Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 6668771, 8 pages
https://doi.org/10.1155/2021/6668771

Research Article

Generative Adversarial Network for Image Raindrop Removal of Transmission Line Based on Unmanned Aerial Vehicle Inspection

Changbao Xu,[1] Jipu Gao,[1] Qi Wen,[1] and Bo Wang[2]

[1]*Electric Power Research Institute of Guizhou Power Grid Co., Ltd., Guiyang 550000, China*
[2]*School of Electrical and Automation Engineering, Wuhan University, China*

Correspondence should be addressed to Bo Wang; whwdwb@whu.edu.cn

Received 9 December 2020; Revised 9 February 2021; Accepted 9 March 2021; Published 23 March 2021

Academic Editor: Mohammad R. Khosravi

Copyright © 2021 Changbao Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the process of UAV line inspection, there may be raindrops on the camera lens. Raindrops have a serious impact on the details of the image, reducing the identification of the target transmission equipment in the image, reducing the accuracy of the target detection algorithm, and hindering the practicability of UAV line inspection technology in cyber-physical energy systems. In this paper, the principle of raindrop image formation is studied, and a method of raindrop removal based on generation countermeasure network is proposed. In this method, the attention recurrent network is used to generate the raindrop attention map, and the context code decoder is used to generate the raindrop image. The experimental results show that the proposed method can remove the raindrops in the image and repair the background image of raindrop coverage area and can generate a higher quality raindrop removal image than the traditional method.

1. Introduction

UAV inspection image is the most important information carrier in Industrial Internet of Things (IIoT). The purpose of intelligent inspection can be achieved through the target detection and fault location of the machine inspection image. Sometimes there are raindrops on the camera in the process of UAV line patrol, which will cover the information of the target object in the background image and reduce the image quality. Raindrops make the transmission line equipment absorb a wider range of environmental light when imaging, and the superposition of these refracted light and the reflected light of the target object causes the image degradation. In addition, the camera should focus on the transmission line equipment when the UAV takes photos during the line patrol, and the presence of raindrops will affect the camera's focus, making the image background virtual, and the image detail information loss is serious, so the follow-up operation of the machine patrol image with raindrops will be extremely difficult. Therefore, the existence of raindrops will lead to the uneven quality of the machine patrol image, which will affect the extraction and utilization of image information and reduce the accuracy and reliability of target detection.

In the field of image processing, single image raindrop removal is an extremely complex technology. There are not many existing methods to carry out relevant technical research for a long time. These methods can be roughly divided into traditional raindrop removal methods and CNN-based raindrop removal methods. The traditional raindrop removal methods are divided into filtering and the learned dictionary plus sparse coding methods. Filtering includes guided filter [1], improved guided filter [2], multi-guided filter [3], LO smoothing filter [4], and nonlocal means filter [5]. The image of raindrop removal generated by filtering is fuzzy, and some raindrops cannot be removed. Fu et al. [6] use the filter to filter the image containing raindrops to get the high-frequency and the low-frequency image, use learned dictionary plus sparse coding to remove raindrops from the high-frequency image, and then combine the high-frequency image and the low-frequency image to get the raindrop image. On this basis, Kang et al. [7] introduced the raindrop HOG feature and used the K-means clustering method to cluster the high-frequency images to obtain the

rain dictionary and the rain-free dictionary and then sparse coding, respectively, to obtain the high-frequency rain-free image and the high-frequency raindrop-free image and the low-frequency image fusion to obtain the raindrop-free image. The image background obtained by this method is clearer than that obtained by Fu's method. Lou et al. [8] proposed a discriminative sparse coding method to remove image raindrops. This coding method has certain discrimination ability, which can reduce the error rate of raindrop discrimination and improve the effect of raindrop removal. In 2013, David et al. [9] first used convolutional neural network for image raindrop removal. Firstly, a sample database containing raindrop-free image pairs was constructed, and the corresponding image was segmented by a sliding window with step length of 1. Then, the network was trained by the mean square error between corresponding image blocks, and finally, the convolutional neural network model capable of raindrop removal was obtained. After that, Fu [10, 11] and others fused the convolution neural network and image decomposition, using the convolution neural network to extract the raindrop feature in the image, as the raindrop feature in the high-frequency component to achieve the raindrop removal in the high-frequency component, and eventually improve the quality of the raindrop removal effect image.

Through the research on the existing methods, we found that most of the traditional methods of raindrop removal are based on the model. The traditional model is used to describe raindrops, rain lines, and background images, respectively, and with the corresponding algorithm, using step by step iterative optimization to remove the raindrop. The traditional method is not ideal for the image processing with dense raindrops; the background image covered by raindrops cannot be repaired precisely. The method based on convolution neural network can fully extract the feature information of the image, and the effect of using this method to remove the raindrop is better.

However, with the increase of network depth, the network is prone to overfitting, and the effect of raindrop removal is difficult to be further improved. In view of the shortcomings of the above algorithm, this paper analyzes the principle of raindrop image generation and then discusses the basic structure of GAN. On this basis, the raindrop image generation model is integrated into the GAN, and a raindrop removal method based on the GAN is proposed. The raindrop image obtained by this method is closer to the real image.

2. Single Image Raindrop Removal Model

2.1. Image Generation Model with Raindrops. In the process of image raindrop removal, the raindrop image is usually modelled as a linear combination of background image and raindrop layer, and the mathematical expression is shown as equal

$$I(x) = (1 - M(x)) \odot B(x) + R(x). \tag{1}$$

I represents the raindrop image taken by the UAV during the line patrol, x is the pixel position in the image, and B is the background image, that is, the UAV takes clear transmission line equipment. R is the impact of raindrops on the image, and M is the binary mask, which is used to represent the impact of raindrops on the background image.

2.2. Generative Adversarial Networks. In recent years, with the continuous development of deep learning, scholars put forward the generative adversarial network (GAN), which has good performance in dealing with complex data distribution and is one of the most promising methods in the field of unsupervised learning. The model contains generating module and discriminating module. In the aspect of image restoration, by the game between the two modules, high-quality images can be output.

The core idea of GAN is game. The generation model is used to generate a realistic sample, and the discrimination model is used to judge the authenticity of the generated image. The discrimination network needs to be able to distinguish whether the input image is a real picture or a picture generated from the generated network. If it is a real picture, output 1; otherwise, output 0. The network generates new pictures according to the pattern of the real pictures. By playing games with the discrimination network, the quality of the generated pictures is as close to the real pictures as possible so that the discriminator cannot recognize the image from the generator. In order to achieve this function, the generation network and the GAN need to be trained alternately and iteratively.

Learning complex data distribution quickly is the strong point of GAN. Also, the network does not need complex constraint functions, and the whole learning process does not need human intervention. Another feature of GAN is that it can update the loss function of the network by itself depending on the distribution of sample data. In the process of training the generative adversarial network, the discriminative network can be used as the loss function of the generative network, which plays a role of supervision and guidance for the optimization of the generated network. The process of judging network parameter updating is also the process of optimizing the network loss function.

2.3. Raindrop Removal Model Based on Generative Adversarial Network. Same as the basic structure of GAN, the raindrop model based on generative adversarial network mainly includes generative network and discriminative network. Under the guidance of attention map, clear and real raindrop removal images are generated as far as possible. The overall architecture of raindrop removal network is shown in Figure 1. The improved generative network and discrimination network will be described in detail below.

The whole loss function of raindrop model based on GAN is shown in

$$\min_G \max_D \left\{ E_{R\sim P\text{clean}}[\log(D(R))] + E_{I\sim P\text{raindrop}}[\log(1 - D(G(I)))] \right\}, \tag{2}$$

where G stands for generating network and D stands for

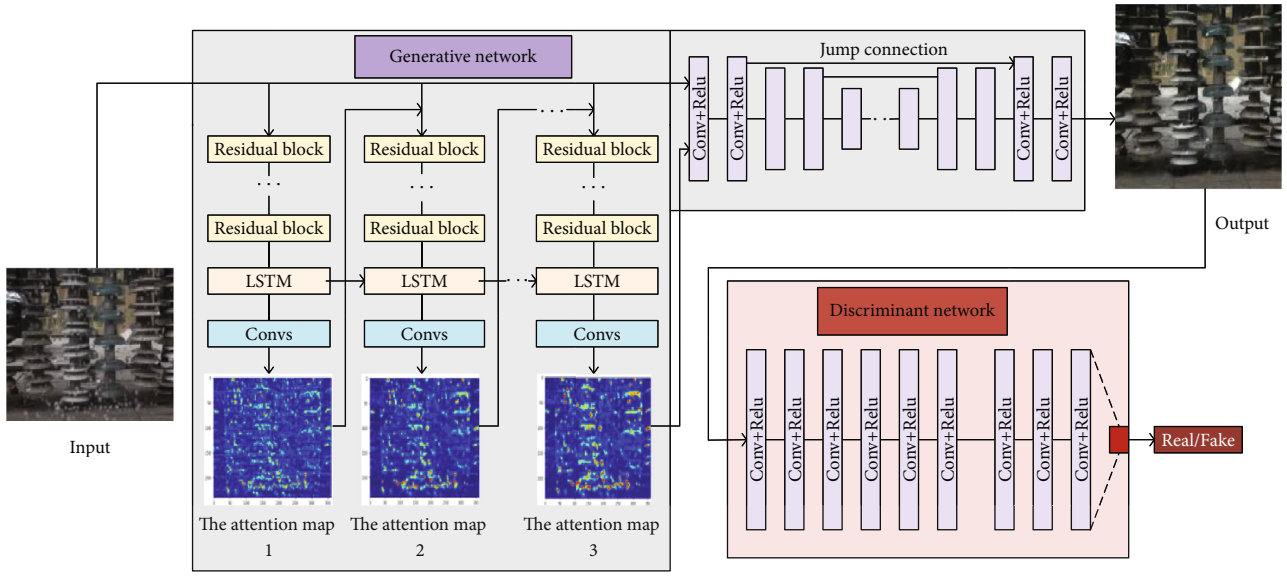


Figure 1: Diagram of the improved generative network consists of two subnetworks: attention recurrent network and context autoencoding decoding network.

discriminating network. I is the image with raindrop, $G(I)$ is the image after raindrop removal, and R is the real sample without raindrop.

2.3.1. Improved Generative Network. As shown in Figure 1, the improved generative network consists of two subnetworks: attention recurrent network and context autoencoding decoding network. LSTM network is included in the attention recurrent network [12], which generates attention map by cyclic iteration. Attention map contains the location and shape information of raindrops in raindrop image, which guides the context codec to focus on raindrops and their surrounding areas.

(1) Attention Recurrent Neural Network. The attention recurrent network is used to locate the target area in the visual attention model to improve the accuracy of target recognition [13–16]. Inspired by this, this paper applies this structure to the raindrop removal network and uses the visual attention guidance generative network and distinguish network to find the location of raindrops in the image. As shown in the generator part of Figure 1, the attention recurrent network consists of four circulation modules, each of which contains a packet residual network [17, 18], an LSTM unit, and a convolution layer, wherein the residual module is used to extract the raindrop feature information from the input image and the attention map generated by the previous recurrent module, and the LSTM unit [19, 20] and the convolution layer are used to generate a 2D attention map.

Binary mask plays a key role in the generation of attention map. There are only two numbers 0 and 1 in the mask. 0 means there is no raindrop in this pixel, and 1 means there is raindrop in this pixel. The mask image and the raindrop image are input into the first recurrent module of the attention cycle network for the generation of the initial attention map. The mask image is obtained by subtracting the clear image from the image with raindrops and then setting a certain threshold value to filter. Although the obtained mask image is relatively rough, it has a great effect on the generation of fine attention map. The biggest distinction between attention graph and mask graph is that the mask graph only contains 0 and 1, and the value of attention graph is [0, 1]. The larger the median value of the attention graph indicates that the more attention should be paid to the pixel, that is, the more likely there are raindrops at the pixel. Even in the same raindrop area, the value of attention map will be different, which is related to the shape and thickness of raindrops and also reflects the influence of raindrops on different pixels of background image.

The attention recurrent network contains a LSTM (Long Short-Term Memory). The LSTM unit includes an input gate i_t, a forgetting gate f_t, an output gate o_t, and a unit status C_t. The interaction between state and gate in time dimension is defined in

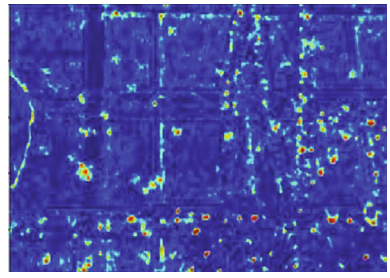
$$\begin{aligned} i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \odot C_{t-1} + b_i), \\ f_t &= \sigma\left(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \odot C_{t-1} + b_f\right), \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tanh\left(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c\right), \\ H_t &= o_t \odot \tanh\left(C_t\right), \end{aligned} \tag{3}$$

where X_t is the image feature generated by the residual network, C_t represents the state feature to be transferred to the next LSTM unit, H_t is the output feature of LSTM unit, $\odot$ is matrix multiplication, and $*$ is convolution operation.

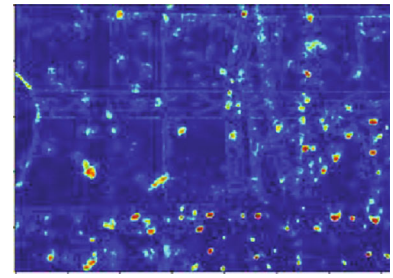
The input of the generated network is an image pair with the same background scene, one with raindrops and one without raindrops. The loss function of each recurrent

(a) The image contains raindrops

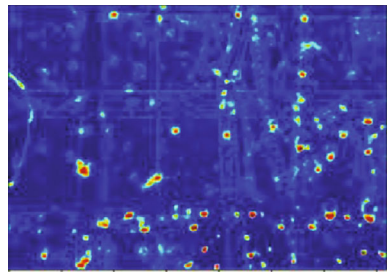
(b) Raindrop removal image



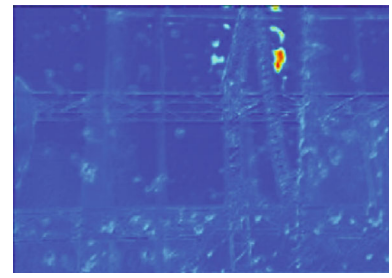
(c) The attention map of raindrops



(d) The attention map of raindrops 2



(e) The attention map of raindrops 3



(f) The attention map of raindrops 4

Figure 2: Raindrop removal image and attention map.

module is defined as the mean square error (MSE) between the output attention map and the binary mask M. For recurrent cycle network, the front-module loss function is given a smaller weight, and the back-module loss function is given a larger weight. The loss function is shown in

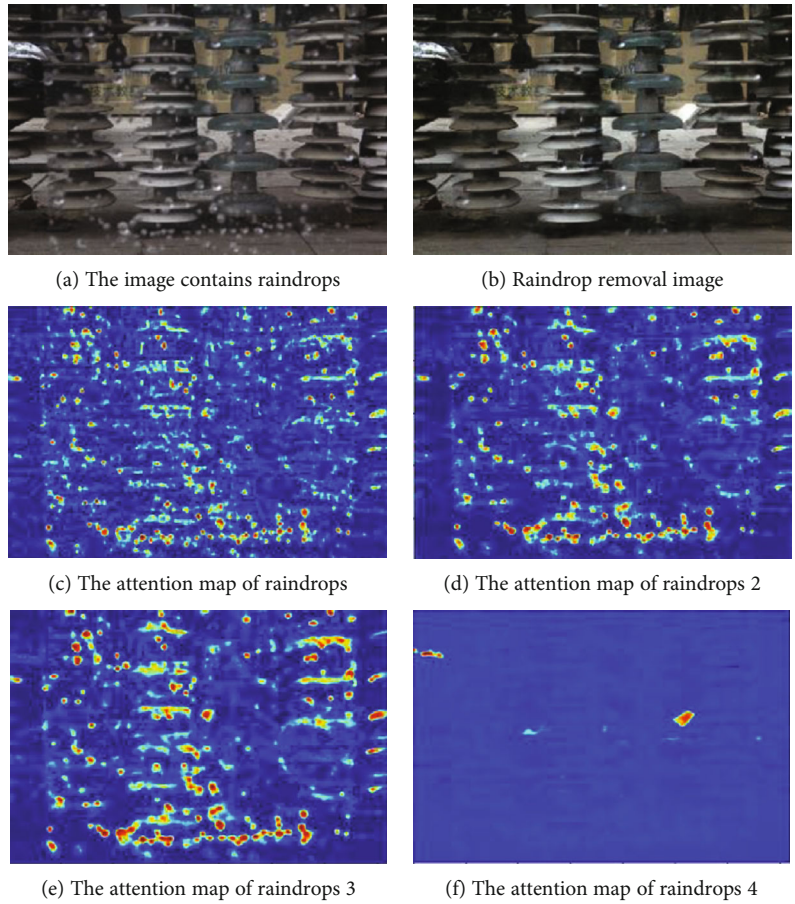
$$L_{\mathrm{ATT}}(\{A\}, M) = \sum_{t=1}^{N} \theta^{N-t} L_{\mathrm{MSE}}(A_t, M), \tag{4}$$

where A_t is the attention graph generated by the cyclic network in time step t. $A_t = \mathrm{ATT}_t(F_{t-1}, H_{t-1}, C_{t-1})$, F_{t-1} represents the fusion of the image with raindrops and the output attention map of the previous recurrent unit. In the whole recurrent network, the larger N is, the finer attention map is generated. But the larger N is, the more memory is needed to store the intermediate parameters. It is found that the network efficiency is the highest when $N = 4, \theta = 0.8$.

(2) Context Automatic Encoder-Decoder. The input of the context auto codec is the attention map generated by the raindrop image and the attention recurrent network. The raindrop removal and background restoration are achieved under the guidance of the attention map. There are 16 conv-relu modules in the context autoencoder-decoder. The structure of coding and decoding is symmetrical. Skip connection is added between corresponding modules to prevent the image from being blurred. There are two loss functions used in the context autoencoder-decoder, multiscale loss and perceptual loss. Multiscale loss function extracts image feature information from different layers of decoder and makes full use of multilevel image information to optimize the model to obtain clear image of raindrop removal. The multiscale loss function is defined as

$$L_M(\{S\}, \{A\}) = \sum_{i=1}^{M} \lambda_i L_{\mathrm{MSE}}\left(S_i, A_{N_i}\right), \tag{5}$$

where S_i represents the image features extracted from the i-th layer of the encoder, A_{N_i} represents the real image which has the same scale with S_i, and $\{\lambda_i\}_{i=1}^{M}$ is the weight of different scales. The design of loss function pays more attention to feature extraction on large-scale image, and the smaller size image contains less information which has little influence on model optimization. The output image sizes of the last layer, the last third layer, and the last fifth layer of the decoder are 1/4, 1/2, and 1 of the original sizes, respectively, and the corresponding weights λ are set to 0.6, 0.8, and 1.0.



(a) The image contains raindrops (b) Raindrop removal image

(c) The attention map of raindrops (d) The attention map of raindrops 2

(e) The attention map of raindrops 3 (f) The attention map of raindrops 4

Figure 3: Raindrop removal image and attention map.

Table 1: PSNR and SSIM of deraindrop image.

Method	PSNR	SSIM
Yang raindrop removal method	19.1538	0.7128
Fu raindrop removal method	19.8693	0.8176
Raindrop removal based on GAN	31.5710	0.9023

In addition to the pixel-based scale loss, this paper also increases the perceptual loss [21] to obtain the global difference between the output of the automatic context encoder-decoder and the corresponding clear picture. Perceptual loss measures the difference between the raindrop removed image and the real image from the global perspective, which will make the raindrop image closer to the real sample. The image global information can be extracted by vgg16, and the network pretraining needs to be completed on the ImageNet data set in advance. The perceptual loss function is defined as

$$L_P(O, T) = L_{\mathrm{MSE}}(\mathrm{VGG}(O), \mathrm{VGG}(T)). \tag{6}$$

VGG is a pretrained CNN, which can complete the feature extraction of a given input image. O is the output image of the automatic encoder, $O = G(I)$, and T is a real image sample without raindrops. To sum up, the loss function of the generated network is defined as

$$L_G = 10^{-2} L_{\mathrm{GAN}}(O) + L_{\mathrm{ATT}}(\{A\}, M) + L_M(\{S\}, \{A\}) + L_P(O, T), \tag{7}$$

where $L_{\mathrm{GAN}}(O) = \log(1 - D(O))$.

2.3.2. Improved Discrimination Network. The function of discriminating network is to distinguish true and false samples. The discriminator in GAN usually uses global discriminator [22–24]. Determine the difference between the image output by the generator and the real sample. Only using global information to judge whether the image is true or false is not conducive to the restoration of local image information by generating network. For image raindrop removal, this method hopes to restore the details of the image as much as possible, so as to carry out the subsequent target detection work. The existing discrimination network cannot be used directly. Therefore, this paper combines the global discriminator and the local discriminator to determine the true and false output samples of the generated network together.

The use of the local discriminator is based on knowing the location information of raindrops in the image. The attention map is generated in the attention cycle network of the image restoration stage, which solves the problem of

Table 2: Target detection results.

Method	Tower failure (AP)	Small size fittings (AP)	Ground conductor failure (AP)	Insulator failure (AP)	mAP
Yang raindrop removal method	0.5164	0.4436	0.5019	0.5482	0.5025
Fu raindrop removal method	0.5548	0.4931	0.5326	0.5931	0.5343
Raindrop removal based on GAN	0.7684	0.5689	0.6143	0.6849	0.6591

location of raindrops in the image. Therefore, attention map can be introduced into the discriminator network to guide the local discriminator to automatically find the raindrop area in the image. CNN is used to extract features from the inner layer of the discriminator. At the same time, it also extracts features from the raindrop image generated by the generator. Then, the loss function of the local discriminator is formed by combining the obtained feature image and attention image. The existence of attention map will guide the discrimination network to pay more attention to the raindrop area in the image. In the last layer of the discrimination network, the full connection layer is used to judge the authenticity of the input image. The overall structure of the discrimination network is shown in the lower right part of Figure 1. The whole loss function of the discrimination network can be expressed as

$$L_D(O, R, A_N) = -\log(D(R)) - \log(1 - D(O)) + \gamma L_{map}(O, R, A_N), \tag{8}$$

where γ is 0.05, the first two terms of the formula are the loss function of the global discriminator, L_{map} represents loss function of local discriminator, and the loss function of local discriminator is shown in

$$L_{map}(O, R, A_N) = L_{\text{MSE}}\left(D_{\text{map}}(O), A_N\right) + L_{\text{MSE}}\left(D_{\text{map}}(R), 0\right). \tag{9}$$

D_{map} represents the two-dimensional attention mask function generated by the discrimination network, and R represents the sample image extracted from the real and clear image database. 0 represents the attention map with only 0 value, which represents there is no raindrop in the real image, so attention map is not required to guide the network to extract features.

The discriminant network in this paper consists of seven convolution layers, the core of which is (3, 3), the full connection layer is 1024, and the single neuron uses the Sigmoid activation function.

3. Model Training

3.1. Data Set Formation. For the training of raindrop removal network proposed in this paper, a set of transmission line equipment image pairs is needed. Each pair of images contains exactly the same background scene, one of which contains raindrops and the other has no raindrops.

Error reporting in order to make the method proposed in this paper suitable for the image raindrop removal in the scene of UAV line patrol, this paper simulates the real scene of transmission line as much as possible when making the data set. UAV carries two cameras with two identical glasses when making the data set, one to spray water and the other to keep clean. Spray water on the glass plate to simulate raindrops on the camera in rainy days. The thickness of the glass plate is 3 mm. Set the distance between the glass and the camera to 2 to 5 cm to produce different raindrop images and minimize the reflection effect of the glass. During the shooting process, keep the relative position of the camera and the glass lens unchanged, and ensure that the background images taken by the two cameras are the same. Also, ensure that the atmospheric conditions (such as sunlight and cloud) and the background objects should be static during the image acquisition process. Finally, 2000 pairs of images including transmission line equipment scenes were taken.

3.2. Raindrop Removal Online Training Details. The 2000 pairs of pictures in the data set are allocated according to 8 : 2, among which 1600 pairs are used as model training sets and 400 pairs are used as model test sets. The super parameters of the model are set, in which the initial learning rate is set to 0.001, the batch size is set to 16, and the number of iterations is set to 40000. Using Adam optimization algorithm, it is found that the rate of gradient descent is relatively low in the process of training. Therefore, it is changed to momentum optimization algorithm, and it is found that the convergence speed of the model is significantly faster. After 40000 times of iterative training, the model is verified by test set, and it is found that the raindrop model based on the network of resistance generation has good portability.

4. Experiment Results

4.1. Comparison of Effect Pictures of Raindrop Removal. Randomly select a picture from the image data set containing raindrops for raindrop removal, and the results are shown in Figures 2 and 3.

The background image in Figures 2 and 3 is the tower and insulator string; Figures 2(a)–2(f) and Figures 3(a)–3(f) are the original image, the raindrop removal image, and the attention map generated by four recurrent networks, respectively. The original image contains dense raindrops. The raindrop removal method proposed in this paper can remove most of the raindrops in the image and repair the background image of the raindrop covered part. It can be seen from the attention map that the location and size of

raindrops in the original image can be clearly determined. From the comparison of Figures 2(a) and 2(b) and Figures 3(a) and 3(b), it can be seen that the contrast, brightness, and target edge information of the raindrop removal image and the original image are basically the same.

4.2. Comparison of Raindrop Removal Image Indexes. Randomly select a picture from the data set containing raindrops, use Yang raindrop removal method [25–27] and the method proposed in this paper to remove raindrops, and calculate the PSNR value and SSIM value of the two methods to obtain the image; the results are shown in Table 1.

It can be seen from Table 1 that the PSNR and SSIM of the image obtained by the method proposed in this paper are higher than those of Yang and Fu, which indicates that the similarity between the raindrop image obtained by the method proposed in this paper and the original clear background image is higher, which proves that the effect of the raindrop method based on the generated antinetwork is better than that of Yang and Fu.

4.3. Target Detection Result Comparison. Randomly select 50 inspection images of transmission line with raindrops including tower fault, small size hardware fault, ground wire fault, and insulator fault from the test set. Yang's raindrop removal method and the raindrop removal method proposed in this paper are, respectively, used for image raindrop removal. The Faster Rcnn target detection algorithm is used to detect the device defect target of raindrop image, Yang raindrop image, and raindrop image of the method proposed in this paper. Then, calculate the AP value of four kinds of faults and the mAP value of each group of images, respectively. The results are shown in Table 2.

From the AP value and the mAP value in Table 2, it can be seen that the target detection accuracy of the image after raindrop removal is higher than that without image enhancement. At the same time, the proposed method is better than the previous methods in the aspects of raindrop removal and image restoration.

5. Conclusion

The discrimination network uses a combination of global and local discriminators to distinguish the generated raindrop images. Using the test set in this paper to test the model, the experiment shows that the method proposed in this paper can completely remove the raindrop in the image and repair the background image, and the raindrop image is closer to the real image. Using the method in this paper to process the image raindrop can restore the image details and improve the accuracy of the target detection algorithm.

Data Availability

The data used to support the findings of this study are included within the article. The project was supported by Science Support Project of Guizhou Province ([2020]2Y039).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Science Support Project of Guizhou Province ([2020]2Y039).

References

[1] J. Xu, W. Zhao, P. Liu, and X. Tang, "Removing rain and snow in a single image using guided filter," in *IEEE International Conference on Computer Science & Automation Engineering*, pp. 304–307, Zhangjiajie, China, 2012.

[2] J. Xu, W. Zhao, P. Liu, and X. Tang, "An improved guidance image based method to remove rain and snow in a single image," *Computer and Information Science*, vol. 5, no. 3, pp. 1–11, 2012.

[3] X. Zheng, Y. Liao, W. Guo, X. Fu, and X. Ding, "Single-image-based rain and snow removal using multi-guided filter," in *International Conference on Neural Information Processing*, pp. 258–265, Berlin, Heidelberg, 2013.

[4] X. Ding, L. Chen, X. Zheng, Y. Huang, and D. Zeng, "Single image rain and snow removal via guided L0 smoothing filter," *Multimedia Tools & Applications*, vol. 75, no. 5, pp. 2697–2712, 2016.

[5] J. H. Kim, C. Lee, J. Y. Sim, and C. S. Kim, "Single-image deraining using an adaptive nonlocal means filter," in *IEEE International Conference on Image Processing*, pp. 914–917, Melbourne, VIC, Australia, 2014.

[6] Y. H. Fu, L. W. Kang, C. W. Lin, and C. T. Hsu, "Single-frame-based rain removal via image decomposition," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 782no. 1, pp. 1453–1456, Prague, Czech Republic, 2011.

[7] L. W. Kang, C. W. Lin, and Y. H. Fu, "Automatic single-image-based rain streaks removal via image decomposition," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1742–1755, 2012.

[8] Y. Luo, Y. Xu, and H. Ji, "Removing rain from a single image via discriminative sparse coding," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3397–3405, Santiago, Chile, 2015.

[9] D. Eigen, D. Krishnan, and R. Fergus, "Restoring an image taken through a window covered with dirt or rain," in *IEEE International Conference on Computer Vision*, pp. 633–640, Columbus, Ohio USA, 2014.

[10] Z. Gao, Y. Li, and S. Wan, "Exploring deep learning for view-based 3D model retrieval," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 16, no. 1, pp. 1–21, 2020.

[11] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, "Clearing the skies: a deep network architecture for single-image rain removal," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2944–2956, 2017.

[12] S. Ding, S. Qu, Y. Xi, and S. Wan, "A long video caption generation algorithm for big video data retrieval," *Future Generation Computer Systems*, vol. 93, pp. 583–595, 2019.

[13] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network,"

in *IEEE Conference on Computer Vision & Pattern Recognition*, pp. 1715–1723, Honolulu, HI, USA, 2017.

[14] Y. Zhao, H. Li, S. Wan et al., "Knowledge-aided convolutional neural network for small organ segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1363–1373, 2019.

[15] S. Ding, S. Qu, Y. Xi, and S. Wan, "Stimulus-driven and concept-driven analysis for image caption generation," *Neurocomputing*, vol. 398, pp. 520–530, 2020.

[16] B. Zhao, X. Wu, J. Feng, Q. Peng, and S. Yan, "Diversified visual attention networks for fine-grained object classification," *IEEE Transactions on Multimedia*, vol. 19, no. 6, pp. 1245–1256, 2017.

[17] Y. Xi, Y. Zhang, S. Ding, and S. Wan, "Visual question answering model based on visual relationship detection," *Signal Processing: Image Communication*, vol. 80, p. 115648, 2020.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, Las Vegas, NV, USA, 2016.

[19] L. Wang, H. Zhen, X. Fang, S. Wan, W. Ding, and Y. Guo, "A unified two-parallel-branch deep neural network for joint gland contour and segmentation learning," *Future Generation Computer Systems*, vol. 100, pp. 316–324, 2019.

[20] Y. Tanaka, A. Yamashita, T. Kaneko, and K. T. Miura, "Removal of adherent waterdrops from images acquired with a stereo camera system," *IEICE Transactions on Information and Systems*, vol. 89, no. 7, pp. 2021–2027, 2006.

[21] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*, pp. 694–711, Cham, 2016.

[22] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–14, 2017.

[23] Z. Gao, H. Xue, and S. Wan, "Multiple discrimination and pairwise CNN for view-based 3D object retrieval," *Neural Networks*, vol. 125, pp. 290–302, 2020.

[24] Y. Li, S. Liu, J. Yang, and M. H. Yang, "Generative face completion," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3911–3919, Honolulu, HI, USA, 2017.

[25] M. Khosravi and S. Samadi, "Reliable data aggregation in internet of ViSAR vehicles using chained dual-phase adaptive interpolation and data embedding," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2603–2610, 2020.

[26] 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 *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1685–1694, Honolulu, HI, 2017.

[27] L. Li, T. T. Goh, and D. Jin, "How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis," in *Neural Computing and Applications*, vol. 32, pp. 4387–4415, Springer, London, 2020.