International Journal on Semantic Web and Information Systems
Volume 19 • Issue 1

A Novel Semantic Segmentation Approach Using Improved SegNet and DSC in Remote Sensing Images

Wanjun Chang, Henan Institute of Technology, China*
Dongfang Zhang, Henan Institute of Technology, China

ABSTRACT

An improved SegNet semantic segmentation model is proposed to address the issue of traditional classification algorithms and shallow learning algorithms not being suitable for extracting information from high-resolution remote sensing images. During the research process, space remote sensing images obtained from the GF-1 satellite were used as the data source. In order to improve the operational efficiency of the encoding network, the pooling layer in the encoding network is removed and the ordinary convolutional layer is replaced with a depth-wise separable convolution. By decoding the last layer of the network to obtain the reshaped output results, and then calculating the probability of each classification using a Softmax classifier, the classification of pixels can be achieved. The output result of the classifier is the final result of the remote sensing image semantic segmentation model. The results showed that the proposed algorithm had the highest Kappa coefficient of 0.9531, indicating good classification performance.

KEYWORDS

Deep Learning, Depth-Wise Separable Convolution, Remote Sensing Image, SegNet, Semantic Segmentation

INTRODUCTION

As a vital component of the earth's ecosystem, vegetation serves an irreplaceable role in climate regulation, maintaining environmental viability and ensuring the earth's ecological balance. As countries strengthen the implementation of sustainable development policies, vegetation coverage is gradually becoming a key focus area. The distribution and accurate identification and classification of vegetation are the premises and foundations for studying vegetation coverage.

The automatic analysis of remote sensing (RS) data is a prerequisite for mining information and transforming RS observations into knowledge (Alsmirat et al., 2019; Kumbhojkar & Menon, 2022; Wang et al., 2020). Its main purpose is to establish a unified, compact, and semantic representation of large RS datasets, thereby laying the foundation for subsequent information mining. The automatic

DOI: 10.4018/IJSWIS.332769 *Corresponding Author

analysis of RS datasets mainly includes data expression, retrieval, and understanding (Kumar et al., 2022; Lv et al., 2022; Stergiou et al., 2021). At present, RS is mostly used to obtain the dynamic information and images of vegetation in real time, but the images obtained are largely affected by external factors such as the local geographical environments, so it is difficult to realize accurate classification (Kadri et al., 2022; Kotaridis & Lazaridou, 2021; Kumbhojkar & Menon, 2022; Wenjuan & Shao, 2021). Consequently, the problem of vegetation classification in RS images is a research hotspot (Chopra et al., 2022; Yu et al., 2020).

In order to realize the classification of vegetation, contemporary approaches usually include decision tree classifiers, support vector machines, random forests, convolutional neural networks (CNNs) and other methods (Nhi & Le, 2022; Wei et al., 2019; Xia et al., 2020; Zhang et al., 2021). However, due to the huge amount of dynamic image data obtained through RS and the relatively complex data types in different regions, there is no single vegetation classification method that can be applied in many different geographical environments (Kattenborn et al., 2021; Wen & Chen, 2020; Yang et al., 2021). In image classification, deep learning models can directly extract and match end-to-end features from raw image data (Kadry et al., 2022; Mandle et al., 2022; Wang et al., 2020). Zhang et al. (2018) designed an integrated multi-layer perceptron and CNN classifier by combining decision fusion theory with multi-layer neural networks. This model performed high-resolution RS image classification but could extract the features in high-dimensional images directly. The deep belief network proposed by Goudarzi et al. (2018) learned the input data in an unsupervised manner, while the Softmax activation function was used to fine tune the network parameters, extract image features, and classify the images. However, this method relied too much on experiments to achieve the best accuracy and is difficult to apply in practice. On basis of the generalized gamma depth confidence network, Zhao et al. (2021) proposed a method for synthetic aperture radar image statistical modeling and land cover classification. However, this method included feature extraction and multiple learning processes and had a long inference time. Chen et al. (2018) created a CNN that used spectral space data and proposed an adaptive classification method for hyperspectral images with a spatial window size selection. However, the accuracy of logistic regression of this method was low, leading to low image classification accuracy. Yu et al. (2022) proposed a SegNet-based RS cage segmentation method (VM-SegNet), which achieved high-precision image classification but had a weak generalization ability. Ren et al. (2022) proposed a land semantic segmentation model (AM-PSPNet) that utilized a pyramid scene analysis network (PSPNet) with an attention mechanism, employed entropy to reduce the uncertainty of pixels, and retained the details of the detected image, but the model incurred a high computational load. The method proposed by Xiao et al. (2022) classified high-resolution RS images through the extraction of super pixels via a multipath feature refinement network. However, this method had low accuracy for RS image boundary processing, and the classification effect was not optimal.

Based on the above analysis, to solve the problem of traditional classification algorithms and shallow learning algorithms not being suitable for extracting information from high-resolution RS images, an improved semantic segmentation model (DSC-SegNet) based on SegNet and depth-wise separable convolution (DSC) (Panigrahi & Raju, 2023) is proposed. The innovations of which are:

1. The vegetation classification algorithm proposed in this paper performs radiometric calibration and geometric, radiometric, and atmospheric corrections on the original RS data and resamples and combines the data according to the cubic convolution interpolation method. This improves to a certain extent the combined vegetation classification effect in RS images.
2. A vegetation classification algorithm based on an improved deep learning SegNet semantic segmentation model is proposed. The 13 layers in the traditional coding network have been replaced with DSC layers, each of which includes six processes: point by point convolution, zero padding, deep convolution, batch normalization, activation, and convolution. This results in an improved computational speed and classification accuracy of the algorithm.

REMOTE SENSING IMAGE DATA SOURCE AND IMAGE PREPROCESSING

Remote Sensing Image Data Sources

In this paper, the space RS images obtained by the GF-1 satellite are used as the data source. The GF-1 satellite is the first satellite launched by the Gaofen special project in April 2013 and was officially put into operation in December 2013. The space RS images obtained by GF-1 have good imaging quality and have been used for key technological breakthroughs in many scientific research directions, such as medium and high spatial resolution imaging, combined multispectral and wide coverage, 5-8-year-old high reliable satellite technology, high-precision and high attitude stability control, etc. Compared with other similar satellites, the width of the RS images obtained by GF-1 is the highest and can be used to meet the needs of resource and environmental analysis, environmental protection, disaster prevention, and emergency lights and provide high-resolution imaging data for Chinese researchers.

The main configuration of gaogao-1 includes two panchromatic multi-spectral (PMS) and four wide field-of-view (WFV) cameras. The specific parameters are listed in Table 1.

The camera configuration in Table 1 allows the gaogao-1 satellite to obtain wide and clear images quickly and efficiently and also gives it a spatial resolution of 2m / 8m / 16m and a wide-range imaging capability. The four multispectral WFV cameras can achieve a width of 800km through different combinations and complementarities, which also means that GF-1 can achieve global coverage in four days without side sway.

For the purposes of this study, the four WFV cameras of the GF-1 satellite were used to obtain 10 random images in February, May, October, and December 2020 in the study area as the data source of RS images. The list of images is shown in Table 2.

Data Preprocessing

The space RS image source data collected using the GF-1 satellite were first preprocessed. First, radiometric calibration and geometric correction were applied for the 10 RS images, and then radiometric correction, atmospheric correction, and cutting were carried out on the preliminarily processed image data. The atmospheric correction module in the SNAP software available for Sentinel data provided by ESA and the ENVI 5.5 software is capable of performing radiometric and atmospheric correction of the image data obtained using the GF-1 satellite. The ENVI 5.5 software

Table 1. Main configuration of GF-1 satellite

Camera	PMS	WFV
Number	2	4
Band and spectral intervals (μm)	Green: 0.52-0.59	Green: 0.52-0.59 Blue: 0.45-0.52 Red: 0.63-0.69 Near infrared: 0.77-0.89
	Blue: 0.45-0.52	
	Red: 0.63-0.69	
	Near infrared: 0.77-0.89	
	Full color: 0.45-0.90	
Spatial resolution (m)	Full color: 2	16
	Others: 8	
Width (km)	60	800
Revisit period (days)	4 (with side swing)	4
	41 (no side swing)	

Table 2. WFV camera images obtained using GF-1

WFV Camera Index	Collection Date	Number of Images
#4	2020 / 02 / 18	2
#2	2020 / 02 / 25	1
#3	2020 / 05 / 12	3
#2	2020 / 10 / 09	1
#1	2020 / 10 / 22	1
#3	2020 / 12 / 14	2

can also directly extract the metadata of the GF-1 satellite sensor, which allows easier processing, so this was the tool of choice for this study's preprocessing prior to analyses.

First, the cubic convolution interpolation method was used to resample the four bands with a resolution of 16m in the 10 source RS images and obtain images with a resolution of 8m. Then, the four bands with 8m resolution obtained by resampling were combined with the existing four bands with 8m resolution. Finally, by combining the bands, it was found that the combination of near-infrared, blue and green bands yielded the best vegetation differentiation effect and could be used to better distinguish different vegetation types.

VEGETATION CLASSIFICATION USING IMPROVED SEGNET SEMANTIC SEGMENTATION

SegNet Model

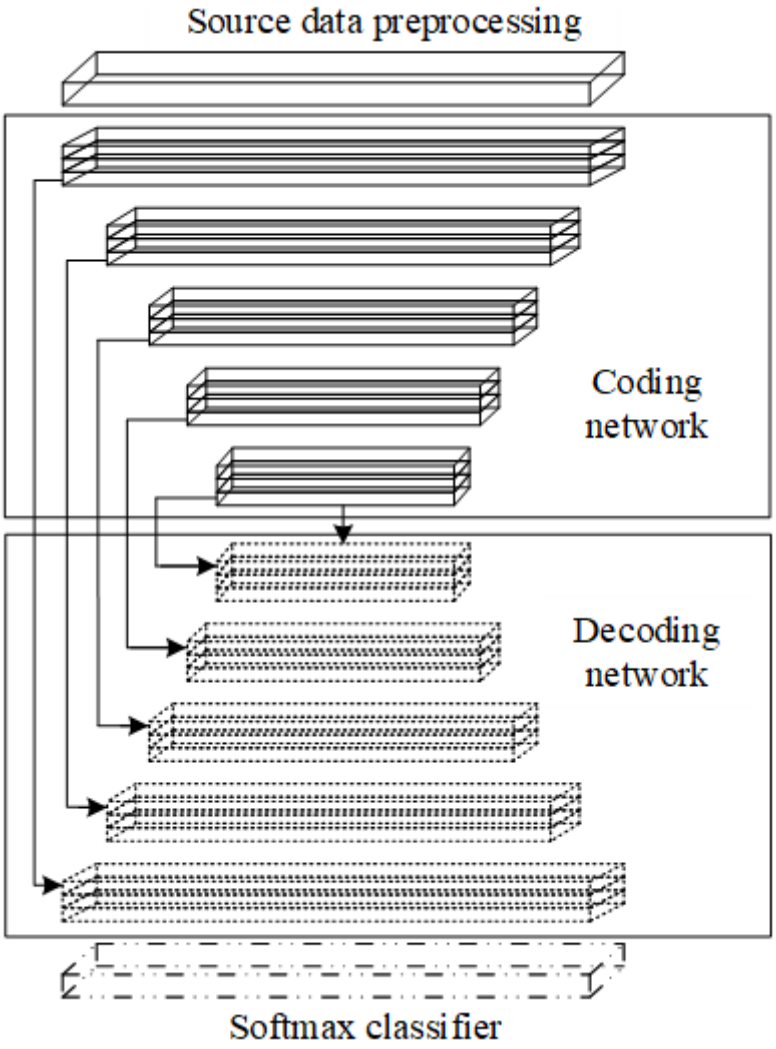
The traditional algorithms for semantic image segmentation are generally based on full CNN models (FCN). FCN models generally do not include a full connection layer, which decreases their complexity and increases the segmentation and calculation speeds during inference (Yu et al., 2022). However, current semantic segmentation models are generally based on slice networks, which are different from the traditional pixel-based models. Slice-based models lack the integrity of features and are inefficient. In this paper, the SegNet model is used to realize image semantic segmentation.

The SegNet model was developed on the basis of FCN models, which are pixel-based end-to-end models. SegNet is a traditional FCN model optimized for semantic segmentation and integrates a coding-decoding structure and a hopping network. Finally, the SegNet model can realize accurate segmentation even when only a small amount of sample data is available. Figure 1 depicts the basic structure of the SegNet semantic segmentation model.

The SegNet semantic segmentation model depicted in Figure 1 consists of three parts:

1. Coding network: The primary function of the coding network is to transform high-dimensional vectors into low-dimensional ones to perform low-dimensional extraction of high-dimensional image features. Multiple max-pooling operations enable the coding network to extract more translation-invariant features but at the expense of some image boundary information, thereby diminishing the accuracy of image segmentation. To avoid the aforementioned issue, the max pooling index is recorded during the max pooling operation in order to preserve the position of the image's maximal feature values. With this approach, the feature map's boundary information can be preserved through up-sampling.
2. Decoding network: The primary function of the decoding network is to map the low-resolution feature maps to high-spatial resolution feature maps. During the process of decoding, the max-pooling index is reused continuously and the mapping process is also based on the max index

Figure 1. Basic structure of the SegNet model



information of the response feature layer saved by the coding network during the down-sampling process. There are two purposes served by the continuous reuse of the max pooling index: (1) in conjunction with a reduction in the number of parameters, end-to-end training can be achieved through the corresponding optimization of the image boundary contour; and (2) all "coding-decoding" models implement up-sampling. The output of the final layer of the decoding network is a high-dimensional feature representing a vector that can serve as training data for the Softmax classifier.

3. Pixel-by-pixel classifier: The main function of the Softmax classifier is to classify each pixel separately. The basis of classification is that the class with the maximum probability for each pixel is regarded as the predicted class during segmentation. Therefore, the output of the classifier is the probability that each pixel is part of a class.

Improved SegNet Model

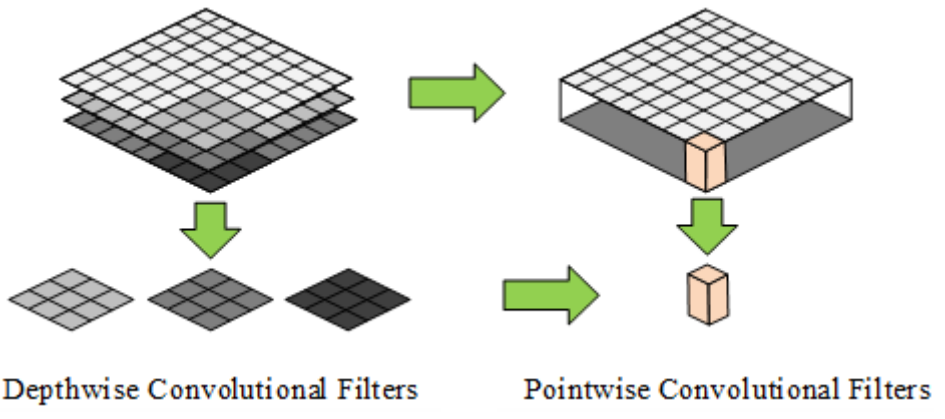
The coding network in SegNet model has 13 layers, which include three types: convolution layers, pooling layers, and batch normalization layers. The primary function of the convolution layer is to capture the local features of the RS images, while the pooling layer performs down-sampling and conveys scale-invariant characteristics to the subsequent layer. The primary purpose of the batch normalization layer is to normalize the distribution of the features and accelerate the rate of learning. There are many parameters in the coding network in the traditional SegNet model, which results in relatively slow coding speed. To enhance the overall performance of the traditional SegNet model, in this paper, several improvements are applied to the coding and decoding networks.

To further enhance the operation efficiency of the coding network, the pooling layer is removed and the ordinary convolution layer is substituted with a depth-wise separable convolution layer. Ordinary convolution involves simultaneously convolving the channels, whereas depth-wise separable convolution includes depth-wise convolution and point-wise convolution. Figure 2 depicts the process of depth-wise separable convolution.

In the improved SegNet model, depth-wise separable convolution is used instead of the 13 convolution layers of the initial network. Each depth-wise separable convolution layer includes six processes: activation, point-by-point convolution, zero padding, deep convolution, batch normalization, and convolution. If a $(H,W,3)$ RS image is input, the shape of the RS image will become a $(H/2,W/2,32)$ map after a series of convolution operations. When the $(H/2,W/2,32)$ map is input to the coding network, the length and width of the shape of each layer of RS image will be reduced to half and the number of channels will double. During the calculation process, only the first 11 depth-wise separable convolution layers can meet the requirements. At this time, the shape of the RS image output by the coding network becomes $(H/16,W/16,512)$.

The decoding network is then improved but without significant changes to its fundamental structure. Similar to the coding network, only the first 11 layers of the decoding network are utilized. The first 10 layers perform up-sampling, zero padding, convolution, and batch normalization operations. For an RS feature map with dimensions of $(H/16,W/16,512)$ obtained by the coding network, after passing through each layer, the length and width will be cut in half and the channels' number will become twice the original. Finally, the shape of the output RS image will become $(H/2,W/2,64)$. The notation $(H/2,W/2,NC)$ represents the output size of the last convolution layer, where NC represents the pixel classification color's number.

Figure 2. Depth-wise separable convolution operation



After decoding, the probability of each pixel's classification category can be obtained through the Softmax layer. The output result of the classifier is the ultimate result of the RS image semantic segmentation model. The overall improved SegNet model consists of the improved coding and decoding networks of the SegNet model working together.

Network Loss Function

For an RS image with *X* pixels, the pixels are defined as: $M = \left\{m_1, m_2, m_3, ..., m_X\right\}$, and the label category corresponding to each pixel is denoted as $N = \left\{n_1, n_2, n_3, ..., n_X\right\}$. For pixel n_i, there are:

$$n_i = \begin{cases} 1, This\ pixel\ is\ the\ vegetation\ area \\ 0, This\ pixel\ is\ not\ the\ vegetation\ area \end{cases} \tag{1}$$

According to the previous analysis, the output of the improved SegNet network model is the likelihood that each pixel depicts a vegetation area, so it can be viewed as a conditional distribution model of two classes. The training process of the model can be regarded as maximizing the likelihood that the label corresponding to the pixel is the correct classification. To achieve this, the following cross-entropy loss function is adopted in this paper:

$$F_L = -\frac{1}{X}\sum_{i=1}^{X} N_i\left(\ln p\right) + \left(1 - N_i\right)\ln\left(1 - p\right) \tag{2}$$

where $p = P\left(N_i = 1 \middle| M_i\right)$ represents the probability that pixel M_i is classified into label category $N_i = 1$.

Vegetation Classification

The vegetation classification flowchart based on the improved SegNet model is depicted in Figure 3.

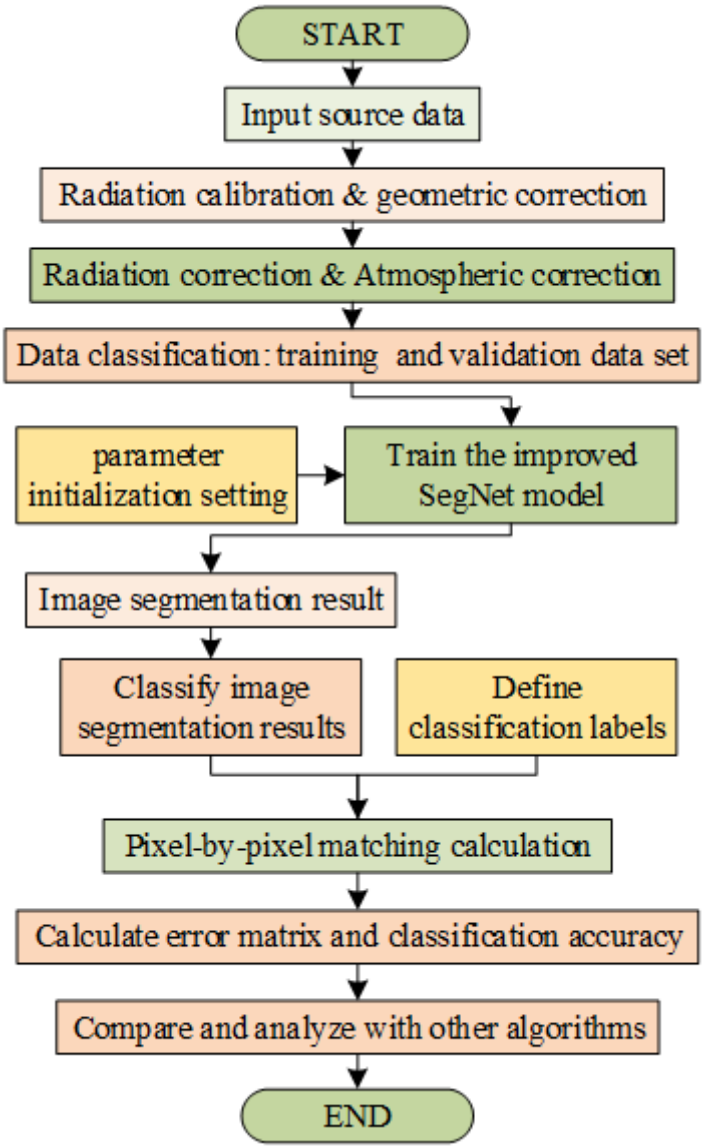
The evaluation of the vegetation classification process applied in this paper mainly included the following four steps:

1. Data preprocessing: Radiometric calibration and geometric correction were applied on the 10 RS images, and then the ENVI 5.5 software was used for radiometric and atmospheric correction to obtain the preprocessed image data.
2. Training of the improved SegNet model: First, the training parameters were initialized, and then the processed RS image data were input into the improved SegNet semantic segmentation model for training and segmentation.
3. Classification of the model's segmentation results: The Softmax classifier was used to classify the model's segmentation results, which were in turn matched with the corresponding labels pixel-by-pixel to calculate the classification results' accuracy.
4. The classification results and accuracy were compared with those obtained using other models for validation.

EXPERIMENT AND ANALYSIS

To validate the proposed method, a series of experiments was conducted on a computer with a i5-9300H CPU, an NVIDIA GeForce GTX 1600 Ti graphics card, and 8GB of RAM. The operating

Figure 3. Vegetation classification based on improved SegNet model



system was Windows 10 64 bit, and the neural network was built using the Keras module of the Tensorflow framework. The software programming environment was Python 3.7.

Model Parameter Setting

Before training the improved SegNet model, the parameters of the model were set, as follows:

Learning Rate

The primary function of the learning rate is to adjust the progress of the SegNet model's training. A value that is too large may cause oscillations and lack of convergence, while a value that is too

small will reduce the calculation and convergence speeds of the model. In this experiment, the model learning rate was set to 0.01.

Learning Rate Change Step

The primary function of the learning rate change step is to modify the rate of learning and ensure that the network converges as training progresses. To achieve smooth adjustment over the maximum range, the learning rate change step was set to 0.1.

Momentum Parameters

The main function of momentum parameter is to increase the model's convergence speed. In this experiment, this parameter was set to 0.9.

Weight Attenuation Value

The main function of weight attenuation value is to reduce the impact of the model's complexity on the loss function to a certain extent. For this experiment, it was set to 0.0005.

Learning Rate Change Frequency

This parameter regulates how often the learning rate changes during training and was set to 2000 for this experiment.

Training and Test Batch Sizes

The sizes of the training and test batches were set to 25 and 15, respectively.

Maximum Number of Iterations

The maximum number of training iterations was set to 50.

Evaluation Indices

The evaluation indices are used to evaluate the classification results. Before evaluating the classification results, the confusion matrix needs to be calculated, as shown in Table 3.

Here, l denotes total number of categories classified, X_{i+} denotes the sum of line i, X_{+i} denotes the sum of column i, and Y denotes the total number of samples.

In this paper, the classification effect of the improved SegNet model was evaluated using the following four evaluation indexes:

1. Overall accuracy: Overall accuracy is the proportion of correctly classified samples relative to the total number of samples. Its purpose is to provide an overall metric of the classification efficacy. The overall accuracy can be expressed using the parameters of the confusion matrix of Table 3 as:

Table 3. Confusion matrix of classification results

Type	Data				
Category	1	2	...	l	Sum
1	X_{11}	X_{12}	...	X_{1l}	X_{1+}
2	X_{21}	X_{22}	...	X_{2l}	X_{2+}
...	...	...	...	...	...
1	X_{l1}	X_{l2}	...	X_{ll}	X_{l+}
Sum	X_{+1}	X_{+2}	...	X_{+l}	Y

$$S_1 = \frac{X_{11} + X_{22} + ... + X_{ll}}{N_{TS}} \tag{3}$$

where N_{TS} is the total number of samples.

2. Producer accuracy: Producer accuracy is defined as the ratio of the number of correct classifications corresponding to a certain category to the total number of samples belonging to this category in the standard data. Taking column *i* as an example, the producer accuracy can be expressed using the elements of the confusion matrix of Table 3 as:

$$S_{2i} = \frac{X_{ii}}{X_{+i}} \tag{4}$$

3. User accuracy: User accuracy refers to the number of correct classifications corresponding to a certain category and compared to the total number of samples classified in that category. Taking row *i* as an example, the user accuracy can be expressed using the elements of the confusion matrix of Table 3 as:

$$S_{3i} = \frac{X_{ii}}{X_{i+}} \tag{5}$$

4. Kappa coefficient: According to equations (3), (4), and (5), the three evaluation indexes of overall accuracy, producer accuracy, and user accuracy use only the sum of the main diagonal elements and rows and columns in the confusion matrix and therefore do not reflect the evaluation efficacy comprehensively and intuitively. The Kappa coefficient is calculated using the elements in the error matrix and is a more comprehensive metric of the evaluation results. The Kappa coefficient can be calculated using equation (6):

$$Ka = \frac{Y\sum_{i=1}^{l} X_{ii} - \sum_{i=1}^{l}\left(X_{i+}X_{+i}\right)}{Y^2 - \sum_{i=1}^{l}\left(X_{i+}X_{+i}\right)} \tag{6}$$

where Ka is the Kappa coefficient, whose value range is $[0,\ 1]$.

Equation (6) shows that all elements of the confusion matrix are used in the calculation of the Kappa coefficient, so the error information in the matrix is utilized comprehensively. It is generally regarded as a comprehensive evaluation index for evaluating classification results. Specifically, Table 4 displays the relationship between the Kappa coefficient calculation results and the accuracy of classification.

Model Training Process

During the model training process, the increase of the accuracy value is used as the standard to stop model training; that is, when the accuracy does not increase after three consecutive iterations, the training is stopped.

Table 4. Relationship between kappa and classification accuracy

Ka	Classification Accuracy
$0 \leq Ka \leq 0.4$	Poor
$0.4 < Ka \leq 0.6$	Middle
$0.6 < Ka \leq 0.8$	High
$0.8 < Ka \leq 1$	Excellent

First, the classification accuracy of the vegetation classification algorithm that utilizes the improved SegNet model was analyzed. Figures 4 and 5 respectively show the change in classification accuracy and loss function as the number of iterations increased.

As depicted in Figure 4 and Figure 5, after 25 training iterations, the classification accuracy did not change substantially, so the training was terminated. At this point, the classification accuracy of the training dataset reached 98.9%, the classification accuracy of the verification dataset reached 93.8%, the loss function value of the training dataset was 0.042 and gradually converged, and the loss function value of the verification dataset converged after it decreased to 0.16. This shows that the proposed vegetation classification algorithm based on the improved SegNet model has good classification performance.

Next, the classification effect of the vegetation classification algorithm was analyzed. Figure 6 shows a sample RS image containing vegetation information captured by the GF-1 satellite used to analyze the classification effect. The left half of Figure 6a is a non-vegetation-covered area, which includes crop cultivated land and fruit land, while the right is a vegetation-covered area. Figure 6b shows a set of feature maps to be identified. The segmented small block shape in the map is cut into a rectangle, while the other areas are zero-padded to identify the features the vegetation pixels. Figure 6c shows the ideal segmentation effect, while Figure 6d shows the final classification result of the algorithm.

Figure 4. Classification accuracy curve

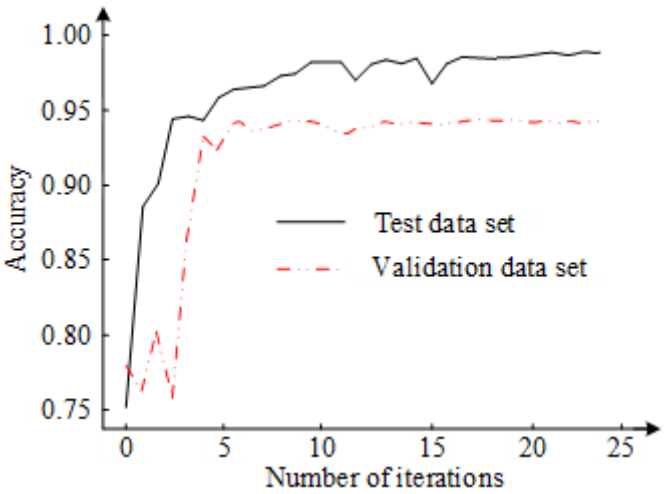
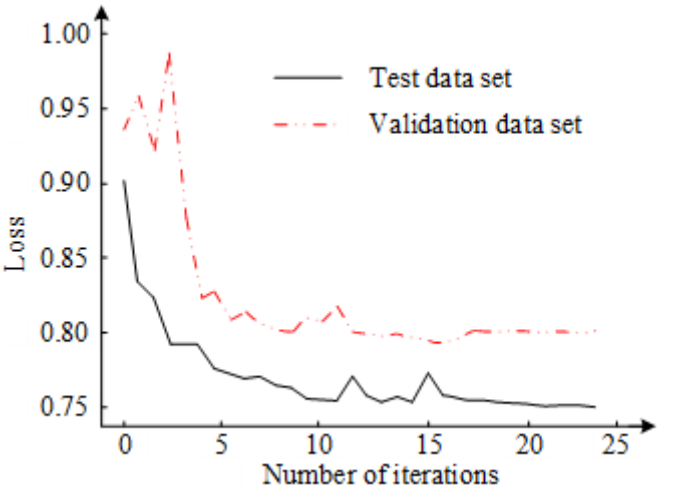


Figure 5. Loss curve



As depicted in Figure 6, the vegetation classification effect map obtained using the improved SegNet-based classification algorithm is slightly different from the ideal result but generally very similar, and the accuracy of classification can fully meet the needs of cartography and other aspects.

Next, the confusion matrix and evaluation index values of the vegetation classification algorithm were analyzed. The results are shown in Tables 5 and 6, respectively.

Figure 6. (a) Vegetation classification results of a sample RS image original image, (b) vegetation classification results of a sample RS image feature map to be identified, (c) vegetation classification results of a sample RS image ideal classification, (d) vegetation classification results of a sample RS Image actual classification result

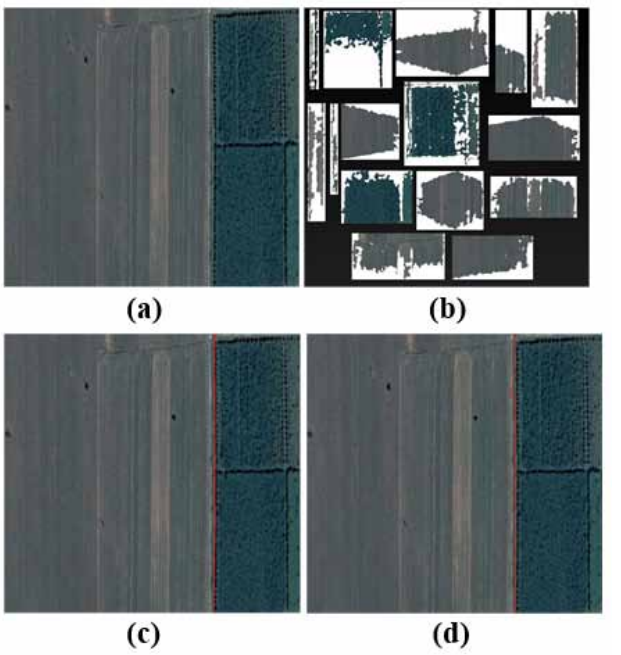


Table 5. Confusion matrix

Type	Data		
Category	Fruit land	Cultivated land	Sum of lines
Fruit land	128639	7937	136576
Cultivated land	6342	342704	349046
Sum of columns	19211	350641	$Y = 369852$

Table 6. Evaluation indices

Evaluation Index	Fruit Land	Cultivated Land
Producer's Accuracy	97.53%	98.86%
User's Accuracy	97.32%	98.45%
Overall Accuracy	98.48%	
Kappa	0.9531	

The values in the error matrix of Table 5 represent the number of pixels. According to Tables 5 and 6, the number of misclassified pixels accounted for less than 1% of the total number, the producer and user accuracies were greater than 97%, the total accuracy was 98.48%, and the Kappa coefficient was 0.9531. This demonstrates that the classification results are extremely accurate, and the classification effect was consistent with the ideal results.

Next, the proposed DSN-SegNet model was compared to VM-SegNet (Yu et al., 2022), AM-PSPNet (Ren et al., 2022), and MFRNet (Xiao et al., 2022) with respect to the total number of parameters, the overall accuracy, and the Kappa coefficient. The final results are shown in Table 7.

From Table 7, it is evident the total number of parameters of the proposed vegetation classification algorithm built on the improved SegNet model is the lowest compared with other traditional algorithms, indicating that the proposed vegetation classification algorithm model is more efficient and quicker to train. This is due to the elimination of the fully connected layer from the improved SegNet model, which significantly decreases the overall complexity of the model and increases the operation speed.

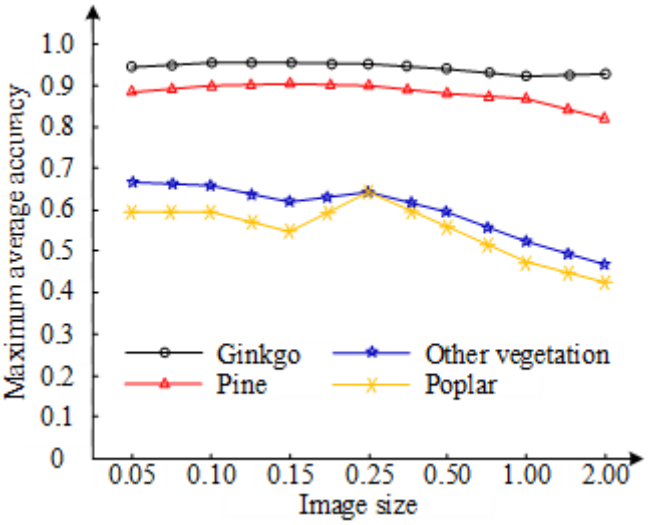
In addition, it can be seen from Table 7 that the overall accuracy of the proposed vegetation classification algorithm and its Kappa coefficient were the highest, the latter reaching 0.9531, indicating that the proposed algorithm has superior overall performance and classification effect relative to other algorithms. This also demonstrates that the improvement of SegNet model proposed in this paper enhances the network's classification accuracy and performance.

In order to further analyze the impact of the RS image resolution and vegetation type on the classification results of the algorithm, Figure 7 depicts the relationship between the maximum average accuracy for different types of vegetation and the spatial resolution of the RS images analyzed.

Table 7. Evaluation indicators of different algorithms

Algorithm	Number of Parameters	Overall Accuracy	Kappa
DSN-SegNet (the authors')	4646831	98.48%	0.9531
VM-SegNet (Yu et al., 2022)	9563487	97.57%	0.9512
AM-PSPNet (Ren et al., 2022)	11652364	97.35%	0.9465
MFRNet (Xiao et al., 2022)	13541265	96.29%	0.9313

Figure 7. Variation of maximum average accuracy for different vegetation types with spatial resolution



From Figure 7, for RS images of different sizes, the maximum average accuracy of ginkgo tree classification was the highest, reaching a maximum of 96% and a minimum of 94%. It was followed by pine, other vegetation types, and poplar. The classification effect and the maximum average accuracy of the various types of vegetation did not change substantially for RS resolutions less than 0.15m. As the spatial resolution decreased, the accuracy of image feature recognition increased, so the significance of vegetation features in RS images became more and more prominent. Compared to the original SegNet network, the improved encoder network using the deep separable convolutional network reduces the number of parameters effectively and significantly improves the inference speed. To improve the robustness of the model, data augmentation was performed on the dataset images to increase the number and diversity of samples.

CONCLUSION

A vegetation classification algorithm based on an improved version of the SegNet semantic segmentation model is proposed for vegetation classification in RS images. The proposed classification algorithm was validated using RS image data obtained from GF-1. The results showed that compared to traditional classification algorithms, the classification accuracy of the proposed algorithm is significantly improved, reaching 98.9%, with a number of model parameters reduced by more than half compared to traditional algorithms. In addition, its classification effect was consistent with the ideal effect, so the algorithm can fully meet the needs of topographical engineering and other applications.

Unmanned aerial vehicle hyperspectral RS images have centimeter-level spatial resolution, which not only enhances visibility but also significantly enhances the generation of isospectral phenomena, introducing more interference factors to fine vegetation classification. At the same time, the increase of information content significantly reduces the efficiency of data processing. In subsequent research, the ultra-fine features contributing to vegetation classification will be fully explored, along with techniques for interference reduction. The requirement for efficient processing using other advanced data analysis technologies such as deep learning will provide more technical insight into the analysis of RS images and the possible applications. In addition, there are still some shortcomings in this

study, such as the lack of a sound theoretical basis for selecting network models and setting training parameters. The next step will focus on the selection of appropriate semantic segmentation models and setting the optimal training parameters based on the characteristics or the classification elements of the RS image dataset.

COMPETING INTERESTS STATEMENT

The authors declare that there is no competing interest for this work, and no funding was received.

DATA AVAILABILITY

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

FUNDING STATEMENT

This work was supported by the scientific and technological key project in Henan Province (232102220018).

REFERENCES

Alsmirat, M. A., Al-Alem, F., Al-Ayyoub, M., Jararweh, Y., & Gupta, B. (2019). Impact of digital fingerprint image quality on the fingerprint recognition accuracy. *Multimedia Tools and Applications*, *78*(3), 3649–3688. doi:10.1007/s11042-017-5537-5

Chen, C., Jiang, F., Yang, C., Rho, S., Shen, W., Liu, S., & Liu, Z. (2018). Hyperspectral classification based on spectral–spatial convolutional neural networks. *Engineering Applications of Artificial Intelligence*, *68*(2), 165–171. doi:10.1016/j.engappai.2017.10.015

Chopra, M., Singh, S. K., Sharma, A., & Gill, S. S. (2022). A comparative study of generative adversarial networks for text-to-image synthesis. *International Journal of Software Science and Computational Intelligence*, *14*(1), 1–12. doi:10.4018/IJSSCI.300364

Goudarzi, S., Kama, M. N., Anisi, M. H., Soleymani, S., & Doctor, F. (2018). Self-organizing traffic flow prediction with an optimized deep belief network for internet of vehicles. *Sensors (Basel)*, *18*(10), 3459–3465. doi:10.3390/s18103459 PMID:30326567

Kadri, O., Benyahia, A., & Abdelhadi, A. (2022). Tifinagh handwriting character recognition using a CNN provided as a web service. *International Journal of Cloud Applications and Computing*, *12*(1), 1–17. doi:10.4018/IJCAC.297093

Kadry, S., Taniar, D., & Meqdad, M. N. (2022). Assessment of brain tumor in flair MRI slice with joint thresholding and segmentation. In R. Chbeir, Y. Manolopoulos, & R. Prasath (Eds.), Lecture notes in computer science: Vol. 13119. *Mining intelligence and knowledge exploration. MIKE 2021*. Springer. doi:10.1007/978-3-031-21517-9_5

Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *173*(3), 24–49. doi:10.1016/j.isprsjprs.2020.12.010

Kotaridis, I., & Lazaridou, M. (2021). Remote sensing image segmentation advances: A meta-analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, *173*(3), 309–322. doi:10.1016/j.isprsjprs.2021.01.020

Kumar, S., Kumar, S., Ranjan, N., Tiwari, S., Kumar, T. R., Goyal, D., & Rafsanjani, M. K. (2022). Digital watermarking-based cryptosystem for cloud resource provisioning. *International Journal of Cloud Applications and Computing*, *12*(1), 1–20. doi:10.4018/IJCAC.311033

Kumbhojkar, N. R., & Menon, A. B. (2022). Integrated predictive experience management framework (IPEMF) for improving customer experience: In the era of digital transformation. *International Journal of Cloud Applications and Computing*, *12*(1), 1–13. doi:10.4018/IJCAC.2022010107

Lv, L., Wu, Z., Zhang, L., Gupta, B. B., & Tian, Z. (2022). An edge-AI based forecasting approach for improving smart microgrid efficiency. *IEEE Transactions on Industrial Informatics*, *18*(11), 7946–7954. doi:10.1109/TII.2022.3163137

Mandle, A. K., Sahu, S. P., & Gupta, G. P. (2022). CNN-based deep learning technique for the brain tumor identification and classification in MRI images. *International Journal of Software Science and Computational Intelligence*, *14*(1), 1–20. doi:10.4018/IJSSCI.304438

Nhi, N. T. U., & Le, T. M. (2022). A model of semantic-based image retrieval using C-tree and neighbor graph. *International Journal on Semantic Web and Information Systems*, *18*(1), 1–23. doi:10.4018/IJSWIS.295551

Panigrahi, S., & Raju, U. S. N. (2023). DSM-IDM-YOLO: Depth-wise separable module and inception depth-wise module based YOLO for pedestrian detection. *International Journal of Artificial Intelligence Tools*, *32*(4), 2350011. doi:10.1142/S0218213023500112

Ren, S., Liu, X., Liu, H., & Wang, L. (2022). Cultivated land segmentation of remote sensing image based on PSPNet of attention mechanism. *International Journal of Robotics and Automation*, *37*(1), 11–19. doi:10.2316/J.2022.206-0730

Stergiou, C. L., Psannis, K. E., & Gupta, B. B. (2021). InFeMo: Flexible big data management through a federated cloud system. *ACM Transactions on Internet Technology*, *22*(2), 1–22. doi:10.1145/3426972

Wang, H., Li, Z., Li, Y., Gupta, B. B., & Choi, C. (2020). Visual saliency guided complex image retrieval. *Pattern Recognition Letters*, *130*, 64–72. doi:10.1016/j.patrec.2018.08.010

Wang, H., Wang, H., & Li, Z. (2020). Visual saliency guided complex image retrieval. *Pattern Recognition Letters*, *130*(Feb), 64–72. doi:10.1016/j.patrec.2018.08.010

Wei, W., Li, X. H., & Liu, J. Z. (2019). Study on remote sensing image vegetation classification method based on decision tree classifier. *2019 IEEE Symposium Series on Computational Intelligence (SSCI),* 2292-2297.

Wen, J., & Chen, W. (2020). Remote sensing image classification by support vector machine based on improved particle swarm optimization. *Jiangsu Keji Daxue Xuebao. Ziran Kexue Ban*, *34*(5), 66–72.

Wenjuan, L. I., & Shao, H. (2021). Landslide susceptibility assessment based on multi-scale segmentation of remote sensing and geological factor evaluation. *The Chinese Journal of Geological Hazard and Control*, *32*(2), 94–99.

Xia, Q., Qin, C. Z., Li, H., Huang, C., Su, F.-Z., & Jia, M.-M. (2020). Evaluation of submerged mangrove recognition index using multi-tidal remote sensing data. *Ecological Indicators*, *113*(5), 1–7. doi:10.1016/j.ecolind.2020.106196

Xiao, T., Liu, Y., Huang, Y., & Yang, G. (2022). MFRNet: A multipath feature refinement network for semantic segmentation in high-resolution remote sensing images. *Remote Sensing Letters*, *13*(10), 1271–1283. doi:10.1080/2150704X.2022.2144778

Yang, H. Y., Jianmin, D. U., & Ruan, P. Y. (2021). Vegetation classification of desert steppe based on unmanned aerial vehicle remote sensing and random forest. *Nongye Jixie Xuebao*, *52*(6), 186–194.

Yu, C., Liu, Y., & Xia, X. (2022). Precise segmentation of remote sensing cage images based on segnet and voting mechanism. *Applied Engineering in Agriculture*, *38*(3), 573–581. doi:10.13031/aea.14878

Yu, C. Q., Wang, L. G., Zhao, J., Hao, L., & Shen, Y. (2020). Remote sensing image classification based on RBF neural network based on fuzzy C-means clustering algorithm. *Journal of Intelligent & Fuzzy Systems*, *38*(4), 3567–3574. doi:10.3233/JIFS-179579

Zhang, C., Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J., & Atkinson, P. M. (2018). A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *140*(6), 133–144. doi:10.1016/j.isprsjprs.2017.07.014

Zhang, H., Wang, L., & Tong, F. (2021). Large remote sensing image segmentation with stitching strategy based on dominant color. *Journal of Software*, *16*(2), 67–79. doi:10.17706/jsw.16.2.67-79

Zhao, Z., Guo, L., Jia, M., & Wang, L. (2021). The generalized gamma-DBN for high-resolution SAR image classification. *Remote Sensing (Basel)*, *10*(6), 878–885. doi:10.3390/rs10060878

Wanjun Chang obtained Master's degree in Computer Science from Guizhou University in 2011. He is currently an Associate Professor at Henan Institute of Technology. His research focuses on data mining and deep learning.

Dongfang Zhang obtained Master's degree in Computer Science from Jiangsu University in 2021. He is currently a teaching assistant at Henan Institute of Technology. His research focuses on data mining and machine learning.