

Received April 12, 2022, accepted May 9, 2022, date of publication May 18, 2022, date of current version June 1, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3175978

Land Cover Classification of Resources Survey Remote Sensing Images Based on Segmentation Model

ZHENYU FAN^{1,2}, (Member, IEEE), TAO ZHAN³, ZHICHAO GAO³, RUI LI^{1,2}, YAO LIU⁴, LIANZHI ZHANG⁴, (Member, IEEE), ZIXIANG JIN⁴, AND SUPENG XU⁴

¹China Aero Geophysical Survey and Remote Sensing Center for Natural Resources, Beijing 100083, China

²Key Laboratory of Airborne Geophysics and Remote Sensing Geology, Ministry of Nature and Resources, Beijing 100083, China

³Institute of Ecological Geology Survey and Research of Heilongjiang Province, Harbin 150030, China

⁴School of Geophysics and Information Technology, China University of Geosciences (Beijing), Beijing 100083, China

Corresponding author: Tao Zhan (hljzhantao@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 41888101, in part by the Eco-Geology Survey Research Institute of Heilongjiang Province, and in part by the China Aero Geophysical Survey and Remote Sensing Center for Natural Resources (AGRS).

ABSTRACT Land type survey is an important task of land resources survey and the basis of scientific management of land resources. With the increasingly prominent problems of population, resources, and environment, there is an urgent need for a fast and accurate classification method of large-scale land use and land cover based on remote sensing data. Traditional machine learning classification methods based on pixel classification achieved sufficient results and are widely used, such as maximum likelihood classification and random forests method. However, with the development of the novel technology of deep learning, in practical application, for multi-classified land resources, how to use the fast and effective classification method of low and medium resolution RS images needs further research. This paper takes the land resource classification of the Tonghe medium resolution RS dataset of the third land survey in China as an example to screen and compare traditional machine learning classification methods and semantic segmentation models FC-DenseNet56, GCN, BiSeNet, U-Net, DeepLabV3, AdapNet, and PSPNet, which aim to select the optimal feature extraction model. The results show that the classification accuracy of the U-Net model can reach 93.62%, which is more accurate and effective than traditional machine learning methods and other semantic segmentation models. It is suitable for multi-classification tasks of land cover resources in low and medium resolution RS images and shows a superior effect in practical application. Besides, the conclusion of this study can provide a demonstration for large-scale land cover resources investigation using low and medium resolution RS images.

INDEX TERMS Land use and land cover, semantic segmentation, multi-classification, deep learning, U-Net.

I. INTRODUCTION

Remote sensing (RS) data is widely used in land use and land cover (LULC) research, and automated image classification is one of the easiest and preferable techniques to prepare LULC of a resource survey area [1]. With the progress of RS satellite observation technology and the improvement of image acquisition convenience, pixel-based multi-classification of land resources has become an important research topic [2]. In the past few decades, the maximum

likelihood classification (MLC) and random forests (RF) of traditional machine learning are fast and feasible RS image classification methods for LULC, but they have high requirements for RS image quality [3]. While the deep learning (DL) technology improves the extraction ability of feature information for LULC classification, has a stronger learning ability and lower requirements for RS image resolution. In RS image recognition, DL can derive complicated, hierarchical, and non-linear features from data and overcome several limitations of traditional methods. Because of their modeling and learning capabilities, DL approaches represent the link between the image object and its real-world

The associate editor coordinating the review of this manuscript and approving it for publication was Jinjia Zhou^{1b}.

elements as closely as possible, allowing for more real-world information [4], [5]. In some situations, DL's classification effect and accuracy may be better than those of traditional machine learning classification methods. The recent research on the DL method mainly uses high resolution RS images for detailed classification. Although these methods can obtain an accurate land cover map, it needs to be explored to use the DL classification method and only rely on medium and low resolution RS images to realize large-scale land cover classification of land resources survey. In the actual LULC classification of resource survey, large-scale land cover resource survey usually manually marked the categories and boundaries of land resources concerning low and Medium Resolution Satellite RS images, which demand complex engineering skills and strongly rely on expert experience. Based on the above, these methods are not practicable for large-scale land cover resources surveys. To automate large-scale land cover resources survey, low and medium RS images and existing DL segmentation methods are applied to address the problems mentioned above.

According to the past LULC practical experience, the RS image classification method based on traditional machine learning is undoubtedly mature and stable. The most widely used methods of LULC are MLC and RF methods. The application of MLC in RS image classification is mainly to monitor the change of vegetation coverage [6], [7] and analyze the causes of vegetation change or damage [8], [9]. However, some comparative analysis results show that the accuracy of MLC is not high compared with other classification methods in some cases [10]. In contrast, the RF machine learning algorithm can achieve almost real-time monitoring and the effect is excellent [11]. Some scholars combine optical and radar imagery and use the RF method for classification [12]. To improve the accuracy of classification, most of them rely on enhancing the resolution of RS data or adding auxiliary information. Although there are many active and established new machine learning classification algorithms based on MLC and RF, it seems unlikely to achieve a breakthrough in these algorithms [13]. Especially in the case of low or medium resolution of RS images and lack of enough auxiliary information, the methods of DL classification may be more effective.

Based on DL classification and semantic segmentation methods, the use of RS image data to carry out resource investigation, application, and analysis of LULC has aroused great interest from researchers in the field of RS science. From the perspective of published articles in recent years, the research focus of scholars can be divided into two aspects as a whole. Many scholars try to improve the ability of classification, segmentation, or detection of the DL algorithm itself, to improve the accuracy of multi-objective classification, improve the effect of detail resolution, solve the problems of insufficient training samples, unbalanced labeling samples, and so on [14], [15]. At present, it is popular to improve based on classical convolutional neural network (CNN) and include

the information processing skills of attention mechanism in the design [16], [17]. Through the performance comparison of experimental metrics, it is proved that the designed new model has advantages [18]–[20]. On the other hand, the application research of LULC which focuses on the combination of DL methods to solve the actual problems of land resources science is less. In fact, the research on the DL algorithm, including model improvement and redesign, does improve the ability of automatic classification of RS images, but most of them are comparative experiments carried out with high resolution public datasets. In many cases, the large-scale LULC resource survey carried out by the government only distributes low and medium resolution satellite RS data, which is more complex and more challenging. The slight accuracy improved by the algorithm research itself is of little significance to the development of actual jobs. In the process of this land resources survey in part of Heilongjiang Province, the LULC work still refers to the RS data to manually mark the category and boundary of land resources. Although expert experience and field investigation and evidence collection can accurately depict the boundary of land resource categories, it is obvious that completely relying on a manual LULC classification map will greatly increase the workload and reduce the work efficiency. Therefore, how to use the existing classical DL semantic segmentation model to solve the LULC of low and medium resolution RS images has become more concern for many geoscientists.

Up to now, there are a variety of classical approaches have been proposed for scene classification. The most commonly used approach is CNN, it has been widely applied to RS image semantic segmentation. Since Ronneberger proposed the U-Net CNN model [21], because the network and training strategy rely on data enhancement to use the available annotation samples more effectively, the improved networks based on it has been widely used in images segmentation and classification [22]–[24]. Zhao *et al.* [25] exploit the capability of global context information by different region-based context aggregation through the pyramid pooling module together with the proposed pyramid scene parsing network (PSPNet), it is well established and has been shown to perform well in image recognition [26]–[28]. Kipf and Welling proposed a semi-supervised learning method based on CNN, which operates the graphics directly [29], and the graph convolutional network (GCN) is an extension of CNN. Compared with the classical CNN method, GCN is considered to be more effective in learning the feature representation of graph-structure data [30]. To improve the performance of semantic segmentation models, researchers have designed a series of CNN algorithm frameworks in recent years, including AdapNet [31], [32], BiSeNet [33], DeepLabV3 [34], and FC-DenseNet [35]. AdapNet architecture is considered to have better performance than other end-to-end semantic segmentation networks. AdapNet is designed to run efficiently on platforms with dynamic resource allocation, which is an adaptive runtime model for executing streaming applications

on multiprocessor systems [36], while BiSeNet is proposed to improve the speed and accuracy of real-time semantic segmentation.

In these CNN networks, AdapNet and BiSeNet are rarely used in RS image classification and recognition, while DeepLabV3 and FC-DenseNet are usually regarded as stable standard CNN models to compare and verify the CNN frameworks designed by researchers. For instance, Zhang *et al.* Proposed an improved algorithm for the high spatial resolution [37] and compared it with DeepLabV3 to prove the improvement of the accuracy of the method. Lin *et al.* proposed a DL model based on the structure of DeepLabV3 and compared the experimental results with FC-DenseNet to prove the improvement of segmentation accuracy [38]. Li *et al.* developed a deep adversarial network, and the generator produced pixel-wise image classification maps using the FC-DenseNet model [39]. Gao *et al.* proposed a semantic segmentation network based on the FC-DenseNet to recognize and map the landslide disasters [40]. Guo *et al.* proposed a deep conditional generative adversarial network integrating the FC-DenseNet for the application of high resolution RS images [41]. For purpose of improving the semantic segmentation performance of land cover in high spatial resolution satellite RS images and obtaining a pixel-wise classification of land cover, combined with the advantages of DenseNet and U-Net, Khan *et al.* proposed a hybrid DL model [42], which demonstrate the most advanced performance on public datasets. Obviously, most of the current proposed DL methods for pixel-based image segmentation and classification in the field of RS are based on the improvement of the above classical network models. The reason why the above semantic segmentation model applies to LULC can be attributed to the following points. Different from the artificial identification features based on specific domain knowledge, they have deep feature extraction neural networks, usually dozens to hundreds of layers. Moreover, many experiments have proved that the computational efficiency of the DL pixel-level segmentation model is very high, and the ability of CNN to learn features is very effective in large-scale target detection and image recognition. Most importantly, the LULC algorithms based on the semantic segmentation of these CNN models are more representative, which realize a series of improvements such as model feature extraction, upsampling and feature fusion.

More and more researchers begin to pay attention to the application of the DL methods in LULC resources surveys. For example, people study the application of generating classified maps of LULC changes over time [43]. Some researchers also use urban hyperspectral RS images for feature extraction and land cover classification [44]. Some people proposed a weakly towards strongly supervised Learning framework for land cover classification to adapt insufficient datasets [45]. Usually, using large-scale and high resolution RS images and an optimized network can get more accurate classification results [46], [47]. However, the above research needs high resolution RS images or hyperspectral RS images.

At present, the national resource survey is still unable to meet the distribution of such large-scale RS satellite data. At present, the large-scale LULC resources survey mainly depends on low and medium resolution RS images. The research on the resources classification methods and implementation process of low and medium resolution RS images are more in line with the current needs. For the application of actual RS data, the performance of CNN with different structures is not alike. It is worth exploring and discussing how to utilize the existing segmentation or classification models to solve practical automating large-scale LULC problems. To further improve the performance of the RS scene classification, we should compare the metrics of the above classical networks. Based on the analysis of their ability to segment and classify, a more appropriate classification model should be selected for our medium and low resolution RS datasets. In general, the major contributions of this paper are as follows:

1. The current situation of semantic segmentation and classification of LULC resources and the limitations of manual labeling and traditional machine learning classification methods are discussed. Then it expounds on the necessity of using the DL segmentation model to segment and classify medium and low resolution RS images in the process of large-scale land resources surveys nationwide, mainly referring to medium and low resolution RS images.

2. Taking the Tonghe RS image of China's third land and resources survey as an example, the data of low and medium resolution RS images are preprocessed, and the results of manual classification are verified. MLC and RF methods of traditional machine learning are used to segment and classify the LULC resources, and the results are briefly analyzed.

3. In order to select the segmentation model suitable for our low and medium resolution RS images, we systematically compare the performance of 7 classical segmentation models to obtain the optimal model.

4. Finally, the experimental results in the test area prove the effectiveness and practicability of our selected segmentation model. This paper focuses on the practical application of low and medium resolution satellite RS LULC. The experimental conclusions can provide a demonstration for the future large-scale LULC resources investigation.

II. BACKGROUND

A land survey is an important means to find out the distribution of land resources. Mastering the current situation and changes in land use, including land types, location, area, and distribution is the main content of China's third land resources survey, in which the category and distribution of farmland and forest are the key content. The RS database in this paper is a part of the land resources survey database in China. This task is based on the purpose of improving the basic data on land use and mastering the current situation of land-use changes in natural resources.

In this section, we briefly introduce the pre-processing and statistics of RS image data, as well as the field verification in

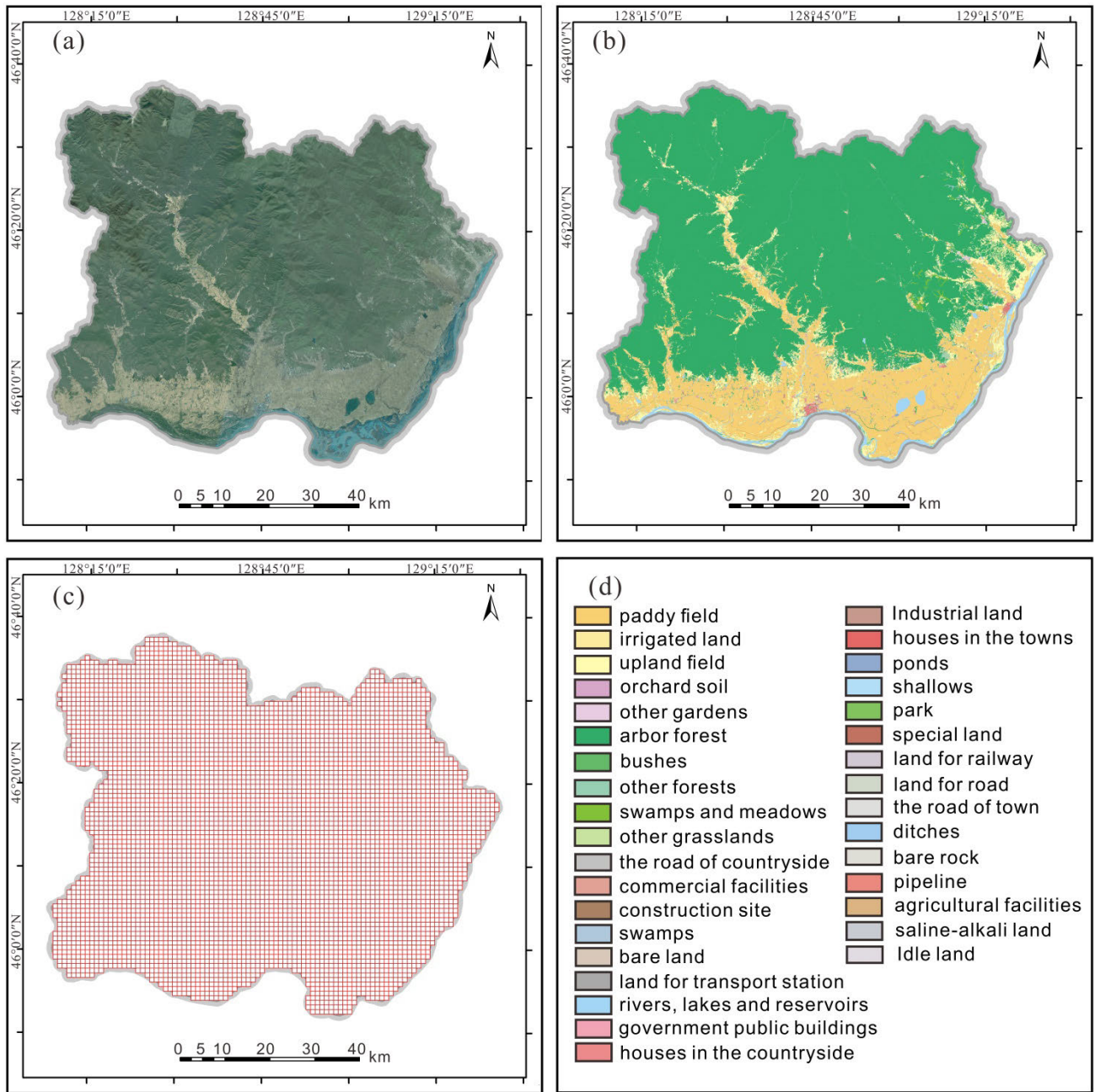


FIGURE 1. Remote sensing data and manual labeling land cover resources classification in Tonghe county. (a) Satellite RS data. (b) The overall artificial map of land cover resources. (c) Fishing net clipping diagram of RS image data and manual annotation data. (d) Types of land-cover resources in the study area marked manually.

the land resources survey. This is the premise of researching the LULC method based on the DL segmentation and classification model.

A. REMOTE SENSING DATABASE

The study area, as shown in Figure 1, is located in northeast China. The RS data is gathered from the third land survey database of Heilongjiang Province, it covers the whole area of the Tonghe County in Heilongjiang Province, located

between 128°09' E - 129°25' E and 45°53' N - 46°40' N. The mountainous area in the north of Tonghe county has dense vegetation and rich forests, and the southern part is the alluvial plain of Songhua River. The ecological environment here is healthy, the soil is fertile, and it also has a national scenic spot. Therefore, the investigation of LULC resources is of great significance for the protection of the ecological environment and the regulation of cultivated land resources.

TABLE 1. Statistics on the type, area, and proportion of land resources in the study area.

Statistics	Arbor forest	Paddy field	Upland field	Rivers, lakes, and reservoirs	The road in the countryside	Ditches	Ponds	Other grasslands	Other forests
Area (km ²)	3888.92	1085.23	361.03	106.13	28.22	23.94	22.18	21.00	18.60
Percentage (%)	68.69	19.17	6.38	1.87	0.50	0.42	0.39	0.37	0.33

Houses in the countryside	Swamps	Shallows	Bushes	Houses in the towns	Highway	Construction site	Industrial land	Other gardens
15.95	14.06	13.28	12.02	10.87	6.62	6.51	6.37	5.62
0.28	0.25	0.23	0.21	0.19	0.12	0.11	0.11	0.10

The road of towns	Swamps and meadows	Government public buildings	Commercial facilities	Orchard soil	Agricultural facilities	Bare land	Park	Irrigated land
3.37	2.87	2.16	1.45	1.35	1.13	0.77	0.55	0.52
0.06	0.05	0.04	0.026	0.024	0.020	0.014	0.010	0.009

Special land	Land for railway	Land for transport station	Idle land	Bare rock	Pipeline	Saline-alkali land	Total
0.37	0.24	0.081	0.053	0.045	0.017	0.014	5661.54
0.007	0.004	0.001	0.0009	0.0008	0.0003	0.00025	100

The third land survey database of Heilongjiang province comes from the RS images taken by the China Gaofen satellite and Resources satellite. It is mainly used in land use dynamic monitoring, mineral resources survey, urban and rural planning monitoring and evaluation, traffic network planning, forest resources survey, and desertification monitoring. The land survey database used this time includes satellite RS images taken by Gaofen-2 and Resources satellite 3 observation systems. After processing, the RS image data is composed of two parts, one with a resolution of 17 m and the other with a resolution of 20 m. The resolution of the RS data we use has been reduced for data confidentiality reasons. In Figure 1 (a), this is the RS image after light correction, splicing, and clipping. In Figure 1 (b), the data of land resource classification is derived through on-screen manual delineation from the RS images. As we all know, manual land types labeling is a laborious and complex process, but it is very important and fundamental. In the last step, the whole study area is marked according to different land resources types, and there are 34 different land resources types. To facilitate the next research, we cut the RS image map and the artificial map of LULC resources together. In Figure 1(c), there are more than 5000 pairs. In particular, the pixels of each cut image is 512×512 , and the corresponding area is $512 \text{ m} \times 512 \text{ m}$.

B. VERIFICATION OF RESOURCES CLASSIFICATION

It should be noted that when the classification map of land cover resources is manually drawn on the computer screen, the investigators need to verify the types of LULC resources on the spot. After an extensive field investigation, we obtained evidence of LULC resources types in the study area. Figure 2 shows some photos of evidence collection, and each photo represents different land resources classifications.

From the pictures, it is not difficult to find that there are certain similarities between many land categories, such as highways and urban roads, as well as buildings and facilities for various purposes. This part of the work not only makes up for the lack of resolution of medium and low resolution RS images but also verifies the accuracy of manually labeled LULC types.

C. STATISTICS OF LAND COVER CLASSIFICATION

In Table 1, the total research area covered by RS images is approximately 5661.54 km^2 , of which the area proportion of growing arbor forest is 68.69%, and the area proportion of cultivated land is 25.55%, and the total proportion of other categories is 5.76%. The cultivated land area can be further divided into paddy fields and upland fields, accounting for 19.17% and 6.38% respectively. From the perspective of traditional LULC, the land types in this area can be roughly divided into five categories, namely forests, rivers and lakes, arable land, roads, and buildings. More detailed multi-classification problems obviously face great challenges.

Through the statistics of land resources classification, we find that various categories are seriously unbalanced, and the area of arbor forest, paddy field, and upland field accounts for more than 90% of the whole region. As shown in Figure 3, there are 30 land cover classifications accounting for less than 1% of the study area.

The process of DL training, controlled by background examples that are easy to classify, will lead to low efficiency of segmentation and classification. Therefore, it is necessary to use a reasonable segmentation model for the comparative test of the sample imbalanced RS image data in the study area.



FIGURE 2. Classification investigation and verification of land cover resources in the study area.

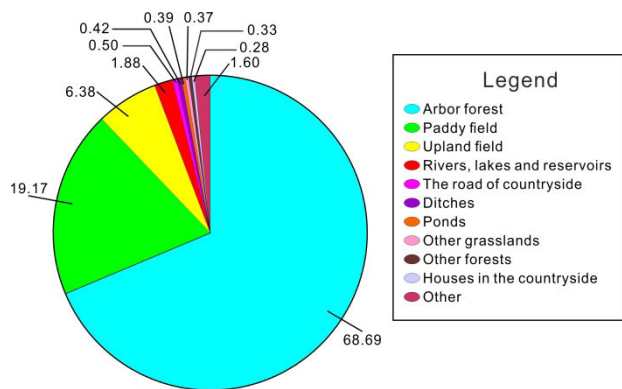


FIGURE 3. Proportional distribution of each land cover category in the dataset.

III. CLASSIFICATION AND SEGMENTATION METHOD

As mentioned above, the current LULC classification mainly relies on manual annotation or traditional machine learning methods to segment the categories and boundaries of land resources. The research content of this section is divided into two parts. Firstly, the feasibility and effect of MLC and RF methods for LULC classification of low and medium resolution RS images are discussed. Then, when the traditional machine learning methods have a poor effect in realizing large-scale automatic classification, the general method of LULC automatic classification based on the DL method is discussed. We use the environment for visualizing images (ENVI) to realize the training of MLC and RF traditional methods supervised classification, and realize deep learning segmentation and classification model construction based on the TensorFlow system.

A. MAXIMUM LIKELIHOOD AND RANDOM FORESTS

Maximum likelihood and random forests have always been important methods to solve the classification of multi-category RS images of land use and land cover. Because the difference in shooting time of RS images will influence the results of MLC and RF classification methods. Therefore, we only select Gaofen-2 satellite data in the region for supervised classification and comparative tests.

We determine the classification system according to the actual land use situation and establish suitable classification samples. The land types distinguish by calculating the separable value Jeffries-Matusita coefficient of the samples. Finally, it is generally marked as five types, namely forests, rivers and lakes, arable land, roads, and buildings. After supervised classification, the majority and minority analysis method is used for noise reduction processing. The final classification results are shown in Figure 4.

The overall accuracy of MLC method classification is 86.7%, while the accuracy of RF method classification is 90.3%. The two methods are effective in the classification of forest categories. However, as the areas marked on the map, their classification results are easy to confuse roads, buildings, and arable land, and it is even more difficult to identify

rivers and Lakes between cities and towns. Of course, the effect of land classification using the MLC or RF method also depends on the human operation, but the level of resolution has a major impact on the classification results.

In fact, after many experiments, using MLC and RF methods, we are still unable to classify the RS image of Resources satellite 3. These two methods have some disadvantages when using supervised classification to extract feature information from Gaofen-2 RS images. It is easy to ignore sporadic small area features, resulting in mixed and wrong classification to a certain extent. It requires high precision and quality of RS images, and there will be a large deviation in classification for images in different seasons.

Observing the several areas delineated by the dotted line in Figure 4, the classification results obtained by the two traditional machine learning methods MLC and RF are very different from the actual land resource types. Most of the arable land is identified as buildings, and the classification effect of the mountain shadow area of RS images by these two methods is poor. Such results are obviously difficult to meet the needs of LULC resources in multiple classifications in this area.

B. THE BASIC ARCHITECTURE OF THE ALGORITHM

As shown in Figure 5, the preprocessed RS image data first passes through the residual network (ResNet-50) [48]. ResNet-50 consists of convolution layers and pool layers, and it uses a connection called shortcut connection. ResNet-50 is simple and practical, widely used in target detection, classification, and recognition. For the dataset of the Tonghe RS image, ResNet-50 is selected as the backbone, which can completely solve the extraction of key information.

As mentioned in the background, the categories of land cover resources are seriously imbalanced, so it is uncertain which segmentation model has the best adaptability for our dataset. Therefore, we choose 7 classical convolutional neural networks as sample training models. RS image data is input into the CNN with RGB 3 channels, and the corresponding result at the output is the manually marked land resources types. The experimental process relies on 7 computers to run independently and is calculated by GeForce RTX 3070 laptop GPU. It takes about 20 days to complete the training of the network model after many times tuning the learning rate, activation function, optimizer, and other parameters and hyperparameters.

The following content is the implementation details that need to be emphasized in the process of this experiment.

To adapt to the serious imbalance between positive and negative samples in the training process, we decided to use the Focal Loss [49] as a unified loss function. It is introduced from the cross-entropy (CE) loss for binary or multi-value classification. The form of Focal Loss we adopt is as follows.

$$FL_{loss}(p_t) = -\alpha_t (1 - p_t)^\gamma \log(p_t) \quad (1)$$

The loss function is used in all the experiments as it yields slightly improved accuracy over the non- α -balanced form,

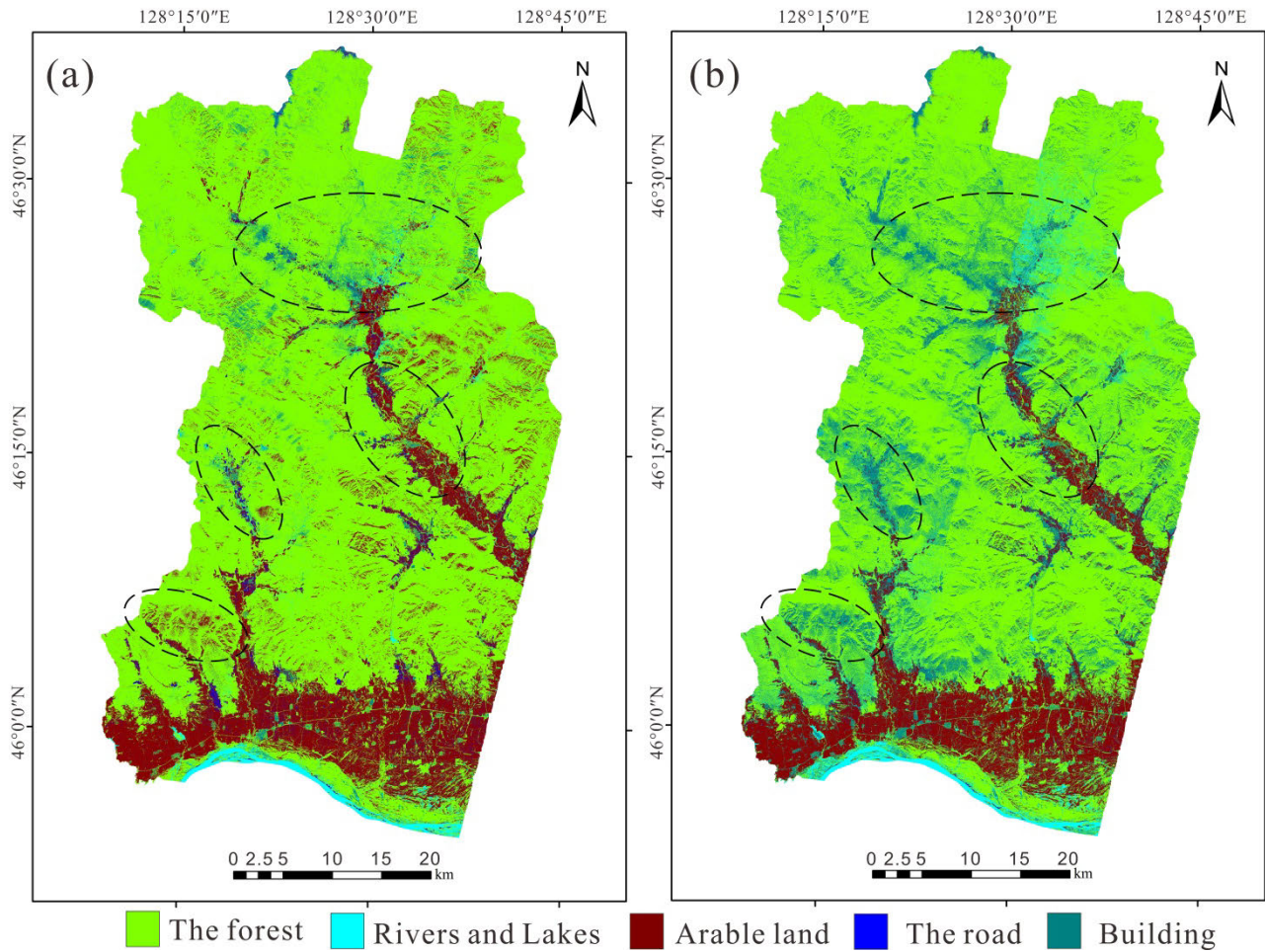


FIGURE 4. (a) Maximum likelihood classification results of Gaofen-2 RS image. (b) Random forests classification results of Gaofen-2 RS image.

and p is the model’s estimated probability for the class. Finally, we determine that when the value of α_t is 0.25 and the value of γ is 5, the effect is the best. Compared with CE loss, the core of Focal Loss is to use an appropriate function to measure the contribution of difficult and easy to classify samples to the total loss.

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise} \end{cases} \quad (2)$$

The selection of an optimization algorithm is also an important part of the model. In order to make up for the defects of the decline of naive gradient, this paper selects adaptive motion estimation (Adam) [50] as the optimizer. The Adam is a first-order optimization algorithm that can replace the traditional Stochastic Gradient Descent (SGD). It can iteratively update the weight of the neural network based on training data, and the learning rate during this training is uniformly given as 10^{-3} .

Obviously, for different datasets, to prevent underfitting and overfitting, the epoch given by us during training is different. The epoch value given in this paper is 200. We think

that the training effect is the best at this time, and it is not in an overfitting state.

C. DATA AUGMENTATION

Due to the limited training samples and random sampling data in nature, there is a large probability of long-tailed distributions, and it is difficult to achieve sample balance for different categories. To eliminate the influence of long-tailed distribution and data augmentation are usually adopted during training to increase data diversity, prevent over-fitting of the training process, and make neural networks have good generalization ability. In the experiment, we adopted the following four data augmentation techniques, which are random flip, random clipping, random rotation, and random translation. A random flip is to flip the original picture vertically or horizontally. The purpose of randomly clipping RS images for regions of interest is to increase random disturbance. By randomly rotating a certain angle or shifting a certain distance, the data can be changed based on the original image, which increases the number of data samples, but the label value of the data does not change.

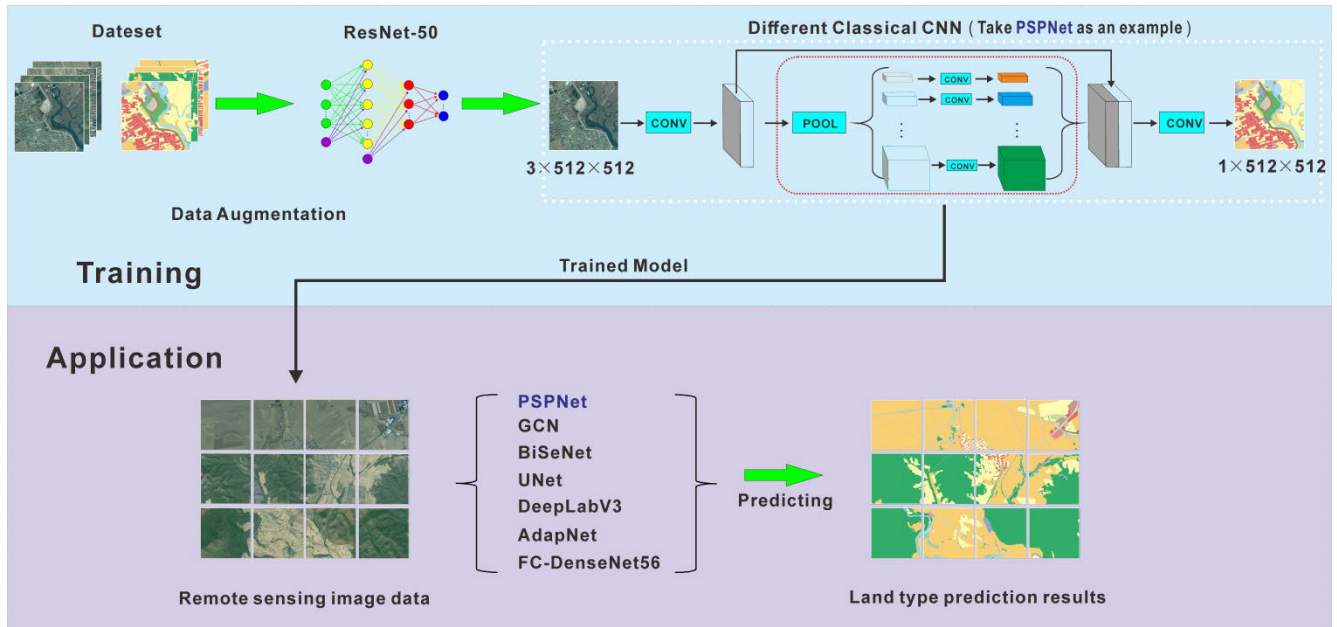


FIGURE 5. Semantic segmentation method of RS image based on CNN and its application.

D. DIVISION OF THE DATA SET

The training set of this experiment is divided into two mutually exclusive subsets by using the common hold-out method, which directly samples the data set randomly according to a certain proportion.

The larger number of samples is used as the training set and the other as the test set. In the process of data division, the sample proportion of each category needs to be consistent. This can maintain the consistency of data distribution and avoid the interference of the deviation introduced by data division on the evaluation results. The existing practical experience of neural network training shows that the training set accounts for two-thirds to four-fifths of the whole data set, which is a reasonable division scheme. Our training set accounts for 70% and the verification set accounts for 30%.

IV. PERFORMANCE COMPARISON OF DL MODELS

In this section, before comparing the optimization model classification results, the major metrics used to evaluate the performance of the DL algorithms will be described. The commonly used metrics include accuracy, precision [51], recall [52], and IoU [53].

Accuracy defines the ability of the model to generate correct quantitative predictions for observations. The definition formula is expressed as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Precision and Recall are defined as follows.

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

where TP and TN denote true positive and true negative. Similarly, FP and FN represent false positive and false negative.

IoU (Intersection over Union) is thought the most popular metric, where B^{gt} is ground truth, and B is the predicted box.

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (6)$$

Finally, the classification algorithms of seven segmentation models are adopted, and the performance evaluation of these algorithms is listed in Table 2. Because forests account for about 69%, the overall classification accuracy will be high, while the MIoU metric can be used to evaluate the overall performance of the model. For the same dataset, we focus on the differences in metrics, to identify the performance of the network models.

TABLE 2. Accuracy, precision, recall, and MIoU with various segmentation models.

Method	Parameters	Accuracy (%)	Precision (%)	Recall (%)	MIoU (%)
FC-DenseNet56	1370843	93.16	95.58	93.16	53.95
GCN	23985766	93.17	95.23	93.17	54.58
BiSeNet	29243083	92.06	94.27	92.06	43.90
U-Net	34970019	93.62	95.23	93.62	55.90
DeepLabV3	27659430	91.57	94.55	91.57	40.81
AdapNet	21069538	90.86	94.04	90.86	41.95
PSPNet	37102307	92.85	95.08	92.91	54.68

Obviously, from the performance of the MIoU metrics, the segmentation models show great differences for the same RS data. It is observed that the U-Net segmentation model exhibits the best performance as compared to others, while the DeepLabV3 model exhibits the worst. In addition, the IoU

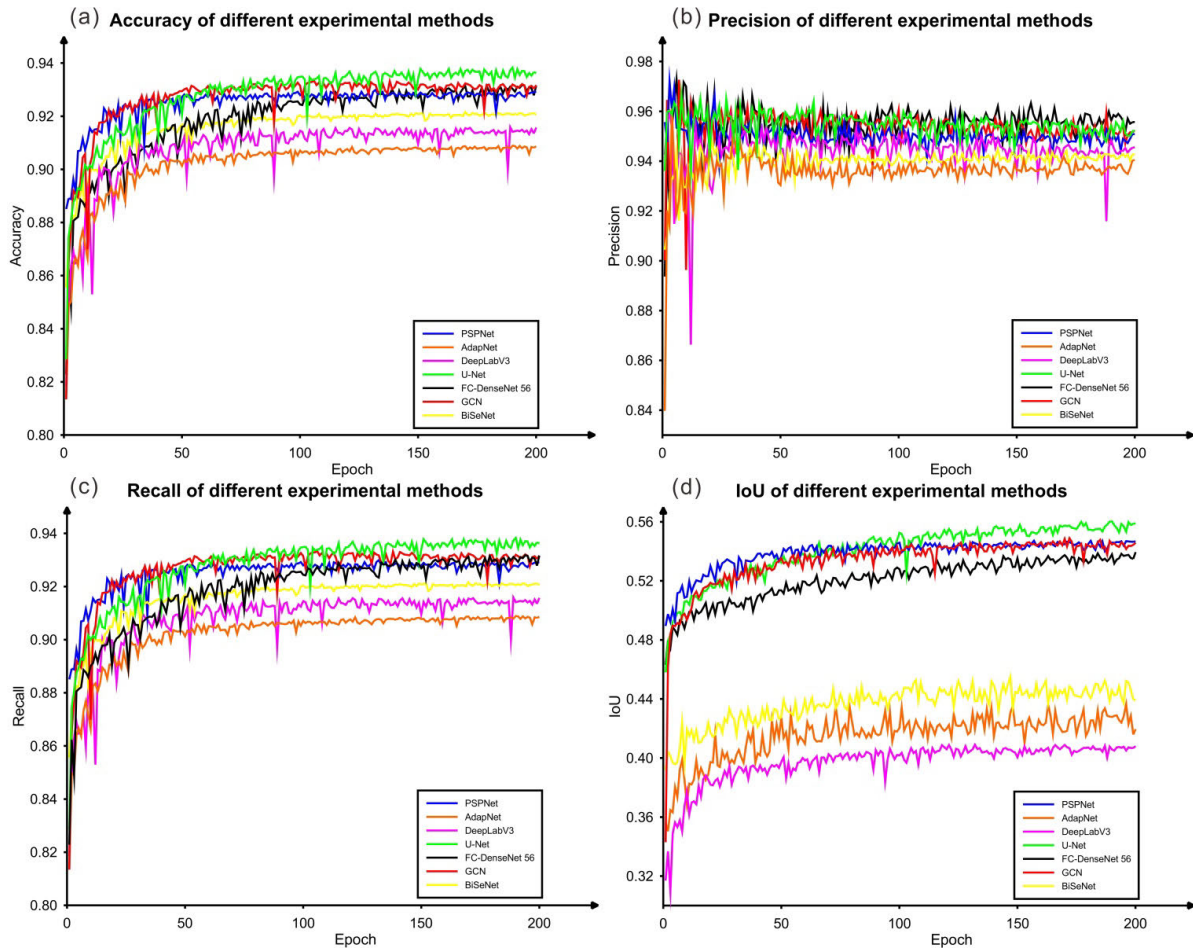


FIGURE 6. The performance curves of different segmentation network models. (a) Accuracy verification curves of different networks. (b) Precision verification curves. (c) Recall verification curves. (d) MIoU verification curves.

metrics of predicting by Bisenet and Adapnet segmentation models are not more than 50%. FC-densenet56, GCN, and PSPNet show almost as good performance as the U-net. Therefore, we should focus on the above four segmentation models. On the other hand, the FC-densenet56 model has the least parameters and the PSPNet model has the most parameters. The metrics of them are almost identical, and it shows the potential of the FC-densenet56 model in solving the classification of low resolution RS images. Figure 6 shows the performance curves of the verification dataset based on different segmentation network models during the RS data experiment in the study area. In this way, we can have the most intuitive understanding of the performance of the segmentation models.

The loss curves of the verification set based on different models are given in Figure 7. As is shown, with the increase of the epoch number, the loss decreases gradually and finally tends to be stable when the number of epochs reaches 200. Specifically speaking, when the number of epochs reaches 200, the loss values of all the segmentation models are below 3×10^{-4} , showing relatively stable performance. It should be noted that the loss value of PSPNet is less than 5×10^{-5} among all the models. The MIoU metric of PSPNet also

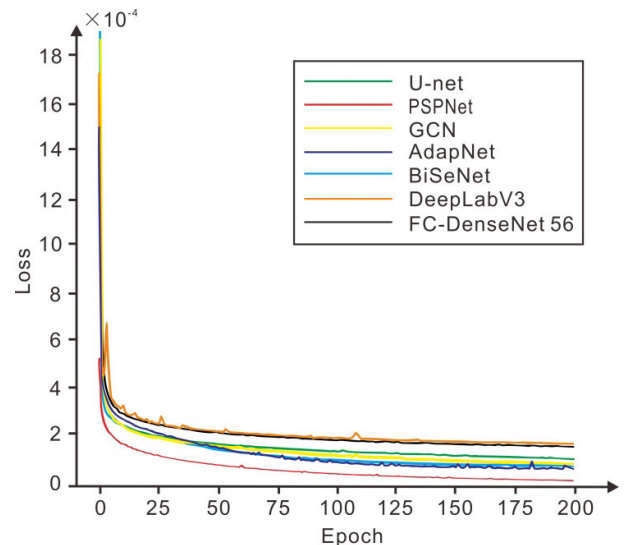


FIGURE 7. Loss curves of different experimental calculation methods.

reflects the excellent performance of the model, but it has not shown its advantages for low resolution RS datasets. The two models with the slowest attenuation of loss curve are

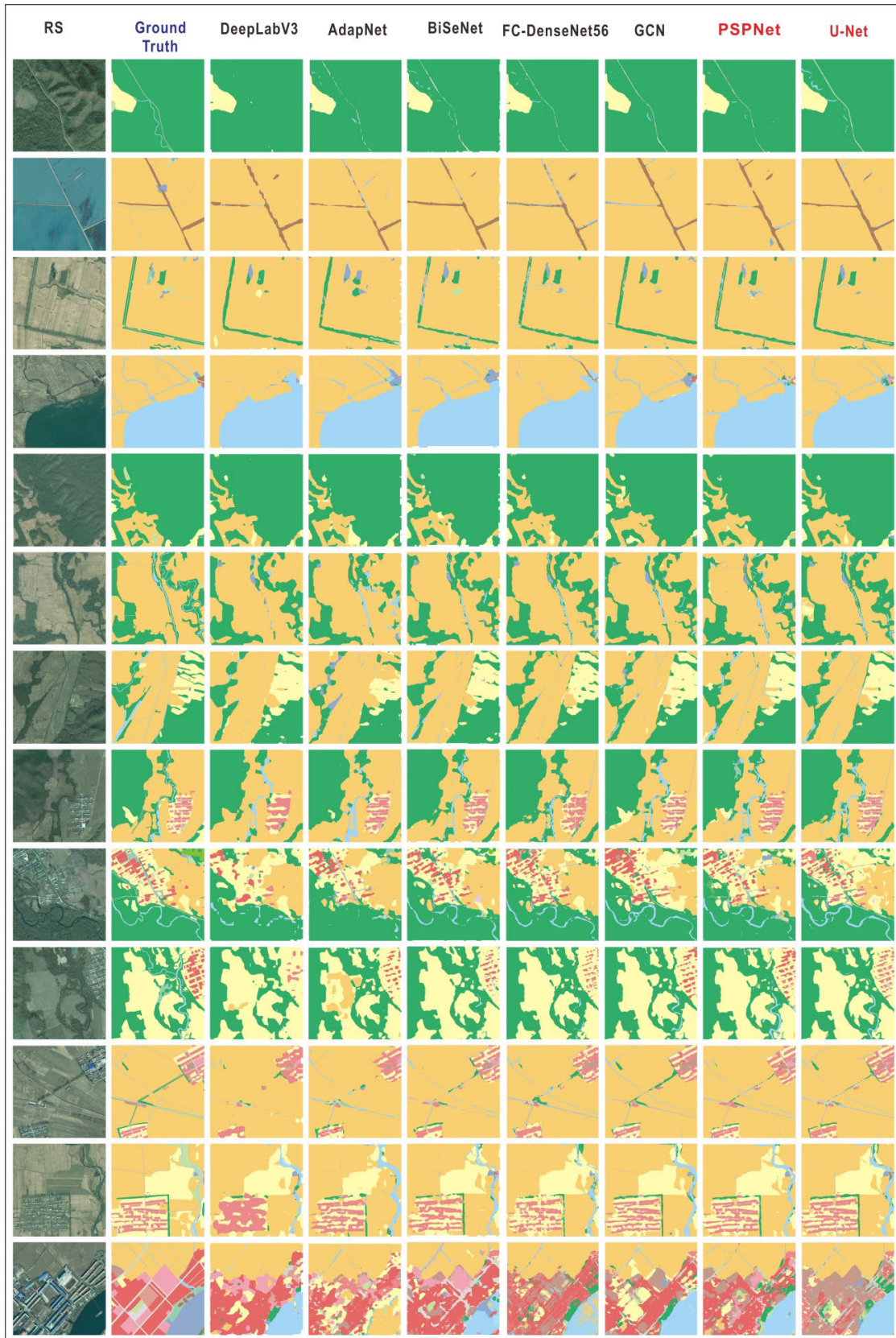


FIGURE 8. Results of classification using different methods for validation dataset.

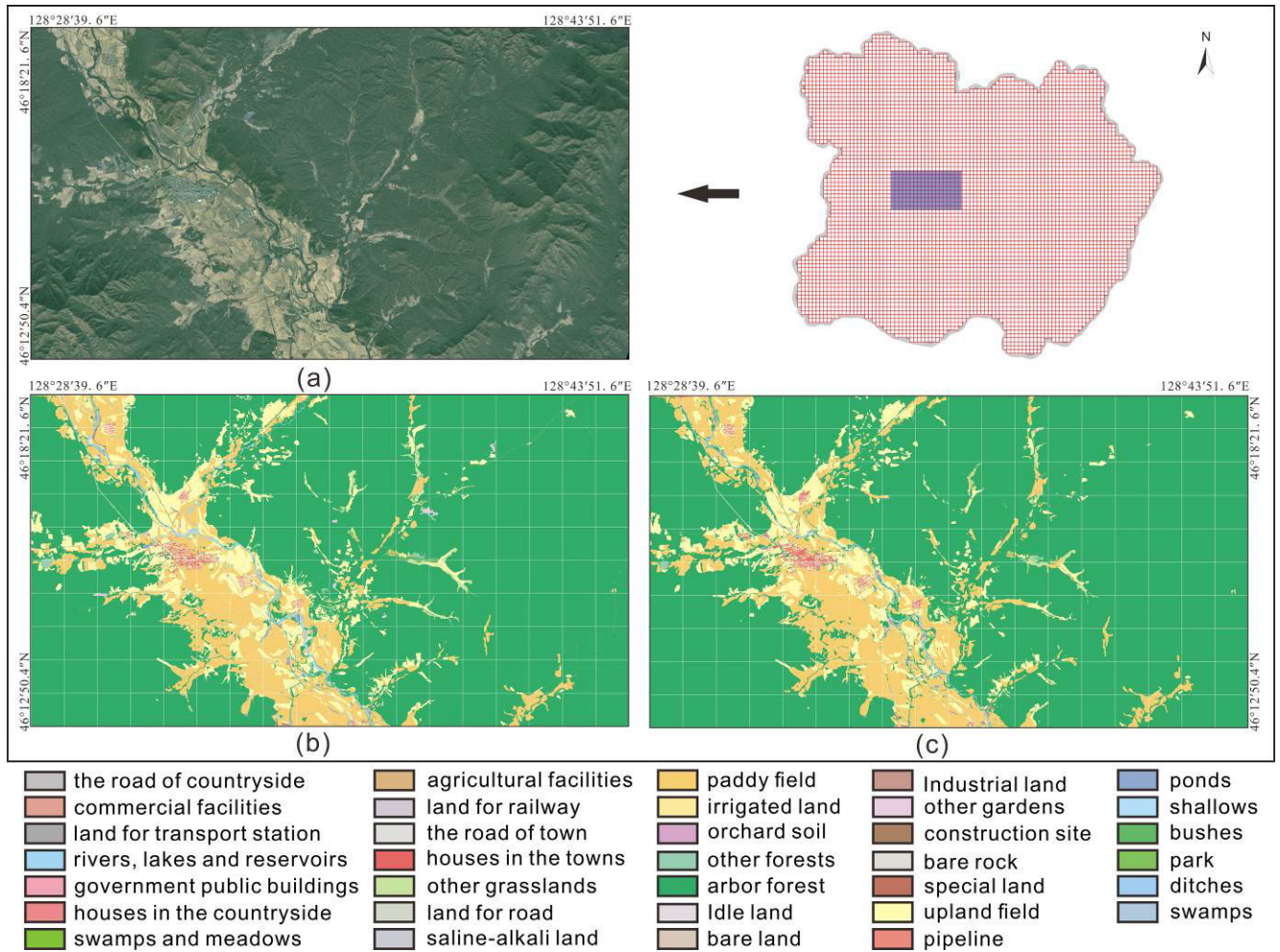


FIGURE 9. Land cover resources classification results of satellite RS images in the test area. (a) Satellite remote sensing image of the test area. (b) Images of land cover resources classification labeled by manual operation software. (c) Prediction results are based on the U-Net segmentation classification model.

DeepLabV3 and DenseNet-56, which shows that the convergence ability of the model is poor during training.

V. RESULTS AND DISCUSSION

The above analysis is only based on the metrics by the models and the loss value curves, while experimental results are the most intuitive to reflect the effect of land cover resources classification, and we are most concerned about it. In the process of training, to monitor the training process in real-time, the prediction results of each epoch model are recorded. We selected the prediction results of 7 kinds of models at the 200 epoch. We picked up the prediction results of 7 kind models at 200 epochs, and selected representative RS images to classify land cover resources from simple to complex.

A. PERFORMANCE ANALYSIS OF DL MODELS

The satellite RS images, the artificial annotation images, and the prediction result images by models are displayed in turn in Figure 8. This paper has the following understanding of

the classification effect of land cover resources based on the prediction results.

Firstly, all the selected segmentation models can classify and segment the low-resolution RS images to a certain extent after training. Particularly, for the arbor forest, paddy field, and upland field, all the models have an obvious effect on land cover resource classification. However, for the classification and segmentation of roads, rivers, and so on, the detailed effect in the prediction results is not good. This result is acceptable because the resolution of RS images is low, and the field survey is carried out when the land resource types are manually marked so that the ground truth labels contain more information.

Secondly, all the selected models, after training, have a poor effect on the detailed classification of buildings. The classification of building categories is the most difficult and the reflection characteristics are the most complex. Building samples are regarded as difficult samples in the process of DL training, which will have a great impact on the metrics of training. Because the classification of building categories

is not calibrated by the shape outline of buildings but determined by manual visits and investigation, the DL models can not learn its characteristic law, and the poor prediction effect is inevitable. Even so, the models selected in this paper have a certain effect on the recognition of building groups in RS images.

Finally, through the analysis and discussion of this computational experiment, we believe that it is more suitable for our RS data to be classified by U-Net, PSPNet, or GCN model. For the task of our land cover resources survey, the U-Net model is the best choice without considering relying on the DL model to segment and classify more details. The classification of specific details still depends on manual actual investigation, especially the classification of building groups use, or the difficult recognition areas caused by the limitation of RS image resolution. The above understanding is based on the images prediction results of the validation set. We should test the application effect of the selected model on the test area, to facilitate the subsequent practical promotion in land cover resources investigation and classification.

B. REMOTE SENSING TEST AREA

As shown in Figure 9, the test area is located in the middle west of the study area, located between 128°28' 39.6" E - 129°43' 51.6" E and 46°12' 50.4" N - 46°18' 21.6" N, which is composed of 190 cropped RS images. The RS data tested in this part are set aside separately and will not participate in the training of the CNN model during the whole process. The area is mainly covered by forests and fields, and there are cities and towns inhabited by human beings, with rivers passing through. The paddy field and upland field are staggered in the test area, and the roads between towns are complex. In the previous section, we conducted detailed experiments on the performance of 7 models and finally selected the U-Net model concerning the viewpoints obtained from our experiments. The trained model includes the backbone ResNet-50, combined with the U-Net convolutional neural network. The U-Net network model after training is put into the practical application of land cover classification in the above test area.

C. THE RESULTS OF USING U-NET DL MODLE

We compare the classification results of artificial land cover resources with the prediction results of the segmentation model U-Net. As shown in Figure 9 (c), it can be found that the prediction results basically restore the overall classification of land cover resources. It can quickly and accurately classify forests, cultivated land, towns, villages, and so on.

However, if we carefully compare the manual annotation results with the classification results of the DL segmentation model, we can see the difference between them. The DL segmentation model can not better predict the panorama of rivers and rural roads, nor can it accurately divide the special uses of buildings as a manual annotation. Nevertheless, even if the results obtained by the DL segmentation model are rough, it is of great significance for land cover resources investigation.

Because manual labeling is time-consuming and laborious work, especially when there is no reference, it is only judged according to satellite RS images, and the DL segmentation model provides great convenience for investigation workers.

VI. CONCLUSION

Traditional supervised classification methods, such as RF and MLC, have the characteristics of low training time, few labeled samples, and stable ability of simple classification. However, they have high requirements for the resolution of RS images and poor ability of feature extraction. To solve the problem of resources classification in the process of land resources investigation, this paper mainly experiments with the performance of the classical CNN segmentation model, including FC-DenseNet56, GCN, BiSeNet, U-Net, Deeplabv3, AdapNet, and PSPNet. The results of various metrics show that the U-Net segmentation model is the most competitive model for low-medium resolution satellite RS images. At the same time, the classification results of the validation set samples are analyzed, and finally, the trained segmentation model U-Net is used in the test area to obtain the land cover resources classification. The results show that the land resources classification images are almost as well as the manual annotation images.

Furthermore, this research work covers the whole process of land resources classification survey, including satellite RS image processing in the early stage, actual exploration and certification of investigators, and careful labeling and classification of land resources by indoor personnel in the later stage. In the process of completing the above investigation work, the experimental result of the optimal CNN segmentation and classification U-Net model is remarkable and can be used as a reference for the fourth land resources survey in the future.

ACKNOWLEDGMENT

The authors truly appreciate the helpful remarks and constructive ideas provided by the academic editors and reviewers.

REFERENCES

- [1] O. Rozenstein and A. Karnieli, "Comparison of methods for land-use classification incorporating remote sensing and GIS inputs," *Appl. Geography*, vol. 31, no. 2, pp. 533–544, Apr. 2011, doi: [10.1016/j.apgeog.2010.11.006](https://doi.org/10.1016/j.apgeog.2010.11.006).
- [2] A. Shafique, G. Cao, Z. Khan, M. Asad, and M. Aslam, "Deep learning-based change detection in remote sensing images: A review," *Remote Sens.*, vol. 14, no. 4, p. 871, Feb. 2022, doi: [10.3390/rs14040871](https://doi.org/10.3390/rs14040871).
- [3] A. Balha, J. Mallick, S. Pandey, S. Gupta, and C. K. Singh, "A comparative analysis of different pixel and object-based classification algorithms using multi-source high spatial resolution satellite data for LULC mapping," *Earth Sci. Informat.*, vol. 14, no. 4, pp. 2231–2247, Dec. 2021, doi: [10.1007/s12145-021-00685-4](https://doi.org/10.1007/s12145-021-00685-4).
- [4] X. X. Zhu, D. Tuia, L. Mou, G. S. Xia, and L. Zhang, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017, doi: [10.1109/MGRS.2017.2762307](https://doi.org/10.1109/MGRS.2017.2762307).
- [5] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, Jun. 2019, doi: [10.1016/j.isprsjprs.2019.04.015](https://doi.org/10.1016/j.isprsjprs.2019.04.015).

- [6] M. Mohajane, A. Essahlaoui, F. Oudija, M. El Hafyani, A. E. Hmaid, A. El Ouali, G. Randazzo, and A. C. Teodoro, "Land use/land cover (LULC) using landsat data series (MSS, TM, ETM+ and OLI) in azrou forest, in the central middle atlas of Morocco," *Environments*, vol. 5, no. 12, p. 131, Dec. 2018, doi: [10.3390/environments5120131](https://doi.org/10.3390/environments5120131).
- [7] J. Kumar, B. Biswas, and S. Walker, "Multi-temporal LULC classification using hybrid approach and monitoring built-up growth with Shannon's entropy for a semi-arid region of Rajasthan, India," *J. Geol. Soc. India*, vol. 95, no. 6, pp. 626–635, Jun. 2020, doi: [10.1007/s12594-020-1489-x](https://doi.org/10.1007/s12594-020-1489-x).
- [8] S. Hussain, M. Mubeen, and S. Karuppappan, "Land use and land cover (LULC) change analysis using TM, ETM+ and OLI landsat images in district of Okara, Punjab, Pakistan," *Phys. Chem. Earth, A/B/C*, to be published, doi: [10.1016/j.pce.2022.103117](https://doi.org/10.1016/j.pce.2022.103117).
- [9] T. Shimrah, P. Lungleng, A. R. Devi, K. Sarma, F. Varah, and Y. S. Khuman, "Spatio-temporal assessment on land use and land cover (LULC) and forest fragmentation in shifting agroecosystem landscape in Ukhrul district of Manipur, Northeast India," *Environ. Monitor. Assessment*, vol. 194, no. 1, p. 14, Jan. 2022, doi: [10.1007/s10661-021-09548-3](https://doi.org/10.1007/s10661-021-09548-3).
- [10] P. Verma, A. Raghubanshi, P. K. Srivastava, and A. S. Raghubanshi, "Appraisal of Kappa-based metrics and disagreement indices of accuracy assessment for parametric and nonparametric techniques used in LULC classification and change detection," *Model. Earth Syst. Environ.*, vol. 6, no. 2, pp. 1045–1059, Jun. 2020, doi: [10.1007/s40808-020-00740-x](https://doi.org/10.1007/s40808-020-00740-x).
- [11] S. Pareeth, P. Karimi, M. Shafiei, and C. De Fraiture, "Mapping agricultural landuse patterns from time series of landsat 8 using random forest based hierarchical approach," *Remote Sens.*, vol. 11, no. 5, p. 601, Mar. 2019, doi: [10.3390/rs11050601](https://doi.org/10.3390/rs11050601).
- [12] C. A. Orieschnig, G. Belaud, J.-P. Venot, S. Massuel, and A. Ogilvie, "Input imagery, classifiers, and cloud computing: Insights from multi-temporal LULC mapping in the cambodian Mekong delta," *Eur. J. Remote Sens.*, vol. 54, no. 1, pp. 398–416, Jan. 2021, doi: [10.1080/22797254.2021.1948356](https://doi.org/10.1080/22797254.2021.1948356).
- [13] R. Manandhar, I. Odeh, and T. Ancev, "Improving the accuracy of land use and land cover classification of landsat data using post-classification enhancement," *Remote Sens.*, vol. 1, no. 3, pp. 330–344, Jul. 2009, doi: [10.3390/rs1030330](https://doi.org/10.3390/rs1030330).
- [14] W. Liu, Y. Zhang, J. Yan, Y. Zou, and Z. Cui, "Semantic segmentation network of remote sensing images with dynamic loss fusion strategy," *IEEE Access*, vol. 9, pp. 70406–70418, 2021, doi: [10.1109/ACCESS.2021.3078742](https://doi.org/10.1109/ACCESS.2021.3078742).
- [15] B. Cui, X. Chen, and Y. Lu, "Semantic segmentation of remote sensing images using transfer learning and deep convolutional neural network with dense connection," *IEEE Access*, vol. 8, pp. 116744–116755, 2020, doi: [10.1109/ACCESS.2020.3003914](https://doi.org/10.1109/ACCESS.2020.3003914).
- [16] G. Sumbul and B. Demir, "A deep multi-attention driven approach for multi-label remote sensing image classification," *IEEE Access*, vol. 8, pp. 95934–95946, 2020, doi: [10.1109/ACCESS.2020.2995805](https://doi.org/10.1109/ACCESS.2020.2995805).
- [17] H. S. Alhichri, A. S. Alswayed, Y. Bazi, N. Ammour, and N. A. Ajlan, "Classification of remote sensing images using efficientNet-B3 CNN model with attention," *IEEE Access*, vol. 9, pp. 14078–14094, 2021, doi: [10.1109/ACCESS.2021.3051085](https://doi.org/10.1109/ACCESS.2021.3051085).
- [18] H. Zhang, Y. Liao, H. Yang, G. Yang, and L. Zhang, "A local-global dual-stream network for building extraction from very-high-resolution remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1269–1283, Mar. 2022, doi: [10.1109/TNNLS.2020.3041646](https://doi.org/10.1109/TNNLS.2020.3041646).
- [19] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2486–2498, May 2017, doi: [10.1109/TGRS.2016.2645610](https://doi.org/10.1109/TGRS.2016.2645610).
- [20] W. Zhang, P. Tang, and L. Zhao, "Fast and accurate land-cover classification on medium-resolution remote-sensing images using segmentation models," *Int. J. Remote Sens.*, vol. 42, no. 9, pp. 3277–3301, May 2021, doi: [10.1080/01431161.2020.1871094](https://doi.org/10.1080/01431161.2020.1871094).
- [21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015, *arXiv:1505.04597*.
- [22] S. Sun, L. Mu, L. Wang, and P. Liu, "L-UNet: An LSTM network for remote sensing image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: [10.1109/LGRS.2020.3041530](https://doi.org/10.1109/LGRS.2020.3041530).
- [23] D. Xiao, L. Yin, and Y. Fu, "Open-pit mine road extraction from high-resolution remote sensing images using RATT-UNet," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: [10.1109/LGRS.2021.3065148](https://doi.org/10.1109/LGRS.2021.3065148).
- [24] Y. Hou, Z. Liu, T. Zhang, and Y. Li, "C-UNet: Complement UNet for remote sensing road extraction," *Sensors*, vol. 21, no. 6, p. 2153, Mar. 2021, doi: [10.3390/s21062153](https://doi.org/10.3390/s21062153).
- [25] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," 2016, *arXiv:1612.01105*.
- [26] R. Zhang, J. Chen, L. Feng, S. Li, W. Yang, and D. Guo, "A refined pyramid scene parsing network for polarimetric SAR image semantic segmentation in agricultural areas," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: [10.1109/LGRS.2021.3086117](https://doi.org/10.1109/LGRS.2021.3086117).
- [27] Y. Li, X. Pei, Q. Huang, L. Jiao, R. Shang, and N. Marturi, "Anchor-free single stage detector in remote sensing images based on multiscale dense path aggregation feature pyramid network," *IEEE Access*, vol. 8, pp. 63121–63133, 2020, doi: [10.1109/ACCESS.2020.2984310](https://doi.org/10.1109/ACCESS.2020.2984310).
- [28] Y. Liu, L. Gross, Z. Li, X. Li, X. Fan, and W. Qi, "Automatic building extraction on high-resolution remote sensing imagery using deep convolutional encoder-decoder with spatial pyramid pooling," *IEEE Access*, vol. 7, pp. 128774–128786, 2019, doi: [10.1109/ACCESS.2019.2940527](https://doi.org/10.1109/ACCESS.2019.2940527).
- [29] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.
- [30] J. Liang, Y. Deng, and D. Zeng, "A deep neural network combined CNN and GCN for remote sensing scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4325–4338, 2020, doi: [10.1109/JSTARS.2020.3011333](https://doi.org/10.1109/JSTARS.2020.3011333).
- [31] A. Valada, J. Vertens, A. Dhall, and W. Burgard, "AdapNet: Adaptive semantic segmentation in adverse environmental conditions," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, Singapore, May/Jun. 2017, pp. 4644–4651, doi: [10.1109/ICRA.2017.7989540](https://doi.org/10.1109/ICRA.2017.7989540).
- [32] S. An, A. Medda, M. N. Sawka, C. J. Hutto, M. L. Millard-Stafford, S. Appling, K. L. S. Richardson, and O. T. Inan, "AdaptNet: Human activity recognition via bilateral domain adaptation using semi-supervised deep translation networks," *IEEE Sensors J.*, vol. 21, no. 18, pp. 20398–20411, Sep. 2021, doi: [10.1109/JSEN.2021.3095176](https://doi.org/10.1109/JSEN.2021.3095176).
- [33] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "BiSeNet: Bilateral segmentation network for real-time semantic segmentation," 2018, *arXiv:1808.00897*.
- [34] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.
- [35] S. Jégou, M. Drozdal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation," 2016, *arXiv:1611.09326*.
- [36] L. Schor, I. Bacivarov, H. Yang, and L. Thiele, "AdaPNet: Adapting process networks in response to resource variations," in *Proc. Int. Conf. Compil., Archit. Synth. Embedded Syst.*, New Delhi, India, Oct. 2014, pp. 1–10, doi: [10.1145/2656106.2656112](https://doi.org/10.1145/2656106.2656112).
- [37] Z. Zhang, T. Jiang, C. Liu, and L. Zhang, "An effective classification method for hyperspectral image with very high resolution based on encoder-decoder architecture," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1509–1519, 2021, doi: [10.1109/JSTARS.2020.3046245](https://doi.org/10.1109/JSTARS.2020.3046245).
- [38] Y. Lin, D. Xu, N. Wang, Z. Shi, and Q. Chen, "Road extraction from very-high-resolution remote sensing images via a nested SE-deeplab model," *Remote Sens.*, vol. 12, no. 18, p. 2985, Sep. 2020, doi: [10.3390/rs12182985](https://doi.org/10.3390/rs12182985).
- [39] X. Li, X. Yao, and Y. Fang, "Building-A-nets: Robust building extraction from high-resolution remote sensing images with adversarial networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 10, pp. 3680–3687, Oct. 2018, doi: [10.1109/JSTARS.2018.2865187](https://doi.org/10.1109/JSTARS.2018.2865187).
- [40] X. Gao, T. Chen, R. Niu, and A. Plaza, "Recognition and mapping of landslide using a fully convolutional DenseNet and influencing factors," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7881–7894, 2021, doi: [10.1109/JSTARS.2021.3101203](https://doi.org/10.1109/JSTARS.2021.3101203).
- [41] X. Guo, Z. Chen, and C. Wang, "Fully convolutional DenseNet with adversarial training for semantic segmentation of high-resolution remote sensing images," *J. Appl. Remote Sens.*, vol. 15, no. 1, Mar. 2021, doi: [10.1117/1.JRS.15.016520](https://doi.org/10.1117/1.JRS.15.016520).
- [42] S. D. Khan, L. Alarabi, and S. Basalamah, "Deep hybrid network for land cover semantic segmentation in high-spatial resolution satellite images," *Information*, vol. 12, no. 6, p. 230, May 2021, doi: [10.3390/info12060230](https://doi.org/10.3390/info12060230).
- [43] N. J. Murray, D. A. Keith, D. Simpson, J. H. Wilshire, and R. M. Lucas, "REMAP: An online remote sensing application for land cover classification and monitoring," *Methods Ecol. Evol.*, vol. 9, no. 9, pp. 2019–2027, Sep. 2018, doi: [10.1111/2041-210X.13043](https://doi.org/10.1111/2041-210X.13043).

[44] C. Cruz-Ramos, B. P. Garcia-Salgado, R. Reyes-Reyes, V. Ponomaryov, and S. Sadovnychiy, "Gabor features extraction and land-cover classification of urban hyperspectral images for remote sensing applications," *Remote Sens.*, vol. 13, no. 15, p. 2914, Jul. 2021, doi: 10.3390/rs13152914.

[45] W. Zhang, P. Tang, T. Corpetti, and L. Zhao, "WTS: A weakly towards strongly supervised learning framework for remote sensing land cover classification using segmentation models," *Remote Sens.*, vol. 13, no. 3, p. 394, Jan. 2021, doi: 10.3390/rs13030394.

[46] Q. Sang, Y. Zhuang, S. Dong, G. Wang, and H. Chen, "FRF-net: Land cover classification from large-scale VHR optical remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1057–1061, Jun. 2020, doi: 10.1109/LGRS.2019.2938555.

[47] H. Shao, Y. Li, Y. Ding, Q. Zhuang, and Y. Chen, "Land use classification using high-resolution remote sensing images based on structural topic model," *IEEE Access*, vol. 8, pp. 215943–215955, 2020, doi: 10.1109/ACCESS.2020.3041645.

[48] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.

[49] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," 2017, *arXiv:1708.02002*.

[50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[51] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation," in *AI 2006: Advances in Artificial Intelligence*, vol. 4304, A. Sattar and B. Kang, Eds. Berlin, Germany: Springer, 2006, pp. 1015–1021, doi: 10.1007/11941439_114.

[52] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation," in *Advances in Information Retrieval*, vol. 3408, D. E. Losada and J. M. Fernández-Luna, Eds. Berlin, Germany: Springer, 2005, pp. 345–359, doi: 10.1007/978-3-540-31865-1_25.

[53] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," 2019, *arXiv:1911.08287*.



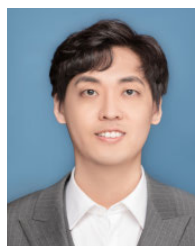
ZHICHAO GAO received the B.S. degree in geophysics from the China University of Geosciences, Beijing, China, and the M.S. degree in human geography from Harbin Normal University. He has been engaged in geophysical surveys and remote sensing imaging interpretation. During the third land resources survey in China, he presided over the geological disaster evaluation and business guidance in most areas of Heilongjiang Province, including the research area of this paper.



RUI LI received the B.S. degree in geophysics and the M.S. degree in geological engineering from the China University of Geosciences, Beijing, China. He is currently a Geophysical Engineer with the China Aero Geophysical Survey and Remote Sensing Center for Natural Resources (AGRS). His research interests include the inversion of airborne gravity data and the processing of remote sensing data, including the Kalman filter and other methods.



YAO LIU received the B.S. degree in measurement and control technology and instrument and the M.S. degree in geophysics from the China University of Geosciences, Beijing, China, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree in geophysical prospecting and information technology. His research interests include machine learning and deep learning, computer vision, gravity, magnetic data processing, and geophysical inverse problem.



LIANZHI ZHANG (Member, IEEE) received the B.S. degree in geophysics from the China University of Geosciences, Beijing, China, in 2015, where he is currently pursuing the Ph.D. degree in geophysical prospecting and information technology. His research interests include computer vision, machine learning, deep learning, gravity and magnetic data processing, and geophysical inverse problem.



ZIXIANG JIN received the B.S. degree in geomatics from Henan Polytechnic University, and the M.S. degree in geophysics from the China University of Geosciences, Beijing, China. His current research interests include deep learning, geomagnetic aiding navigation, and point cloud registration.



SUPENG XU received the M.S. degree in geophysics from the China University of Geosciences, Beijing, China, in 2015, where he is currently pursuing the Ph.D. degree. His research interests include machine learning and deep learning, geomagnetic navigation, and geophysical inverse problem.

...



ZHENYU FAN (Member, IEEE) received the B.S. and M.S. degrees in geophysics and the Ph.D. degree in geophysical prospecting and information technology from the China University of Geosciences, Beijing, China, in 2013, 2016, and 2020, respectively.

He is currently an Engineer. He is also working with the China Aero Geophysical Survey and Remote Sensing Center for Natural Resources (AGRS), responsible for the research of airborne

geophysics inversion methods and the processing of geophysical data and remote sensing image data. His research interests include computer vision, machine learning, deep learning, and geophysical inverse problem. It is committed to solving earth science problems by combining deep learning or machine learning methods.



TAO ZHAN received the Ph.D. degree from the Institute of Geology and Geophysics, Chinese Academy of Sciences.

For a long time, he has been an Outside School Professor at Heilongjiang University, a Project Evaluation Expert at the National Natural Science Foundation of China, and an Expert at the Bureau of Geology and Mineral Resources of Heilongjiang Province. He is currently a Professor-Level Senior Engineer majoring in

hydrological engineering and environment. In addition, he has written two monographs and published more than 30 scientific research papers. His research interests include quaternary geology, global change, geothermal geology, clean energy heating, natural geography, and remote sensing.