

# Conditional GAN-based Remote Sensing Target Image Generation Method

Liu Haoyang

School of Computer Science and Engineering  
Xi'an Technological University  
Xi'an, 710021, ShaanXi, China  
E-mail: 502339341@qq.com

Hu Zhiyi

Engineering Design Institute  
Army Research Laboratory  
Beijing, 100000, China  
E-mail: 763757335@qq.com

Yu Jun

School of Computer Science and Engineering  
Xi'an Technological University  
Xi'an, 710021, ShaanXi, China  
E-mail: yujun@xatu.edu.cn

Gao Shouyi

School of Computer Science and Engineering  
Xi'an Technological University  
Xi'an, 710021, ShaanXi, China  
E-mail: 478204287@qq.com

**Abstract**—In view of the uncontrollable process of traditional GAN generating remote sensing target images, the shortcomings of generated samples are similar, and lack diversity. This paper proposes a generative confrontation network model based on background conditions. First, the computer vision attention mechanism is introduced into the generative confrontation network. Choose a learning target model so that the GAN network only learns the target model during training, and ignores other non-target information. Reduce the dependence on the number of samples in the GAN training process; secondly, use the U-net network as a generator to restore other non-target information when generating the remote sensing image of the target as much as possible; again, distinguish by different colors of the conditional mask The category of the generated target; at the same time, the L1 regularization loss is added to the loss term of the generator model, and finally, the remote sensing target image is generated. The experimental results show that the peak signal-to-noise ratio (PSNR) of the remote sensing image generation algorithm proposed in this paper reached 18.512, and the structural similarity (SSIM) reached 88.47%, which is better than the comparison test model where the generator is an ordinary autoencoder.

**Keywords**-Remote Sensing Target Image; Conditional GAN; U-net; Conditional Mask

## I. INTRODUCTION

In recent years, due to the advancement and maturity of satellite remote sensing technology, target detection based on deep neural network methods [1-2] has played a more significant role in the search and rescue of sea ships and the detection and recognition of sea ship types [3-6]. However, an accurate and generalized remote sensing target detection algorithm needs to be trained on the largest possible data set. However, in reality, the cost of acquiring a large number of remote sensing images is extremely high, so it is necessary to use data enhancement technology to expand the existing image samples.

The image in the original sample undergoes rotation, translation, cropping, etc., to generate new image training data. Although this method increases the number of training samples, the samples are highly similar, without increasing the diversity of the sample set. Although the generator model of the Variational Autoencoder (VAE) [7-9] can increase the diversity of the data set, the generated image is blurred and the effect is not good in the process of training the network, so it is not suitable for increasing the remote sensing image data Set method.

Ian J. Goodfellow et al. proposed Generative adversarial nets (GANs) [10]. The powerful fitting ability is provided by the multilayer neural network. In

theory, it is possible to model any data distribution and build a complex data model [11]. But this kind of network needs more target samples when generating remote sensing target images, and the generation process is uncontrollable. It is easy to cause problems such as high similarity of generated samples and lack of diversity in samples.

To optimize the above problems, this paper proposes a conditional generative adversarial network [15] (Conditional Generative Adversarial Nets, CGAN for short). Here, this article takes the ship target on the sea as an example. When there are only a few remote sensing image samples of the ship, the generated ship samples are controlled by the ocean background conditions. This method not only increases the number of samples of remote sensing images of sea ships, but also improves the imaging quality of ship samples, thereby effectively expanding the data set of remote sensing target images, improving the generalization capabilities of classification and detection models, and making them more Good application in the field of target detection.

## II. CONDITIONAL GENERATIVE COUNTERMEASURE NETWORK MODEL

### A. The basic principle of conditional generative countermeasure network

The traditional generative confrontation network has the characteristics of directly obtaining data distribution. "Real data" can be generated without pre-modeling the data distribution in advance, but the process of its generation is uncontrollable, and the generated samples are often similar and lack diversity. To make the generated samples efficient and controllable, experts and scholars have made improvements based on traditional GAN. As shown in Figure 1, both the generator and the discriminator add a constraint  $y$ , which can be any meaningful information. Input  $y$  as an additional input into the generator together with the prior noise  $z$ , which affects the generated data. The discriminator will also give a prediction under the influence of the constraint  $y$ , to achieve control of the generated samples. The objective optimization function definition of the conditional generative confrontation network is shown in formula 1:

$$\min_G \max_D V(G, D) = E_{x \sim P_{\text{data}(x)}} [\log D(x|y)] + E_{z \sim P_z} [\log(1 - D(G(z|y)))] \quad (1)$$

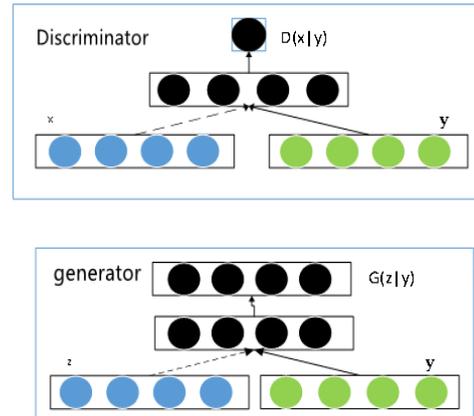


Figure 1. Conditional generative countermeasure network structure

### B. Model design of conditional generative countermeasure network

Due to the high cost of acquiring remotely sensed sea target image samples, the traditional DGAN model is not ideal when applied to sea target image generation. To reduce the dependence on the number of samples, this paper proposes a conditional generative confrontation network model for ocean background. This model introduces a visual attention mechanism. Use the mask to synthesize the conditional mask image with the ordinary sea surface image, and then control the generated data by using the conditional image as a constraint condition. For example, use the color value of the mask to control the type of ship-generated. The network only needs to learn the data distribution of the ship model during the training process, while ignoring the non-target sea background information, so that it can generate ship images that meet the conditional category with a small number of sample data sets. The network encodes the mask in the conditional image during training. A kind of mapping is formed in the hidden space, and then the corresponding target is generated by decoding. The principle of the network model in this paper is shown in Figure 2.

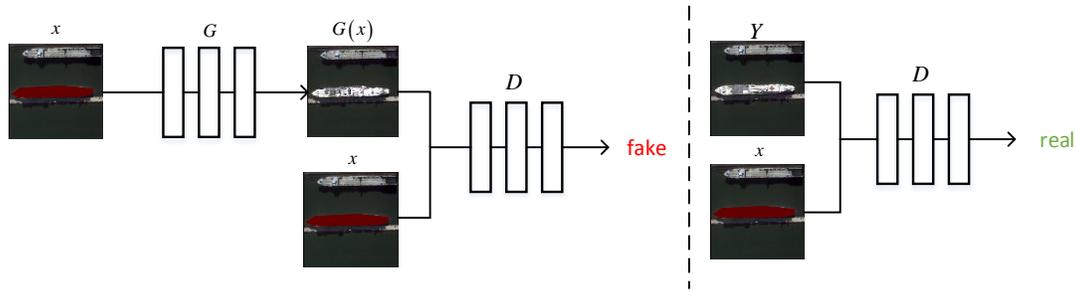


Figure 2. Working principle of marine background condition generated countermeasure network model

In order to realize the attention mechanism of task encoding→decoding, the generator of this paper chooses the network model of the U-net structure. The discriminator uses the structure of an area discriminator.

#### 1) Network structure of generator model based on U-NET network

In order to obtain valuable information from the conditional background picture, the network structure of the generator in this paper adopts the U-net network structure of cascaded encoding and decoding. The U-net network was first proposed by Olaf et al. to be applied to the full convolutional network (FCN) of medical image segmentation. The U-Net generative model is similar to the traditional self-encoding model, including encoding and decoding structures. At the same time, a skip connection method is used between the  $i$ -th layer and  $n-i$ . The feature map of the  $i$ -th layer is connected with the output of the  $n-i-1$ th layer as the input of the  $n$ -ith layer. During output, the underlying image information that needs to be preserved in the background is restored, and the part of the ship mask that needs to be generated is reconstructed. The network structure of the U-Net generation model is shown in Figure 3.

Figure 3 shows that for the  $256 \times 256$  size high-resolution remote sensing sea surface ship images in this article, the coding structure is an 8-layer down-sampling convolution operation. Each time the sample is down, the length and width of the feature map are reduced by half. The number of channels of the image sequentially becomes  $\{64, 128, 256, 512, 512, 512, 512, 512\}$ . The size of the input image is fixed at  $256 \times 256 \times 3$ . After 8 layers of downsampling, the output size is  $1 \times 1 \times 512$ . The decoding structure is an 8-layer up-sampling transposed convolution operation. Each up-sampling is connected to the output of the corresponding coding layer, the length and width of the feature map are doubled, and the number of image channels becomes  $\{512, 1024, 1024, 1024, 1024, 512,$

$256, 128, 3\}$ . Finally, a target generated image with a size of  $256 \times 256 \times 3$  is obtained.

The Unet network structure is divided into two parts: encoding and decoding. The coding network uses convolutional layers for down-sampling and extracts the semantic information in the conditional image. The decoding network uses transposed convolutional layers. Input a conditional background image into the encoding network. After extracting the image semantic information feature, it is output through the decoder. Generate qualified remote sensing images of ships. The specific parameters of the network are shown in Table 1.

#### 2) Network structure of discriminator model

The discriminator network as a whole is still a traditional deep convolutional neural network. Different from the general discriminator, the output of the network is not a single prediction value of the true or false of the image, but a prediction matrix. But the discriminator is still a two-class network. For the output prediction matrix, the error from the label is calculated by the mean value, and the error is averaged and passed to the network as a loss. The network then updates the parameters according to the loss. Specifically, each item in the prediction matrix represents an  $N \times N$  image block in the input. The discriminator finally gives the final prediction result of the input image by averaging the prediction results of all image blocks.

First, the input of the discriminator is to cascade the conditional image and the discriminant image to get an input tensor of  $256 \times 256 \times 6$ . The network extracts features through the five-layer convolutional layer to obtain  $30 \times 30 \times 1$  output. When the identification network is trained, the input for the combination of the generated picture and the conditional mask is a negative sample, and the real picture and the conditional mask image are used as a positive sample.

In order for the network to generate good images, this article has made many attempts on the data set. It is finally determined that when the output size is 30×30,

the effect of generating 256×256 remote sensing sea surface ship images is the best. The discriminator network structure is shown in Table 2 and Figure 4.

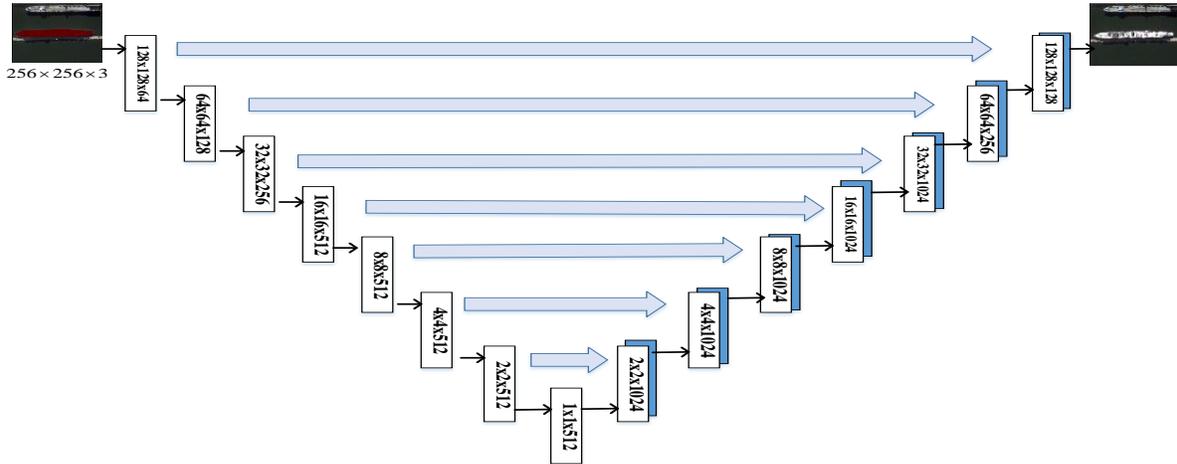


Figure 3. U-net generator network structure and coding and decoding structure diagram

TABLE I. GENERATOR MODEL NETWORK STRUCTURE PARAMETER TABLE

Inputs	Type	Kernel	Batch Normalization	Activation Function	Outputs
256x256	conv	4x4	YES	RELU	128x128
128x128	conv	4x4	YES	RELU	64x64
64x64	conv	4x4	YES	RELU	32x32
32x32	conv	4x4	YES	RELU	16x16
16x16	conv	4x4	YES	RELU	8x8
8x8	conv	4x4	YES	RELU	4x4
4x4	conv	4x4	YES	RELU	2x2
2x2	conv	4x4	YES	RELU	1x1
1x1	deconv	4x4	YES	RELU	2x2
2x2	deconv	4x4	YES	RELU	4x4
4x4	deconv	4x4	YES	RELU	8x8
8x8	deconv	4x4	YES	RELU	16x16
16x16	deconv	4x4	YES	RELU	32x32
32x32	deconv	4x4	YES	RELU	64x64
64x64	deconv	4x4	YES	RELU	128x128
128x128	deconv	4x4	YES	RELU	256x256

TABLE II. DISCRIMINATOR MODEL NETWORK STRUCTURE PARAMETER TABLE

Inputs	Type	Kernel	Batch Normalization	Activation Function	Outputs
256x256	conv	4x4	YES	LeakyReLU	128x128
128x128	conv	4x4	YES	LeakyReLU	64x64
64x64	conv	4x4	YES	LeakyReLU	32x32
32x32	conv	4x4	YES	LeakyReLU	16x16
16x16	conv	4x4	YES	LeakyReLU	8x8
8x8	conv	4x4	YES	LeakyReLU	4x4
4x4	conv	4x4	YES	LeakyReLU	2x2
2x2	conv	4x4	YES	LeakyReLU	1x1

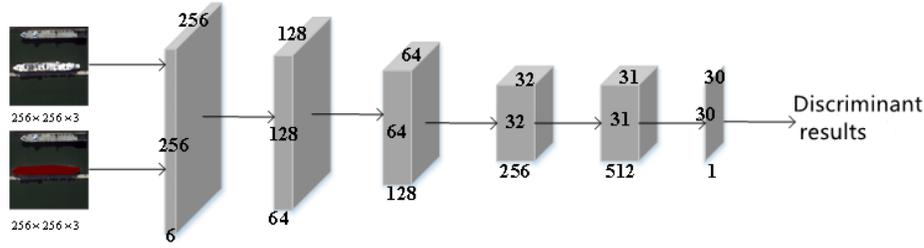


Figure 4. Network structure diagram of discriminator

### 3) Design of loss function

In this paper, the least square loss is used to improve the training instability problem in the generative confrontation network. For the generative model  $G$  and its corresponding discriminant model  $D$ , the loss function is defined as shown in formula 2.

$$L_{cGAN}(G, D) = \frac{1}{2} (E_{x \sim P_{data}(y)} [(D_Y(y) - 1)^2] + E_{x \sim P_{data}(x)} [(D_Y(G(x|y)))^2]) \quad (2)$$

In addition to the maximum and minimum optimization of the original loss of the generative confrontation network, the network also needs to generate the corresponding type of ship based on the mask in the conditional image, and also consider filling in the background image in the generated image. The sea background in the conditional image, so in order to minimize the difference between the generated image and the original background, the L1 regularization loss is added to the loss term of the generated model  $G$ . The loss function is defined as shown in formula 3.

$$L_{L1}(G) = E_{x,y,z} (\|y - G(x|y)\|_1) \quad (3)$$

The final objective loss function becomes as shown in formula 4.

$$G^* = \arg \min_G \max_D L_{cGAN}(G, D) + \lambda L_{L1}(G) \quad (4)$$

## III. VERIFICATION EXPERIMENT AND COMPARATIVE ANALYSIS

In order to verify whether the conditional generation model in this paper can generate remote sensing target images in non-training samples, and to verify the quality of the generated images. We will generate random samples and compare and evaluate the sample images generated by the model in this article and the normal model.

### A. Preparation of experimental environment

This experiment uses Linux operating system and python deep learning framework. The configuration of the specific experimental environment is shown in Table 3.

### B. Design of experimental scheme

Figure 5 shows the flow of this experimental program. The experiment is mainly divided into the following main steps:

Firstly, the data source image is preprocessed; secondly, the trained model and the comparative experimental model are separately trained; thirdly, the randomly selected mask image is used as the input of the model to generate remote sensing target images; finally, the two models are generated Comparative evaluation of images. They are described separately below.

TABLE III. EXPERIMENTAL ENVIRONMENT

<b>Operating system</b>	Ubuntu 18.04 LTS 64bit
<b>CPU</b>	Intel(R )Xeon(R) Gold 5118 CPU@2.30GHz
<b>GPU</b>	Nvidia GeForce TITAN Xp
<b>Memory</b>	32G
<b>programming language</b>	Python3.6.1
<b>compiler</b>	Pycharm2018.3
<b>Deep learning framework</b>	pytorch 0.4

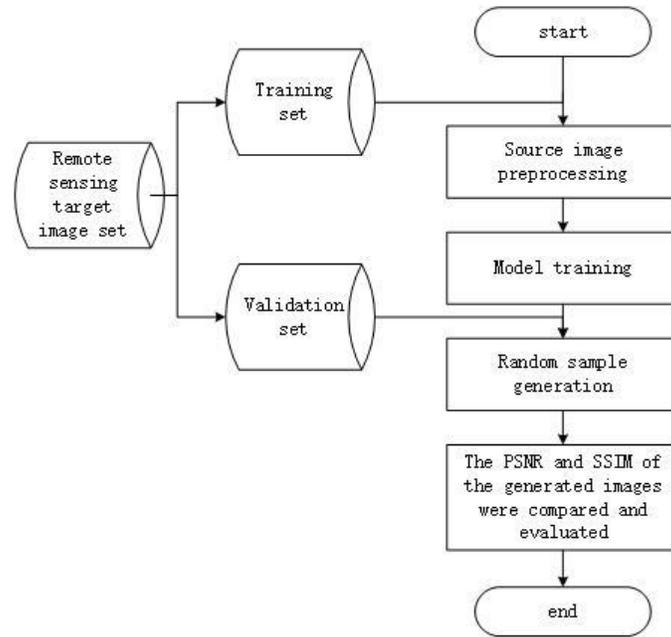


Figure 5. Experimental process

### 1) Preprocessing of data source image

In this paper, dozens of high-resolution remote sensing images of ships have been collected, and these images are relatively large. After appropriately cropping the image, resize it to 256\*256. In order to be able to generate ship images in given background conditions. We masked the ships and aircraft carriers in the sample ship image by category, that is, the RGB values of the mask image pixels of the ship are (100, 0, 0) respectively. The pixel RGB values of the mask image of the aircraft carrier are respectively (0, 100, 0), and finally the mask image is paired with the target picture. The preprocessed 420 data sets are divided into training sets (336 sheets) and validation sets (84 sheets) at the ratio of 0.8 to 0.2. The preprocessed original image and conditional mask image are shown in Figure 6. Figure 6(a) is the real remote sensing target image, including sea background and ships, and Figure 6(b) is the conditional mask image corresponding to (a), in which the part of the ship is marked with a red mask.

### 2) Model training

It can be seen from Figure 7 that the generator learns the original image and the discriminator feedback information, and uses the mask image to generate the target image. The discriminator

discriminates the image generated by the generator under the influence of the mask image until the discrimination prediction rate converges to 0.5.

In the model training process, different models use the same parameters when training on the remote sensing image data set. In order to make good use of each sample. The batch size during training is 1. Make the network fully learn the characteristics of ships in each sample; the initial learning rate is set to learning rate to 0.0002; the number of training iterations epoch is set to 1000; the Adam method is used as the training optimizer, and the momentum parameter is set to 0.5. The training time of the network proposed in this paper is about 5 hours.

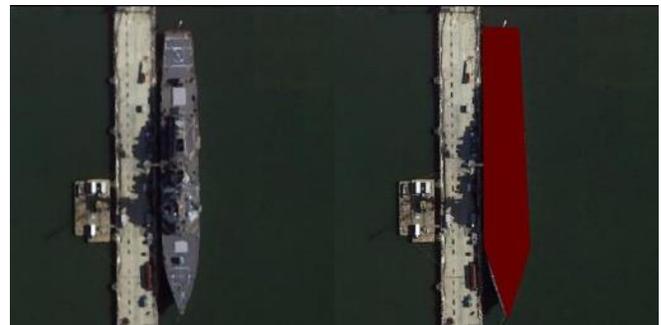


Figure 6. Preprocessing of high resolution ship remote sensing image

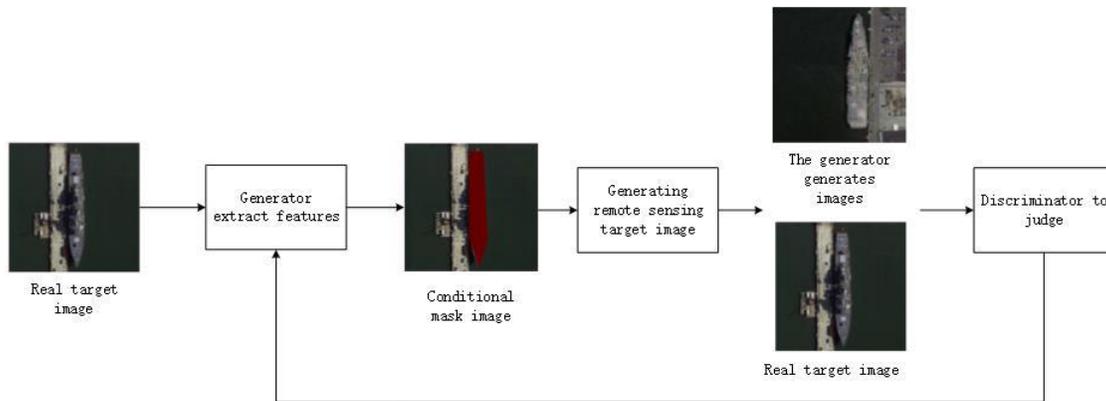


Figure 7. Model training process

### 3) Random sample generation

In order to test whether the generative confrontation network model proposed in this paper can generate remote sensing target images that are not in the training set. We will randomly select images from the images that are not involved in training and add any number of masks. Generate new remote sensing sea surface target images through the model in this paper. The masks of the ships and aircraft carriers used are derived from the set of ship masks saved when the training samples were made. An example of labeling is shown in Figure 8.

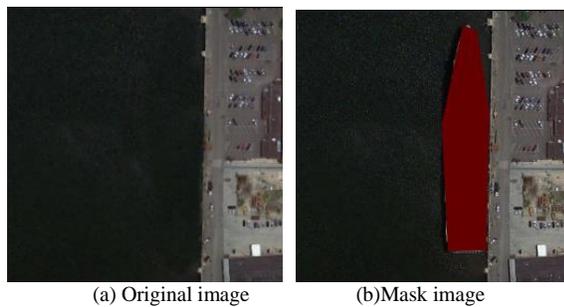


Figure 8. Sample random sample conditional mask

### 4) Comparison and evaluation of generated images

To detect the quality of the generated image, we need to compare and evaluate the sample image generated by the model in this article and the ordinary model. The evaluation is mainly to compare the peak signal-to-noise ratio and structural similarity. Randomly select images from the data validation set. After the remote sensing sea surface target image is generated, the peak signal-to-noise ratio and structural similarity value of the real image are calculated.

Compared with the experimental model, the U-net in the model is replaced with an ordinary autoencoder. The self-encoding network is divided into three parts: encoder, converter and decoder. The function of the encoder is to extract the feature and semantic information of the conditional background image. The converter maps the feature and semantic information extracted by the encoder to the hidden space and saves the feature information and semantic information. The decoder network will decode the activated features and semantic information from the converter network, generate pictures layer by layer, and get our final generated pictures.

The network structure of the self-encoding generator is shown in Figure 9.

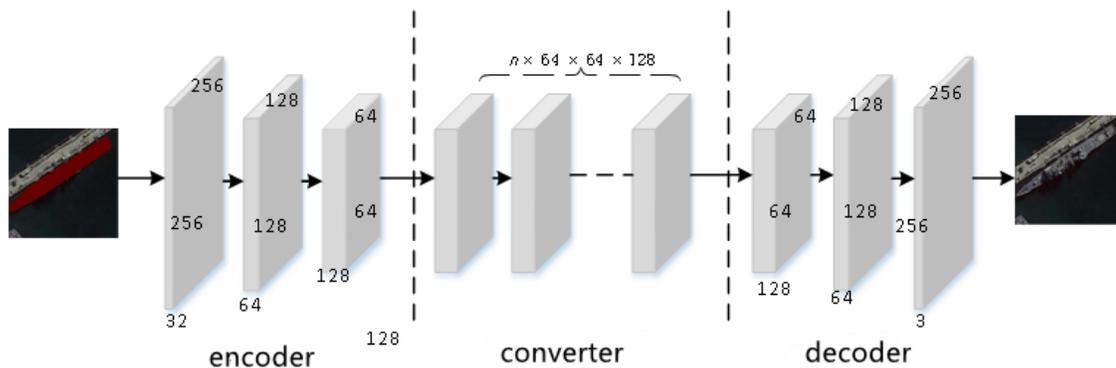


Figure 9. Generator structure of contrast experiment model

The encoder part also uses a deep convolutional network. Whenever down-sampling is performed through the convolutional layer, the scale of the network output is halved. The number of feature channels has been doubled. The decoder part also uses a deep transposed convolutional network. After transposing the convolutional layer once, the size of the image output by the network is doubled and the number of channels is halved. In classification and recognition tasks, Resnet (deep parameter transfer) and Densenet (dense) are conventional network design ideas. Both structures can deepen the depth of the network. Strengthen the transfer of features, make full use of features, and at the same time alleviate the problem of gradient disappearance due to depth during network training. Because of these advantages, the residual network structure used in the converter part is designed to contain 9 and 6 residual block converters respectively. The residual block network structure in the converter is shown in Figure 10.

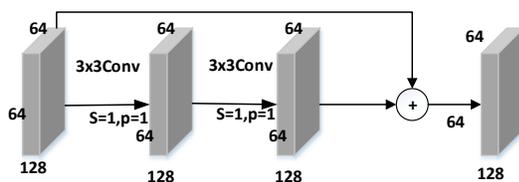


Figure 10. Residual block network structure in converter

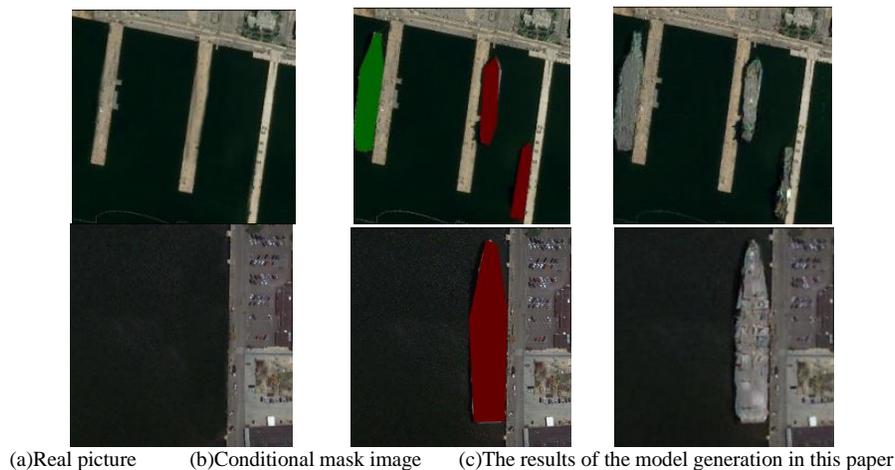


Figure 11. Random conditional sample generation results

In the comparative evaluation experiment, the peak signal-to-noise ratio and structural similarity of the images generated by this model and the comparative experimental model were calculated.

As shown in Table 4, the peak signal-to-noise ratio values of different generative model algorithms are

The specific steps for the comparison and evaluation of the generated images are as follows: (1) 30 images are randomly selected from the 84 samples in the verification set, and the corresponding masks are marked. (2) The conditional mask image is used to generate remote sensing sea surface target images through the model and comparison model in this paper. (3) Calculate the peak signal-to-noise ratio and structural similarity of the images produced by different algorithms, and calculate the average value as the final evaluation.

### C. Experimental results and analysis

Figure 11 shows a sample image of a ship of the corresponding category generated by the method in this paper through the given condition image. The generation algorithm proposed in this paper, after learning the previous samples, adds conditional masks under the new background, the model can generate new remote sensing sea ship images according to the mask type. It is proved that in the case of a few samples, more remote sensing sea surface target images can be generated by the method in this paper.

compared. It can be seen from the table that the same training set and test set are used. Under the same experimental conditions, the image generation algorithm proposed in this paper is higher than the other two algorithms in terms of image authenticity, and the noise level is low.

TABLE IV. COMPARISON OF PSNR EVALUATION RESULTS

Image generation network	PSNR
The method of this paper	18.512
A generation model with nine residual blocks	17.596
A generating model with six residual blocks	17.089

As shown in Table 5, the structural similarity between the generated image of different model algorithms and the real image is compared. It can be seen from the table that training under the same training set, the structural similarity between the generated image and the real image in this article is better than other comparative experimental models.

TABLE V. COMPARISON OF PSNR EVALUATION RESULTS

Image generation network	SSIM
The method of this paper	88.47%
A generation model with nine residual blocks	81.65%
A generating model with six residual blocks	76.31%

#### IV. CONCLUSION

This paper proposes a conditional generative countermeasure network that can generate corresponding category high-resolution remote sensing ship image samples based on conditional semantic information in the background under small samples. And carried out the relevant experimental verification, at the same time, the objective and quantitative evaluation of this algorithm based on the conditional background image generation. It briefly introduces the relevant basic theory of the evaluation method. The image generation algorithms of different generator structures are compared through experiments. Under the same conditions of the background image, the peak signal-to-noise ratio (PSNR) of the image generation algorithm proposed in this paper reached 18.512, and the structural similarity (SSIM) reached 88.47%, which is better than the general autoencoder. The comparative test model.

#### REFERENCES

- [1] Tan Kun, Wang Xue, Du Peijun. Remote sensing image classification based on deep learning and semi supervised learning [J]. Chinese Journal of image graphics, 2019, 24 (11): 1823-1841Zhang Zemiao, Huo Huan, Zhao Fengyu. Review of Object detection algorithms for deep convolutional neural networks [J]. Minicomputers, 2019, 40(09): 1825-1831.
- [2] Liu Xiaobo, Liu Peng, Cai Zhihua, Qiao Yulin, Wang Ling, Wang min. research progress of optical remote sensing image target detection based on deep learning [J / OL]. Acta automatica Sinica: 1-13 [2019-12-23] <https://doi.org/10.16383/j.aas.c190455>.
- [3] Chen Huiyuan, Liu Zeyu, Guo Weiwei, Zhang Zenghui, Yu Wenxian. Fast detection method of ship target in large scene remote sensing image based on cascaded convolutional neural network [J]. Acta radari Sinica, 2019, 8 (03): 413-424
- [4] Yin ya, Huang Hai, Zhang Zhixiang. Research on ship target detection technology based on optical remote sensing image [J]. Computer science, 2019, 46 (03): 82-87
- [5] Xu Fang. Research on key technologies for automatic detection of sea surface targets in visible remote sensing images [D]. University of Chinese Academy of Sciences (Changchun Institute of optics, precision machinery and physics, Chinese Academy of Sciences), 2018
- [6] Yang Miao. Research on ship detection algorithm in port based on optical remote sensing image [D]. Xihua University, 2018
- [7] Kingma, Diederik P, Welling, Max. Auto-Encoding Variational Bayes[J].
- [8] Kingma, Diederik P, Rezende, Danilo J, Mohamed, Shakir, and so on. Semi-Supervised Learning with Deep Generative Models[J]. Advances in Neural Information Processing Systems, 2014, 4:3581-3589.
- [9] Bodin, Erik, Malik, Iman, Ek, Carl Henrik, and so on. Nonparametric Inference for Auto-Encoding Variational Bayes[J].
- [10] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.
- [11] Gan LAN, Shen Hongfei, Wang Yao, Zhang Yuejin. Data set enhancement method based on improved dcgan [J / OL]. Computer application: 1-11 [2020-11-10] <http://kns.cnki.net/kcms/detail/51.1307.TP.20201015.1715.017.html>.
- [12] Lin Zhipeng, Zeng Libo, Wu qiongshui. Cervical cell image data enhancement based on generative antagonism network [J]. Science, technology and engineering, 2020,20 (28): 11672-11677
- [13] Yang Lanlan, Gao Mingyu, Wang Chenning, Feng Dongjie, LV Xinrui. Research on facial expression recognition based on data enhancement [J]. Computer products and circulation, 2020 (11): 128-129
- [14] Mehdi Mirza, Simon Osindero. Conditional generative adversarial nets. [DB/OL]. (2014-11-06). CoRR. abs/1411.1784
- [15] Lu Chuanwei, sun Qun, Zhao Yunpeng, sun Shijie, Ma Jingzhen, Cheng mianmianmian, Li Yuanfu. A path extraction method based on conditional generative countermeasure network [J / OL]. Journal of Wuhan University (Information Science Edition): 1-10 [2020-11-10] <https://doi.org/10.13203/j.whugis20190159>.