Neural Computing and Applications (2023) 35:8667–8675
https://doi.org/10.1007/s00521-022-07737-w

S.I.: AI BASED TECHNIQUES AND APPLICATIONS FOR INTELLIGENT IOT SYSTEMS

HRNet- and PSPNet-based multiband semantic segmentation of remote sensing images

Yan Sun[1] · Wenxi Zheng[1]

Received: 10 April 2022 / Accepted: 17 August 2022 / Published online: 3 September 2022

Abstract

High-resolution remote sensing images have become mainstream remote sensing data, but there is an obvious "salt and pepper phenomenon" in the existing semantic segmentation methods of high-resolution remote sensing images. The purpose of this paper is to propose an improved deep convolutional neural network based on HRNet and PSPNet to segment and realize deep scene analysis and improve the pixel-level semantic segmentation representation of high-resolution remote sensing images. Based on hierarchical multiscale segmentation technology research, the main method is multiband segmentation; the vegetation, buildings, roads, waters and bare land rule sets in the experimental area are established, the classification is extracted, and the category is labeled at each pixel in the image. Using the image classification network structure, different levels of feature vectors can be used to meet the judgment requirements. The HRNet and PSPNet algorithms are used to analyze the scene and obtain the category labels of all pixels in an image. Experiments have shown that artificial intelligence uses the pyramid pooling module in the classification and recognition of CCF satellite images. In the context of integrating different regions, PSPNet affects the region segmentation accuracy. FCN, DeepLab and PSPNet are now the best methods and achieve 98% accuracy. However, the PSPNet object recognition algorithm has better advantages in specific areas. Experiments show that this method has high segmentation accuracy and good generalization ability and can be used in practical engineering.

Keywords Remote sensing image · Semantic segmentation · Ensemble learning · Convolutional neural network

1 Introduction

The characteristics of a large quantity of data, strong ambiguity and rich semantic details determine the higher requirements for the segmentation efficiency and segmentation effect of remote sensing images. There are dozens or even hundreds of bands of hyperspectral image data, and there is considerable information redundancy [1]. Therefore, effective band information needs to be extracted for hyperspectral image data and cannot be fully analyzed because redundant information leads to low efficiency. By effectively extracting the band information effect and efficiency of hyperspectral remote sensing image classification, the original band image can be processed by dimensionality reduction. As an important attribute of remote sensing images, semantics play a unique role in image analysis [2]. The classic image semantic analysis methods are mostly for single-band or color images, but the semantic analysis methods of multispectral and hyperspectral remote sensing images need to be strengthened [3], including images defined by generalized semantics, two-dimensional image spatial mapping mode of spatial characteristics (or other), spectral feature point distributions of image semantic features (or other) or map expression, and special cases of nonsemantic area and semantic area single-band images. Visual band image semantics are a special

✉ Yan Sun
sunyan@hrbeu.edu.cn

Wenxi Zheng
zhengwenxi@hrbeu.edu.cn

[1] College of Information and Communication Engineering,
Harbin Engineering University,
Harbin 150001, Heilongjiang, China

case [4]. Extracting two-dimensional images is particularly good, and single-band image semantics can be used for analysis, while for multiband images, more sensitive semantic segmentation technology is required for identification.

FCNs face many problems. At present, improving FCN technology is still focused on improving the algorithm, and there is no overall improvement in the FCN technology. This paper adopts the following methods to solve the problem. First, the subspace is used to divide the relevant frequency bands. Then, the adaptive frequency band method is adopted in each subspace to select the frequency band with more information. Finally, the J-M distance model is introduced to distinguish the spectral separability of foreign objects in the same spectrum. Finally, deep convolutional neural networks, FCN, DeepLab, HRNet, and PSPNet are applied to the semantically classified. The work can be divided into complete target labeling, label image coloring, image data enhancement and image edge extraction on the acquired remote sensing images. The existing convolutional neural network is referenced, and the network structure is modified to obtain a complete network consisting of a convolutional layer and a deconvolutional layer using three different network training schemes.

The main contribution of this paper is to propose an improved HRNet- and PSPNet-based deep convolutional neural network for segmentation and deep scene analysis to improve pixel-level semantic segmentation representation of high-resolution remote sensing images. Compared with existing research, this paper utilizes the advantages of HRNet and PSPNet to perform fast and efficient segmentation of high-resolution remote sensing images. The traditional satellite image algorithm is based on the traditional semantic segmentation algorithm, which cannot accurately distinguish the elements in the image. The algorithm in this paper not only improves segmentation accuracy but also improves processing efficiency.

The structure of this paper is as follows. Chapter 1 introduces the current application and shortcomings of semantic segmentation and introduces the current processing methods of satellite high-resolution images. Combined with the current background, the significance of this research is proposed, and the ideas for this paper are introduced. The second chapter summarizes the research on remote sensing image analysis and semantic segmentation algorithms in recent years. Based on previous research, the innovation of this paper is proposed. The third section introduces the analysis principle of remote sensing images, optimizes the semantic segmentation algorithm based on a neural network, and introduces the optimization process in detail. The fourth section tests the algorithm proposed in this paper. The results also prove the effectiveness of the algorithm in this paper. The fifth chapter is a detailed discussion of the experimental results, describing the research results of this paper in detail. Finally, the main work of this paper is summarized, and the shortcomings and prospects of the work are presented.

2 Related work

Image processing belongs to the category of remote sensing technology [5]. In other words, a cluster with similar features is identified as a cluster in feature space. Unsupervised classification has many advantages in image processing [6]. Commonly used methods include the K-means algorithm, ISODATA algorithm, and fuzzy clustering algorithm. Information is transformed into a series of homogeneous regions based on heterogeneous references, and the polygon entities obtained after segmentation are the objects that participate in the extraction and processing of information [7, 8]. Objects include spectral information, spatial information, and semantic information. In our country, object-oriented technology started late, but the technology has been developing rapidly, and many researchers have joined the object-oriented technology research [9]. The object-oriented method is used to extract the seismic damage information of the building, and it is divided into three levels: basic integrity, damage and complete collapse. The extraction results are good and meet the requirements of rapid seismic damage assessment [10, 11].

With the progress in society and the development of various sciences, remote sensing technology is also advancing with the times. With the launch of an increasing number of high-resolution satellites, the resolution of remote sensing images continues to increase [12]. Image information processing technology is becoming more mature as it gradually improves. The research of remote sensing science is often based on good remote sensing images, so the quality of remote sensing images greatly affects the subsequent image information processing [13]. The remote sensing image, as multiband information, is obtained by multiple different spectral sensors. Generally, the light images corresponding to each band in the multiband image can be processed separately to decompose a group of multiband images into multigray-scale images for processing [14]. However, the choice of bands should meet the requirements of a large amount of information, low correlation between bands, and good spectral separability. However, the above methods cannot meet the above three requirements simultaneously, due to the difficulty of remote sensing image scene analysis[15], the most advanced mismatch, confusion and less classification [16].

3 Multiband semantic segmentation method of remote sensing images based on HRNet (PSPNet)

3.1 Multiband remote sensing image semantic segmentation

Multiband remote sensing, also known as multispectral remote sensing, is a remote sensing technology that uses sensors with more than two spectral channels to simultaneously image ground objects. It divides the electromagnetic wave information reflected or radiated by objects into several spectral bands for reception and recording. The multiband algorithm is a method and the key to describing the heterogeneity of two image objects [17]. For the *d*-dimensional feature space, assuming that the feature values of two adjacent objects are f_{1d} and f_{2d}, the heterogeneity is defined as:

$$h = \sqrt{\sum_{d} (f_{1d} - f_{2d})^2} \tag{1}$$

The spectral feature and shape feature of the image target can be regarded as the one-dimensional feature space [18]. In Formula (2), find the standard deviation of each Witt and further standardize the characteristic space distance:

$$h = \sqrt{\sum_{d} \left[(f_{1d} - f_{2d})\sigma_{fd}\right]^2} \tag{2}$$

This process iteratively merges adjacent image objects into larger image objects. In the segmentation process, the spectral features and shape features of the image target are processed simultaneously. First, the segmentation parameters, including the band threshold and the weight of each feature, are set as the termination condition of the image target merge; then, the segmentation process is started. During each segmentation, the neighboring image objects are searched, and the neighboring image objects are merged into a larger image object according to the principle of least heterogeneity. Due to the high variability of remote sensing images, when searching for adjacent images, images with consistent bands are preferentially selected. In the first segmentation, a single pixel is used as the smallest image object [19]. In the second and subsequent segmentation processes, the image objects generated in the previous segmentation process are used as the basis for calculating the heterogeneity to determine the relationship between the heterogeneity *h* and the scale threshold. If h $\leq$ scale threshold, continue the nth $(n > 2)$ segmentation; otherwise, the segmentation will end.

(1) Vector MRF semantic model.

In remote sensing image interpretation, there is a strong correlation between the category attribute of a point and its neighborhood attribute, that is, the locality of the image [20]. MRF has the characteristics of investigating local features, which is in-line with the objective conditions of remote sensing images. The definition method of gray image MRF has become very mature.

Due to the equivalence relationship between the Gibbs distribution and MRF, the vector Gibbs distribution can be used to describe the definition of the vector form and the Markov properties of the image with the vector *X*:

$$p(X = x) = \exp\sum_{c\in C} V_c(x) / \sum_{c\in\partial} \exp - \sum_{c\in C} V_c(x) \tag{3}$$

Among them, C is called a subgroup, which is a collection of pixels or adjacent pixels. C is the set of all subgroups on l, and the definition of a subgroup is the same as the scalar Gibbs distribution. Different C subgroups correspond to different Gibbs parameters, which reflect the orientation and thickness of the remote sensing image semantics. Through the size of the parameters, we can estimate the band peaks and data of the remote sensing image and deduce the semantic direction and thickness of the image. If there are too many subgroups, the computational complexity will greatly increase. The Markov model generally uses a second-order neighborhood system [21]. Additionally, due to the randomness in semantic direction, this paper adopts an isotropic second-order neighborhood system. Vc(x) is a potential function that only depends on the state of each point in subgroup C.

(2) Hierarchical Markov random field and multiband image semantic segmentation model.

For multispectral remote sensing images defined on a grid point L in a finite two-dimensional space, different semantics represent different objects in the image; they contain many different types of semantics and form one or more regions on L. Derin proposed a hierarchical Markov random field model, which uses a high random field to represent the coding region mapping and a low random field to represent a gray image [22].

In this paper, the high level represents the coding area map random field $Y = (Y_1,Y_2\ldots,Y_N)$, and the low level represents the spectral vector random field $X = (X_1,X_2\ldots,X_N)$ [23]. Both *X* and *Y* are random fields defined on *L* (the definition of X is as described in §1.1), and the image space range is the same. *Y* is a scalar random field, $y = (y_1,y_2\ldots,y_N)$ is an implementation of random field *Y*, and *X* is a vector random field. Assuming that there are *K* types of image semantics, $Yi = k$ $(k = 1, 2\ldots,K)$ denotes that in the realization y of random field *Y*, point *i* corresponds to the *k* semantics. The given multispectral remote sensing image data are an implementation of vector random field $X = x_0$. Obviously, the segmentation result is unknown in advance, and the segmentation process is to estimate Y based on *X*. According to the maximum posterior probability estimation (MAP), starting from the random field *X*, seeking the best realization of *Y y3*, the posterior probability distribution *P(Y = y|X = × 0)* should be maximized, and the semantic segmentation problem is transformed into solving a problem with image MAP:

$$p(Y = y^{*}|X = x_0) = \max p(Y = y|X = x_0) \tag{4}$$

According to the Bayes formula:

$$p(Y = y|X = x_0) = p(X = x_0|Y = y) \cdot p(Y = y)/p(X = x_0) \tag{5}$$

In the formula, for a given multispectral image × 0, P(X = × 0) is a constant. Therefore:

$$p_y = p(X = x_0|Y = y) \cdot p(Y = y) \tag{6}$$

If *Py* is the largest, the maximum posterior segmentation can be performed on the image. In the formula, *P(Y = Y)* reflects the probability of any random field segmentation pattern. Once the segmentation mode is determined, *K* groups of Gibbs corresponding parameter vectors are generated. The parameter vectors of each semantic type are denoted as *K (K = 1,2…)*. The condition under the Gibbs parameter is determined by the division method, *Y*.

(3) Algorithm principle.

To evaluate the accuracy, this paper optimizes the input and output of the algorithm. The input information is optimized with the characteristics of the fuzzy matrix, and finally, high evaluation and classification accuracy is achieved. According to certain rules and principles, all subspace bands that meet the conditions are selected for optimal combination in each subspace [24]. By calculating the correlation coefficients of adjacent bands and their transfer correlation vectors, the hyperspectral data space is divided into appropriate subspaces. Its advantage is that it fully reflects the local characteristics of the data. The correlation coefficient of the adjacent bands of the image is calculated as follows:

$$R_{i,j} = \frac{\sum_{k=1}^{n}(x_{ik} - \bar{x})(y_{jk} - \bar{y})}{\left(\sum_{k=1}^{n}(x_{ik} - \bar{x})^2 \sum_{k=1}^{n}(y_{jk} - \bar{y})^2\right)^{\frac{1}{2}}} \tag{7}$$

For each band, a mathematical model was established. The selected band contains more information:

$$I_i = \sigma_i/\left(\left(R_{i-1} + R_{i,i+1}\right)/2\right) \tag{8}$$

$$\sigma_i = \left[(1/(M \times N)) \sum_{x=1}^{M} \sum_{y=1}^{N} \left(f_i(x,y) - \bar{f}_i\right)^2\right]^{\frac{1}{2}} \tag{9}$$

$$R_{i,j}(x,y) = \left(E\left\{\left(f_i(x,y) - \bar{f_i}\right)\left(f_j(x,y) - \bar{f_j}\right)\right\}\right) \tag{10}$$

$$R_{i,j} = \left(E\left\{\left(f_i(x,y) - \bar{f_i}\right)\right\}^2\right)^{\frac{1}{2}} \times \left(E\left\{\left(f_j(x,y) - \bar{f_j}\right)\right\}^2\right)^{\frac{1}{2}} \tag{11}$$

In the formula, *Ii* is the index image *I*; *I* is the standard deviation *I* band; *Ri-1, I, Ri, I* + 1 are the correlation coefficients of the two bands before and after the adjacent bands of the *I* band; *Fi (x, y)* denote image I; *Fi* is the average pixel value; and E{} is the mathematical expectation of the *i* image pixel.

3.2 PSP-NET network structure of semantic segmentation

(1) Construct a hollow convolution residual structure model.

The remaining network building blocks are shown in Fig. 1. Set the input as X and the activation function as ReLU. The residual function F(x) is obtained by fitting and superimposed with the map x to obtain the output feature map Y, expressed as:

$$y = F(x) + x \tag{12}$$

In the network design, there are three convolutional layers inside each residual module, and the entire neural network contains seven residual learning blocks for a total of 21 convolutional layers. The empty convolution, also called the unfolded convolution, is shown in Fig. 1.

When the cavity convolution is completed, S is introduced. The loss of network structure may lead to serious consequences. However, without pools, deep networks are meaningless. Therefore, the use of empty convolutions can expand the perceptual space. During our learning process, zero elements are not adjusted [25].

(2) PPS-NET network structure.

In functional analysis, convolution is a mathematical operation that generates a third function through two functions f and g, the integral of the product of the overlapped part function value with the translation over the overlap length. This network has two main parts. The first part is the global feature extraction and training the image input network. The size of the convolution kernel is 7 × 7, and the stride of the convolution is 2 (stride = 2). This layer performs the same convolution, the output image size remains unchanged, and the number of channels becomes 64 dimensions. The image output by the convolution layer is downsampled once so that the convolution kernel is a

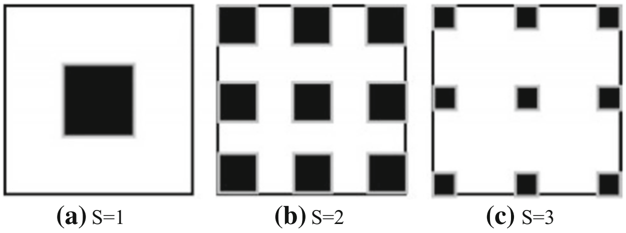
(a) S=1 **(b)** S=2 **(c)** S=3

Fig. 1 Empty convolution, also called unfolded convolution

3 × 3 local maximum pool, and the step size of the pool is also 2. Through pooling, abstract image features are extracted. After this layer is output, the image size is reduced to over 12. After the same convolution and pooling, the number of channels (number of dimensions) becomes 64. After pooling the output image, the remaining structure is entered with a dashed frame structure. The left and right residual structures are 1 × 1, 3 × 3, and 1 × 1 convolution kernels, and they all perform the same convolution. After three convolutions, the number of channels is increased to 256 dimensions, but the size of the feature map does not change (this part is repeated three times). The next step is to replace the pooling layer with the empty convolution (s = 2). After this empty convolution layer, the value is again halved to 14 before the initial input. Then, the feature map enters the residual structure again, which is another three-layer convolution kernel. The kernel is shown in the second remaining structure on the right (the dashed box). This process is performed four consecutive times, and the number of channels changes accordingly. Finally, the network structure on the left is complete.

(3) Network structure diagram.

Figure 2 is a network diagram. The CNN here is not a simple convolutional neural network but a residual network of hollow convolution. The input image size is 512 × 512 × 1, and the initial global feature is 64 × 64 × 512. Based on this global feature, the features extracted from different receptive domains are subjected to dimensionality reduction processing and pooled into feature maps of different sizes through the PPM model. Finally, these dimensionless features are combined with the existing global features to obtain a more detailed feature map.

4 Experimental setup

4.1 Experimental dataset

CCF AI's satellite image classification and recognition data are from 2015 high-resolution UAV remote sensing images in southern China, with meter resolution and spectral visible light bands (R, G, B). The training samples provided are divided into five categories: vegetation (marked 1), buildings (marked 2), water bodies (marked 3), roads (marked 4), others (marked 0), and cultivated land. The initial training set is composed of two large PNG images of 7,939*7,969, and the prediction set is composed of three large PNG images of 5,190*5,204. The training set includes three large PNG images, and the prediction set includes three PNG images. Compared with data from other types of competitions, remote sensing image visualization is more convenient, allowing a comprehensive and intuitive understanding of remote sensing images. Figure 3 shows the initial training prediction set. In this study, combined with the actual situation of the test area, the ground features are divided into seven categories according to the characteristics of remote sensing image data, the color of the image, and the degree of distinction and contrast in the image: lush wheat, sparse wheat, woodland, idle land, rural roads, residential land and asphalt roads.

4.2 Experimental environment

The server software and hardware configuration server content used in this experiment, the CPU Intel Xeon E5-2660 memory 96 GB GPU graphics card NVIDIA GeForce GTX TITANX operating system Linux Ubuntu16.04 LTSserver Cuda Cuda8.0 with cuDNN data processing Python2.7, MATLAB2014b, and sklearn. In this experiment, the calculation is based on a GPU to improve the calculation speed.

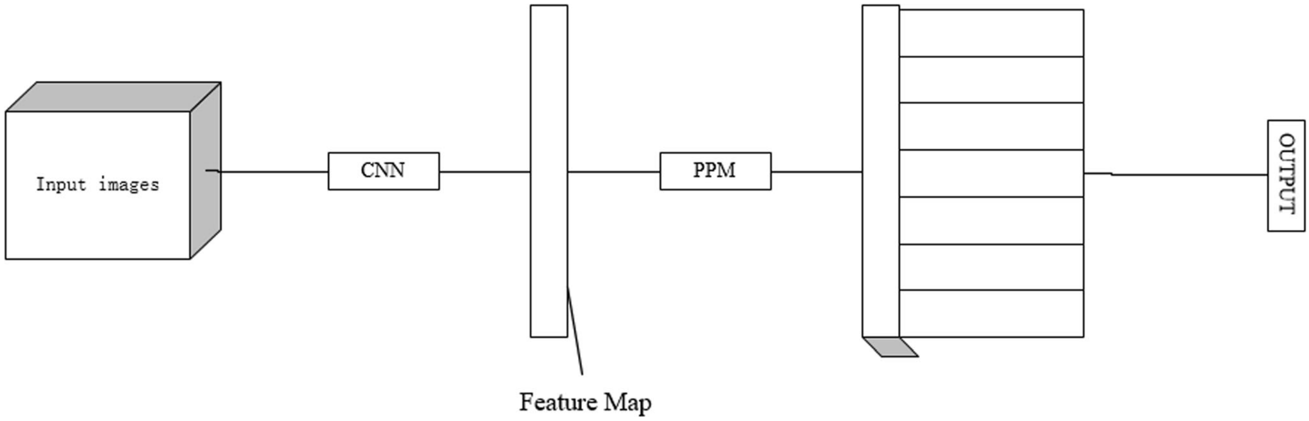

Fig. 2 Schematic diagram of network structure

Fig. 3 Initial training set prediction set

4.3 Experimental Procedure

The test process is divided into five steps:

(1) Subspace division. Calculate the correlation coefficient matrix between the frequency bands, divide the frequency bands with high correlation into a group, and obtain several subspaces.

(2) In each divided subspace, the adaptive band selection method is used to calculate the index of each band, and the band with high information content and high index is selected. This band can better capture the semantics of the image and better segment the image.

(3) Combining the optimal band selected from each subspace, calculate the J-M distance of each band combination, and filter out the band combined with the best separability between rural roads and bare land and asphalt roads and residential land.

(4) Train the HRNet classifier of semantic segmentation and input the best band combination of HRNet classification of semantic segmentation.

(5) Comparative analysis of HRNet (PSPNet) classification results without selective semantic segmentation.

5 Discussion

5.1 Experimental results of remote sensing image semantic segmentation

To solve the problem of semantic segmentation of remote sensing images, this paper tested four methods: HRNet and PSPNet, the mean-shift algorithm, and AlexNet. Semantic segmentation of remote sensing images is a basic task in remote sensing image understanding. There is an obvious "salt and pepper phenomenon" in the existing high-resolution remote sensing image semantic segmentation methods (some pixels in a single feature are identified as other features), resulting in the main reason for this phenomenon being that remote sensing objects have intraclass inconsistency (the same object label but different external

Table 1 Semantic segmentation results of PASCAL VOC 2011 test set

	HRNet	PSPNet	Mean-shift	Alexnet
Mean IU	39.8	45.32	56.34	49.6
Forward time	50 ms	54 ms	63 ms	74 ms
Rf.size	355	342	452	453
Max stride	32	32	32	32

features) and no difference between classes (two adjacent objects have different labels but similar external features).

Table 1 shows the semantic segmentation results of these four networks on the standard semantic segmentation test set PASCAL VOC2011. The observation results show that, as shown in Fig. 4, the four networks achieved good segmentation results on the standard test set. Therefore, it is considered feasible in this paper to use these four networks.

By comparing the data in the table and combining network training scheme ii with the color label image, the best semantic segmentation model can be obtained. In addition, among the three CNN networks, the HRNet has the best segmentation effect semantic tasks, and three evaluation indicators, pixel accuracy PA, average accuracy PA and average matching degree MIOU, are the highest. In addition to the objective data results, the following is an example of the results. As an example, the results obtained by training the HRNet network and color label images using the network training scheme are given. Image semantic segmentation results are shown in Fig. 5.

5.2 Classification results and accuracy evaluation experimental results and comparison

(1) Based on the above-mentioned multiband segmentation method, the study is divided into nine categories: shoals, marsh vegetation, xerophyte vegetation, estuaries, rivers, fish ponds, salt pans, roads and residential areas. The classification result is shown in Fig. 6. To evaluate its accuracy, this study randomly selects test points automatically, establishes a confusion matrix, and evaluates the accuracy of the classification results. The results are shown in Table 2.

(2) In this paper, the software adopts the adaptive threshold processing of the smart image edge detection operator, and the software's environmental image and the original image edge detection registration, edge detection and image as the original image new band. A new remote sensing image not only has the original image but also adds the edge. In image segmentation, edge information can be added to segmentation parameters to improve the utilization of shape features and spatial features. Under the same parameters, the original image and the image segmentation results are compared with edge information, as shown in Fig. 7.

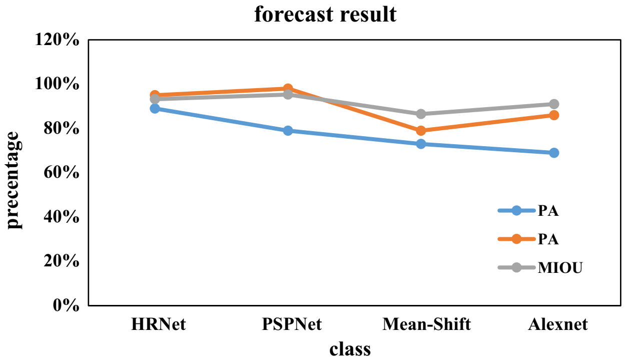

Fig. 4 Forecast results of different algorithms

6 Conclusions

The semantic segmentation method cannot describe the region boundary as clearly as the PSPNet method. This is because the PSPNet method only divides the spectral information, and the research object is a single pixel, which reflects the microscopic characteristics of the spectral distribution so the boundary is very clear. The semantic feature is a two-dimensional space feature that directly depends on the neighborhood and needs to be expressed in a certain two-dimensional space to reflect the macroscopic characteristics of the spectral distribution. Combined with spectral segmentation coding, the result of semantic segmentation coding can be used as effective information for subsequent image classification and interpretation. In the HRNet segmentation map with edge detection information, the original image can segment buildings more accurately. In the original image, the structure of the object is smaller than the actual object or irregular polygon, and the rectangle of the actual object is added to the image segmentation image edge information, including the structure of the polygon object, and basically matches the actual object. The shape of these objects is more similar to the real object. The house and road are not divided into one object, which is the segmentation method.

This study uses aerial data sources to classify coastal wetlands based on multiscale segmentation. It can be seen from the characteristics of the research field that various types of wetlands are related to each other and have different degrees of similarity, similar to the feature values of the metaspectral characteristic spectrum, which are difficult to distinguish. Through multiscale segmentation, the entity objects are at different scales. Different levels, spectral features, shape factors, and texture features are combined to extract different types of wetland information, effectively reducing the “salt and pepper” phenomenon in remote sensing images and achieving better classification accuracy.

This article only conducts a preliminary study of change detection in phase remote sensing images, and there are still many issues that need further research and discussion.

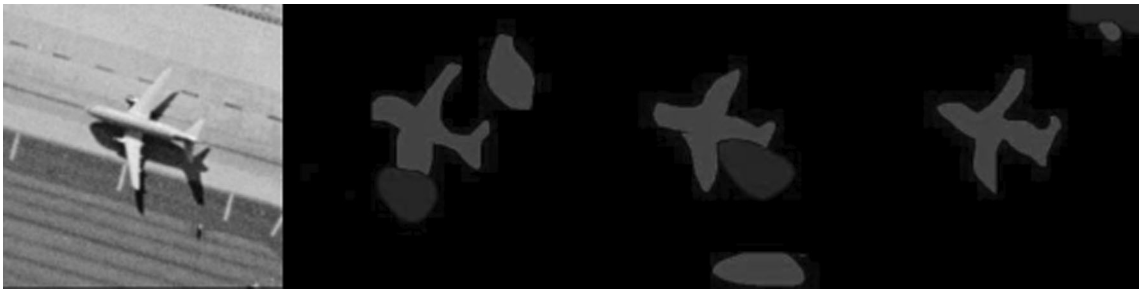

Fig. 5 Image semantic segmentation results

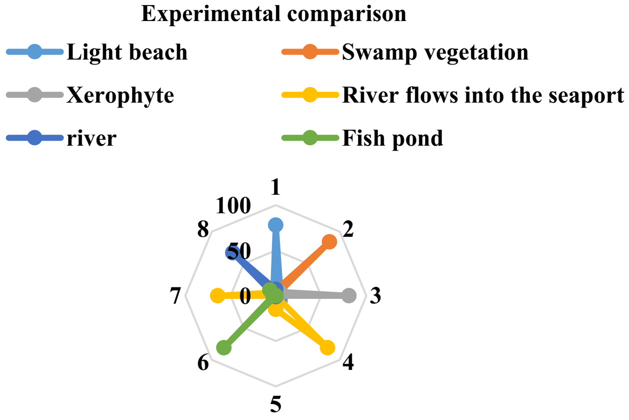

Fig. 6 Accuracy evaluation results of classification results

Table 2 Accuracy evaluation of classification results

Number of samples	Classification result					
	1	2	3	4	5	6
1	78	6	0	0	7	0
2	7	84	6	0	0	0
3	0	0	81	0	0	0
4	7	0	0	81	0	0
5	0	0	4	15	1	0
6	0	0	0	0	0	81

Based on the research work of this article, the following aspects can be used as the research focus to strengthen theoretical research and discussion. Although remote sensing change detection technology has been applied to and has borrowed some theoretical knowledge and model methods in disciplines such as mathematics and physics, remote sensing change detection does not form a systematic theoretical system, but is a relatively independent and complete theoretical system of established development and popularization image change detection technology. A combination multiple methods is needed. Although various change detection methods have been proposed, each method has advantages and disadvantages, and no one method is the best. Therefore, combining a variety of methods with image characteristics and actual requirements will effectively improve the change detection performance, and how to complement different detection methods remains to be further studied. Currently, most change detection methods only deal with changes in a single data source. How the various data sources change requires further research and experimentation.

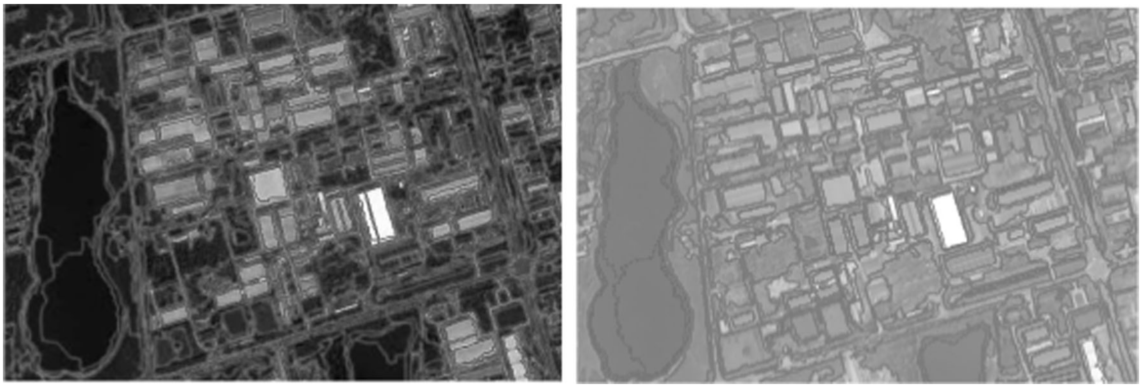

Fig. 7 Segmentation results with edge information added

Funding This paper is funded by National Natural Science Foundation of China under Grant No. 62071136.

Data availability No data were used to support this study.

Declarations

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

References

1. Liu Y, Ren Q, Geng J et al (2018) Efficient patch-wise semantic segmentation for large-scale remote sensing images. Sensors 18(10):3232
2. Zhongbin Su, Li W, Ma Z, Gao R (2022) An improved U-Net method for the semantic segmentation of remote sensing images. Appl Intell 52(3):3276–3288
3. Moradkhani K, Fathi A (2022) Segmentation of waterbodies in remote sensing images using deep stacked ensemble model. Appl Soft Comput 124:109038
4. Kemker R, Salvaggio C, Kanan C (2018) Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. ISPRS J Photogramm Remote Sens 145:60–77
5. Huang Y, Wang Q, Jia W et al. (2019) See More Than Once–Kernel-Sharing Atrous Convolution for Semantic Segmentation. arXiv preprint: arXiv:1908.09443
6. Ogohara K, Gichu R (2022) Automated segmentation of textured dust storms on mars remote sensing images using an encoder-decoder type convolutional neural network. Comput Geosci 160:105043
7. Panboonyuen T, Jitkajornwanich K, Lawawirojwong S et al (2019) Semantic segmentation on remotely sensed images using an enhanced global convolutional network with channel attention and domain specific transfer learning. Remote Sens 11(1):83
8. Zhang Z, Huang J, Jiang T et al (2020) Semantic segmentation of very high-resolution remote sensing image based on multiple band combinations and patchwise scene analysis. J Appl Remote Sens 14(1):016502
9. Liu Y, Shen C, Yu C et al. (2020) Efficient Semantic Video Segmentation with Per-frame Inference. arXiv preprint: arXiv:2002.11433
10. Jamali-Rad H, Szabo A, Presutto M. (2020) Lookahead adversarial semantic segmentation. arXiv preprint: arXiv:2006.11227
11. Alam M, Wang J-F, Cong G, Lv Y, Chen Y (2021) Convolutional neural network for the semantic segmentation of remote sensing images. Mob Networks Appl 26(1):200–215
12. Dong R, Pan X, Li F (2019) DenseU-net-based semantic segmentation of small objects in urban remote sensing images. IEEE Access 7:65347–65356
13. Ding L, Zhang J, Bruzzone L (2020) Semantic segmentation of large-size vhr remote sensing images using a two-stage multiscale training architecture. IEEE Trans Geosci Remote Sens 58(8):5367–5376
14. Mohammadimanesh F, Salehi B, Mahdianpari M et al (2019) A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem. ISPRS J Photogramm Remote Sens 151:223–236
15. Jacquemart D, Mandin JY, Dana V et al (2021) A multispectrum fitting procedure to deduce molecular line parameters: application to the 3–0 band of 12C16O[J]. Eur Phys J D 14(1):55–69
16. Guo Z, Wu G, Song X et al (2019) Super-resolution integrated building semantic segmentation for multi-source remote sensing imagery. IEEE Access 7:99381–99397
17. Diakogiannis FI, Waldner F, Caccetta P et al (2020) Resunet-a: a deep learning framework for semantic segmentation of remotely sensed data. ISPRS J Photogramm Remote Sens 162:94–114
18. Wang L, Li D, Zhu Y et al. (2020) Dual super-resolution learning for semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3774–3783
19. Wang Z, Tang Z, Li Y et al. (2020) GSTO: Gated scale-transfer operation for multi-scale feature learning in pixel labeling. arXiv preprint: arXiv:2005.13363
20. Sun K, Zhao Y, Jiang B et al. (2019) High-resolution representations for labeling pixels and regions. arXiv preprint: arXiv:1904.04514
21. M Rustowicz R, Cheong R, Wang L et al. (2019) Semantic segmentation of crop type in africa: A novel dataset and analysis of deep learning methods. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 75–82
22. Zheng Xu, Kamruzzaman MM, Shi J (2022) Method of generating face image based on text description of generating adversarial network. J Electron Imag 31(5):051411
23. Venugopal N (2020) Automatic semantic segmentation with deeplab dilated learning network for change detection in remote sensing images. Neural Process Lett, pp 1–23
24. Yang H, Yu B, Luo J et al (2019) Semantic segmentation of high spatial resolution images with deep neural networks. GIScience Remote Sens 56(5):749–768
25. Broni-Bediako C, Murata Y, Mormille LH et al. (2021) Evolutionary NAS for aerial image segmentation with gene expression programming of cellular encoding. Neural Comput Appl.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.