

Improving the Interpretability and Reliability of Regional Land Cover Classification by U-Net Using Remote Sensing Data

WANG Xinshuang¹, CAO Jiancheng², LIU Jiange¹, LI Xiangwu¹, WANG Lu³, ZUO Feihang¹, BAI Mu¹

(1. Shaanxi Geomatics Center of Ministry of Natural Resources, Xi'an 710054, China; 2. The First Topographic Surveying Brigade of Ministry of Natural Resources, Xi'an 710054, China; 3. The First Geographic Information Cartography Institute of Ministry of Natural Resources, Xi'an 710054, China)

Abstract: The accurate and reliable interpretation of regional land cover data is very important for natural resource monitoring and environmental assessment. At present, refined land cover data are mainly obtained by manual visual interpretation, which has the problems of heavy workload and inconsistent interpretation scales. Deep learning has greatly improved the automatic processing and analysis of remote sensing data. However, the accurate interpretation of feature information from massive datasets remains a difficult problem in wide regional land cover classification. To improve the efficiency of deep learning-based remote sensing image interpretation, we selected multisource remote sensing data, assessed the interpretability of the U-Net model based on surface spatial scenes with different levels of complexity, and proposed a new method of stereoscopic accuracy verification (SAV) to evaluate the reliability of the classification result. The results show that classification accuracy is more highly correlated with terrain and landscape than with other factors related to image data, such as platform and spatial resolution. As the complexity of surface spatial scenes increases, the accuracy of the classification results mainly shows a fluctuating declining trend. We also find the distribution characteristics from the SAV evaluation results of different land cover types in each surface spatial scene. Based on the results observed in this study, we consider the distinction of interpretability and reliability in diverse ground object types and design targeted classification strategies for different surface scenes, which can greatly improve the classification efficiency. The key achievement of this study is to provide the theoretical basis for remote sensing information analysis and an accuracy evaluation method for regional land cover classification, and the proposed method can help improve the likelihood that intelligent interpretation can replace manual acquisition.

Keywords: land cover classification; stereoscopic accuracy verification; U-Net; remote sensing; interpretability; reliability

Citation: WANG Xinshuang, CAO Jiancheng, LIU Jiange, LI Xiangwu, WANG Lu, ZUO Feihang, BAI Mu, 2022. Improving the Interpretability and Reliability of Regional Land Cover Classification by U-Net Using Remote Sensing Data. *Chinese Geographical Science*, 32(6): 979–994. https://doi.org/10.1007/s11769-022-1315-z

1 Introduction

Land cover data are indispensable basic data in global climate change research, environmental assessments, natural resource monitoring, and the construction of

broad-area geographic information resources. With socioeconomic development and the intensification of human activities, land cover has undergone significant changes. At this stage, multiplatform, multitemporal, multispectral, multiangle, and multimode integrated re-

Received date: 2022-04-26; accepted date: 2022-06-30

Foundation item: Under the auspices of National Natural Science Foundation of China (No. 41971352), Key Research and Development Project of Shaanxi Province (No. 2022ZDLSF06-01)

Corresponding author: CAO Jiancheng. E-mail: caojc@snsn.mnr.gov.cn; BAI Mu. E-mail: baimu123@163.com

© Science Press, Northeast Institute of Geography and Agroecology, CAS and Springer-Verlag GmbH Germany, part of Springer Nature 2022

remote sensing imaging systems have been developed. However, the processing capability of remote sensing information is insufficient (Quartulli and Olaizola, 2013; Li et al., 2014). This problem is particularly prominent in regional land cover classification when using remote sensing data. Therefore, it is of vital importance to study the rapid and accurate acquisition of Earth's surface features from multisource heterogeneous remote sensing data and improve the intelligent interpretation efficiency.

Large-scale land cover classification is mainly divided into two categories: research experiment-based and engineering operation-based classifications (Weng, 2011; Yuan et al., 2020). The United States and European Union have developed global land cover data products using remote sensing methods, and the spatial resolution has gradually increased from the initial 1° to 8 km, 1 km, and 300 m (De et al., 1998; Loveland et al., 2000; Bicheron et al., 2011). To meet the demand for high-resolution land cover data in global change and earth system research, Chen et al. (2017) developed the first set of global land cover datasets, named GlobeLand 30, using an organic combination of pixels, objects and knowledge; the data have a spatial resolution of 30 m. Zhang et al. (2021) adopted multitemporal and random forest classification methods and developed refined 30 m resolution global land cover classification results. Gong et al. (2019) used a global training set developed in 2015 at a 30 m resolution to classify 10 m resolution images acquired in 2017 and performed the first 10 m resolution land cover mapping tasks based on a random forest classifier. The above macroscopic research results have made important contributions to the maintenance of the global environment and the realization of the United Nations' Sustainable Development Goals. However, engineering applications generally focus on national-scale natural resource management and the provision of accurate geographic information. Moreover, such applications involve many technical factors and have extremely high data quality requirements. The existing image classification algorithms are still unable to meet the needs of large-scale, high-precision land cover mapping and mainly involve visual interpretation methods that require the manual delineation of the boundaries of land cover types. The existing large-scale land cover products are limited by their coarse grain size and insufficient geometric accuracy; con-

sequently, to improve work efficiency, it is necessary to incorporate intelligent interpretation methods that are highly compatible with business processes into practical applications.

Traditional land cover classification methods based on middle- and low-level features provide limited semantic support and low classification accuracy (Guo et al., 2018; Carranza-García et al., 2019). In recent years, deep learning methods that support semantics, such as deep neural networks, have achieved excellent classification results for remote sensing images (Hinton et al., 2006; Zhang et al., 2016), and many methods based on convolutional neural networks (CNNs) have been used in land cover classification (Kussul et al., 2017; Xu et al., 2017; Rezaee et al., 2018; Ma et al., 2020). The related research results showed that CNNs, as deep hierarchical classifiers, can be used to explore the complex spatial patterns hidden in images, extract the semantic features of ground objects, improve the classification ability of land cover compared to traditional automatic interpretation methods, and provide strong generalization capabilities (Meng et al., 2018). At present, the networks used for the semantic segmentation of remote sensing images are extensions of CNNs, and they include fully convolutional networks (FCNs) (Shelhamer et al., 2016), U-Net (Ronneberger et al., 2015), SegNet (Badrinarayanan et al., 2017), DeepLab (Chen et al., 2017; Chen et al., 2018), PSPNet (Zhao et al., 2017), and DeconvNet (Noh et al., 2015). Many scholars have focused on improving and optimizing each network model. Guo replaced standard convolution with dilated convolution to segment high-resolution remote sensing images based on an FCN. After enhancing the training data, they used a Conditional Random Field (CRF) to optimize the boundaries of the segmentation results (Guo et al., 2018). Based on DeepLabv3, Chen segmented high-resolution remote sensing images using an improved Atrous Spatial Pyramid Pooling (ASPP) approach, a fully connected fusion path and a pretrained encoder (Chen et al., 2019). Based on U-Net, Pan integrated a channel attention mechanism and a confrontation network to realize the extraction of building information (Pan et al., 2018). The optimized deep learning model may improve the classification accuracy in a small lab area. However, for a large scale classification with various land cover types, the results show little diversity among different networks. Therefore, U-Net is

chosen because it can obtain effective classification results with a smaller sample size. In addition, the U-Net model has achieved excellent segmentation results in various applications (Zhang *et al.*, 2018; Shamsolmoali *et al.*, 2019; Wang *et al.*, 2020) and has become the preferred model for various remote sensing image segmentation applications.

When implementing land cover classification by using deep learning methods in the engineering field, various problems are encountered, such as those associated with complex and changeable scenes, the poor generalization of samples and training models, the amounts of data required for full coverage, and data differences in spatial resolution and quality. However, the influence of various complex factors on image classification accuracy and the best way to improve data processing efficiency have not been fully researched, and an accuracy evaluation method for the engineering field is not currently available. In order to provide the theoretical basis for the complex factors of the deep learning land cover classification, especially the influence of surface spatial scene, and an accuracy evaluation method for classification results in the engineering field, we selected several study areas with typical topographic features and surface landscapes, using remote sensing data from different satellite platforms with different spatial resolutions. Then, the U-Net architecture was used for training classification models to analyze the interpretability and reliability of land cover extraction in remote sensing images based on deep learning technology. Additionally, we also proposed a stereoscopic accuracy verification (SAV) method suitable for operational applications to evaluate the reliability of classification results. On this basis, we presented an intelligent image interpretation scheme suitable for engineering applications based on a deep learning method. The research could provide a scientific basis for improving the efficiency of remote sensing image interpretation for the regional land cover classification.

2 Data and Methods

2.1 Study areas

Considering the distribution of terrain and landscape that may be encountered in large-scale engineering applications, we select four types of surface spatial scenes reflecting the gradual transition of terrain and landscape

from simple to complex. Two typical study areas are selected for each surface spatial scene. The distribution of the study areas is shown in Fig. 1, and detailed information is shown in Table 1. The topography of the study area covers plains, hills, terraces, mountainous terrain and topographically heterogeneous areas. The land cover types include forest, cropland, water bodies, residential land, mudflats, vegetated wetlands and other different surface landscapes. The surface landscape of the study area exhibits a uniformly distributed and well-proportioned spatial pattern, as well as a clustered and hierarchical spatial pattern. Additionally, economically developed areas with complex and diverse features are present, as are areas with sparse human habitation and a predominant desert ecosystem. In short, the selected study areas cover almost all surface forms common in the task of regional land cover classification.

2.2 Data and processing

Several remote sensing images are selected that cover eight study areas from large datasets formed in our department. The satellite images for training and testing mainly include ZY3, GF1, GF2, WV2 and Sentinel 2, which are widely used in various project types. The spatial resolution of the data ranges from 0.5 m to 10 m. The image phase is the growing season; the cloud cover is less than 5%. We use images that include four bands (blue, green, red, NIR) of spectrum information to perform the model training and classification task in all study areas. There is no obvious banding or noise in the images, and the greyscale image is generally normally distributed. Detailed information is shown in Table 1. To ensure the validity of accuracy verification, test data usually can not overlap with the sample data used in model training. Therefore, a scene image is divided into two nonoverlapping areas: one area for model training and one for testing.

To ensure the high precision of sample labelling, we use the manually annotated method to obtain the sample data since the reliability of labelled data can impact deep learning model performance. In the process of sample labelling, we select continuous or multiple independent areas that can represent the landscape form of the whole test area based on remote sensing images, and each pixel in the region has corresponding category attribute information. At the same time, we strictly abided by scientific and reasonable labelling rules for the best

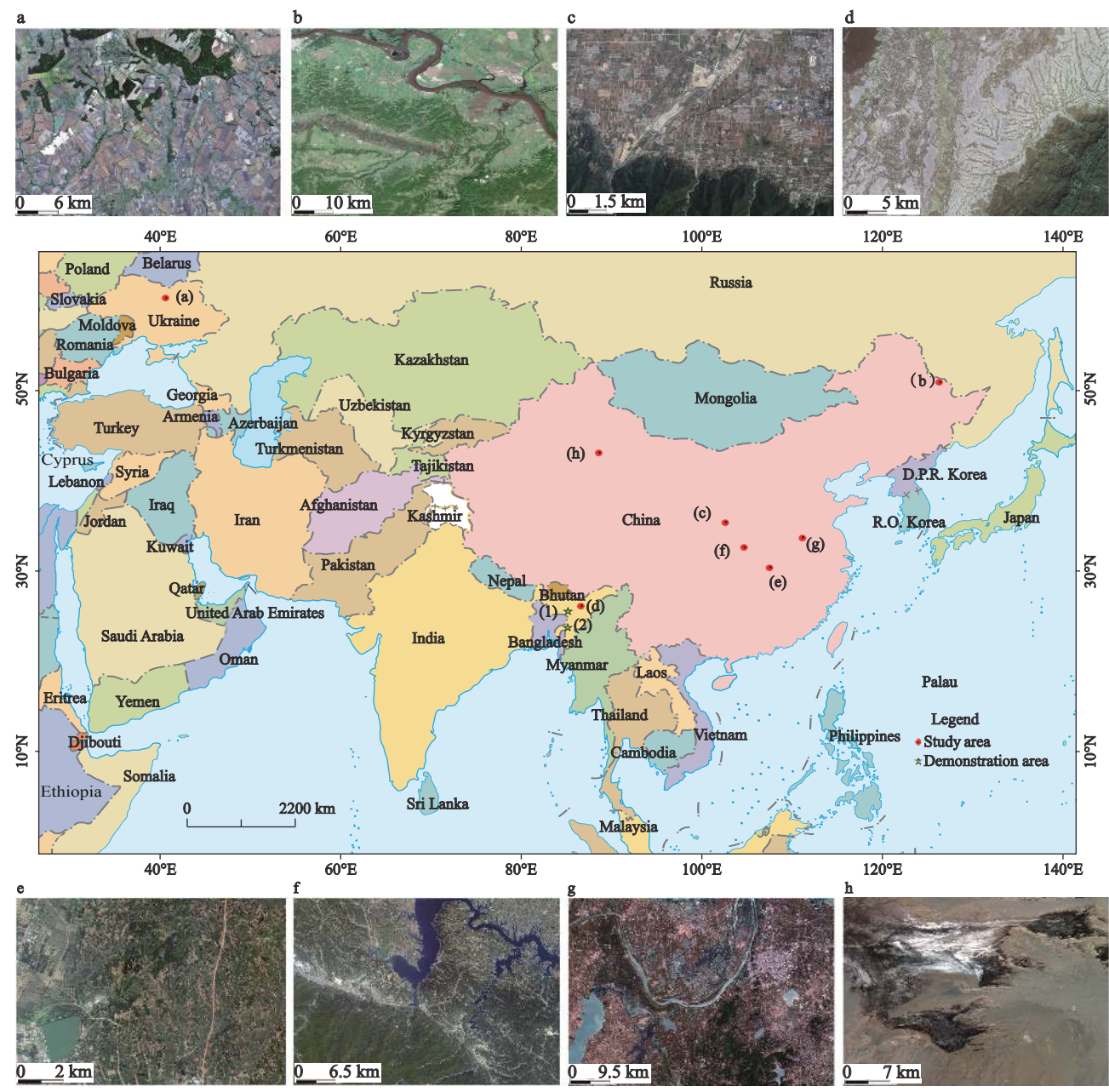


Fig. 1 The distribution of study areas

Table 1 Detailed information on the study areas and remote sensing data

Study area	Location	Terrain	Dominant land cover type	Sensor	Spatial resolution / m
a	Cherkasy, Ukraine	Plain	Cropland, forest, residential land, water body	ZY3	2.0
b	Heilongjiang, China	Plain	Cropland, forest, residential land, water body, grassland	Sentinel 2	10.0
c	Shaanxi, China	Plain, mountain	Cropland, residential land, forest	GF1	2.0
d	Assam/Nagaland, India	Plain, hill	Cropland, forest, grassland, residential land, water body	ZY3	2.0
e	Hubei, China	Hill, terrace	Forest, paddy field, cropland, residential land, water body	WV2	0.5
f	Henan, China	Mountain, valley	Forest, water bodies, mud flat, cropland	GF1	2.0
g	Jiangsu, China	Plain	Cropland, forest, residential land, water body	GF2	0.8
h	Gansu, China	Plain, Gobi desert	Grassland, vegetated wetland, barren land	ZY3	2.0

Note: study areas a–h see Fig. 1

semantic consistency, such as the representativeness, the equilibrium among classes, the homogeneity within classes, and the consistent sample size in each study area. The main land cover types of the sample data are cropland, forest, grassland, residential land, water bodies and roads, which reflect actual landscape features.

2.3 Methodology

2.3.1 Semantic segmentation model

We select the U-Net (Ronneberger et al., 2015) semantic segmentation model for training based on sample data. Each hidden layer of the U-Net model has multiple feature dimensions, making it conducive for learning diverse and comprehensive features. The U-shaped architecture of the model makes the image clipping and mosaicking process intuitive and reasonable. The combination of high-level features and low-level feature maps, as well as the repetition and continuity of convolution, enables the model to combine contextual information and detailed information to obtain accurate input feature maps.

The network architecture we used in this paper is the traditional U-Net which is composed of a contracting path and an expansive path. The contracting path consists of four blocks, and each block has the repeated application of two convolutions with a given size kernel, followed by the Rectified Linear Unit (ReLU) function and a 2×2 max pooling operation for downsampling. The process of the expansive path has two operations: an upsampling of the feature map followed by a 2×2 convolution that halves the number of feature channels with the correspondingly cropped feature map from the contracting path, and two convolutions that are each followed by a ReLU. The final layer is a one-by-one convolution to map each feature vector to the desired number of classes. Instead of using the soft-max method as the loss function, it is computed by a pixelwise soft-max over the final feature map combined with the cross-entropy loss function.

Considering various research needs, this paper segments the cropland, forest, water body, residential land and other ground object categories from satellite images, and each set of interpretation experiments uses the same amount of sample data. This research is based on the TensorFlow framework. After fine-tuning of this U-Net architecture, the corresponding classification model is trained for every sample data; ultimately, the test data

are used to obtain the final classification result.

2.3.2 Stereoscopic accuracy verification

We propose the stereoscopic accuracy verification (SAV) method to evaluate the reliability of land cover classification results based on the overall attribution accuracy (OAA) and edge localization accuracy (ELA) of land cover types obtained from remote sensing images. This method is highly compatible with most application requirements and can comprehensively evaluate the accuracy of land cover classification in multiple scenarios from multiple dimensions (Ronneberger et al., 2015).

A confusion matrix is used to evaluate the OAA of the classification results. According to previous studies (Congalton, 1988; 1991; Pugh and Congalton, 2001), random sampling or stratified random sampling of sufficient points is a suitable sampling strategy for the accuracy assessment of image classification considering the spatial autocorrelation of errors. Considering this, the random sampling points are generated uniformly in a grid of 8×8 pixels in a scene from a prediction image in our study. According to the actual landscape, we ensure that there are no less than 200 sampling points in each scene, and the true attribute value is obtained by manual annotation based on high-resolution images.

To meet the application requirement of land cover classification accuracy assessment, the ELA method is used to evaluate the usability of the result. We selected five to six sampling areas in each group of predicted classification results and manually labelled the ground truth. The minimum root mean square (MRMSD) distance between each prediction edge pixel and labelled edge pixels was calculated. Based on the standard edge accuracy of two pixels, ELA is the percentage of the number of predicted edge pixels for which the MRMSD is smaller than two, as shown in Equation (1).

$$ELA = \frac{\sum_{i=1}^{n_p} D_i}{n_p}$$

$$D_i = \begin{cases} 1, & \min \left(\sqrt{(p_i(x) - gt_j(x))^2 + (p_i(y) - gt_j(y))^2} \right) \leq 2, \\ & j = 1, 2, \dots, n_{gt}, i = 1, 2, \dots, n_p \\ 0, & \text{else} \end{cases} \quad (1)$$

where $p_i(x)$ and $p_i(y)$ denote the column and row position of the i th point $p_i(x, y)$ from prediction edge, respectively. $gt_j(x)$ and $gt_j(y)$ denote the column and row

position of the j th point $gt_j(x, y)$ from labelled edge. n_p and n_{gt} represent the number of edge points of the prediction edge and labelled edge, respectively. The range of ELA is from 0 to 1. The ELA reaches its best value at 1 and worst at 0.

The low OAA can infer that ELA must be low. However, the high OAA result can not represent a good ELA result. Therefore, a comprehensive index from OAA and ELA was proposed to evaluate the result. The SAV was obtained by a weighted average of OAA and ELA, as shown in Eq. (2). Considering the importance of usability of the results, the ELA was given a larger weight of 0.7. Therefore, the range of SAV is the same as those of OAA and ELA, from 0 to 1. SAV reaches its best value at 1 and worst at 0.

$$SAV = 0.7 \times ELA + 0.3 \times OAA \quad (2)$$

2.3.3 Overview research route

Land cover complexity varies noticeably across countries of the world. This paper divides surface spatial scenes into four types based on the actual complexity of the Earth's surface and selects eight typical study areas as cases. Based on these cases, a systematic classification tests are implied to figure out the relationship between landscape and deep learning classification ability. The complexity is defined based on based on the four principles of topography conditions, compositional complexity, configurational complexity and temporal complexity. Specifically, from scene 1 to scene 4, the number of land cover categories on a landscape is more abundant, the patches of the landscape are increasing,

and the attribute diversity of land cover types is more flexible. In each set of tests, the U-Net semantic segmentation model is used to train and predict images. And the interpretability and reliability of the classification results are analyzed based on the SAV method. Through comparing the deep learning interpretation ability of remote sensing images for different terrain and landscape conditions and classification reliability of the model from different satellite platforms and different spatial resolutions, we assess the performance of the deep learning model in interpreting land cover in remote sensing images and further explore the efficient application of this model in large-scale land cover remote sensing image scene classification. The overall research framework is shown in Fig. 2.

For each test, the method mainly consists of three steps, preprocessing the images and labels, training the classification model by U-Net and testing the interpretability and reliability of the trained model. Since the limitation of GPU (Graphic Processing Unit), the large size remote sensing images can not process in one pass. Therefore, first, images and labels are split into smaller patches of $512 \times 512 \times n$ (n represents the number of image channels) pixels and 512×512 pixels, respectively, with a sliding window method. The patches are the inputs to train and test the model. To increase the number of training samples, a stride of 128 is given. The total number of samples is different for different images but at least ten thousand for each test. We use the same model architecture for each test. The kernel size of calculating features is important to the accuracy of the res-

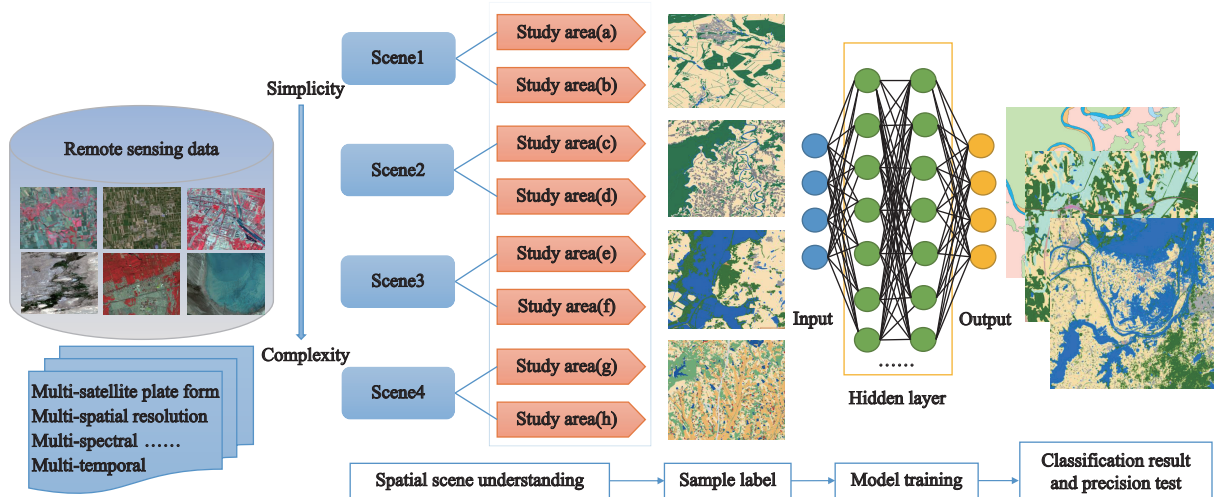


Fig. 2 Overall framework for the interpretability and reliability analysis of land cover classification on different complexity scene

ult, especially in remote sensing image processing. To provide a reasonable trade-off between the receptive field and the computation, a constant kernel size of 5×5 over all the layers is chosen. According to the memory of the GPU, we set the batch size and epoch to 10 and 30, respectively. The stochastic gradient descent (SGD) method is used for training, and the learning rate is set to 0.01. To smooth the results and remove the discontinuities along the boundary, the stride of the sliding window must be smaller than the patch size when testing an image. In the test, the stride is 128, and the result is obtained by averaging the predictions of overlapping pixels.

3 Results

3.1 Qualitative description of classification results

The results of land cover classification based on remote sensing data and U-Net model are shown in Fig. 3. Scene 1 (Figs. 3a–3b) is dominated by plains, with low topographic relief, low ground feature patch fragmentation, regular land cover types, distinct gradation, and clear object outlines. The land cover types mainly include cropland, forest and residential land, and satisfactory classification results are obtained. In scene 2

(Figs. 3c–3d), there are plains, mountains and hills, the distribution of ground objects is relatively regular. The classification results for cropland, forest and residential land are generally good, but the classification results for grassland and roads are relatively poor due to the lack of samples, which results in imbalanced model training. Scene 3 (Figs. 3e–3f) is a complex area with interconnected hills, terraces, mountains and river valleys and large terrain undulations. Each land cover type is present in fragmented patches, which are often misclassified. However, the mudflats distributed along the river valley display clear differences in colour and texture from other land cover types. Therefore, a good classification result is achieved. Scene 4 (Figs. 3g–3h) is the most complex surface landscape in land cover classification based on remote sensing images in this study. This region includes economically developed suburban areas, where the distribution of land cover types is complicated and land use patches are fragmented. The phenomenon of the same objects but with different spectra is serious in remote sensing images, and the classification results for ground objects are relatively poor; therefore, the accuracy requirements for land cover classification in the engineering field are not met. A natural landscape that has not been disturbed by humans, in this

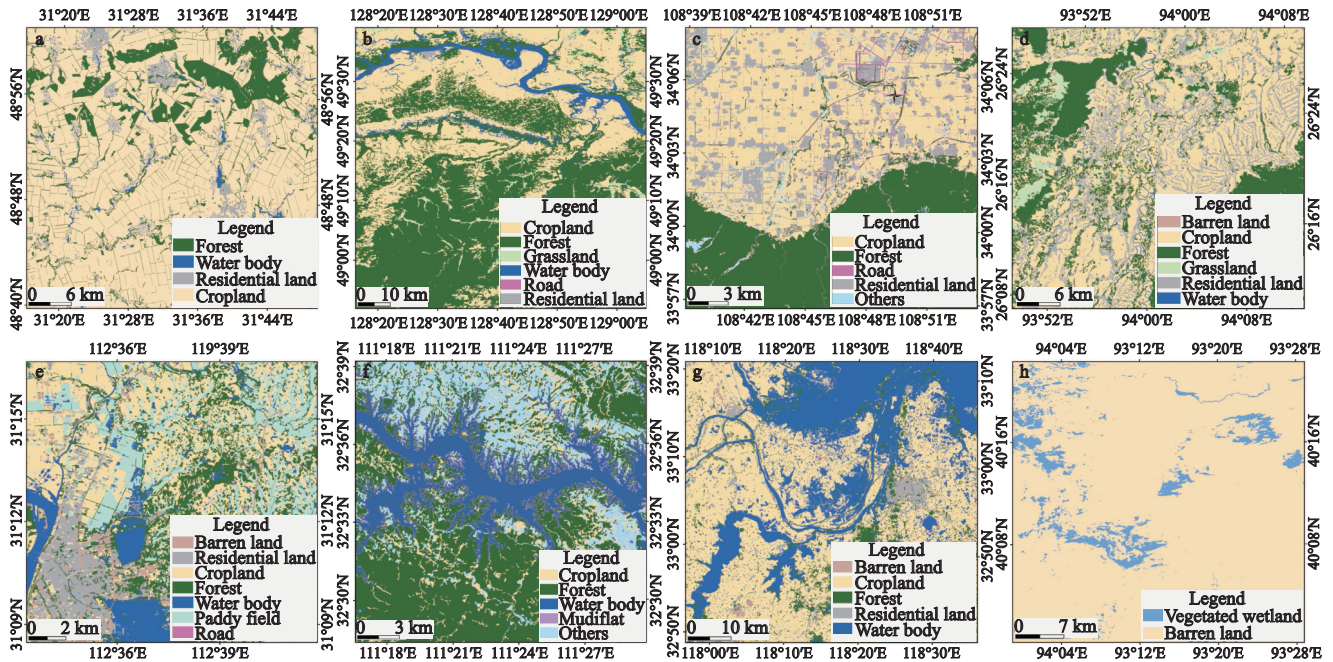


Fig. 3 Land cover classification results on different complexity scene: Cherkasy based on ZY3 (a) and Heilongjiang based on Sentinel 2 (b) of scene 1, Shaanxi based on GF1 (c) and Assam/Nagaland based on ZY3 (d) of scene 2, Hubei based on WV2 (e) and Henan based on GF1 (f) of scene 3, Jiangsu based on GF2 (g) and Gansu based on ZY3 (h) of scene 4

case, the Gobi Desert, is also considered; this landscape includes desert vegetation belts, vegetated wetlands, and other features. The boundaries of ground objects are blurred, and it is difficult to manually label samples. Therefore, the accuracy of the automatic classification results is relatively poor for this scene.

Figs. 3a–3b show the satisfactory classification result in scene 1, both in test area Cherkasy covered with a 2 m spatial resolution image and in Heilongjiang covered with a 10 m spatial resolution image, and almost all forest, water body and cropland are correctly classified. However, as shown in Figs. 3e–3f, the classification results of water bodies, forests and residential areas are better in the 2 m spatial resolution images than in the 0.5 m spatial resolution images. This is because the abundance of categories in test area Hubei is higher than that in Henan. The classification results for a more categories are vulnerable to the complex scene with unclear boundaries and fragmented land types, such as barren land, paddy fields and roads. Through comparing the classification results obtained based on commonly used data sources of different resolutions (0.5 m, 0.8 m, 2 m, 10 m) and satellite platforms (WV2, GF1, GF2, ZY3, Sentinel 2) from Figs. 3a–3h, the classification result is mainly affected by land cover categories and landscape composition.

3.2 Quantitative accuracy assessment of classification results

The SAV method is used to evaluate the accuracy of the classification results for eight study areas with four types of surface spatial scenes (Table 2). In general, as landscape diversity increases from scene 1 to scene 4, the complexity of the surface spatial scene increases, the degree of fragmentation of classes also increases, and the phenomenon of ‘same spectrum but different objects and same objects but different spectra’ appears frequently. Therefore, the accuracy of the classification results mainly shows a fluctuating declining trend.

Table 2 shows that high SAV accuracy is achieved for water bodies in all scenes, most OAA and ELA are larger than 0.85 and the SAV accuracy is larger than 0.900 in every scene. This means that most of the water body classification results can be used in applications directly. However, as in the case that many water-filled pits and ponds with different shapes and spectral information are distributed in residential parts of study

area Assam and Nagaland, the OAA of the classification result for water bodies is relatively poor, with a value of 0.831. This is because most ponds are surrounded by dense vegetation that produce shadows that influence the accuracy of water body classification. High SAV classification accuracy is achieved for cropland and forest in the two study areas in scene 1, and the automatic classification results can be used in engineering applications to improve the work efficiency. However, the corresponding classification accuracy in scene 2 and 3 is relatively low. The cropland in study area Henan is mostly distributed on both sides of valleys with obvious topographical undulations, and the shape of these areas is complex and variable. Therefore, the SAV classification accuracy is the lowest among all scenes, and the SAV is only 0.737. Similarly, the OAA and ELA of the cropland and forest classification results in scene 4 can not meet the requirements of engineering practice. The classification results of residential land are fairly good among all scenes in which the SAVs are all higher than 0.800. This land cover type can be extracted using deep learning methods according to specific situational requirements. Grassland areas display the same spectrum but contain either different objects or the same objects. However, they display different spectra in most scenes; thus, the classification accuracy for most scenes is low, and it is difficult to obtain accurate extraction results. Due to the large differences in the distribution of roads in various regions in large-scale scenarios, the road samples are heterogeneous, and roads are easily obscured by the tree canopy along streets. Therefore, the classification accuracy of roads in all scenes is relatively low. Although the spatial scene in study area Henan is complex, the classification results of mudflats are relatively ideal (SAV is 0.904) because of the consistent spectral characteristics of mudflats, which are quite different from other ground object types.

On the other hand, the results of land cover types in Heilongjiang show higher classification accuracy; however, the spatial resolution of remote sensing images covering this area is 10 m, which is the lowest in this study. Further analysis in detail shows that the accuracies for forest and residential land are higher than those in study area Shaanxi to Gansu, and the spatial resolutions of the images covering these areas include 0.5 m, 0.8 m and 2 m, all of which are lower than 10 m.

Table 2 Accuracy of land cover classification on different complexity scene

Land cover type	Scene 1						Scene 2						Scene 3						Scene 4					
	Area a		Area b		Area c		Area d		Area e		Area f		Area g		Area h									
	OAA	ELA	SAV	OAA	ELA	SAV	OAA	ELA	SAV	OAA	ELA	SAV	OAA	ELA	SAV	OAA	ELA	SAV	OAA	ELA	SAV	OAA	ELA	SAV
Barren land	–	–	–	–	–	–	–	0.698	0.681	0.686	0.625	0.520	0.552	–	–	0.607	0.451	0.498	0.561	0.392	0.443	–	–	–
Cropland	0.994	0.952	0.965	0.829	0.852	0.845	0.894	0.825	0.846	0.971	0.810	0.858	0.858	0.810	0.824	0.766	0.725	0.737	0.782	0.722	0.740	–	–	–
Forest	0.917	0.961	0.948	0.915	0.891	0.898	0.798	0.819	0.813	0.836	0.783	0.799	0.887	0.741	0.785	0.713	0.740	0.732	0.702	0.697	0.699	–	–	–
Grassland	–	–	–	0.508	0.718	0.655	–	–	–	0.336	0.451	0.417	–	–	–	–	–	–	–	–	–	–	–	–
Mudflats	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.885	0.912	0.904	–	–	–	–	–	–	–
Paddy fields	–	–	–	–	–	–	–	–	–	–	–	–	0.755	0.850	0.822	–	–	–	–	–	–	–	–	–
Residential land	0.944	0.919	0.927	0.882	0.859	0.866	0.821	0.811	0.814	0.826	0.800	0.808	0.812	0.823	0.820	–	–	–	0.803	0.815	0.811	–	–	–
Roads	–	–	–	0.111	0.22	0.187	0.183	0.225	0.212	–	–	–	0.502	0.310	0.368	–	–	–	–	–	–	–	–	–
Vegetated wetlands	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.875	0.856	0.862	–	–	–
Water body	0.993	0.975	0.980	0.887	0.958	0.937	–	–	–	0.831	0.972	0.930	0.867	0.920	0.904	0.915	0.931	0.926	0.872	0.926	0.910	–	–	–

Notes: OAA means the overall attribution accuracy; ELA means the edge localization accuracy; SAV means the stereoscopic accuracy verification calculated from OAA and ELA; – means that there is no corresponding land cover type in the test area

Cropland and grassland also show good classification accuracy. Since the road is narrow and will be blocked by vegetation, it is easy to lose information in the process of dimensionality reduction, and there are missing and inaccurate points in the classification results; therefore, low spatial resolution will have a certain impact on the classification accuracy, although the scene has the most significant impact on the classification results. We then analyze the classification accuracy of land cover types at different resolution scales in the same scene. As shown in Figs. 3e–3f for scene 3, the SAV accuracy of water bodies is 0.926 in area Henan and 0.904 in area Hubei. Other accuracy verification results show that the higher the spatial resolution of the image is, the higher the classification accuracy in the same scene. The analysis results show that under certain conditions of the scene, most of the classification results are proportional to the spatial resolution, and some have no relationship with the resolution, such as water bodies, which are mainly affected by shadows that are easily confused with water bodies in different test areas. We also analyze the classification accuracy of land cover types in different scenes at the same spatial resolution scale of 2 m. From the accuracy evaluation results of area Cherkasy, area Shaanxi, area Henan and area Gansu, it can be found that with the increase in landscape diversity, the SAV classification accuracy of each land type decreases, such as barren land, cropland, forest, grassland, residential land and water body.

4 Discussion

Land cover classification based on remote sensing images is an important research direction, and it is of practical application value to further improve the accuracy of automatic classification results. He proposed a multispectral land cover classification method based on a deep learning model and effectively improved the accuracy of multispectral image classification (He and Wang, 2021). In the refined classification of complex surface scenes that include diverse information, single-modality-dominated deep networks are inevitably limited in classification tasks. Hong provided a baseline solution to the aforementioned issue by developing a general multimodal deep learning (MDL) framework, evaluated different fusion strategies, trained deep networks and built a new network architecture (Hong et al.,

2020). At this stage, most research has focused on improving the classification accuracy of land cover types in areas with complex surfaces through model modification. However, the original objective of our study was to simplify the representation of earth's complex surface and the related classification problems. We conducted an in-depth analysis of surface spatial scenes and land cover types and explored how to combine deep learning with manual visual inspection to improve the efficiency of engineering work while ensuring a high interpretation accuracy. Therefore, from the perspective of improving the accuracy of land cover classification, our work is a complement to other methods. To demonstrate the validity of our research results, we further explored application cases involving engineering projects.

4.1 The effect of image quality on classification results

We use sharpness as an evaluation metric to further assess the interpretability of images with different qualities. Many satellite images are required to cover the working area in regional land cover classification at a large scale. Generally, cloud cover is less than 5%, there is no obvious band or noise interference, and the grey-scale range is normally distributed. However, in reality, poor-quality images are generally present in datasets. Therefore, it is necessary to conduct an in-depth analysis of image interpretability and reliability to better guide engineering operations. It is difficult to quantitatively evaluate remote sensing images of poor quality (Guo et al., 2020), although scholars have investigated many evaluation indicators related to image quality, such as the ground sampling distance, signal-to-noise ratio, and information entropy (Wang et al., 2004; Yuan et al., 2014). These indicators can only describe some of the characteristics of an image, and currently, no indicators can be used to fully evaluate the information in remote sensing images. Therefore, this paper adopts a visual subjective evaluation method to intuitively assess image quality based on image sharpness (Gastaldo et al., 2005).

With the data and classification results for Henan study area, we select GF1 scene data from an adjacent orbit with the same spatial resolution and time phase for a comparative analysis. One of the scenes is characterized by good sharpness, and the ground features are highly distinguishable. The other scene is characterized

by relatively poor image quality. In the overlapping area of the two images, the same samples and methods are used for model training and classification, and SAV evaluation is performed to assess the impact of image quality on deep learning classification. The classification results are shown in Fig. 4, and the accuracy statistics are shown in Table 3. The results show that for remote sensing images with different qualities, the classification results based on deep learning are highly variable. High-quality images yield higher classification accuracy than low-quality images, and low-quality images lead to comparatively more misclassifications for land cover types other than water.

4.2 The effect of sample size on classification results

We take the highest efficiency as the goal and discuss the optimal sample size. Cevikalp showed that the classification accuracy improved as the size of the training dataset increased (Cevikalp et al., 2020). However, as the amount of data included in model training increases indefinitely, the robustness of the model begins to decline beyond a specific point. Therefore, we used data for study area Assam and Nagaland to determine the optimal sample size in practice. We select sample sizes of 12.5%, 25.0%, 50.0% and 75.0% of the whole image. To ensure that the test results are not affected by other factors, mountainous and urban areas with a uniform

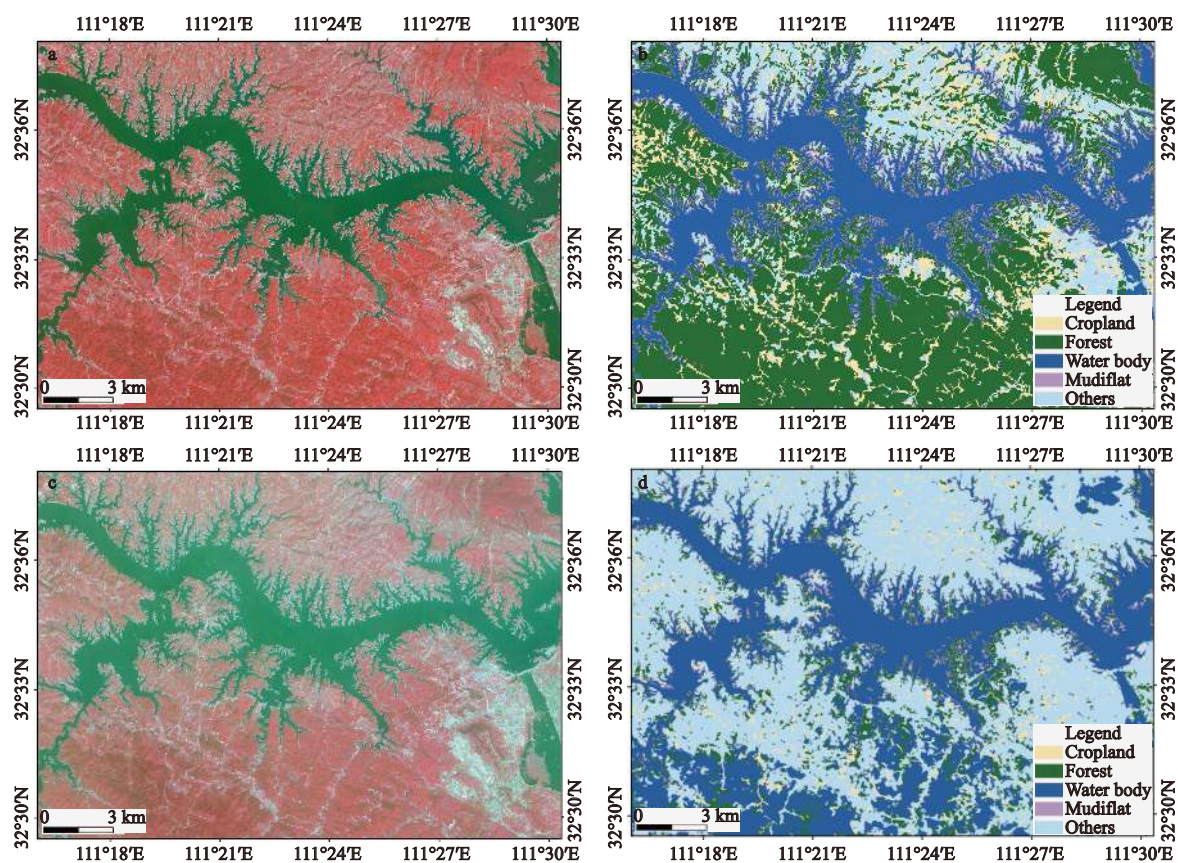


Fig. 4 Classification results with different image quality: (a) high-quality image (R: NIR, G: Red, B: Green), (b) classification result of image (a), (c) low-quality image (R: NIR, G: Red, B: Green), (d) classification result of image (c)

Table 3 Accuracy statistics for the classification of images with different qualities

Image qualities	Cropland			Forest			Water body			Mudflats			Other		
	OAA	ELA	SAV	OAA	ELA	SAV	OAA	ELA	SAV	OAA	ELA	SAV	OAA	ELA	SAV
High quality	0.981	0.922	0.9397	0.987	0.927	0.945	0.993	0.952	0.9643	0.885	0.882	0.8829	0.784	0.756	0.7644
Low quality	0.374	0.285	0.3117	0.374	0.306	0.3264	0.931	0.904	0.9121	0.321	0.283	0.2944	0.608	0.401	0.4631

Note: meanings of OAA, ELA and SAV see Table 2

distribution and located outside the sampling area are selected as the classification areas; then, the model training and classification prediction tasks are performed. The classification results based on different sample sizes are shown in Fig. 5, and the accuracy evaluation results are shown in Table 4. The main conclusion is that the classification effect is significantly improved as the sample size increases. When the sample size is small, we obtain a poor classification result with

a high degree of fragmentation. With increasing sample size, the classification results become increasingly complete, and the boundaries among land cover types become more accurate. However, we find that, the classification accuracy increases only slightly, and the accuracy of individual land cover types may even slightly decrease in some areas. Therefore, the relationship between the relevant workload and the required accuracy must be considered in the process of land cover

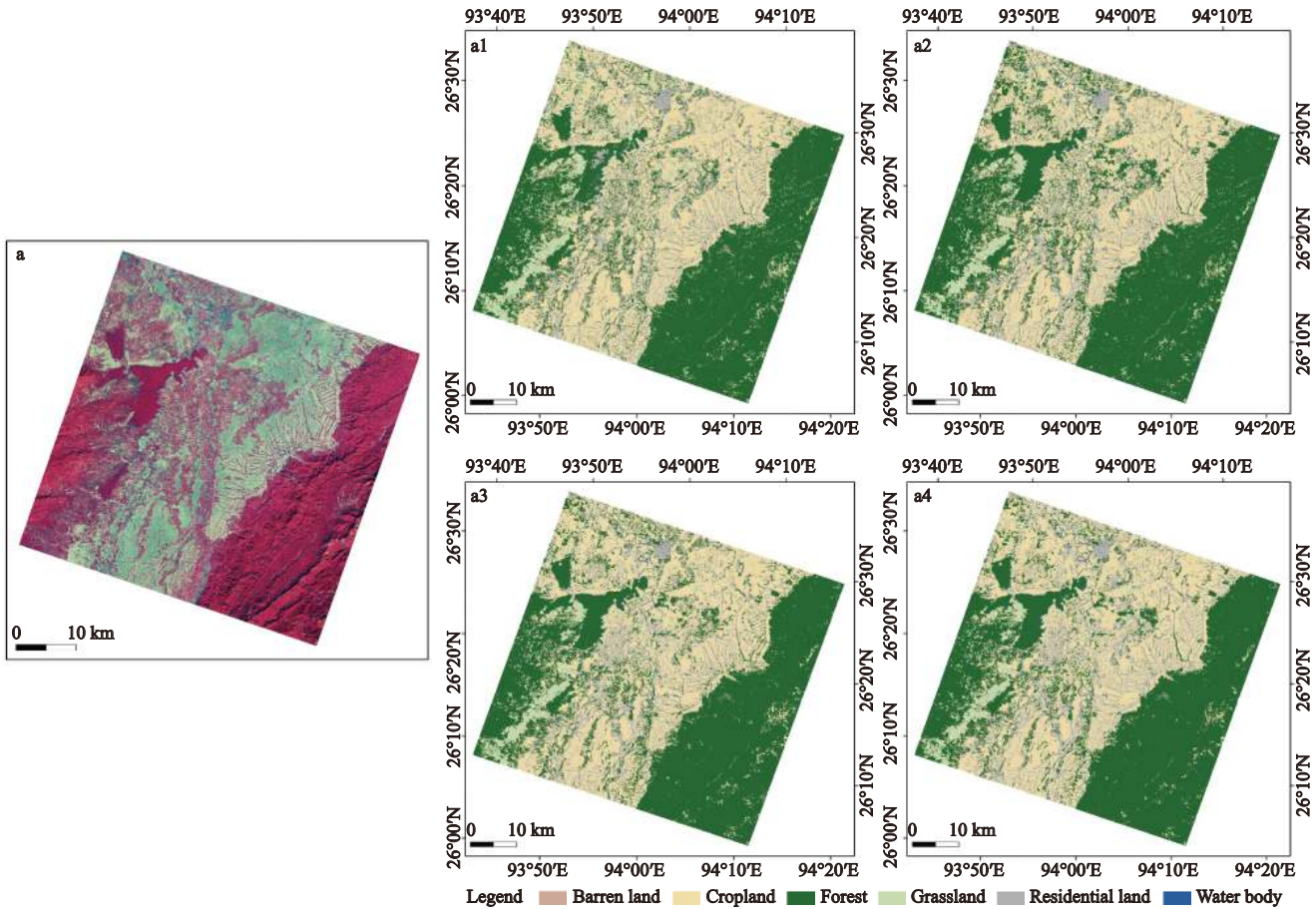


Fig. 5 Classification results with different sample size in the same region: (a) training and testing image (R: NIR, G: Red, B: Green), a1–a4 are the classification results with different sample size of 12.5%, 25.0%, 50.0% and 75.0% of entire image (a), respectively

Table 4 Accuracy statistics for the classification of images with different sample size

Sample size / %	Barren land			Cropland			Forest			Grassland			Residential land			Water body		
	OAA	ELA	SAV	OAA	ELA	SAV	OAA	ELA	SAV	OAA	ELA	SAV	OAA	ELA	SAV	OAA	ELA	SAV
12.5	0.487	0.405	0.430	0.669	0.701	0.691	0.725	0.755	0.746	0.252	0.311	0.293	0.805	0.811	0.809	0.562	0.502	0.521
25.0	0.543	0.552	0.549	0.872	0.823	0.838	0.787	0.802	0.798	0.305	0.408	0.377	0.811	0.819	0.817	0.611	0.551	0.569
50.0	0.665	0.712	0.698	0.970	0.935	0.946	0.836	0.813	0.820	0.331	0.456	0.419	0.825	0.852	0.844	0.823	0.852	0.843
75.0	0.698	0.705	0.703	0.971	0.932	0.944	0.831	0.811	0.817	0.336	0.452	0.417	0.826	0.852	0.844	0.811	0.853	0.840

Notes: Sample size means the proportion of sample of an image

classification. It is necessary to comprehensively consider the time cost and select the most appropriate sample size to maximize efficiency.

4.3 Engineering demonstration areas for application

Finally, we apply the research results in two engineering demonstration areas (Fig. 1), analyze the surface spatial scenes and determine the relevant classification rules. In demonstration area one, most of the residential land types are covered by canopies, and roads are relatively concentrated. Therefore, by merging the residential land with forest, roads and adjacent ground objects based on automatic interpretation with deep learning and manually modifying the land cover types with lower

accuracy, a final result that meets the relevant technical specifications can be obtained (Fig. 6). Compared with that of fully manual visual interpretation, the efficiency of the combined approach is 10% higher.

The demonstration area two is a seaside area with few types of ground objects, but large numbers of paddy fields and breeding ponds with high surface areas are interspersed in the region; consequently, the manual interpretation workload is very high. Therefore, after comprehensive evaluation, we automatically extracted only the paddy fields and performed binary classification (Fig. 7). Notably, only a small amount of manual intervention is needed to obtain the final classification result, and the efficiency of this approach is 30% higher than that of fully manual visual interpretation.

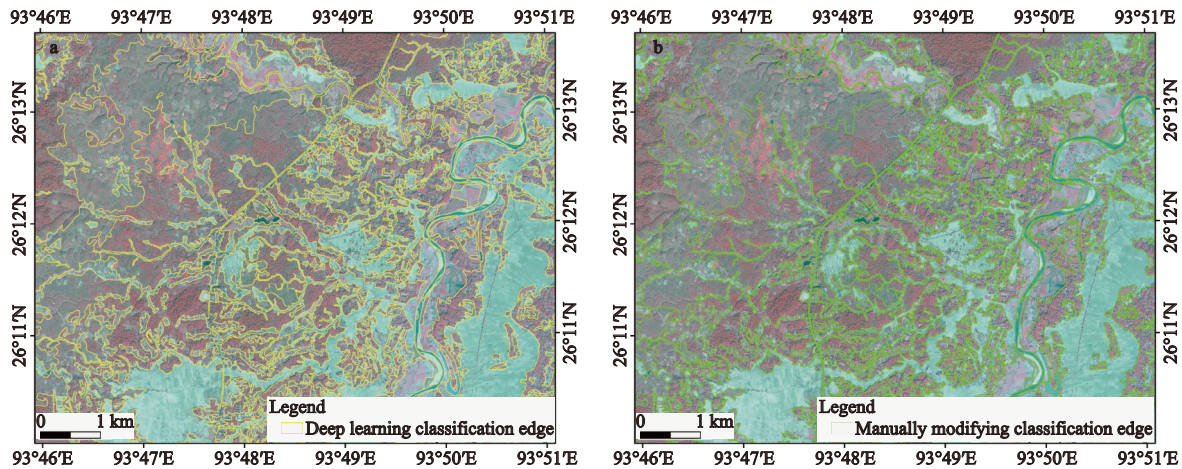


Fig. 6 Classification result of demonstration area one in Fig. 1: (a) deep learning classification result based on U-Net, (b) manually modifying result based on (a) (R: NIR, G: Red, B: Green)

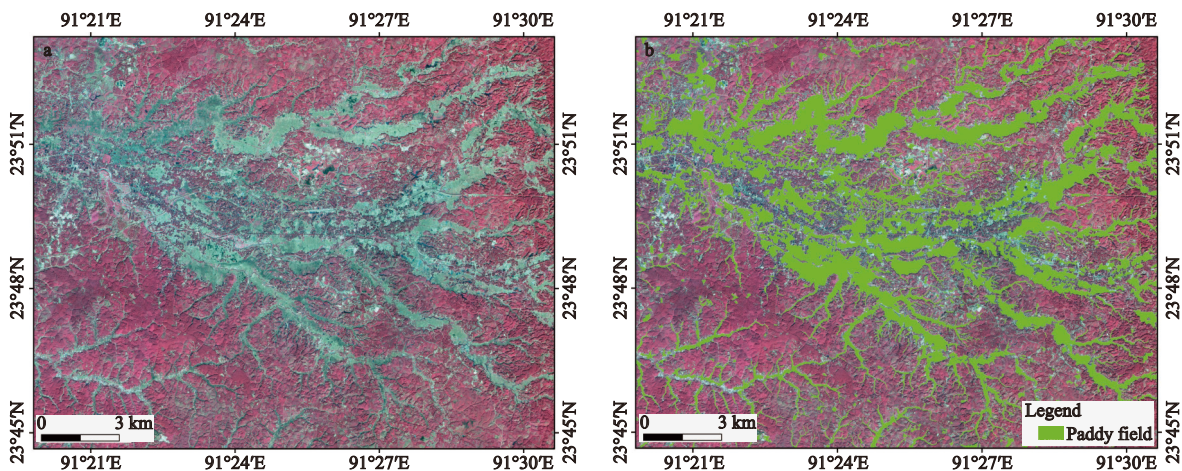


Fig. 7 Classification result of demonstration area two in Fig. 1: (a) image with large numbers of paddy fields, (b) result based on U-Net (R: NIR, G: Red, B: Green)

The insights gained from this study may help to improve the application of artificial intelligence methods in remote sensing interpretation in the engineering field. The classification of large-scale land cover datasets should be based on the core principle of quality over quantity. It is necessary to conduct comprehensive and detailed data analyzes of the surface landscape and extract the land cover types with few limiting phenomena, clear distributions, and high semantic consistency for automatic interpretation. On this basis, manual subdivision and attribute labelling can be performed according to the relevant technical requirements to minimize the workload of manual correction for misclassified results. Additionally, it is necessary to weigh the relationship between work efficiency and sample size in practice to meet the classification requirements for a given situation. The results can be used to minimize misclassification issues and be efficiently integrated with the results of manual visual interpretation, which demonstrates a strong effect of improving image interpretation efficiency and reducing the workload of manual collection.

5 Conclusions

In this study, we focused on the interpretability and reliability of regional land cover classification performed by U-Net using remote sensing data. The interpretability was assessed based on different surface spatial scenes, and the reliability of interpretation for each land cover type was quantitatively analyzed by the proposed SAV method. Our results provide a theoretical and scientific basis for quickly and accurately obtaining ground feature information from multisource heterogeneous remote sensing data in the engineering field, and provide ideas for the study of the private network model of ground object recognition. The main conclusions can be drawn as follows:

(1) The interpretation ability of remote sensing images is highly related to terrain and landscape; the accuracy of classification results is mainly affected by land cover categories and landscape composition and not highly correlated to satellite platform and spatial resolution. When the data with same spatial resolution and platform, it could be found that as the landscape diversity increase, the SAV classification accuracy of each land type decreased.

(2) The proposed accuracy verification method of

SAV was used to evaluate the classification results of each surface spatial scene. We found that areas with few ground objects, clear distributions, distinct boundaries, and few land use homeomorphisms are generally associated with good classification accuracy. Areas with complex geographical categories, relatively unclear boundaries and the widespread phenomenon of the same objects but with different spectra, the accuracy of the classification results is generally poor. The experiment results also showed that SAV is an effective method to evaluate the remote sensing intelligent interpretations from the perspective of engineering application requirements of large-scale classification tasks.

(3) The interpretability and reliability of remote sensing images are highly correlated with the ground object being classified, and the correlations with satellite platform and spatial resolution are relatively low. Better classification results can be obtained for water bodies with spectral characteristics that differ from those of other land cover types, such as cropland, forestland and residential land, with regular shapes. Satisfactory classification results can also be obtained for nondominant land cover types, such as mudflats and vegetated wetlands, which are relatively different from other ground objects based on their spectral characteristics and distributions. The classification accuracy is relatively low for land cover types distributed in areas with long and narrow shapes; that are easily shaded by canopy vegetation (e.g., roads, canals). Moreover, areas characterized by the same spectrum but different objects or the same objects but different spectra are difficult to classify (e.g., grassland, shrubs, gardens).

(4) The research provides an effective way to make the best use of deep learning method in the large area of the land cover classification task: the high dominance and good discrimination land cover types were retained, and other minor types were discarded. It is not only an efficient way to simplify the remote sensing information model and improving the universality and reliability of the deep learning classification model, but also optimizing the efficiency of land cover classification application.

References

Adam H, Chen L C, Papandreou G et al., 2018. Encoder-decoder with atrous separable convolution for semantic image segment-

- ation. *Proceedings of the European Conference on Computer Vision*, 801–818. doi: [10.1007-978-3-030-01234-2_49](https://doi.org/10.1007-978-3-030-01234-2_49)
- Badrinarayanan V, Kendall A, Cipolla R, 2017. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2481–2495. doi: [10.1109-TPAMI.2016.2644615](https://doi.org/10.1109-TPAMI.2016.2644615)
- Bicheron P, Defourny P, Brockmann C et al., 2011. GLOBCOVER: products description and validation report. *Foro Mundial De La Salud*, 17(3): 285–287.
- Carranza-García M, García-Gutiérrez J, Riquelme J C, 2019. A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sensing*, 11(3): 274. doi: [10.3390-rs11030274](https://doi.org/10.3390-rs11030274)
- Cevikalp H, Benligiray B, Gerek O N, 2020. Semi-supervised robust deep neural networks for multi-label image classification. *Pattern Recognition*, 100: 107164. doi: [10.1016-j.patcog.2019.107164](https://doi.org/10.1016-j.patcog.2019.107164)
- Chen G S, Li C, Wei W et al., 2019. Fully convolutional neural network with augmented atrous spatial pyramid pool and fully connected fusion path for high resolution remote sensing image segmentation. *Applied Sciences*, 9(9): 1816. doi: [10.3390-app9091816](https://doi.org/10.3390-app9091816)
- Chen Jun, Liao Anping, Chen Jin et al., 2017. 30-Meter global land cover data product- globe land30. *Geomatics World*, 24(1): 1–8. (in Chinese)
- Chen L C, Papandreou G, Kokkinos I et al., 2017. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4): 834–848. doi: [10.1109-TPAMI.2017.2699184](https://doi.org/10.1109-TPAMI.2017.2699184)
- Congalton R G, 1988. Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 54(5): 587–592. doi: [10.1109-36.3037](https://doi.org/10.1109-36.3037)
- Congalton R G, 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1): 35–46. doi: [10.1016-0034-4257\(91\)90048-B](https://doi.org/10.1016-0034-4257(91)90048-B)
- De Fries R S, Hansen M, Townshend J R G et al., 1998. Global land cover classifications at 8 km spatial resolution: the use of training data derived from landsat imagery in decision tree classifiers. *International Journal of Remote Sensing*, 19(16): 3141–3168. doi: [10.1080-014311698214235](https://doi.org/10.1080-014311698214235)
- Gastaldo P, Zunino R, Heynderickx I et al., 2005. Objective quality assessment of displayed images by using neural networks. *Signal Processing:Image Communication*, 20(7): 643–661. doi: [10.1016-j.image.2005.03.013](https://doi.org/10.1016-j.image.2005.03.013)
- Gong P, Liu H, Zhang M N et al., 2019. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Science Bulletin*, 64: 370–373. doi: [10.1016-j.scib.2019.03.002](https://doi.org/10.1016-j.scib.2019.03.002)
- Guo Chongzhou, Li Ke, Li He, 2020. Deep convolution neural network method for remote sensing image quality classification. *Geomatics and Information Science of Wuhan University*, 1–9. (in Chinese)
- Guo R, Liu J B, Li N et al., 2018. Pixel-wise classification method for high resolution remote sensing imagery using deep neural networks. *ISPRS International Journal of Geo-Information*, 7(3): 110. doi: [10.3390-ijgi7030110](https://doi.org/10.3390-ijgi7030110)
- Guo Y M, Liu Y, Georgiou T et al., 2018. A review of semantic segmentation using deep neural networks. *International Journal of Multimedia Information Retrieval*, 7(2): 87–93. doi: [10.1007-s13735-017-0141-z](https://doi.org/10.1007-s13735-017-0141-z)
- He T D, Wang S X, 2021. Multi-spectral remote sensing land cover classification based on deep learning methods. *The Journal of Supercomputing*, 77(3): 2829–2843. doi: [10.1007-s11227-020-03377-w](https://doi.org/10.1007-s11227-020-03377-w)
- Hinton G E, Osindero S, Teh Y W, 2006. A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7): 1527–1554. doi: [10.1162-neco.2006.18.7.1527](https://doi.org/10.1162-neco.2006.18.7.1527)
- Hong D F, Gao L R, Yokoya N et al., 2020. More diverse means better: multimodal deep learning meets remote-sensing imagery classification. *IEEE Transactions on Geoscience and Remote Sensing*, 59(5): 4340–4354. doi: [10.1109-TGRS.2020.3016820](https://doi.org/10.1109-TGRS.2020.3016820)
- Kussul N, Lavreniuk M, Skakun S et al., 2017. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5): 778–782. doi: [10.1109-LGRS.2017.2681128](https://doi.org/10.1109-LGRS.2017.2681128)
- Li Deren, Zhang Liangpei, Xia Guisong, 2014. Automatic analysis and mining of remote sensing big data. *Acta Geodaetica et Cartographica Sinica*, 43(12): 1211–1216. (in Chinese)
- Loveland T R, Reed B C, Brown J F et al., 2000. Development of a global land cover characteristics database and IGBP DIS cover from 1 km AVHRR data. *International Journal of Remote Sensing*, 21(6–7): 1303–1330. doi: [10.1080-014311600210191](https://doi.org/10.1080-014311600210191)
- Ma H J, Liu Y L, Ren Y H et al., 2020. Improved CNN classification method for groups of buildings damaged by earthquake, based on high resolution remote sensing images. *Remote Sensing*, 12(2): 260. doi: [10.3390-rs12020260](https://doi.org/10.3390-rs12020260)
- Meng X R, Zhang S Q, Zang S Y, 2018. Remote sensing classification of wetland communities based on convolutional neural networks and high resolution images: a case study of the Honghe wetland. *Scientia Geographica Sinica*, 38: 1914–1923. doi: [10.13249-j.cnki.sgs.2018.11.019](https://doi.org/10.13249-j.cnki.sgs.2018.11.019)
- Noh H, Hong S, Han B, 2015. Learning deconvolution network for semantic segmentation. *Proceedings of the IEEE International Conference on Computer Vision*, 1520–1528. doi: [10.1109-ICCV.2015.178](https://doi.org/10.1109-ICCV.2015.178)
- Pan X R, Gao L R, Zhang B et al., 2018. High-resolution aerial imagery semantic labeling with dense pyramid network. *Sensors*, 18(11): 3774. doi: [10.3390-s18113774](https://doi.org/10.3390-s18113774)
- Pugh S A, Congalton, 2001. Applying spatial autocorrelation analysis to evaluate error in new England forest-cover-type maps derived from landsat thematic mapper data. *Photogrammetric Engineering and Remote Sensing*, 67(5): 613–620. doi: [10.1007-s001900100173](https://doi.org/10.1007-s001900100173)

- Quartulli M, Olaizola I G, 2013. A review of EO image information mining. *ISPRS journal of Photogrammetry and Remote Sensing*, 75: 11–28. doi: [10.1016-j.isprsjprs.2012.09.010](https://doi.org/10.1016/j.isprsjprs.2012.09.010)
- Ronneberger O, Fischer P, Brox T, 2015. U-net: convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham*, 234–241. doi: [10.1007-978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- Rezaee M, Mahdianpari M, Zhang Y et al., 2018. Deep convolutional neural network for complex wetland classification using optical remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(9): 3030–3039. doi: [10.1109-JSTARS.2018.2846178](https://doi.org/10.1109/JSTARS.2018.2846178)
- Shamsolmoali P, Zareapoor M, Wang R et al., 2019. A novel deep structure U-Net for sea-land segmentation in remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(9): 3219–3232. doi: [10.1109-JSTARS.2019.2925841](https://doi.org/10.1109-JSTARS.2019.2925841)
- Shelhamer E, Long J, Darrell T, 2016. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4): 640–651. doi: [10.1109-TPAMI.2016.2572683](https://doi.org/10.1109-TPAMI.2016.2572683)
- Wang Yahui, Chen Erxue, Guo Ying et al., 2020. Deep U-net optimization method for forest type classification with high resolution multispectral remote sensing images. *Forest Research*, 33(1): 11–18. (in Chinese)
- Wang Z, Bovik A C, Sheikh H R et al., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612. doi: [10.1109-TIP.2003.819861](https://doi.org/10.1109-TIP.2003.819861)
- Weng Q H, 2011. *Advances in Environmental Remote Sensing: Sensors, Algorithms, and Applications*. New York: CRC Press.
- Xu X D, Li W, Ran Q et al., 2017. Multisource remote sensing data classification based on convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2): 937–949. doi: [10.1109-TGRS.2017.2756851](https://doi.org/10.1109-TGRS.2017.2756851)
- Yuan Q Q, Shen H F, Li T W et al., 2020. Deep learning in environmental remote sensing: achievements and challenges. *Remote Sensing of Environment*, 241: 111716. doi: [10.1016-j.rse.2020.111716](https://doi.org/10.1016-j.rse.2020.111716)
- Yuan T, Zheng X Q, Hu X et al., 2014. A method for the evaluation of image quality according to the recognition effectiveness of objects in the optical remote sensing image using machine learning algorithm. *PloS One*, 9(1): e86528. doi: [10.1371-journal.pone.0086528](https://doi.org/10.1371-journal.pone.0086528)
- Zhang L P, Zhang L P, Du B., 2016. Deep learning for remote sensing data: a technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2): 22–40. doi: [10.1109-MGRS.2016.2540798](https://doi.org/10.1109-MGRS.2016.2540798)
- Zhang X, Liu L Y, Chen X D et al., 2021. GLC_FCS30: global land-cover product with fine classification system at 30 m using time-series landsat imagery. *Earth System Science Data*, 13(6): 2753–2776. doi: [10.5194-essd-13-2753-2021](https://doi.org/10.5194-essd-13-2753-2021)
- Zhang Z X, Liu Q J, Wang Y H, 2018. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5): 749–753. doi: [10.1109-LGRS.2018.2802944](https://doi.org/10.1109-LGRS.2018.2802944)
- Zhao H S, Shi J P, Qi X J et al., 2017. Pyramid scene parsing network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6230–6239. doi: [10.1109/CVPR.2017.660](https://doi.org/10.1109/CVPR.2017.660)