Hindawi
Journal of Environmental and Public Health
Volume 2022, Article ID 9974914, 12 pages
https://doi.org/10.1155/2022/9974914

Research Article

Landscape Classification Method Using Improved U-Net Model in Remote Sensing Image Ecological Environment Monitoring System

Jing Wang

Art Academy, Northeast Agricultural University, Harbin, Heilongjiang 150030, China

Correspondence should be addressed to Jing Wang; wangjing0720@neau.edu.cn

Received 29 July 2022; Revised 15 August 2022; Accepted 25 August 2022; Published 21 September 2022

Academic Editor: Pengjiang Qian

Aiming at the problems of low classification accuracy and time-consuming properties in traditional remote sensing image classification methods, a remote sensing image classification method of ecological garden landscape based on improved U-Net model is proposed. Firstly, the remote sensing images of ecological garden landscape are collected by s185 multirotor unmanned aerial vehicle (UAV) system and preprocessed by min-max standardization and data enhancement. Then, the asymmetric convolution block and attention mechanism are used to improve the U-Net model to form the Att-Unet network model, so as to overcome the problems of easy overfitting of the model and incomplete small target detection. Finally, the fully connected conditional random field is introduced into the classification postprocessing to refine the segmentation results. Based on the Keras learning framework, the proposed method is experimentally demonstrated. The results show that the recall, precision, F1 value, and accuracy of the proposed method in the remote sensing image of ecological garden landscape are 0.854, 0.801, 0.836, and 0.982, respectively, and the classification test time is 8.9s. The overall performance is better than other comparison methods, which can provide theoretical support for the dynamic monitoring of the development of ecological garden.

1. Introduction

The classification and identification of vegetation are the basis for the study of the status and dynamic changes of the ecological garden landscape. Early vegetation remote sensing image classification is often carried out through large-scale remote sensing images, which is more suitable for northern ecological gardens with simple vegetation types and large plots [1, 2]. The southern ecological garden landscape has the characteristics of complex structure and fast growth, which brings great difficulties to the fine classification of vegetation. The traditional remote sensing image classification process usually includes three steps. First, the image preprocessing technology is applied to register and denoise the image, so as to eliminate the image difference caused by imaging factors. Then, the difference image is generated by image difference, ratio, and other methods. Finally, the difference image is classified, and detailed features are extracted from it for classification [3]. Generally speaking, the basis for identifying the types of ecological garden landscapes is the difference in the spectral characteristics of vegetation. The analysis of vegetation spectral characteristics and species identification based on the measured reflectance spectrum data is an important content of remote sensing theoretical research. It helps to grasp the spectral separability of different vegetation types, so as to more effectively carry out species identification [4,5].

In recent years, with the rapid development of high-resolution satellites, high-resolution remote sensing image data has increased dramatically, which has provided convenience for the application and research of remote sensing images. But at the same time, there is a problem that, for the original remote sensing image interpretation processing speed, it is difficult to meet the existing needs. Therefore, research on an efficient and accurate remote sensing image classification and recognition model has become an urgent need [6,7]. At present, there have been many researches on

remote sensing image classification technology, and the application of land vegetation cover classification has been relatively mature, but the classification application of remote sensing image of ecological garden landscape is still less [8]. There are mainly two remote sensing image classification methods, namely, traditional classification models and classification models based on deep learning [9].

The traditional computer classification method is to extract spectral information on the basis of pixels to determine the category of the pixels [10]. The basic idea is that, in the same feature space, pixels of the same type of features are clustered together, while pixels of different types of features are separated from each other. The classification effect of remote sensing images is closely related to the classifier and classification algorithm. Commonly used classification algorithms include supervised methods such as maximum likelihood method, minimum distance, and Mahalanobis distance, and unsupervised methods such as k-means [11]. Yuan et al. (2019) proposed a method based on rearranging local features to solve the problem of high correlation between remote sensing image categories and local features [12]. By fusing side classes to combine global and local features to enhance image representation, the accuracy of image classification still needs to be improved. Dano U *L* et al. (2020) compared and analyzed three image classification algorithms such as the minimum distance based on the application of remote sensing and geographic information system (GIS) computer programs [13]. Hu S et al. (2021) proposed an evolutionary expansion and contraction method for remote sensing image data processing [14]. After expanding multiple data streams into subspaces, data stream mining and image bound model learning are dynamically completed, which improve the accuracy of image recognition. However, the performance of image classification for complex ecological garden vegetation remains to be verified.

With the development of deep learning, its advantages such as strong ability to automatically extract features, less manual intervention, and being not limited by the input size of the image have gradually become prominent. And this advantage provides a new idea for the classification of remote sensing images of ecological gardens [15]. ShujunLiang et al. (2019) proposed a maximum likelihood classification model for soil remote sensing images combined with deep learning network [16]. Extract soil targets in remote sensing images through deep learning network and use maximum likelihood method to classify soil remote sensing images. However, the problem of image category diversity and category similarity is still not well resolved. H. Song et al. (2020) proposed a new dual-channel densely connected convolutional network based on deep learning and multisource remote sensing data to automatically classify surface remote sensing images [17]. The dual channel dense connection convolution network carries out feature extraction and integrates hyperspectral and radar features to output accurate classification results. However, the detailed feature processing of the feature image needs to be improved. Zhang C. et al. (2019) proposed a multiscale dense network for hyperspectral remote sensing image (HRSI) classification [18]. It makes full use of and combines different scale information in the entire network structure to realize the feature extraction and classification of two-dimensional remote sensing images. However, some global or local information of HRSI is ignored. Convolutional neural network (CNN) is an emerging computer processing model. Many scholars have applied it to remote sensing image classification processing and achieved good research results. Zhao F. et al. (2018) used the pretrained CNN model as a feature extractor to extract deep-level features from the fully connected layer to complete the accurate classification of HRSI images [19]. But the robustness and computational efficiency of the model are not good. Cheng *G.* et al. (2018) proposed a deep CNN to improve remote sensing image scene classification performance [20]. By optimizing the new discriminant objective function for training, the regularization learning is strengthened, and the classification model is more discriminative. However, the degree of discrimination is not high for objects with high similarity in the scene.

Aiming at the problem that most of the existing remote sensing image classification methods do not easily meet the complex and changeable ecological garden landscape, a remote sensing image classification method of ecological garden landscape using an improved U-Net model is proposed. The innovations are summarized as follows:

(1) Since the U-Net model is prone to overfitting during training, the proposed Att-Unet model uses asymmetric convolution blocks instead of standard convolution operations to enhance the robustness of the convolution kernel and the central skeleton of the network. In addition, the attention mechanism is introduced to strengthen the learning of change characteristics to solve the problems of complicated remote sensing image background and easy-to-miss detection of small target changes.

(2) Considering that many vegetations in ecological gardens are relatively similar, the probability of adjacent pixels belonging to the same category is higher. The conditional random field is introduced into the Att-Unet model, and the extracted feature map is input as the conditional random field to improve the fineness of target edge segmentation.

The remaining chapters of this paper are arranged as follows: the second chapter introduces the remote sensing image data source and image preprocessing process. The third chapter introduces the classification method of remote sensing images based on the improved U-Net model. In Chapter 4, experiments are designed to verify the performance of the proposed method. The fifth chapter is the conclusion.

2. Remote Sensing Image Data Source and Image Preprocessing

2.1. Remote Sensing Image Acquisition. The acquisition of remote sensing images of an ecological garden in the suburbs of Harbin, Heilongjiang Province, mainly uses the S185 unmanned aerial vehicles (UAV) system, as shown in

Figure 1. The system mainly includes Cubert S185 hyperspectral data acquisition system, six-rotor electric unmanned aerial vehicle system (maximum load is about 6 kg, endurance time is 15min-30 min), and three-axis stabilization gimbal and data processing system.

The remote sensing image data acquisition system is mainly composed of the German Cubert S185 airborne high-speed imaging spectrometer and a micro control unit (used for data acquisition and data storage). During flight operations, the S185 is precalibrated by black and white radiation, and the exposure time is automatically matched. When the altitude is 100m, the flight speed is 4.8 m/s, the sampling interval is 0.8s, the heading overlap rate is about 80%, and the side overlap rate is about 70%, it can simultaneously acquire 125 effective bands of hyperspectral data (450nm-946 nm) and clear images with a spatial resolution of about 2.6 cm.

The collection of UAV hyperspectral data should be carried out on sunny days to avoid cloud shadows affecting image quality. And the deflection angle should not be too large during the acquisition time to avoid too large shadow area in the image. The experimental data collection time was within 10 : 00–13 : 00 on October 8th and October 10th, 2019. The weather was fine and slightly clouded. A total of 12 UAV hyperspectral images were acquired in 3 survey areas. In the experiment, it was ensured that the amount of cloud and the intensity of sunlight in the sky had little difference during the collection of each sortie.

2.2. Data Preprocessing. A data normalization operation is required to control the data distribution within a certain range. Data normalization is an important preprocessing step before model training [21, 22].

At present, the commonly used normalization methods are min-max standardization and z-score standardization. Min-max standardization is also called dispersion standardization, which maps data values to [0, 1]. This method is suitable for data distributed in a limited range. The standardized data x' is calculated as follows:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{1}$$

where x is the data before standardization, $x_{\max}$ is the maximum value of the sample data, and $x_{\min}$ is the minimum value of the sample data.

Z-Score standardization is to use the mean and standard deviation of the original data for standardization. This method is suitable for situations where there is no obvious boundary. The data standardized by z-score conforms to the standard normal distribution; that is, the mean is 0 and the standard deviation is 1. The calculation is as follows:

$$x' = \frac{x - \overline{x}}{\sigma}, \tag{2}$$

where $\overline{x}$ is the mean value of the data and σ is the standard deviation of the data.

Appropriate data normalization methods have a great impact on the effectiveness and accuracy of model training.

Figure 1: S185 multirotor UAV system.

Select min-max standardization to normalize the training and verification images and reduce the data range of each channel from the interval [0, 255] to the interval [0, 1].

In addition, data enhancement is an essential step in deep learning image classification tasks. Commonly used image enhancement methods can be roughly divided into three categories: color transformation, geometric transformation, and cropping [23]. It is easy to lose important information in cropping operation, and geometric transformation is more suitable for remote sensing images with different shapes and angles. Therefore, geometric transformation is mainly adopted, and the data enhancement operation of random flip (including horizontal and vertical flip) is performed on the sample to be trained after cropping.

3. Remote Sensing Image Classification Based on Improved U-Net Model

3.1. Modeling. The process of using Att-Unet model and fully connected conditional random fields (CRFs) to classify remote sensing images of ecological garden landscape is mainly divided into two phases: training phase and classification and postprocessing phase. The entire classification process is shown in Figure 2.

The upper part of Figure 2 is the training phase. The training samples composed of multisource remote sensing images and ground real data are input into the Att-Unet model for feature learning, and the predicted probability distribution map is obtained. Then, the cross entropy function is used to measure the loss value between the predicted classification result and the ground truth data. The Adam optimization algorithm is used to reduce the loss value, and the parameters in the Att-Unet model are continuously updated iteratively until the loss value is reduced to a given threshold range; then, the training ends and the optimal Att-Unet model is obtained. The lower part of Figure 2 is the classification and postprocessing stage. It uses the trained Att-Unet model to classify images to be classified and obtains preliminary classification results. Then, combined with the original image to be classified, fully connected CRFs are used to adjust and optimize the classification results to improve the misclassification phenomenon and

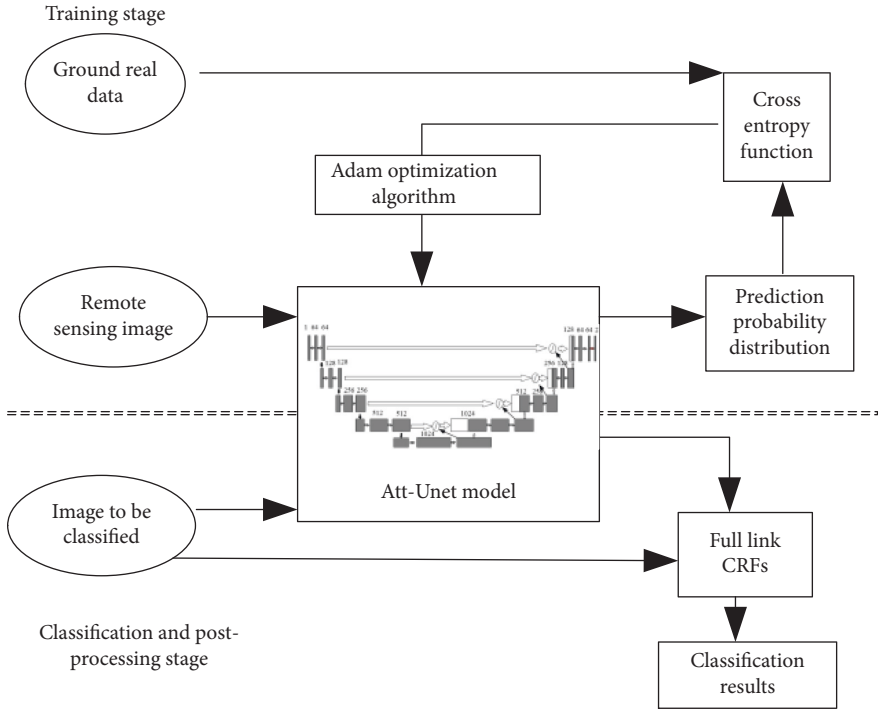

Figure 2: Classification process structure of the proposed model.

refine the edges of the features to obtain a more detailed and accurate classification.

3.2. Att-Unet Model Training

3.2.1. Asymmetric Convolution Block. The U-Net network has gone through 5 coding blocks in the feature extraction part, and ten 3×3 standard convolution operations. Repeated convolution operations will cause information loss in the feature extraction part of the network, which is prone to overfitting and affects the accuracy of detection [24]. The internal structure of the U-Net network has been improved. In the feature extraction process, an asymmetric convolution block (ACB) is used to replace the 3×3 standard convolution to improve the accuracy of network change detection. The structure of the ACB module is shown in Figure 3.

ACB is a convolution operation obtained by accumulating convolution results of a set of convolution kernels of 3×3, 1×3, and 3×1. It is equivalent to adding two single convolution operations with 1×3 and 3×1 convolution kernels at the center of the 3×3 convolution kernel to obtain an equivalent output. Using ACB to replace the standard 3×3 convolution of the feature extraction part can enrich the feature space during the training process. The knowledge learned by the model is incorporated into the square kernel, the central skeleton part of the square convolution kernel is enhanced, and the information loss caused by the convolution operation is reduced. The robustness of the model to rotation distortion is enhanced without adding additional parameters and calculations, thereby improving the accuracy of the model.

3.2.2. Attention Mechanism. The remote sensing image contains a variety of features such as buildings, vegetation, bare land, farmland, and waters. The proposed model only focuses on ecological gardens, and other types of ground objects are treated as background. The background situation is more complicated, which greatly interferes with the accuracy of the classification results. Therefore, the attention mechanism is introduced in the step connection part of the U-Net network to adjust the feature weights and suppress the model learning features that are not related to the changing pixels. Focus on learning features related to changeable pixels and strengthen the model's extraction of features of changeable ecological gardens. The structure of the attention mechanism is shown in Figure 4.In the structure, $\mathbf{d}$ is the feature map matrix of the decoding part, $\mathbf{e}$ is the feature map matrix of the coding part, H, W, and C, respectively, represent the length, width, and number of channels of the feature map, and $\omega_{\mathbf{d}}$ and $\omega_{\mathbf{e}}$ are the feature weight matrix. The specific operation of the attention mechanism is divided into three steps.

3.2.3. Feature Weight Extraction

$$\begin{aligned} \boldsymbol{\omega}_{\mathbf{e}} &= \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} e(i, j), \\ \boldsymbol{\omega}_{\mathbf{d}} &= \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} d(i, j), \end{aligned} \tag{3}$$

where i and j correspond to the pixel positions in the feature map and $e(i, j)$ and $d(i, j)$ are the pixels in the encoder and decoder, respectively.

By performing global average pooling on the feature map $\mathbf{e}$ of the encoding part and the feature map $\mathbf{d}$ of the decoding

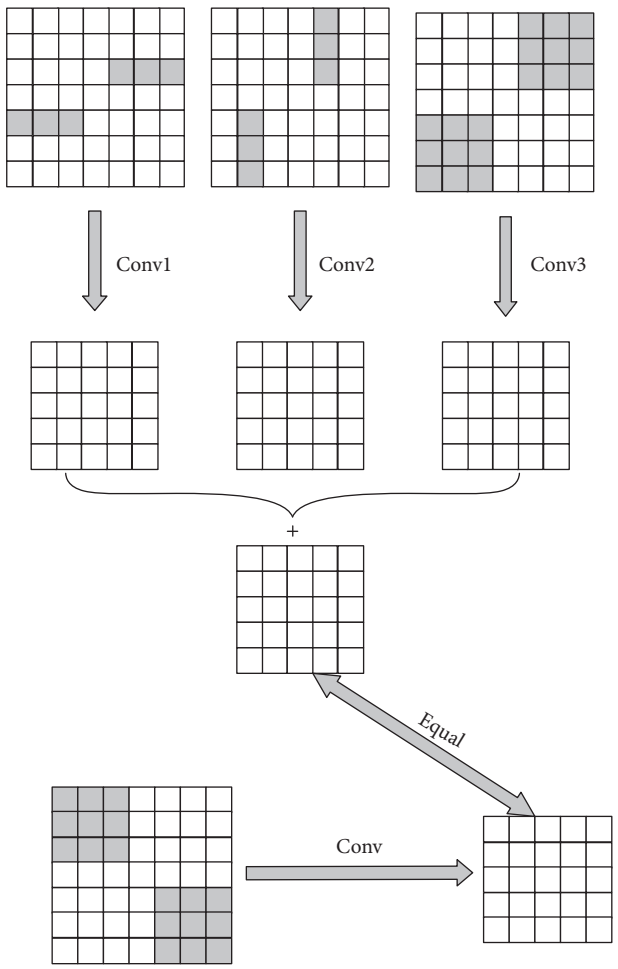

Figure 3: Structure of ACB module.

part, respectively, the weight matrices $\mathbf{w_d}$ and $\omega_\mathbf{e}$ of the feature maps containing C channel information are obtained.

(1) Feature weight ω update is

$$\begin{aligned} Q_{att} &= \boldsymbol{\psi}^T\left(\delta_1\left(\boldsymbol{\omega}_\mathbf{e}^T\mathbf{e} + \boldsymbol{\omega}_\mathbf{d}^\mathbf{T}\mathbf{d}\right)\right), \\ \boldsymbol{\omega} &= \delta_2\left(Q_{att}\left(\mathbf{e}; \Theta_{att}\right)\right), \end{aligned} \tag{4}$$

where δ_1 is the activation function of Rectified Linear Unit (ReLU), δ_2 is the Sigmoid function, and Θ_{att} represents the proportion of backpropagation learning.

The attention mechanism realizes the update of feature weights through two fully connected layers. First of all, by multiplying $\omega_\mathbf{e}$ by $\mathbf{e}$ and $\omega_\mathbf{d}$ by $\mathbf{d}$, the full connection operation of the encoding part of the feature map and the decoding part of the feature map is realized, reducing the amount of parameter calculation. Then, the result of the fully connected layer is summed and then passed through the ReLU layer, and the result is multiplied by the ψ point to make a full connection again. The weight matrices $\omega_\mathbf{d}$ and $\omega_\mathbf{e}$ are learned through backpropagation, and the importance of each element in the $\mathbf{d}$ and $\mathbf{e}$ matrices is obtained.

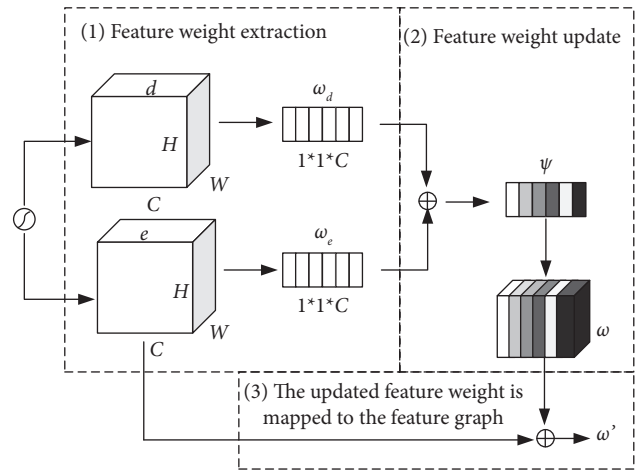

Figure 4: Attention mechanism module.

Accordingly, the proportion of the $\mathbf{d}$ and $\mathbf{e}$ matrices to continue forward propagation is adjusted. Finally, the weight of each pixel is redistributed, and the weight matrix ω after the feature weight update is obtained through the Sigmoid layer.

(2) The updated feature weights are mapped to the feature map:

$$e' = \mathbf{e} \cdot \boldsymbol{\omega}. \tag{5}$$

Multiply the updated weight matrix ω by the feature map $\mathbf{e}$. Increase the weight of the channel related to change pixels in the feature map, and decrease the weight of the channel related to other pixels. Obtain the feature map with the attention mechanism, and step-connect it with the feature map $\mathbf{d}$ to enter the next decoding layer.

3.2.4. Att-Unet Model and Model Training. The attention mechanism is introduced into the U-Net network, and the resulting Att-Unet network structure is shown in Figure 5. Att-Unet introduces the attention gate in the skip connection part of the U-Net network. A channel-level attention control is performed on the underlying information and the characteristics of the current channel. The characteristics of different channels can be linked, and the characteristics of the same type have mutual restrictions. Compared with the image restored by direct upsampling, it is more refined, and the classification accuracy is also improved [25].

Suppose the image is divided into K categories. For the pixel $n\{n = 1, 2, \ldots, N\}$ in each sample image, N is the total number of pixels. Its true category label is expressed as $y_k^n\{k = 0, 1, \ldots, K-1\}$. The K-dimensional output feature vector obtained by forward propagation of the sample is denoted as $O_\mathrm{k}^n\{k = 0, 1, \ldots, K-1\}$. Then, the process of finding the optimal solution of the model parameters can be transformed into a process of narrowing the gap between the output value O_k^n and the ground truth data y_k^n. Firstly, for multiclass problems, the softmax function is usually used to

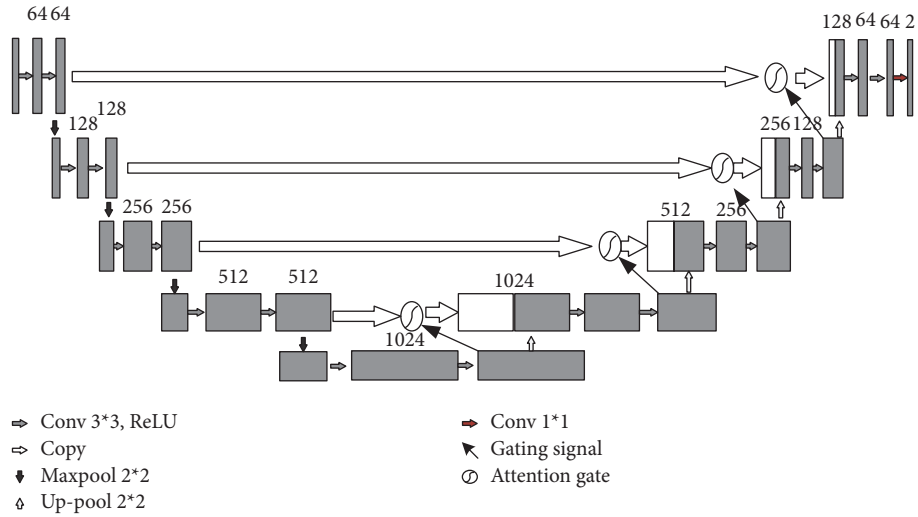

Figure 5: Structure of Att-Unet model.

convert the linear prediction values of all categories in the feature vector O_k^n into probability values. Then, the calculation formula of the predicted probability that the pixel n belongs to the K-th category is

$$p_{\mathrm{k}}(x_n) = \frac{\exp(O_k^n)}{\sum_{k=1}^{K} \exp(O_k^n)}. \tag{6}$$

After obtaining the probability value, use the loss function to calculate the loss value between the ground truth data and the predicted probability to quantify the difference between the two. When the *loss* value is smaller, the classification is more accurate. The cross entropy function is used to calculate the loss value. The formula is as follows:

$$\mathrm{loss} = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{C} y_k^n \log p_k(x^n). \tag{7}$$

The process of model training is the process of optimizing the *loss* function and reducing the value of loss, that is, the process of adjusting and updating the Att-Unet model parameters, also known as backward propagation. The experiment uses Adam optimization algorithm for model training and updates the parameters in the model layer by layer. Adam algorithm is easy to implement, has high computational efficiency and low memory requirements, and is currently one of the commonly used optimization algorithms in deep learning. When the loss value reaches a certain threshold, the training stops.

3.3. Model Prediction. Model prediction refers to the forward propagation process after the parameters of the model are determined. The final model is used to solve the probability that each pixel in the image to be classified belongs to each category. Then, use the argmax function to find the dimension to which the maximum probability belongs, that is, the pixel category label. The specific method is that, for each pixel $n\{n = 1, 2, \ldots, N\}$ in the sample, the predicted probability of belonging to the K-th category is obtained and denoted as $\widehat{p}_k(x^n)$. Then, calculate the category label K_n of the pixel n as follows:

$$E(K) = \sum_{n=1} \psi_U(K_n) + \sum_{E} \psi_P(K_n, K_m). \tag{8}$$

In the model prediction process, in order to prevent memory overflow, the image to be classified is usually cropped into fixed-size image blocks for prediction. Then, stitch together into the whole image. However, due to the convolution operation, the boundary of the image block is filled with 0. Therefore, the prediction method will make the prediction accuracy of the boundary pixel of each image block lower than the prediction accuracy of the center pixel. The classified images obtained after splicing have obvious splicing traces. In order to obtain higher prediction results, a marginal abandonment strategy is adopted. A sliding window is used to obtain image blocks with a certain overlapping area. Then, for each predicted image block, the classification result of a certain area in the middle is retained, and the result of inaccurate edges is discarded and then spliced in sequence. In this way, obvious splicing marks can be avoided and the image prediction effect can be improved.

3.4. Fully Connected CRFs Postprocessing. Upsampling is performed in the Att-Unet network decoder. This step can restore the feature map to the original size. But it also causes the loss of features, and the problem of blurred boundaries of ground objects [26]. In addition, the convolution operation is locally connected, which can only provide information in a rectangular area around a pixel. Although repeated downsampling convolution operations can gradually increase the area of the rectangle, even in the last convolution layer, the correlation between one pixel and all other pixels in the entire image cannot be obtained. In order to solve the above problems and improve the accuracy of classification, the Att-Unet network and fully connected CRFs are combined, by calculating the similarity between two pixels to determine whether they belong to the same category. In the model test,

the output probability distribution diagram of the last layer of the decoder is used as the unary potential energy of fully connected CRFs. The position and color information in the binary potential energy is provided by the original image. The result of image postprocessing is used as the final output result.

The energy function of fully connected CRFs is calculated as follows:

$$E(K) = \sum_{n=1} \psi_U(K_n) + \sum_{E} \psi_P(K_n, K_m). \tag{9}$$

The first term $\psi_U(K_n)$ of the energy equation is a unary potential energy function. It is used to measure the probability of the pixel point belonging to the category label K_n when the color value of the pixel point n is C_n. The second term of the energy equation is a paired potential energy function $\psi_P(K_n, K_m)$, which is used to measure the probability $P(K_n, K_m)$ of two events occurring at the same time, and describes the relationship between each pixel and other pixels. The color and the pixels that are relatively close together are classified into one category, and the calculation formula is as follows:

$$\psi_P(K_n, C_m) = \underbrace{U(K_n, C_m) \sum_{g=1}^{G} \omega^g \kappa_\Delta^g(\mathrm{f}_n, \mathrm{f}_m)}_{\kappa(\mathrm{f}_n, \mathrm{f}_m)}, \tag{10}$$

where U is the label probability function, which calculates the probability that the pixel n and the pixel m belong to the same class. If $K_n \neq C_m$, then $U(K_n, C_m) = 1$; otherwise, it is 0. ω^g is used to balance the function. κ_Δ^g is the Gaussian kernel function. The $\kappa_\Delta^g(f_n, f_m)$ expression is

$$\kappa_\Delta^g(f_n, f_m) = \exp\left(-\frac{1}{2}(f_n, f_m)^{\mathrm{T}} \Lambda^{(g)}(f_n, f_m)\right), \tag{11}$$

where f_n and f_m represent the feature quantity of pixel n and pixel m.

ω^g in (10) is the weight of Gaussian κ_Δ^g. Each Gaussian kernel κ_Δ^g is characterized by a symmetric positive precision matrix $\Lambda^{(g)}$, which defines the shape.

For remote sensing image classification problems, $\kappa(f_n, f_m)$ is usually used in dual-core potential, and the expression is

$$\begin{aligned}\kappa(f_n, f_m) = \omega^{(1)} \exp\left(-\frac{\|L_n - L_m\|^2}{2\vartheta_\alpha^2} - \frac{\|\mathrm{I}_n - \mathrm{I}_m\|^2}{2\vartheta_\beta^2}\right) \\ + \omega^{(2)} \exp\left(-\frac{\|L_n - L_m\|^2}{2\vartheta_\gamma^2}\right),\end{aligned} \tag{12}$$

where L_n and L_m are the pixel position and I_n and I_m are the amount of pixel color. The first item on the right side of the formula is called the appearance kernel, and the second item is called the smooth kernel. The appearance kernel assumes that adjacent pixels with similar colors are likely to belong to the same category, and the function of the smoothing kernel is to eliminate isolated small areas. The function of formula (12) is to judge whether similar pixels belong to the same category. If the pixels belong to the same category, the energy function value is relatively small. Conversely, if the pixels do not belong to the same category, the energy function is relatively large.

In the classification of ecological garden remote sensing images, the use of this energy function can make the classification of garden features and neighboring features more accurate. When pixels in similar areas are judged to be of different types, the energy function value will become larger. When the areas with differences are judged to be the same type, a larger energy value will also be produced. Through multiple iterations, the value of the energy function is minimized to obtain the final result. In this way, the information of the entire image is used to refine the edge of the garden and improve the accuracy of classification.

4. Experiment and Analysis

Att-Unet network will perform a lot of calculations and consume a lot of memory and video memory during training, which requires high hardware. However, due to the price and experimental environment, a balance is pursued in terms of platforms. The proposed model is built based on the deep learning framework Keras, and the deep learning experimental environment is built according to the mainstream configuration environment. The basic software and hardware configuration are shown in Table 1. A total of 1200 images were collected. After preprocessing, 900 images were randomly selected for model training, and the remaining 300 images were used for model testing.

4.1. Evaluation Index. F1-Score is used as an evaluation index, which is an important index to measure the accuracy of classification problems and is the harmonic average of recall and precision. When using F1-score to evaluate model accuracy, F1-score and accuracy rate *Acc* are calculated as follows:

$$\begin{aligned} F1 &= 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \\ \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ Acc &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \end{aligned} \tag{13}$$

where TP represents the number of positive categories that are correctly classified, TN represents the number of negative categories that are correctly classified, FP represents the number of misclassified positive categories, and FN represents the number of negative categories that are misclassified. In the experiment, the positive category is the number of pixels of the change category, and the negative category is the number of pixels of the nonchange category.

4.2. Training Curve. Through multiple experiments, considering the model calculation efficiency, result accuracy,

Table 1: Configuration of basic hardware and software system.

Hardware configuration	Parameter	Software configuration	Parameter
System	Ubuntu 16.04	GPU-Driver	384
CPU	Intel E5-1630	CUDA	10.0
Memory	16 GB	*Python*	3.6
Hard disk	1T	Keras	2.2.4
Graphics card	NVIDIA GTX970	Tensorflow	1.4.0

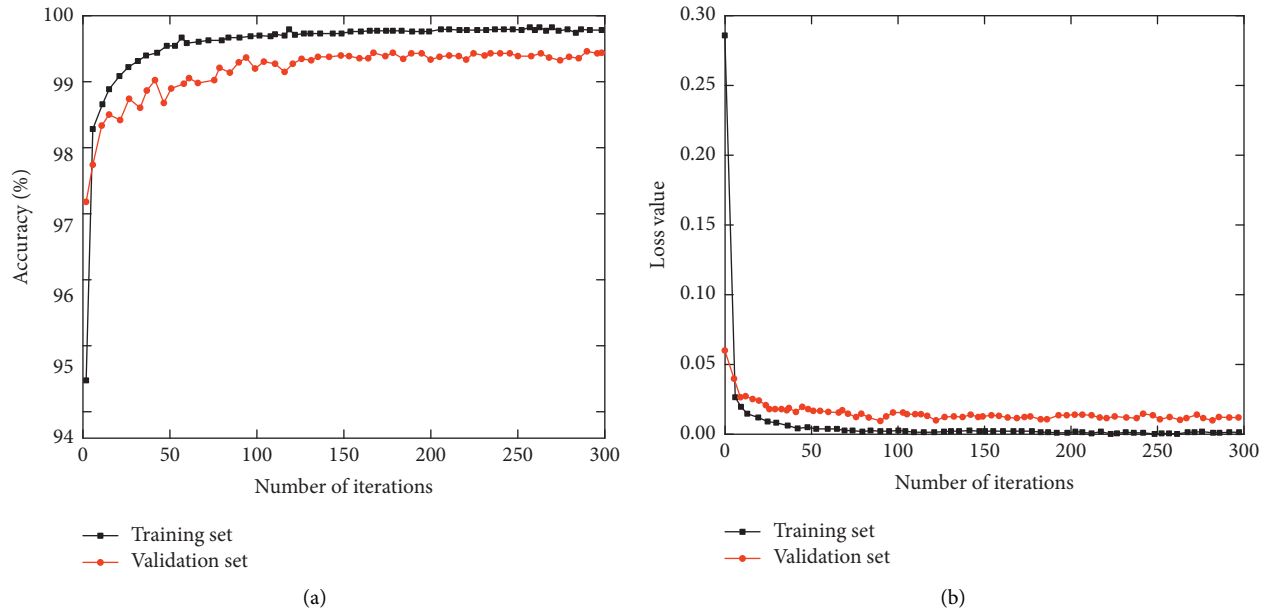

Figure 6: Training curve of improved U-Net model.

and hardware, the experiment finally set the number of iterations to 300 and the batch size to 25. Adadelta was chosen as the optimizer, and the initial learning rate was set to 0.01. The accuracy and loss value changes obtained by the proposed model training are shown in Figure 6.

It can be seen from Figure 6 that when the number of iterations is 50, the proposed model tends to converge. When the number of iterations exceeds 100, the model has steadily converged. At this time, the classification accuracy and loss value on the training set are close to 99.8% and 0.005, respectively. And the classification accuracy and loss value on the validation set are about 99.3% and 0.02, respectively. It can be demonstrated that the proposed model has good convergence performance, fast convergence speed, and ideal classification performance.

4.3. Att-Unet Classification Results. Typical vegetation in an ecological garden in the suburbs of Harbin, Heilongjiang Province, includes rape, sunflower, and reed. The Att-Unet model is used to classify ecological garden landscapes of different ages. The result is shown in Figure 7.

It can be seen from Figure 7 that the distribution of typical vegetation in a certain ecological garden in the suburbs of Harbin with different ages basically remains unchanged. In the early years, most of the gardens were wasteland and the ecological environment was harsh. Reeds are distributed only near the water source. With the improvement of the ecological environment, the wasteland began to be covered with grass, forming grassland. With the continuous development of human activities and the driving of natural evolution, the planting of artificial vegetation also represents human intervention in garden vegetation, changing the natural distribution pattern of garden vegetation. In recent years, large areas of rapeseed and sunflowers have gradually appeared. Among them, the distribution of rape flowers is concentrated in strips and shows a trend of expansion. The change in the distribution area of reeds showed an area that first decreased and then increased, possibly due to the impact of earlier human activities. The garden landscape distribution law obtained by the Att-Unet model is consistent with the actual distribution law. Therefore, the proposed model is effective. It is used to analyze the evolution of the spatial distribution of typical garden vegetation to infer its habitat changes and driving factors so as to realize the dynamic monitoring and protection of the ecological garden landscape.

4.4. Comparison of U-Net and Att-Unet Classification Results. In order to demonstrate the classification performance of the Att-Unet network model, it is compared with the U-Net model. The two classification results of ecological garden remote sensing images are shown in Figure 8.

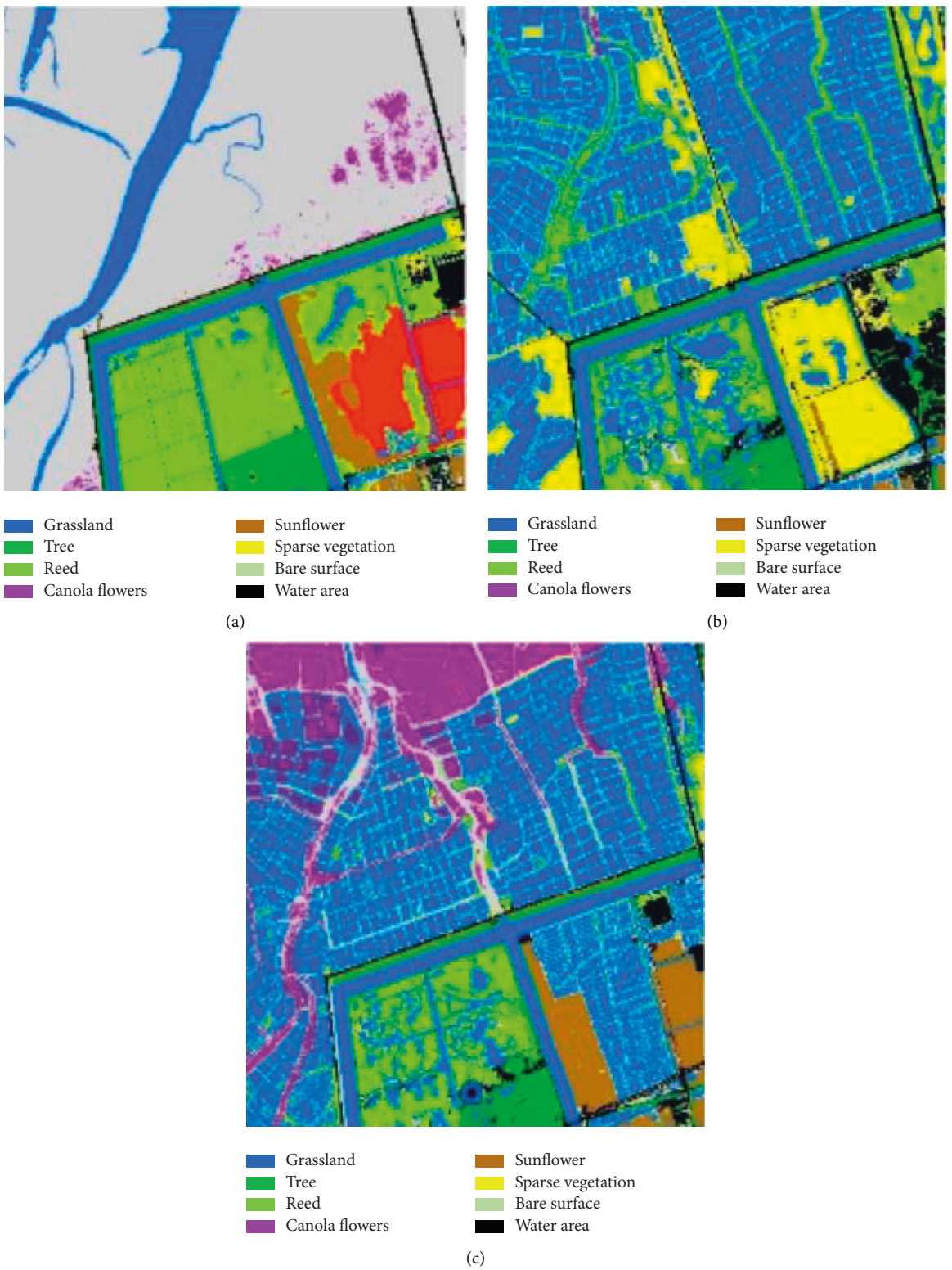

Figure 7: Classification results of Att-Unet model.

It can be seen from Figure 8 that the image learned by the U-Net network has salt and pepper phenomenon. Moreover, the junction of different features is the mixed pixel, and the classification situation is more complicated due to the lack of clear and effective spatial information. The Att-Unet network introduces an attention mechanism and is processed by fully connected CRFs, which can better handle small target classification. The boundary conditions are obviously optimized, and the boundary of reed classification is clearer.

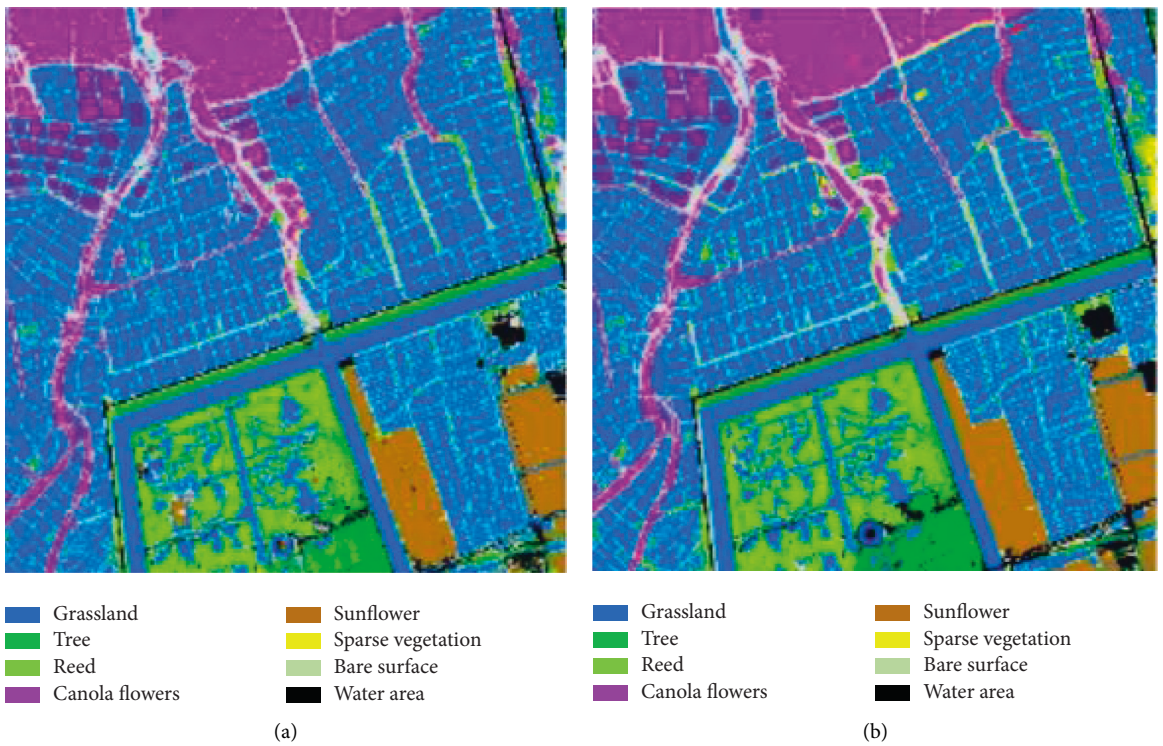

Figure 8: Comparison of classification results of remote sensing images.

Through comparison, we can find that there are more misclassifications of reeds. The main reason is that reeds are scattered sporadically and staggered with other vegetations, and the boundary characteristics between different vegetations are not obvious. This affects the U-Net network to accurately extract the boundaries of the reeds. The Att-Unet network joins the attention mechanism to avoid such problems to a certain extent. Rapeseed flowers are widely distributed and formed into patches, which are obviously different from other vegetations, so there are very few mistakes. However, because the image characteristics of early rape plants are similar to those of reeds, the boundary information is more blurred. In the U-Net network classification model, the pixels are independent of each other, leading to inconsistent classification results of some adjacent pixels, and the rape plants are mistakenly classified as reeds. However, after the fully connected CRFs are processed, it can effectively overcome the effects of different spectra or foreign objects of the same spectrum, make up for the defects of pixel-based classification, and improve classification accuracy.

4.5. Comparison with Other Methods. The proposed method improves U-Net network by introducing attention mechanism and ACB convolution block. The improved Att-Unet model detects changes in the ecological garden landscape in remote sensing images. The training time comparison results of different methods are shown in Table 2.

It can be seen from Table 2 that [13] uses the minimum distance method to classify remote sensing images. The method is more traditional and the calculation is simple, so the overall training time is 28 minutes. Reference [16] proposed a maximum likelihood classification method for remote sensing images combined with deep learning network. Reference [17] fuses deep learning and multisource remote sensing data to propose a new dual-channel densely connected convolutional network for automatic classification of remote sensing images. The network scale of the two methods is relatively large, and the parameter update takes a long time. The proposed method uses the Att-Unet model, which introduces an attention mechanism in the U-Net model. The feature weight extraction and update in the attention mechanism increase the amount of model parameters, thereby increasing the model training time to 42 min.

Table 2: Comparison of training time of different methods.

Method	Training time/min
Ref. [13]	28
Ref. [16]	59
Ref. [17]	71
Proposed method	42

In order to demonstrate the performance of the proposed method, it is compared with [13], [16], and [17]. The results of each evaluation index are shown in Table 3.

It can be seen from Table 3 that, compared with other methods, the proposed method has the best classification accuracy. The recall rate, precision, F1 value, and accuracy rate are 0.854, 0.801, 0.836, and 0.982, respectively. Because the proposed method adopts the Att-Unet network model, which introduces the attention mechanism and fully

Table 3: Comparison of experimental results of different methods.

Method	Ref. [13]	Ref. [16]	Ref. [17]	Proposed method
Recall	0.758	0.791	0.829	0.854
Precision	0.725	0.763	0.786	0.801
F1	0.698	0.760	0.805	0.836
Accuracy	0.894	0.939	0.961	0.982
Testing time/s	7.6	11.7	12.5	8.9

connected CRFs, it can better extract small target landscapes from remote sensing images of ecological gardens and achieve more detailed classification. And it uses the ACB convolution block to replace the traditional convolution structure, which simplifies the network model and can speed up the classification to a certain extent. Therefore, the test time is 8.9s, and the overall performance is relatively ideal. Reference [13] uses the minimum distance method for remote sensing impact classification, which is simple and easy to implement, and the test time is only 7.6s. But the classification accuracy is lower than 0.9, and the overall performance is poor. Reference [16] extracts remote sensing image targets through a deep learning network and uses the maximum likelihood method to classify remote sensing images. However, the classification effect of similar remote sensing images is not good, and its F1 value is 0.760, which is 0.076 lower than the proposed method. Reference [17] proposed a new dual-channel densely connected convolutional network for automatic classification of remote sensing images. Among them, the dual-channel densely connected convolutional network is used for feature extraction, and hyperspectral and radar features are merged to output accurate classification results. The model is complex, and the test time is up to 12.5s. However, the classification accuracy has been improved, which is only 0.015 lower than the proposed method.

5. Conclusion

The remote sensing image records the detailed shape, geometric structure, texture, and other characteristic information of the ground object. While providing high-quality information, it also poses new challenges for efficient and accurate remote sensing image classification. For this reason, a classification method for remote sensing images of ecological garden landscape using an improved U-Net model is proposed. Among them, an asymmetric convolution block and an attention mechanism are introduced to improve the U-Net model. And the improved Att-Unet model is used for remote sensing image classification of ecological garden landscape. At the same time, fully connected CRFs are used for classification post-processing to achieve more refined remote sensing image classification. Experiments demonstrate that the classification results of the proposed method are clearer, especially for small landscape targets. And the recall rate, precision, F1 value, and accuracy rate obtained are 0.854, 0.801, 0.836, and 0.982, respectively. The classification test time is 8.9s, and the overall performance is better than other comparison methods.

It has obvious advantages in dynamic monitoring of ecological garden landscape. However, the improved model network framework is larger and the number of parameters increases, which leads to a longer model training time. Therefore, follow-up research should be carried out in the direction of further improving the accuracy of the model and accelerating the speed of model training.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

References

[1] W. Jing, Q. Ren, J. Zhou, and H. Song, "AutoRSISC: automatic design of neural architecture for remote sensing image scene classification," *Pattern Recognition Letters*, vol. 140, no. 3, pp. 186–192, 2020.

[2] H. T. A. El-Hamid, W. Wang, and Q. Li, "Landscape evaluation based on gaofen satellite in the southern part of the nile delta, Egypt," *Journal of Geoscience and Environment Protection*, vol. 7, no. 7, pp. 47–60, 2019.

[3] P. Dou, H. Shen, Z. Li, X. Guan, and W. Huang, "Remote sensing image classification using deep–shallow learning," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, no. 8, pp. 3070–3083, 2021.

[4] Y. Boualleg, M. Farah, and I. R. Farah, "Remote sensing scene classification using convolutional features and deep forest classifier," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 12, pp. 1944–1948, 2019.

[5] J. M. Liu and M. H. Yang, "Deep learning-based classification of remote sensing image," *Journal of Computers*, vol. 13, no. 1, pp. 44–48, 2018.

[6] L. Wang, P. Marzahn, M. Bernier, and R. Ludwig, "Mapping permafrost landscape features using object-based image classification of multi-temporal SAR images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 141, no. 07, pp. 10–29, 2018.

[7] 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, pp. 215943–215955, 2020.

[8] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F Pla, "Deep pyramidal residual networks for spectral–spatial hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 740–754, 2019.

[9] Y. Gao, J. Shi, J. Li, and R. Wang, "Remote sensing scene classification based on high-order graph convolutional network," *European Journal of Remote Sensing*, vol. 54, no. sup1, pp. 141–155, 2021.

[10] S. Wang, Y. Guan, and L. Shao, "Multi-granularity canonical appearance pooling for remote sensing scene classification," *IEEE Transactions on Image Processing*, vol. 29, no. 4, pp. 5396–5407, 2020.

[11] S. Song, H. Yu, Z. Miao, Q. Zhang, Y. Lin, and S. Wang, "Domain adaptation for convolutional neural networks-based

remote sensing scene classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 8, pp. 1324–1328, 2019.

[12] Y. Yuan, J. Fang, X. Q. Lu, and Y. Feng, "Remote sensing image scene classification using rearranged local features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1779–1792, 2019.

[13] U. L. Dano, M. Alhefnawi, and F. Al-Shihri, "Assessing the accuracy of image classification algorithms using during-flood TerraSAR-X imagery," *Disaster Advances*, vol. 13, no. 8, pp. 23–33, 2020.

[14] S. Hu, S. Fong, L. Yang et al., "Fast and accurate terrain image classification for ASTER remote sensing by data stream mining and evolutionary-EAC instance-learning-based algorithm," *Remote Sensing*, vol. 13, no. 6, p. 1123, 2021.

[15] H. Gao, M. Wang, Y. Yang, X. Cao, and C. Li, "Hyperspectral image classification with dual attention dense residual network," *International Journal of Remote Sensing*, vol. 42, no. 15, pp. 5604–5625, 2021.

[16] S. J. Liang, J. Cheng, and J. W. 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.

[17] H. Song, S. Dai, and H. Yuan, "Multi-source remote sensing image classification based on two-channel densely connected convolutional networks," *Mathematical Biosciences and Engineering*, vol. 17, no. 6, pp. 7353–7377, 2020.

[18] C. Zhang, G. Li, and S. Du, "Multi-scale dense networks for hyperspectral remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9201–9222, 2019.

[19] F. Zhao, X. Zhang, X. Mu, Z. Yi, and Z. Yang, "Learning multimodality features for scene classification of high-resolution remote sensing images," *Journal of Computer and Communications*, vol. 06, no. 11, pp. 185–193, 2018.

[20] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 5, pp. 2811–2821, 2018.

[21] H. Dong, L. Zhang, and B. Zou, "PolSAR image classification with lightweight 3D convolutional networks," *Remote Sensing*, vol. 12, no. 3, p. 396, 2020.

[22] S. Pan, H. Guan, Y. Chen et al., "Land-cover classification of multispectral LiDAR data using CNN with optimized hyperparameters," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, no. 9, pp. 241–254, 2020.

[23] J. Guo, L. Wang, D. Zhu, C. Y. Hu, and C. Y. Xue, "Subspace learning network: an efficient ConvNet for PolSAR image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 12, pp. 1849–1853, 2019.

[24] C. Cho, Y. H. Lee, J. Park, and S. Lee, "A self-spatial adaptive weighting based U-net for image segmentation," *Electronics*, vol. 10, no. 3, p. 348, 2021.

[25] X. Bai, R. Xue, L. Wang, and F. Zhou, "Sequence SAR image classification based on bidirectional convolution-recurrent network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9223–9235, 2019.

[26] S. S. Heydari and G. Mountrakis, "Meta-analysis of deep neural networks in remote sensing: a comparative study of mono-temporal classification to support vector machines," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, no. 6, pp. 192–210, 2019.