


Object-Oriented Method Combined with Deep Convolutional Neural Networks for Land-Use-Type Classification of Remote Sensing Images

Baoxuan Jin¹ · Peng Ye^{2,3}  · Xueying Zhang^{2,3} · Weiwei Song⁴ · Shihua Li¹

Received: 30 October 2018 / Accepted: 9 January 2019 / Published online: 16 January 2019
© The Author(s) 2019

Abstract

Land-use information provides a direct representation of the effect of human activities on the environment, and an accurate and efficient land-use classification of remote sensing images is an important element of land-use and land-cover change research. To solve the problems associated with traditional land-use classification methods (e.g., rapid increase in dimensionality of data, inadequate feature extraction, and low running efficiency), a method that combines object-oriented approach with deep convolutional neural network (COCNN) is presented. First, a multi-scale segmentation algorithm is used to segment images to generate image segmentation regions with high homogeneity. Second, a typical rule set of feature objects is constructed on the basis of the object-oriented segmentation results, and the segmentation objects are classified and extracted to form a training sample set. Third, a convolutional neural network (CNN) model structure is modified to improve classification performance, and the training algorithm is optimized to avoid the overfitting phenomenon that occurs during training using small datasets. Ten land-use types are classified by using the remote sensing images covering the area around Fuxian Lake as an example. By comparing the COCNN method with the method based solely on CNN, precision and kappa index were selected to evaluate the classification accuracy of the two methods. For the COCNN method, on the basis of the classification statistics, precision and kappa index coefficients are 96.2% and 0.96, respectively, which are 8.98% and 0.1 higher than those of the method based solely on CNN. Experimental results show that the COCNN method reasonably and efficiently combines object-oriented and deep learning approaches, thereby effectively solving the problem of the inaccurate classification of typical features with better classification accuracy than the simple use of CNN.

Keywords Land-use-type classification · Object-oriented · Convolutional neural networks · Deep learning · Multi-scale segmentation

Introduction

Land-use and land-cover change (LUCC), which is closely related to global climate change and changes in ecosystems and biodiversity, reflects the effects of human activities and climate change on the ecological environment of the Earth's surface (Blasi et al. 2008; Yang et al. 2014). Since the 1990s, the Food and Agriculture Organization, International Geosphere-Biosphere Project, International Institute for Applied Systems Analysis, and other research institutions have launched a series of LUCC-related projects (Sands and Leimbach 2003). The international community attaches importance to placing LUCC as the core content of global environmental change research. Remote sensing is an effective tool for monitoring the Earth's

✉ Peng Ye
yep730@163.com

¹ Information Center, Department of Natural Resources of Yunnan Province, Kunming 650224, Yunnan, China

² Key Laboratory of Virtual Geographical Environment (Ministry of Education), Nanjing Normal University, Nanjing 210023, Jiangsu, China

³ Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, Jiangsu, China

⁴ Department of Geoinformation Science, Kunming University of Science and Technology, Kunming 650504, Yunnan, China

surface and a basic element of applications that use classification and recognition technologies to investigate land-use status (Song et al. 2012).

Numerous types of land-use classification standards exist. These systems include several classes and account for the complex features of land-use and land-cover types, and these characteristics pose difficulties for accurate classification. In the classification of remote sensing images, determining the classification strategy first is necessary, followed by selecting the appropriate classifier. Classification strategies include supervised or unsupervised classification, the direct use of original spectral information or extraction of other features from spectral information, and hard or soft classification. In particular, classification strategies can be divided into pixel-based and object-oriented classifications due to differences in the basic unit of classification (Zheng et al. 2010). In terms of classifier selection, the traditional approach used is the statistical method for low-level feature extraction, including distance (Tzeng 2006), *K*-nearest neighbor (Meng et al. 2007), maximum likelihood (Bruzzone and Prieto 2001), and logistic regression (Lee 2005) classifiers. With the rapid development of aerospace, sensor, and computer technologies over the past decade, high-resolution remote sensing (HRRS) images have been increasingly applied in land-use classification (Hu et al. 2015). The diversity of objects within a given class increases as does the similarity of objects in different classes due to spectral confusion in HRRS images. These properties reduce the effectiveness of traditional classification methods based on low-level features (Paisitkiangkrai et al. 2016). Therefore, the method based on mid-level feature modeling has been developed on the basis of the low-level feature method (Bosch et al. 2007). Three types of mid-level feature extraction methods describe image semantics, namely the bag-of-visual-words (BoVW), latent Dirichlet allocation (LDA), and machine learning models. However, in practical applications, the performance of BoVW-based methods relies on the extraction of handcrafted local features (Alkhawani et al. 2015). LDA modeling methods rely on *K*-means clustering to produce a visual dictionary. Thus, the expression of mid-level semantic features in an image is limited. The machine learning models independently perform data expression and feature extraction (Campsvalls 2008) and discard the pattern of the extracted features in accordance with pre-determined rules (Tuia et al. 2013; Lin et al. 2017); thus, they obtain improved classification results when applied to complex images. The commonly used machine learning methods include sparse coding (Jiang et al. 2014), neural networks (Yuan et al. 2009), support vector machine (Blanzieri and Melgani 2006; Dai et al. 2007), and deep learning (Zhang et al. 2016). The deep learning networks are composed of multiple nonlinear mapping layers, which

represent a new method of intelligent pattern recognition and are an important new direction in the field of remote sensing image processing (Zhao et al. 2015).

Convolutional neural networks (CNNs) are a basic deep learning model representing biologically inspired multi-stage architectures composed of convolutional–pooling–fully connected layers (Längkvist et al. 2016). A CNN uses the low-level features contained in an image to form a high-level feature through the multilayer abstraction mechanism (Zhao et al. 2016), which effectively reduces the gap between low-level image and high-level semantic features. Research applying CNN to remote sensing images has emerged in recent years. The Hinton team won an overwhelming victory in the ImageNet image classification competition and reduced the top-5 classification error rate of 1000 images from 26.2 to 15.3% (Krizhevsky et al. 2017). Hu et al. used a CNN model to classify HRRS images for the first time. Chen et al. adopted a CNN classification method that incorporates pixel spectral information and spatial information and studied the importance of spatial information in classifying HRRS images (Chen et al. 2016). Qi et al. (2017) presented a Multiscale Deeply Described Correlation-based algorithm that jointly incorporates appearance and spatial information at multiple scales to perform land-use-type classification. Therefore, CNN has surpassed traditional pattern recognition and machine learning algorithms and has achieved superior performance and accuracy.

In general, the classification method based on CNN is executed by pixel. The classification results can easily be confused for the transitional zones between land types because land-use types are numerous and their spatial distribution is mixed with each other. This approach is not conducive to the type identification of small land blocks. To overcome these difficulties, traditional methods increase the training set size of deep learning or increase the model depth and the number of nodes, thereby causing tremendous pressure on manual labeling (Lin et al. 2016). Object-oriented classification strategies classify objects on the basis of homogeneous multi-pixels and use spectral, spatial, shape, and other features of images to perform type judgments together, thereby breaking through the limitations of pixel-based classification. In addition, the improvement of training sample set construction and deep learning method can reduce the dependence of deep learning model on training sample size. Therefore, this study improves from two aspects: classification strategy and deep learning model. The major contributions of this research are as follows:

1. The object-oriented method is combined with the deep learning method. On the one hand, the object-oriented method is used to construct a multi-scale sample set to

provide high-precision training data for deep learning model training. On the other hand, on the basis of the object-oriented concept, the method avoids the processing of mixed pixels in the classification process and enhances the typicality of classification objects in deep learning.

2. The CNN model structure is modified to improve classification performance, and the training algorithm is optimized to avoid the overfitting phenomenon that occurs during training using small datasets.

The remainder of this paper is organized as follows: The “**Methodology**” section introduces the proposed framework that combines object-oriented approach with deep convolutional neural network (COCNN) for use in land-use-type classification. The “**Experiment and Results**” section presents the experimental results and analysis. The “**Conclusions**” section offers concluding remarks and perspectives on future work.

Methodology

The general process of remote sensing image classification mainly consists of feature extraction and classification based on image features. The traditional object-oriented method establishes fuzzy rules in accordance with the feature differences of various class objects, focusing on the improvement of feature extraction. The object features include color, spectral characteristics [e.g., luminance value, normalized difference water index (NDWI), and normalized difference vegetation index (NDVI)], and shape–texture (e.g., boundary index, compactness, and aspect ratio) (Chen et al. 2006; Su et al. 2007; Robertson and King 2011). However, the feature extraction of the object-oriented method cannot cover all feature types. Therefore, supporting the classification and recognition of class objects that only rely on the extracted feature information is insufficient when the performance of the classifier is not improved. Deep learning combines low-level features to form a further abstract high-level representation, which has strong expressive capability and outstanding classification performance. The characteristics of a multi-band of remote sensing images are not fully considered because deep learning is often performed by RGB images. In addition, deep learning requires a large number of labels; thus, the manual identification workload is large. Table 1 compares and analyzes the advantages and disadvantages of the two methods.

The advantage of combining object-oriented approach with deep learning method includes two aspects. On the one hand, through the object-oriented method for constructing the feature rule set, the land-use object can be

initially extracted, and the training sample sets required for deep learning can be further constructed by the object. On the other hand, the performance of deep learning is affected by the number of features in practical application, especially when the size of the sample set is relatively small (Mares et al. 2016). After the combination of object-oriented method, large-scale spatial context information can be considered by extracting object units, and additional feature rules and prior knowledge can be integrated in the deep learning process (Zeiler and Fergus 2014). In addition, the classification result can be corrected in accordance with the feature rule set of the object-oriented method. The optimization of feature extraction strategy is conducive to the further improvement of the classification effect. A land-use-type classification method (COCNN) based on the technical characteristics of object-oriented and deep learning approaches is proposed on the basis of the analysis of the advantages of the two methods. This method is explained in detail in the following section.

COCNN Land-Use-Type Classification Framework

The general flowchart of the COCNN framework (Fig. 1) illustrates three features. First, after the preprocessing of remote sensing images, such as image fusion, the multi-scale segmentation algorithm is used to segment the image. Second, on the basis of the object-oriented segmentation results, the typical rule set of construction land, roads, water bodies, vegetation, and other land-use types is constructed, and the segmentation objects are classified and extracted to obtain training samples to form a typical object sample set. Finally, the CNN model training is performed in accordance with the sample set, and the multi-scale segmentation results are further classified on the basis of the training model.

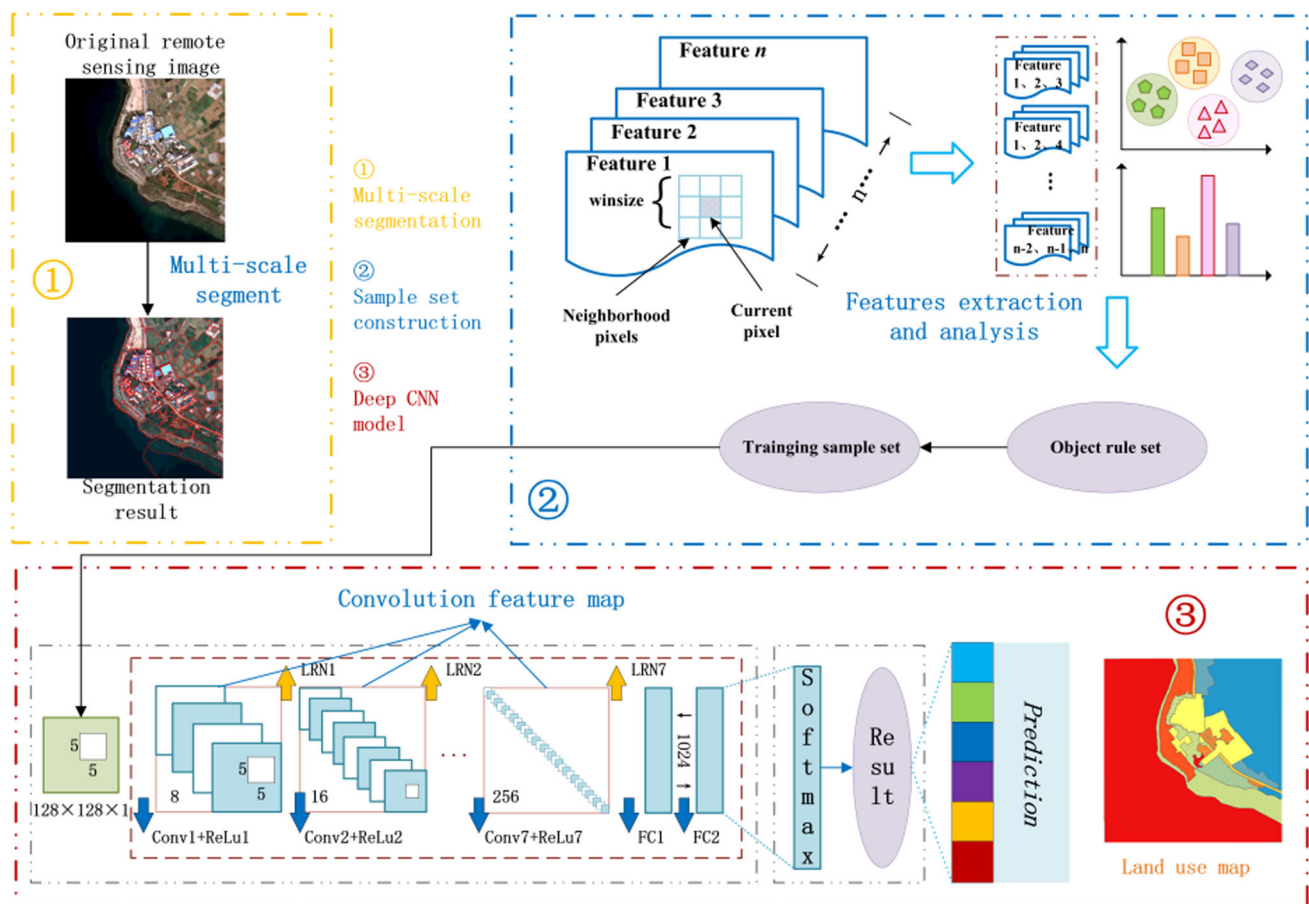
Multi-scale Image Segmentation Based on Object-Oriented Method

Image Preprocessing

Preprocessing of remote sensing images includes radiometric calibration, geometric correction, and image fusion. The main purpose of preprocessing is to express the information contained in the process of imaging synthesis further to render it closest to the actual image state. Radiation correction is used to eliminate image distortion caused by radiation errors. Geometric correction requires that the absolute error of the corrected position is less than one pixel. The image fusion algorithm selects NNDiffuse Pan Sharpening, which can effectively preserve the color, texture, and spectral information of the image.

Table 1 Advantages and disadvantages of object-oriented and deep learning methods

Classification method	Advantages	Disadvantages
Object-oriented approach	The limitation of cell classification method is broken by using the “homogeneous” multiple pixels as the basis for classification. Various object feature rule sets can be established to extract land-use types by dividing different land uses into various objects	The description of shape and texture is incomprehensive and inaccurate, and the amount of information is insufficient to support accurate land-use classification and identification
Deep learning method	The high-level characteristics of different objects can be mastered to distinguish different land use accurately and divide the land use in accordance with the training results	The image classification model requires a large amount of data to train. Especially for the image segmentation model, a large number of tags are required to identify different features, and the manual identification workload is considerably large. The result of deep learning is often a raster image; thus, the final result is difficult to correct

**Fig. 1** Flowchart of the COCNN method

Multi-scale Image Segmentation

The multi-scale object-oriented segmentation algorithm considers an image to be a region adjacency graph consisting of topological relationships between regions (Wang and He 2011). The algorithm can segment the image in accordance with the specified scale for ensuring that the

image segmentation region (image object) with high homogeneity (or minimal heterogeneity) is generated, which is suitable for the optimal separation and representation of the object (Woodcock and Strahler 1987).

The algorithm is roughly divided into two steps during execution: (1) initial segmentation and (2) object merging. In the initial segmentation step, starting from a single pixel,

the difference measure is calculated with the neighboring cells, and the image segmentation is conducted in accordance with the heterogeneity (Jin et al. 2018). This heterogeneity is determined by the difference in spectrum and geometry between objects, and the calculation for heterogeneity follows formula (1).

$$f = w_1x + (1 - w_1)y. \quad (1)$$

In the formula, f is the heterogeneity value; w_1 represents the weight, $0 \leq w_1 \leq 1$; x denotes the spectral heterogeneity; and y refers to the shape heterogeneity. The calculation of x and y follows formulas (2) and (3).

$$x = \sum_{i=1}^n p_i \sigma_i, \quad (2)$$

$$y = w_2u + (1 - w_2)v. \quad (3)$$

In the formula, p_i is the weight of the i th image layer; σ_i indicates the standard deviation of the i th image layer spectral value; u represents the overall tightness of the image region; v denotes the image region boundary smoothness; and w_2 stands for the weight, $0 \leq w_2 \leq 1$. The calculation of u and v follows formulas (4) and (5).

$$x = \frac{E}{\sqrt{N}}, \quad (4)$$

$$y = \frac{E}{L}. \quad (5)$$

In the formula, E is the actual boundary length of the image region; N denotes the total number of pixels of the image region; and L represents the total length of the rectangular boundary, including the range of the image region.

The object merging step starts with each region in the region adjacency graph. The region pairs that satisfy the local optimal merge condition are determined, the two regions are merged, and the feature values of all regions connected to the original two regions are updated. When the adjacent two regions are merged, the heterogeneity of the newly generated large image region object is calculated using formula (6).

$$f' = w_1x' + (1 - w_1)y'. \quad (6)$$

In the formula, f' is the heterogeneity value of the newly merged large image region object; x' and y' represent the spectral and shape heterogeneities of the newly merged large image region, respectively. The calculation of x and y follows formulas (7) and (8).

$$x' = \sum_{i=1}^n p_i [N' \sigma'_i - (N_1 \sigma_{i1} + N_2 \sigma_{i2})], \quad (7)$$

$$y' = w_2u' + (1 - w_2)v'. \quad (8)$$

In the formula, N' denotes the total number of pixels in the merged image region; σ'_i refers to the standard deviation of the i th layer spectral value of the merged image; N_1 and N_2 are the total numbers of image pixels in adjacent regions 1 and 2 before the merge, respectively; σ_{i1} and σ_{i2} refer to the standard deviations of the spectral values of the i th layer of adjacent regions 1 and 2 before the merge, respectively. The calculation of u' and v' follows formulas (4) and (5).

$$u' = N' \frac{E'}{\sqrt{N'}} - \left(N_1 \frac{E_1}{\sqrt{N_1}} + N_2 \frac{E_2}{\sqrt{N_2}} \right), \quad (9)$$

$$v' = N' \frac{E'}{L'} - \left(N_1 \frac{E_1}{L_1} + N_2 \frac{E_2}{L_2} \right). \quad (10)$$

In the formula, E' and L' are the actual boundary length of the merged image region and the total length of the circumscribed rectangle boundary of the region range, respectively; E_1 and L_1 denote the actual boundary length of adjacent region 1 before the merge and the total boundary length of region 1s circumscribed rectangle, respectively; E_2 and L_2 indicate the actual boundary length of adjacent area 2 before the merge and the total boundary length of region 2s circumscribed rectangle, respectively. Figure 2 shows the results of image segmentation at different scales.

Sample Set Construction Based on Multi-scale Rules

The hierarchical structure of the sample classification is established through the correspondence between the feature information of the object and the land use. Multi-scale hierarchical segmentation is used, and different land uses are segmented by different scales. Then, classification rules are set in accordance with the spectral, geometric texture, and topological features of the land-use object.

In the large-scale segmentation layer, the index of brightness, NDWI, and NDVI are used as the basis for the assessment (Zhu et al. 2017), and the first classes, such as construction land, road, water body, and vegetation, are initially extracted. On the image objects of the first classes, the appropriate segmentation scale is selected, and the subclasses in the first classes are segmented by considering the shape index of the objects, such as the boundary index, compactness, and aspect ratio. Table 2 shows the multi-scale object rule set.

A set of typical remote sensing image features, including cultivated land, woodland, water, roads, and buildings, is established on the basis of the object judgment rules, by tracking the sample boundaries under each category, and the training sample set is obtained. Table 3 shows the training sample set example.

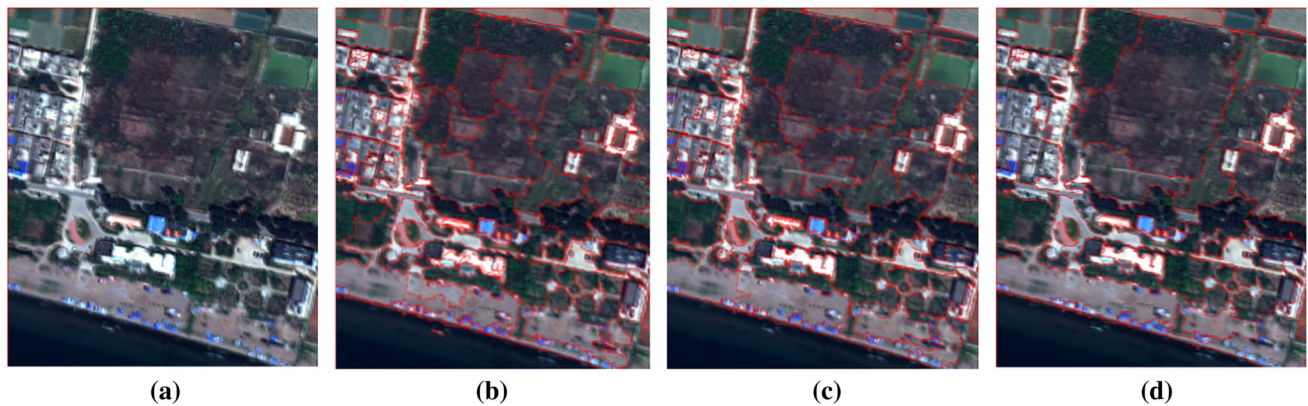


Fig. 2 Multi-scale segmentation result diagram. **a** Original image, **b** the segmentation scale is 75, the shape weight is 0.3, and the compactness weight is 0.5, **c** the segmentation scale is 120, the shape

weight is 0.3, and the compactness weight is 0.5, **d** the segmentation scale is 300, the shape weight is 0.3, and the compactness weight is 0.5

Construction of Deep CNN Model

Modeling Method

The deep CNN model is selected for deep learning by using the sample images in the sample set as the training data. The characteristics of the samples are automatically obtained through deep learning, and the object-oriented segmentation results are used to realize the automatic classification of typical land-use types. Therefore, the structural design of the CNN is the key issue.

A deep CNN is formed by stacking multiple basic network structures. To obtain further accurate classification results, adding additional nodes to the model is necessary. However, the model complexity requires additional training samples for support, but the training samples that can be used in practical applications are limited. The comparison of the traditional methods in CNN training shows that improving the learning algorithm in training is further effective.

The important structural parameters and training strategies are optimized to improve the classification effect of deep CNN:

1. Rectified linear unit (ReLU) activation function accelerates model convergence.

The ReLU function is one of the most popular neuronal activation functions in the deep learning field (Shang et al. 2016). In comparison with other activation functions, the commonly used sigmoid function is a nonlinear activation function that displays a saturation effect, thereby causing a loss of gradient information for large and small input data values (Chen et al. 2013). The output gradient of the sigmoid function is not centered on zero, resulting in convergence fluctuations during the gradient descent phase. When the number of layers is relatively large, the gradient

to the front layer becomes small, and the network weight is ineffectively updated. The tanh activation function also has a small gradient value at saturation, leading to inefficient training (Nambiar et al. 2014; Gulcehre et al. 2016).

The gradient constant of the ReLU function is equal to 1 when $x > 0$. Thus, the problem of gradient disappearance is alleviated during backpropagation (Zhang et al. 2017). Moreover, the ReLU function is sparsely activated by simple thresholding activation. In comparison with other activation functions (e.g., sigmoid and tanh functions), the ReLU function increases the convergence speed of CNN.

2. Use of regularization to prevent overfitting.

Regularization reduces the model complexity by restricting the parameter's ranges to reduce the disturbances caused by noisy inputs. This procedure reduces overfitting to a certain extent (Fanany 2017). The L2 regularization is realized by modifying the cost function, whereas the dropout technique is realized by modifying the neural network. The key concept of the dropout technique involves randomly suppressing neurons in the target layer with a certain probability during every iteration of model training (Zheng et al. 2017). This process considerably reduces the complex mutual adaptation among neurons and achieves the suppression of overfitting.

3. Local response normalization (LRN) layer enhancement generalization.

The LRN layer mimics the side inhibition mechanism of biological nervous systems and creates a competitive environment for the activity of local neurons (Li et al. 2015). This behavior enhances the relatively large response values and suppresses other neurons with small feedback, thereby elevating the model's generalization capability. Furthermore, as LRN selects large feedback from the responses of multiple convolution kernels of the nearest

Table 2 Multi-scale object rule set

No.	Scale	Category	Judgment basis	Rule	Remark
1	Large-scale segmentation layer (segmentation scale of 300, shape weight of 0.3, and compactness weight of 0.5)	Construction land	Lightness value	Lightness value > 338	The building spectrum is bright, and the variance is large
2		Roads	Width and aspect ratio	Width < 17.8, aspect ratio > 12.5	It is strip-shaped and has a large ratio of length to width
3		Water bodies	NDWI	NDWI > 0.5	NDWI is used to extract the water body information in images, and the effect is improved
4		Vegetation	NDVI	NDVI > 0	NDVI is the best indicator of vegetation growth and coverage
5	Small-scale segmentation layer (segmentation scale of 75, shape weight of 0.3, and compactness weight of 0.5)	Residential buildings	Extracted from construction land, in accordance with shape rules	Boundary index > 1.0, homogeneity index < 12	Consider the relationship between the internal parent and child objects, the boundary index of the object, the lightness value, and the homogeneity
6		Industrial land	Extracted from construction land, in accordance with shape rules and spectral value index	Boundary index > 0.8, lightness value < 398	
7		Other construction land	Extracted from construction land, except residential building and industrial land	–	–
8		Cultivated land	Extracted from vegetation, in accordance with shape rules	Boundary index > 1.5, homogeneity index < 12	Consider the relationship between the internal parent and child objects, the boundary index of the object, the ecological index, and the homogeneity
9		Garden plots	Extracted from vegetation, based on spectral value index and NDVI	0.30 < NDVI < 0.45, homogeneity index > 24	
10		Grassland		0.15 < NDVI < 0.30, homogeneity index > 24	
11		Forest land		NDVI > 0.45, homogeneity index > 32	
12		Bare land	Extracted from vegetation, except cultivated land, garden plot, grassland, and forest land	–	–

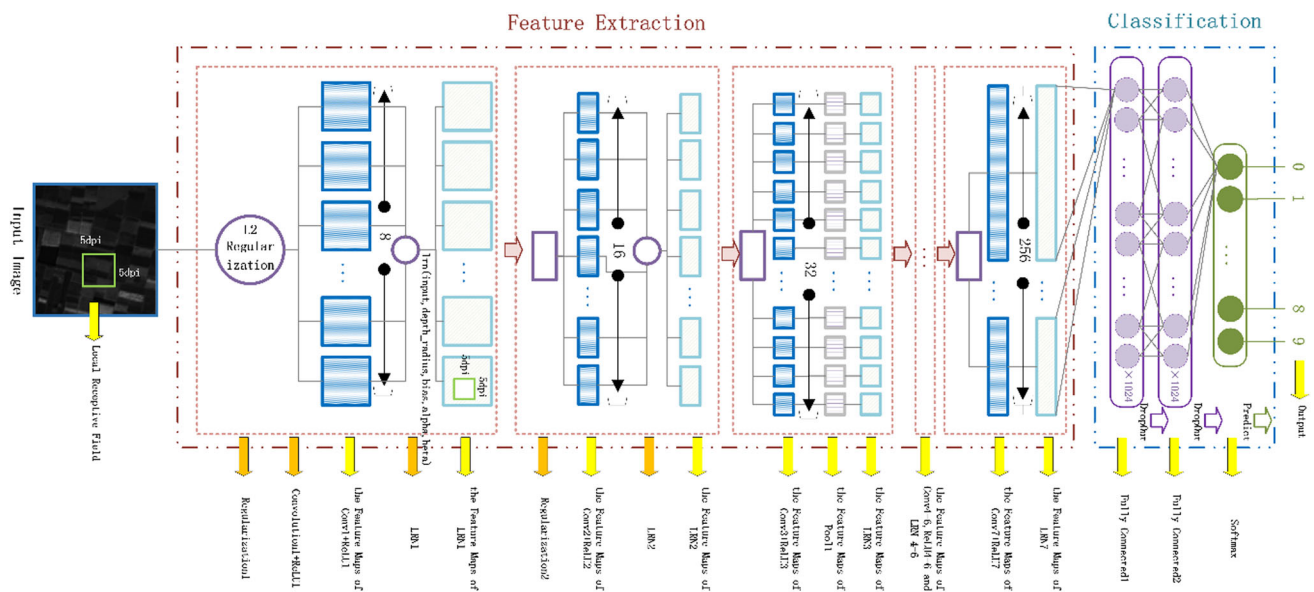
neighbor, it applies to the ReLU activation function without an upper bound.

CNN Model Structure

Figure 3 presents the deep CNN framework constructed in this paper. The initial weights of the network are extracted from the Gaussian distribution with a standard deviation of 0.01 and a mean value of 0. At the training stage, the

Table 3 Training sample set example

Category	Sample example			Category	Sample example		
Residential buildings				Grassland			
Industrial land				Forest land			
Other construction land				Bare land			
Cultivated land				Roads			
Garden plots				Water bodies			

**Fig. 3** Structure of the deep CNN

sliding step length is 1, and a gradient descent is performed with a constant learning rate of 0.0005. The core component of the deep CNN is composed of 7 convolutional layers (Conv1–Conv7), 1 pool layer (pooling1), and 7 LRN layers (norm1–norm7).

Four of the important parameters and functions in the deep CNN-based model are described as follows: (1) size and number of the local receptive fields and activation functions. The convolution kernels are 3×3 , 5×5 , or 7×7 pixel blocks. Convolutional layers 1–7 have 8, 16, 32, 64, 128, 256, and 256 kernels. Different sizes and

numbers of convolution kernels are used to investigate the effects of the characteristic sampling density on the model performance. After the convolution operation, the ReLU activation function is used. (2) Initial weight regularization. The L2 regularization is added to the initialization parameters of each layer in the network. (3) Fully connected layer with dropout. The model consists of two fully connected layers, each of which has 1024 outputs. The dropout technique is applied to the FC1–2 layer to control overfitting given that the fully connected layer FC1–2 is densely connected. The optimal dropout probability is

optimized within the range of 0.5–0.9. (4) Loss function of the classification layer. The softmax loss function constructs the corresponding classifier in the classification layer. Each node in the output of the CNN represents the probability that the input information belongs to a certain class I as follows:

$$P(Y = i|x, w, b) = \text{softmax}(wx + b) = \frac{e^{w_i x + b_i}}{\sum_j e^{w_j x + b_j}}, \quad (11)$$

where w is the weight parameter in the last layer and b denotes the corresponding bias parameter.

Experiment and Results

Experimental Data and Environment

Experimental data containing the optical remote sensing image, a high-quality land-use classification vector layer, and the classified field information are derived from the land-cover classification results of the National Geoinformation Survey. In particular, the remote sensing image has a scale of 1:10,000, measures 8386×5772 pixels, and has a pixel resolution of 1 m. The image (Fig. 4) shows the area surrounding Fuxian Lake in Yunnan Province, China. On the basis of the classification system of the National Geoinformation Survey, the land-use types shown in the image are divided into ten classes: residential building, industrial land, other construction lands, cultivated land, garden plots, grassland, forest land, bare land, roads, and water bodies. The sample set is constructed on the basis of

multi-scale rules, and a part of it is selected as test data. Table 4 shows the data volume of the sample and test sets for different land-use types.

The indexes for evaluating the experimental results include precision (P) and kappa index (K). P and K are calculated using formulas (11) and (12) (Wang et al. 2012), in which n_{st} refers to the same quantity between annotation result s and classification result t , n_t denotes the quantity of results of the classification result t , r represents the number of rows in the table, x_{ii} is the quantity of type combinations on the diagonal part of the table, x_{i+} indicates the number of observations in line I , x_{+i} refers to the number of observations in column I , and N stands for the number of cells in the table.

$$P(s, t) = \frac{n_{st}}{n_t}, \quad (12)$$

$$K = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \times x_{+i})}. \quad (13)$$

The experiment uses Windows Server 2012 R2 operating system, with a NVIDIA Tesla K80 for GPU acceleration. Other major hardware elements include an Intel (R) Xeon (R) CPU E5-2630 processor and 128 GB of memory. The deep CNN model is developed on the basis of the TensorFlow open-source framework. The main versions of the software are CUDA 8.0, cuDNN 6.0, and tensorflow_gpu_1.2.0.

Fig. 4 Remote sensing image of the study area

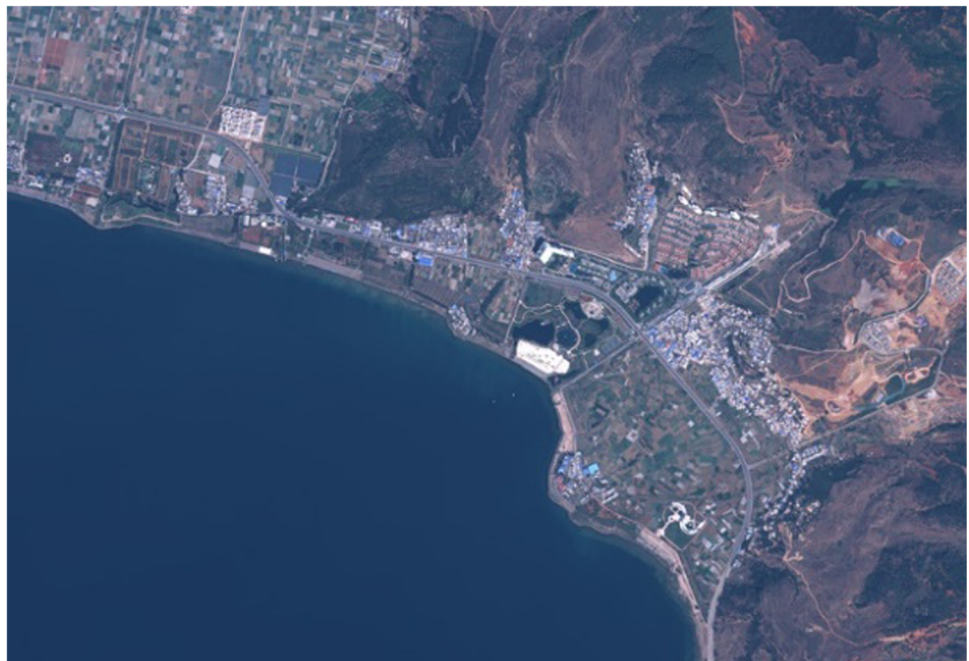


Table 4 Data volume of the datasets

Scene	Sample set	Test set	Scene	Sample set	Test set
Cultivated land	863	216	Garden plots	806	202
Forest land	807	202	Grassland	738	185
Residential buildings	826	207	Roads	835	209
Other construction lands	854	214	Industrial land	773	194
Bare land	772	193	Water bodies	782	196

Setting of the Experiment

In this experiment, the COCNN method is compared with the method based solely on CNN. For the CNN method, the land-use classification of images is based solely on the deep CNN model. The window size of 30×30 is selected to extract spatial information in the images for land-use-type classification. Then, the entire image is scanned by moving the window.

The experimental results are compared from two perspectives. On the one hand, under the condition that the structure and parameters of the deep CNN model remain unchanged, the difference of classification accuracy between the COCNN method and the method based solely on CNN is compared. On the other hand, on the basis of the joint object-oriented method, the structure and parameters of the deep CNN model are adjusted, and the classification effects under different structural and parameter conditions are compared.

The baseline parameter configuration for COCNN (Table 5) assumes different frame selections and parameter settings that affect the classification accuracy. Comparative experiments are conducted to change some parameters while keeping the remaining settings unchanged. Model training uses 100 elements of the training data for each iteration, and 1500 training iterations are performed. The network state is tested 100 times per iteration.

Results and Analysis

Influence of Classification Strategies on Classification Results

The classification results of the images are obtained, and the accuracy is evaluated on the basis of the COCNN method. The land-use map of the area surrounding Fuxian Lake (Fig. 5) contains ten land-use classes with class-specific confusion matrixes (Table 6). The P and K coefficients of the land-use classes are 96.2% and 0.96, respectively. The water body type has the highest producer's accuracy (99.5%), whereas the industrial land type has a relatively low value of 91.2%.

When based solely on the CNN method, the P and K of the classification results are 87.22% and 0.86, respectively. Thus, the classification accuracy (Table 7) of the CNN method is lower than that of the COCNN method. In addition to the slight decrease in the accuracy of water bodies, the classification accuracy of other land-use types has been remarkably reduced. Combined with the observation of land-use map, the classification results based on COCNN classification method are relatively complete, and few faults are found in large-scale plaques, such as vegetation and construction land. In addition, COCNN is useful for solving the problems of the incomplete extraction of linear features and the small plots of crops. The accuracy of land-use information extracted based on CNN methods is slightly insufficient, and several plaques with inaccurate classification types are found. Moreover, the classification results are relatively fragmented; thus, the classified plots have evident spatial heterogeneity. The errors are a consequence of errors among land-use types with the same

Table 5 Basic parameter settings for COCNN

Configuration item	Parameter	Configuration item	Parameter
Size of the convolution kernels	3×3	Number of convolution kernels (Conv1–Conv7)	8–16–32–64–128–256–256
Activation function	ReLU	Pool method	1-max pooling
Learning rate	0.00005	Dropout	0.5
LRN-n/2	4	LRN-k	1.0
LRN- α	0.001/9.0	LRN- β	0.75

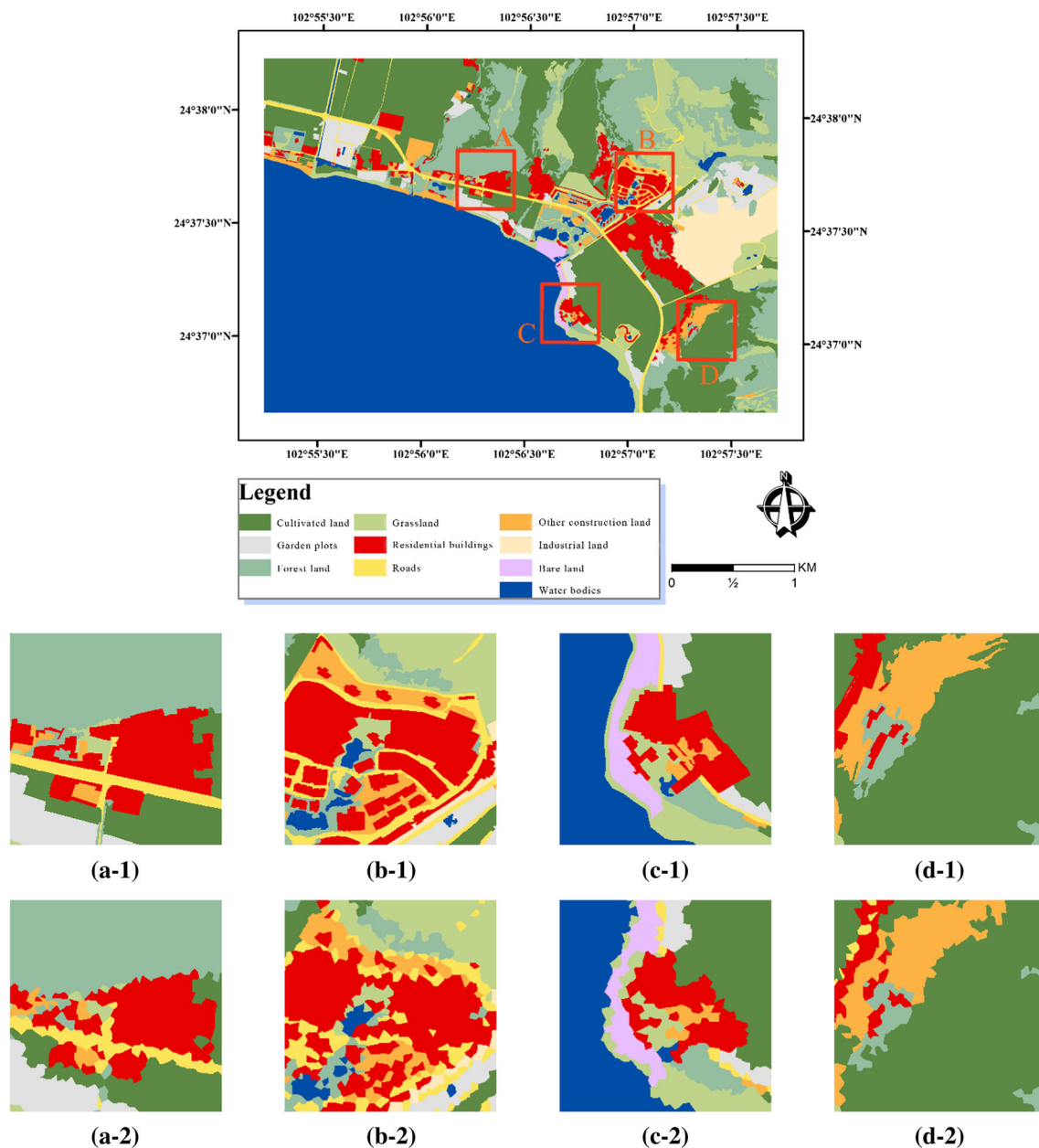


Fig. 5 Detailed land-use map of the area surrounding Fuxian Lake. **a-1, b-1, c-1, and d-1** Classification of regions A, B, C, and D based on COCNN, respectively. **a-2, b-2, c-2, and d-2** Classification of regions A, B, C, and D based solely on CNN, respectively

natural attributes. For example, when the residential building type displays a classification error, the inaccurately selected classification type is often another construction land. Research results show that the COCNN method not only fully uses the spectral information of remote sensing images but also considers the spatial distribution characteristics and correlation of geographic objects. On the one hand, noise generated due to heterogeneity and spectral differences in pixel classification is effectively avoided. On the other hand, a multi-feature

sample set is constructed by correlation rules, which can assist the deep learning effect of CNN.

Influence of Deep CNN Structure on Classification Results

1. Influence of network parameters on different convolution kernel parameters.

The convolution kernel is regarded as the most sensitive element of the CNN and is responsible for directly extracting the lowest-level features from the original input image. The effects of the size and number of convolution

Table 6 Confusion matrix of the land-use-type classification based on COCNN

	Classification results										Total	Precision (%)
	CL	GP	FL	GL	RB	R	OC	IL	BL	W		
<i>Standard results</i>												
CL	213	0	0	2	0	0	1	0	0	0	216	98.61
GP	0	186	7	4	0	0	0	0	0	5	202	92.08
FL	0	1	199	2	0	0	0	0	0	0	202	98.51
GL	2	0	0	181	0	0	0	1	0	1	185	97.84
RB	0	0	0	0	203	2	2	0	0	0	207	98.07
R	0	0	0	0	4	200	5	0	0	0	209	95.69
OC	12	0	0	2	0	0	197	2	1	0	214	92.06
IL	0	0	4	9	0	0	2	177	2	0	194	91.24
BL	0	0	0	0	0	0	2	1	190	0	193	98.45
W	0	0	0	1	0	0	0	0	0	195	196	99.49
Total	227	187	210	201	207	202	209	181	193	201	2018	96.18

CL cultivated land, GP garden plots, FL forest land, GL grassland, RB residential building, R roads, OC other construction lands, IL industrial land, BL bare land, W water bodies

Table 7 Confusion matrix of the land-use-type classification based solely on CNN

	Classification results										Total	Precision (%)
	CL	GP	FL	GL	RB	R	OC	IL	BL	W		
<i>Standard results</i>												
CL	194	7	0	3	0	0	10	2	0	0	216	89.81
GP	4	167	16	12	0	0	3	0	0	5	202	82.67
FL	6	3	179	14	0	0	0	0	0	0	202	88.61
GL	5	5	7	165	0	2	0	1	0	0	185	89.19
RB	0	0	0	0	184	8	11	0	4	0	207	88.89
R	0	0	0	9	4	181	12	1	2	0	209	86.60
OC	16	0	0	2	12	6	171	4	3	0	214	79.90
IL	0	0	8	14	6	0	7	156	3	0	194	80.41
BL	1	0	0	0	0	4	6	3	177	2	193	91.71
W	4	0	0	2	0	0	0	0	4	186	196	94.90
Total	230	182	210	221	206	201	220	167	193	193	2018	87.22

CL cultivated land, GP garden plots, FL forest land, GL grassland, RB residential building, R roads, OC other construction lands, IL industrial land, BL bare land, W water bodies

kernels on the recognition accuracy of COCNN (Fig. 6) show that the model performance increases as the convolution kernel size decreases. When the convolution kernel size is 3×3 , the verification accuracy reaches its highest value. When the convolution kernel size is large, the mixing of information from coarse-grained features (e.g., edge features) occurs, and excessive detail is lost from the information that passed to the convolution kernel of the high layers because the distinction between similar land-use types often depends on the description of local textures.

When the fixed convolution kernel size is 3×3 , the experiment verifies that models with smaller numbers of convolution kernels and greater numbers of layers display higher classification effectiveness than models with greater

numbers of convolution kernels and smaller numbers of layers. The seven-layer model with 8, 16, 32, 64, 128, 256, and 256 convolution kernels is more accurate than the four-layer model with 64, 128, 256, and 512 convolution kernels. The CNN network requires a sufficient number of low-level features to ensure the capability to fit the data to overcome the data complexity caused by factors, such as the variety of land types, because the dataset covers a relatively small number of species and samples. Therefore, increasing the depth of the CNN improves the network performance.

In the CNN, the feature map of this layer is a different combination of the feature maps extracted by the previous layer. Thus, the output data of the previous layer are the

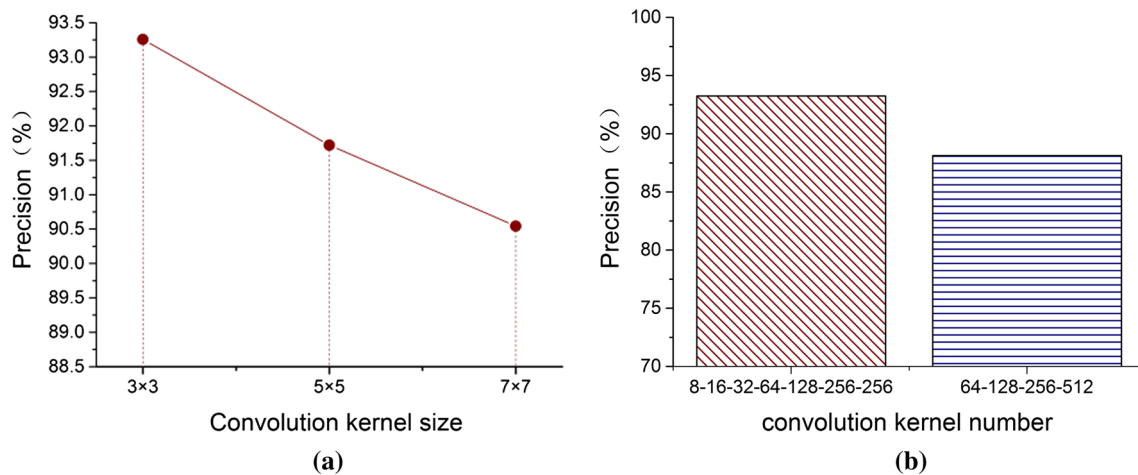


Fig. 6 Effects of different sizes (a) and numbers (b) of convolution kernels on model performance

input data of this layer. To further verify that no redundancy exists in the convolution result for each layer, the entire convolution kernel is visualized (Fig. 7). No repetitive or random convolution kernel is found after comparing the visualization results. Thus, the convolution kernel is effectively trained on all cases.

2. Influence of network parameters on the use of regularization and dropout to suppress overfitting.

In COCNN, regularization and dropout suppress overfitting in model training, and the effectiveness of the two methods was tested separately. The model without L2 regularization displays the effects of overfitting when trained 900 times. The L2 regularization term has no effect on the updating of bias b in each layer of the model but affects the updating of weight w (Fig. 8). When w is positive, the updated w decreases, and when w is negative, the updated w becomes large. The effect of L2 regularization is to bring w closer to 0. Thus, the weights in the network

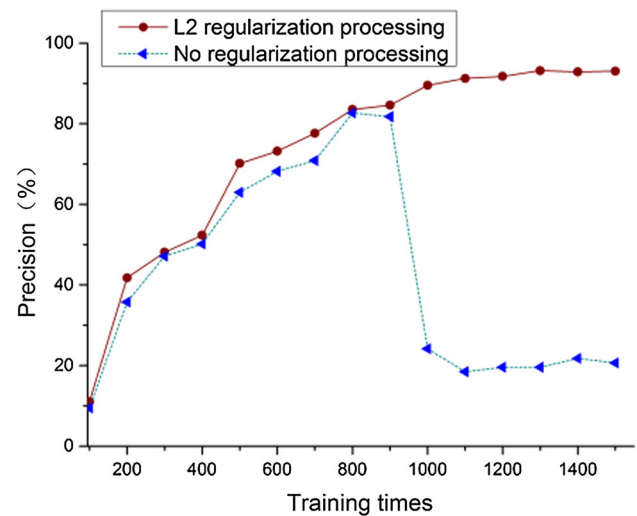


Fig. 8 Effects of L2 regularization on model performance

approach 0. This behavior is equivalent to reducing the weight of the network and changing the complexity of the network, thereby avoiding overfitting.

The effects of the dropout probability on model performance show that the accuracy of model classification reaches its peak when the dropout probability is 0.50 (Fig. 9). When the dropout value is large and insufficient training data are used, excessive amounts of feature information extracted from the model are retained, resulting in overfitting. As the dropout value decreases, the model performance also decreases. This outcome is a consequence of excessive deleted neurons, resulting in insufficient trained subnetwork and leading to reductions in the capability of the model to fit the data. Then, the model experiences difficulties in effectively establishing mapping relationships between the image data and land-use types. By combining the effects of regularization and dropout, overfitting avoidance is limited by relying on regularization

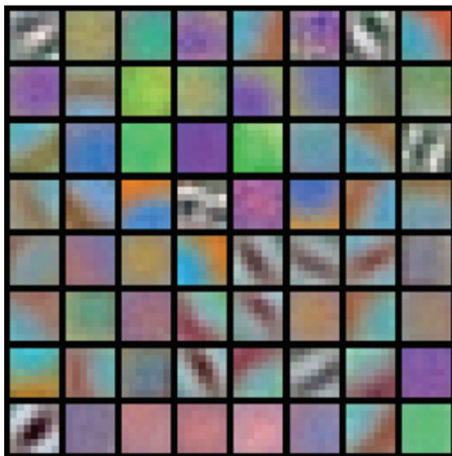


Fig. 7 Visualization result of convolution kernel

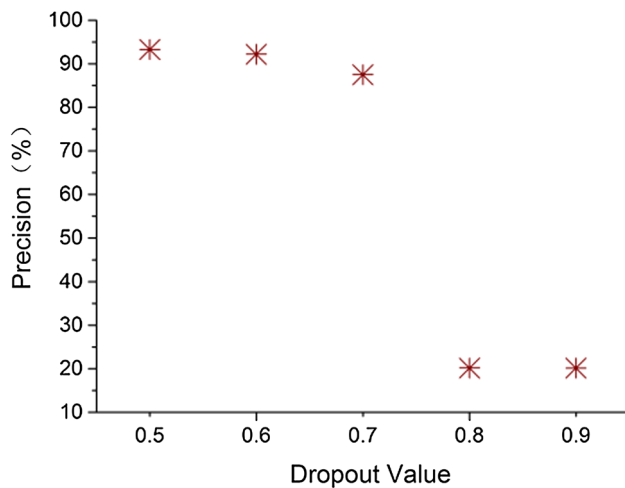


Fig. 9 Effects of different dropout probabilities on model performance

or dropout alone; improved effectiveness is achieved when regularization and dropout are combined.

Conclusions

To realize the accurate recognition of land-use types based on remote sensing images, a method that uses COCNN is proposed. The COCNN method constructs a set of typical feature samples obtained by the general rule set on the basis of the multi-scale segmentation of images. Then, sample training and feature extraction are further performed by the deep learning method. Finally, the sample characteristics after learning are applied to the segmentation results to complete the land-use classification of remote sensing images. For the classification statistics, the P and K coefficients are 96.2% and 0.96, respectively. For the influence of deep CNN structure, increasing the depth of the CNN improves the network performance. In addition, overfitting avoidance is limited by relying on regularization or dropout alone; improved effectiveness is achieved when regularization and dropout are combined. Experimental results show that the COCNN method reasonably and efficiently combines object-oriented and deep learning approaches and can comprehensively utilize the spectral, geometric, and texture information of image objects. This method not only can fully use the correlation between neighboring pixels, to obtain the small spatial heterogeneity of the classification results, but also has strong anti-noise capability, which effectively reduces the phenomenon of pixel spectral confusion. However, the existing research remains imperfect in the general rule set construction and deep learning structure design and requires further improvement.

Acknowledgements This research is supported by the National Natural Science Foundation of China (41661086) and National Key Research and Development Program of China (2017YFB0503602).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Alkhawani, M. M., Elmogy, M., & Elbakry, H. M. (2015). Content-based image retrieval using local features descriptors and bag-of-visual words. *International Journal of Advanced Computer Science & Applications*, 6(9), 212–219.
- Blanzieri, E., & Melgani, F. (2006). An adaptive SVM nearest neighbor classifier for remotely sensed imagery. In *International geoscience and remote sensing symposium* (pp. 3931–3934).
- Blasi, C., Zavattero, L., Marignani, M., Smiraglia, D., Copiz, R., Rosati, L., et al. (2008). The concept of land ecological network and its design using a land unit approach. *Plant Biosystems*, 142(3), 540–549.
- Bosch, A., Muñoz, X., & Martí, R. (2007). Which is the best way to organize/classify images by content? *Image and Vision Computing*, 25(6), 778–791.
- Bruzzone, L., & Prieto, D. F. (2001). Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2), 456–460.
- Campsalls, G. (2008). New machine-learning paradigm provides advantages for remote sensing. *Spie Newsroom*, 46(1), 1–3.
- Chen, Y. H., Feng, T., Shi, P. J., & Wang, L. F. (2006). Classification of remote sensing image based on object-oriented and class rules. *Geomatics and Information Science of Wuhan University*, 31(4), 316–320.
- Chen, Y. S., Jiang, H. L., Li, C. Y., Jia, X. P., & Chamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10), 6232–6251.
- Chen, Z. X., Zhu, H. Y., & Wang, Y. G. (2013). A modified extreme learning machine with sigmoidal activation functions. *Neural Computing and Applications*, 22(3–4), 541–550.
- Dai, C. G., Huang, X. B., & Dong, G. J. (2007). Support vector machine for classification of hyperspectral remote sensing imagery. *Fuzzy Systems and Knowledge Discovery*, 4, 77–80.
- Fanany, M. I. (2017). Handwriting recognition on form document using convolutional neural network and support vector machines (CNN–SVM). In *Information and communication technology (ICICT), 2017 5th international conference on* (pp. 1–6). IEEE.
- Gulcehre, C., Moczulski, M., Denil, M., & Bengio, Y. (2016). Noisy activation functions. In *International conference on machine learning* (pp. 3059–3068).
- Hu, F., Xia, G. S., Hu, J. W., & Zhang, L. P. (2015). Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing*, 7(11), 14680–14707.
- Jiang, C., Zhang, H., Shen, H., & Zhang, L. (2014). Two-step sparse coding for the pan-sharpening of remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(5), 1792–1805.

- Jin, Y. T., Yang, X. F., Gao, T., Guo, H. M., & Liu, S. M. (2018). The typical object extraction method based on object-oriented and deep learning. *Remote Sensing for Land and Resources*, 30(1), 22–29.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.
- Långkvist, M., Kiselev, A., Alirezaie, M., & Loutfi, A. (2016). Classification and segmentation of satellite orthoimagery using convolutional neural networks. *Remote Sensing*, 8(4), 329–339.
- Lee, S. (2005). Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *International Journal of Remote Sensing*, 26(7), 1477–1491.
- Li, H., Lu, H., Lin, Z., Shen, X., & Price, B. L. (2015). LCNN: Low-level feature embedded CNN for salient object detection. arXiv preprint [arXiv:1508.03928](https://arxiv.org/abs/1508.03928).
- Lin, D., Fu, K., Wang, Y., Xu, G., & Sun, X. (2017). MARTA GANs: Unsupervised representation learning for remote sensing image classification. *IEEE Geoscience and Remote Sensing Letters*, 14(11), 2092–2096.
- Lin, Z., Lanchantin, J., & Qi, Y. (2016). MUST-CNN: a multilayer shift-and-stitch deep convolutional architecture for sequence-based protein structure prediction. In *Thirtieth AAAI conference on artificial intelligence*. AAAI Press (pp. 27–34).
- Mares, M. A., Wang, S., & Guo, Y. (2016). Combining multiple feature selection methods and deep learning for high-dimensional data. *Transactions on Machine Learning and Data Mining*, 9, 27–45.
- Meng, Q., Cieszewski, C. J., Madden, M., & Borders, B. E. (2007). K nearest neighbor method for forest inventory using remote sensing data. *GI Science & Remote Sensing*, 44(2), 149–165.
- Nambiar, V. P., Khalil-Hani, M., Sahnoun, R., & Marsonob, M. N. (2014). Hardware implementation of evolvable block-based neural networks utilizing a cost efficient sigmoid-like activation function. *Neurocomputing*, 140(9), 228–241.
- Paisitkriangkrai, S., Shen, C., & Av, H. (2016). Pedestrian detection with spatially pooled features and structured ensemble learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(6), 1243–1257.
- Qi, K., Yang, C., Guan, Q., Wu, H., & Gong, J. (2017). A Multiscale deeply described correlations-based model for land-use scene classification. *Remote Sensing*, 9(9), 917–927.
- Robertson, L. D., & King, D. J. (2011). Comparison of pixel-and object-based classification in land-cover change mapping. *International Journal of Remote Sensing*, 32(1), 1505–1529.
- Sands, R. D., & Leimbach, M. (2003). Modeling agriculture and land use in an integrated assessment framework. *Climatic Change*, 56(1–2), 185–210.
- Shang, W., Sohn, K., Almeida, D., & Lee, H. (2016). Understanding and improving convolutional neural networks via concatenated rectified linear units. In *International conference on machine learning* (pp. 2217–2225).
- Song, X., Duan, Z., & Jiang, X. (2012). Comparison of artificial neural networks and support vector machine classifiers for land cover classification in Northern China using a SPOT-5 HRG image. *International Journal of Remote Sensing*, 33(10), 3301–3320.
- Su, W., Li, J., Chen, Y. H., Zhang, J. S., Hu, D. Y., & Liu, C. M. (2007). Object-oriented urban land-cover classification of multi-scale image segmentation method: A case study in Kuala Lumpur City Center, Malaysia. *Journal of Remote Sensing*, 11(4), 521–530.
- Tuia, D., Munozmari, J., Gomezchova, L., & Malo, J. (2013). Graph matching for adaptation in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 51(1), 329–341.
- Tzeng, Y. C. (2006). Remote sensing images classification/data fusion using distance weighted multiple classifiers systems. *Journal of Chinese Institute of Engineers*, 31(4), 639–647.
- Wang, W. H., & He, M. (2011). Multi-scale segmentation in land-use information extraction based on object-oriented method. *Science of Surveying and Mapping*, 36(4), 160–161.
- Wang, H., Wang, X., & Dou, A. (2012). Study on the precision evaluation method for a specific category in the classification of remote sensing image. In *International geoscience and remote sensing symposium* (pp. 978–981).
- Woodcock, C. E., & Strahler, A. H. (1987). The factor of scale in remote sensing. *Remote Sensing of Environment*, 21(3), 311–332.
- Yang, J., Chen, F., Xi, J., Xie, P., & Li, C. (2014). A multitarget land use change simulation model based on cellular automata and its application. *Abstract and Applied Analysis*, 2014(6), 1–11.
- Yuan, H., Van Der Wiele, C. F., & Khorram, S. (2009). An automated artificial neural network system for land use/land cover classification from Landsat TM imagery. *Remote Sensing*, 1(3), 243–265.
- Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Proceedings of the European conference on computer vision* (pp. 818–833). Berlin: Springer.
- Zhang, S., Gong, Y. H., & Wang, J. J. (2017). The development of deep convolution neural network and its applications on computer vision. *Chinese Journal of Computers*, 40(9), 1–29.
- Zhang, L., Zhang, L., & 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.
- Zhao, H., Gu, F., Huang, Q., Garcia, J. A., Chen, Y., Tu, C., et al. (2016). Connected fermat spirals for layered fabrication. *International Conference on Computer Graphics and Interactive Techniques*, 35(4), 100–110.
- Zhao, W., Guo, Z., Yue, J., Zhang, X., & Luo, L. (2015). On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery. *International Journal of Remote Sensing*, 36(13), 3368–3379.
- Zheng, Y., Wu, F. D., & Liu, Y. F. (2010). A feature analysis approach for object-oriented classification. *Geography and Geo-Information Science*, 26(2), 19–22.
- Zheng, Z., Zheng, L., & Yang, Y. (2017). A discriminatively learned CNN embedding for person reidentification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(1), 13–23.
- Zhu, Y. S., Zeng, Y. N., & Zhang, M. (2017). Extract of land use/cover information based on HJ satellites data and object-oriented classification. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 33(14), 258–265.