Journal of Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

Using convolutional neural network to identify irregular segmentation objects from very high-resolution remote sensing imagery

Tengyu Fu
Lei Ma
Manchun Li
Brian A. Johnson

SPIE.

Tengyu Fu, Lei Ma, Manchun Li, Brian A. Johnson, "Using convolutional neural network to identify irregular segmentation objects from very high-resolution remote sensing imagery," *J. Appl. Remote Sens.* **12**(2), 025010 (2018), doi: 10.1117/1.JRS.12.025010.

Using convolutional neural network to identify irregular segmentation objects from very high-resolution remote sensing imagery

Tengyu Fu,[a,b] Lei Ma,[a,b,*] Manchun Li,[a,b,*] and Brian A. Johnson[c]
[a]Nanjing University, Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing, China
[b]Nanjing University, School of Geographic and Oceanographic Sciences, Nanjing, China
[c]Natural Resources and Ecosystem Services Area, Institute for Global Environmental Strategies, Hayama, Kanagawa, Japan

Abstract. Convolutional neural network (CNN) has shown great success in computer vision tasks, but their application in land-use type classifications within the context of object-based image analysis has been rarely explored, especially in terms of the identification of irregular segmentation objects. Thus, a blocks-based object-based image classification (BOBIC) method was proposed to carry out end-to-end classification for segmentation objects using CNN. Specifically, BOBIC takes advantage of CNN to automatically extract complex features from the original image data, thereby avoiding the uncertainty caused by the manual extraction of features in OBIC. Additionally, OBIC compensates for the shortcomings of CNN whereby it is difficult to delineate a clear right boundary for ground objects at the pixel level. Using three high-resolution test images, the proposed BOBIC was compared with support vector machine (SVM) and random forest (RF) classifiers, and then, the effect of image blocks and mixed objects on classification accuracy was evaluated for the proposed BOBIC. Compared with conventional SVM and RF classifiers, the inclusion of CNN improved the OBIC classification performance substantially (5% to 10% increases in overall accuracy), and it also alleviated the effect derived from mixed objects. [DOI: 10.1117/1.JRS.12.025010]

Keywords: object-based image classification; convolutional neural network; multiresolution segmentation; irregular segmented object; image block.

Paper 171050 received Dec. 10, 2017; accepted for publication May 2, 2018; published online May 17, 2018.

1 Introduction

Object identification in very high-resolution (VHR) remote sensing imagery has always been a fundamental but challenging issue. In the past few decades, various methods for the identification of different types of objects have been proposed, including the template matching-based method,[1–3] knowledge-based method,[4–6] object-based image analysis (OBIA) method,[7–9] and machine learning-based method.[10,11] Among them, the OBIA method can be easily combined with geographical information system (GIS) techniques, which allows for more complete mapping of land-use types for GIS analyses.[12] Thus, OBIA has attracted the attention of many scholars.[12–14] The first step in OBIA is to segment the images into relatively homogeneous regions (segmentation objects),[15] and then, the statistical information for the segmentation objects is employed for image analyses (e.g., object-based image classification, hereafter, OBIC). As compared with pixels, the segmented objects not only exhibit rich spectral and textural features, but also provide shape and contextual information,[16] which can improve the classification performance for various types of objects.

*Address all correspondence to: Lei Ma, E-mail: maleinju@gmail.com; Manchun Li, E-mail: limanchunnju@163.com

However, the sharp increase in the feature number for each segmentation object renders the determination of optimal features as an uncertain or subjective process. For example, Weston et al.[17] and Guyon and Elisseeff[18] pointed out that reducing feature dimensions could improve support vector machine (SVM) classification accuracy, whereas Melgani and Bruzzone[19] and Pal and Mather[20] deemed that SVM was insensitive to the number of data dimensions. Likewise, Duro et al.[21] found that feature selection could improve the classification performance of the random forest (RF) classifier,[22] whereas Ma et al.[23] deemed that RF was a relatively stable classification model, as they found that there were no significant differences among its classification accuracies irrespective of the use of feature selection. Presently, the feature selection process is always associated with an uncertainty factor during OBIC using traditional classification models. Emerging deep learning[24] methods are famous for their ability to carry out automatic feature extraction on raw data, and therefore, such methods could potentially be used to optimize the process of feature extraction and selection in OBIC. However, deep learning methods have not been extensively tested in land-use type classifications, especially within the framework of OBIA.

As deep learning was proposed,[24] it has received extensive attention from many scholars because it can automatically generate complex and abstract high-level features in a hierarchical manner.[25] High-level features have proven to be highly effective in representing complex objects (e.g., high-resolution images).[26] The convolutional neural network (CNN) is one of the algorithms with most rapid development in deep learning and was specially designed for image classification tasks.[27,28] Images served as the input at the lowest layer in the CNN's hierarchical structure, and each layer obtains the features of the upper layer through a convolution filter.[29] Moreover, with increased hierarchical depth, features became more and more robust and complex. This allows for salient features of translation-, scaling-, and rotation-invariant data to be obtained.[30] However, a major drawback is that the input of the CNN framework must be image blocks of a fixed size. This poses a certain challenge in terms of combining CNN with object-based remote sensing image classification because the minimum processing unit of OBIA is usually irregular segmentation objects.

Despite the above problems, the continuing success of CNN in the field of image recognition[31,32] has motivated researchers in the remote sensing community to investigate its potentials for OBIA. Guirado et al.[33] compared state-of-the-art OBIA methods with CNN-based methods for the detection of plant species of conservation concern and reasoned that adopting the CNN-based methods could further improve OBIA methods. Zhao et al.[34] proposed a two-step OBIC framework using a combination of handcrafted and deep CNN features. In their work, however, CNN only served as a feature descriptor of segmentation objects, which makes the process of feature selection in OBIA still uncertain. Liu et al.[35] implemented end-to-end classifications of wetland land cover under the OBIA framework and tested the classification performance of the model using different training samples. However, their work did not systematically assess the geometric relationship of the irregular segmentation objects to the input image blocks of the CNN; it only focused on the identification of wetland land cover. All of the above studies show that the CNN can effectively improve the OBIC classification performance in specific contexts, so work is urgently needed to systematically evaluate the availability of classifying irregular segmentation objects using CNN.

In a similar way, this paper considers that including CNN in an OBIA framework could take advantage of the benefits of both methods, e.g., OBIA segmentation to delineate homogeneous areas and CNN for classification. Hence, a blocks-based object-based image classification (BOBIC) method is proposed to combine OBIA with CNN. In this work, the multiresolution segmentation (MRS) algorithm was employed to generate highly irregular segmentation objects.[36] Image blocks were subsequently generated according to the center of gravity (CG) of the segmentation objects, thereby combining irregular objects with the CNN. Furthermore, the differences between this method and conventional classifiers were compared systematically at three study sites, and the effects of segmentation object shape and mixed objects on the classification accuracy were also analyzed. The remaining parts of this paper are organized as follows: Sec. 2 introduces the three study sites that were used in the experiments. Section 3 elaborates on how to apply CNN in OBIA and the experimental procedures

used in this paper. The experimental results are presented in Sec. 4, and Sec. 5 contains a discussion of the experimental results. Finally, Sec. 6 summarizes the entire paper.

2 Study Area

In this work, unmanned aerial vehicle (UAV) images and International Society for Photogrammetry and Remote Sensing (ISPRS) standard datasets corresponding to agricultural areas and urban areas, respectively, were employed for the experiments. Images for study site 1 were sourced from the high-resolution image acquisition project in Deyang City, Sichuan Province, China.[37] This project adopted a fixed-wing UAV equipped with a Canon EOS 5D Mark II digital camera. At 80% heading overlap and 60% side overlap and with an average flight altitude of 750 m, the UAV captured raw image data for the built-up area and suburban area of Deyang City with a total area of 400 km^2 in August 2011. Furthermore, a digital orthophoto map (DOM) with a resolution of 0.2 m was finally obtained using digital photogrammetric techniques. In this work, a standard-sized UAV DOM (500 m × 500 m) [Fig. 1(a)] was randomly selected, where crop (41%), woodland (46%), buildings (6%), roads (2%), and bareland (5%) were distributed.

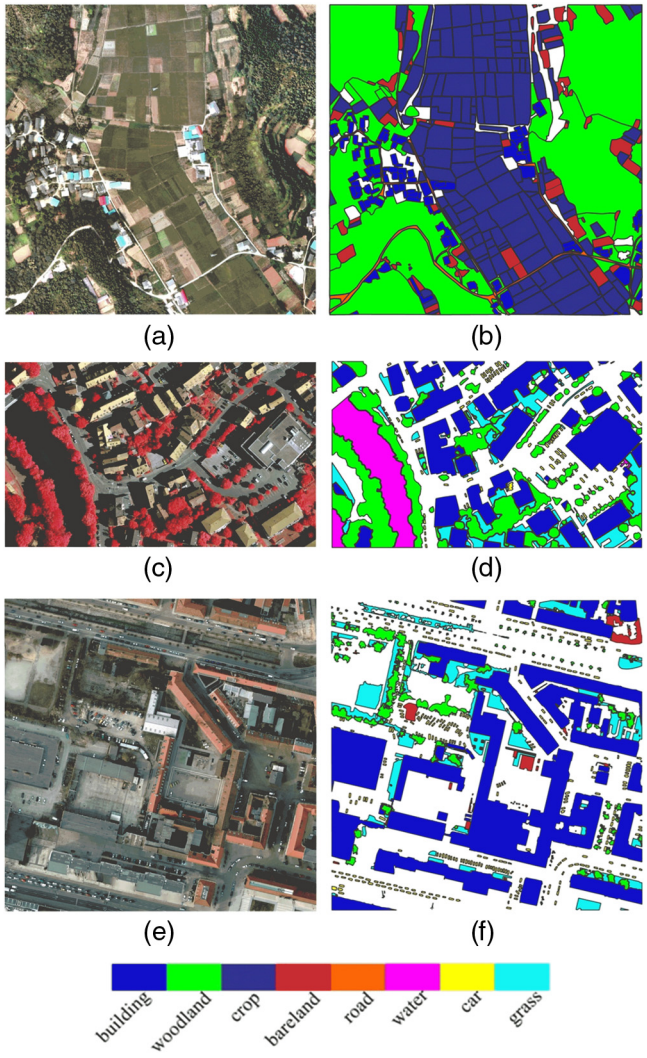

Fig. 1 Images of the study sites in this work and their corresponding reference layers. (a), (c), and (e) The images of the three study sites; (b), (d), and (f) the corresponding reference (labeled) layers of three study sites.

Study sites 2 and 3 employed Vaihingen and Potsdam datasets provided by the ISPRS Commission III, respectively. These datasets can be downloaded freely from the ISPRS website.[38] The Vaihingen dataset contains a total of 33 aerial images of varying sizes (average size of 2494 pixels × 2064 pixels), 16 of which also have visually interpreted reference (labeled) polygons, and the spatial resolution for each aerial image is 9 cm. In this work, one image (region 26) was randomly selected from the 16 visually interpreted images for study site 2 [Fig. 1(c)], where buildings (42%), woodland (29%), water (12%), cars (3%), and grass (14%) were distributed. The Potsdam dataset comprises a total of 38 aerial images (each image size was 6000 pixels × 6000 pixels), 24 of which have visually interpreted reference polygons, and the spatial resolution for each aerial image is 5 cm. Likewise, one image (region 07_12) was randomly selected from the 24 visually interpreted images for study site 3 [Fig. 1(e)], where buildings (69%), woodland (9%), bareland (3%), cars (4%), and grass (15%) were distributed. Images of the three study sites and their corresponding visually interpreted polygon layers are shown in Fig. 1.

3 Methods

As mentioned in Sec. 1, traditional OBIA methods require a large number of image features to be empirically designed, which is time-consuming and often fails to lead to accurate representations. In contrast to traditional methods, the CNN can perform automatic feature extraction on raw images, and deep features extracted by the CNN are generally effective for complex image pattern descriptions.[31,32] However, CNN often fail to capture the precise contour of real-world objects in the images, and suffer from the "pepper and salt" effect because the output features of CNN are highly abstract. Thus, it is natural to consider that including CNN in an OBIA

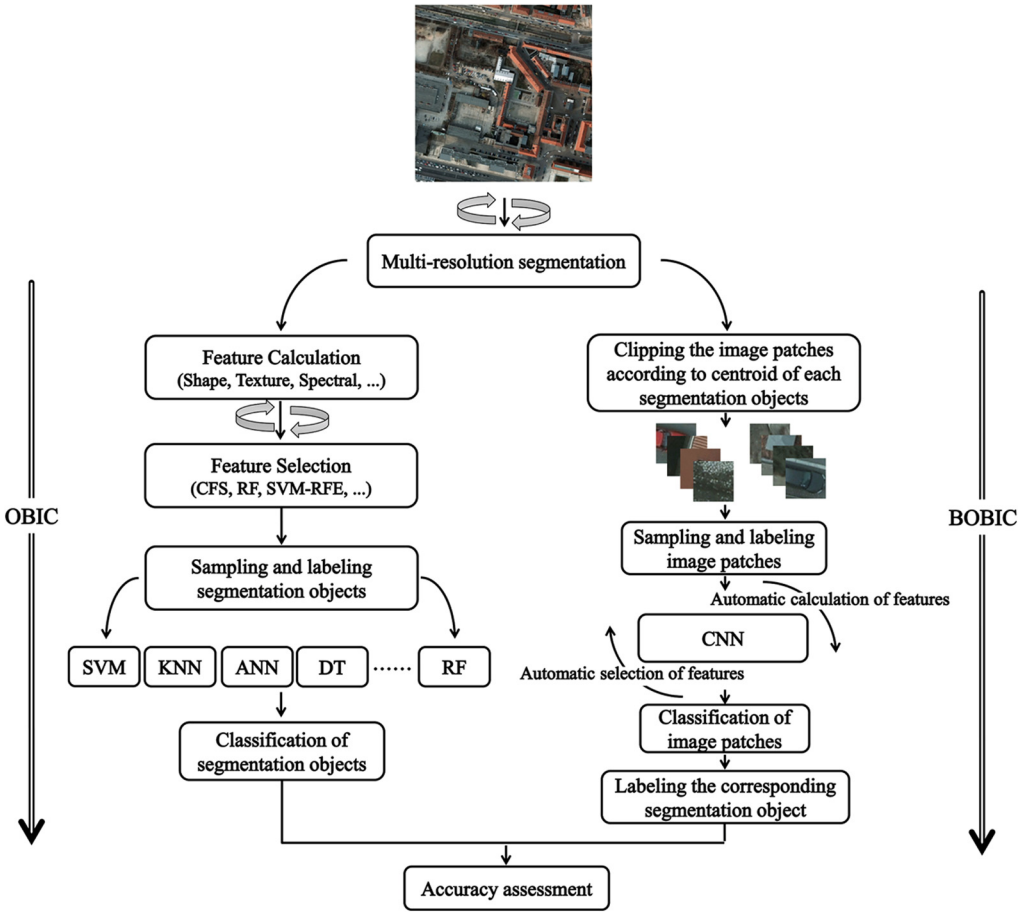

Fig. 2 Flowchart of the comparison between the OBIC method and BOBIC methods.

framework can take advantage of the benefits of both methods, i.e., CNN for object classification and OBIA segmentation to provide accurate edge realizations. However, the CNN framework requires fixed-sized image blocks as input, which limits its development in the OBIA framework. In consideration of this issue, in this paper, we try to propose a BOBIC method to classify irregular segmentation objects using CNN. Figure 2 summarizes the technical roadmaps of OBIC and the proposed BOBIC.

As shown in Fig. 2, OBIA involves two steps, namely image segmentation and object classification. The proposed BOIBC method involves applying CNN to the object classification step so as to improve the OBIC method. Therefore, image segmentation is the common step of these two methods, and this is described in detail in Sec. 3.1. Object classification is divided into two parts, namely OBIC (Sec. 3.2) and BOBIC (Sec. 3.3). Furthermore, the object classification process of the traditional OBIC method mainly includes the following two steps: feature calculation and selection (Sec. 3.2.1) and classifier selection (Sec. 3.2.2). The proposed BOBIC method can automatically perform the feature calculation and selection of images using the CNN, but there is a need to generate a unique image block corresponding to each segmentation object. The generation of image blocks for segmentation objects is elaborated on in Sec. 3.3.1, and Sec. 3.3.2 presents the structure of the CNN used in this paper. In addition, the sampling and accuracy assessment methods are described in Sec. 3.4.

3.1 *Image Segmentation*

Image segmentation is the first step and a necessary prerequisite for generating the basic classification unit of OBIA.[39–41] MRS has been proven to be one of the rather successful segmentation algorithms in OBIA.[42,43] In this paper, image segmentation was performed for three study sites in a unified manner using MRS implemented with eCognition 8.7 software (eCognition Software® Definiens, 2011),[36] and subsequently, irregular segmentation objects were generated. The following three parameters need to be set for the MRS: color/shape ratio, smoothness/compactness ratio, and segmentation scale parameter (SSP). The color/shape ratio defines what percentage of the homogeneity of spectral values is weighted against the homogeneity of shape. The smoothness/compactness ratio is used to determine the smoothness or compactness of each object. In this work, to make the spectral information have a dominant role during segmentation, the color/shape ratio was set to 0.9/0.1. The smoothness/compactness ratio was configured to 0.5/0.5, because we did not want to favor compact or noncompact segments.

The most important parameter for MRS is the SSP, which controls the internal heterogeneity of each object. Specifically, use of a small SSP results in smaller and more homogeneous objects, i.e., fewer pixels per object. However, using an overly small object size (i.e., over-segmentation) may affect the quality of the information extracted from each object[44] and increase the computational burden of the subsequent classification process. Conversely, an overly large SSP (i.e., under-segmentation) will produce objects containing multiple different classes (i.e., this leads to the generation of mixed objects[45]). Automated identification/selection of

Table 1 The number of segmentation objects for various land-use types at three study sites; data were derived using segmentation scales of 50 and 110.

Study sites	Class	Count (50)	Count (110)	Study sites	Class	Count (50)	Count (110)	Study sites	Class	Count (50)	Count (110)
1	Bareland	286	89	2	Grass	331	74	3	Grass	1069	263
	Woodland	2832	608		Woodland	1030	263		Woodland	740	159
	Building	439	143		Building	1381	435		Building	6758	1732
	Crop	1148	290		Car	119	27		Car	1212	469
	Road	145	39		Water	116	35		Bareland	437	102
	Total	4850	1169		Total	2977	834		Total	10216	2725

the "appropriate" SSP(s) for segmentation (i.e., those which can minimize under- and over-segmentation) is still an active research topic.[16,46,47] In this research, two SSPs (50 and 110), selected based on visual analysis, were employed for image segmentation to enrich the experimental results.

Additionally, if the area of a primary class that was encompassed by the segmentation object accounted for over 60% of the total area of this segmentation object, then this segmentation object was labeled with this class (otherwise the segmented object was left unlabeled). Here, the proportion of the primary class was set to 60% with reference to the research by Verbeeck et al.[48] and Ma et al.[23] The numbers of segmentation objects for various classes at the three study sites are shown in Table 1.

3.2 *Object-Based Image Classification*

3.2.1 *Feature calculation*

Features of segmentation objects need to be calculated to employ conventional OBIC algorithms (e.g., SVM or RF). In this paper, eCognition 8.7 software was adopted to calculate commonly used shape, textural, and spectral features. The shape features included the area, density, roundness, compactness, border index, shape index, main direction, elliptic fit, rectangular fit, and asymmetry; the textural features included the gray-level co-occurrence matrix (GLCM) entropy, GLCM std. dev., GLCM contrast, GLCM dissimilarity, GLCM homogeneity, GLCM mean, GLCM ang.2nd moment, and GLCM correlation that were computed according to the GLCM[49,50] as well as the gray-level difference vector (GLDV) entropy, GLDV contrast, GLDV mean, and GLDV ang.2nd moment that were derived from the GLDV;[51] the spectral features included the mean blue, mean green, mean red, max difference, standard deviation blue, standard deviation green, standard deviation red, and brightness. Considerable uncertainty exists concerning feature selection with regard to different classifiers.[52,53] Hence, feature selection has not been performed for the above-mentioned features.

3.2.2 *Selection of conventional classifiers*

The SVM and RF classifiers have been extensively applied, and such studies have demonstrated their classification advantages in OBIA multiple times.[23,52,54–57] Hence, in this work SVM and RF classifiers were utilized to classify the extracted features in Sec. 3.2.1. The SVM used the LIBSVM library that was developed by Chang and Lin,[58] and we employed the radial basis function (RBF)[59] as its kernel function. The RBF involved penalty parameter C and kernel parameter γ. The accuracy of each cross validation was tested by using the grid-search method, and thus, the parameters with the highest cross-validation accuracy could be identified as the penalty parameter and kernel parameter. The RF classifier used the "randomforest" package in R language. Roughly speaking, constructing an RF classifier requires the following two parameters: (1) n is the number of features when each decision tree is constructed, (2) k is the total number of decision trees. Based on the results obtained by Rodriguez-Galiano et al.,[60] k was set to 479, and n was equivalent to one single random segmentation variable; the intent was to reduce the generalization error and the correlation between trees and prevent over-fitting in the classification process as much as possible.

3.3 *Blocks-Based Object-Based Image Classification*

3.3.1 *Generation of image blocks for segmentation objects*

Image blocks of a fixed size have to be generated for each segmentation object to use CNN in OBIC. The size of an image block is constrained by the depth of the CNN network and the capacity of computer memory.[61] With respect to subsequent experiments in this work, supervised classification tests were conducted mainly using a small sample size, where the ultra-large scale CNN framework could not be adopted. Hence, 32×32 and 64×64 pixel shapes were selected as the size of the image block. In addition, in this paper the CG for the segmentation object

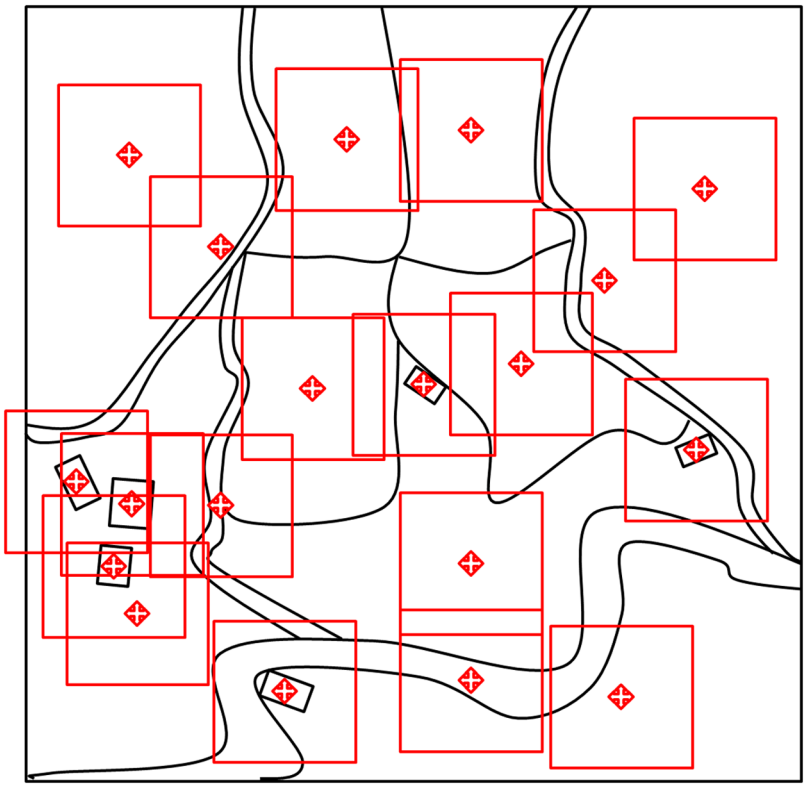

Fig. 3 Schematic generation of image blocks for irregular segmentation objects. (Black lines denote the segmentation boundaries of irregular segmentation objects, red cross points represent the CG of irregular segmentation objects, and red square boxes indicate the range of a sampled image block.)

served as the center point of an image block. Each segmentation object corresponded to one unique image block. In addition, the class of an image block was in good agreement with that of the corresponding segmentation object. Figure 3 shows a schematic of the generation of image blocks for irregular segmentation objects, where black lines denote the segmentation boundaries of irregular segmentation objects, red cross-points represent the CG of irregular segmentation objects, and red square boxes indicate the range of a sampled image block.

It can be seen from Fig. 3 that the CG of a convex polygon, in most cases, fell inside the polygon. However, with respect to a nonconvex polygon, its CG exhibited a certain shift. This presented a challenge with regard to the application of the proposed BOBIC method. Hence, we summarized in detail the geometric relationship of irregular segmentation to the input image blocks of the CNN. First, when the CG of a segmentation object fell within the segmentation object, there existed a total of the following three situations:

1. The CG fell inside the segmentation object, and the image block entirely encompassed the segmentation object.
2. The CG fell within the segmentation object, and the segmentation object entirely encompassed the image block.
3. The CG fell inside the segmentation object, and the image block encompassed a portion of the segmentation object.

Second, under circumstances where the CG fell outside a segmentation object, it was impossible for the segmentation object to encompass the image block. In addition, the CG was likely to either fall within the segmentation object of the same type or fall inside the segmentation object of a different type. No difference existed in the former case between situations 1 and 3. This was because the center point of the image block always fell on the land cover of the same type, and the class of the segmentation object that corresponded to the image block remained unchanged. Hence, this situation was not listed separately, i.e., the situation where the CG fell within the land

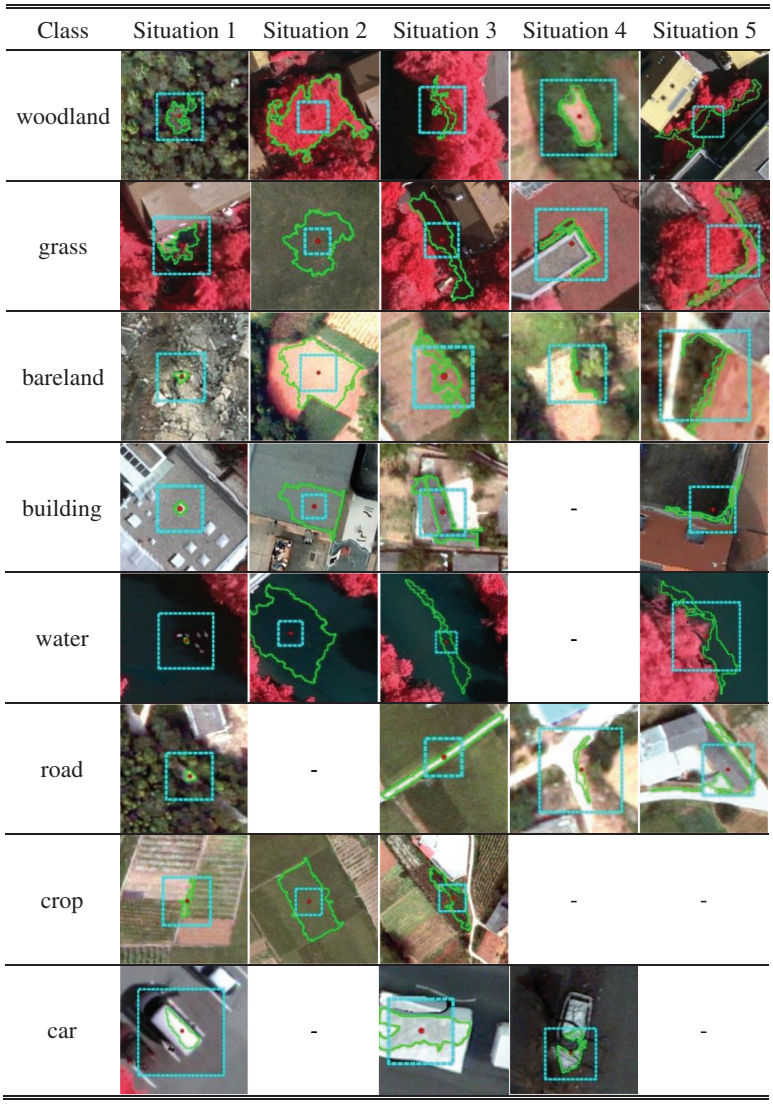

Fig. 4 Five situations amongst image blocks and segmentation objects. ("–" indicates that this situation does not exist with respect to the current land cover class, image blocks are enveloped by bright blue dotted boxes, bright green solid boxes depict the range of segmentation objects, and red points are the CG of segmentation objects.)

cover of the same type was included in situations 1 and 3 correspondingly. Then, the remaining situations were as follows:

1. The CG fell outside the segmentation object, and it fell inside different types of segmentation objects, where the image block entirely encompassed the segmentation object.
2. The CG fell outside the segmentation object, and it fell inside different types of segmentation objects, where the image block encompassed a portion of the segmentation object.

The above five situations with different types of land covers are shown in Fig. 4.

3.3.2 *Convolutional neural network*

The CNN consisted mainly of three different types of hierarchical structures, specifically, convolution layers, pooling layers, and fully connected layers. Convolution layers, also known as feature extraction layers, constitute the primary layers of CNN architecture. The input of convolution layers comprises a set of two-dimensional (2-D) feature maps of a fixed size. In the

convolution phase, trainable filter W (convolution kernel) performs the convolution operation by using a sliding window technique.[62,63] Assume the convolution kernel is $i \times j$ in size, and then, the output feature map Y that corresponds to X can be written as follows:

$$Y_{m,n} = f\left(b + \sum_{i=0} \sum_{j=0} W_{i,j} X_{m+i,n+j}\right), \tag{1}$$

where m, n denote the row and column number of a hidden neuron in the 2-D feature map, b is a trainable bias parameter, and f represents the particular nonlinear activation function.

Pooling layers are down-sampling layers in the CNN architecture, which can enhance the spatial-invariance property of the convolutional architecture.[64] A down-sampling operation was performed for each 2-D feature map normally through max pooling.[65] The max pooling operation aims to compute the maximum value of a neuron within the local region, which is expressed as

$$Y = \max_{1<m<i,1<n<j} X_{m,n}, \tag{2}$$

where (i, j) denotes the size of the local region X, m, n represents the row and column number of a neuron inside the local region, and Y is the output of the max pooling operation, respectively.

Fully connected layers generally constitute the last few layers of the CNN architecture, which accept all neurons in a 2-D feature map and connect them to one-dimensional neurons. With regard to a multiclass problem, the number of neurons for the last fully connected layer equals the number of classes for the final classification. In addition, the last fully connected layer is normally followed by the Softmax layer,[66] which can be used to obtain the discrimination probability for each class. The equation is given as

$$Y_i = \frac{\exp(X_i)}{\sum_{j=1}^{k} \exp(X_j)}, \tag{3}$$

where X_i denotes the output of class i in the last fully connected layer, k is the number of classes, and Y_i represents the discrimination probability for class i, respectively.

In this work, the architecture of VGG-Net[67] was used as a reference. The end-to-end training was performed for image blocks of segmentation objects using the CNN architecture as shown in Fig. 5.

The CNN architecture shown in Fig. 5 is comprised of four convolution layers (blue layers as shown in Fig. 5). Each convolution layer used a 3×3 convolution kernel, and convolution operations were performed with stride 1 for the 2-D feature maps in the previous layer. The first two convolution layers produced 32-dimensional output, whereas the latter two generated

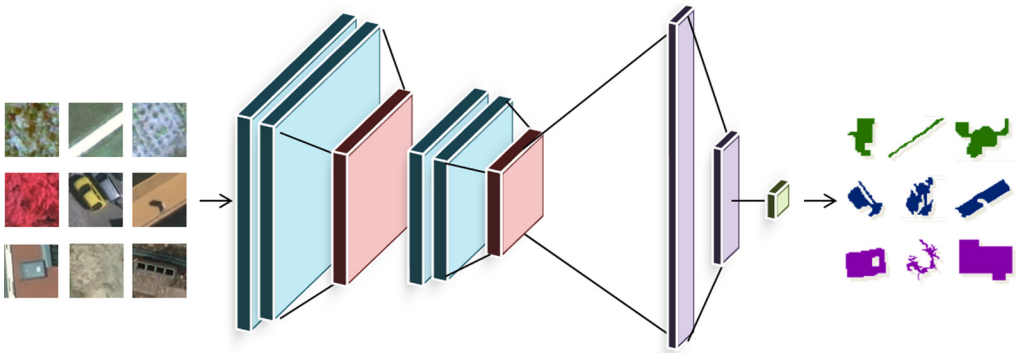

Fig. 5 CNN architecture employed in this work. Image blocks that were generated in Sec. 3.2.1 served as the CNN input, and the output of CNN was comprised of the classes of segmentation objects that corresponded to the image blocks, blue layers represent convolution layers (using ReLU as the activation function), red layers are pooling layers (using the max pooling layer), purple layers denote fully connected layers, and green layers are Softmax layers.

64-dimensional output. A rectified linear unit (ReLU)[25] can address the gradient disappearance phenomenon well.[68,69] Therefore, ReLU was adopted as the activation function for each convolution layer. Every two convolution layers were followed by a 2×2 max pooling layer (red layers as shown in Fig. 5). The first purple layer in Fig. 5 shows a fully connected layer that was comprised of 512 neurons, whereas the number of neurons for the last fully connected layer (the second purple layer in Fig. 5) was equal to the number of land-use types in the three study sites, all being 5 in this work. Finally, the Softmax function was applied after the last fully connected layer, which allowed for the generation of the green class output as shown in Fig. 5.

To avoid the risk of overfitting,[70] the following strategies were adopted in this work:

1. Employ the dropout technique after the pooling layers and fully connected layers.[71] The dropout technique aims to avoid co-adaptation of neurons during training. It randomly "shuts down" a given percentage of neurons during CNN training, thereby reducing the overfitting risk. In this work, the dropout percentage after the max pooling layer was set to 20%, and it was set to 50% after the fully connected layer.
2. Apply the early stopping technique that monitors a certain value (normally the loss value). The CNN training stops when this value does not increase or decrease after multiple epochs. In this paper, the loss values of training samples were monitored. When these values were all <0.1 within 20 epochs, the CNN training was stopped.
3. Data augmentation can extend data without increasing the number of training samples. The commonly used enhancement strategies include random image rotation, random image scaling, horizontal image shift, and noise injection. To maintain the high resolution of images, only random rotation was performed on the images.

In addition, all the weights in convolution layers and fully connected layers were initialized using the He normal distribution.[68] In this work, the CNN was trained from scratch using the end-to-end method.

3.4 *Sampling and Accuracy Evaluation*

Regardless of whether the base unit of classification is a segmented object or an image block generated based on the segmentation object, it makes no difference from the perspective of sampling. Hence, the random sampling method was adopted in the experiments. Proportions amounting to 10%, 20%, 30%, 40%, and 50% of the total number of segmentation objects in three study sites were sampled as training sample sets, whereas the remaining samples served as test sample sets. The classification accuracy was derived by dividing the number of correctly classified segmentation objects in the test sample set by the total number of segmentation objects in the test sample set. Twenty-time random samplings were performed with respect to each sampling ratio, and then, statistics were collected for the classification accuracies with regard to 20-time samplings. Finally, the mean value and standard deviation of classification accuracies were computed.

In addition, Welch's t-test[72] was used to test whether significant differences existed between two sets of data. Specifically, Welch's t-test was performed on the classification accuracies with respect to adjacent sampling ratios, thereby allowing us to assess whether significant differences existed in terms of the classification accuracies of adjacent sampling ratios. *P*-values were derived from the Welch's *t*-test, and significant differences were deemed to exist between two sets of data when the *p*-value was <0.05.

4 Results

This section contains a complete description of the classification performance of the conventional OBIC method and the proposed BOBIC method. First, to test whether the BOBIC method could achieve higher land-use type classification accuracy than the OBIC methods, we compared the two methods at the three study sites using five sampling ratios and two different SSPs (results presented in Sec. 4.1). Second, as discussed in Sec. 3.3.1, the geometric relationships between the segmented objects and image blocks were complex. Often the image block did not entirely

contain its corresponding segmented object, which presented a challenge during the application of the proposed method. Therefore, the classification error rates of different geometric relationships were calculated in Sec. 4.2 to assess the influence of these geometric relationships on classification accuracy. In addition, the mixed objects were a special but easily overlooked issue in the framework of OBIA. On the one hand, the classification accuracy of the mixed objects tended to be lower, because they often contained pixels belonging to many different land-use classes. On the other hand, the existence of mixed objects could not be avoided because of the limitation of the current segmentation algorithm. So, we counted the classification accuracy of mixed and pure objects in Sec. 4.3 to evaluate the applicability of the proposed method to mixed objects.

4.1 *Comparison of OBIC and BOBIC in Terms of the Classification Effect*

Based on the sampling and accuracy evaluation methods described in Sec. 3.4, final classification results were obtained using the OBIC method and BOBIC method, and these results are shown in Tables 2 and 3. In addition, the classification objects for SVM and RF classifiers were extracted features of the segmentation objects described in Sec. 3.2.1, which represents OBIC; additionally, the classification objects for the CNN were image blocks that were generated using the CGs of segmentation objects in Sec. 3.3.1, which represents the proposed BOBIC. Table 2 shows the mean value and standard deviation of classification accuracies for 20-time random

Table 2 The mean value and standard deviation of classification accuracies for 20-time random samplings based on different sampling ratios with a segmentation scale of 50 for three study sites.

	OBIC (SVM)		OBIC (RF)		BOBIC (32×32)		BOBIC (64×64)	
Sample ratio (%)	Accuracy (Mean)	Accuracy (Std)	Accuracy (Mean)	Accuracy (Std)	Accuracy (Mean)	Accuracy (Std)	Accuracy (Mean)	Accuracy (Std)
				Study site 1				
10	0.7947	0.0070	0.7791	0.0093	0.8380	0.0052	0.8594	0.0069
20	0.8216	0.0065	0.7957	0.0078	0.8592	0.0049	0.8881	0.0037
30	0.8304	0.0073	0.8024	0.0073	0.8740	0.0036	0.8908	0.0052
40	0.8369	0.0054	0.8095	0.0059	0.8765	0.0034	0.9023	0.0044
50	0.8385	0.0061	0.8148	0.0063	0.8851	0.0041	0.9096	0.0030
				Study site 2				
10	0.8322	0.0142	0.8301	0.0097	0.8771	0.0056	0.9001	0.0041
20	0.8503	0.0079	0.8453	0.0060	0.9042	0.0061	0.9253	0.0059
30	0.8644	0.0088	0.8512	0.0065	0.9117	0.0034	0.9288	0.0046
40	0.8746	0.0064	0.8555	0.0081	0.9207	0.0037	0.9370	0.0040
50	0.8807	0.0100	0.8605	0.0099	0.9257	0.0039	0.9529	0.0038
				Study site 3				
10	0.7825	0.0055	0.7437	0.0063	0.8361	0.0032	0.8865	0.0046
20	0.8050	0.0051	0.7670	0.0055	0.8633	0.0042	0.9109	0.0050
30	0.8159	0.0055	0.7784	0.0044	0.8749	0.0046	0.9270	0.0043
40	0.8219	0.0051	0.7843	0.0065	0.8835	0.0046	0.9364	0.0046
50	0.8264	0.0048	0.7928	0.0038	0.8919	0.0018	0.9412	0.0017

Table 3 The mean value and standard deviation of classification accuracies for 20-time random samplings based on different sampling ratios with a segmentation scale of 110 for three study sites.

	OBIC (SVM)		OBIC (RF)		BOBIC (32 × 32)		BOBIC (64 × 64)	
Sample ratio (%)	Accuracy (Mean)	Accuracy (Std)	Accuracy (Mean)	Accuracy (Std)	Accuracy (Mean)	Accuracy (Std)	Accuracy (Mean)	Accuracy (Std)
				Study site 1				
10	0.7249	0.0240	0.7231	0.0191	0.7878	0.0105	0.7974	0.0115
20	0.7667	0.0162	0.7585	0.0169	0.8131	0.0082	0.8235	0.0108
30	0.7889	0.0129	0.7721	0.0155	0.8298	0.0132	0.8474	0.0105
40	0.8045	0.0136	0.7875	0.0178	0.8337	0.0096	0.8558	0.0091
50	0.8131	0.0150	0.7952	0.0192	0.8296	0.0068	0.8626	0.0090
				Study site 2				
10	0.8421	0.0170	0.8355	0.0271	0.8557	0.0109	0.8920	0.0110
20	0.8561	0.0120	0.8557	0.0100	0.8924	0.0071	0.8975	0.0071
30	0.8633	0.0100	0.8648	0.0121	0.8979	0.0077	0.9181	0.0076
40	0.8760	0.0150	0.8706	0.0164	0.9079	0.0064	0.9196	0.0064
50	0.8818	0.0121	0.8765	0.0143	0.9054	0.0061	0.9351	0.0065
				Study site 3				
10	0.7530	0.0164	0.7410	0.0171	0.7880	0.0075	0.8361	0.0096
20	0.7939	0.0126	0.7695	0.0099	0.8220	0.0067	0.8662	0.0107
30	0.8149	0.0083	0.7887	0.0080	0.8410	0.0074	0.8905	0.0047
40	0.8270	0.0098	0.7924	0.0111	0.8476	0.0051	0.8888	0.0073
50	0.8347	0.0096	0.7996	0.0126	0.8665	0.0059	0.9043	0.0063

samplings on five sampling ratios using four classification methods with a segmentation scale of 50.

Meanwhile, Table 3 shows the mean value and standard deviation of classification accuracies for 20-time random samplings on five sampling ratios, using four classification methods with a segmentation scale of 110.

According to the results shown in Tables 2 and 3, the following observations can be made. (1) The classification accuracies of the proposed BOBIC on five sampling ratios were all superior to the OBIC method. (2) The classification accuracy of image blocks with 64 pixels × 64 pixels was obviously superior to that of image blocks with 32 pixels × 32 pixels. (3) The BOBIC method was characterized by better classification stability. The variance of its classification accuracies under corresponding sampling ratios remained less than that of the two conventional classifiers.

Based on the BOBIC experimental results presented in Tables 2 and 3, the Welch's *t*-test was conducted for adjacent sampling ratios (Sec. 3.4), and these results are shown in Table 4. From a vertical perspective of Tables 2 and 3 as well as in combination with Table 4, when the sampling ratio increased from 10% to 20%, the classification accuracy of the BOBIC exhibited a marked increase (most of the p-values were all <0.05). With regard to the remaining adjacent sampling ratios, the improvement in classification accuracy did not exhibit an obvious pattern.

Graphical representations of the classification performance for the three study sites were prepared with respect to the optimal classification results of 20-time random samplings by

Table 4 Welch's t-test results for the BOBIC with respect to adjacent sampling ratios.

	Study site 1		Study site 2		Study site 3	
	p-value	*p*-value	*p*-value	*p*-value	*p*-value	*p*-value
Adjacent ratio	(32 × 32)	(64 × 64)	(32 × 32)	(64 × 64)	(32 × 32)	(64 × 64)
			SSP (50)			
10%/20%	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05
20%/30%	<0.05	>0.05	<0.05	>0.05	<0.05	<0.05
30%/40%	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05
40%/50%	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05
			SSP (110)			
10%/20%	<0.05	<0.05	<0.05	>0.05	<0.05	<0.05
20%/30%	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05
30%/40%	>0.05	<0.05	<0.05	>0.05	<0.05	>0.05
40%/50%	>0.05	>0.05	>0.05	<0.05	<0.05	<0.05

Note: A *p*-value <0.05 indicates that a significant difference exists between the two sets of data.

using a sampling ratio of 50% (Fig. 6). It can be observed from Fig. 6 that, compared with the OBIC method (SVM and RF), the classification performance of the proposed BOBIC was more "clear-cut," i.e., it overcame the so-called "pepper and salt" effect. Specifically, different land cover types were characterized by more clear boundaries, e.g., woodland, farmland, and barren land in study site 1; water bodies and buildings in study site 2; and buildings, woodland, barren land, and grassland in study site 3, respectively. In summary, the proposed BOBIC method improved the overall classification performance of the traditional OBIC.

4.2 *Classification Effect of Different Geometric Relationships between Image Blocks and Segmented Objects*

The geometric relationship between image blocks and segmentation objects forms an important part of the proposed BOBIC method. So this section provides a further statistical analysis of the five situations summarized in Sec. 3.3.1. Table 5 presents the number of segmentation objects in the three study sites under different situations.

With a sampling proportion of 50%, the classification error rates for each situation were calculated, as shown in Table 6.

The following could be clearly observed from Tables 5 and 6. (1) The probability for the occurrence of situation 4 and 5 remained extremely low, but their error rates were very high. (2) The error rate of situation 2 remained very low; however, the number of training samples for situation 2 was very small. (3) The numbers for situation 1 and 3 accounted for the vast majority of the total number of segmentation objects, and the error rates of these two situations were close.

4.3 *Effects of the BOBIC Method on the Classification of Mixed Objects*

The effects of the BOBIC method on the classification of mixed objects are discussed in this section. The ratio of the area of the primary class in a segmentation object to the total area of the segmentation object [referred to as the primary class proportion (PCP)] was employed as an indicator to measure the mixed degree of the segmentation objects. When the PCP was 100%, the segmentation object was a pure object. Lower PCP values reflect the greater mixed degree of the segmentation objects. Then, statistics were collected for the ratios of sample sizes for the different intervals of the PCP to the total sample size, as shown in Fig. 7.

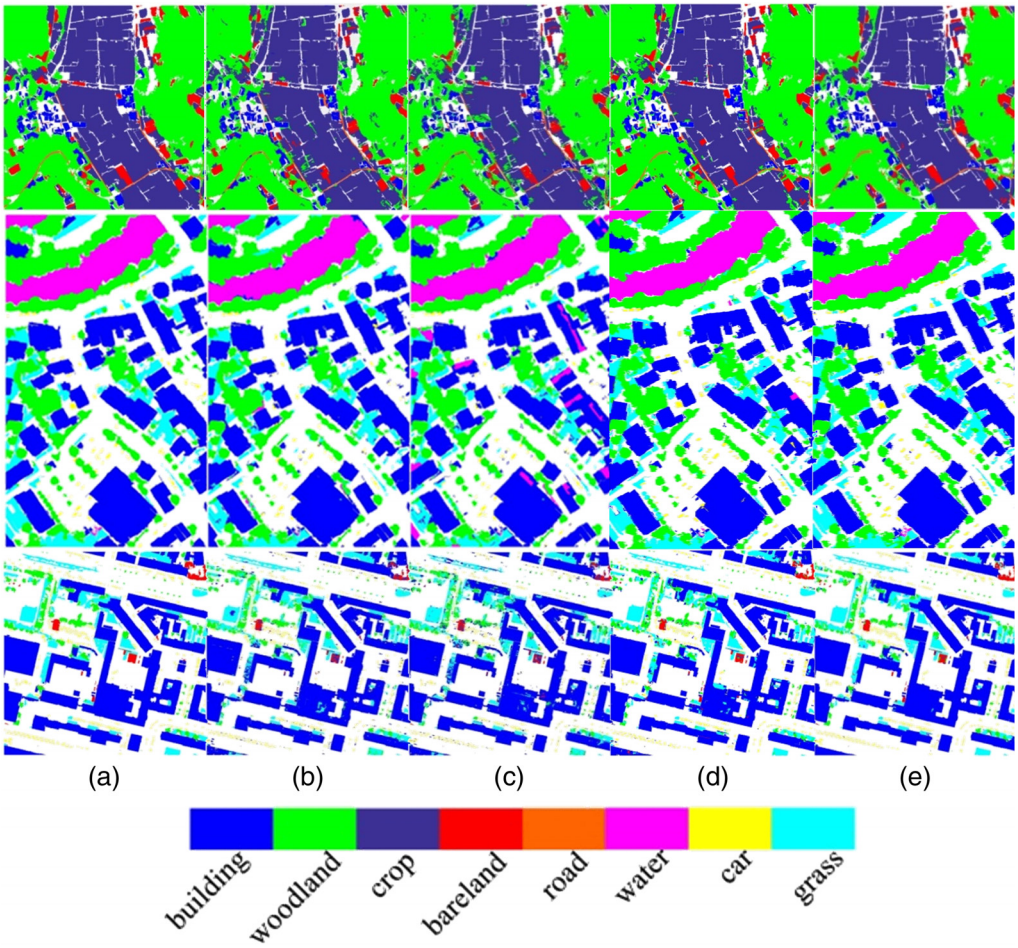

Fig. 6 Graphical representations of the classification performance for different study sites using a sampling ratio of 50%. [For the different study sites, (a) is the vector graph of the fully correct classification, (b) is the vector graph classified by using the SVM classifier, (c) is the vector graph classified by using the RF classifier, (d) is the vector graph of the BOBIC with an image block size of 32 pixels × 32 pixels, and (e) is the vector graph of the BOBIC with an image block size of 64 pixels × 64 pixels.]

Smaller SSP values were associated with more severe over-segmentation. Therefore, the number of pure objects with a segmentation scale of 50 was obviously larger than that with a scale of 110 in Fig. 7. In addition, with decreasing levels of the mixed degree (increases in the PCP), the number of segmentation objects increased gradually. We used the classification model with a sampling rate of 10% described in Sec. 4.1 to classify all segmentation objects in the study area, and then, we computed the classification accuracies for the different intervals of the PCP. The sampling ratio of 10% was selected to minimize the difference that different classifiers would impose varying levels of fitting on training samples. Figure 8 shows a combo line and column chart for the classification accuracies of the different intervals of the PCP at the three study sites.

First, as observed from Fig. 8, the classification accuracies of the BOBIC method over different intervals of the PCP were almost all superior to those of the SVM and RF classifiers, in particular with respect to image blocks of 64 pixels × 64 pixels. Second, the proposed method improved the classification accuracy of mixed objects substantially. Moreover, with an increased level in the mixed degree (decreases in the PCP), the BOBIC method demonstrated a more obvious advantage. Finally, the proposed method also exhibited more superior performance when classifying pure objects, in particular with respect to a segmentation scale of 50.

Table 5 Number of segmentation objects under different situations.

Situation	Study sites	Count (32×32)	Count (64×64)	Study sites	Count (32×32)	Count (64×64)	Study sites	Count (32×32)	Count (64×64)
					SSP (50)				
1	1	743	2087	2	621	1503	3	1700	3925
2		113	13		80	3		432	43
3		3938	2696		2261	1456		8060	6224
4		0	12		0	3		0	1
5		56	42		15	12		24	23
Total		4850	4850		2977	2977		10,216	10,216
					SSP (110)				
1	1	23	159	2	79	210	3	248	717
2		134	30		91	9		410	109
3		992	959		657	608		2059	1891
4		0	0		0	0		0	0
5		20	21		7	7		8	8
Total		1169	1169		834	834		2725	2725

Table 6 Classification error rates of segmentation objects under different situations.

Situation	Study sites	Error rate (32×32) (%)	Error rate (64×64) (%)	Study sites	Error rate (32×32) (%)	Error rate (64×64) (%)	Study sites	Error rate (32×32) (%)	Error rate (64×64) (%)
					SSP (50)				
1	1	4.98	4.22	2	0.00	0.00	3	4.41	2.68
2		2.65	0.00		1.25	0.00		4.86	0.00
3		5.26	3.93		4.64	3.50		5.42	2.84
4		—	50.00		—	66.67		—	100.00
5		37.50	23.81		60.00	41.67		45.83	34.78
Total		5.53	4.33		3.86	1.95		5.32	2.85
					SSP (110)				
1	1	8.70	5.66	2	3.80	3.33	3	5.65	2.79
2		8.96	0.00		3.30	0.00		8.05	5.50
3		7.06	5.94		4.87	2.63		6.31	4.81
4		—	—		—	—		—	—
5		60.00	23.81		42.86	28.57		87.50	12.50
Total		8.21	6.07		4.92	3.00		6.75	4.33

Note: "—" denotes that segmentation objects do not exist under the current situation.

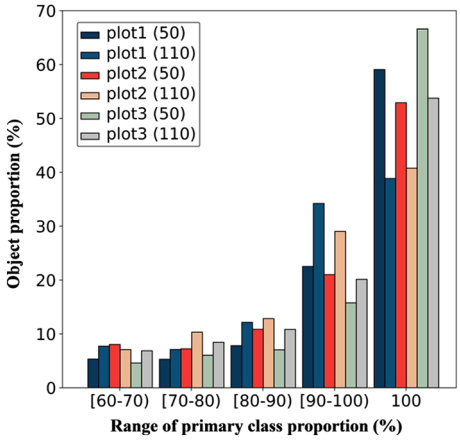

Fig. 7 The ratios of the segmentation object quantity for the different intervals of PCPs in the three study sites to the total quantity of segmentation objects. (The PCP represents the ratio of the area of the primary class in a segmentation object to the total area of the segmentation object.)

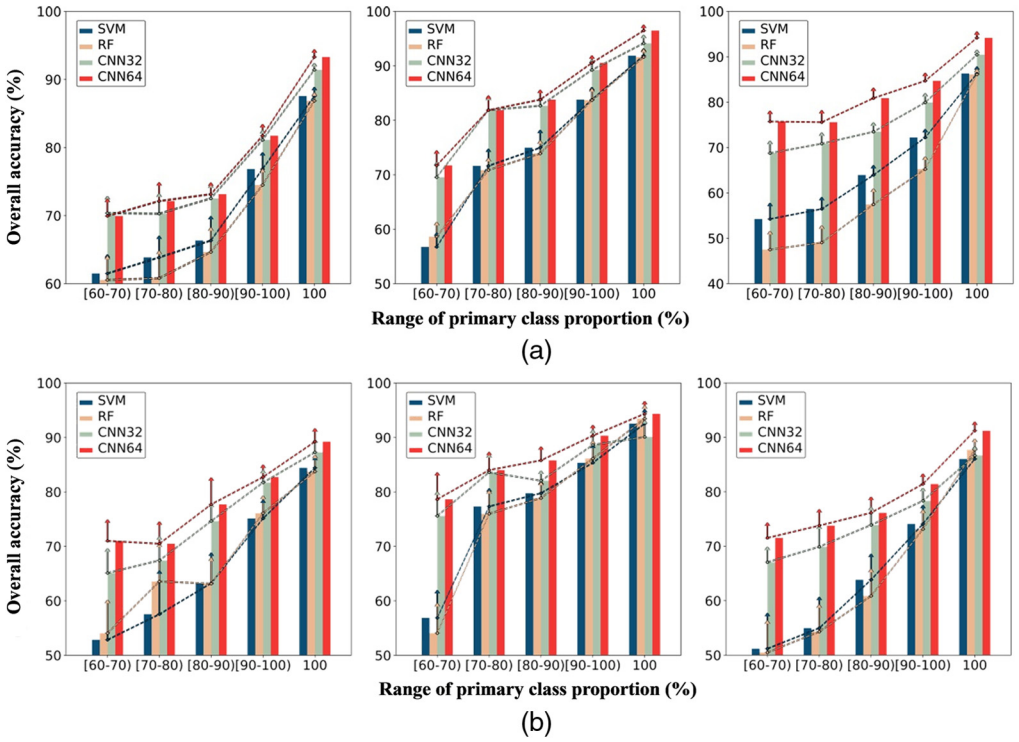

Fig. 8 Under a sampling proportion of 10%, the combo line and column chart for the classification accuracies of different intervals of the PCP in the three study sites. (a) and (b) The results when the segmentation scale was 50 and 110, respectively. (In the subfigures, the *Y*-coordinates of the top end of each bar and node represent the mean value of 20 classification accuracies, the error bar on the node denotes the standard deviation of 20 classification accuracies, and the PCP represents the ratio of the area of the primary class in a segmentation object to the total area of the segmentation object.)

5 Discussion

The proposed BOBIC method exhibited better classification accuracy than the conventional OBIC method in the three study areas, and the results confirmed the feasibility of using the proposed method for land-use type classifications. We also found that the geometric relationship of image blocks to segmented objects was important for the proposed BOBIC method. This was

because, in terms of remote sensing images, segmented objects of different land cover types would exhibit varying features. For example, the single area of vehicular segmented objects was normally small, whereas segmentation objects of rural roads were generally strip-shaped. Irregular shapes of segmented objects resulted in situations where the image block of a fixed-sized often encompassed only a portion of the segmented object, or even was enclosed by the segmented object. In our experimental results, the numbers of situations 1, 2, and 3 accounted for the vast majority of the total number of segmented objects. Moreover, situations 2 and 3 did not exhibit a higher error rate than situation 1, which demonstrates that the classification accuracy of CNN would not be affected by the situation where the image block only encompasses a portion of the segmentation object. This finding further confirms the feasibility of using the proposed BOBIC method.

Furthermore, another key point of the proposed BOBIC method was that it improved the classification effect of mixed objects, which can be attributed to the way that it generates samples, i.e., by generating image blocks using CGs of segmentation objects. First, the image block itself was a mixed object, which could substantially narrow the gap between mixed and pure segmentation objects. Second, owing to the fact that the CG was the center of object mass, the center point of the image block exhibited a tendency to fall on or approach the region of the primary class in the mixed object. Moreover, as the PCP became greater, this tendency became more pronounced. Certainly, only the CNN can overcome the fact that the complexity of VHR images can cause traditional human-dependent classification models to fail due to the limited representation power of handcrafted features,[34] thereby obtaining class information from complex image blocks. It can be concluded that the proposed BOBIC method was successful at applying the CNN to OBIC, which also proves the hypothesis of Guirado et al.[33] that stated that the inclusion of CNN-models could further improve OBIA methods.

Finally, we need to mention that there was a disadvantage in relation to the use of the proposed method in that the center point of an image block fell onto different types of land covers in a few rare cases (i.e., situations 4 and 5, and in particular, with respect to the road under situation 5, where its image block represented not a road but a building). As discussed in Sec. 4.2, the error rates of situations 4 and 5 were very high, but the probability of the occurrence of situations 4 and 5 remained extremely low. This was because only if the boundary line between two types of land covers exhibited a larger curvature, the CG of land cover on the outward side of the boundary line (in the direction opposite to the side where the curvature center was located) fell within the land cover on the inward side of the boundary line (on the side where the curvature center was positioned). Meanwhile, the CG of land cover on the inward side of the boundary line still fell onto the land cover of the same type. Even so, how to generate more appropriate image blocks for the segmented objects of situations 4 and 5 will be an important focus topic for us in the future.

6 Conclusions

In this work, a blocks-based OBIC (BOBIC) method was proposed for applying a CNN to OBIC. Compared with traditional classification methods, the proposed method utilizes the ability of CNN to automatically extract high-level features, thereby achieving end-to-end classification for irregular segmentation objects within the framework of OBIA. To evaluate the feasibility of the proposed BOBIC method, we systematically summarized the geometric relationships of segmented objects to image blocks and tested the method at three study sites using two segmentation scales and two types of image block sizes. Experimental results showed that the BOBIC method could substantially improve the OBIC classification effect and alleviate the effect derived from mixed objects. However, there was a drawback to the proposed method in that erroneous samples could be generated when the boundary line between two types of land covers exhibited a large curvature, which will be the focus topic of our future research. In summary, the proposed BOBIC exhibited an excellent classification effect compared with the OBIC. Moreover, this approach successfully reduced the uncertainty associated with OBIA during classification, which is mainly comprised of uncertainty during feature selection and that of mixed objects.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (No. 2017YFB0504205), the National Natural Science Foundation of China (No. 41701374), Natural Science Foundation of Jiangsu Province of China (No. BK20170640), China Postdoctoral Science Foundation (No. 2017T10034, 2016M600392), and the funding provided by the Alexander von Humboldt Foundation. We are also grateful to anonymous reviewers and members of the editorial team for advice.

References

1. K. Stankov et al., "Detection of buildings in multispectral very high spatial resolution images using the percentage occupancy hit-or-miss transform," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**, 4069–4080 (2014).
2. Y. Lin et al., "Rotation-invariant object detection in remote sensing images based on radial-gradient angle," *Remote Sens. Lett.* **12**, 746–750 (2015).
3. G. Liu et al., "Interactive geospatial object extraction in high resolution remote sensing images using shape-based global minimization active contour model," *Pattern Recognit. Lett.* **34**, 1186–1195 (2013b).
4. S. Leninisha et al., "Water flow based geometric active deformable model for road network," *ISPRS J. Photogramm. Remote Sens.* **102**, 140–147 (2015).
5. A. O. Ok, "Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts," *ISPRS J. Photogramm. Remote Sens.* **86**, 21–40 (2013).
6. A. O. Ok et al., "Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery," *IEEE Trans. Geosci. Remote Sens.* **51**, 1701–1717 (2013).
7. D. G. Goodin et al., "Mapping land cover and land use from object-based classification: an example from a complex agricultural landscape," *Int. J. Remote Sens.* **36**, 4702–4723 (2015).
8. X. Li et al., "Identification of forested landslides using LiDar data, object-based image analysis, and machine learning algorithms," *Remote Sens.* **7**, 9705–9726 (2015b).
9. D. Contreras et al., "Monitoring recovery after earthquakes through the integration of remote sensing, GIS, and ground observations: the case of L'Aquila (Italy)," *Cartogr. Geogr. Inf. Sci.* **43**, 115–133 (2016).
10. X. Yao et al., "A coarse-to-fine model for airport detection from remote sensing images using target-oriented visual saliency and CRF," *Neurocomputing* **164**, 162–172 (2015).
11. D. Zhang et al., "Weakly supervised learning for target detection in remote sensing images," *IEEE Geosci. Remote Sens. Lett.* **12**, 701–705 (2015a).
12. D. Arvor et al., "Advances in geographic object-based image analysis with ontologies: a review of main contributions and limitations from a remote sensing perspective," *ISPRS J. Photogramm. Remote Sens.* **82**, 125–137 (2013).
13. H. Costa et al., "Combining per-pixel and object-based classifications for mapping land cover over large areas," *Int. J. Remote Sens.* **35**, 738–753 (2014).
14. T. Blaschke et al., "Geographic object-based image analysis—towards a new paradigm," *ISPRS J. Photogramm. Remote Sens.* **87**, 180–191 (2014).
15. U. C. Benz et al., "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information," *ISPRS J. Photogramm. Remote Sens.* **58**, 239–258 (2004).
16. B. Johnson et al., "Unsupervised image segmentation evaluation and refinement using a multi-scale approach," *ISPRS J. Photogramm. Remote Sens.* **66**, 473–483 (2011).
17. J. Weston et al., "Feature selection for SVMS," in *Proc. of the 13th Int. Conf. on Neural Information Processing Systems*, pp. 668–674 (2000).
18. I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.* **3**, 1157–1182 (2003).
19. F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.* **42**, 1778–1790 (2004).

20. M. Pal and P. Mather, "Some issues in the classification of dais hyperspectral data," *Int. J. Remote Sens.* **27**, 2895–2916 (2006).
21. D. C. Duro et al., "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery," *Remote Sens. Environ.* **118**, 259–272 (2012).
22. A. Puissant et al., "Object-oriented mapping of urban trees using random forest classifiers," *Int. J. Appl. Earth Obs.* **26**, 235–245 (2014).
23. L. Ma et al., "Training set size, scale, and features in geographic object-based image analysis of very high resolution unmanned aerial vehicle imagery," *ISPRS J. Photogramm. Remote Sens.* **102**, 14–27 (2015).
24. G. E. Hinton et al., "A fast learning algorithm for deep belief nets," *Neural Comput.* **18**, 1527–1554 (2006).
25. A. Krizhevsky et al., "ImageNet classification with deep convolutional neural networks," in *26th Annual Conf. of Neural Information Processing Systems*, Lake Tahoe, Nevada (2012).
26. O. A. Penatti et al., "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *IEEE Conf. of Computer Vision and Pattern Recognition Workshops*, Boston, Massachusetts (2015).
27. Y. LeCun et al., "Backpropagation applied to handwritten zip code recognition," *Neural Comput.* **1**, 541–551 (1989).
28. Y. LeCun et al., "Gradient-based learning applied to document recognition," *Proc. IEEE* **86**, 2278–2324 (1998).
29. A. M. Cheriyadat, "Unsupervised feature learning for aerial scene classification," *IEEE Trans. Geosci. Remote Sens.* **52**, 439–451 (2014).
30. D. Ciresan et al., "In deep neural networks segment neuronal membranes in electron microscopy images," in *Proc. of the 25th Int. Conf. on Neural Information Processing Systems*, Vol. 2, pp. 2843–2851 (2012).
31. Y. Jia et al., "Caffe: convolutional architecture for fast feature embedding," in *ACM Int. Conf. on Multimedia*, Orlando, Florida (2014).
32. H. Lee et al., "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *26th Annual Int. Conf. on Machine Learning*, Montreal, Canada (2009).
33. E. Guirado et al., "Deep-learning versus OBIA for scattered shrub detection with Google earth imagery: Ziziphus Lotus as case study," *Remote Sens.* **9**(12), 1220 (2017).
34. W. Zhao et al., "Object-based convolutional neural network for high-resolution imagery classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **99**, 3386–3396 (2017).
35. T. Liu et al., "Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system," *GIsci. Remote Sens.* **55**, 243–264 (2018).
36. M. Baatz and A. Schäpe, "Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation," *Angew. Geogr. Informationsverarb.* **12**, 12–23 (2000).
37. L. Ma et al., "Using unmanned aerial vehicle for remote sensing application," in *21st Int. Conf. of Geoinformatics*, Kaifeng, China, pp. 20–23 (2013).
38. ISPRS, "ISPRS 2D Semantic Labeling–Vaihingen data," 2013 http://www2.isprs.org/commissions/comm3/wg4/2d-sem-label-vaihingen.html.
39. L. Drăguţ et al., "Automated parameterisation for multi-scale image segmentation on multiple layers," *ISPRS J. Photogramm. Remote Sens.* **88**, 119–127 (2014).
40. J. P. Ardil et al., "Context-sensitive extraction of tree crown objects in urban areas using VHR satellite images," *Int. J. Appl. Earth Obs. Geoinf.* **15**, 57–69 (2012).
41. L. Dragut et al., "ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data," *Int. J. Geogr. Inf. Sci.* **24**, 859–871 (2010).
42. C. Witharana and D. L. Civco, "Optimizing multi-resolution segmentation scale using empirical methods: exploring the sensitivity of a supervised discrepancy measure," *ISPRS J. Photogramm. Remote Sens.* **87**, 108–121 (2014).

43. T. G. Whiteside et al., "Area-based and location-based validation of classified image objects," *Int. J. Appl. Earth Obs.* **28**, 117–130 (2014).
44. M. Kim et al., "Multi-scale Geo-obia with very high spatial resolution digital aerial imagery: scale, texture and image objects," *Int. J. Remote Sens.* **32**, 2825–2850 (2011).
45. D. Liu and F. Xia, "Assessing object-based classification: advantages and limitations," *Remote Sens. Lett.* **1**, 187–194 (2010).
46. G. M. Espindola et al., "Parameter selection for region—growing image segmentation algorithms using spatial autocorrelation," *Int. J. Remote Sens.* **27**, 3035–3040 (2006).
47. H. Zhang et al., "Image segmentation evaluation: a survey of unsupervised methods," *Comput. Vision Image Understanding* **110**, 260–280 (2008).
48. K. Verbeeck et al., "External geo-information in the segmentation of VHR imagery improves the detection of imperviousness in urban neighborhoods," *Int. J. Appl. Earth Obs.* **18**, 428–435 (2012).
49. R. M. Haralick, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **SMC-3**, 610–621 (1973).
50. R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Comput. Vision Graphics Image Process.* **29**, 100–132 (1985).
51. J. S. Weszka, C. R. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification," *IEEE Trans. Syst. Man Cybern.* **SMC-6**, 269–285 (1976).
52. M. Li et al., "A systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments," *Int. J. Appl. Earth Obs.* **49**, 87–98 (2016).
53. M. Pal et al., "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. Geosci. Remote Sens.* **48**, 2297–2307 (2010).
54. T. Liu et al., "A novel transferable individual tree crown delineation model based on fishing net dragging and boundary classification," *ISPRS J. Photogramm. Remote Sens.* **110**, 34–47 (2015).
55. M. A. Ahmed et al., "Spatially-explicit modeling of multi-scale drivers of aboveground forest biomass and water yield in watersheds of the Southeastern United States," *J. Environ. Manage.* **199**, 158–171 (2017).
56. S. Lee et al., "Detection of deterministic and probabilistic convection initiation using Himawari-8 advanced Himawari imager data," *Atmos. Meas. Tech.* **10**, 1859–1874 (2017).
57. J. Im et al., "Downscaling of AMSR-E soil moisture with MODIS products using machine learning approach," *Environ. Earth Sci.* **75**, 1120–1139 (2016).
58. C. C. Chang and C. J. Lin, "Libsvm: a library for support vector machines," *ACM Trans. Intell. Syst. Technol.* **2**, 1–27 (2011).
59. C. Hsu et al., *A Practical Guide to Support Vector Classification*, Department of Computer Science, National Taiwan University, Taipei, Taiwan (2010).
60. V. F. Rodriguez-Galiano et al., "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS J. Photogramm. Remote Sens.* **67**, 93–104 (2012).
61. K. He et al., "Deep residual learning for image recognition," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 770–778 (2016).
62. D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Phys.* **160**, 106–154 (1962).
63. Y. LeCun et al., "Convolutional networks and applications in vision," in *IEEE Int. Symp. on Circuits and Systems (ISCAS)*, Paris, France, pp. 253–256 (2010).
64. D. Scherer et al., *Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition*, Springer, Berlin/Heidelberg, Germany (2010).
65. T. Serre et al., "On the role of object-specific features for real world object recognition in biological vision," *Lect. Notes Comput. Sci.* **2525**, 387–397 (2002).
66. C. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York (2006).
67. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Int. Conf. of Learning Representations*, San Diego, California (2015).

68. K. He et al., "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *IEEE Int. Conf. on Computer Vision*, Boston, Massachusetts, pp. 1026–1034 (2015).
69. B. Xu et al., "Empirical evaluation of rectified activations in convolutional network," *Comput. Sci.* **5**, 12 (2015).
70. I. V. Tetko et al., "Neural network studies. 1. Comparison of overfitting and overtraining," *J. Chem. Inf. Comput. Sci.* **35**, 826–833 (1995).
71. N. Srivastava et al., "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
72. B. L. Welch, "The generalisation of student's problems when several different population variances are involved," *Biometrika* **34**, 28–35 (1947).

Biographies for the authors are not available.