Research Article

Semantic Segmentation of Remote Sensing Image Based on GAN and FCN Network Model

Liang Tian ¹, Xiaorou Zhong ¹, and Ming Chen ²

¹*Institute of Computer and Information Engineering, Xinxiang University, XinXiang, Henan 453003, China*

²*College of Software Engineering, Zhengzhou University of Light Industry, ZhengZhou, Henan 450000, China*

Correspondence should be addressed to Liang Tian; tianliang@xxu.edu.cn

Received 17 June 2021; Revised 30 August 2021; Accepted 25 October 2021; Published 3 November 2021

Academic Editor: Pengwei Wang

Copyright © 2021 Liang Tian 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.

Accurate remote sensing image segmentation can guide human activities well, but current image semantic segmentation methods cannot meet the high-precision semantic recognition requirements of complex images. In order to further improve the accuracy of remote sensing image semantic segmentation, this paper proposes a new image semantic segmentation method based on Generative Adversarial Network (GAN) and Fully Convolutional Neural Network (FCN). This method constructs a deep semantic segmentation network based on FCN, which can enhance the receptive field of the model. GAN is integrated into FCN semantic segmentation network to synthesize the global image feature information and then accurately segment the complex remote sensing image. Through experiments on a variety of datasets, it can be seen that the proposed method can meet the high-efficiency requirements of complex image semantic segmentation and has good semantic segmentation capabilities.

1. Introduction

Image segmentation technology is to divide the image into different types of uniform areas according to the internal characteristics of the image [1]. It is required that the segmentation edges between regions must be accurately delineated, and the internal features of the segmented objects have consistency or similarity. Each area belongs to the same category, and different areas belong to different categories. Image understanding belongs to the high-level content of computer vision. It studies the types of objects in the image and the interrelationship between the objects to obtain the understanding of the image content and the description and interpretation of the scene, that is, the semantic features of the image. Therefore, the research content of image understanding includes not only low-level data processing and analysis, but also high-level knowledge expression reasoning.

Optical remote sensing image segmentation technology, as an important branch of image segmentation research technology, aims to classify remote sensing images at the pixel level based on actual semantic information and divide

it into a series of areas with features such as roads, farmland, villages, and industrial areas [2, 3]. The spatial resolution of remote sensing images is getting higher and higher, and it has reached the submeter level. In recent years, with the advancement of high-resolution earth observation technology, image segmentation through massive observation image data obtained by remote sensing satellites is the processing basis for urban planning, disaster monitoring, target recognition, and other applied researches [4–7]. It should be pointed out that because there are a large number of objects with different shapes in the image, it is difficult to achieve object segmentation.

The accurate identification of remote sensing images has practical guiding significance for human life [8]. Many relevant personnel have launched corresponding discussions on it. Traditional image semantic segmentation methods are mostly based on image surface features for segmentation and recognition [9]. Common methods include threshold segmentation, edge detection segmentation, and area segmentation [10–12]. The traditional method relies on the recognition of the basic features of the image, the principle is simple and does not require prior knowledge, and it has the

advantages of low computational cost and time-consuming. However, it is precisely because it only relies on the basic features as a basis, and the spatial information in the image is not used, and it is easy to lose detailed information. The segmentation result is greatly affected by the setting of initial parameters, and it is difficult to deal with the shortcomings of curve topology changes. When processing remote sensing images with complex features, the noise resistance is poor, and the segmentation effect is not ideal.

Deep learning proposes a new solution for semantic segmentation of remote sensing images. Deep network itself is a multilayer network structure [13]. The multilayer segmentation network of remote sensing images is constructed by continuously training and learning the data samples of the remote sensing images through the multilayer network. Then, based on this semantic segmentation network, the efficient processing of remote sensing images can be realized.

Deep learning uses multilayer networks to continuously train and learn and can achieve efficient segmentation of remote sensing image information. Many researchers have found that deep learning networks can extract features that are useful for image segmentation, so as to achieve accurate segmentation of remote sensing images [14]. Reference [15] proposed a DeepLab-v3+ image segmentation method based on encoder and decoder, which can extract edge features with strong robustness in remote sensing images. In [16], facing the needs of massive remote sensing image processing, a remote sensing image extraction method based on U-net network was proposed to realize the semantic segmentation of high-resolution images. Reference [17] proposed a classification method based on deep extreme learning machine to ensure efficient recognition of complex images. Reference [18] uses the residual network (ResNet) to achieve image segmentation and maintains the high resolution of the input image based on the extended spatial pyramid pool module and the deconvolution network. Reference [19] first shows multilabel segmentation of images based on adaptive clustering algorithm and uses complex-valued neural networks to complete complex image semantic recognition. However, when the above image segmentation methods are used to segment complex remote sensing images, there is often too much element information in the image, which makes the image information extraction incomplete. As a result, these methods are difficult to meet the needs of efficient semantic recognition.

However, it should be noted that the current image segmentation method is too complicated to set the model parameters due to the complicated network model structure. As a result, the semantic recognition network has insufficient image feature extraction capabilities [20]. The remote sensing image has the characteristics of too many image elements [21,22]. The above reasons will make it difficult for deep networks to achieve effective image feature extraction. This also leads to larger scene recognition errors for complex remote sensing image processing.

In order to solve the problem of insufficient accuracy of current image segmentation methods, we think that the main advantage of GAN network is that it can generate new data according to the characteristics of real data. The two

networks progress in the confrontation and continue to confront after progress. The data obtained by the generative network becomes more and more perfect, approaching the real data, so that the desired data can be generated. The FCN network has two major advantages. First, it can accept input images of any size. The second is more efficient, because it avoids the problem of repeated storage and calculation convolution due to the use of pixel blocks. So we thought of combining the advantages of the two into one [23]. Reference [24] performs semantic segmentation based on GAN network and FCN, but the target is ordinary images. Our article is aimed at the semantic segmentation of remote sensing images and uses multiple loss. Based on the GAN and FCN models, this paper proposes a new semantic segmentation method:

- (1) Aiming at the problem that complex remote sensing images are difficult to segment accurately, this paper builds a deep semantic segmentation network based on FCN, ensuring that the image receptive field is enlarged and then realizing accurate semantic segmentation of remote sensing images.
- (2) The residual structure is added to the FCN-GAN model, so that the network has a deeper layer, thereby enhancing the learning ability of the model.
- (3) The model uses multiple loss strategies. Pixel loss is combined with adversarial loss and segmentation loss to obtain the loss function of the FCN-GAN model.

The rest of this article is arranged as follows. Section 2 elaborates on deep network models, such as FCN network and GAN network. In Section 3, the FCN-GAN semantic segmentation network model is described accordingly. Section 4 is an example verification of the method proposed in this article. Section 5 is the conclusion of this article.

2. Deep Network Model

As a kind of deep network, convolutional neural network can realize the semantic segmentation of remote sensing image with the help of training and learning of multilayer network structure [25–27]. GAN network can realize the depth and accurate extraction of complex image features [28]. The fusion of FCN and GAN can further improve the accuracy of semantic segmentation of complex remote sensing images.

2.1. VGGNet. VGGNet can realize continuous iterative training and learning for sample datasets due to its deep network model structure. The distribution of its network parameters is mainly concentrated on the fully connected network structure. The essence of the VGGNet network is a convolutional neural network, and the parameters and structure of the convolutional layer are particularly important for network performance. A reasonable convolutional layer can improve the performance of input data feature extraction and transmission. The calculation formula is

$$O = \left\lceil \frac{W - K + 2P}{S} \right\rceil + 1, \quad (1)$$

where O is the output feature size, W represents the input feature size, K is the size of the convolution kernel, P is the number of pixels filled, and S stands for stride.

2.2. Fully Convolutional Network. The fully convolutional network structure includes a convolutional layer and a pooling layer but does not include a fully connected layer. The FCN network reduces the image feature size through pooling operation and reduces the feature space dimension, thereby reducing the amount of calculation, increasing the receptive field, and at the same time preventing overfitting. Then, through the upsampling operation of the deconvolution layer, the reduced-size feature map is restored to the original image size to achieve end-to-end output.

The FCN model is a semantic segmentation model for arbitrary size image input. This model includes two parts: image feature extraction and coding and label image decoding and generation. Image feature extraction and coding is a process of abstracting high-level semantic features of an image. Among them, multiple convolution and pooling operations can further delete redundant information and extract the essential information of the image [29]. Decoding and generating labeled images is a process of reconstructing semantic features. The original size of the image is restored through the upsampling operation, and the category to which each pixel belongs is obtained.

As shown in Figure 1, this paper uses the VGG16 network model as the basic network of the FCN semantic segmentation network. FCN replaces the fully connected layer of the VGG16 network with a 1×1 convolutional layer to realize image input of any size. Since the image feature extraction process has undergone 5 maximum pooling, the original image is reduced by 32 times. Therefore, it is necessary to use deconvolution for upsampling to retain the information of the original image.

As shown in Figure 2, the FCN network can realize the extraction and fusion of image features with the help of the deconvolution module. Usually, the feature map in the intermediate step of feature extraction contains rich shallow features, such as edges and textures. Combining this shallow feature with the finally extracted deep abstract feature can segment the object more accurately.

In order to make FCN network semantic segmentation and recognition more accurate, this paper uses the mean square error loss function to improve the original FCN model, as shown in formula (2), where y_i is the predicted label value and $\hat{y}_i$ is the true label value. The purpose of using this loss function is to use regression to make the sample elements of the final feature map more prominent. The general position of the image area is accurately segmented by the large difference between the foreground and the background:

$$\text{loss} = \frac{1}{n} \sum_i^n (y_i - \hat{y}_i)^2. \quad (2)$$

However, it needs to be pointed out that the FCN semantic segmentation network uses the pooling operation to

FIGURE 1: FCN basic network model diagram.

reduce the size of the feature map and then enlarges it through deconvolution, which is easy to cause partial loss of image feature information. This article uses GAN network to improve.

2.3. Generative Adversarial Network. The FCN semantic segmentation network has the problem of partial loss of image feature information due to its own pooling and reoperation process. In this paper, the FCN semantic segmentation network is improved based on the generative confrontation network: comprehensive global information, in-depth mining of image information, and more accurate and in-depth realization of semantic segmentation of remote sensing images.

The GAN network is composed of two parts: generator and discriminator. The discriminator is essentially a CNN model. Through the iterative image feature analysis of the real image and the fake image, the probability value of the interval $[0, 1]$ is finally output. The generator is an inverse CNN model, which is upsampled through a series of deconvolution layers. The actual input parameter is the actual error between the real image and the corresponding label. The network updates the network by continuously iteratively calculating errors. GAN network parameter optimization is based on the calculation of cross-entropy loss function [30]. The process can be expressed as

$$\min_G \max_D V(D, G) = E_{x \sim P_{\text{data}}(x)} \{1b[D(x)]\} + E_{z \sim P_{G(z)}} \{1b[1 - D[G(z)]]\}, \quad (3)$$

where $\min_G \max_D V(D, G)$ is the objective function, E is the mean value, D is the discriminator in the GAN network, G is the generator in the GAN network, P is the vector item, x is the actual image input, and z is the white noise input. When the discriminator input is an actual image, and $D(x)$ parameter takes the value of 1; otherwise, $D(x)$ parameter takes the value 0. Since the GAN network has two subnetwork models, a stepwise crossover training method is adopted.

For the GAN network, because the performance of the discriminator in the default network is in the best state,

$$\max_D V(D, G) = E_{x \sim P_{G(x)}} \{1b[1 - D[G(z)]]\}. \quad (4)$$

Since the GAN network has the function of simulating and generating images, the GAN network is integrated into the FCN remote sensing image semantic segmentation network model and introduces the difference discrimination branch in the discriminator of the GAN network, increases

FIGURE 2: FCN network structure.

the pixel loss, and adds more detailed information for the semantic segmentation of remote sensing images. The image is transformed into a reprocessed image that can be easily segmented by the semantic segmentation model.

3. FCN-GAN Semantic Segmentation Network Structure

In this paper, the FCN-GAN network model is used to realize the semantic recognition of remote sensing images. The main network model structures are generator network structure and discriminator network structure.

3.1. Generator Network Structure. The generator network structure in the GAN network uses $x \in R^{w \times h \times c}$ as the input image, $w = h = 512$, $c = 3$. In the traditional GAN network structure, the encoder-decoder structure mode is adopted, and the data information in the image samples in the network is inefficient.

In order to make the generator avoid this kind of situation, this paper improves the model based on U-shaped network. The network structure of the FCN-GAN generator is shown in Figure 3. For the n -layer network structure, connect the output of the i -layer network and the output of the $n - i$ -th layer as the input of the $n - i + 1$ -th layer nodes. Unlike the traditional GAN network, the generator input is not random noise, but the actual collected image. Firstly, the feature extractor performs a 3×3 convolution operation on the input image x_i ($i = 1, 2, \dots, n$) with a step size of 1. Secondly, the convolution operation of 4 modules is performed, respectively, and the size of the convolution kernel in each network layer is 3×3 . Once again, the average pooling operation is performed after the convolution module, and the window size is 2×2 . In order to prevent overfitting during the training process, dropout is used after the average pooling operation. Finally, the effective classification of image information is realized. The classifier is composed of two fully connected layers and a Softmax layer. The classification loss E_c of the classifier is defined as

$$E_c(\theta_f, \theta_c) = \sum_{i=1,2,\dots,N} L_c(G_c(G_f(x_i; \theta_f)\theta_c), y_i). \quad (5)$$

In the formula, x_i is the original input image, θ_f is the network parameter, G_f is the feature extractor, θ_c is the

parameter of the classifier, G_c is the classifier, y_i is the real label, and L_c is the classification loss calculation function.

When dealing with the problem of semantic segmentation of remote sensing images, the complexity of the model is often increased by adding a residual structure to the generation network. It can effectively solve overfitting and improve the generalization ability of the network.

Adding the residual structure after the convolutional layer of the FCN-GAN network makes the network have a deeper layer and adds nonlinear features, thereby greatly improving the learning ability of the model [31], in order to further improve the generator's ability to generate images and enhance the nonlinear fitting function of the network so that the deep convolution can be fitted to obtain the features of the shallow convolution output and the different features are merged while preventing the network from appearing the problem of gradient disappearance. We expand the 16 residual blocks of the FCN-GAN generation network to 32.

Because both the adversarial loss and the segmentation loss obtain the optimized gradient information used to generate the image from the perspective of image feature analysis. For the model, the equivalent pixel distribution can be obtained. However, due to the different values at different positions, the effect of image presentation varies greatly. Therefore, in order to make up for the visual shortcomings caused by the difference in pixel distribution, when optimizing the image generated by the generator, a pixel loss term is added, which can be expressed as

$$L_p = |G(z) - x|. \quad (6)$$

3.2. Discriminator Network Structure. As shown in Figure 4, this paper builds the discriminator network in the FCN-GAN network based on the VGG16 network model. For this network structure, we apply the Leaky-ReLU function to each network layer. In this way, the convolution with stride can replace the effect of the pooling layer.

Different from the traditional discriminator, the discriminator model consists of two branches, namely, the discriminant branch and the semantic segmentation branch. The main function of the discrimination branch is to determine the authenticity of the image. The semantic segmentation branch has two functions. One is the result of

FIGURE 3: Schematic diagram of generator network structure. (a) Generator network structure. (b) Convolution of different variants.

FIGURE 4: Schematic diagram of the discriminator network structure.

semantic segmentation to complete network detection of images. The other is to obtain the segmentation loss and adjust the distribution of the image generated by the generator together with the adversarial loss obtained by the discriminator.

In order to guide the image generated by the network to more perfectly fit the real image, the difference discrimination branch is introduced in the FCN-GAN network. The discriminator is no longer split to discriminate between the two images but to discriminate the difference between the two. The formula is as follows:

$$\text{dis}[x, G(z)] = |F(x) - E\{F[G(z)]\}|_{L_1}, \quad (7)$$

where F represents the FCN-GAN discriminative network feature extraction process. Therefore, $F(x)$ represents the features extracted from the real image after passing through the discriminant network. $F[G(z)]$ represents the features extracted from the generated image after the discriminant network. The output of the traditional discriminator is mapped to the probability value of the interval $[0, 1]$ by the

activation function $D = \sigma(F(\cdot))$, where σ is the activation function. The loss calculation formula of GAN network is

$$L_{\text{adv}} = E_{x \sim \text{data}(x)} \{\text{lb}\{D[\text{dis}[x, G(z)]]\}\} + E_{x \sim P_{G(z)}} \{\text{lb}\{1 - D[\text{dis}[G(z), x]]\}\}. \quad (8)$$

The split branch in the FCN-GAN network uses an independent split structure model. As shown in Figure 5, the first row represents the general structure of the network. It is composed of dense block structure, transition structure, convolutional layer, and deconvolutional layer. The composition of dense block, layer and transition down structure is shown in the second row of Figure 5. The arrow indicates that the output of the layer where the arrow starts and the output of the layer where the arrow ends are spliced.

First, the image is generated by the generative model, and the image and its corresponding semantic segmentation label map are sent to the segmentation branch together. The process of calculating the loss through the given semantic segmentation label map can be expressed as

FIGURE 5: Segmentation branch structure diagram.

$$L_{\text{seg}} = \sum_{ij} [y_{ij} \ln x_{ij} + (1 - y_{ij}) \ln (1 - x_{ij})], \quad (9)$$

where x_{ij} is the category of the pixels corresponding to the positions i and j in the network prediction image and x_{ij} is the category of the pixel corresponding to the position i and j of the label image.

Pixel loss is combined with adversarial loss and segmentation loss to jointly control the distribution of the image, and the loss function of the FCN-GAN model is obtained as

$$L = L_{\text{adv}} + \lambda L_{\text{seg}} + \mu L_p, \quad (10)$$

where L_{adv} is the adversarial loss, L_{seg} is the segmentation loss, λ is the segmentation parameter, and μ is the pixel parameter.

The peak signal-to-noise ratio (PSNR) is used to compare and evaluate the images that contain only L_{adv} , L_{adv} , and L_p and L_{adv} , L_{seg} , and L_p in this article. The results are shown in Figure 6. It can be seen that the image quality generated by the network model corresponding to the $L_{\text{adv}} + L_{\text{seg}} + L_p$ curve is much higher than the other two models.

4. Experimental Simulation and Analysis

The experiments in this article are all carried out under the same platform environment, and the hardware environment is NVIDIA Geforce MX450 graphics card, Intel Core i7 1165G7 processor. The software environment is Linux system and Pytorch 1.0 development environment. The experimental network frameworks are all implemented and built with the Tensorflow framework.

4.1. Network Parameter Setting. The FCN-GAN generator structure is more complicated, so a step-by-step training strategy is adopted.

The first step is to train the generator network separately, first define the input and label of the semantic segmentation network, and calculate the pixel loss L_p . In the training experiment, the SGD optimizer is used, and the momentum

FIGURE 6: The impact of different losses on the PSNR of the generated image.

is set to 0.9; the learning rate is initially set to 0.03, the minimum is 0.00015, and it is reset every 100 times. The learning rate decay method is cosine annealing, as shown in the following formula, and the number of iterations is 1000:

$$lr = lr_{\min} + \frac{1}{2} (lr_{\max} - lr_{\min}) \left(1 + \cos \left(\frac{T_{\text{cur}}}{T_{\max}} \pi \right) \right), \quad (11)$$

where $lr_{\min}$ is the minimum setting value of the learning rate, $lr_{\max}$ is the maximum setting value of the learning rate, T_{cur} is the current iteration number, $T_{\max}$ sets the maximum number of iterations for the simulation, and $lr_{\max} = 0.001$, $lr_{\min} = 0.00015$, $T_{\max} = 100$.

The second step is to train the segmentation branch of the discriminator separately. The selected 3000 remote sensing images are used as the input of the segmentation branch, and the corresponding semantic segmentation results manually labeled are used as the label of the segmentation branch. We calculate the segmentation loss L_{seg} . The higher the value of L_{seg} , the less accurate the segmentation of the input image by the segmentation branch, based on Adam gradient descent method to obtain the network model solution.

The third step is to read the training parameters of the first two steps respectively. On this basis, the discriminant branches of the generator and the discriminator are jointly trained. The artificially collected images and the generator model are used to generate images as the input of the discriminant branch of the discriminator. The label of the generated sample is 0, and the label of the real sample is 1. The implementation steps are as follows:

- (1) Real samples are used to train discriminant branch D_{adv} .
- (2) Keep the parameters of the generator model unchanged and use the generated samples to train the discriminant branch D_{adv} .
- (3) Keep the parameters of the discrimination branch D_{adv} unchanged, input the generated image into D_{adv} , and calculate the loss value of D_{adv} . Thus, the update parameters of the generation network and the gradient information for adjusting the distribution of the generated image are obtained, and the gradient information is used to complete the update.
- (4) Keep the parameters of the segmentation branch D_{seg} unchanged and input the generated image into D_{seg} . After calculating the loss value of D_{seg} , return the value to complete the update.
- (5) Repeat steps 1 to 4 and complete the update after iterative training 15,000 times.

4.2. Evaluation Index. At this stage, the performance evaluation of the graph semantic segmentation algorithm mainly includes accuracy and speed. Among them, the speed is the number of pictures that can be processed per second. The accuracy indicators mainly include Pixel Accuracy (PA), Mean Pixel Accuracy (mPA), Intersection over Union (IoU), and Mean Intersection over Union (Mean Intersection over Union, $mIoU$).

p_{ij} represents the number of pixels of category i that were misjudged as category j . p_{ii} represents the situation where the category i is correctly discriminated. p_{ji} and p_{jj} are the same.

PA is expressed as the proportion of pixels that are correctly predicted to the total pixels:

$$PA = \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}}. \quad (12)$$

mPA is a more accurate evaluation method, which calculates the average of the accuracy of all categories of pixels:

$$mPA = \frac{1}{k+1} \sum_{i=0}^k \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}}. \quad (13)$$

IoU represents the intersection of the prediction area and the real area than the union of the prediction area and the real area:

$$IoU = \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}}. \quad (14)$$

$mIoU$ is expressed as the average of all categories' intersection ratio:

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}}. \quad (15)$$

In the experiment, we use mPA and $mIoU$ as a measure of the pros and cons of the simulation experiment results.

4.3. Network Convergence Analysis. During the training process, the training dataset is randomly shuffled. The objective function of the network selects the cross-entropy loss function and introduces the L2 regularization method to further avoid the occurrence of overfitting. In addition, we also realize the balance processing of the pixel information of the dataset based on the median frequency equalization method.

Take the Vaihingen dataset as an example. Since the Vaihingen dataset is relatively small, the image reprocessing method is used to expand the sample. In this experiment, the semantic categories of the dataset are divided into 15 categories, and 15 types of targets will be segmented during the test. As shown in Figure 7, the FCN-GAN network model achieves the optimization of network performance after 110 epochs, and the $mIoU$, mPA , and Loss have reached the best state in the iterative process.

4.4. General Dataset Simulation Verification. The simulation experiment in this paper chooses Vaihingen, Potsdam, and Deep Globe Road Extraction dataset as general datasets in order to realize the experimental research of semantic segmentation of different volume datasets.

The Vaihingen dataset contains 3 channels of IRRG (Infrared, Red, Green) images, DSM (Digital Surface Model) images, and NDSM (Normalized Digital Surface Model) images. The average size of the images is 2494×2064 , and the total number of images is 33. Seventeen of them were used as a test set.

The Potsdam dataset contains IRRG images, IRGB images, DSM images, and NDSM images. The image size is 6000×6000 , and the total number of images is 38, of which 14 are used as the test set.

The size of the images in the Deep Globe Road Extraction dataset is 1024×1024 , and the total number of images is 6226. In this paper, the dataset is randomly divided into training set and validation set according to the ratio of 4:1.

4.5. Vaihingen Dataset. The simulation experiment environment was built as described above, and each image segmentation method segmented the Vaihingen dataset in the same experimental scene. Table 1 shows the semantic segmentation of Vaihingen dataset under different methods.

As shown in Table 1, the FCN-GAN semantic segmentation method proposed in this paper has mPA accuracy

FIGURE 7: FCN-GAN training process.

TABLE 1: Vaihingen dataset experimental results.

Method	Evaluation parameter (%)	
	<i>mPA</i>	<i>mIoU</i>
The proposed method	87.2	94.8
Reference [18]	85.4	92.3
Reference [16]	84.9	91.9
Reference [15]	84.1	91.3
SegNet	85.3	91.7
PSPNet	86.1	92.7

of 87.2% and *mIoU* of 94.8% for the semantic segmentation index of the Vaihingen dataset. The image segmentation results of the proposed methods are higher than other methods. It is confirmed that the proposed method has certain performance advantages when performing image segmentation on the Vaihingen dataset.

Figure 8 shows the simulation results of part of the Vaihingen dataset. It can be seen from Figure 8 that the method proposed in this paper can well identify various image elements in the Vaihingen dataset. The effect of dividing lawns and buildings is excellent.

4.6. Potsdam Dataset. Similarly, different methods are used for comparative analysis with the methods proposed in this article. Table 2 shows the experimental results of the Potsdam dataset under different methods.

Table 2 indicates that the method proposed in this paper has a better semantic segmentation effect on the Potsdam dataset than the comparison method. *mPA* of the FCN-GAN semantic segmentation method is 84.2%, and *mIoU* is 90.5%.

Figure 9 shows the simulation results of part of the Potsdam dataset. It can be seen from Figure 8 that the method proposed in this paper can well identify various image elements in the Potsdam dataset. However, [16, 18] misjudged grassland as buildings. Reference [16] does not make effective judgments on stationary vehicles. Among them, SegNet and PSPNet have better segmentation results.

Combined with the above theoretical analysis, it can be seen that there are two main reasons for the excellent semantic segmentation performance of the method proposed in this paper. One is that the FCN network model can reduce the dimension of the feature space and increase the receptive field through pooling operations. The second is that the GAN network itself has the characteristics of continuous confrontation generation, and the introduction of the residual network structure further improves the fitting ability of the network, realizes the deep and accurate extraction of image features, and guarantees the accuracy of semantic segmentation.

4.7. Deep Globe Road Extraction Dataset. The Deep Globe Road Extraction dataset is the largest dataset in this article, and its segmentation results are shown in Table 3.

It can be seen from Table 3 that the FCN-GAN image semantic segmentation method proposed in this paper also has advantages for image segmentation of large-scale datasets. For the Deep Globe Road Extraction dataset, the recognition result *mPA* is 89.5% and *mIoU* is 91.9%. At the same time, compared with [18], the image discrimination performance *mPA* and *mIoU* have achieved 3.0% and 2.4% improvement, respectively. This verifies that the proposed method is also suitable for semantic segmentation of large-scale image datasets.

In summary, compared with the current semantic segmentation methods, the method in this paper can be applied

FIGURE 8: Vaihingen dataset experimental results.

TABLE 2: Experimental results of Potsdam dataset.

Method	Evaluation parameter (%)	
	mPA	$mIoU$
The proposed method	84.2	90.5
Reference [18]	76.4	80.3
Reference [16]	80.3	82.9
Reference [15]	82.1	86.3
SegNet	85.1	88.7
PSPNet	84.0	88.3

FIGURE 9: Potsdam dataset experimental results.

TABLE 3: Experimental results of deep globe road extraction dataset.

Method	Evaluation parameter (%)	
	<i>mPA</i>	<i>mIoU</i>
The proposed method	89.5	91.9
Reference [18]	86.5	89.5
Reference [16]	84.3	85.1
Reference [15]	84.6	86.8
SegNet	87.2	87.9
PSPNet	88.7	90.3

to the semantic segmentation of image datasets of various sizes and has the advantage of precise semantic segmentation.

5. Conclusion

Aiming at the problem of low accuracy of semantic segmentation of remote sensing images, this paper combines the diversity of data generated by GAN network and the high-efficiency advantages of FCN network to propose a new method. Applying the FCN network model to semantic segmentation of remote sensing images can effectively increase the receptive field of the model. At the same time, the effective fusion of GAN network can realize the depth and accurate extraction of complex image features. Experiments show that in comparison with other commonly used methods, the proposed method can achieve the best results and can satisfy the accurate semantic segmentation of complex remote sensing images. However, the experimental part of this article uses general datasets. For actual samples, the collected images are easily affected by the weather environment. Therefore, the future research direction will be oriented to more complex actual scenes to achieve efficient image segmentation.

Data Availability

The data included in this study are available without any restriction.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

This work was supported by Key R&D and Promotion Special Project of Henan Province (no. 212102210104).

References

- [1] C. Shu and L. Sun, "Automatic target recognition method for multitemporal remote sensing image," *Open Physics*, vol. 18, no. 1, pp. 170–181, 2020.
- [2] Y. Liu, Z. Zhang, and F. Li, "Key frame extraction of multi-resolution remote sensing images under quality constraint," *Open Physics*, vol. 17, no. 1, pp. 871–878, 2019.
- [3] L. Chen and M. Bushell, "The extraction method for 3D dynamic characteristics of remote sensing images," *Recent Advances in Electrical & Electronic Engineering*, vol. 10, no. 2, pp. 179–183, 2017.
- [4] Y. P. Zhang, Y. Zhang, N. Yang, X. Luo, and Q. Luo, "A fast spectral clustering algorithm for hyperspectral remote sensing images," *Bulletin of Surveying and Mapping*, vol. 1, no. 12, pp. 60–64, 2019.
- [5] S. Liang, J. Cheng, and J. Zhang, "Maximum likelihood classification of soil remote sensing image based on deep learning," *Earth Sciences Research Journal*, vol. 24, no. 3, pp. 357–365, 2020.
- [6] T. Chen, Y. Zhao, and Y. Guo, "Sparsity-regularized feature selection for multi-class remote sensing image classification," *Neural Computing & Applications*, vol. 32, no. 11, pp. 6513–6521, 2020.
- [7] G. Wang and Y. H. Zhang, "Feature extraction method of multi-source remote sensing image after landslide disaster based on sensor network," *Disaster science*, vol. 35, no. 3, pp. 71–74, 2020.
- [8] J. Zhang, X. P. Shi, H. L. Zhang, and S. Geng, "Sparse denoising reconstruction of high noise remote sensing image," *Journal of Harbin Institute of Technology*, vol. 51, no. 10, pp. 47–54, 2019.
- [9] Z. Zhang, J. Chen, and Z. Liu, "SLIC segmentation method for full-polarised remote-sensing image," *Journal of Engineering*, vol. 2019, no. 20, pp. 6404–6407, 2019.
- [10] F. Y. Nie, J. Q. Li, P. F. Zhang, and T. Y. Tu, "An image segmentation threshold selection method based on Tsallis relative entropy," *Progress in laser and optoelectronics*, vol. 54, no. 7, pp. 137–144, 2017.
- [11] Q. W. Li, G. C. Ma, G. Y. Huo, Y. Zhou, and H. Sheng, "A new method of side scan sonar image segmentation based on NSCT domain edge detection," *Journal of Instrumentation*, vol. 34, no. 8, pp. 1795–1801, 2013.
- [12] Q. H. Zhao, H. Y. Zhang, and Y. Li, "Regional image segmentation based on Kolmogorov Smirnov statistics," *Chinese Journal of image graphics*, vol. 20, no. 5, pp. 678–686, 2015.
- [13] J. Gu, X. Sun, Y. Zhang, K. Fu, and L. Wang, "Deep residual squeeze and excitation network for remote sensing image super-resolution," *Remote Sensing*, vol. 11, no. 15, pp. 1–16, 2019.
- [14] X. R. Zhang, X. Wang, X. Tang, H. Y. Zhou, and C. Li, "Description generation for remote sensing images using attribute attention mechanism," *Remote Sensing*, vol. 11, no. 6, pp. 1–18, 2019.
- [15] L. Yuan, J. S. Yuan, and D. Z. Zhang, "Remote sensing image classification based on deep lab-v3 +," *Progress in laser and optoelectronics*, vol. 56, no. 15, pp. 236–243, 2019.
- [16] J. Li, M. Xu, and H. Xiu, "U-net network for building information extraction of remote-sensing imagery," *International Journal of Online and Biomedical Engineering (ijOE)*, vol. 14, no. 12, pp. 179–190, 2018.
- [17] F. Lu and M. Han, "Hyperspectral remote sensing image classification based on deep extreme learning machine," *Journal of Dalian University of Technology*, vol. 58, no. 2, pp. 166–173, 2018.
- [18] N. Jiang and J. Li, "An improved semantic segmentation method for remote sensing images based on neural network," *Traitement du Signal*, vol. 37, no. 2, pp. 271–278, 2020.
- [19] M. Barr, "A novel technique for segmentation of high resolution remote sensing images based on neural networks," *Neural Processing Letters*, vol. 52, no. 1, pp. 679–692, 2020.
- [20] Y. Li, Y. Zhang, X. Huang, and J. Ma, "Learning source-invariant deep hashing convolutional neural networks for cross-source remote sensing image retrieval," *IEEE Transactions on*

- Geoscience and Remote Sensing*, vol. 56, no. 11, pp. 6521–6536, 2018.
- [21] S. Zhang, Q. Yuan, J. Li, J. Sun, and X. Zhang, “Scene-adaptive remote sensing image super-resolution using a multiscale Attention network,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 7, pp. 4764–4779, 2020.
 - [22] W. Zhou, D. Ming, L. Xu, H. Bao, and M. Wang, “Stratified object-oriented image classification based on remote sensing image scene division,” *Journal of Spectroscopy*, vol. 2018, no. 1, 11 pages, Article ID 3918954, 2018.
 - [23] D. Wu, S. F. Shang, X. Cai et al., “Quantum differential evolution with cooperative coevolution framework and hybrid mutation strategy for large scale optimization,” *Knowledge-Based Systems*, vol. 224, Article ID 107080, 2021.
 - [24] L. X. Zhao, “Image semantic segmentation method based on GAN network and FCN model,” *Journal of Engineering*, pp. 1–9, 2021.
 - [25] H. Chen, F. Miao, and X. Shen, “Hyperspectral remote sensing image classification with CNN based on quantum genetic-optimized sparse representation,” *IEEE ACCESS*, vol. 8, no. 1, pp. 99900–99909, 2020.
 - [26] C. Yu, L. Wang, J. Zhao, L. Hao, and Y. Shen, “Remote sensing image classification based on RBF neural network based on fuzzy C-means clustering algorithm,” *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 4, pp. 3567–3574, 2020.
 - [27] Y. Wang, “Remote sensing image semantic segmentation algorithm based on improved ENet network,” *Scientific Programming*, vol. 2021, Article ID 5078731, 10 pages, 2021.
 - [28] H. Li, “Image semantic segmentation method based on GAN network and ENet model,” *Journal of Engineering*, vol. 2021, no. 10, pp. 594–604, 2021.
 - [29] V. K. Shrivastava and S. Bera, “Effect of pooling strategy on convolutional neural network for classification of hyperspectral remote sensing images,” *IET Image Processing*, vol. 14, no. 3, pp. 480–486, 2019.
 - [30] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., “Generative adversarial networks,” *Advances in Neural Information Processing Systems*, vol. 3, pp. 2672–2680, 2014.
 - [31] L. Xu and Q. Chen, “Remote-sensing image usability assessment based on ResNet by combining edge and texture maps,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 6, pp. 1825–1834, 2019.