Land Cover Classification Using High Resolution Satellite Image Based on Deep Learning

Ming Zhu [1, 2, *], Bo Wu [2], Yongning He [2], Yuqing He [2]

[1] Institute of Geoscience and Resources, China University of Geosciences, Beijing, 100083, China - zhuming@cugb.edu.cn
[2] Geographic Information Center of Guangxi, Nanning, 530023 China

KEY WORDS: Deep Learning, Convolutional Neural Networks, Land Cover Classification, High Resolution Satellite Image, Semantic Segmentation

ABSTRACT:
In the coming era of big data，the high resolution satellite image plays an important role in providing a rich source of information for a variety of applications. Land cover classification is a major field of remote sensing application. The main task of land cover classification is to divide the pixels or regions in remote sensing imagery into several categories according to application requirements. Recently, machine interpretation methods including artificial neural network and decision tree are developing rapidly with certain fruits achieved. Compared with traditional methods, deep learning is completely data-driven, which can automatically find the best ways to extract land cover features through high resolution satellite image.
This study presents a detailed investigation of convolutional neural networks for the classification of complex land cover classes using high resolution satellite image. The main contributions of this paper are as follows: (1)Aiming at the uneven spatial distribution of surface coverage, we study the training errors caused by this uneven distribution. An improved SMOTE algorithm is designed for automatic processing the task of sample augmentation. Through experimental verification, the improver algorithm can increase 2-5% classification accuracy by the same network structure. (2)The main representations of the network are also shared between the edge loss reinforced structures and semantic segmentation, which means that the CNN simultaneously achieves semantic segmentation by edge detection.(3)We use Beijing-2 satellite (BJ-2) remote sensing data for training and evaluation with Integrated Model, and the total accuracy reaches 89.6%.

0 Introduction

Land cover classification is the basis for monitoring land cover change and further studying land resource management and ecological environment change. With the development of high-resolution remote sensing satellite and its application in China, it is possible now to use these satellites with favorable characteristics of high spatial resolution and short revisit period to carry out land cover classification, which, being helpful in the conducting of a land use survey or land spatial database updating, is of great significance to geographic condition monitoring and digital city construction(Feng L,2014).
In recent years, a series of breakthroughs have been made on Deep Convolutional Neural Network (DCNN) in various fields, such as image classification(He K M,2016), target detection (Girshick R,2014), image semantics segmentation (Long J,2015) and facial recognition(Parkhi O M,2015). Compared with traditional classification methods, DCNN has stronger ability of feature learning and expression. Thus, it has become a hotspot in the research of land cover classification based on remote sensing images. Aiming at the difficulty of getting accurate identification of rice area in complex surface landscape area through remote sensing, Zhao S(Zhao S,2018) adopted the strategy of hierarchical classification. Based on the preliminary classification of remote sensing images by using Convolutional Neural Network (CNN) with pre-training mechanism, the precise identification of rice information was realized by combining phenological information. This method combines time feature and deep abstract feature, and the accuracy of rice area recognition is improved to 90%. Using support vector machine as classifier, Zhang W (Zhang W,2017) classified the multispectral images of 16-meter spatial resolution taken by WFV camera of GF-1 satellite. Three different DCNN models are introduced and analysis on features of different layers of DCNN and effects of the size of neighboring window for feature extraction on the classification results are analyzed. Jianhao Tai(Tai J H,2017) proposed a high-resolution remote sensing image classification method based on FCN, constructed an overall framework of the FCN-based classification method, introduced the classification process of the method in detail, and focused on the sample preparation, model training and network parameters setting.
The spatial distribution of land cover is uneven. So how to effectively improve the classification accuracy of weak land types in remote sensing images, and how to improve the training efficiency by using multi-model integrated method are questions to be answered. However little research has been done concerning these questions at present. In view of this situation, this paper, taking advantage of the features of high-solution satellite images, proposes an application scope of geometric enhancement and pixel transformation enhancement and augments the sample data by improving the SMOTE augmentation and screening algorithm. The relative balance of various types of samples is ensured, and the classification accuracy can be improved. Secondly, this paper proposes a new convolutional neural network for land cover classification based on high-resolution satellite images, which is able to operate input images of any size. Thirdly, an integrated algorithm of heterogeneous models is proposed to improve the training efficiency and accuracy of land cover classification.

1. Data Augmentation

Data augmentation is an important means to expand the training sample data when the number of training samples is insufficient. The data is expanded without changing the label category and the generalization ability of the neural network to be trained is improved. Image data augmentation includes geometric augmentation of pixel coordinates and numerical augmentation of pixel values. Geometric augmentation includes translation, distortion, rotation, cropping, flipping, scaling and so on. Digital augmentation includes color transformation, random noise, saturation, brightness and so on. In the process of data augmentation, one or several of these methods are usually used to expand data.

As a special image, the surface classification sample data can not be randomly augmented in the process of data augmentation. The augmented data should be close to or conform to the corresponding features of real objects in texture and geometry. In the process of augmentation processing, a reasonable interval range must be specified in data transformation in order to form an effective augmented data set. The main methods of sample expansion are geometric transformation and pixel value transformation.

1.1 Data Augmentation Through Geometric Transformation

The data augmentation through geometric transformation is realized mainly by the following methods, such as flip, rotation, scaling, shear and affine transformation. Different transformations can be achieved through different transformation matrices.

$$[x \quad y \quad 1] = [x_0 \quad y_0 \quad 1]T_{Transformation\ Matrix} \quad (1)$$

x_0 and y_0 are coordinates of the original image. x_0 and y_0 are coordinates of the augmented data image. T is a transformation matrix for different purposes. The main transformation matrices used in the augmentation process is as follows.

$$T_{rotation} = \begin{bmatrix} \cos\beta & \sin\beta & 0 \\ -\sin\beta & \sin\beta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$T_{scaling} = \begin{bmatrix} k_x & 0 & 0 \\ 0 & k_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad T_{flip\ horizontal} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$T_{flip\ vertical} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} T_{shear} = \begin{bmatrix} 1 & a_1 & 0 \\ a_2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$T_{affine\ transformation} = \begin{bmatrix} a_1 & a_2 & 0 \\ a_3 & a_4 & 0 \\ a_5 & a_6 & 1 \end{bmatrix} \quad (2)$$

$T_{rotation}$is used to achieve rotation transformation of polygon. By random rotation of β, the rotation range being 0-180 degrees,the classification map in different directions is simulated. $T_{scaling}$ is to scale images according to specified scaling coefficient. k_x and k_y are respectively the scaling coefficients of direction x and direction y. This matrix is applied to simulate classification polygons of all sizes. To take the actual area of polygons on map into consideration and to ensure no loss of corresponding information in neural network encoding, the minimum coefficient shall be no less than the scaling ratios of down-sampling network. $T_{flip\ horizontal}$ and $T_{flip\ vertical}$ are used to augment semantic samples of different textures, while T_{shear} and $T_{affine\ transformation}$ are used to obtain complex transformation of polygons by setting the value of a_i.

The distortion transformation mainly uses sinusoidal distortion. The use of distortion transformation augmentation aims mainly to effectively increase the number of samples of line-type polygon, such as, rivers and roads, and to simulate different distortion patterns of such polygon. Generally, distortion transformation is not used to augment data for structures. The transformation formula is as follows.

$$T_{\text{sinusoidal distortion}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & k\sin(\omega x_0) & 1 \end{bmatrix} \quad (3)$$

In the formula, k is the accommodation coefficient for amplitude and ω is the accommodation coefficient for frequency. In order to make the distortion more natural, the values of k and ω is limited, that is, $k \in [0,2]$ and $\omega \in [0,2]$. The seven transformation matrices shown in (2) and (3) can be combined to form new geometric variations. However, in principle, a picture can only undergo three image transformations at most in order to avoid the possible distortion of the original image in form and texture.

Bilinear interpolation algorithm is used to fill the void in images that has gone through geometric transformation. The interpolation formula is as follows.

$$C = k[x_2 - x \quad x - x_1]\begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix}\begin{bmatrix} y_2 - y \\ y - y_1 \end{bmatrix},$$

$$where\ k = \frac{1}{(x_2 - x_1)(y_2 - y_1)} \quad (4)$$

In the formula, $C_1(x_1,y_1)$, $C_2(x_2,y_1)$, $C_3(x_1,y_2)$ and $C_4(x_2,y_2)$ are pixel values of the four neighboring points around $C(x,y)$. The bilinear interpolation is performed and the interpolated value of C is calculated separately according to each channel of sample data, and the void is thus filled. The effect of geometric augmentation is shown in Fig. 1.

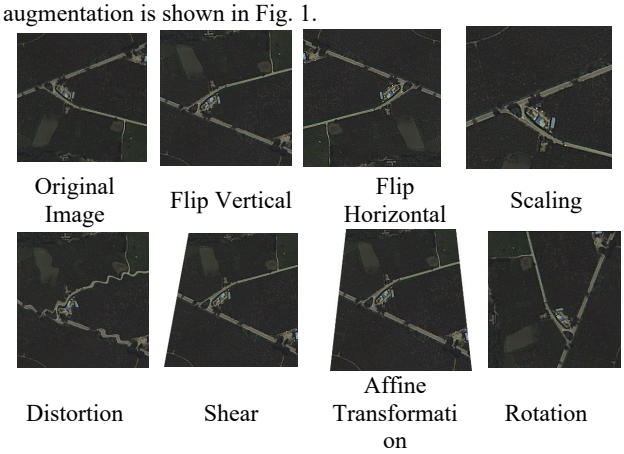

Fig.1 Effect of Geometric Augmentation

The augmented data that meets the above conditions can effectively simulate various forms of land types, but the ultimate goal of data augmentation is to strengthen the generalization ability of deep learning model and improve the accuracy of semantic segmentation of the deep learning model. Therefore, the augmented data must be screened. In the process of augmentation, priority should be given to expanding samples that can not be effectively recognized by the deep learning model. For samples that can be well recognized by the deep learning model, the augmentation of such samples should be reduced. Therefore, after the original sample generates the augmented samples, it is necessary to discriminate and screen the augmented samples, select the augmented samples which are more conducive to the expansion of the generalization ability, and increase the number of such augmented samples.

1.2 An Improved Screening Method for SMOTE Augmentation

The screening of augmented samples is mainly based on accuracy of semantic segmentation, that is IoU. Before the screening, the original samples are used to train the deep learning model. And then the screening begins after the training is completed. The single augmented sample is input into the deep learning model one by one. IoU is calculated on the predicted polygons outputted by each augmented sample. When the IoU of a polygon is below the selected threshold, the augmented sample passes through the screening. And then a number of samples would be generated by using the same transformation method with different random parameters. The generation algorithm is as follows.

The SMOTE method(Dina Elreedy,Amir F,2019) based on interpolation is used to synthesizes new samples for small

sample class. The main idea of this method is as follows.

(1) Define the feature space. And then correspond each sample to a point in the feature space and determine the sampling rate N according to the unbalanced proportion of samples.

(2) For each sample(x, y) from small sample class, K nearest neighbor samples are selected according to Euclidean distance, among which a sample point is randomly selected and assumed to be (x_n, y_n). A random point is selected on the line segment between sample points and its nearest neighbor sample points in the feature space as a new sample point, which shall be done according to the following formula.

$$(x',y') = (x,y) + rand(0-1) * ((x_n - x),(y_n - y)) \quad (5)$$

The SMOTE method is extended to the vector space of samples, and the eigenvector of an augmented sample is defined as v= (rotation angle, distortion, brightness, R, G, B, contrast ratio). When an augmented sample is generated, its validity is judged. If the sample is valid, its eigenvector is used as a template, based on which, the vector scalar is randomly adjusted to automatically generate new effective augmented data. The algorithm is shown as follows.

Algorithm 1: Algorithm for Screening Augmented Samples

Input:

Training Set $T = \{(x_1,y_1),\cdots,(x_n,y_n)\}$

Augmented Sample Set $T' = \{(x'_1,y'_1),\cdots,(x'_n,y'_n)\}$

Steps:

1. for i=0 to n do:
2. For an augmented sample (x'_i,y'_i), input the model and compute the IoU of the augmented class.
3. if IoU<I do:
4. Comparing IoU with threshold I, if IoU<I, merge (x'_i,y'_i) into training set T.
5. Obtaining training set T with Augmented Samples.

2 .Deep Learning Model for Land Cover

2.1 Design of Convolutional Neural Network

As a core part of the deep learning model for land cover, the convolutional neural network is designed to accomplish the semantic segmentation of images with high accuracy and to further improve the classification accuracy by connecting, connecting in series and connecting in parallel with other network parts.

The convolutional neural network must achieve the following design objectives:

(1) The objects to be processed are mainly images shot by 0.8m high resolution satellites, such as GF- 2 and BJ-2 satellites.

(2) An end-to-end processing of semantic segmentation shall be realized. Data of images of any size can be inputted and processed. Besides data preprocessing and necessary post-processing, there shall be no need for manual intervention in the process of land cover classification.

(3) The smallest polygon that it is able to divide shall be greater than 16 pixels. Better classification accuracy shall be achieved, with an overall accuracy of large-scale land cover classification being above 90%.

According to the requirements put in the design objectives, in order to achieve processing of images of any size, the deep learning model can not contain full connection layer and all operations are completed by convolution. In order to achieve the polygon segmentation ability of greater than 16 pixels, it is necessary to reduce the shallow information loss in the process of information extraction when designing the deep learning model. In principle, the minimum feature map in the coding stage must be larger than 1/8 of the original size, so the number of classical pooling layer must be reduced as much as possible. But in the meantime, other network structures with the same function shall be used to replace some of the classical pooling layers. In order to have a high processing efficiency, it is necessary choose a structure with fewer parameters for the deep learning model to reduce the amount of system operation. Based on the above principles, a deep learning model is designed. And the overall architecture of it is shown in Fig. 2.

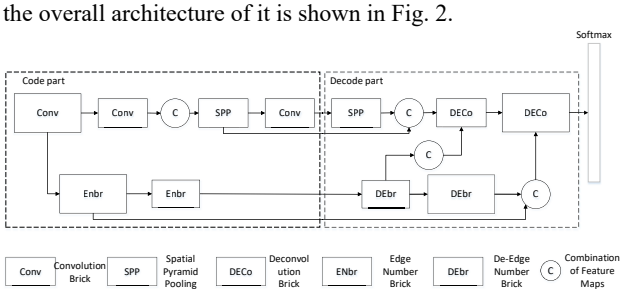

Fig.2 Overall Architecture of the Deep Learning Model

2.2 Multi-model Integration Algorithm for Classification

Based on the classification results of multiple independent models, the multi-model land cover classification method is to fuse the results of each model according to a certain algorithm and determine the final classification. The main logic structure of the multi-model classification method is shown in Fig. 3. Module A is the predicted output module of each model and module B is the multi-model result integration module constructed by a certain algorithm. And the integrated result is output to be the final classification result. Among the three main modules, the result integration module is the most important one, and is seen as the core of the structure for that the ability of the integration algorithm directly determines the classification accuracy of the whole integrated learning architecture.

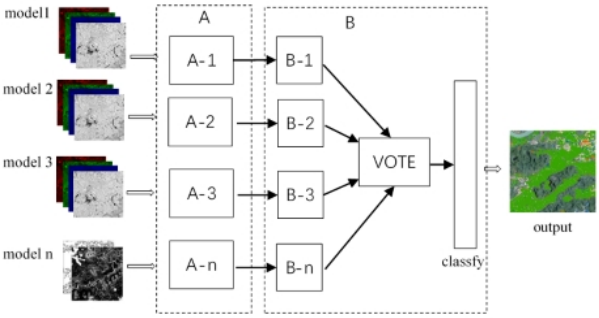

Fig. 3 Structure of the Multi-model Integration Classification Method

The general idea of the multi-model integration algorithm is put as follows. The integrated model contains several N learning classifiers, that is, M=(N_1, N_2,……, N_n). For a sample containing i classifications, the result vector of each classification algorithm is N=(a_1, a_2,……, a_i) . And then the algorithm f = merge (N_1,N_2,……, N_n) is given, which maximizes the probability of the corresponding result output.

From the idea of multi-model integration, it can be concluded that the key to model integration is the classifiers that are integrated and the final integration algorithm. Classifiers can be integrated by means of homogeneity and heterogeneity to form an integration model. A homogeneous model is an integrated model synthesized by several classifiers that are based on the same classification algorithm but uses discrepant training data.

A heterogeneous model is the integration of n classifiers with different classification algorithms. The integration algorithm is designed to make the integrated model acquire higher accuracy and generalization performance.

2.3 The Model Integration Algorithm

The model integration algorithm is a multi-order integration algorithm, which can be divided into two stages in terms of function. The initial stage of the model is composed of several heterogeneous and unrelated primary learning models, such as $N_1,N_2,N_3,\cdots,N_k$ and so on. In the initial stage, the training dataset $T=\{(x_1,y_1),\cdots,(x_n,y_n)\}$ is used to train the primary learners respectively. After the training is completed, a set of primary learners is formed. In this paper, 8 models, including heterogeneous model convolutional neural network, FCN16s、AttU_Net and DenseASPP, are used as primary learners. And these heterogeneous models are trained separately using the training set of balanced classes after data augmentation.

After training, the training samples are input again into the 8 heterogeneous primary learners at the same time. Each classification outputs one result. 8 results are regarded as an output vector Z, which is used as a training data set for high-order integrator, that is, $Z=\{(z_1,y_1),\cdots,(z_n,y_n)\}$. The trained integrator is connected with the primary learner to receive the output of the primary learners and complete the classification.

Algorithm No.2: Heterogeneous Models Integration Algorithm

Input:
Training Set: $T=\{(x_1,y_1),\cdots,(x_n,y_n)\}$
Heterogeneous Models: $N_1,N_2,N_3,\cdots,N_k$
Integrator: S
Steps:
1. for i=0 to n do:
2. Use sample (x_n,y_n) to train heterogeneous models separately:$N_1,N_2,N_3,\cdots,N_k$, $Z_i=\{N_1(x_i,y_i),\cdots,N_k(x_i,y_i),\}$ is output.
3. Z_i add to set $Z=\{(z_1,y_1),\cdots,(z_n,y_n)\}$.
4. Use training set $Z=\{(z_1,y_1),\cdots,(z_n,y_n)\}$ to train the integrator S
5. Integrate S with heterogeneous integrators.

Algorithms 2 uses a stacking integration method. In the specific integration process, the learning algorithm of integrator S chooses fully connected layer to realize connection according to the multi-classification tasks of land cover classification. As the number of integrated models in reality is usually not too large, the use of fully connected network has little impact on the overall efficiency of the system.

3. Experiment and Analysis

In the experiment, we selected data from BJ-2 satellite for training. The whole area covers about 4,200 square kilometers, just as shown in the following figure.

Figure 4 BJ-2 Image Data of the Experimental Site

The distribution proportion of main land types in the experimental area is shown in the table below. There are great differences in the proportion of land cover types.

Table 1 Distribution Proportions of Main Land Types in the Experimental Site

	Water	Construction	Hardened Ground	Forestland/ Grassland
Distribution Proportion	3.90%	8.03%	5.69%	31.85%
	Garden Plot	**Dryland**	**Paddy Field**	**Bare Land**
Distribution Proportion	23.78%	19.62%	4.71%	2.43%

3.1 Comparisons Between the Improved SMOTE Algorithm and the Traditional SMOTE Algorithm

It can be seen from Table 2 that after the improved SMOTE algorithm screens the augmented data to achieve classification balance, the overall gap between the various classifications is controlled within 2%, which means that the data is relatively balanced.

Table 2 Distribution Proportions of Land Types in All Test Groups

	Water	**Construction**	**Impervious Surface**	**Forestland/ Grassland**
Original Data	3.90%	8.03%	5.69%	31.85%
Under sampling	11.77%	12.03%	12.61%	13.85%
Over Sampling	12.84%	12.25%	11.89%	12.96%
	Garden	**Dryland**	**Paddy Field**	**Bare Land**
Original Data	23.78%	19.62%	4.71%	2.43%
Under sampling	12.78%	13.42%	12.11%	11.43%
Over Sampling	13.13%	12.04%	12.38%	12.51%

As shown in Fig. (5), contrast experiments were done between the traditional SMOTE algorithm and the improved SMOTE algorithm on accuracy and efficiency. The experimental data shows that the improved SMOTE augmentation algorithm has better accuracy than the traditional algorithm when using training set of the same size, indicating that the augmented data is deficient without SMOTE algorithm, with invalid or inefficient augmented data mixed in. As the improved algorithm limits the effective range and filters the invalid augmented data. Therefore, the augmented data can effectively expand the feature space of the classifications and effectively compensate for the accuracy loss caused by data imbalance. However, due to the addition of the screening step, the efficiency of data expansion decreases, and it takes longer to complete the expansion of data of the same scale compared with the traditional algorithm.

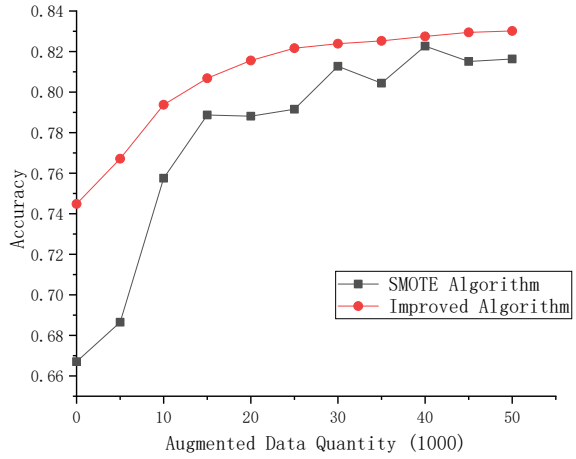

Fig. 5 Contrast Between Traditional SMOTE Algorithm and the Improved SMOTE Algorithm

In addition, in order to find out the effect of the volume of augmented data on the accuracy of algorithm, a training set of augmented data from 5000 to 50000 was designed as a comparative test object.

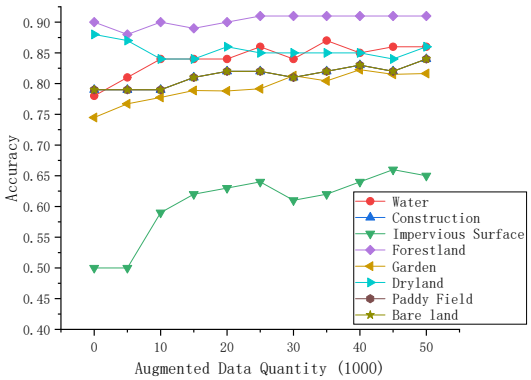

Fig. 6 Change of Accuracy with Different Volume of Augmented Data

It can be seen from the graph that the effect on accuracy is the best when the volume of data augmentation reaches about 30,000. Data augmentation of more than 30,000 sheets has less effect on accuracy improvement. Along with the increase of volume, the accuracy of unbalanced land types is improved greatly. Accuracy of the advantaged land types that is already easy to be classified before the data augmentation is improved in a relatively small scale. This shows that the data augmentation effectively improves the classification accuracy of weak land types.

3.2 Result of the Classification Experiment

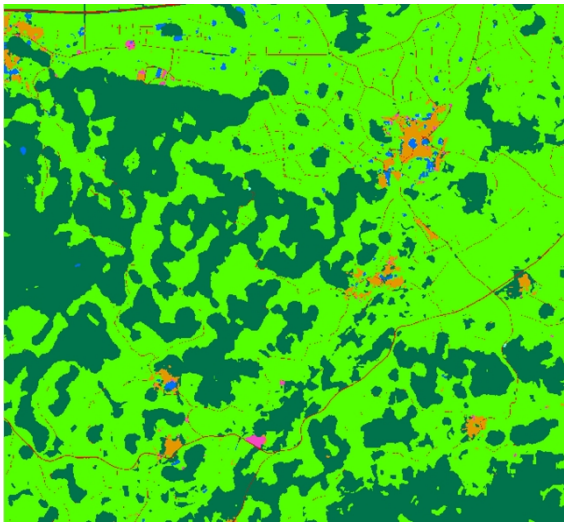

Fig. 7 Result of Classification

8 models, including heterogeneous model convolutional neural network, FCN16s、 AttU_Net and DenseASPP, are used as primary learners. And the 8 heterogeneous models are trained separately using the training set of balanced classes after data augmentation. which is shown in the following table.

Table 3 Contrast on the Classification Accuracy of Land Cover

	acc	iou	f1	recall
AttU_Net	0.78771	0.463661	0.590089	0.607812
BiSeNet	0.72796	0.396862	0.529237	0.525610
DANet	0.73493	0.401882	0.533852	0.565942
DenseASPP	0.78287	0.439216	0.564753	0.603061
FCN16s	0.80739	0.480383	0.602365	0.600168116
NestedUNet	0.77022	0.450123	0.578265	0.5881124
PSPNet	0.73947	0.424897	0.557446	0.542100369
Integrated Model	0.89601	0.531859	0.660428	0.670872

4 Conclutions

Firstly, the improved SMOTE augmentation algorithm has better accuracy of data augmentation. The augmented data can effectively expand the feature space of classifications and effectively compensate for the accuracy loss due to data

imbalance. Secondly, the number of data augmentation has a certain impact on the classification results. Augmented data of about 30,000 pieces is the most effective for classification and plays a positive role in improving the classification accuracy of weak land types. Thirdly, heterogeneous model integration algorithm can achieve the results integration and classification output of primary learners through a learnable integrator. The learnable integrator is driven entirely by the features of the land cover data and the efficiency of each primary classifier. This avoids a manual design of voting weights and enables the whole classifier to have good flexibility and generalization. Compared with the traditional classification method based on support vector machine, the classification method proposed in this paper can achieve a higher classification accuracy of land cover.

References

Feng L. Change detection of urban typical land coverage based on high-resolution satellite remote sensing[D]. Northeastern University, 2014.

He K M,Zhang X Y,Ren S Q,et a1．Deep residual learning for image recognition[C]//Proceedings of2016 IEEE Conference Computer Vision and Pattern Recognition．Las Vegas, Nevada, USA：IEEE,2016：770-778．

Girshick R,Donahue J,Darrell T,et a1．Rich feature hierarchies for accurate object detection and semantic segmentation[C] //Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition．Columbus,OH,USA：IEEE,2014：580-587．

Long J,Shelhamer E,Darrell T．FuHy convolutional networks for semantic segmentation[C] //Proceedings of 2015 IEEE Confer．ence on Computer Vision and Pattern Recognition．Boston,Massachusetts,USA：IEEE,2015：3431-3440．

Parkhi O M,Vedaldi A,Zisserman A．Deep face recognition[C]//Proceedings of 2015 British Machine Vision Conference．Swansea,UK：BMVA,2015：41．141．12．[DOI：10．5244/C．29．41]

Zhao S. Precise mapping and spatiotemporal analysis of p addy rice area in complex surface landscapes[D]. China U niversity of Geosiences, 2018.

Zhang W. Land cover classification with extracted deep fe atures of deep convolutional neural network[D]. Institute o f Remote Sensing and Digital Earth, Chinese Academy of Sciences, 2017.

Tai J H. Research on the application of deep learning in remote sensing image target detection and land cover classification[D]. Wuhan University, 2017.

Dina Elreedy,Amir F. Atiya. A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance[J]. Information Sciences,2019,505.