Generation of digital artistic images based on GANs

Fang Wang

(Henan Vocational College of Agriculture, Zhengzhou, Henan, 451450, China)

wangfang212165@163.com

Abstract. This research proposes an end-to-end defogging algorithm based on Conditional Generative adversarial network (CGAN), aiming at the problem that the image quality of visible light observation equipment is significantly reduced in the case of fog (haze). This algorithm innovatively constructs a new generator that replaces traditional maximum pooling with soft pooling in the convolutional layer of the encoder to improve the ability to extract fine-grained features. At the same time, a global average pooling layer is introduced between the encoder and decoder to eliminate the oscillation effect of image edges, thereby improving the clarity of the defogged image. In addition, the structure of the discriminator based on PatchGAN is simplified, and the design of the Loss function and the selection of the weight value are optimized, thus improving the training speed and quality of the network model. The experimental results show that the algorithm exhibits stable performance in both synthetic and real foggy images. The image after defogging has an improvement in subjective perception of clarity, edge sharpness, color naturalness, and detail restoration. In terms of quantitative indicators, compared to traditional algorithms, this algorithm has improved to varying degrees in terms of structural similarity, peak signal-to-noise ratio, and image information entropy. In summary, the CGAN based end-to-end defogging algorithm proposed in this study has significant effects on improving image quality in foggy (hazy) conditions. By optimizing the network structure, Loss function and weight selection, the algorithm shows good performance in quantitative indicators and target recognition tasks.

Keywords: defogging algorithm; Generating a countermeasure network; Image generation; loss function

1 Introduction

Computer vision, also known as machine vision, refers to the use of computers and related equipment to simulate human vision, identify, detect and track targets, so that machines can observe and understand the world like humans. With the development of related research in the field of computer vision, the task of image conversion has gradually attracted extensive attention in academic circles. Image conversion refers to the conversion of an image from one representation scene to another while keeping the content of the image unchanged [1]. Many problems in computer vision, computer graphics and image processing can actually be understood as image conversion prob-

R. Appleby et al. (eds.), *Proceedings of the 2nd International Conference on Intelligent Design and Innovative Technology (ICIDIT 2023)*, Atlantis Highlights in Intelligent Systems 10,
https://doi.org/10.2991/978-94-6463-266-8_47

lems, for example, image coloring problems. It can be regarded as converting the gray scale image into its corresponding color image; The problem of image restoration can be regarded as the completion of the missing image to its corresponding complete image; The problem of image high resolution can be regarded as transforming a low-resolution image into its corresponding high-resolution image; The problem of image style transfer can be regarded as transforming an image into another style image while retaining the basic characteristics of the original image, and so on. There are still many such works. These tasks are widespread in our production and life. Traditional algorithms mostly aim at specific problems to design models for specific application scenarios [2]. Later, with the development of Convolutional Neural Networks (CNNs), more and more image conversion problems have been solved. However, due to the different modeling methods and optimal design of different tasks, the task burden of image conversion is still very huge. In recent years, GANI2I has become a hot research content in academic circles, and has made remarkable achievements in the field of computer vision. The image conversion problem has also become one of the most important application fields of GAN [3]. Because of its better generating ability and flexible design, GAN and its derivative models have been widely used in image conversion problems, and great achievements have been made in these problems. Many experimental studies show that more realistic images can be generated by incorporating GAN. Under the condition of fog (haze), affected by tiny particles in the air, the overall color of the image collected by the imaging system tends to be gray, and the image content is blurred. In essence, this image degradation can be explained as follows: the contrast reduction and color distortion of the image make it tend to be gray-white, and with the depth of field getting farther, the contrast attenuation intensifies, and the scattering effect of tiny particles becomes more obvious, which will aggravate the image blur.

Digital art image generation based on Generative adversarial network (GANs) is an exciting technology. It realizes the computer generation of artistic images through machine learning and deep learning algorithms. Generative adversarial network (GANs) are antagonistic models composed of generators and discriminators. The generator is responsible for generating false images, while the discriminator aims to distinguish between the generated image and the real image. Through iterative confrontation training, the generator gradually improves the quality of the generated image to deceive the discriminator as much as possible[4-5]. This adversarial process prompts the generator to learn the features and styles of real images, thereby generating realistic artistic images. In the generation of digital art images, GANs technology can imitate the style and creative characteristics of famous artists, generating similar art images. Through the training generator network, it can learn the texture, color, line, composition and other features in the art works, and generate new art images based on them. The generated art image can have a variety of styles and forms, such as oil painting style, Impressionism style, watercolor style, etc. By learning a large-scale dataset of art works, the generator can infer the unique style of the artist and present it in the generated images. The digital art image generation technology based on GANs has potential application value in many fields. It can provide inspiration and auxiliary tools for artistic creation, helping artists achieve creativity. In addition, this technology can also be applied to fields such as virtual reality, game development, and film production, providing rich possibilities for the generation and design of visual effects.

2 Methods

2.1 Defogging Algorithm Based on Conditional Generation Confrontation Network

Aiming at the problems of incomplete defogging, image color distortion and detail loss after defogging in traditional defogging algorithms, the author proposes a defogging algorithm based on CGAN (CGAN-DN), in which the condition is a foggy image, and the algorithm process is as follows: firstly, input the foggy image into a generator for defogging, and output the defogged image; The foggy image and the defogged image are sent to a discriminator, and the discriminator outputs a discriminating result, and the result is between 0 and 1, indicating the discriminating probability of the discriminator; Then the foggy image and the real fog-free image are sent to the discriminator, and the discrimination result is output again; Calculate the probability difference value of the second discrimination result, and feed back the discrimination error to the generator and discriminator respectively for iterative learning. When the probability difference between the two discriminant results is 0 or the conditional optimal solution, it shows that the discriminator can no longer distinguish the fog-removed image from the real fog-free image, indicating that the network successfully converges, and the CGAN-DN network is obtained after the algorithm training. The algorithm process is shown in Figure 1 [6].

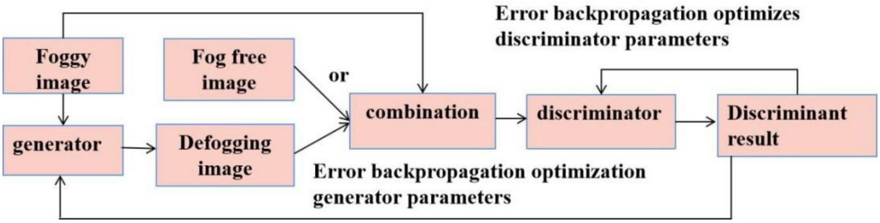

Fig. 1. Process of defogging algorithm based on CGAN

(1) Generator based on global average pooling

The function of the generator is to process the input foggy image into a defogged image. The generator of the algorithm proposed by the author adopts symmetric convolutional neural network structure, the first half is an encoder and the second half is a decoder, which consists of 12 convolution layers, 6 pooling layers and 1 global average pooling layer. See Table 1 for specific network parameters.

Table 1. Generator network parameters

number of plies	Convolution kernel size/pixel	step length	Edge filling	Activation function
Convolution 1	3×3×1×64	1	1	LReLU
Convolution 2	3×3×64×64	1	1	LReLU
Convolution 3	3×3×64×1	1	1	Tanh
Deconvolution 1	3×3×1×64	2	1	ReLU
Deconvolution 2	3×3×64×64	2	1	ReLU
Deconvolution 3	3×3×64×1	2	1	Tanh
Soft pool layer	2×2	2	0	-
Global pooling layer	64×64	1	0	-

(2) the discriminator structure based on PatchGAN.

The structure of the discriminator consists of multi-layer convolution with different scales. Its input is an image, and its output is a scalar value, which represents the probability that the input image is a defogged image. The difference between the two discriminant results of the discriminator in the same round of learning will be fed back to the generator and discriminator for iterative optimization of the network. Traditional discriminators usually use a large number of fully connected layers to discriminate pixel by pixel, and only one evaluation value (true or false) is output as a result, so there are shortcomings such as low operation efficiency and extreme evaluation results [7]. PatchGAN uses block discrimination, uses sliding window mode, outputs discriminant matrix after block comparison, and finally takes the average value of discriminant matrix as discriminant probability output. Through a large number of experiments, it is proved that when the sliding window, that is, the size of the receptive field is 70×70, the effect is the best. The author simplifies the structure of the original PatchGAN and reduces the number of layers in order to further improve the operation efficiency. The discriminator consists of four convolution layers, and then the discriminant probability is given by S-shaped curve activation function (sigmoid). See Table 2 for specific network parameters.

Table 2. Discriminator network parameters

layer	Convolution kernel size/pixel	step length	Edge filling	Activation function
Convolution 1	3×3×1×64	2	1	LReLU
Convolution 2	3×3×1×128	2	1	LReLU
Convolution 3	3×3×1×256	2	1	LReLU
Convolution 4	3×3×1×512	2	1	LReLU

2.2 loss function design

Designing and optimizing the loss function plays an important role in accelerating the convergence speed of the network model and improving the defogging ability of the algorithm [8]. The loss functions in the algorithm include: L1 norm loss function (L1-Loss), L2 norm loss function (L2-Loss), Lp perceptualloss function, and LAdv generator counteracts loss function.The overall loss function of the generator is shown in formula (1), and the optimization goal is to get LG to the minimum.

$$L_G = \omega_{L_1} L_{L1} + \omega_{L_2} L_{L_2} + \omega_P L_P + \omega_{Adv} L_{Adv} \tag{1}$$

In the process of determining the corresponding weight of each loss function in formula (1), the separation strategy is adopted, that is, the numerical value of each loss function at the time of convergence is obtained separately. The numerical value can represent the proportion of the loss function in LG. Through calculation, it is found that there is $L_{Adv} > L_{L_2} > L_P > L_{L_1}$ when each loss function converges, so when each loss function is controlled at the same order of magnitude, the corresponding weight is $\omega_{L_1} > \omega_{L_P} > \omega_{L_2} > \omega_{Adv}$. Then assign the respective weight values. The assignment process refers to the grid search method, and the parameter assignment and exhaustive calculation are carried out within a certain range to find out the optimal combination within a certain range. For example, four weight values can be added to equal 1, and assigned 0.1, 0.2, 0.3 and 0.4 respectively. Then, using the list exhaustive method, all possible permutations and combinations are listed, and a more accurate weight value is determined under the condition that LG takes the minimum value.

3 Experimental test and result analysis

3.1 Data set, parameter setting and experimental environment

In order to establish the mapping from foggy images to foggy images, the defogging algorithm requires that foggy and foggy images appear in pairs in the same scene. However, it is very difficult to establish this data set under natural conditions, so the large-scale data sets commonly used in defogging algorithms are mostly artificially synthesized. In this paper, the realistic single image de-fogging (RESIDE) data set 19 is adopted, which is based on the NYUv2 data set and contains 1399 real clear images, and 10 foggy blurred images are synthesized for each clear image to train the network model. RESIDE's test set is divided into Synthetic Objective Testing Set (SOTS) and hybrid subjective testing set (HSTS), and the images in the test set are also synthesized by algorithms and used for subjective and objective comprehensive evaluation[9].

At the same time, in order to verify the applicability of visible light image guidance in sea fog conditions, a self-made sea fog image data set containing ship targets is also made. On the one hand, using this data set can test the ability of the algorithm to filter the real fog (haze) in nature, on the other hand, combined with the target detection and recognition algorithm, we can compare the difference of the target detection and recognition ability before and after fog removal, and comprehensively test the practical application ability and potential of the algorithm. The data set was collected from the vicinity of Yantai Port and shot by telephoto lens, with a total of 388 images, and the resolution was uniformly adjusted to 512×512.

In the Conditional Generation Antagonistic Network, the input image is 512×512, the initial learning rate is 1×10-4, the batchsize of each batch is 2, and the training is carried out for 100 cycles. Adam optimizer is used to lose the function weight value $\xi_{Adv} = 0.01, \omega_{L_2} = 0.2, \omega_{L_P} = 3, \omega L_1 = 4$.

3.2 Comparison of target detection effect before and after defogging

YOLOv5 model is trained on PASCALvOC2007 data set, and it is migrated and learned on the self-made data set of ship targets and their key parts. There are 1554 images of ship targets and their key parts, which are marked with 7 kinds of targets (including aircraft carriers, destroyers, masts, cockpit, phased array radar antenna, steering engine room and civilian ships), with an average of 3.37 targets per image. At the same time, the foggy images on the sea containing ship targets are labeled according to PASCALVOC format, and each image contains 3.28 targets on average. From the results in Table 3, it can be seen that the defogged image improves the recognition target mAP by 4.13% compared with the foggy image, and the defogging algorithm can effectively improve the detection and recognition effect [10].

Table 3. Comparison of target detection effect before and after defogging

Detection object	mAP/%
Original foggy image	72.43
Deformed image	76.56

4 Conclusion

Aiming at the problem that fog (haze) will significantly reduce the imaging quality of visible light detection equipment based on image-guided weapons, thus interfering with the accurate identification of targets, a single image defogging algorithm based on conditional generation countermeasure network is proposed. Soft pooling operation is used in the down-sampling of the generator to improve the extraction ability of fine-grained features: the global average pooling layer is added to eliminate the shock effect of image edges and improve the clarity of defogged images; Simplify the structure of discriminator, optimize the method of determining the weight value of loss function, and improve the training efficiency of network model. The experimental results show that the defogged image is clear and sharp, and the color is natural. It is superior to the classical defogging algorithm in objective and quantitative indicators such as structural similarity, peak signal-to-noise ratio and image information entropy, and the average accuracy of target detection for defogged image is improved by 4.13%.

The follow-up work will continue to optimize the network model, tap the potential of the algorithm, test the performance of the algorithm under the worse conditions such as dense fog and irregular smoke, and further expand the application scope of the algorithm.

References

1. Shen, T. , Hao, K. , Gou, C. , & Wang, F. Y. . (2021). Mass image synthesis in mammogram with contextual information based on gans. Computer Methods and Programs in Biomedicine, 202(5), 106019.
2. Li, A. , Zhao, L. , Zuo, Z. , Wang, Z. , Chen, H. , & Lu, D. , et al. (2021). Diversified text-to-image generation via deep mutual information estimation. Computer vision and image understanding: CVIU(211-), 211.
3. Mao, C. , Huang, L. , Xiao, Y. , He, F. , & Liu, Y. . (2021). Target recognition of sar image based on cn-gan and cnn in complex environment. IEEE Access, PP(99), 1-1.
4. Xie, H. , Ni, J. , & Shi, Y. Q. . (2021). Dual-domain generative adversarial network for digital image operation anti-forensics. IEEE Transactions on Circuits and Systems for Video Technology, PP(99), 1-1.
5. Song, S. , Mukerji, T. , & Hou, J. . (2021). Bridging the gap between geophysics and geology with generative adversarial networks. IEEE Transactions on Geoscience and Remote Sensing, PP(99), 1-11.
6. Hu, J. , Yan, P. , Su, Y. , Wu, D. , & Zhou, H. . (2021). A method for classification of surface defect on metal workpieces based on twin attention mechanism generative adversarial network. IEEE Sensors Journal, PP(99), 1-1.
7. Zhao, C. , Shuai, R. , Ma, L. , Liu, W. , & Wu, M. . (2021). Dermoscopy image classification based on stylegan and densenet201. IEEE Access, PP(99), 1-1.
8. Yeh, W. C. , Jiang, Y. , Tan, S. Y. , & Yeh, C. Y. . (2021). A new support vector machine based on convolution product. Complexity, 2021(4), 1-19.
9. Chen, M. Y. , Wei, W. , Chao, H. C. , & Li, Y. F. . (2021). Robotic musicianship based on least squares and sequence generative adversarial networks. IEEE Sensors Journal, PP(99), 1-1.
10. Li, B. , Xian, Y. , Su, J. , Zhang, D. Q. , & Guo, W. L. . (2021). I-gans for infrared image generation. Complexity, 2021(7), 1-11.