Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 5983045, 10 pages
https://doi.org/10.1155/2022/5983045

Research Article

Image Semantic Segmentation Method Based on Deep Learning in UAV Aerial Remote Sensing Image

Min Ling,[1] **Qun Cheng**,[1] **Jun Peng**,[2] **Chenyi Zhao**,[3] **and Ling Jiang**[2]

[1]*Shanghai Urban Construction Engineering School (Shanghai Gardening School), Shanghai 200232, China*
[2]*School of Geographic Information and Tourism, Chuzhou University, Chuzhou 239000, China*
[3]*Hongya Education and Technology (Shanghai) Co., Ltd., Shanghai 200241, China*

Correspondence should be addressed to Jun Peng; ipengjun2020@163.com

Received 16 March 2022; Revised 8 April 2022; Accepted 12 April 2022; Published 26 April 2022

Academic Editor: Ramin Ranjbarzadeh

The existing semantic segmentation methods have some shortcomings in feature extraction of remote sensing images. Therefore, an image semantic segmentation method based on deep learning in UAV aerial remote sensing images is proposed. First, original remote sensing images obtained by S185 multirotor UAV are divided into smaller image blocks through sliding window and normalized to provide high-quality image set for subsequent operations. Then, the symmetric encoding-decoding network structure is improved. Bottleneck layer with 1×1 convolution is introduced to build ISegNet network model, and pooling index and convolution are used to fuse semantic information and image features. The improved encoding-decoding network gradually strengthens the extraction of details and reduces the number of parameters. Finally, based on ISegNet network, five-classification problem is transformed into five binary classification problems for network training, so as to obtain high-precision image semantic segmentation results. The experimental analysis of the proposed method based on TensorFlow framework shows that the accuracy value reaches 0.901, and the F1 value is not less than 0.83. The overall segmentation effect is better than those of other comparison methods.

1. Introduction

With the rapid development of remote sensing technology in recent years, especially the rise of high-resolution remote sensing images, remote sensing technology has become a necessary method for timely regional Earth observation [1]. The emergence of various high-resolution remote sensing satellites has led to the increase of remote sensing image collection sources and the expansion of the scale of multimodal datasets [2]. In the face of massive, multimodal remote sensing image data, the traditional image processing and analysis methods do not perform well in the big data environment [3, 4]. Deep learning has the ability to extract main features from massive data and can achieve real-time data processing. Therefore, remote sensing image semantic segmentation based on deep learning has gradually become the focus of research [5].

Semantic segmentation of remote sensing image is a transition link and key step to realize object-oriented extraction from data to information [6]. As a bridge to advanced tasks, semantic segmentation is widely used in the field of computer vision and remote sensing, such as automatic driving, attitude estimation, and remote sensing image interpretation. Previous image segmentation was usually based on the image pixel itself and the representation of image low-order visual information, such as pixel clustering segmentation and graph segmentation [7–9]. Although these methods are with low computational complexity because of lacking model training based on dataset, due to the lack of sufficient manual annotation

information, the segmentation results are not ideal in image segmentation for complex tasks [10]. The rise of deep learning makes image segmentation enter a new stage, such as image block segmentation using convolutional neural network [11].

The existing semantic segmentation methods have the problem of insufficient extraction depth in remote sensing image feature extraction, and the massive small targets in the image cause new difficulties in using deep learning to obtain robust feature representation [12]. Therefore, in order to effectively solve the above problems, an image semantic segmentation method based on deep learning is proposed to realize semantic segmentation and target recognition in UAV aerial remote sensing images.

2. Related Work

The basic process of semantic segmentation of remote sensing image is to preprocess the data of remote sensing image and then extract tensor features of image according to the preprocessed data. The input of the model is the tensor features. The model is initialized with a given training network model and the segmentation and prediction of remote sensing image targets are completed [13].

Early semantic segmentation is in the stage of traditional image segmentation, which requires participants to annotate the features manually, but the accuracy of annotation greatly affects the segmentation results [14]. The human and material resources consumed cannot meet the high-efficiency requirements of large-scale applications. With the emergence of full convolution neural network, semantic segmentation has officially entered the era of deep learning. The goal of image semantic segmentation is to mark each pixel of the image with the corresponding class. In order to understand the research status, existing problems, and development prospects of image semantic segmentation, [15] introduced the mainstream image semantic segmentation methods on the basis of extensive investigation. The segmentation results of common image semantic segmentation algorithms were summarized and compared. Based on the summary of common image semantic segmentation datasets and evaluation standards, the development trend of image semantic segmentation in the future was prospected. Reference [16] proposed Generative Adversarial Networks for image semantic segmentation, including edge adversarial network and semantic segmentation adversarial network, which effectively improved the segmentation accuracy in fog, but the segmentation efficiency needs to be improved. Reference [17] studied image segmentation in lane departure system and compared traditional processing methods with deep learning semantic segmentation methods, and the results showed that deep neural network image semantic segmentation had better robustness and efficiency, but it is still slightly insufficient in the processing performance of high-order task set. Reference [18] designed a fusion network for small target image segmentation, that is, extracting the feature information of RGB image and depth image to complement each other, but it has high requirements for image acquisition.

At the same time, in order to enhance the segmentation performance of the deep learning network, some improvements are made. For example, [19] proposed a superpixel enhanced depth neural model to solve the problems of distinguishing surface image features and classifying ground objects in the process of image segmentation. It adopted an end-to-end method combined with the depth convolution neural network to achieve better classification accuracy. However, for a large number of small targets in remote sensing images, the effect of segmentation and recognition needs to be improved. In order to solve the challenge of learning spatial context of deep convolution neural network in high-resolution image semantic segmentation, [20] proposed a new segmentation model, which deduced a symbolic distance map for each semantic class from the real label map to improve the segmentation accuracy. However, the training process is complex, and the processing efficiency of a large number of remote sensing images needs to be improved. Reference [21] proposed a new pixel-level feature extraction model with convolutional encoder and decoder for medical image recognition, which improved the global and average accuracy through dataset training, and the improvement of feature extraction enhanced the accuracy of semantic segmentation, but it cannot consider the processing efficiency. Reference [22] proposed a depth convolution neural network based on U-Net to realize end-to-end semantic segmentation. The model fusion strategy effectively improved the segmentation accuracy, but the segmentation effect is not good for partial occlusion or fog. Aiming at the problem that traditional segmentation methods find it difficult to deal with high-resolution remote sensing images containing complex ground objects, an image semantic segmentation method based on deep learning in unmanned aerial vehicle (UAV) aerial remote sensing images is proposed. Its contributions are summarized as follows:

(1) Considering that the traditional SegNet method has low segmentation accuracy and long training time, the proposed method expands the convolution layer and adds the Bottleneck layer to obtain the Improved SegNet (ISegNet) network, so that it can express more complex features.

(2) In order to avoid the problem that classification results of a single classifier are greatly affected by misclassification, the proposed method converts five-classification problem into five binary classification problems for network training based on ISegNet network, so as to improve the segmentation accuracy of the model.

3. Proposed Research Method

3.1. Classification Model Establishment. The main flow of the proposed semantic segmentation method is shown in Figure 1. The whole process is more concise and effective than the traditional image semantic segmentation based on regional features. There is no need to search the region containing the target image in the early stage, and there is no need to merge similar regions in the later stage.

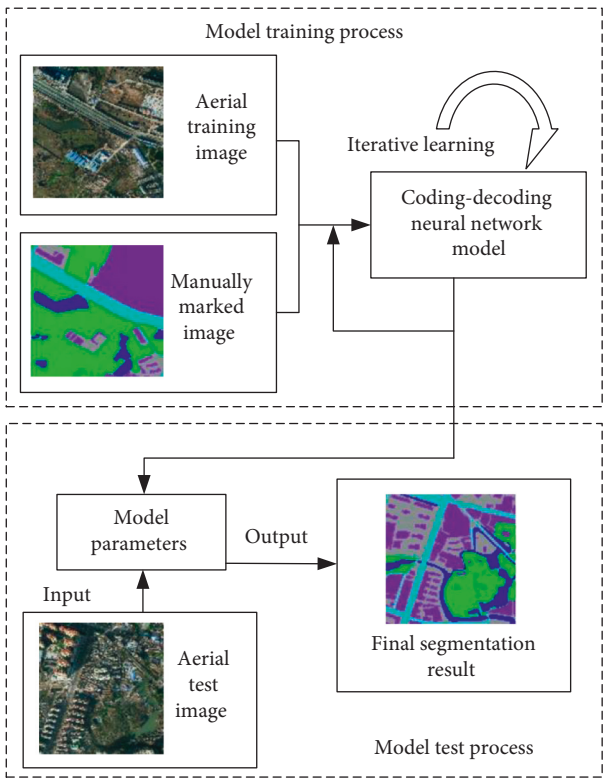

Figure 1: Semantic segmentation process based on deep learning.

The aerial image and the corresponding manually marked image are input into the encoding-decoding network, the optimal model parameters are obtained through multiple iterative learning, and the model and corresponding parameters are saved. In the inference stage, the final segmentation result can be obtained by directly inputting the aerial test image into the saved model. However, directly convoluting the image will make the image smaller, and the contribution of the information at the edge of the original image to the image content is small. Therefore, padding is used to solve this problem; that is, 0 is filled in the image edge to expand the image size in the process of convolution, so that the image size remains unchanged after multiple convolution operations [23].

3.2. UAV Hyperspectral Image Acquisition. In order to fully understand the differences of the research objects in hyperspectral images, a spectral information database was established, and the hyperspectral images of UAVs in two small areas in the field of Shanghai suburb were obtained. In one area, discontinuous UAV images of four sorties were obtained to fully cover the object types in the study area. In the other area, relatively continuous UAV images of five sorties were obtained to provide a data source for the classification and identification of typical object types. The S185 UAV system was used in the test, which mainly included Cubert S185 hyperspectral data acquisition system, six-rotor electric UAV system (maximum load is about 6 kg; flying time is 15 min–30 min), triaxial stabilized camera, and data processing system, as shown in Figure 2.

The acquisition of UAV hyperspectral data should be carried out in sunny days to avoid cloud shadow on the image and affecting the image quality, and the solar deflection angle should not be too large during acquisition to avoid too large shadow area in the image. The experimental data collection time is between 10:00 and 13:00 on December 18 and December 20, 2020. The weather is sunny and cloudy. Hyperspectral images of 9 UAVs in two research areas are obtained. In the experiment, it is ensured that the cloud amount and sunlight intensity in the air have little difference in the collection process of each sortie.

3.3. Data Preprocessing. High-resolution remote sensing images are usually large in size and cannot be processed by convolution [24]. For example, the average size of ISPRS images from the Vaihingen dataset is 2493×2063 pixels, while most convolution operations support resolution of 256×256. In view of the memory limitation of the current graphics processing unit (GPU), the proposed method uses a sliding window to segment the original remote sensing image into smaller image blocks. If the convolution step is smaller than the image block size, in the case of overlapping continuous image blocks, multiple predictions are averaged to obtain the final classification of the overlapping pixels [25]. This operation can smooth the prediction of each image block boundary and eliminate possible discontinuities.

The goal of the proposed method is to apply the typical artificial neural network structure to earth observation data. Therefore, using the artificial neural network originally designed for RGB data, the processed image must comply with 3-channel format. The three channels in the ISPRS dataset will be processed into RGB images. The dataset contains the data of digital surface model (DSM) obtained from the aerial laser sensor. Normalized digital surface model (NDSM) will also be used, and then normalized difference vegetation index (NDVI) will be calculated from near-infrared and infrared channels. Finally, DSM, NDSM, and NDVI information will be used to build a corresponding composite image for each IRRG image.

3.4. Network Structure Design. Because the traditional SegNet method has low segmentation accuracy and long training time, an ISegNet network is proposed, and its structure is shown in Figure 3. The original SegNet convolution layers are expanded from 26 layers to 27 layers, and Bottleneck is added. Rectified linear unit (ReLU) and exponential linear unit (ELU) are introduced, respectively, into activation functions to increase the nonlinearity of the network, so that the network can express more complex characteristics.

ISegNet includes 5-layer encoding structure and 5-layer decoding structure. The 5-layer encoding structure is composed of 13 convolution layers with kernel size 3×3, 5 batch normalization (BN) layers, and 4 Maxpooling layers with kernel size 2×2 and step size 2; the 5-layer decoding structure consists of 13 convolution layers with kernel size 3×3, 5 BN layers, and 4 upsampling layers, which is completely symmetrical with the encoding structure.

Figure 2: S185 multirotor UAV system.

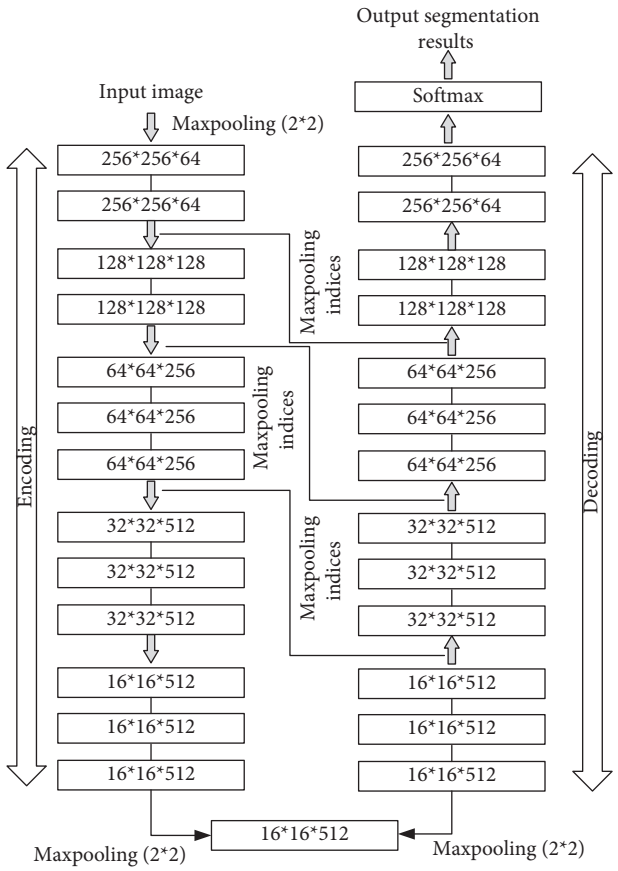

Figure 3: Structure of ISegNet remote sensing image semantic segmentation network.

Convolution layer with kernel size 1×1 is set at the connection between the encoding structure and the decoding structure to further extract nonlinear features of the encoded image data, which deepens the network depth and reduces the amount of parameters.

3.4.1. Encoder Structure. The images normalized to 256×256 pixels are used as the input of the network, and the image features are extracted by the encoder composed of convolution layer and pooling layer. ISegNet uses the convolution of same mode to ensure that the size of image

remains unchanged. Image features are transmitted to the decoder through the Maxpooling indices for nonlinear upsampling to obtain the lost image spatial semantic information in the coding process [26]. The BN layer is added after each Maxpooling layer, and the value range of the features after nonlinear function approaches the saturation region, so as to standardize the distribution of output features. BN mainly includes normalization and transformation reconstruction.

(1) Normalization. The normalization expression of relevant parameters of input sample data in the network is

$$\mu = \frac{1}{k}\sum_{i=1}^{k} x_i,$$
$$\sigma^2 = \frac{1}{k}\sum_{i=1}^{k} (x_i - \mu)^2, \quad (1)$$
$$\widehat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}},$$

where k is the input batch size; μ is the average value of the input; σ^2 is variance; ε is a constant set to maintain parameter stability; x_i is the ith input sample value; $\widehat{x}_i$ is the normalized value corresponding to the ith input sample value. After this step, the feature parameters learned by the network change, and the feature distribution needs to be reconstructed.

(2) Feature Distribution Reconstruction. The expression of feature distribution reconstruction is

$$z_i = \omega \widehat{x}_i + b, \quad (2)$$

where ω and b are the weight and bias, respectively. BN layer can effectively prevent overfitting in model training and speed up learning.

In addition, ISegNet uses ReLU and ELU as the activation function of the network, respectively. ReLU is a commonly used activation function, which changes all negative values to 0 and positive values retain the output, so that neurons are sparsely activated. ReLU has stronger feature mining ability. ReLU is used to fit the training data and to further prevent the gradient from disappearing. The ReLU function is

$$y = \begin{cases} 0, & x \le 0, \\ x, & x > 0, \end{cases} \quad (3)$$

where x is neuron input and y is neuron output.

ELU is also a kind of correction activation function, which adds a nonzero output for negative input. Unlike ReLU, ELU activation function includes a negative exponential term. The function expression of ELU is

$$y = \begin{cases} a[\exp(x) - 1], & x < 0, \\ x, & x \ge 0, \end{cases} \quad (4)$$

where a is a constant value.

3.4.2. Decoder Structure. The decoder is composed of upsampling layer, convolution layer, and pooling layer. The upsampling layer can recover part of the lost information in Maxpooling layer and recover the pixel position information according to the pooling index. The convolution of same mode is also used in the decoder. The upsampling process is shown in Figure 4.

The decoder uses the index stored in each Maxpooling layer to upsample the corresponding feature map. In order to combine coarse and fine textures and prevent parameter redundancy, a Bottleneck layer is set at the connection of encoding-decoding structure. By introducing convolution layer with kernel size 1×1, the encoder output unit achieves feature dimension reduction. The quantity of encoder output parameters can be expressed as

$$D = \sum_{i=1}^{n} R_{l_i}^2 \cdot R_{c_i} \cdot h,$$
$$D' = h \cdot R + \sum_{i=1}^{n} R_{l_i}^2 \cdot R_{c_i} \cdot R, \quad (5)$$

where D is the output parameter quantity of the encoder; D' is the output parameter quantity after adding 1×1 convolution kernel; R_{l_i} is the kernel size of ith convolution layer; R_{c_i} is the kernel number of ith convolution layer (filter depth); l_i is the neuron at ith layer; c_i is the ith convolution kernel; h is the depth of input feature; R is the number of 1×1 convolution kernels. In general, $R < h$. Therefore, after adding 1×1 convolution kernel, the quantity of calculated parameters is reduced, and each pixel after encoding is linearly combined on different channels, which not only retains the original structure of the encoded image but also expands and widens the network.

Fully connected layer, previous layer of output layer, is replaced with a convolution layer to improve the efficiency of network forward propagation. The features output from the last convolution layer are input into the Softmax classifier, and the classifier finally outputs 5 probabilities, which, respectively, predict the probability that the sample belongs to each category; that is, the number of classification labels is 5. The predicted segmentation corresponds to the category with the maximum probability at each pixel.

3.5. Classifier Design. For the semantic segmentation of remote sensing images, the ensemble learning method is further used to improve the segmentation accuracy. The classification results obtained by Softmax classifier are combined through some strategies to achieve better performance than a single classifier. This model fusion method is essentially a relearning process. Even if one classifier makes an erroneous prediction, other classifiers can correct the error. Generally speaking, this ensemble learning method can improve the segmentation effect to a certain extent, but the process of multimodel learning and relearning will increase the cost of calculation [27]. For the combination strategy of ensemble learning, the proposed method adopts the Plurality Voting method. Assuming that the prediction category is $\{g_1, g_2, \ldots, g_m\}$, for any

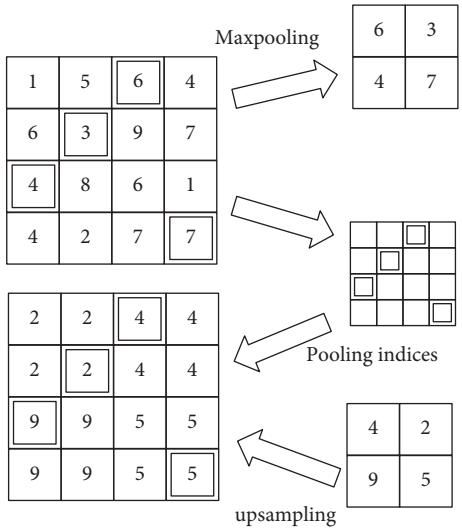

Figure 4: Maxpooling index and upsampling.

prediction sample x, the prediction results of T classification models are $(y_1(x), y_2(x), \ldots, y_T(x))$, respectively. Select the category g_i with the largest number of prediction results of T classification models for sample x as the final classification category. If more than one category gets the highest vote at the same time, select one at random as the final category. Using this ensemble idea, by finding a compromise between multiple classifiers, the worst classifier can be avoided, and some pixels with obvious classification errors can be effectively improved, so as to improve the prediction ability of the model. The ensemble learning process based on multiple classifiers is shown in Figure 5.

First, SegNet is used to complete the semantic segmentation of aerial images, and the prediction results are obtained. Second, the network is improved on the basis of SegNet. The feature maps of different scales in the encoding module are copied to the corresponding upsampling part by skip connection. The improved network model (ISegNet) is trained again, and the prediction results based on this model are obtained. Then, based on ISegNet, five-classification problem is transformed into five binary classification problems for training, and the prediction results of each category are combined to obtain the prediction result. Finally, the segmentation results obtained from the above three different models are combined through ensemble learning, and the ensemble prediction results are obtained by using the Plurality Voting method.

4. Experiments and Analysis

In the experiment, the training equipment of deep learning network is 4-core 8-thread Intel i7-7700K CPU, 32 GB memory, NVIDIA GTX1080 video card, and 8G video memory. The software environment is Ubuntu 16.04.01 operating system. The development platform is Anaconda 4.3.1. The built-in Python version is 3.6.1. The deep learning software framework is TensorFlow 1.2. At the same time,

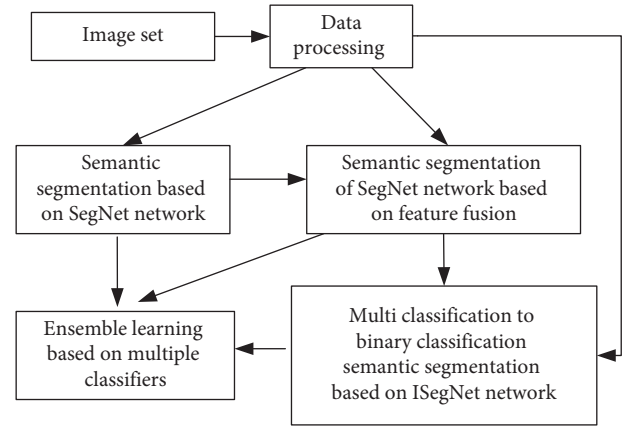

Figure 5: Ensemble learning process based on multiple classifiers.

the experimental dataset is divided according to 8 : 1, and the learning rate is set to 0.01 and the number of iterations is 60.

4.1. Evaluation Index. Kappa coefficient has an important application in the accuracy evaluation of remote sensing classification images. Value range of Kappa coefficient is (–1, 1). A value greater than 0.8 means good classification, and a value of 0 or lower means poor classification. Kappa coefficient is calculated as follows:

$$\varphi = \frac{p_0 - p_e}{1 - p_e},$$

$$p_0 = \frac{\sum_{i=1}^{r} y_{ii}}{N}, \tag{6}$$

$$p_e = \frac{\sum_{i=1}^{r} (y_{i+} \cdot y_{i+})}{N^2},$$

where p_0 is the observation precision ratio, which reflects the proportion of correctly segmented cells; p_e is the contingency consistency ratio, which indicates the proportion of wrong segmentation caused by accidental factors. N is the total number of samples; y_i is the segmented sample.

In addition, the F1 value is used to evaluate the experimental results, and the calculation is as follows:

$$F_{1t} = 2 \times \frac{\text{precision}_t \times \text{recall}_t}{\text{precision}_t + \text{recall}_t},$$

$$\text{recall}_t = \frac{p_{t0}}{Q_t}, \tag{7}$$

$$\text{precision}_t = \frac{p_{t+}}{P_t},$$

where p_{t+} and p_{t0} are the numbers of pixels correctly recognized and recalled in category t, respectively. Q_t is the number of pixels belonging to category t; P_t is the number of pixels in category t identified by the model. In addition, the boundary of the object in the test label image is eroded by a

circular area with a radius of 3 pixels. These eroded areas are ignored in the evaluation process to reduce the impact of uncertain boundary definition. Therefore, the performance of the test set is slightly better than that of the validation set.

4.2. Network Loss and Accuracy Curve. When training the proposed image semantic segmentation network model, the experimental loss and accuracy curve are shown in Figure 6. Generally speaking, the lower the loss and the higher the accuracy, the better the segmentation performance of the network.

As can be seen from Figure 6, when the number of iterations reaches 35, the loss value and accuracy value of the proposed network begin to stabilize. When the iteration is completed, the two values are 0.08 and 0.901, respectively. Combined with the training oscillation degree, network segmentation effect, convergence speed, and other factors, the comprehensive performance of the proposed network is ideal. The deep image features are extracted through ISegNet to reduce the loss of detail information. At the same time, the ensemble idea is introduced to aggregate the classification results to further ensure the classification accuracy. The ensemble idea makes the network model more stable, reduces the training oscillation frequency, and makes the remote sensing image segmentation more accurate.

4.3. Remote Sensing Image Semantic Segmentation Results. Before the detailed quantitative evaluation of remote sensing image semantic segmentation results, it is necessary to qualitatively show the results of remote sensing image semantic segmentation by the proposed method and the methods in [17, 22] to demonstrate the performance of the proposed method. The comparison results of the three methods are shown in Figure 7.

As can be seen from Figure 7, compared with other comparison methods, the segmentation result of the proposed method is closest to the artificial mark, and the overall visual perception is the best. Because the ISegNet includes ReLU and ELU activation functions, the nonlinearity of the network can be increased, more complex features can be expressed, and the fused classifier design further reduces the segmentation error, especially the image details. Reference [17] used deep neural network for image semantic segmentation, but the segmentation effect of complex images was not ideal. For example, it is difficult to distinguish vegetation and water, and the results contain misclassification. Reference [22] realized pixel level end-to-end semantic segmentation based on the improved U-Net deep convolution neural network. The segmentation result is clear and the loss of targets is less. However, compared with the proposed method, it lacks the reclassification of classifier. Therefore, for the segmentation of small targets, its edge effect needs to be improved.

4.4. Semantic Segmentation Quality Evaluation of Different Categories. In order to better evaluate the performance of the proposed method, three methods are quantitatively analyzed, and the results are shown in Figure 8.

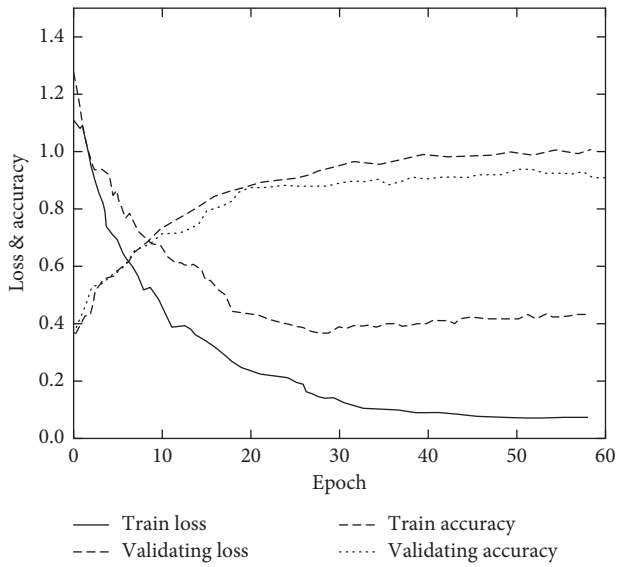

Figure 6: Network loss and accuracy curve.

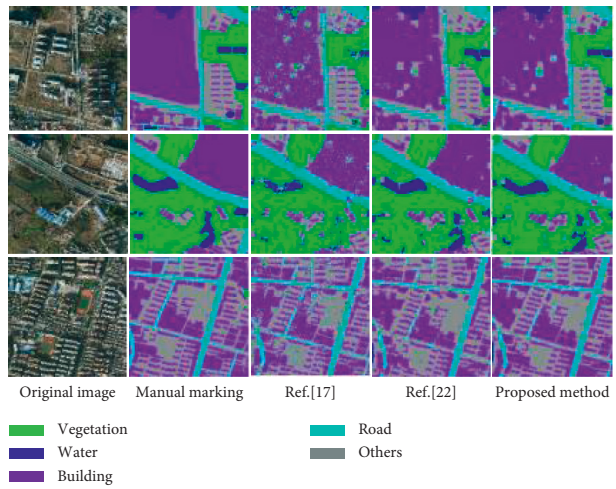

Figure 7: Semantic segmentation results of remote sensing images.

As can be seen from Figure 8, the proposed method is more ideal for the segmentation results of buildings and water bodies, while it is weak for the image segmentation of roads and other categories. Because the roads are narrow and long and easy to be blocked by trees, there may be missing points, resulting in low segmentation result values. However, the proposed method performs relearning classification based on the segmentation results obtained by ISegNet. Compared with other comparison methods, its segmentation result is more ideal. Taking the F1 value as an example, the segmentation results of the proposed method for vegetation, buildings, and water bodies exceed 0.85 and are generally not less than 0.83. Reference [17] used deep neural network for image segmentation, but the segmentation model lacked the ability to extract and learn complex features, so the segmentation result of road and other types of images was lower than 0.73. Reference [22] realized image segmentation based on improved U-Net depth convolution neural network. The segmentation results of some image

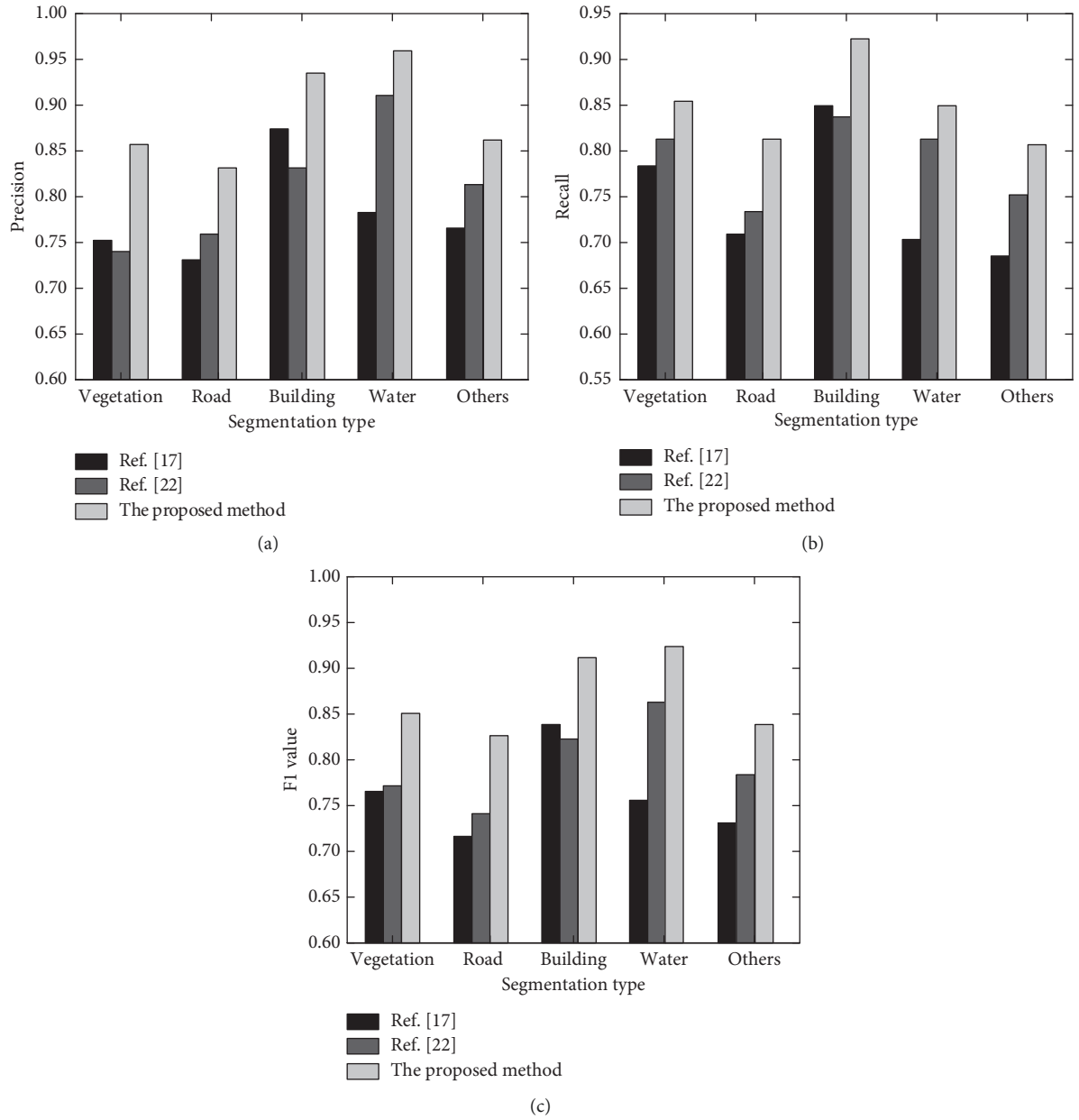

Figure 8: Evaluation index values of different semantic segmentation methods of remote sensing images. (a) Precision. (b) Recall. (c) F1 value.

categories such as water body were similar to those of the proposed method, but the segmentation result of some small buildings was not ideal. Taking precision value as an example, it is lower than 0.85. Overall, the proposed method has the best segmentation result for five categories of images.

5. Conclusions

Image segmentation is an important basic part of remote sensing interpretation. UAV remote sensing image contains complex ground object information, and the application of traditional segmentation methods is greatly limited. Therefore, an image semantic segmentation method based on deep learning in UAV aerial remote sensing image is proposed. The preprocessed image is input into ISegNet for learning and analysis, and five-classification problem is transformed into 5 binary classification problems for training in the classifier, so as to output high-precision image semantic segmentation results. The experimental results show the following:

(1) ISegNet uses pooling index and 1×1 Bottleneck layer to further extract image details, so the segmented image is closest to the artificial standard, and the accuracy value reaches 0.901.

(2) The proposed method improves the classifier and relearns the classification problem to ensure the segmentation result. Its F1 value is not less than 0.83,

and the overall segmentation effect is better than those of other comparison methods.

Although the proposed method can obtain high segmentation precision, the model used is more complex and has more parameters. When the extracted features are more abstract, the model complexity is higher and the training time is longer. In practical application, efficiency needs to be further considered, such as parallel processing. Further optimization will be carried out from the model itself or distributed computing using deep learning framework.

Data Availability

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

Conflicts of Interest

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

Acknowledgments

This work was supported by Natural Science Research Projects in Colleges and Universities of Anhui Province under Contract no. KJ2020A0719; Chuzhou Science and Technology Planning Program under Contract no. 2021ZD010; key project of research and development in Chuzhou Science And Technology Program under Contract no. 2020ZG016; Open Fund of Hunan Provincial Key Laboratory of Geo-Information Engineering in Surveying, Mapping and Remote Sensing, Hunan University of Science and Technology under Contract no. E22136; and Innovation program for Returned Overseas Chinese Scholars of Anhui Province under Contract no. 2021LCX014.

References

[1] Q. Zhou, Y. Wang, J. Liu, X. Jin, and L. J. Latecki, "An open-source project for real-time image semantic segmentation," *Science China (Information Sciences)*, vol. 62, no. 12, pp. 246-247, 2019.

[2] L. Jing, Y. Chen, and Y. Tian, "Coarse-to-Fine semantic segmentation from image-level labels," *IEEE Transactions on Image Processing*, vol. 29, no. 2, pp. 225–236, 2019.

[3] B. Yan, X. Niu, B. Bare, and W. Tan, "Semantic segmentation guided pixel fusion for image retargeting," *IEEE Transactions on Multimedia*, vol. 22, no. 3, pp. 676–687, 2020.

[4] Y. Li, T. Shi, Y. Zhang, W. Chen, Z. Wang, and H. Li, "Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 175, no. 3, pp. 20–33, 2021.

[5] S. Y. Wang, S. C. Wang, and D. Wang, "A road extraction method for remote sensing image based on encoder-decoder network," *Journal of Surveying and Mapping: English Edition*, vol. 2020, no. 2, pp. 16–25, 2020.

[6] H. Liu, H. Du, D. Zeng, and Q. Tian, "Cloud detection using super pixel classification and semantic segmentation," *Journal of Computer Science and Technology*, vol. 34, no. 3, pp. 622–633, 2019.

[7] X. Zhang, Z. Xiao, D. Li, M. Fan, and L. Zhao, "Semantic segmentation of remote sensing images using multiscale decoding network," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 9, pp. 1492–1496, 2019.

[8] K. Zhang, Y. Han, J. Chen, Z. Zhang, and S. Wang, "Semantic segmentation for remote sensing based on RGB images and lidar data using model-agnostic meta-learning and partical swarm optimization," *IFAC-PapersOnLine*, vol. 53, no. 5, pp. 397–402, 2020.

[9] C. Peng, Y. Li, L. Jiao, Y. Chen, and R. Shang, "Densely based multi-scale and multi-modal fully convolutional networks for high-resolution remote-sensing image semantic segmentation," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 8, pp. 2612–2626, 2019.

[10] X. Sun, B. Wang, Z. Wang, H. Li, H. Li, and K. Fu, "Research progress on few-shot learning for remote sensing image interpretation," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, no. 7, pp. 2387–2402, 2021.

[11] Y. Li, B. Peng, L. He, K. Fan, and L. Tong, "Road segmentation of unmanned aerial vehicle remote sensing images using adversarial network with multiscale context aggregation," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2279–2287, 2019.

[12] X. Li, F. Xu, L. Xin et al., "Dual attention deep fusion semantic segmentation networks of large-scale satellite remote-sensing images," *International Journal of Remote Sensing*, vol. 42, no. 9, pp. 3583–3610, 2021.

[13] L. Zabawa, A. Kicherer, L. Klingbeil, R. Töpfer, H. Kuhlmann, and R. Roscher, "Counting of grapevine berries in images via semantic segmentation using convolutional neural networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 164, no. 5, pp. 73–83, 2020.

[14] M. Wurm, T. Stark, X. X. Zhu, M. Weigand, and H. Taubenböck, "Semantic Segmentation of slums in satellite images using transfer learning on fully convolutional neural networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 150, no. 4, pp. 59–69, 2019.

[15] W. Song, N. Zheng, R. Zheng, X.-B. Zhao, and A. Wang, "Digital image semantic segmentation algorithms: a survey," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 10, no. 1, pp. 196–211, 2019.

[16] K. Liu, Z. Ye, H. Guo, D. Cao, L. Chen, and F.-Y. Wang, "FISS gan: a generative adversarial network for foggy image semantic segmentation," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 8, pp. 1428–1439, 2021.

[17] W. Chen, W. Wang, K. Wang, Z. Li, H. Li, and S. Liu, "Lane departure warning systems and lane line detection methods based on image processing and semantic segmentation: a review," *Journal of Traffic and Transportation Engineering*, vol. 42, no. 6, pp. 25–51, 2020.

[18] W. Liu, "Real-time obstacle detection based on image semantic segmentation and fusion network," *Traitement du Signal*, vol. 38, no. 2, pp. 443–449, 2021.

[19] L. Mi and Z. Chen, "Superpixel-enhanced deep neural forest for remote sensing image semantic segmentation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 159, no. 8, pp. 140–152, 2020.

[20] A. Dc, B. Sn, and D. Jhc, "Aerial image semantic segmentation using DCNN predicted distance maps," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 161, no. 2, pp. 309–322, 2020.

[21] M. Shahzad, A. I. Umar, M. A. Khan, S. H. Shirazi, Z. Khan, and W. Yousaf, "Robust method for semantic segmentation of whole-slide blood cell microscopic image," *Computational*

and Mathematical Methods in Medicine, vol. 2020, no. 3, Article ID 4015323, 13 pages, 2020.

[22] M. Pai, V. Mehrotra, U. Verma, and R. M. Pai, "Improved semantic segmentation of water bodies and land in SAR images using generative adversarial networks," *International Journal of Semantic Computing*, vol. 14, no. 1, pp. 55–69, 2020.

[23] F. Mohammadimanesh, B. Salehi, M. Mandianpari, E. Gill, and M. Molinierd, "A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 151, no. 5, pp. 223–236, 2019.

[24] E. Collier, S. Mukhopadhyay, K. Duffy et al., "Semantic segmentation of high resolution satellite imagery using generative adversarial networks with progressive growing," *Remote Sensing Letters*, vol. 12, no. 5, pp. 439–448, 2021.

[25] A. Kumthekar and G. R. Reddy, "An integrated deep learning framework of U-Net and inception module for cloud detection of remote sensing images," *Arabian Journal of Geosciences*, vol. 14, no. 18, pp. 1–13, 2021.

[26] H. Shao, Y. Li, Y. Ding, Q. Zhuang, and Y. Chen, "Land use classification using high-resolution remote sensing images based on structural topic model," *IEEE Access*, vol. 8, no. 6, Article ID 215943, 2020.

[27] I. Kotaridis and M. Lazaridou, "Remote sensing image segmentation advances: a meta-analysis," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, no. 3, pp. 309–322, 2021.