

0.1 times the previous learning rate is used to adjust the learning rate until the network convergence. During the training process, the generator first conducts two separate training sessions for epochs and then adds the discriminator to the training process. The rain removal network in this paper is trained on a computer equipped with NVIDIA1080TiGPU, and the specific network framework is built on PyTorch. Each batch contains 4 sets of images during training. α and β in the loss function (Eq. (1)) are set to 1 and 0.001 respectively.

A. DATASET

In the experiment, a training set was prepared according to the method in reference [23]. This training set consisted of 20800 groups of training samples, and the rain images in each training sample were synthesized by Screen Blend Mode.

In order to more fairly test the performance and extensibility of the proposed method, the benchmark set in literature [8], [16], [10] constitutes the first test set of the algorithm in this paper, Rain-I. Literature [6] syntheses two test sets Rain100L and Rain100H, which respectively contain 100 sample pairs. In this paper, Rain100H, which is more challenging, is selected as the second test set Rain-II. In addition, it is difficult to remove wider rain lines and blurry edges. Therefore, the third test data set, Rain-III, is composed of another 400 samples containing wider rain-lines and blurry rain-line edges. The actual rain images in the experiment were taken from previous rain removal literature or downloaded from Google. They contain rich image content, including natural scenery, urban buildings, human faces and so on.

B. COMPARATIVE EXPERIMENT

Table 1 shows the comparison test results between the selected method and the method proposed in this paper. In Rain-I and Rain-III, the method in this paper is superior to other methods in terms of two objective indicators. On Rain-II, the objective index of this algorithm is the second best. The best performance comes from the literature [10], because this literature is trained on the training set corresponding to this test set. In addition, the literature [7] is also trained on the training set corresponding to Rain-II, but the performance of the method in this paper exceeds that of this algorithm on this test set.

In order to compare the performance of different methods more fairly, the selected literature is retrained on the training set used by the algorithm in this paper, and the test results are shown in Table 2. Note that reference [7], [9] requires additional ground truth configuration during training, so it is impossible to train on the training set in this paper. Although literature [8] also needs ground truth of rain density in training, in the process of retraining, the parameters of rain density

TABLE 1. A comparative experiment on three test sets.

Dataset	Rain-I		Rain-II		Rain-III	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Deep Detail [16]	**29.27**	0.889	21.94	0.715	30.03	0.874
Multi-stream Dense [8]	26.98	**0.901**	19.88	0.702	24.99	0.868
Context Aggregation Net [10]	27.55	0.880	**25.96**	**0.851**	27.28	0.879
Deep Joint [7]	28.62	0.849	23.04	0.450	28.87	0.880
depth-guided GAN [9]	18.03	0.083	18.11	0.552	19.01	0.614
Spatial Attentive [11]	29.11	0.872	22.73	0.739	**30.32**	**0.921**
DRPRN [24]	26.34	0.823	22.58	0.754	26.35	0.854
Feature-Supervised GAN[25]	25.89	0.817	21.36	0.722	24.91	0.833
Ours	**30.31**	**0.913**	24.29	0.836	**31.84**	**0.946**

TABLE 2. A comparative experiment on three test sets after re-trained.

Dataset	Rain-I		Rain-II		Rain-III	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Deep Detail [16]	27.59	0.838	20.68	0.689	**30.52**	**0.909**
Multi-stream Dense [8]	22.98	**0.901**	18.34	0.671	26.98	0.868
depth-guided GAN [9]	27.03	0.882	22.76	0.754	28.77	0.899
Spatial Attentive [11]	**28.56**	0.850	**23.55**	**0.819**	30.49	0.904
DRPRN [24]	27.32	0.879	22.35	0.762	29.32	0.863
Feature-Supervised GAN[25]	27.10	0.864	21.66	0.713	28.55	0.814
Ours	**30.19**	**0.909**	24.78	0.828	**31.86**	**0.938**

estimation module are kept unchanged, and only the network parameters of rain removal module are updated through training. Therefore, table 2 contains the comparison results of the four literatures. It can be seen that when training on the training set in this paper, the proposed method produces better objective performance on all three data sets.

C. ASSESSMENT OF THE VISUAL QUALITY OF RAIN ELIMINATION

Figure 4 shows some of the results on the composite rain image and their corresponding horizontal gradients. These three images are from the three test sets in this chapter, and it can be found that the method presented in this chapter produces the best visual effects. References [16] and [8] are good at processing the synthetic Rain image in the test set
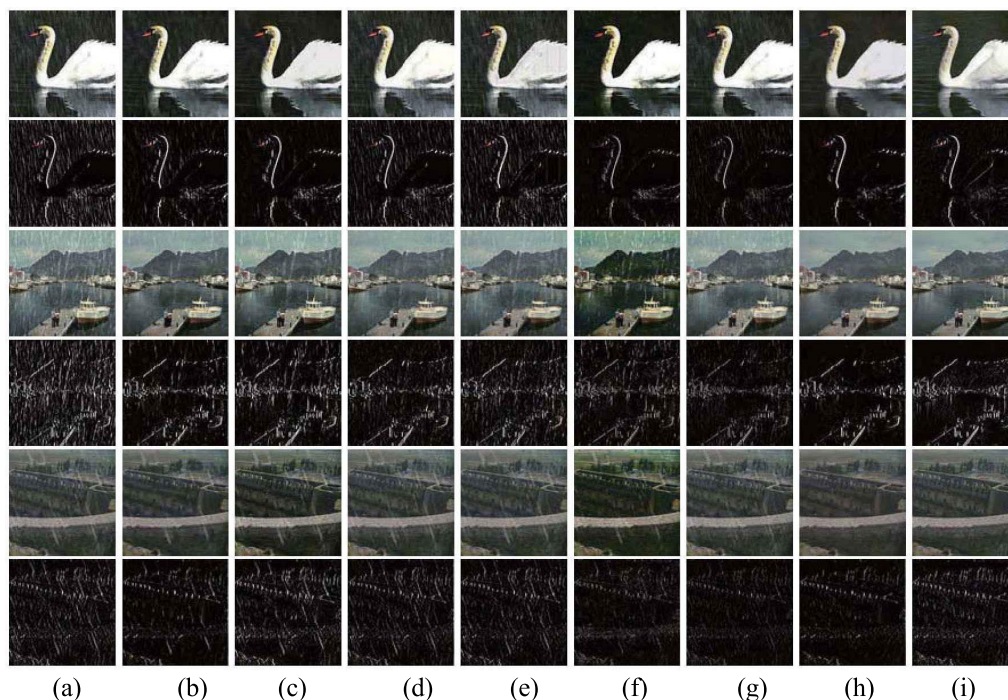
FIGURE 4. The results of different methods on the synthetic rain image, (a) composite rain image; (b)-(h) Results of rain elimination using [16], [7]–[11] and method proposed; (i) ground truth. In order to see the texture of the image more clearly, the gradient of the corresponding image in the horizontal direction is shown here.

Rain-I, because the Rain pattern in this data set is relatively fine. In the case of rain-II data set with high pixel intensity or wide edges with fuzzy rain-lines, these two methods cannot produce good rain-removing effect. The literature [7], [10] is specially trained on the training set corresponding to Rain-II

of the test set, so the literature [80] can remove most Rain patterns from the test set Rain-II, but the literature [7] still leaves out some obvious Rain patterns. These two literatures do not produce very good results when encountering dense drizzle, such as the first rain image. In addition, all the selected literatures failed to remove relatively wide rain stripes. As can be seen from the third figure, fine rain stripes were removed, but wide rain stripes were still left in the final rain removal result. References [9] will obviously change the color tone of some of the dewatering images, and some distinct rain patterns will remain in the final dewatering image. References [11] is not very good for some challenging rain streaks. In contrast, the approach presented in this chapter produces better rain pattern processing, with most of the rain pattern removed, and the image details restored with significantly more fidelity. These visual results demonstrate the good scalability of the proposed method.

The results of rain removal and the corresponding gradient on the three challenging real rain images are shown in Figure 5. The first rain image contains a face covered by the rain. Although the rain lines in this image are thinner, the edges of the rain lines are blurry. The second rain image has wider rain streaks with the same blurry edges. The rain lines in the third image are relatively thin, but they all overlap. With the exception of literature [16], [10], [11] which

produced good results for the third image, the other methods selected did not produce good results for the three images. In addition, the results of [9] are affected by the block effect, such as the second figure. By comparison, the methods in this chapter produce better rain removal effects and demonstrate better scalability.

D. ABLATION EXPERIMENT
The network in this paper adopts the combination of residual network Resnet-18 and RASPP as the backbone network of the generator network. Gradient is used to assist depth feature extraction, network structure optimization, and discriminator discrimination. Finally, adversarial training is introduced to further improve the performance of the network. Therefore, two additional ablation experiments can be performed to prove the effectiveness of the trunk generator network, gradient, and discriminator. The first ablation experiment is gradient removal auxiliary and discriminator adversarial training, retaining only the role of generator, which is named RASPP. The second is a guide that adds a gradient during training, which is called GRASPP. The third was added to the adversarial training, and was named GRASPP-GAN.

Table 3 shows the PSNR and SSIM of RASPP, GRASPP and GRASPP -GAN on the three test sets respectively. It can be seen that RASPP on the main network produces satisfactory rain-removing results, and its performance even exceeds that of some selected literatures, such as literature [8]. This is because the residual network effectively restricts the gradient disappearance in the network training process, so as to obtain better rain-removal features. In addition, the
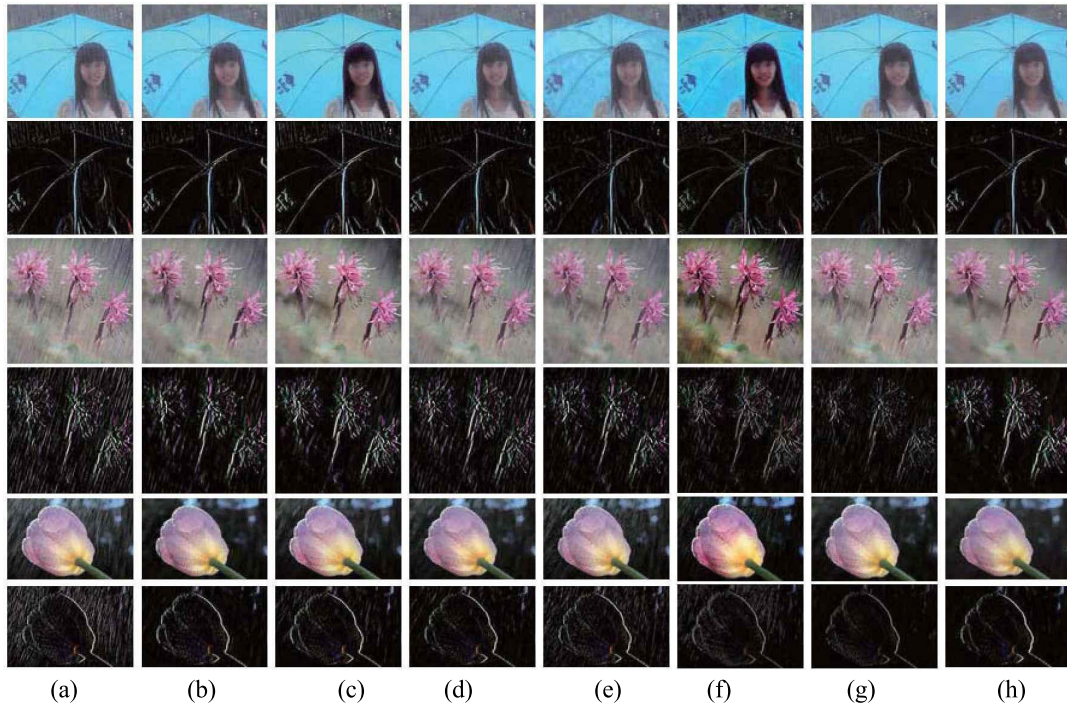
| (a) | (b) | (c) | (d) | (e) | (f) | (g) | (h) |

FIGURE 5. The results of different methods on real rain image, (a) real rain image; (b)-(h) Results of rain elimination using [16], [7]–[11] and method proposed. In order to see the texture of the image more clearly, the gradient of the corresponding.

TABLE 3. Results of ablation experiment.

Dataset	Rain-I		Rain-II		Rain-III	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
RASPP	28.31	0.891	22.49	0.779	**28.99**	**0.918**
GRASPP	28.89	0.897	24.01	0.831	**30.86**	**0.925**
GRASPP-GAN	30.13	0.908	24.96	0.842	**31.77**	**0.951**

modified RASPP network structure can obtain multi-scale depth characteristics reflecting the shape and size of rain. After the introduction of gradient information (GRASPP), PSNR and SSIM achieved consistent improvement in the three data sets.

The gradient accelerates the training process of the network. In the absence of gradient, the network converges after 45 epochs, and the addition of gradient information only requires 13 epochs, which means that the gradient helps the encoder to form the characteristics of rain removal faster. Adversarial training often has a positive effect on performance improvement by progressively improving the quality of the rain-removal results generated by the generator by continuously determining whether the rain-removal results contain residual rain-lines. Therefore, the performance of GRASPP-GAN has been further improved.

Figure 6 and Figure 7 show the rain eliminating results of the ablation subjects on the synthesized and actual rain images, respectively. For most rain images, GRASPP-GAN and its two variants can obtain good rain removal results. Consider the second image in Figure 6 and the two actual rain images in Figure 7. However, there are still significant
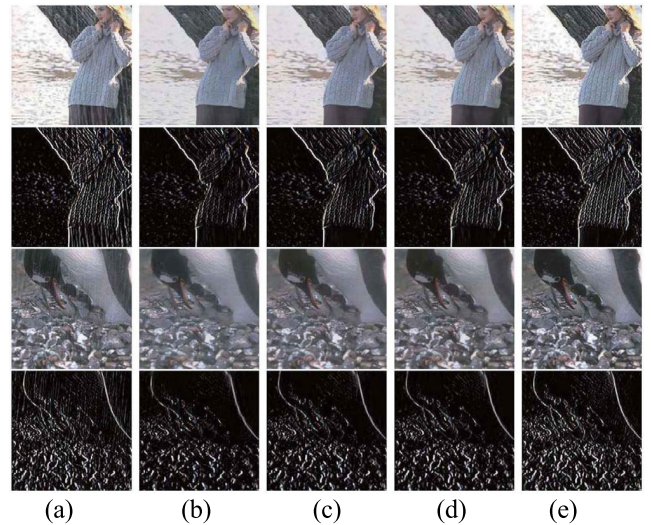

| (a) | (b) | (c) | (d) | (e) |

FIGURE 6. The visual results of the different subjects in the ablation experiment on the composite rain image, (a) input image; (b) The result of RASPP; (c) The result of GRASPP; (d) The result of GRASPP-GAN; (e) Ground truth. The gradient is also shown here to better illustrate the restoration of the image texture.

differences among the three for some images, such as the first image in Figure 6. It was found that RASPP lost some of the image details and the lost details were gradually recovered with the addition of gradient and adversarial training.

E. THE INFLUENCE OF LARGE-SCALE FEATURES ON THE TASK OF RAIN REMOVAL

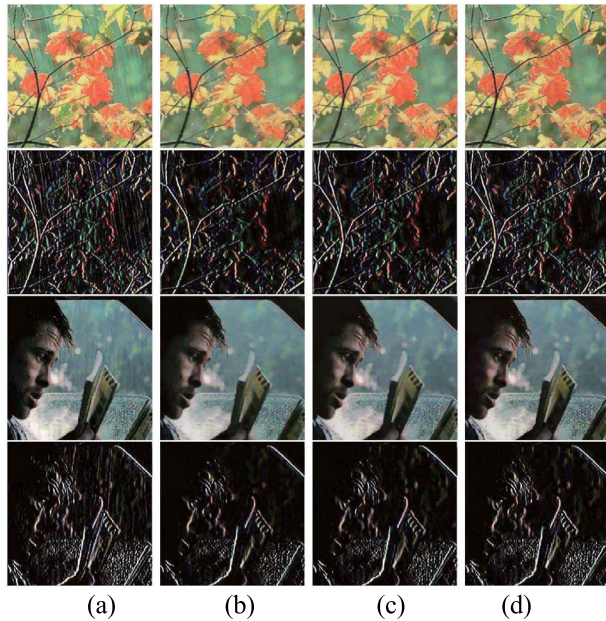Through extensive experiments, it is found that global features and large-scale features will cause bad effects in

FIGURE 7. The visual results of the different subjects in the ablation experiment on real rain image, (a) Input image; (b) The result of RASPP; (c) The result of GRASPP; (d) The result of GRASPP-GAN;. The gradient is also shown here to better illustrate the restoration of the image texture.

low-level computer vision tasks. This effect will seriously affect the PSNR and SSIM indexes of the algorithm as well as the visual quality. The proposed algorithm modifies the original ASPP structure by removing the global scale and large scale feature extraction layer. By comparing the performance of the original ASPP with that of the modified RASPP, the effect of large-scale characteristics on the rain removal task of low-level computers was studied. The experimental results are shown in Table 4, and the visual results are shown in Figure 8. It can be seen from Table 4 that the objective index of ASPP is significantly less than RASPP. The reason is that large-scale features will produce abnormal black block effect in the final result of rain removal, as shown in the last three columns in Figure 8. Therefore, large scale features are not conducive to the task of rain removal sensitive to local image details.
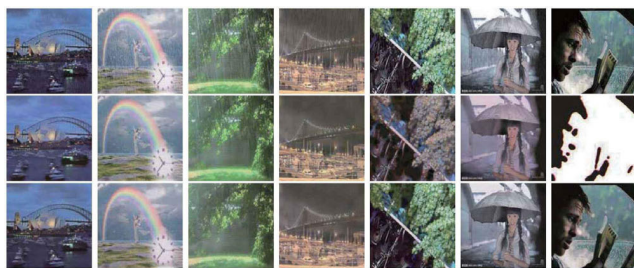

FIGURE 8. Effects of large-scale depth features on low-level computer rain removal tasks, the first line is an actual image of rain; the second row is the result of using the original ASPP; the third row is the result given by RASPP.

In this algorithm, the last two layers of the original ASPP structure are deleted, and the expansion rates of the remaining

TABLE 4. Comparison of PSNR and SSIM between ASPP structure and modified RASPP structure.

Dataset	Rain-I		Rain-II		Rain-III	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
ASPP	23.98	0.846	19.19	0.703	24.31	0.902
RASPP	30.17	0.908	23.99	0.832	31.66	0.938

three layers are set to 1, 2 and 4. These expansions are not randomly selected, but the expansions that are most beneficial to rain removal performance and avoid abnormal image effects are verified by extensive experiments. In order to verify the impact of expansion ratio combination on the performance of ASPP module, we designed a group of comparative experiments to compare the performance of network models with different expansion ratio combinations. The experimental results are shown in Table 5. It can be seen from table 5 that the network performs best when the expansion rate is 1, 2 and 4. Large scale features are not conducive to the task of rain removal which is sensitive to local image details.

TABLE 5. Comparison of PSNR and SSIM between different expansion rates of ASPP.

Rate	Rain-I		Rain-II		Rain-III	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1,2,4	**28.43**	**0.932**	**24.49**	**0.909**	**28.86**	**0.923**
2,4,8	27.31	0.910	23.14	0.885	27.18	0.902
3,6,9	24.73	0.864	20.24	0.810	22.41	0.863
4,8,12	19.62	0.821	18.32	0.768	20.37	0.821
5,10,15	15.73	0.708	14.24	0.632	15.29	0.699

V. CONCLUSION

In this paper, a gradient-assisted generative adversarial network (GRASPP-GAN) is proposed to remove the influence of rain from a single rain image. The modified ResNet network structure can effectively constrain the gradient disappearance in the backward propagation. In order to adapt to the varied shapes and sizes of rain patterns, a modified RASPP structure was used to extract multi-scale depth features of rain images. In addition, the rain pattern has more obvious features in the gradient domain, so the extraction and optimization of the depth features of the generator are completed under the assistance of the gradient, so as to generate better rain-removal features. For image gradient extraction, the gradient convolution layer of Sobel is specially designed to facilitate the establishment of the network. Finally, the objective performance and visual quality of rain removal are further improved by gradient-guided discriminator. Extensive experiments show that the method proposed in this paper produces better rain removal performance than existing methods.

REFERENCES

[1] M. Li, Y. Du, Z. Gao, Y. Zhang, and Z. Qin, "Research and application of object recognition method of visual grasping robot based on deep learning," in *Proc. 3rd Int. Conf. Electron Device Mech. Eng. (ICEDME)*, May 2020, pp. 652–656, doi: 10.1109/ICEDME50972.2020.00154.

[2] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Robust visual tracking via hierarchical convolutional features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2709–2723, Nov. 2019, doi: 10.1109/TPAMI.2018.2865311.

[3] Guo, Guodong, and N. Zhang, "A survey on deep learning based face recognition," *Comput. Vis. Image Understand.*, vol. 189, Dec. 2019, Art. no. 102805, doi: 102805.1-102805.37.

[4] K. Garg and S. K. Nayar, "Vision and rain," *Int. J. Comput. Vis.*, vol. 75, no. 1, pp. 3–27, Jul. 2007.

[5] S. K. Nayar and S. G. Narasimhan, "Vision in bad weather," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Sep. 1999, pp. 820–827, doi: 10.1109/ICCV.1999.790306.

[6] W. Yang, R. T. Tan, J. Feng, Z. Guo, S. Yan, and J. Liu, "Joint rain detection and removal from a single image with contextualized deep networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 6, pp. 1377–1393, Jun. 2020, doi: 10.1109/TPAMI.2019.2895793.

[7] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, "Deep joint rain detection and removal from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1685–1694, doi: 10.1109/CVPR.2017.183.

[8] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 695–704, doi: 10.1109/CVPR.2018.00079.

[9] R. Li, L.-F. Cheong, and R. T. Tan, "Heavy rain image restoration: Integrating physics model and conditional adversarial learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1633–1642, doi: 10.1109/CVPR.2019.00173.

[10] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha, "Recurrent squeeze-and-excitation context aggregation net for single image deraining," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), vol. 11211, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-030-01234-2_16.

[11] T. Wang, X. Yang, K. Xu, S. Chen, Q. Zhang, and R. W. H. Lau, "Spatial attentive single-image deraining with a high quality real rain dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12262–12271, doi: 10.1109/CVPR.2019.01255.

[12] W. Yang, J. Feng, J. Yang, F. Zhao, J. Liu, Z. Guo, and S. Yan, "Deep edge guided recurrent residual learning for image super-resolution," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5895–5907, Dec. 2017, doi: 10.1109/TIP.2017.2750403.

[13] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.1109/TPAMI.2017.2699184.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015, doi: 10.1109/TPAMI.2015.2389824.

[16] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1715–1723, doi: 10.1109/CVPR.2017.186.

[17] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, "Clearing the skies: A deep network architecture for single-image rain removal," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2944–2956, Jun. 2017, doi: 10.1109/TIP.2017.2691802.

[18] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3943–3956, Nov. 2020, doi: 10.1109/TCSVT.2019.2920407.

[19] M. Khalid, J. Baber, M. K. Kasi, M. Bakhtyar, V. Devi, and N. Sheikh, "Empirical evaluation of activation functions in deep convolution neural network for facial expression recognition," in *Proc. 43rd Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2020, pp. 204–207, doi: 10.1109/TSP49548.2020.9163446.

[20] C. LC, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), vol. 11211, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-030-01234-2_49.

[21] N. Kanopoulos, N. Vasanthavada, and R. L. Baker, "Design of an image edge detection filter using the sobel operator," *IEEE J. Solid-State Circuits*, vol. 23, no. 2, pp. 358–367, Apr. 1988, doi: 10.1109/4.996.

[22] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.

[23] S. Li, W. Ren, J. Zhang, J. Yu, and X. Guo, "Single image rain removal via a deep decomposition-composition network," *Comput. Vis. Image Understand.*, vol. 186, pp. 48–57, Sep. 2019, doi: 10.1016/j.cviu.2019.05.003.

[24] T. T. Gong and J. S. Wang, "Wavelet based deep recursive pyramid convolution residual network for single image rain removal," *IEEE Access*, vol. 8, pp. 195870–195882, 2020, doi: 10.1109/ACCESS.2020.3034238.

[25] P. Xiang, L. Wang, F. Wu, J. Cheng, and M. Zhou, "Single-image de-raining with feature-supervised generative adversarial network," *IEEE Signal Process. Lett.*, vol. 26, no. 5, pp. 650–654, May 2019, doi: 10.1109/LSP.2019.2903874.

[26] X. Huang, B. Du, and W. Liu, "Multichannel color image denoising via weighted schatten P-norm minimization," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, Jul. 2020, pp. 1–8.

[27] Z. Hu, Z. Huang, X. Huang, F. Luo, and R. Ye, "An adaptive non-local Gaussian prior for hyperspectral image denoising," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1487–1491, Sep. 2019, doi: 10.1109/LGRS.2019.2896888.

MA JINGYI has received his master's degree in Information and Communication. He has graduated from the school of University of Science and Technology of China in 2008. He is currently working in the Gansu Branch of China Meteorological Administration Training Centre. He has more than 10 years of research experience in the field of Information and Communication. And he has published more than 20 academic papers in this field in peer-reviewed journals at home and abroad.

ZHANG TIEJUN is currently working with the Institute of Arid Meteorology, China Meteorological Administration.

JING GUODONG is currently working with the China Meteorological Administration Training Center, Beijing.

YAN WENJUN is currently working with the Gansu Branch of China Meteorological Administration Training Centre, Lanzhou, China.

YANG BIN is currently working with the Gansu Branch of China Meteorological Administration Training Centre, Lanzhou, China.

● ● ●