Hindawi
Journal of Sensors
Volume 2022, Article ID 8240900, 10 pages
https://doi.org/10.1155/2022/8240900

Research Article

Application Research of Deep Learning Technology in Natural Landscape Animation Design

Lili Xu and Lilei Wen

School of design art, Xijing University, Xi'an, Shaanxi Province, China 710123

Correspondence should be addressed to Lili Xu; 20130104@xijing.edu.cn

Received 12 January 2022; Revised 13 February 2022; Accepted 19 February 2022; Published 19 April 2022

Academic Editor: Wen Zeng

Due to the limitation of technology and cost, the animation design of natural landscape in the past was often dealt with relative simplicity. With the increasing level of audience appreciation and the continuous development of animation technology, new requirements are put forward for the design of natural landscape animation. In order to make the animation design effect of natural landscape more real, the synthetic aperture radar image is firstly analyzed to obtain the location of mountains, farmland, rivers, villages, roads, and buildings. Considering the superiority of U-Net network in image semantic segmentation, this paper constructs a semantic segmentation model based on U-Net structure. In this model, dense connection module is introduced in downsampling, and spatial void pyramid structure is introduced in upsampling to retain more image features, to achieve accurate segmentation of satellite images. Experimental results show that the proposed algorithm has higher segmentation accuracy than other algorithms. After accurate classification of natural scene images, it can provide a guarantee for designing more real natural landscape animation design effects.

1. Introduction

Most of the landforms in animated films about natural landscapes are fond of mixing mountains, hills, forests, and rivers. On the one hand, it makes artists develop their imagination better, and on the other hand, it also enriches the visual experience of the audience. However, plain terrain is not much. When the art design is carried out to the production level, the actual terrain proportion in the natural environment should be considered first. At the same time, paying attention to the proportion of plants to characters and scenes, each plant has its own volume. We should not only consider the beauty of individual plants but also focus on the unity and harmony of the group effect. Too characteristic monomer design put together may not be able to achieve beautiful group effect. In the production of the environment, the vegetation is programmed to be copied and arranged in the scene. Dense plants are more likely to produce visual repetition [1]. If the vegetation in the animated film is too inconsistent with real life, it will produce a "sense of repulsion" in the vision and reduce the beauty of the picture. Therefore, how to carry out animation design according to the proportion of real terrain environment has important research significance.

The elements of landscape environment design mainly include the following five parts. The first is the terrain. According to the terrain scale, it can be divided into large terrain, small terrain, and micro terrain. Among them, plains, hills, mountains, and other terrain with a larger area of land are the large terrain. Landforms with small geographical areas such as ramps, tablelands, and flat lands are small landforms. The terrain with small fluctuation, such as grassland and sand dune, belongs to microtopography. Water body is one of the important elements in landscape design. According to its morphology, it can be divided into moving water and still water. Rivers, waterfalls, and streams are moving water, while pools are still water. The change of moving water and standing water in form makes the design of landscape environment also have infinite change. The third is plants; plants are full of vitality, which has a high ornamental value. Plant morphology can change with climate and environment. By making full use of plants, the

design effect of landscape environment can be greatly enhanced. The fourth is the sky scene, that is, the atmospheric environment. Atmospheric environment is changeable; day and night changes and seasonal changes belong to the sky scene. The sky scene endows the landscape environment with the beauty of time and space interaction. The fifth is the landscape facilities; the artificial facilities that people need in the process of leisure, life, and entertainment are all landscape facilities [2]. Landscape facilities can effectively meet people's needs, more importantly through the design of the landscape.

In the real world, the main factors affecting vegetation are the geographical location and climate of growth. Vegetation design should be created on the premise of respecting the objective reality. Common plants include trees, shrubs, ferns, grasses, weeds, and mosses. The more detailed the plant species are planned, the richer and more vivid the natural environment of the art design will be. There are too many plants in nature. As an important part of the rich picture, the stone and earth blocks of different shapes and sizes, broken branches, and fallen leaves are indispensable. They will be scattered randomly on the ground to make the environment more vivid. If artists can classify common and representative plant events according to botanical theory, it would be very worthwhile to use it as a reference for art design. In order to carry out animation design according to the proportion of real terrain, the natural environment can be classified according to image classification and image segmentation technology and then according to the classified data, to make animation design and to achieve high-quality animation design effect.

In recent years, deep learning has made great achievements in computer vision, image classification, and image segmentation. The special network structure of deep learning can transfer extracted feature values through neurons, and each layer can extract and learn the features transferred from the previous layer to continue transmission, to extract the optimal feature values [3]. Methods in the field of image semantic segmentation include AlexNet, FCN, U-Net, SegNet, and ResNet. AlexNet has achieved successful application of ReLU, Dropout, and local response normalization (LRN) in convolutional neural networks, etc. [4]. FCN changed AlexNet and VGG full connection layer to convolution layer [5]. FCN downsampling carries out feature extraction for the image and deconvolution for upsampling. This ensures that the output image is the same size as the input image. Based on FCN network architecture, literature [6] proposed to learn a multilayer deconvolution network to replace simple bilinear interpolation. Literature [7] proposed that SegNet is based on encoder-decoder architecture. The convolution layer and pooling layer constitute the encoder, while the convolution layer and upsampling layer constitute the decoder. The function of the encoder is to extract the feature image, and the function of the decoder is to return the feature image to the same size as the input image.

The similarity between SegNet and deconvolution network is that the network structure is similar. The difference is that SegNet removes the two fully connected layers in the middle of the network and uses the batch normalization method and Softmax classifier. The advantages of SegNet are high efficiency and low memory consumption, while the disadvantages are low accuracy. Pooling layer is introduced in FCN and SegNet networks. The advantage of this method is that the image size is reduced while the receptive field is increased, but the disadvantage is that part of the position information is lost. Literature [8] designs a network dedicated to image pixel prediction. The network does not contain pooling layer, and the convolution layer adopts extended convolution. The advantage of this network is that the extended convolution increases the receptive field of the convolution kernel. It can fuse the multiscale context information of captured image and improve the accuracy of pixel prediction.

The advantage of ResNet residual block model is to reduce the gradient disappearance problem caused by the increase of neural network depth. It uses the cascade operation between encoder and decoder to fuse the high-level information with the shallow level information. Its advantage is to avoid the loss of high-level semantic information and preserve image features as much as possible. Deep learning can realize object segmentation and extraction by computer. It learns features of lower and higher levels through special network models and has higher learning efficiency [9]. Deep learning segmentation of image feature elements can achieve better experimental results through many data experiments. Therefore, deep learning algorithm can be used to segment and extract feature images from satellite images.

Aiming at the problems of confusion and unclear recognition in segmentation, this paper proposes an image segmentation algorithm based on deep learning based on U-Net. The innovations and contributions of this paper are listed below. (1) In order to improve the accuracy of image segmentation, the specific part of image is input by attentional supervision mechanism during the fusion of image feature information. (2) After accurate classification of natural scene images, it can provide a guarantee for designing more real natural landscape animation design effects.

The structure of this paper is listed as follows. The related theories are described in the next section. The proposed method is expressed in Section 3. Section 4 focuses on the experiment and analysis. Section 5 is the conclusion.

2. Related Theories

2.1. Early Remote Sensing Image Scene Classification Method. Before the rise of deep learning, scene classification of high-resolution remote sensing images was based on manual features. Among them, there are color histogram, scale-invariant feature transformation, universal search tree, and other classic manual features. However, manual feature design needs a lot of prior knowledge and is time-consuming and laborious, and the effect is poor. In order to obtain a higher accuracy of scene classification, manual coding features appear later. The main idea of this method is to further abstract the image based on manual features. The most typical feature of manual coding is the visual word bag model [10]. Although hand-coded features can improve classification accuracy, it is limited by the upper limit of

underlying features. Therefore, there are obvious shortcomings of weak generalization ability and low classification accuracy when only using low-level features in scene classification tasks of high-resolution remote sensing images.

2.2. Deep Neural Network. The concept of neural networks was inspired by the 1943 model of artificial neurons. Then, the perceptron algorithm is proposed and the MCP model is used to successfully classify multidimensional data. However, subsequent experiments show that this model can only deal with linear classification problems. Until 1986, Hinton, the father of neural network, invented back propagation (BP) algorithm. It uses Sigmoid to carry out nonlinear mapping, so the nonlinear classification problem is solved. However, the neural network is still faced with problems such as gradient disappearance, time-consuming training network, and difficulty in local optimization.

A solution to the problem of gradient disappearance is proposed in deep network training. First, the unsupervised method is used to pretrain the network, so that the network weight has a good initial value. Then, the network is optimized in a more detailed and supervised way to further improve the network performance. Subsequently, ReLU activation function, AlexNet [11], and a series of new technologies and network architectures were proposed. It makes deep neural network really receive wide attention.

Deep neural networks can be roughly divided into two categories: one is deep neural network (DNN) with one-dimensional vector input; the other is the input of two-dimensional image or three-channel color image DNN. The former is represented by deep belief network (DBN), while the latter is represented by convolutional neural network (CNN).

2.3. Research Status of Deep Learning in Remote Sensing Image Classification. The scene classification methods of high-resolution remote sensing images based on deep learning can be divided into three categories according to the supervision methods: full supervision method, semisupervision method, and weak supervision method.

2.3.1. Fully Supervised Classification Methods. Fully supervised learning, also known as supervised learning, is a method of learning data and its corresponding labels and then used for network training. At present, most of the high-resolution remote sensing scene classification methods based on deep learning can be classified as full supervision. Subject-based model is an effective method. Literature [12] proposed an adaptive deep sparse semantic model. It makes full use of the multilevel semantics of remote sensing image scenes. At the semantic level, sparse thematic features and deep features are effectively integrated, which effectively improves the ability of feature representation and thus achieves a higher level of classification.

It is also a common method to improve the scene classification accuracy of remote sensing images by fusing multilayer deep features. Literature [13] realized that most existing CNN methods only use feature vectors of the last fully connected layer for scene classification, which ignores local information of images. Although some images have similar global characteristics, they belong to different categories. The reason is that the category of the image may be highly correlated with local features rather than global features. Therefore, the features of the last convolutional layer and the last fully connected layer of the deep neural network are firstly extracted as local features and global features, respectively. Then, the clustering method is used to cluster the global features into multiple sets. Then, the local features are rearranged according to the similarity between the local features and the cluster center. Finally, the global and local remote sensing image features can be obtained through the fusion of the two.

2.3.2. Semisupervised Classification Method. Semisupervised learning can make use of many unlabelled samples, so the need for label samples is reduced, which to some extent solves the problem of insufficient label samples in the field of deep learning. From the perspective of enlarging tag sample size, literature [14] proposed a generation framework based on semisupervised deep learning features. The framework can be trained to automatically expand the number of label samples. Firstly, the tagged samples were used to fine-tune the pretrained CNN, and then, the deep features extracted from the fine-tuned CNN were used to train the support vector machine (SVM). At the same time, the method combines multiple support vector machines to identify easily confused category samples, which effectively improves the labelling accuracy and the number of label samples. Therefore, it can effectively improve the generalization ability and classification accuracy of the network.

2.3.3. Weakly Supervised Classification Methods. A combination of weak supervision and deep learning is also widely used. In literature [15], features extracted from labelled images are taken as the source domain, while features extracted from unlabelled images are taken as the target domain. Then, it is used for network training and optimization of specified loss functions to classify labelled and unlabelled data.

2.3.4. Qualitative Comparison of Supervision Methods. The classification method based on total supervision has remarkable effect and high accuracy. However, all the above monitoring methods require many labelled samples to train the classification network, and labelled samples are often difficult to obtain. It takes a lot of time and energy to label unlabelled images, which limits the further development of full supervision methods. The semisupervised classification method can train the network with many unlabelled samples so that the network can obtain more "extra" information, thus improving the robustness of the network. However, only unlabelled samples can be used to refine the feature space constructed by labelled samples, without significantly increasing the discriminant information, thus limiting the classification accuracy.

3. Image Segmentation Algorithm Based on U-Net Structure

Considering the superiority of U-Net network in image semantic segmentation and dense connection network, this

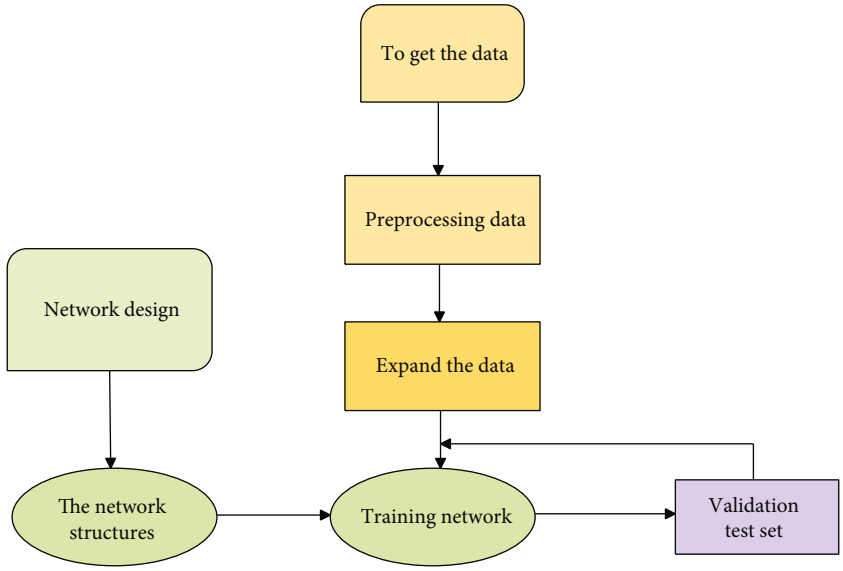


Figure 1: Flow of the proposed algorithm.

paper constructs a semantic segmentation model based on U-Net structure. Its network structure consists of encoder, decoder, dense connection module, ASPP, and CBAM.

(1) Encoder includes dense connection module and maximum pooling layer. The dense connection module extracts the semantic features of the image through the convolution layer, and the maximum pooling layer performs downsampling operation on the feature information of the image

(2) Dense connection module reuses the image features of the previous layer. ASPP is introduced to increase the receptive field of image feature information and improve the robustness of image feature. Introducing CBAM to conduct attention supervision when learning deep feature information can effectively extract feature information of elements

(3) Decoder part includes dense connection module and deconvolution layer. The deconvolution layer upsamples the image feature information so that the size of the input image and the output image remain unchanged

3.1. Algorithm Flow. The algorithm implementation process includes network design, network construction, data acquisition, data preprocessing, and data expansion. The algorithm flow is shown in Figure 1.

3.2. Network Structure. Based on the original U-Net network, the dense connection module, ASPP, and CBAM modules are introduced in this paper. Its network structure is shown in Figure 2.

The input image size is 480×480. The dense connection module includes two convolution and two feature fusion, and the image size does not change to 480×480. After the maximum pooling layer, the image size is 1/2 of the original, that is, 240×240. Record the dense connection module and the maximum pooling operation as one operation unit. Then, the image size becomes a feature map of 30×30 after three times of operation. ASPP and CBAM are introduced before deconvolution (upsampling) of feature images. ASPP can fuse deeper image details. CBAM does not affect the size of the feature graph. Network learning integrates the feature graph and weight graph of channel and spatial attention model by dot product and then inputs the fused result graph into the deconvolution layer (upsampling). Deconvolution is performed during upsampling, and the image size is doubled, i.e., 60×60. After intensive connection module, the image size is still 60×60. Finally, the image size was restored to 480×480 after three operations.

3.2.1. Dense Connection. In deep learning networks, the problem of gradient disappearance becomes more and more obvious with the deepening of network depth. In this paper, the dense connection module is introduced by referring to the concept of dense connection in DenseNet. All layers of the network are connected while ensuring maximum information transmission between layers. In order to ensure the feedforward characteristics, each layer splices the input of all previous layers and then transmits the output feature graph to all subsequent layers.

The operation process is as follows. The input image size is 480×480. The dense connection module includes two convolution and two feature fusion. The convolution kernel is 3 by 3. The step size is 2. The number of convolution kernels is 64. After four intensive connection modules and pooling operations, a feature map with a size of 30×30 was obtained.

Batch normalization (BN) layer is added between each convolutional layer and activation function, which normalizes data from each batch during each random gradient decline. As a result, the mean value and variance of data from each channel in the output feature graph are 0 and 1,

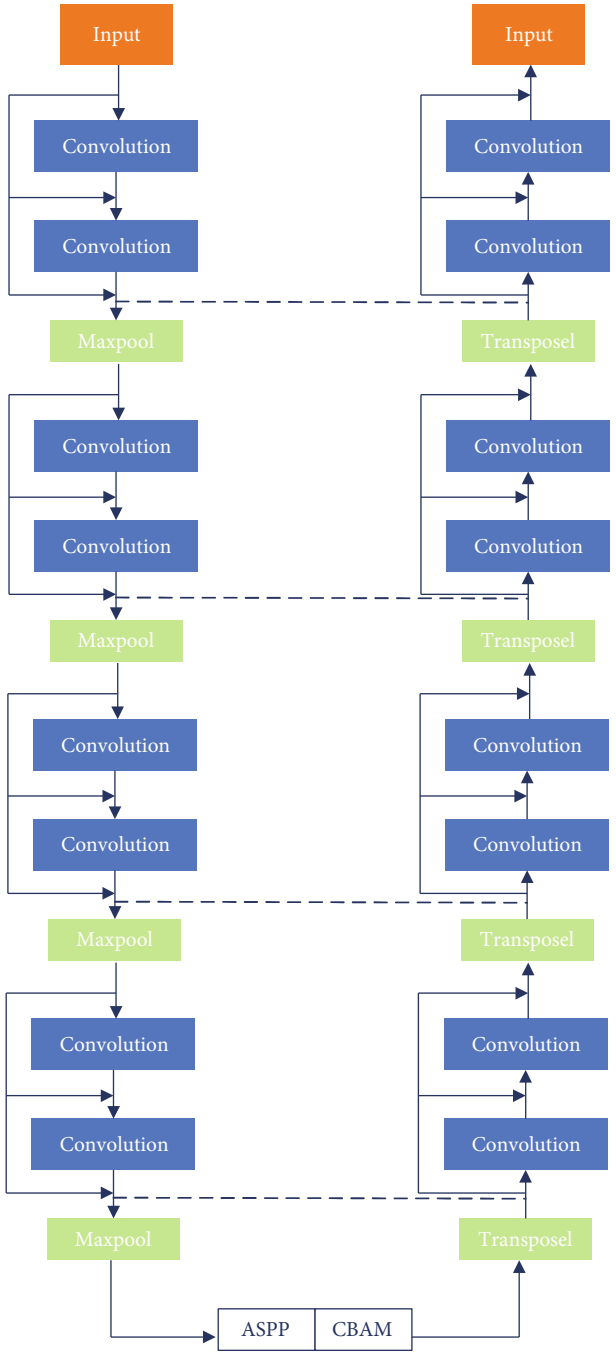


Figure 2: The network structure of the proposed model.

which reduces the gradient disappearance during network learning. ReLU activation layer is introduced to activate features.

3.2.2. Pyramid Structure of Void Space. The structure of aurous spatial pyramid pooling (ASPP) is introduced before upsampling. The basic element of the pyramid structure of void space is the convolution of void with different expansion rates. The advantages of empty convolution are as follows: (1) Without increasing parameters, the receptive field of convolution kernel is increased. (2) Under the condition of no loss of information in downsampling, the convolution output range of spatial information is larger, so that the network retains more image features. The convolution kernel of the empty convolution in this paper is 3×3. The empty convolution with expansion rates of 2, 4, 8, and 16 replaces the original ordinary convolution. This makes the model segmentation image clearer, as shown in Figure 3.

ASPP performs six convolution operations on the input feature graph. Convolution kernel size of the first convolution is 1×1, and that of the second to the fifth is 3×3. The empty convolution with expansion rates of 2, 4, 8, and 16 replaced the original ordinary convolution to obtain multiscale characteristic information. The average pooling method is introduced into the model. After global mean pooling of the input feature graph, the feature graph is fed into the convolution kernel with a size of 1×1. Use BN operation and upsample to image original size. Then, the feature images fused with six multiscales are sent into the convolution layer with a size of 1×1. Finally, the output feature map is fed into the attentional mechanism model.

3.2.3. Attention Mechanism Module. CBAM is introduced after ASPP and before upsampling. CBAM module (including channel attention and spatial attention two submodules) is shown in Figure 4.

After the input feature image passes through the channel attention submodule and the spatial attention submodule, the feature image multiplied by the output results of the two submodules is sent to the coding stage for upsampling. CBAM can effectively extract feature information of elements by attentional supervision when learning deeper feature information through space and channel. The realization process is as follows.

$$F' = W_c(F)F \otimes F'' = W_s\left(F'\right)F', \tag{1}$$

where the characteristic graph of F, after being operated by channel attention module and spatial attention module, is F' and F'', respectively. $\otimes$ means multiply elements by elements.

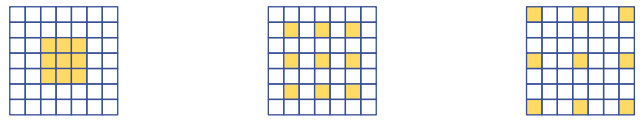
The realization process of channel attention module is as follows. Global average pooling (AvgPool) and global maximum pooling (MaxPool) were performed on input feature graph F, and the results were A1 and B1, respectively. Then, feature elements g and h were obtained by adding g_1 and h_1 through multilayer perceptron (MLP) to obtain C_1. The feature of channel attention is the result of fusion of C_1 and F. The operation process is as follows.

$$W_c\left(F'\right) = (\mathrm{MLP}(\mathrm{AvgPool}(F) + \mathrm{ML\,P}(\mathrm{MaxPool}))), \tag{2}$$

$$W_c\left(\mathrm{F}'\right) = (M_1(M_0(g_1)) + M_1(M_0(h_1)))W_c\left(F'\right) = (G + H), \tag{3}$$

$$W_c\left(F'\right) = (C_1), \tag{4}$$

where M_0 and M_1 are the two-layer parameters of MLP.



(a) The expansion rate is 1 (b) The expansion rate is 2 (c) The expansion rate is 3

Figure 3: The structure of ASPP.

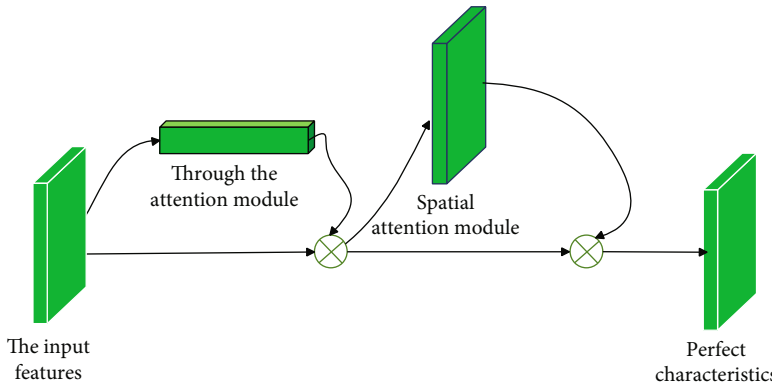


Figure 4: Overall structure of attention model.

Sigma is the sigmod activation function. M_0 needs to be activated using the ReLU function.

The realization process of spatial attention module is as follows: the results of average pooling and maximum pooling of input feature maps are g_2 and h_2. Then, g_2 and h_2 are fused for 3×3 convolution operation to obtain feature graph C_2. The spatial attention feature is the result of C_2 and F fusion. The operation is as follows.

$$W_s(F) = (f([\text{AvgPool}(F); \text{MaxPool}(F)])), \quad (5)$$

$$W_s(F) = (f(g_2; h_2)), \quad (6)$$

$$W_s(F) = (C_2), \quad (7)$$

where F represents the 3×3 convolution operation.

3.3. Coordinate Point Data. In this paper, the obtained satellite image information is optimized and the images without elements are eliminated. First, OpenStreetMap captures coordinates of the image based on latitude and longitude. Then, the corresponding position in Mapbox is located according to the endpoint coordinate data, and Labelme is used to annotate the data. In order to solve the problems of different image sizes, unclear image content, large amount of data, and slow training speed, the captured image was preprocessed to adjust the size to 480×480 to improve the training speed of the model. In this paper, random flipping, image noise, image brightness, and other methods are used to enhance the data, to avoid the overfitting phenomenon of the network, and to achieve better training of the network model.

4. Experimental Results and Analysis

4.1. Experimental Data. Two real polarimetric SAR data are used to verify the algorithm. The first data is the data image

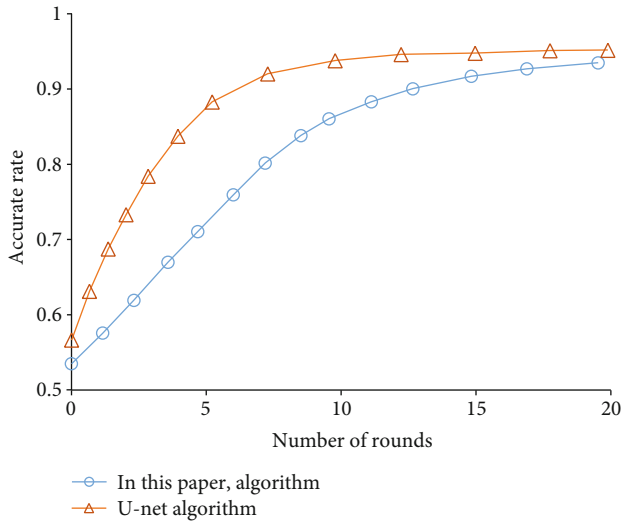


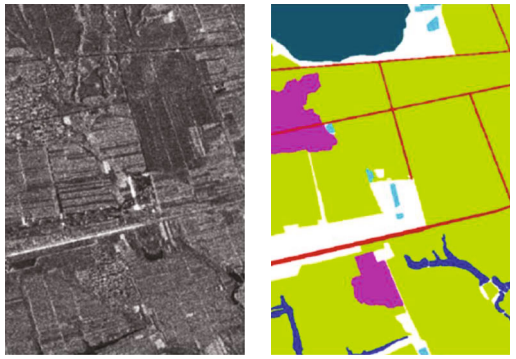
Figure 5: The change of accuracy rate in network training stage.

obtained by RADARSAT-2. The second data is a NASA/JPLAIRSAR image of the San Francisco area. The image contains five types of ground objects, namely, mountains, farmland, rivers, villages, roads, and buildings.

4.2. Performance Specifications. Accuracy, recall, and precision were used to evaluate the segmentation effect of the proposed algorithm.

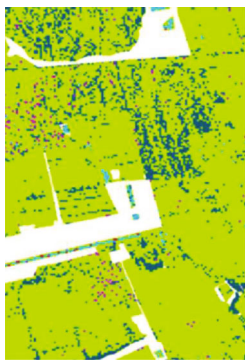
$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (8)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (9)$$

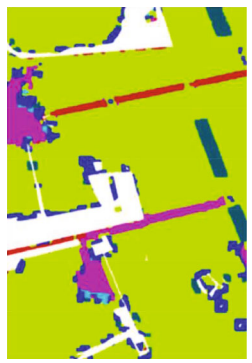


(a) Original

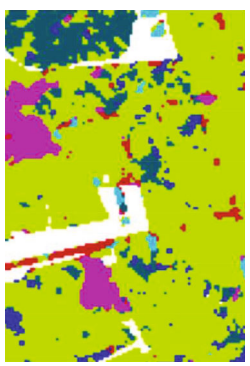
(b) Ground truth



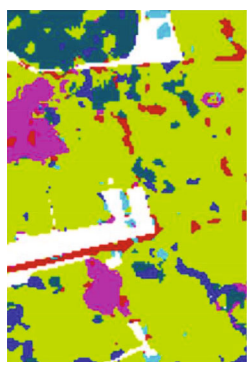
(c) SVN



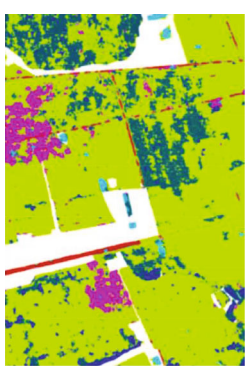
(d) Wishart



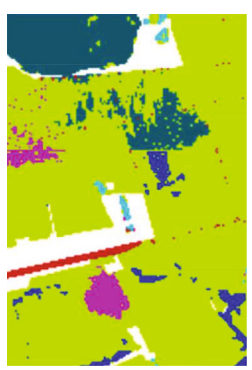
(e) Algorithm [16]



(f) Algorithm [17]



(g) CNN



(h) Algorithm [18]

Figure 6: Continued.

(i) Proposed

Figure 6: The classification results.

Table 1: Classification accuracy on RADARSAT-2 image.

Algorithm	Mountain	Farmland	River	Village	Road	Building
SVM	97.71	97.06	78.83	63.44	90.8	93.71
Wishart	51.64	95.15	96.17	94.97	95.02	87.8
Algorithm [16]	99.6	98.05	96.35	91.19	94.33	89.6
Algorithm [17]	99.79	99.38	98.57	96.6	98.03	96.34
CNN	98.77	96.79	88.78	92.47	89.63	93.84
Algorithm [18]	92.52	96.12	92.9	94.03	94.06	86.33
Proposed	99.89	99.57	99.01	98.96	99.65	98.7

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{10}$$

where TP is a true case, TN is a true negative case, FP is a false positive case, and FN is a false negative case.

The accuracy analysis results of U-Net and the algorithm presented in this paper in the training process are shown in Figure 5.

As can be seen from Figure 5, after 20 iterations in the training process, the accuracy of the algorithm in this paper is always higher than that of U-Net. This proves the feasibility of the proposed algorithm for image segmentation.

4.3. Comparison and Analysis of Algorithms

4.3.1. Experimental Results of RADARSAT-2 Data. The proposed algorithm was compared with other 6 typical algorithms, including SVM algorithm, Wishart algorithm, algorithm [16], algorithm [17], CNN, and algorithm [18]. Figure 6 and Table 1 are the classification results and classification accuracy values.

As can be seen from Figure 6(c), SVM algorithm has serious misclassification, especially in the upper and left part of the image. Figure 6(d) is the classification result of the Wishart algorithm, in which bare land and water are seriously confused. In addition, the classification confusion of farmland and river also exists. Figures 6(e) and 6(f) are the classification results of literature [16] and literature [17] algorithms. There are many misclassified pixels in both class plots. At the same time, there is the problem of internal pixel discontinuity in the image.

Figure 6(g) shows the classification results of CNN algorithm. The results are clearer than those of the previous algorithms. However, the category of farmland excessively affects the classification of the whole image, and many pixels that do not belong to the category of farmland are classified as the category of farmland. Figure 6(h) shows the classification results of literature [18] algorithm, and the classification confusion of village and road categories is serious. Figure 6(i) shows the classification results of the proposed algorithm, and its classification performance is greatly improved compared with other algorithms. The classification image of the proposed algorithm is cleaner and has better spatial connectivity.

As shown in Table 1, the overall classification accuracy of the proposed algorithm is higher than that of other algorithms. The classification accuracy of the Wishart algorithm is 1.25% lower than that of the algorithm in this paper. However, it should be noted that the algorithm uses 5% of real marker pixels as training samples, while the algorithm in this paper only uses 1% of real marker pixels.

4.3.2. Experimental Results of San Francisco Area Data. Table 2 shows the classification results and classification accuracy values of data in San Francisco area by SVM algorithm, Wishart algorithm, this paper algorithm, and CNN algorithm, respectively. Table 2 shows that the overall

Table 2: Classification accuracy on San Francisco image.

Algorithm	Mountain	Farmland	River	Village	Road	Building
SVM	85.63	57.56	24.5	59.33	12.56	49.97
Wishart	93.83	88.11	45.17	39	59.11	66.32
CNN	97.1	83.56	78.11	82.33	87.67	89.45
Proposed	98.89	91.2	95.58	98.11	96.56	96.76

classification accuracy of the proposed algorithm is higher than that of other algorithms.

5. Conclusion

Powered by the continuous development of human civilization, landscape design has become an important issue for more and more professionals. Animation technology also plays an increasingly important role in landscape design. In order to realize that the animation design effect of natural landscape is close to the real effect, the synthetic aperture radar image is firstly analyzed to obtain the location of mountains, farmland, rivers, villages, roads, and buildings. In this paper, we design an improved U-Net network structure, which can induce dense connection modules in downsampling. Before upsampling, ASPP and CBAM are introduced to segment satellite image road elements accurately. Experimental results show that the segmentation accuracy of the proposed algorithm is higher than that of other comparison algorithms. The feasibility of the proposed algorithm in landscape design is verified. Although through the algorithm in this paper, the location of mountains, farmland, rivers, villages, roads, and buildings in natural scenes can be obtained. However, the actual scene environment will be more complex, and how to achieve high-precision image classification in a more complex environment is the follow-up research.

Data Availability

The labelled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no competing interests.

Acknowledgments

This work is supported by the Xijing University.

References

[1] A. Chirico, F. Ferrise, L. Cordella, and A. Gaggioli, "Designing awe in virtual reality: an experimental study," *Frontiers in Psychology*, vol. 8, p. 2351, 2018.

[2] X. Ziya, G. Xiaru, L. Haohao, L. Naixin, and F. Yanhua, "Landscape environment design of long shadow historical section," *Chinese & Overseas Architecture*, vol. 7, 2018.

[3] X. Li, W. Zhang, and Q. Ding, "Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction," *Reliability Engineering & System Safety*, vol. 182, pp. 208–218, 2019.

[4] K. Lee, S. H. Sung, D.-h. Kim, and S.-h. Park, "Verification of normalization effects through comparison of CNN models," in *2019 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, pp. 1–5, Ho Chi Minh City, Vietnam, 2019.

[5] J. Zhang, X. Hu, Z. Ning et al., "Energy-latency tradeoff for energy-aware offloading in mobile edge computing networks," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2633–2645, 2018.

[6] S. Zhao, M. Hu, Z. Cai, Z. Zhang, T. Zhou, and F. Liu, "Enhancing Chinese character representation with lattice-aligned attention," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2021.

[7] A. Mittal, R. Hooda, and S. Sofat, "LF-SegNet: a fully convolutional encoder–decoder network for segmenting lung fields from chest radiographs," *Wireless Personal Communications*, vol. 101, no. 1, pp. 511–529, 2018.

[8] Y. Zhang and A. Jatowt, *Image Tweet Popularity Prediction with Convolutional Neural Network*, Springer, Cham, 2019.

[9] Q. Zhang, Z. Cui, X. Niu, S. Geng, and Y. Qiao, *Image Segmentation with Pyramid Dilated Convolution Based on ResNet and U-Net*, Springer, Cham, 2017.

[10] X. Liu, S. Zhang, T. Huang, and Q. Tian, "E^2BoWs: an end-to-end bag-of-words model via deep convolutional neural network for image retrieval," *Neurocomputing*, vol. 395, pp. 188–198, 2020.

[11] S. Sengan, L. Arokia Jesu Prabhu, V. Ramachandran, V. Priya, L. Ravi, and V. Subramaniyaswamy, "Images super-resolution by optimal deep AlexNet architecture for medical application: a novel DOCALN1," *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 6, pp. 8259–8272, 2020.

[12] Q. Zhu, Y. Zhong, L. Zhang, and D. Li, "Adaptive deep sparse semantic modeling framework for high spatial resolution image scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, pp. 1–16, 2018.

[13] Y. Yuan, J. Fang, X. 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.

[14] Z. Ning, X. Hu, Z. Chen et al., "A cooperative quality-aware service access system for social Internet of vehicles," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2506–2517, 2017.

[15] S. Zhao, M. Hu, Z. Cai, and Z F Liu, "Dynamic modeling cross-modal interactions in two-phase prediction for entity-relation extraction," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[16] X. Huang, C. Liao, M. Xing et al., "A multi-temporal binary-tree classification using polarimetric RADARSAT-2 imagery," *Remote Sensing of Environment*, vol. 235, article 111478, 2019.

[17] Q. Xie, J. Wang, C. Liao et al., "On the use of Neumann decomposition for crop classification using multi-temporal RADARSAT-2 polarimetric SAR data," *Remote Sensing*, vol. 11, no. 7, p. 776, 2019.

[18] H. Xiping, J. Cheng, M. Zhou et al., "Emotion-aware cognitive system in multi-channel cognitive radio ad hoc networks," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 180–187, 2018.