IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 23, 2020, accepted March 18, 2020, date of publication March 26, 2020, date of current version April 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2983436

Multi-Scale Weighted Fusion Attentive Generative Adversarial Network for Single Image De-Raining

XIAOJUN BI[1] AND JUNYAO XING[2]
[1]College of Information Engineering, Minzu University of China, Beijing 100081, China
[2]College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China

Corresponding author: Xiaojun Bi (bixiaojun@hrbeu.edu.cn)

This work was supported in part by the Research on many-objective optimization design method for ship hydrodynamic performance with green and safety under grant #51779050.

ABSTRACT With the rapid development of outdoor vision system, removing rain streaks from a single image has attracted considerable attention as rain streaks can affect the quality of the image taken in rainy days, and interrupt the key information, which will greatly reduce the use value of the image, thus affecting the performance of traffic, safety monitoring and other facilities. Although the deep learning methods have achieved satisfying performance in single image de-raining, there are still two problems: First, the rain streaks contained in one dataset we can use are limited, and in the case of real rainy days, the rain streak density is diverse, it is impossible to accurately classify them. Therefore, the existing rain removal models cannot remove rain streaks properly for images with different rain streak density which attend to over or under rain removal. Secondly, the results of single image after rain removal model often appear the phenomenon of variegated spots, image contrast saturation change and even unsmooth rain streak after rain removal. We use a three-way multi-scale weighted fusion module to enhance the feature extraction, and then generate an attention map through the improved spatial attentive module to accurately locate the location of the rain streaks. After the combination of the two, we will obtain the foreground information of the rain streaks. Through the characteristic of mutual game in the training mechanism of GAN, we can enhance the rain streak location recognition and effectively remove the rain at the same time. Through the training mechanism of the GAN network game, we can enhance the rain line location recognition and effectively remove the rain at the same time. Experiments show that our network achieves superior performance, it has high generalization for different rain streak density, and ensures that the contrast and saturation of the image are not changed.

INDEX TERMS Improved spatial attentive mechanism, single image rain removal, condition generative adversarial networks, multi-scale weighted fusion.

I. INTRODUCTION

Various outdoor vision tasks require accurate scene detection. However, the visibility of the scenery is greatly reduced under rain conditions, and the images are greatly degraded by rain streaks, it causes that various details of images cannot be recognized, and the value of image usage is debased. Moreover, the existing video-based de-raining methods are difficult to deal with the scene with fast background change. The single image rain removal technology can be used as an auxiliary basic work to better process each frame of such video for video-based de-raining, but there is a great challenge that the available information of single image rain removal is limited.

The associate editor coordinating the review of this manuscript and approving it for publication was Vladimir M. Mladenovic.

Therefore, rain removal from input images is an important research, which has a high value of theoretical and practical. In this paper, we aim to propose a method to deal with the single image rain removal problem.

We focus on two difficulties in the single image rain removal problem: the first is that unlike video-based de-raining problem, we cannot use the temporal prior information in a single image. Although rain streaks can show thin and bright lines in the image, it is difficult to accurately detect the rain streaks due to several factors such as wind direction, and the distribution of rain streaks is variable and complex under real conditions. Second, it often leads to over de-rain or under de-rain in the process of rain removal, it is important to retain the original details of the image while removing rain streaks.

(a) (b) (c)

FIGURE 1. **Image de-raining results. (a) Input image. (b) Our method. (c) Result from [2].**

Recently, several deep learning based de-raining methods achieve promising performance. Nonetheless, there are still limitations of the existing methods. For example, there is only one kind of rain streak is simulated in a training dataset, the trained network can only deal with the situation which is similar to the training dataset. For the extreme situation of the real-world images, they often tend to under de-rain or leave unsmooth streaks in the image after rain removal. Zhang and Patel [1] synthetic a large-scale dataset consisting of 12000 images with different rain-density levels (heavy, medium and light), and these networks estimate the rain-density level firstly and fuse the estimated density information into the de-raining network to get the final output. But in reality, we cannot divide all the rain densities into three levels, in some extreme cases, there is still the problem of over de-rain or under de-rain. We still need to improve the generalization ability of the model to deal with this problem. For example, Chen *et al.* [2] proposed a gated context aggregation network for image de-hazing and de-raining, but as shown in Figure 1, it leads to change the contrast and saturation.

In order to address these issues, we propose a Multi-scale Weighted fusion Attentive Generative Adversarial Network called MWA-GANet that can use an attention map to guide the generative adversarial network to generate the rain removal image from a single rainy image. The multi-scale weighted fusion module is used to assist the spatial attentive mechanism to generate a more accurate attention map for improving the detection of variable and complex rain streaks. The rain streaks removal task is based on a generative adversarial network, due to the superiority of GAN in image generation and translation, it has the ability to generate a more realistic image with the guidance of attention map that can retain detail information. A number of evaluation results show that our proposed network achieves an excellent performance of single image de-raining.

Hence, this paper makes the following contributions:

1.We propose a new generative adversarial network to remove rain streaks from a single image, through which we can remove rain streaks while preserving image details. The generator network consists of the multi-scale weighted fusion module, the improved spatial attentive module and the contextual auto-encoder.

2.Under the condition of training a kind of rain streaks density, we can remove rain streaks from a single image of a variety of densities. In this way, we just need to train different rain streaks shapes instead of a large number of datasets to include several rain streaks.

3.Our network achieves superior performance on both the synthetic dataset and real-world images. Thus, we can keep the original saturation and contrast of the background image as much as possible while removing the rain streaks.

II. RELATED WORKS

In the past research, most researchers devoted their attention to the problem of removing rain streak from video [3]–[10]. Other traditional methods regard the problem of removing rain streaks from a single image as a separate problem. They model a rain image as a linear combination of a clean background image x and rain streaks map r to recover x by y. For instance, Kang *et al.* [11] extract the high-frequency layer of a rain image, and separate the rain streaks from it by sparse coding. There are still some methods that include non-local mean filter methods [12], Gaussian mixture model methods [13] and low-rank representation-based methods [14]. However, due to the lack of temporal prior and other key information in the single image de-raining task compared with the video-based de-raining task, the results of rain removal often have problems of over-smoothness that leads to loss of image details.

In recent years, the advantages of deep learning in feature extraction make it make great progress in the problem of removing rain from a single image. The method of removing rain from a single image based on the deep learning model has gradually replaced the traditional image processing method, and become the hotspot of the research on the problem of removing rain from a single image. In 2017, Fu *et al.* [15] and others first introduced the deep learning method into the single image rain removal. They first divided the rain image into high-frequency and low-frequency layers, and constructed an end-to-end single image de-raining network by using the depth residual module to operate the high-frequency layer. To some extent, this method optimizes the image restoration quality after rain removal, but the model is simple and the generalization ability is low. In the same year, Yang *et al.* [16] proposed a multi-task context dilated network to gradually extract the rain streak map to generate the image after rain removal, and adopted the structure of return rain detection to solve the problem of incomplete rain removal, but at the same time, for the real rain day image, the situation of excessive rain removal or lack of rain removal was brought to varying degrees, resulting in the image becoming fuzzy.

With the appearance of GAN network, the game training mechanism between generator and discriminator has made a great breakthrough in the quality of image generation. Therefore, some scholars have gradually applied GAN network to the task of removing rain from a single image and achieved relatively ideal results. In 2017, Zhang *et al.* [17] first use conditional generation countermeasure network to study the task of removing rain from a single image, and add perceptual loss to improve the visual quality of the generated image, which to a certain extent improved the description of image details. In 2018, Qian [18] proposed a GAN model with attentive mechanism built by LSTM to remove raindrops from a single image, and achieved ideal effect. However, raindrops are different from rain streaks. Raindrops are a kind of interference attached to the lens, which can affect the overall quality of the image. However, because most of the raindrops are transparent, some background information can still be reflected through raindrops, thus providing valuable information for image restoration.

Although the above algorithms have made some improvement on the removal effect of rain line of a single image, for different rainfall density, it is easy to lead to excessive or incomplete rain removal. In 2018, Zhang and Patel [1] put forward a 12000 pieces of labels with three kinds of rainfall density, which are currently the most widely used and the synthetic effect is the closest to the real rain streak feature synthesis datasets, and training a residual perception rainfall density classifier to first classify the images of different rainfall density to guide the model to the same degree of rain removal. This method solves the problem of excessive or insufficient rain removal in the previous model, and the effect of image restoration is the best. However, there is a full-connected layer in the network, which requires the input image size to be consistent. For other sizes of rain map, the first step is to scale, which results in the loss of image details when the original size is restored after rain removal. In addition, in the real rain images, we can not divide the rain distribution into several grades accurately, and because of the influence of wind and background, there are often different distribution of rain lines in different parts of an image. Therefore, it is obviously not enough to simulate the rain lines of three rainfall densities.

In 2019, Chen *et al.* [2] proposed an end-to-end gated context aggregation network to solve the problem of removing rain and fog from a single image, and achieved the best results in the training and testing on the data set it used. However, when the real rain image is used for testing, it is found that the network is basically unable to effectively remove rain from the image, and will lead to the change of image contrast saturation. In conclusion, although the deep learning model has made a great breakthrough in the field of single image rain removal, there are still problems such as low generalization ability of the model and poor image detail recovery. Therefore, it is still necessary to improve the generalization ability of the model for rain line recognition and enhance the recovery of image details.

III. PROPOSED METHOD

As the direction and distribution of rain streaks are complex and changeable, we propose a Multi-scale Weighted fusion spatial Attentive Generative Adversarial Network (MWA-GANet) to effectively retain background details while removing rain streaks. Following the idea of generative adversarial networks [19], our backbone network consists of the generative and discriminative networks. The generative network attempts to generate a high-quality image as far as possible that is not affected by rain streaks. The discriminative network will discriminate whether the image generated by the generative network is real. The entire network architecture of the MWA-GANet is shown in Figure 2.

A. GENERATIVE NETWORK

Our generative network consists of three modules: (a) multi-scale weighted fusion module, (b) improved spatial attentive module, (c) contextual auto-encoder. First, the multi-scale weighted fusion module aims to extract effective a feature map for follow-up modules, and then the improved spatial attentive module is used to locate the position of rain streaks accurately to enhance the generalization ability. Finally, we use the attention map to guide the contextual auto-encoder to carry out rain removal operation.

1) MULTI-SCALE WEIGHTED FUSION MODULE

For the task of removing rain in a single image, the position of rain streaks, the shape characteristics of rain streaks and the filling of background information of the position of rain streaks are very important, but the available information in a single image is limited. In fact, the detection of rain streak in a single rain image can be regarded as a small target detection problem. Generally, in the CNN structure of target detection, the convolution layer is used for feature extraction, and the pooling layer is used for feature fusion, so that the model has a certain degree of translation invariance. But there are some problems in this structure. Because in the task of target detection, the receptive field is very important. First of all, generally, how many pixels a point on the feature map can map to the original map determines the upper limit of the target size that the network can detect [20]. An important way to ensure that the receptive field is large enough is through down sampling. However, the problem caused by such operations is that small targets are not easy to detect.

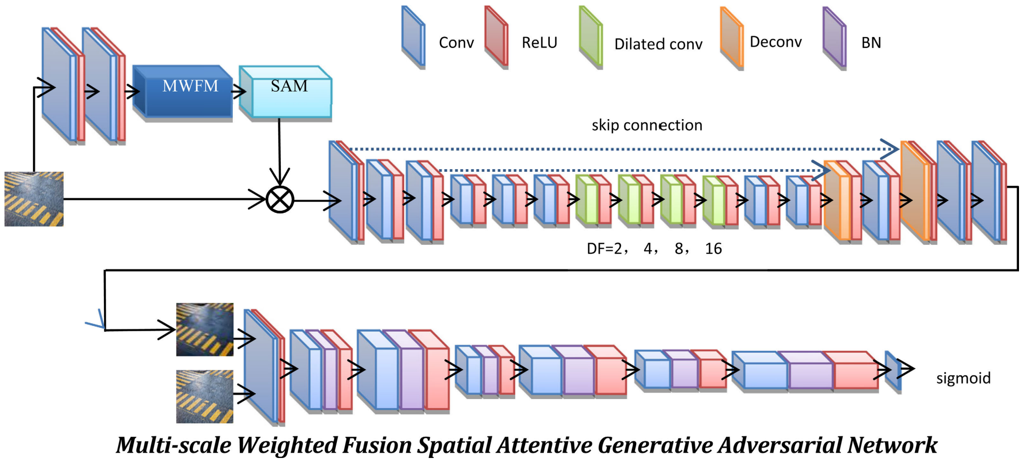

FIGURE 2. **The overview of the proposed MWA-GANet method. The generator consists of three parts: (a) multi-scale weighted fusion module (MWFM), (b) spatial attentive module (SAM), (c) contextual auto-encoder. The discriminator consists of a series of convolutions, batch-normalization and ReLU layers, guided by attention map.**

But if we do not use downsampling, we need to increase the number of convolutions to expand the receptive field, but such an operation will undoubtedly increase the calculation of the network. Therefore, the purpose of dilated convolution is to expand the receptive field of convolution layer without increasing parameters and losing resolution. However, since the convolution kernels of the dilated convolution are spaced, using the continuous dilated convolution with the same dilated factor will lead to the problem of mesh artifacts. In order to solve this problem, we use three-path dilated convolution, each path uses different dilated factors, and then fuse the feature maps extracted from these paths to solve the problem of mesh artifacts. For example, Yang *et al.* [16] use a three-way multi-scale dilated convolution to extract features, and then fuse it to aggregate the extracted rain streak features, and get a satisfied feature extraction effect.

According to the idea of three-path multi-scale dilated convolution, the structure of the residual intensive connection module is used, that is, connect the feature map output by each current basic residual module with the feature map output by the subsequent basic residual module, so as to ensure that the shallow features can be transmitted in the network all the time. And the dilated factor of each path is 1, 2 and 3 to achieve multi-scale feature extraction. Combined with multi-scale feature fusion, on the one hand, it solves the problem of mesh artifacts, on the other hand, it can make the final feature map contain different scale information, so that the network can extract more abundant rain line features.

Because the contribution of feature information of different scales to the final generation of attention map is different, and the role of information of some scales to the network is relatively weak, so we should reduce the proportion of these information in the transmission process, and increase the proportion of other information to achieve the dynamic adjustment of feature map fusion of different paths. Therefore, when the three-path multi-scale feature maps are fused, the idea of weighted fusion is introduced into the model in this chapter. Give different weights to the feature maps of each path, so that the network can selectively integrate features of different scales. The structure of multi-scale weighted fusion module is shown in Figure 3. We first use two consecutive Conv-ReLU layers to generate a preliminary feature map called feature map1, and then enter the following three-way multi-scale weighted fusion module, finally, the combined feature map generated by three-way multi-scale weighted fusion module called feature map2 is fed into the spatial attentive block to generate the attention map.

2) SPATIAL ATTENTIVE MODULE

In general, neural network recognizes objects through a large number of data to train the network's ability to recognize objects, but the whole value of the pictures in the trained network is equal, that is to say, these features are not different in the eyes of neural network, and the network will not pay too much attention to a certain area, so the basic idea of attention mechanism in computer vision is to make the network have ability to ignore irrelevant information and focus on key information. Because of the good performance of recurrent neural networks (RNN) and identity matrix initialization (IRNN) on modeling long-range dependencies generally [21], [22], such as natural language processing task and image processing task to ensure that information can be effectively spread in the whole image, RNN and IRNN which can transfer context information are of great significance to improve visual recognition task. In [23], two consecutive IRNN modules are used to model the contextual information around one pixel to promote the performance of shadow detection. Moreover, we use recurrent neural network (RNN) with context information to construct attention mechanism network generally, and the attention mechanism has achieved remarkable results in rain streaks location in a single image. For example, Qian *et al.* [18] proposed an attentive mechanism recurrent network structured by a convolutional LSTM

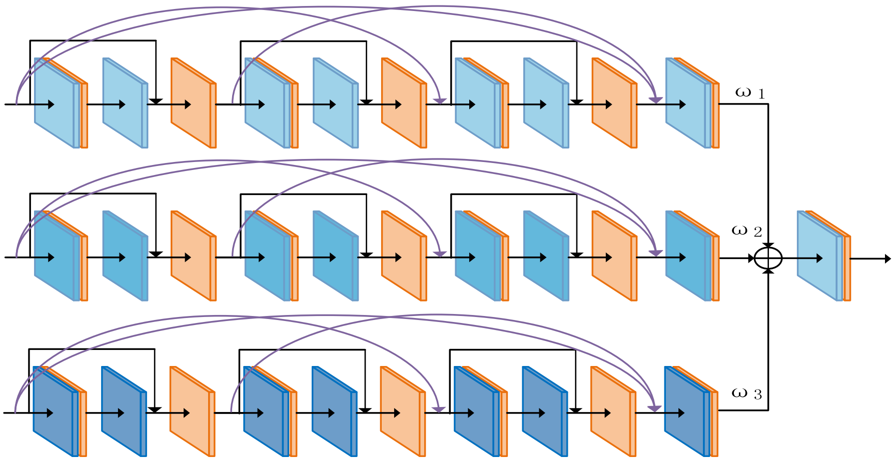

FIGURE 3. The schematic illustration of the multi-scale weighted fusion module. There are three successive residual blocks to refine the feature maps progressively in the three-way multi-scale weighted fusion module for feature extraction. From top to bottom, the dilated factors is 1, 2, 3 respectively. In addition, we use an extra gated fusion sub-network to fuse the feature maps from different levels in each branch. Finally, the combined feature map generated by fusing the three-way output called feature map2 is fed into the spatial attentive block to generate the attention map.

unit. Wang *et al.* [24] confirm that the direction-aware attention mechanism constructed by IRNN also plays an important role in the single image de-raining problem.

Denote fi,j as the feature at pixel (i, j), we summarize one step of data compute operation to the left as:

$$f_{i,j} \leftarrow \max(\alpha_{left} f_{i,j+1} + f_{i,j}, 0) \tag{1}$$

where α left denotes the weight parameter in the recurrent convolution layer for the left direction. From the perspective of the receptive field, the entire module achieves global perception through the following operations. One pixel in the feature map originally can only get its own information, through the first round of IRNN, the pixel in the same position achieves the information of its previous pixel, which achieves the information of its previous pixel, the original-centered cross context information is obtained after repeating the above operation n times in four directions. Then, repeat the whole process, each pixel in the center can feel its own cross-context information, and then generate the whole spatial context.

In [23], spatial attentive mechanism is introduced to detect shadows, and good detection results are obtained. In this paper, two rounds of spatial RNN are used to aggregate the global information of the image. In order to make the spatial RNN selectively utilize the features in different directions, a weight mapping in four directions is trained, and the features in four directions are multiplied by the weight first and then fused to achieve the selective utilization. However, the weight learning is not necessary for the task of removing rain in a single image. In shadow detection, because the shadow target covers more pixels and can use more background correlation information, the context information in different directions is more helpful to determine whether the pixel is a shadow. But the rain streak is relatively small relative to the target, covering fewer pixels, and the complexity of the rain streak direction and its common distribution in the whole image determines that the context information in four directions is equally important. Therefore, it is not necessary for training weight to fuse the feature images in four directions, and the amount of calculation is large. Therefore, in this paper, the weight learning is removed, and the feature images in four directions are fused directly, which reduces the computation while not affecting the performance. The structure of the spatial attentive module is shown in Figure 4.

3) CONTEXTUAL AUTO-ENCODER

We propose a contextual auto-encoder as a generative network. The single image degraded by rain streaks multiplied with attention map is used as the input of the generative network. That is to say, we learn rain streak features as foreground information, and use attention map to guide the image generation. Similar to traditional U-net structure, there are 16 conv-ReLU blocks to perform two downsampling operations and two upsampling operations on the input, and skip connections are added to prevent the loss of high-resolution features during the downsampling. Differently, in the middle of contextual auto-encoder, four successive dilated convolution layers with different dilated factors are used to increase the receptive field to varying degrees without increasing the parameters. In this way, we can not only remove the rain from the pixels of the rain line position, but also fill the occluded part more accurately by using the correlation between the rain line position and the background image. At the same time, connecting different sizes of receptive fields means that different sizes of information can be obtained between each layer, and solve the problem of grid artifacts. The schematic illustration of the Contextual auto-encoder is shown in the top right of Figure 2.

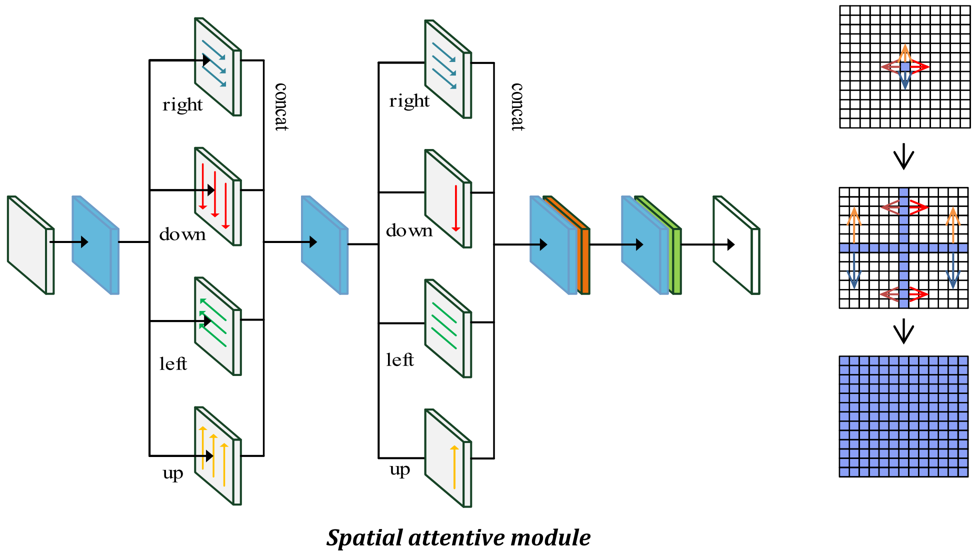

FIGURE 4. The schematic illustration of the improved spatial attentive module. The right represents how spatial context information propagates in two-round IRNN.

B. DISCRIMINATIVE NETWORK

It has been proved that L1 loss can only capture low-frequency information, which may produce a blurry result on image generation. To solve this problem, PatchGAN [25] is used to capture high-frequency information. In an image degraded by rain streaks, different color background areas are degraded by rain streaks to different degrees (for example, light color areas are less degraded). If using the traditional GAN discriminator to average each pixel of the whole picture, the result of individual areas may be weakened and the performance of the discriminator will be affected. Specifically, the N×N size of patch is used as input of discriminator instead of the whole image. This discriminator classifies whether each patch in an image is real, we run discriminator on the whole image and average all the results to avoid the extreme output of the discriminator.

We summarize the calculation formula of the receptive field as:

$$S_{input} = (S_{output} - 1)Stride + Size_{kernel} \tag{2}$$

where Sinput denotes the receptive field size of output node, Stride is step size, and Sizekernel is the size of convolution kernel. Through this formula, it can be calculated that the size of the input receptive field corresponding to the output of the discriminator is 70 × 70.

C. TRAINING DETAILS

1) LOSS FUNCTION

According to the original loss function of the generative adversarial networks as follow:

$$\min_G \max_D E_{O\sim P_{rain}}[\log(1 - D(G(O)))] + E_{B\sim P_{background}}[\log(D(B))] \tag{3}$$

where G is the generative network, and D is the discriminative network. O denotes the output of the generator, and B denotes the clean background. We adopt the following loss function to train our generator:

$$L_G = L_1 + (1 - L_{SSIM}) + L_A + L_{GAN} \tag{4}$$

$$L_{GAN} = \log(1 - D(G(O))) \tag{5}$$

We use the L1 loss to measure the per-pixel reconstruction accuracy to capture low-frequency information, and L1 loss can prevent gradient explosion. L_{SSIM} [26] is used to constrain the structural similarities, the value of SSIM [27] is between 0 and 1, the higher the structural similarity is, the higher the value is. L_A presents the MSE distance between the attention map and the binary mask map, and the binary mask map is obtained by calculating the difference between the rain image and original clean image that used to constrain the generation of the attention map.

We adopt the loss function of the discriminator as follow:

$$L_D = \log(D(B)) + \log(1 - D(B)) \tag{6}$$

2) TRAINING METHOD

The traditional training method of GAN is to alternate between one gradient descent step on D, then one step on G. But in the process of training, we found that the loss of discriminator decreases rapidly due to the strong capability of the discriminator, we train the network by updating the generator twice and updating the discriminator once alternately.

3) DATASET

We use the most realistic and widely used synthetic dataset which is proposed by [1] for training and testing. Its training dataset includes 12000 images of three rain streaks density levels, light, medium and heavy, and there are 4000 images per density level. Similarly, the testing dataset includes 1200 images which degraded by rain streaks with different directions and density level, samples synthetic images in three different conditions show in Figure 5.

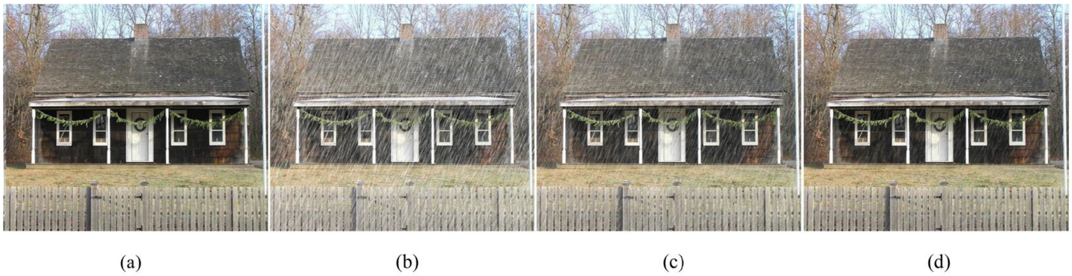

FIGURE 5. (a) Clean background image. (b) Heavy rain streak density level image. (c) Medium rain streak density level image. (d) Light rain streak density level image.

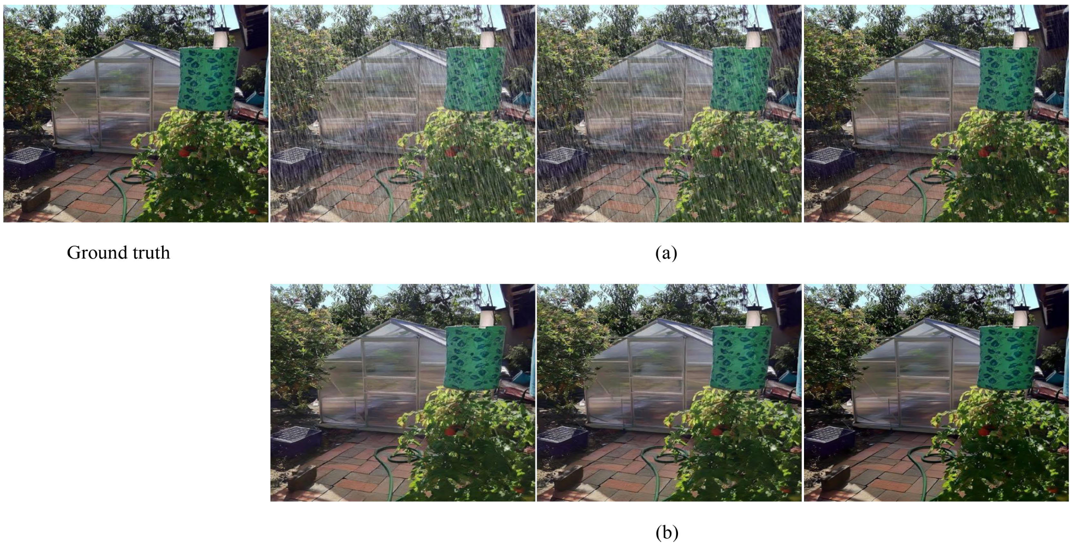

FIGURE 6. Rain-streak removal results on sample images from the synthetic datasets Test1 in three different conditions. (a) Three rain streak density level images. (b) Corresponding to the rain removal results of the images above.

Different form the training method of using the whole training dataset to classify and de-raining with lables, we only choose 2000 images of the medium density level from the training dataset to train our network and test on the whole testing dataset to verify the generalization ability of our network and the effectiveness our method on small sample dataset.

IV. EXPERIEMENTS

In this section, we validate our method on the synthetic testing dataset in terms of PSNR and SSIM compared with the state-of-the-art single-image de-raining methods. Then we analyze the contribution of each component of the proposed network and test the generalization ability of our network on another each component of the proposed network and test the generalization ability of our network on other synthetic dataset. Lastly, the effectiveness of our method is further verified by comparing it with other methods of on real-world images.

A. QUANTITATIVE EVALUATION ON THE PROPSED NETWORK

We test our network on a test dataset containing three rain streak density levels, which is proposed by [1], called *Test1*, as shown in Figure 6. Although we only select 4000 images with medium rain streak density level from the training dataset to train our network, our method has also achieved good rain removal effect for the other two rain streak density level images. And there is no different rain removal result with the different rainfall density. The clarity of the image and the recovery of the details are consistent. The results show that without a large number of training datasets composed of different rain line density layers, our network can process different degrees of rain streak degradation images, which proves that our network can carry out high-quality rain removal for images with different rainfall density without classification.

Then, we compare the quantitative performance of our method with previous state-of-the-art single image de-raining methods on *Test1*. From Figure 7, we can see that some networks are not able to achieve the effect of rain removal, like [16], and some networks cannot recover clear background images after removing rain streaks, for example, the images generated by [15] have faults, and some methods tend to change the contrast of images, like [1]. However, the proposed MWA-GANet can identify the rain streaks

FIGURE 7. Rain-streak removal results on sample images from the synthetic datasets Test1 compared with other state-of-the-art single-image de-raining methods.

TABLE 1. Quantitative results evaluated in terms of average SSIM and PSNR. Part of the results are cited from [1]

Methods	Input	GMM[13] (CVPR'16)	CNN[28] (TIP'17)	JORDER[16] (CVPR'17)	DDN[15] (CVPR'17)	JBO[29] (ICCV'17)	DID[1] (CVPR'18)	MWA-GANet
PSNR(dB)	21.15	22.75	22.07	24.32	27.33	23.05	27.95	29.05
SSIM	0.7781	0.8352	0.8422	0.8622	0.8978	0.8522	0.9087	0.9181

accuracy and remove them to generate a smooth clear background without losing details. In addition, our method can better retain the original contrast and saturation of the images. For the quantitative evaluation index, we achieve a superior de-raining performance in the terms of both PSNR and SSIM, as shown in Table 1, and we can better retain the contrast and saturation of the original image visually.

B. ABLATION ANALYSIS FOR EACH COMPONENT

In order to verify the effectiveness of each component added in our structure, we carry out the ablation analysis for different combinations of each component on another dataset. As shown in Table 2, GAN is our baseline module, the second row indicates whether there is an improved attention mechanism module or not, the third row indicates there is a three-way multi-scale dilated convolution network or a single-way network without dilated convolution, and the fourth row indicates that the network has gated fusion or outputs directly after three dilated residual modules without shallow feature fusion. The results show that each component of our network improves the final result to a certain extent.

In order to more intuitively observe the rain removal effect of each model structure in the ablation experiment from the visual effect, the rain removal effect of the ablation experiment in other data sets of this method is shown in Figure 8. Compared with the test results of (c) GAN and (e) GAN + multi-scale weighted fusion module, it can be seen that the multi-scale weighted fusion module can effectively improve the description of image details, but due to the lack of spatial attentive module to enhance the recognition and location of rain streaks, only a few rain streaks with similar features can be removed. By comparing the test results of (d) GAN + improved spatial attentive module structure with the complete structure of (f) GAN + multi-scale weighted fusion module + improved spatial attentive module proposed in this chapter, it can be seen that after adding multi-scale weighted fusion module, on the one hand, it can help the improved spatial attentive module to better identify and locate the rain streaks, on the other hand, it can improve the quality of the image after rain removal, and the ability of depicting image details is stronger. Through ablation experiments, the effectiveness of

TABLE 2. Quantitative results of ablation study evaluated in terms of average PSNR.

GAN	√	√	√	√
Improved spatial attentive mechanism	-	√	-	√
Multi-scale weighted fusion	-	-	√	√
PSNR（dB）	25.2860	27.4345	26.4364	28.2316

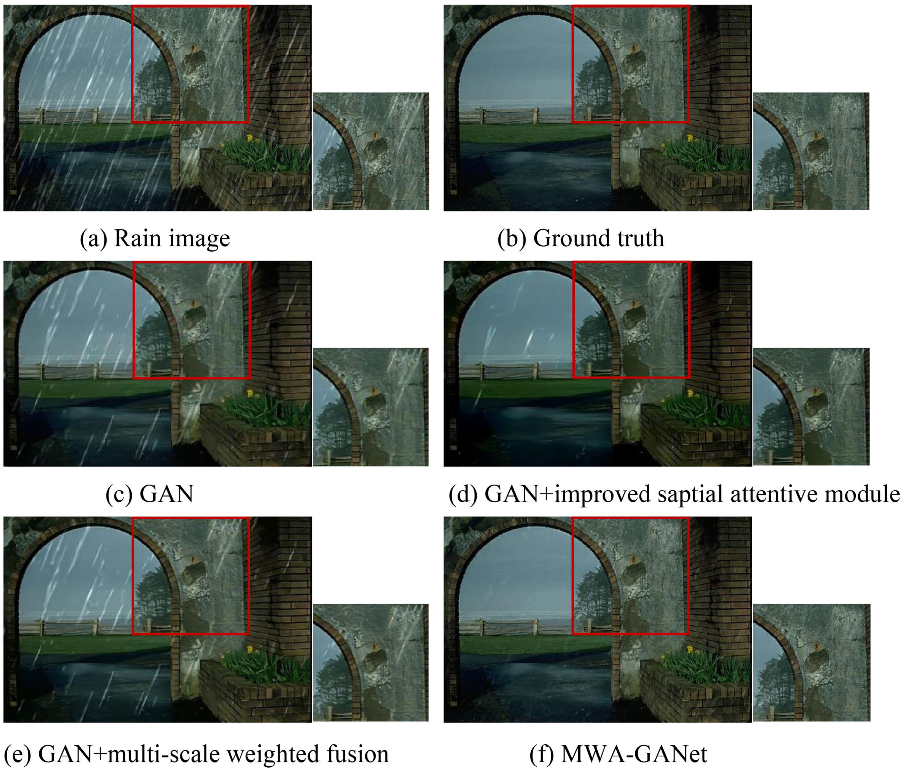

FIGURE 8. Rain-streak removal results of ablation study on sample images from another synthetic datasets.

each module in the proposed method is verified, which can effectively remove different rain streaks and restore the image details better.

C. QUALITATIVE EVALUATION ON THE PROPOSED NETWORK

In order to verify that the single image rain model proposed in this paper can also be applied to real-life rain images and highlight the practical application value of the model. In this paper, we test the real rain images taken in rainy days by downloading them randomly from the network. The test results are shown in Figure 9. The first line is the rain image taken on a real rainy day, and the second line is the image after the rain is removed by the method in this chapter. From the test results, it can be seen that the single image rain removal model proposed in this paper can effectively remove rain from the real world rain images under various rainfall conditions while retaining the image details, and the contrast saturation of the image after rain removal is consistent with the original image, and the obtained rain removal image is clear and the quality of the realistic image is higher. It solves the problem that the current single image rain removal network does not have high generality for different rain lines, and the fuzziness or even the inability to remove rain lines for real rain images.

We also evaluated the proposed MWA-GANet on real-world images and compare with the previous state-of-the-art single image de-raining methods. The Figure 10 shows the effect of rain removal on real-world images, we can see that our MWA-GANet can identify the rain streaks accurately under different rainfall conditions and generate a clearer image with less blur, and closer to input images in contrast and saturation compared with other methods. Moreover, the image generated by our method has less mottled area.

FIGURE 9. The results of removing rain in real rainy days.

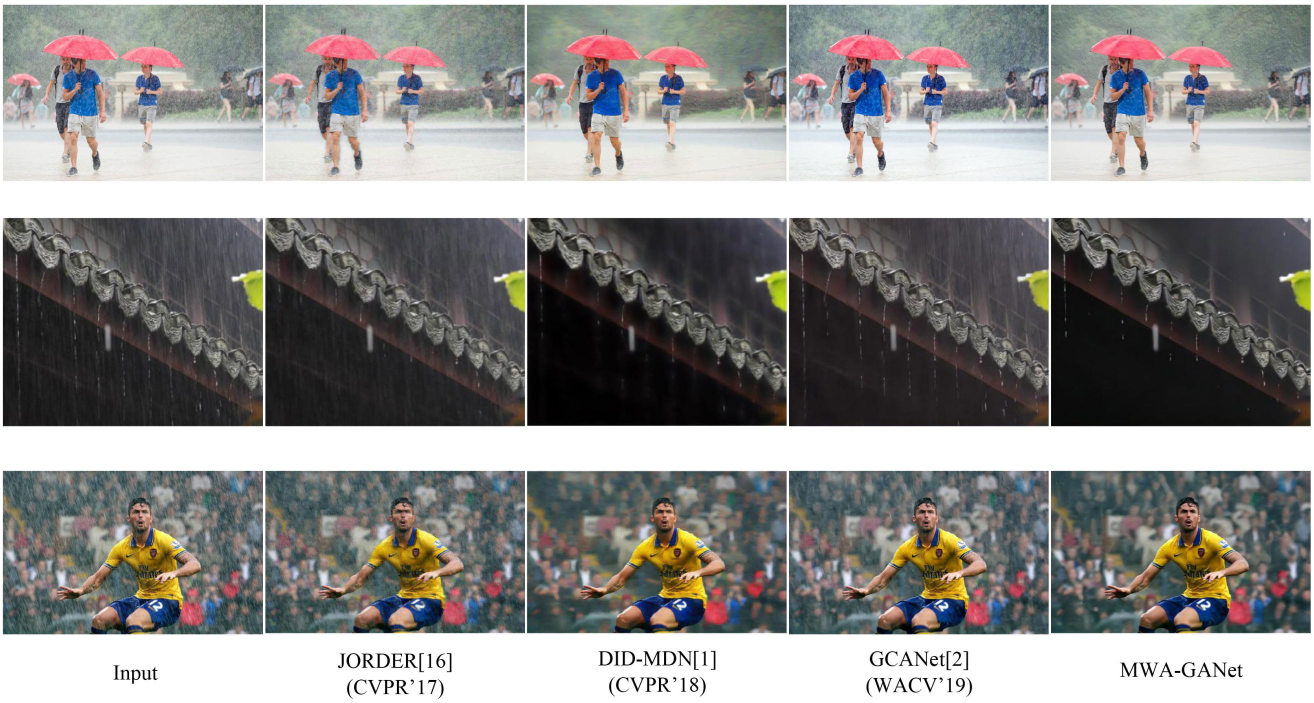

FIGURE 10. Rain-streak removal results on sample real-world images compared with other state-of-the-art single-image de-raining methods.

V. CONCLUSION

In this paper, we proposed a multi-scale weighted fusion attentive generative adversarial network for single image de-raining. The generative network uses the multi-scale weighted fusion module to generate the feature map to provide valuable rain streak features for the follow-up spatial attentive module, and then the spatial attentive module generates an attention map to guide the contextual auto-encoder to generate the rain removal background image. Afterwards, the discriminative network classifies whether the image is real or not. We solve the problem that the existing networks require training datasets with a variety of rain streaks density level images to avoid over de-raining or under de-raining. We only use the dataset with one rain streaks density level to train our network and achieve excellent rain removal results for images of the other several rain streaks density levels. At the same time, it improves the quality of the image after rain removal and enhances the ability of detail recovery. Experiments and comparisons performed on two synthetic datasets and real-world images prove that the proposed MWA-GANet is superior to the state-of-the-art existing single image de-raining methods. In addition, we carry out the ablation analysis for different combinations of each component to illustrate the contribution of each component and use another dataset to prove the effectiveness of the spatial attentive mechanism module in visual effect.

However, our network has limitations. For instance, rainy days are often accompanied by fog. For real-world rainy images, we can only remove the rain streaks, but we cannot restore the clear background image without fog. Because it is impossible for us to obtain the real-world rainy-rainless image pairs, and in the synthetic images, only the rain streaks are synthesized, the fog is not simulated. In further research, we can improve the synthetic image pairs in order to train the networks to restore clear images after removing the fog while removing the rain streaks.

In this paper, through the analysis and research on the current situation of single image rain removal, we find that there are still some problems waiting for further research. Although it is difficult to obtain real rain - free images, it is not difficult to take high-quality pictures of rainy days. With the development of unsupervised / semi supervised training network, a semi supervised network is designed for the task of removing rain from a single image [30]. It can be used as the next research direction to add real rain images into the network and train together with other data sets, which has high research value.

ACKNOWLEDGMENT

The authors would like to thanks to the editors and reviewers for time and effort in this article.

REFERENCES

[1] 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.
[2] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, "Gated context aggregation network for image dehazing and deraining," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 1375–1383.
[3] S. Starik and M. Werman, "Simulation of rain in videos," in *Proc. ICCV*, Oct. 2003, pp. 406–409.
[4] X. Zhang, H. Li, Y. Qi, W. Leow, and T. Ng, "Rain removal in video by combining temporal and chromatic properties," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jul. 2006, pp. 461–464.
[5] X. Zhao, P. Liu, J. Liu, and T. Xianglong, "The application of histogram on rain detection in video," in *Proc. 11th Joint Conf. Inf. Sci. (JCIS)*, 2008, pp. 1–8.
[6] M. Zhou, Z. Zhu, R. Deng, and S. Fang, "Rain detection and removal of sequential images," in *Proc. Chin. Control Decis. Conf. (CCDC)*, May 2011, pp. 615–618.
[7] S. Krishnan, "Restoration of video by removing rain," *Int. J. Comput. Sci., Eng. Appl.*, vol. 2, no. 2, pp. 19–28, Apr. 2012.
[8] N. L. N. Brewer, "Using the shape characteristics of rain to identify and remove rain from video," in *Structural, Syntactic, and Statistical Pattern Recognition*. Berlin, Germany: Springer, 2008.
[9] V. Santhaseelan and V. K. Asari, "Utilizing local phase information to remove rain from video," *Int. J. Comput. Vis.*, vol. 112, no. 1, pp. 71–89, Mar. 2015.
[10] K. Garg and S. K. Nayar, "Detection and removal of rain from videos," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2004, p. 1.
[11] L.-W. Kang, C.-W. Lin, and Y.-H. Fu, "Automatic Single-Image-Based rain streaks removal via image decomposition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1742–1755, Apr. 2012.
[12] J.-H. Kim, C. Lee, J.-Y. Sim, and C.-S. Kim, "Single-image deraining using an adaptive nonlocal means filter," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 914–917.
[13] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, "Rain streak removal using layer priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2736–2744.
[14] Y.-L. Chen and C.-T. Hsu, "A generalized low-rank appearance model for spatio-temporally correlated rain streaks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1968–1975.
[15] 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. 3855–3863.
[16] 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. 1357–1366.
[17] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," 2017, *arXiv:1701.05957*. [Online]. Available: http://arxiv.org/abs/1701.05957
[18] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2482–2491.
[19] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[20] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. [Online]. Available: https://arxiv.org/abs/1511.07122
[21] S. Bell, C. L. Zitnick, K. Bala, and R. Girshick, "Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2874–2883.
[22] Q. V. Le, N. Jaitly, and G. E. Hinton, "A simple way to initialize recurrent networks of rectified linear units," 2015, *arXiv:1504.00941*. [Online]. Available: http://arxiv.org/abs/1504.00941
[23] X. Hu, L. Zhu, C.-W. Fu, J. Qin, and P.-A. Heng, "Direction-aware spatial context features for shadow detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7454–7462.
[24] 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. 12270–12279.
[25] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.
[26] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.
[27] 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.
[28] 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.
[29] L. Zhu, C.-W. Fu, D. Lischinski, and P.-A. Heng, "Joint bi-layer optimization for single-image rain streak removal," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2526–2534.
[30] J. Xing and X. Bi, "Dilated residual spatial attentive generative adversarial network for single image de-raining," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 790, Apr. 2020, Art. no. 012033.

XIAOJUN BI was born in 1964. She received the M.S. degree from the Harbin Institute of Technology, Harbin, China, in 1990, and the Ph.D. degree from Harbin Engineering University, Harbin, in 2006. She is currently a Professor and the Ph.D. Supervisor with the College of Information Engineering, Minzu University of China, Beijing, China. Her research interests include evolutionary computation, multiobjective optimization, and data mining.

JUNYAO XING received the bachelor's degree in communication engineering from Harbin Engineering University, in 2017, where she is currently pursuing the master's degree in information and communication engineering. Her research interest includes image processing technology based on deep learning.