An Urban Land Cover Classification Method Based on Segments' Multidimension Feature Fusion

Zhongyi Huang, Jiehai Cheng, Guoqing Wei, Xiang Hua, and Yuyao Wang

***Abstract*—Using object-based deep learning for the urban land cover classification has become a mainstream method. This study proposed an urban land cover classification method based on segments' object features, deep features, and spatial association features. The proposed method used the synthetic semivariance function to determine the hyperparameters of the superpixel segmentation and subsequently optimized the image superpixel segmentation result. A convolutional neural network and a graph convolutional neural network were used to obtain segments' deep features and spatial association features, respectively. The random forest algorithm was used to classify segments based on the multidimension features. The results showed that the image superpixel segmentation results had the significant impact on the classification results. Compared with the pixel-based method, the segment-based methods generally yielded the higher classification accuracy. The strategy of multidimension feature fusion can combine the advantages of each single-dimension feature to improve the classification accuracy. The proposed method utilizing multidimension features was more efficient than traditional methods used for the urban land cover classification. The fusion of segments' object features, deep features, and spatial association features was the best solution for achieving the urban land cover classification.**

***Index Terms*—Graph convolutional neural network (GCN), high spatial resolution remote sensing image, land cover classification, segments' multidimension features, superpixel.**

I. Introduction

Urban land cover plays an important role in urban resource management, urban planning, and environmental protection [1], [2], [3]. In recent years, with the abundance of high spatial resolution remote sensing images, it has become a common practice to generate urban land cover data by using high spatial resolution remote sensing images [4], [5], [6].

High spatial resolution remote sensing images have high spatial detail and the characteristics of increase of the intraclass variation and decrease of the interclass variation [7], [8]. As a result, the traditional pixel-based image classification techniques always produce salt and pepper effect and have low classification accuracy [9]. Object-based image analysis methods (OBIA) can effectively overcome the above shortcomings and greatly improve the urban land cover classification accuracy [10], [11]. OBIA takes segments as the basic units [12], [13], [14] and extracts segments' object features to classify the segments [4], [15], [16]. However, in a complex urban scene, only the segments' object features are used for classification, which may lead to misclassification. This is because the segments' object features are essentially the statistical information of pixels within the segments, which belong to the shallow features of the segments [17], [18]. In addition, the spatial association information between segments are not utilized [19].

Segments' deep features were proposed for remote sensing image classification in many studies. Fu et al. [20] set the segments' centroids as the centers of the image patches and used CNN to extract the deep features of each segment and classify them. Compared to the approach of utilizing segments' object features, this approach significantly improved the image classification accuracy. Zhang et al. [16] proposed to decompose each segment into multiple image patches based on the moment bounding box, and the deep features of the segment were jointly expressed by the deep features of these image patches. Similar approaches were proposed by Lv et al. [21] and Martins et al. [22], which further considered the influence of the image patch size on classification results. In addition, some studies fused segments' object features with deep features. Zhao et al. [23] fused segments' object features with deep features extracted by CNN and obtained better classification results compared to using only object features or deep features. Zhang et al. [24] fused superpixels' deep features with object features of segments obtained by multiresolution segmentation and showed that fusing multidimension features can obtain better classification results.

Tobler's [25] first law of geography states that everything is related to everything else, and that things that are close together have a greater degree of relatedness. In urban scenes, the land cover in remote sensing images shows more obvious spatial association. For example, buildings are connected to buildings or adjacent to roads; vegetation is located in the center or on both sides of the road. Fully and reasonably utilizing spatial association information can improve the performance of classifiers [26]. Recently, graph convolutional neural networks (GCNs) in the deep learning field have been applied to urban land cover classification work [27], [28]. Researchers applied GCN to obtain spatial association information between segments and neighboring segments [26], [29]. Zhang et al. [19] used segments as graph nodes and the adjacency relationship between

Manuscript received 25 November 2023; revised 17 January 2024 and 2 February 2024; accepted 9 February 2024. Date of publication 23 February 2024; date of current version 6 March 2024. This work was supported in part by the National Science Foundation of China under Grant 41671514 and in part by the Science and Technology Research Project of Henan Province under Grant 212102311149 and Grant 232102320244. *(Corresponding author: Jiehai Cheng.)*

The authors are with the School of Surveying & Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China (e-mail: 1329890640@qq.com; chengjiehai@hpu.edu.cn; 3168222317@qq.com; 1369927703@qq.com; 1056898277@qq.com).

Digital Object Identifier 10.1109/JSTARS.2024.3367626

segments as edges to construct GCN for image classification, and proved that using spatial association information for image classification was a feasible approach. Liu et al. [30] combined CNN and GCN while considering deep features and spatial association features to improve image classification accuracy.

Although multidimension feature fusion is considered suitable for improving the urban land cover classification accuracy, few studies have fully considered segments' object features, deep features, and spatial association features. In addition, segmentation results can directly affect urban land cover classification results. However, the segmentation results in most studies were determined by manually setting segmentation parameters or exhaustive enumeration, which are subjective and uncertain. To address these problems, this study proposes an urban land cover classification method based on segments' multidimension feature fusion. The main contributions of this study include the following.

1) Developing a novel urban land cover classification method that can be easily integrated into an object-based deep learning paradigm and takes advantage of multidimension features that can be extracted from segments.
2) Determining if the proposed method utilizing multidimension features is more efficient than traditional methods used for the urban land cover classification.
3) Exploring the significant impact of image superpixel segmentation results on classification results.

The rest of this study is organized as follows. Section II illustrates the principle of the proposed method. The experimental results and analysis is shown in Section III. Section IV discusses the experimental results. Finally, Section V concludes this article.

II. Method

The flowchart of the proposed method is illustrated in Fig. 1. Zero parameter version of simple linear iterative cluster (SLICO) algorithm [31], [32] is adopted for the image superpixel segmentation. The final image segmentation result is obtained through automatic optimization of segmentation parameters. Taking each segment on the segmentation result as the basic unit, the object features, deep features, and spatial association features of the segment are extracted. Segments are classified to obtain the urban land cover classification result by fusing the above three type features.

A. Image Segmentation

The SLICO algorithm clusters similar pixels based on their spectral and distance information to generate the segmentation result, where the shape and size of the segments are relatively uniform

$$S = \sqrt{N/K} \tag{1}$$

where S represents the expected superpixel size; K represents the number of presegmented superpixels; and N represents the number of pixels in the image.

As shown in Formula (1), the hyperparameter K need to be set. In existing studies, K was usually set manually. If S is determined, K can be estimated. S is related to the image land cover variation characteristics. The synthetic semivariance function [33] is used to set S. The synthetic semivariance in Formula (2) is the mean of the horizontal and vertical semivariances

$$\Upsilon_s\ (l_i) = \frac{[\Upsilon_h\,(l_i) + \Upsilon_v\,(l_i)]}{2} \tag{2}$$

$$\Delta\Upsilon_s\ (l_i) = \Upsilon_s\ (l_i) - \Upsilon_s\,(l_{i-1}) \tag{3}$$

where $\Upsilon_h(l_i)$,$\Upsilon_v(l_i)$ and $\Upsilon_s(l_i)$ represent the horizontal semivariance, vertical semivariance, and synthetic semivariance, respectively, when the lag is l_i; $\Delta\Upsilon_s(l_i)$ represents the increment in the synthetic semivariance when the lag changes from l_i to l_{i-1}

$$m = 2 * l_s + 1 \tag{4}$$

where l_s is the lag when $\Delta\Upsilon_s(l_i)$ in Formula (3) is 0 or less than 0 for the first time.

S is set by the value of m in Formula (4).

B. Segments' Object Features

In this study, the segments' object features were obtained from three aspects: spectral features, spatial features, and texture features in Table I.

C. Segments' Deep Features

The CNN model was used to extract segments' deep features shown in Fig. 2. The model included four convolution layers, two pooling layers, one flattening layer, and two fully connected layers (including 1024 and 512 nodes, respectively). The specific parameters of the model were shown in Table II. The image patch size was set by m in Formula (4). The output of the last fully connected layer was taken as the segments' deep features.

D. Segments' Spatial Association Features

The graph can be defined as Graph $= (V, E)$ where $V = \{v_1, v_2, \ldots, v_k | 1 \leq k \leq n\}$ represents nodes, $E = \{e_{ij} | 1 \leq i \leq n, 1 \leq j \leq n\}$ represents edges, i.e., the adjacency relationship between nodes, and n is the number of nodes. The segments were regarded as graph nodes. The segment adjacencies were regarded as edges. In this way, the GCN model was used to extract the segments' spatial association features (shown in Fig. 3).

The GCN model included an input layer, two hidden layers, and an output layer. The GCN input had two parts: the segment feature matrix X and the segment adjacency matrix A. In this study, the segments' object features and deep features were used to construct the feature matrix X. The adjacency matrix A was constructed in the following way: 1) Creating a zero matrix of $n \times n$ (n is the number of segments); and 2) completing the adjacency analysis between the segments. If two segments were adjacent, the corresponding position in the zero matrix was assigned to 1.

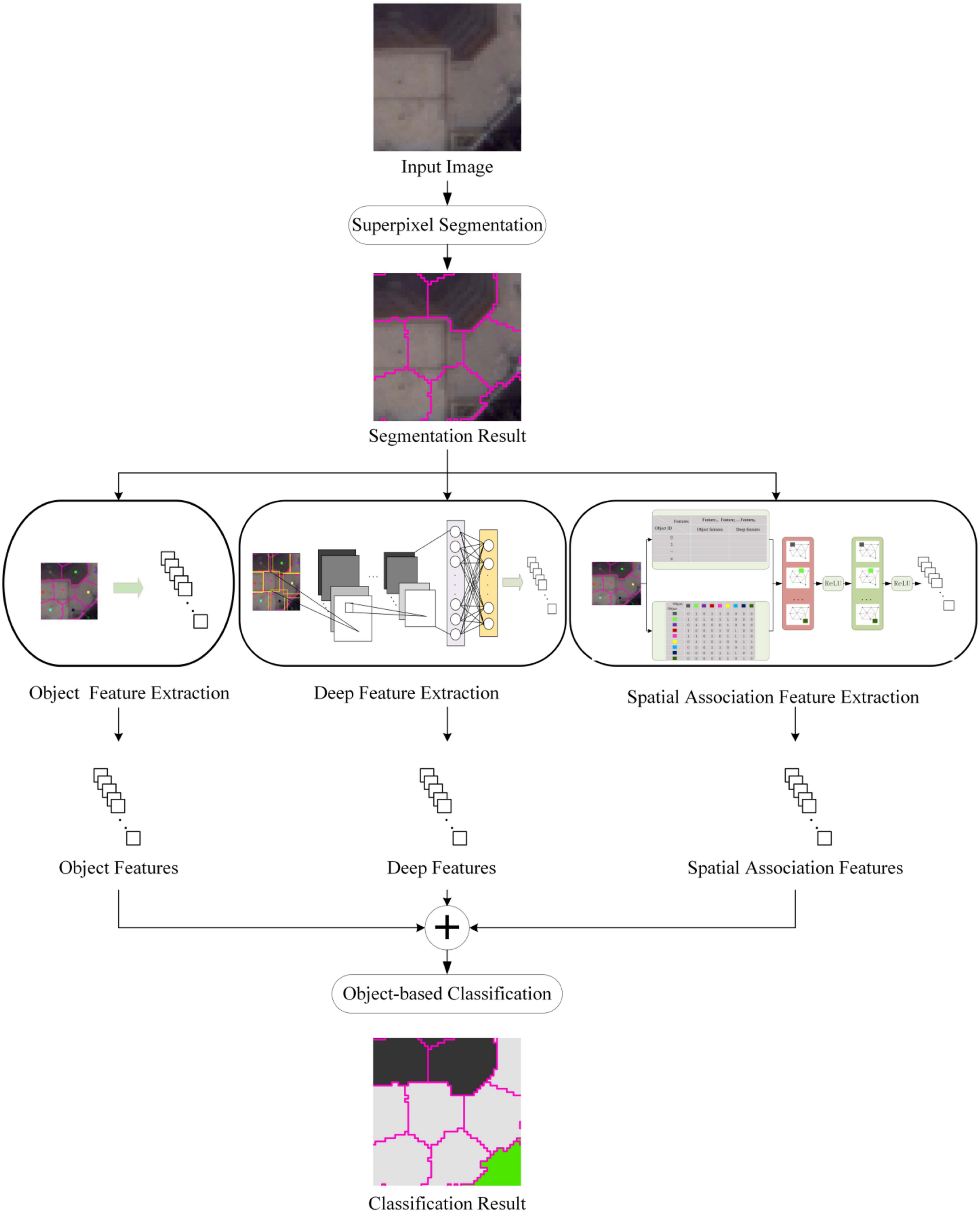


Fig. 1. Flowchart.

TABLE I
SEGMENTS' OBJECT FEATURES

Feature type	Name
Spectral features	NDVI, Mean, Standard deviation, Skewness, Brightness
Spatial features	Area, Border length, Length, Asymmetry, Border index, Compactness, Main direction, Roundness, Shape index
Texture features	GLCM (gray-level co-occurrence matrix) homogeneity, GLCM contrast, GLCM dissimilarity, GLCM entropy, GLCM ang.2nd moment, GLCM mean, GLCM standard deviation, GLCM correlation (0°, 45°, 90°, 135°)

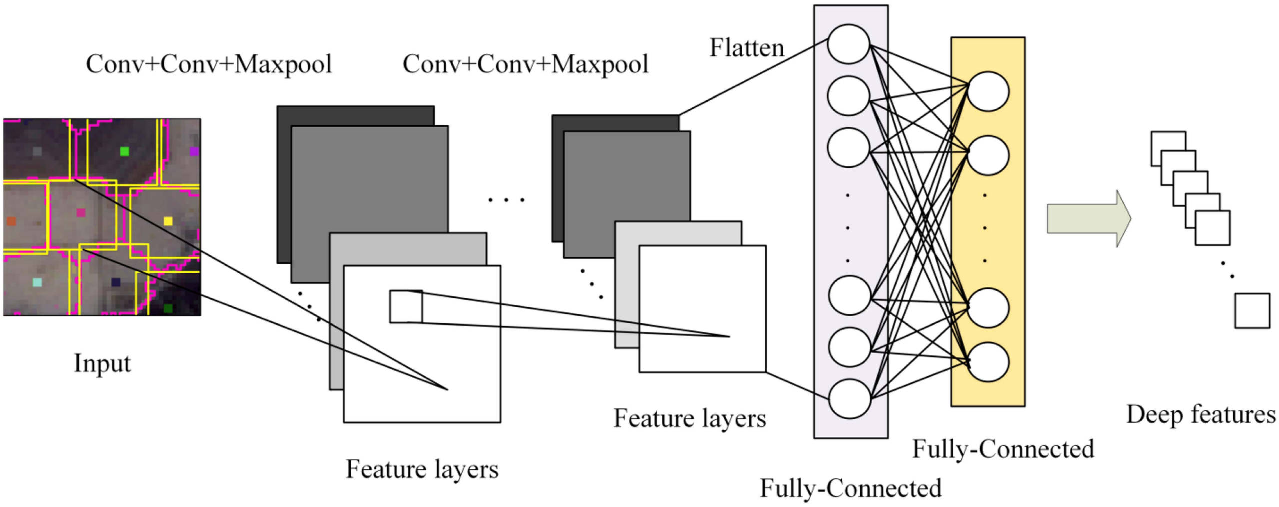


Fig. 2. Deep feature extraction model.

TABLE II
ARCHITECTURE OF THE DEEP FEATURE EXTRACTION MODEL

Layer	Convolution kernel	Steps	Activation function
Convolution	5×5, 64	1	ReLU
Convolution	3×3, 32	1	ReLU
Max pooling	2×2, -	2	-
Convolution	5×5, 64	1	ReLU
Convolution	3×3, 32	1	ReLU
Max pooling	2×2, -	2	-
Flatten	-	-	-
Fully Connected	-	-	ReLU
Fully Connected	-	-	ReLU

In the hidden layers, the GCN aggregates node features and neighborhood node features through convolution operations to generate new node features containing spatial association information. The output of the lth hidden layer in the GCN can be represented by Formula (5). In the literature [34], it was found that the number of hidden layers in the GCN model should not be set too high; usually, 2–3 layers are sufficient. Therefore, two hidden layers are set in this study

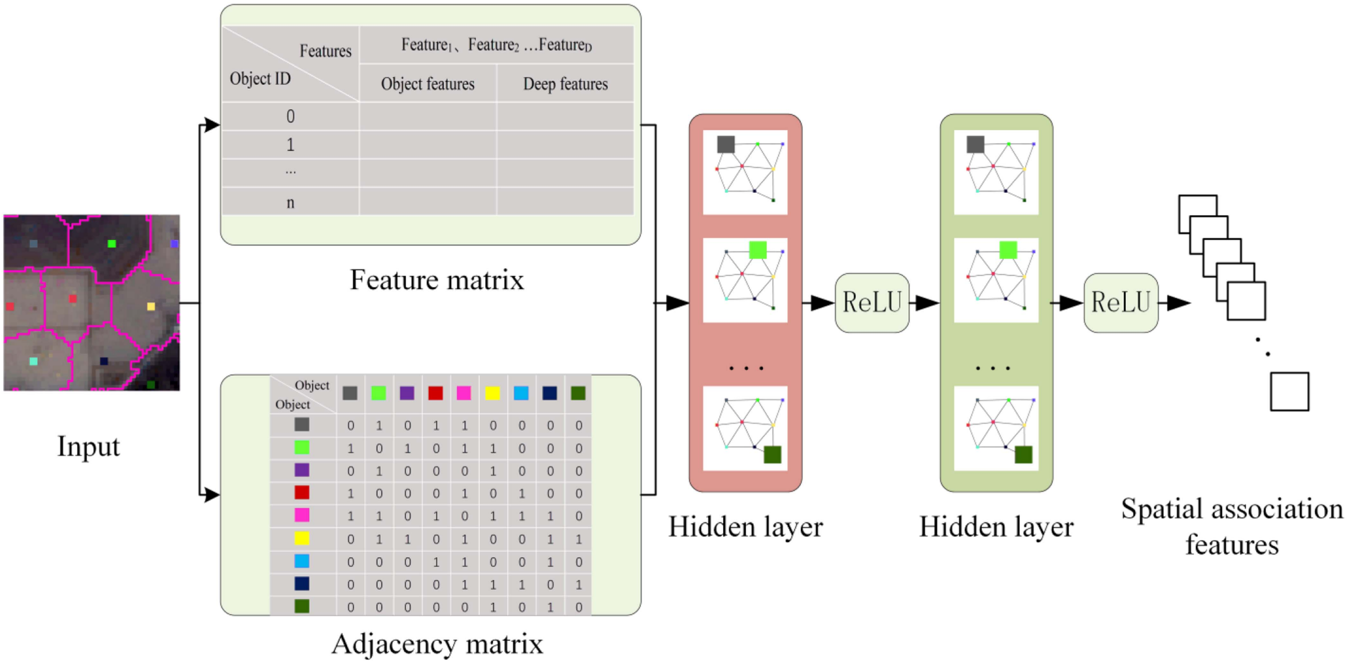
$$f_s^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}} f_s^{(l)} W^l\right) \tag{5}$$

where $f_s^{(l)}$ represents the feature matrix of the nodes in the lth hidden layer, $f_s^{(0)} = X$; $\tilde{A}$ represents the adjacency matrix with self-rings, i.e., $\tilde{A} = A + I$, I is the identity matrix; $\tilde{D}$ represents the degree matrix of $\tilde{A}$; W^l represents the matrix of trainable parameters in the lth hidden layer; and $\sigma(\cdot)$ represents the nonlinear activation function.

The output of the last hidden layer in the GCN model was taken as the segments' spatial association features.

E. Urban Land Cover Classification

1) Feature Fusion: If the image is segmented, N segments are generated. The segments' object features are denoted as $f_o \in R^{N\times D_o}$. The segments' deep features are denoted as $f_d \in R^{N\times D_d}$. The segments' spatial association features are denoted



Object ID \ Features	Feature₁、Feature₂ ...Feature_D	
	Object features	Deep features
0		
1		
...		
n		

Object \ Object	■	■	■	■	■	■	■	■	■
■	0	1	0	1	1	0	0	0	0
■	1	0	1	0	1	1	0	0	0
■	0	1	0	0	0	1	0	0	0
■	1	0	0	0	1	0	1	0	0
■	1	1	0	1	0	1	1	1	0
■	0	1	1	0	1	0	0	1	1
■	0	0	0	1	1	0	0	1	0
■	0	0	0	0	1	1	1	0	1
■	0	0	0	0	0	1	0	1	0

Fig. 3. Spatial association feature extraction model.

as $f_s \in R^{N \times D_s}$. D_o, D_d, and D_s denote the dimensions of the object features, deep features, and spatial association features, respectively. The acquired object features, deep features, and spatial association features were fused by $F = \{f_o, f_d, f_s\} \in R^{N \times D}$ and $D = D_o + D_d + D_s$ to obtain the classification features of the segments.

2) Classifier Determination: Random forest, support vector machine, and K-nearest neighbor classifiers were used to classify urban land cover in existing studies. Ramezan et al. [35] compared the classification accuracy of several supervised classifiers. The study showed that the random forest classifier had fewer parameters, good stability, and high classification accuracy. In this study, the random forest classifier was selected for urban land cover classification.

III. Experiments and Results

A. Data

The data in this study were WorldView-3 high spatial resolution remote sensing images of Zhengzhou City, Henan Province, China. The panchromatic and multispectral bands (red, green, blue, and near-infrared bands) of the data were fused to generate a multispectral image with a spatial resolution of 0.5 m (see Fig. 4). The image size was 2728 × 2728, and the land cover types included building, vegetation, water, road, bare land, parking lots, and shadow. For the convenience of application, bare land and parking lot were combined into road.

B. Results

1) SLICO Superpixel Segmentation: The lag in the synthetic semivariance was set to [2 ,50] with a step of 1. The increment in the synthetic semivariance was calculated and plotted, as shown in Fig. 5. The increment in the synthetic semivariance appeared to be less than 0 for the first time at $l_i = 8$, i.e., l_s was 8. According to Formula (4), m was 17, i.e., the expected superpixel size S was 17. N was 2728 × 2728. According to Formula (1), the number of SLICO presegmented superpixels was 25 751.

After SLICO superpixel segmentation, the actual number of segments obtained was 25 181. The segmentation result was shown in Fig. 6. The image patches matched well with the corresponding segments in terms of morphology, which can ensure that the deep features of the segments were not lost to the greatest extent.

2) Multidimension Feature Extraction and Classification of Segments: According to Table I, 55 object features for each segment were calculated.

The deep features for each segment were extracted using the model shown in Fig. 2. The image patch size was set to 17 × 17. The learning rate was set to 0.0001. For each land cover category, 30% of the segments, except for water, were randomly selected as the training samples. Due to the small number of water segments, 60% of the segments were randomly selected as training samples. To ensure sample balance and maintain sample resolution, sample enhancement (random rotation) was performed on the water training samples. For each land cover category, 10% of the segments were randomly selected as the validation samples. After model training, all the segments were fed into the model, and the output of the last fully connected layer was used as the segments' deep features extracted by the model. The dimensions of the deep features were 25181 × 512.

In the spatial association feature extraction model, the feature matrix consisted of object features and deep features of segments

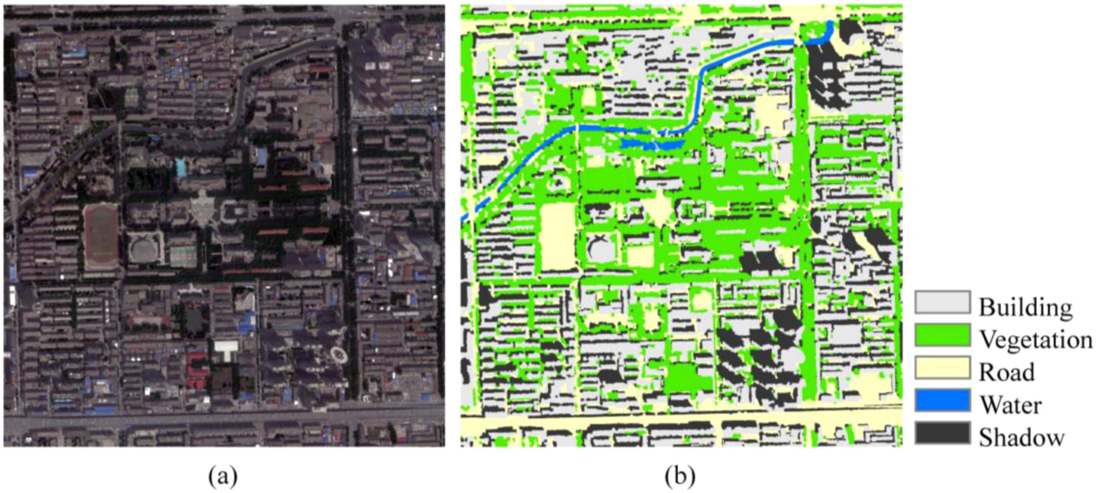


Fig. 4. Study area image and corresponding reference classification.

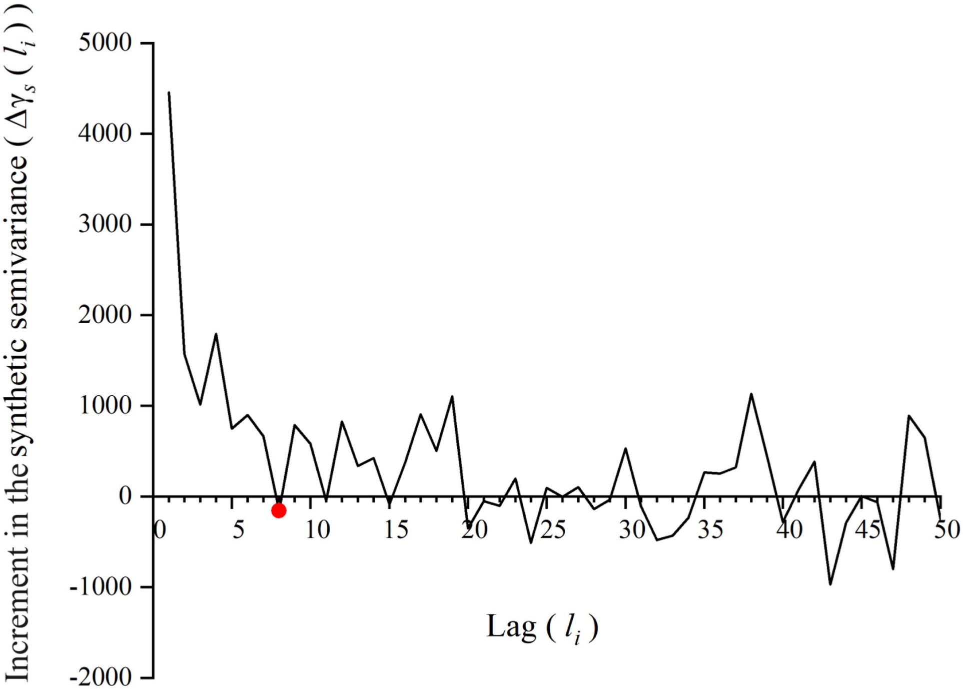


Fig. 5. Increment in the synthetic semivariance.

with dimensions of 25181 × 567. The number of nodes in the hidden layer of the model was set to 128 and the learning rate was set to 0.01. For each land cover category, 30% of the segments, except for water, were randomly selected as the training samples, 60% of the water segments were randomly selected as training samples. After model training, all the segments were fed into the model, and the output of the last hidden layer was used as the segments' spatial association feature extracted by the model. The dimensions of the spatial association features were 25181 × 128.

The object features, deep features, and spatial association features for each segment were fused to generate a total of 695 features. A random forest classifier was used to achieve urban land cover classification.

C. Comparison Analysis

To evaluate the performance of our method, on the basis of the image superpixel segmentation results, the corresponding urban land cover classification results were obtained based on the object features, deep features, spatial association features, object features + deep features, object features + spatial association features, deep features + spatial association features, and object

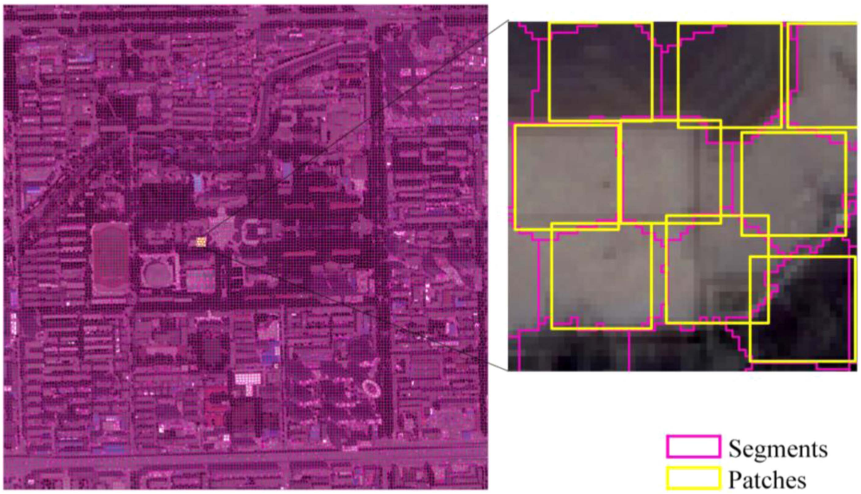


Fig. 6 Segmentation result.

TABLE III
CLASSIFICATION ACCURACY COMPARISON OF DIFFERENT METHODS

Classification method	F1-Score	OA (%)	Kappa	OA for each land cover category (%)				
				Building	Vegetation	Road	Water	Shadow
O_F	0.7843	79.59	0.7277	73.50	87.18	76.23	80.53	81.11
D_F	0.7887	80.09	0.7344	73.76	87.57	76.61	83.11	82.04
S_F	0.7661	78.45	0.7126	73.91	83.25	78.71	90.42	77.72
(O+D)_F	0.7947	80.50	0.7398	74.68	87.53	77.25	82.96	82.13
(O+S)_F	0.8174	82.18	0.7619	**79.32**	86.78	80.73	89.10	81.03
(D+S)_F	0.8178	82.31	0.7639	78.51	86.92	80.96	89.59	82.43
Pixel_CNN	0.7362	77.47	0.6987	73.37	**90.68**	61.11	**91.66**	78.78
OCNN	0.7758	79.11	0.7209	74.85	87.15	71.26	88.69	80.65
MULTI_F_{14}	0.8168	81.88	0.7578	78.48	87.46	80.28	82.78	80.67
MULTI_F_{34}	0.7114	72.49	0.6321	70.60	81.40	73.66	67.08	62.68
MULTI_F	**0.8183**	**82.54**	**0.7669**	78.59	87.56	**81.00**	89.81	**82.47**

The bold value in each column represents the maximum value in that column for the convenience of readers.

features + deep features + spatial association features of the segments, which were noted as O_ F, D_ F, S_ F, (O+D)_ F, (O+S)_ F, (D+S)_ F, and MULTI_ F, respectively. In addition, the following methods were selected as comparison methods: 1) Pixel_CNN: The method took a pixel as the basic unit, and set the image patches with the fixed size centered on pixels. The land cover classification was carried out according to the pixel deep features extracted from the image patches. 2) OCNN [23]: The method fused the pixel-level deep features extracted by the CNN with the object features, and classified land cover by majority voting by counting the pixel categories within each segment.

Considering the impact of different superpixel segmentation results on the classification results, S in Formula (1) was also set to 14 and 34, and the superpixel segmentation was performed on the image according to the calculated K value. Based on the object features + deep features + spatial association features, the corresponding urban land cover classification results were obtained and denoted as MULTI_F_{14} and MULTI_F_{34}.

The classification results of different methods were shown in Fig. 7, and the classification accuracy comparison of different methods was shown in Table III. The description of the accuracy evaluation metrics was shown in the following:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$F1 = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 2 \tag{8}$$

$$F1 - \text{Score} = \frac{\sum_{a=1}^{n} F1\_a}{n} \tag{9}$$

$$\text{OA} = \frac{\sum_{i=1}^{n} X_{ii}}{N} \tag{10}$$

$$\text{Kappa} = \frac{N\sum_{a=1}^{n} X_{aa} - \sum_{a=1}^{n}\left(\sum_{i=1}^{n} X_{ia} \times \sum_{i=1}^{n} X_{ai}\right)}{N^2 - \sum_{a=1}^{n}\left(\sum_{i=1}^{n} X_{ia} \times \sum_{i=1}^{n} X_{ai}\right)} \tag{11}$$

where TP represents true positive; FP represents false positive; FN represents false negative; $F1\_a$ represents $F1$ of class a; n represents the number of classes; N represents the total number of the samples; X_{ai} represents the number of the samples whose actual and predicted classes are a and i, respectively.

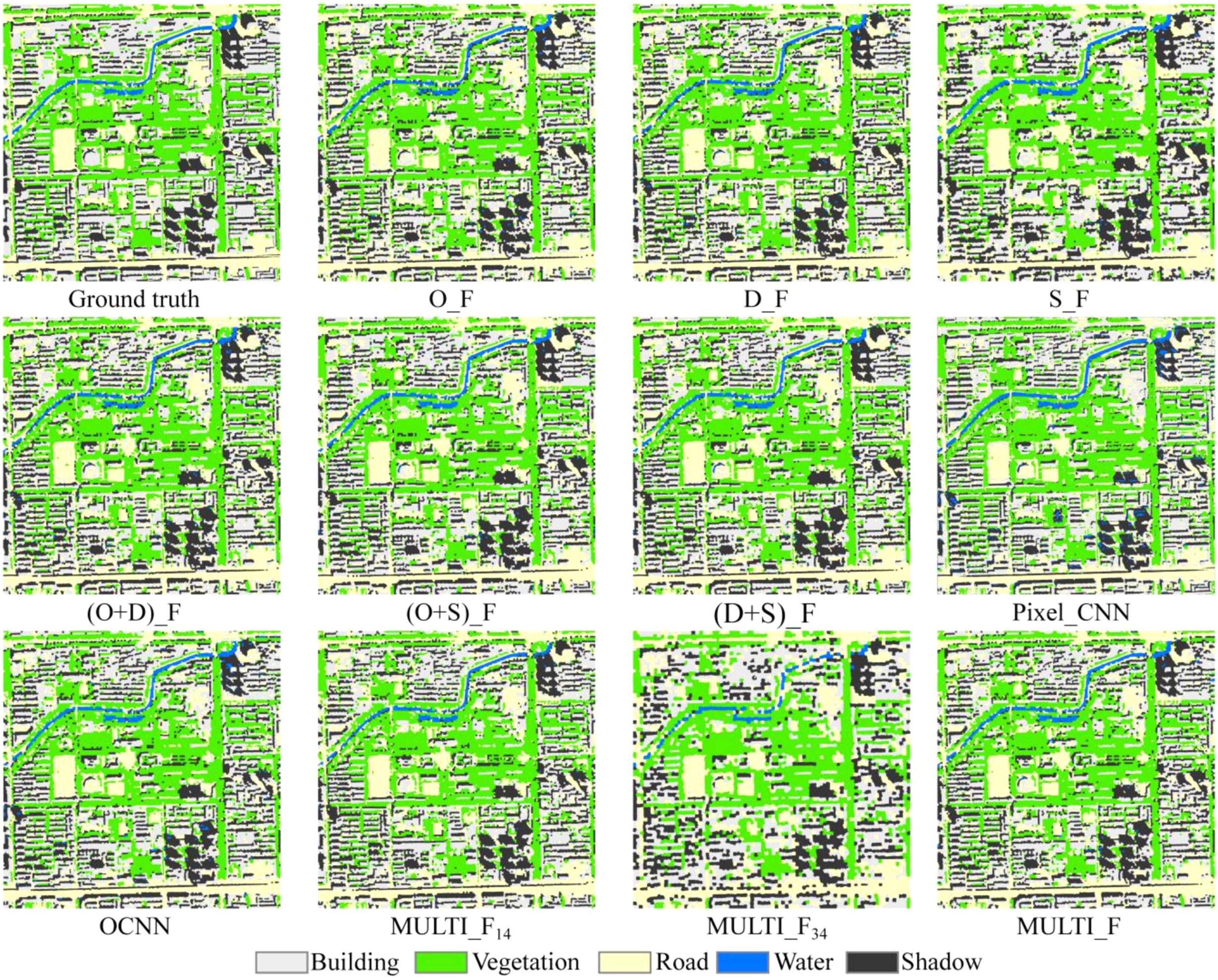


Fig. 7. Classification results of different methods.

In Table III, F1-Score, OA, and Kappa of the MULTI_F classification result in this study were greater than those of the Pixel_CNN and OCNN methods. In terms of the classification accuracy for each land cover category, road (OA is 81.00%) and shadow (OA is 82.47%) had the highest classification accuracy in the MULTI_F classification result.

IV. Discussion

The multidimension feature fusion was suitable for improving the urban land cover classification accuracy. As shown in Table III, in comparison with other methods, our method, which fused multidimension features of segments achieved the higher classification accuracy. This proved the effectiveness of the proposed method.

A. *Segment-Based Versus Pixel-Based Methods*

A comparison with Pixel_CNN was performed to compare the differences between segment-based and pixel-based urban land cover classification. Table III showed that the segment-based methods generally yielded the higher classification accuracy than the pixel-based method. Similar to pixel-based method, D_F used the CNN to obtain the deep features and classified the urban land cover based on the deep features. The classification accuracy of D_F (F1-Score 0.7887, OA 80.09%, and Kappa 0.7344) was significantly better than that of Pixel_CNN (F1-Score 0.7362, OA 77.47%, and Kappa 0.6987). Fig. 8 showed that there were many misclassified phenomena in the Pixel_CNN classification result, especially when roads were misclassified as buildings. In addition, there was a severe salt-and-pepper phenomenon in the Pixel_CNN classification result. This indicated that using segments as the analysis unit for classification is more effective.

Notably, the Pixel_CNN classification result had a higher classification accuracy for vegetation and water in Table III. Through analysis, it was found that there was some isolated vegetation in the image. The image had shadows around some water, and the two were similar in spectrum and shape. Pixel_CNN took pixels as the analysis unit and can correctly classify some vegetation and water in these areas with higher accuracy. During image segmentation, isolated vegetation and a small amount of water were segmented into the same segments that were mainly covered

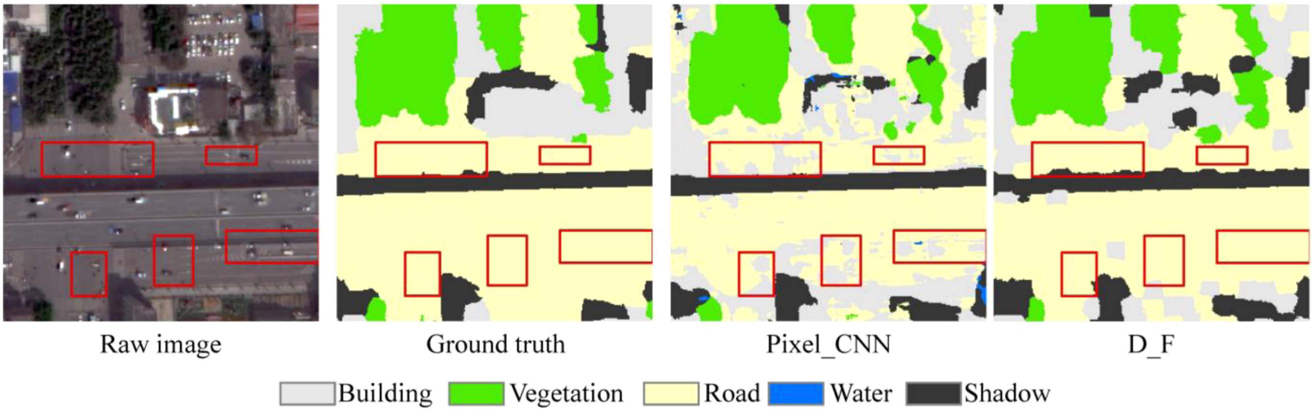


Fig. 8. Local classification results of Pixel_CNN and D_F.

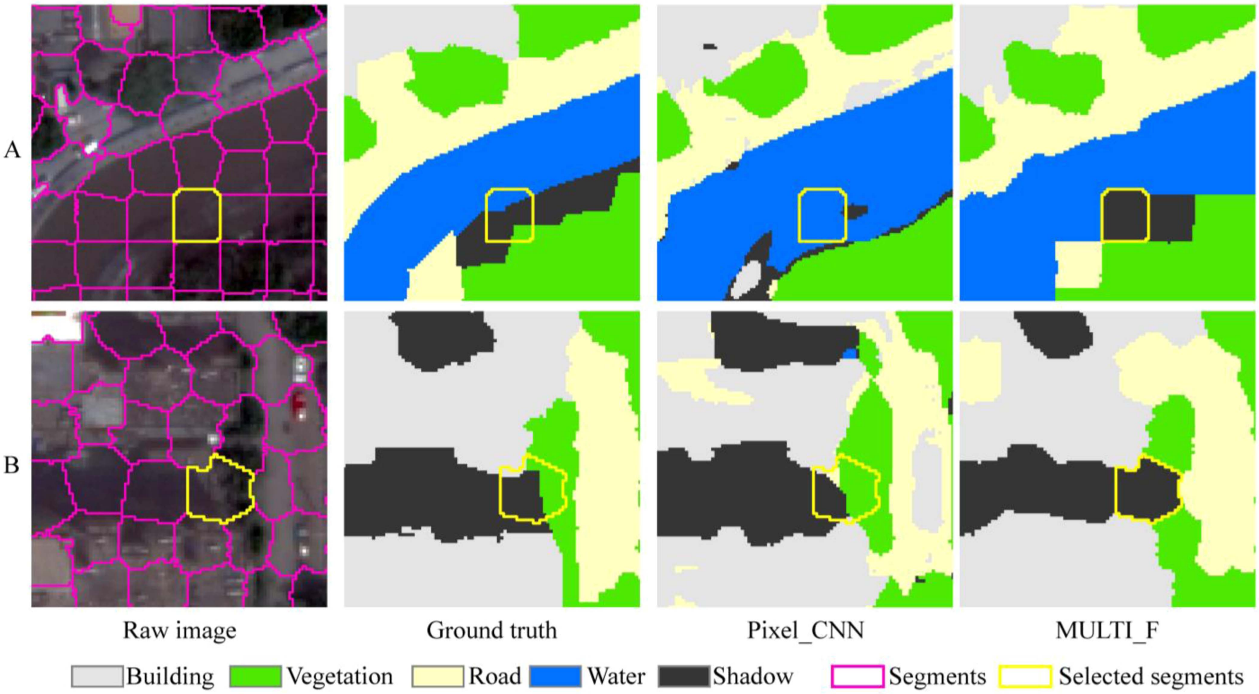


Fig. 9. Local classification results of Pixel_CNN and MULTI_F.

by other land covers, resulting in a small amount of vegetation and water being misclassified via the segment-based method, as shown in Fig. 9. In general, the pixel-based analysis method has certain advantages in some details. However, the overall classification accuracy of pixel-based methods is not high, and these methods are very time-consuming. In addition, the author believes that the segment-based classification accuracy will be further improved by optimizing the image segmentation result.

For the segment-based methods, the accuracy of the classification results based on different single-dimension features (i.e., O_F, D_F, and S_F) was different. In areas A and B of Fig. 10, D_F and S_F existed the misclassification phenomena in the red rectangular areas. The vegetations were misclassified as the shadows, roads or buildings, and the buildings were misclassified as the roads or vegetations. However, the classification results of O_F were consistent with the reference classification results. The raw image showed that the red rectangular area of A mainly belonged to vegetation, and the connections between vegetation categories were not tight. The buildings in the red rectangular area of B were surrounded by a large amount of vegetation and can be easily ignored. The classification results of O_F in the red rectangular areas of A and B demonstrated the advantages of object features and proved the necessity of object features in urban land cover classification.

As shown in areas C and D of Fig. 10, O_F and D_F existed the misclassification phenomena in the red rectangular areas. The roads were misclassified as buildings, and the buildings were misclassified as the roads. However, the classification results

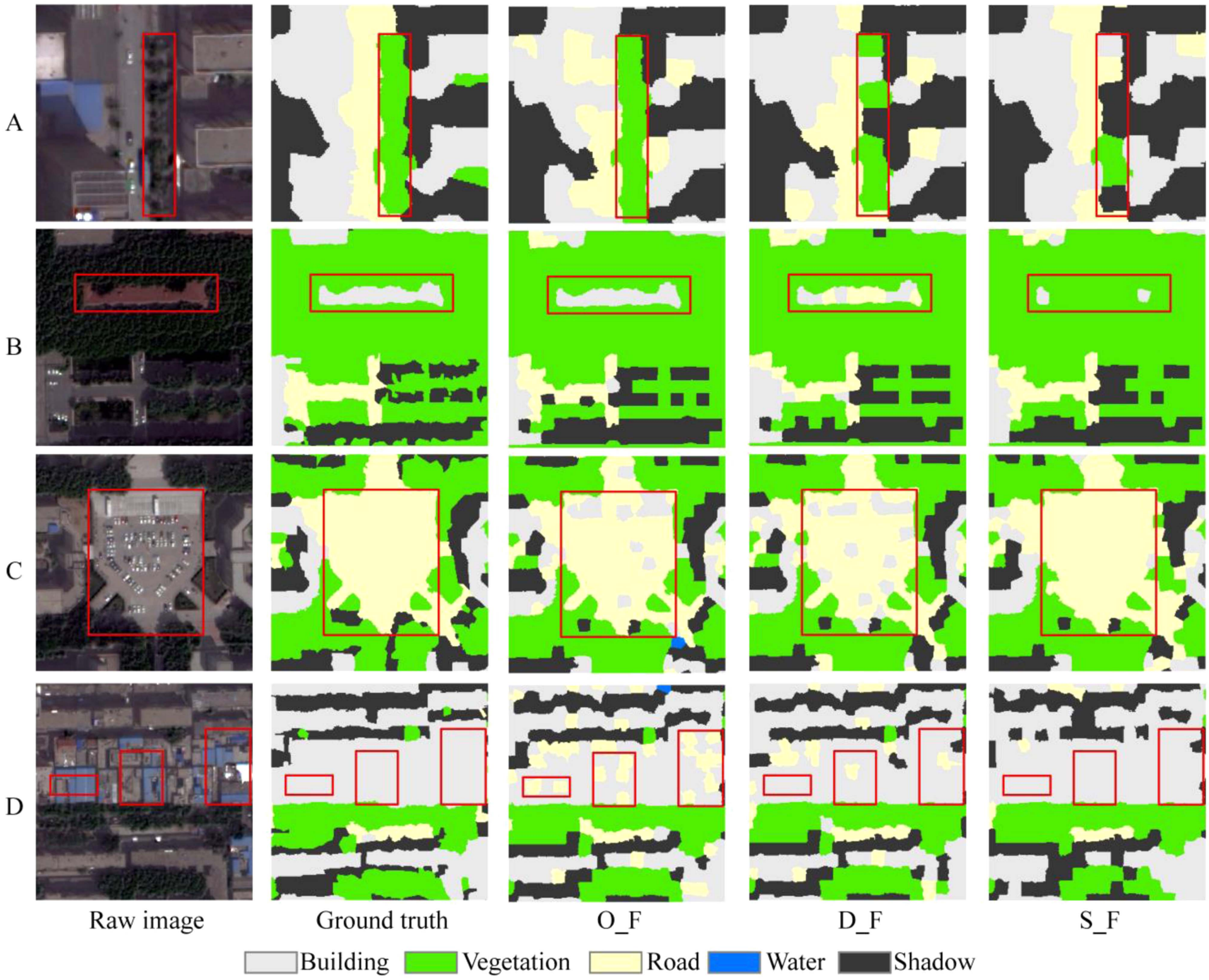


Fig. 10. Local classification result of each single-dimension feature classification method.

of S_F were consistent with the reference classification results. The raw image showed that there were many cars in the red rectangular area of C (parking lot) with complex spectral information, which was close to that of some buildings composed of high-reflectivity materials. In the red rectangular area of D, there were buildings with roofs of different materials and large intraclass spectral differences. The classification results of O_F and D_F in the red rectangular areas of C and D showed that the object features and deep features focusing on the segments' internal information still had limitations in the complex urban scene for the land cover classification. The spatial association features in the red rectangular areas of C and D were very obvious, such as vehicles were on the road; buildings were adjacent to each other; etc. Utilizing the spatial association features can compensate for the deficiency of object features and deep features. Although both O_F and D_F existed the misclassification phenomena in the red rectangular area of C, the classification result of D_F was significantly better than that of O_F in the red rectangular area of area D. This indicated that deep features had stronger generalization ability than object features in the areas with large intraclass spectral differences.

B. Impact of Image Superpixel Segmentation Results on Classification Results

As shown in Table III, the classification accuracy of MULTI_F$_{14}$, MULTI_F$_{34}$, and MULTI_F was different. This demonstrated that the image superpixel segmentation results have a direct impact on the classification results.

As shown in area A of Fig. 11, in the segmentation result of $S = 17$, the combination of the five selected segments can represent the corresponding road; however, in the segmentation result of $S = 14$, the number of segments representing the road increased to 6, among which some of the segments were misclassified as buildings. The reason was that the spectral characteristics of buildings and roads in this area was similar, while the internal spectral difference in the road was large. When S decreased, the image superpixel segmentation results tended to the state of oversegmentation. The difference between road segments increased, and the difference between some road and building segments decreased. Therefore, roads can be misclassified as buildings. The oversegmentation results can bring additional noise in the urban land cover classification

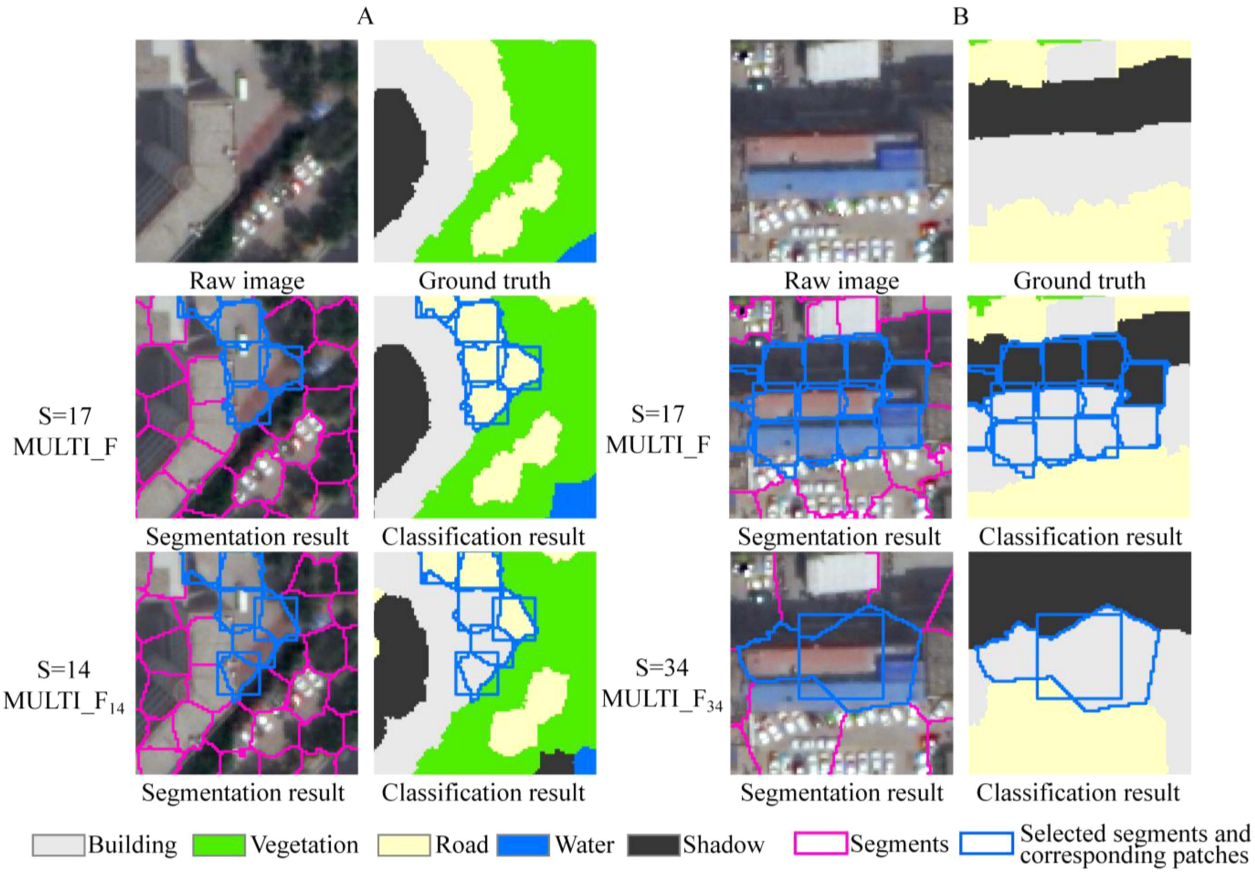


Fig. 11. Local segmentation results for $S = 17$, $S = 14$, and $S = 34$ and local classification results for MULTI_F, MULTI_F$_{14}$, and MULTI_F$_{34}$. First row: Raw images and ground truth; Second row: segmentation results for $S = 17$ and classification results for MULTI_F; Third row: segmentation result for $S = 14$ and classification result for MULTI_F$_{14}$, segmentation result for $S = 34$ and classification result for MULTI_F$_{34}$.

process, and lead to the "salt-and-pepper phenomenon" similar to the what often occurs in pixel-level land cover classification methods.

As shown in area B of Fig. 11, in the segmentation result of $S = 17$, the selected segments contained two land cover categories: buildings and shadows; however, in the segmentation result of $S = 34$, the number of the selected segment reduced to 1, and the selected segment also contained buildings and shadows. According to the classification results of MULTI_F$_{34}$, some shadows were misclassified as buildings, and some buildings were misclassified as shadows. The reason was that when S increased, the image superpixel segmentation result tended to the state of undersegmentation, and multiple land covers can appear in the same segment, resulting in a higher classification error rate. It can also be seen from Fig. 11 that when S = 17, the segments and the image patches of the deep feature extraction model were very consistent in morphology.

This study used a synthetic semivariance function to determine the hyperparameters of the superpixel segmentation. The obtained segmentation result had the advantages of a single land cover category within each segment and the good matching degree between the segment and image patch in morphology. This strategy was beneficial for improving the accuracy of urban land cover classification.

C. *Effectiveness of the Multidimension Feature Fusion Strategy*

Table III showed that the classification accuracy of fusing two-dimension features was better than that of single-dimension features. Among the methods for fusing two-dimension features, (O+D)_F improved the classification accuracy compared with each single-dimension feature classification method but did not show an obvious advantage. The classification accuracies of (O+S)_F and (D+S)_F were obviously better than that of (O+D)_F. This indicated that spatial association features were important for improving the land cover classification accuracy. As shown in Table III, spatial association features had the large impact on building, road, and water categories. After fusing two-dimension features, the classification accuracy of building, road, and water increased from 73.76%, 76.61%, and 83.11% (the highest accuracy in O_F or D_F) to 79.32%, 80.96%, and 89.59% (the highest accuracy in (O+S)_F or (D+S)_F), respectively. Deep features had the large impact on vegetation and shadow categories. After fusing two-dimension features, the classification accuracy of vegetation and shadow increased from 87.18% and 81.11% (the highest accuracy in O_F or S_F) to 87.53% and 82.43% (the highest accuracy in (O+D)_F or (D+S)_F), respectively.

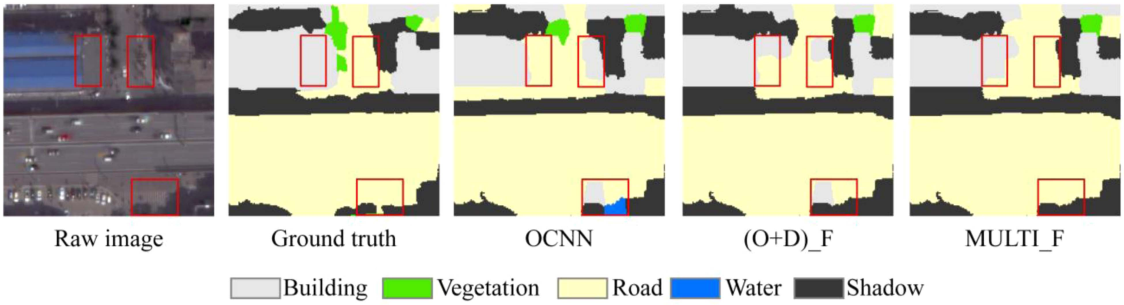


Fig. 12. Local classification results of OCNN, (O+D)_F and MULTI_F.

Table III also showed that the classification accuracy of fusing three-dimension features was better than that of fusing two-dimension features. In addition, after fusing three-dimension features, the classification accuracy of each land cover category was close to or even higher than the highest result of this land cover category for all single-dimension features and two-dimension features. This verified the effectiveness of the multidimension feature fusion strategy in this study.

As shown in Table III, the classification accuracy of MULTI_F (F1-Score 0.8183, OA 82.54%, and Kappa 0.7669) was significantly better than that of OCNN (F1-Score 0.7758, OA 79.11%, and Kappa 0.7209). It can be seen in Fig. 12 that the OCNN classification result was prone to misclassification at the junction of buildings and roads, while the classification result of MULTI_F was significantly optimized at the same scene. In addition, the units for extracting object features and deep features in the OCNN method were inconsistent. Directly fusing the features of different units may increase invalid information and make it difficult to effectively improve classification accuracy. (O+D)_F also adopted the strategy of fusing object features and deep features, but the units for extracting object features and deep features were consistent. This ensured that the fusion of object features and deep features can provide multiangle and valuable information for segments, thus improving the classification accuracy. Comparing with OCNN, it was further demonstrated that the three-dimension feature fusion strategy in this study was more suitable for the urban land cover classification.

V. Conclusion

This study proposed an urban land cover classification method based on segments' multidimension feature fusion. Compared with other methods, our method had a higher accuracy. The specific conclusions were as follows.

1) Compared with the pixel-based method (F1-Score 0.7362, OA 77.47%, and Kappa 0.6987), the segment-based methods generally yielded the higher classification accuracy. For the segment-based methods, the classification accuracy from different single-dimension features was different.
2) The image superpixel segmentation results had the significant impact on the classification results. Using the synthetic semivariance function to determine the hyperparameters of the superpixel segmentation was proposed. The obtained segmentation result had the advantages of a single land cover category within each segment and the good matching degree between the segment and image patch in morphology.
3) Single-dimension features based on segments' object features, deep features, and spatial association features had their own advantages for the urban land cover classification, with the classification accuracies of (F1-Score 0.7843, OA 79.59%, and Kappa 0.7277), (F1-Score 0.7887, OA 80.09%, and Kappa 0.7344), and (F1-Score 0.7661, OA 78.45%, and Kappa 0.7126), respectively. The strategy of multidimension feature fusion can combine the advantages of each single-dimension feature to improve the classification accuracy. Three-dimension feature fusion was the best solution for achieving the urban land cover classification, with the classification accuracy of (F1-Score 0.8183, OA 82.54%, and Kappa 0.7669).

The proposed method shows great potentials in the urban land cover classification. In future research, it can be tried to filter and optimize the multidimension features when reducing the data redundancy. Using deep neural networks to learn multidimension features of segments and constructing the improved models for the urban land cover classification can also be attempted. In addition, it is possible to explore how to adaptively apply the idea of fusing multidimension features of segments based on the optimal segmentation results from different segmentation algorithms.

References

[1] C. Qiu, L. Mou, M. Schmitt, and X. X. Zhu, "Local climate zone-based urban land cover classification from multi-seasonal Sentinel-2 images with a recurrent residual network," *ISPRS J. Photogrammetry Remote Sens.*, vol. 154, pp. 151–162, Aug. 2019.

[2] W. Zhou, D. Ming, X. Lv, K. Zhou, H. Bao, and Z. Hong, "SO–CNN based urban functional zone fine division with VHR remote sensing image," *Remote Sens. Environ.*, vol. 236, Jan. 2020, Art. no. 111458.

[3] J. G. Chang, Y. Oh, and M. Shoshany, "Regional and Global-scale LULC mapping by synergetic integration of NDVI from optical data and degree of polarization from SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 2622–2628, 2023.

[4] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogrammetry Remote Sens.*, vol. 65, no. 1, pp. 2–16, Jan. 2010.

[5] B. Huang, B. Zhao, and Y. Song, "Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery," *Remote Sens. Environ.*, vol. 214, pp. 73–86, Sep. 2018.

[6] X. Wei, W. Zhang, Z. Zhang, H. Huang, and L. Meng, “Urban land use land cover classification based on GF-6 satellite imagery and multi-feature optimization,” *Geocarto Int.*, vol. 38, no. 1, Jul. 2023, Art. no. 2236579.

[7] P. Corcoran, A. Winstanley, and P. Mooney, “Segmentation performance evaluation for object-based remotely sensed image analysis,” *Int. J. Remote Sens.*, vol. 31, no. 3, pp. 617–645, Feb. 2010.

[8] C. Liu, D. Zeng, H. Wu, Y. Wang, S. Jia, and L. Xin, “Urban land cover classification of high-resolution aerial imagery using a relation-enhanced multiscale convolutional network,” *Remote Sens.*, vol. 12, no. 2, Jan. 2020, Art. no. 311.

[9] T. Blaschke et al., “Geographic object-based image analysis – Towards a new paradigm,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 87, pp. 180–191, Jan. 2014.

[10] T. Liu and A. Abd-Elrahman, “Multi-view object-based classification of wetland land covers using unmanned aircraft system images,” *Remote Sens. Environ.*, vol. 216, pp. 122–138, Oct. 2018.

[11] S. Ye, R. G. Pontius, and R. Rakshit, “A review of accuracy assessment for object-based image analysis: From per-pixel to per-polygon approaches,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 141, pp. 137–147, Jul. 2018.

[12] U. C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen, “Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 58, no. 3/4, pp. 239–258, Jan. 2004.

[13] J. Cheng, Z. Huang, J. Wang, and S. He, “The automatic determination method of the optimal segmentation result of high-spatial resolution remote sensing image,” *Acta Geodaetica et Cartographica Sin.*, vol. 51, no. 5, pp. 658–667, May 2022.

[14] M. D. Hossain and D. Chen, “Segmentation for object-based image analysis (OBIA): A review of algorithms and challenges from remote sensing perspective,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 150, pp. 115–134, Apr. 2019.

[15] B. Johnson and Z. Xie, “Unsupervised image segmentation evaluation and refinement using a multi-scale approach,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 66, no. 4, pp. 473–483, Jul. 2011.

[16] C. Zhang et al., “An object-based convolutional neural network (OCNN) for urban land use classification,” *Remote Sens. Environ.*, vol. 216, pp. 57–70, Oct. 2018.

[17] Y. Chen, D. Ming, and X. Lv, “Superpixel based land cover classification of VHR satellite image combining multi-scale CNN and scale parameter estimation,” *Earth Sci. Inform.*, vol. 12, no. 3, pp. 341–363, Sep. 2019.

[18] Z. Tang, M. Li, and X. Wang, “Mapping tea plantations from VHR images using OBIA and convolutional neural networks,” *Remote Sens.*, vol. 12, no. 18, Sep. 2020, Art. no. 2935.

[19] X. Zhang, X. Tan, G. Chen, K. Zhu, P. Liao, and T. Wang, “Object-based classification framework of remote sensing images with graph convolutional networks,” *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8010905.

[20] T. Fu, L. Ma, M. Li, and B. Johnson, “Using convolutional neural network to identify irregular segmentation objects from very high-resolution remote sensing imagery,” *J. Appl. Remote Sens.*, vol. 12, no. 2, May 2018, Art. no. 025010.

[21] X. Lv, D. Ming, T. Lu, K. Zhou, M. Wang, and H. Bao, “A new method for region-based majority voting CNNs for very high resolution image classification,” *Remote Sens.*, vol. 10, no. 12, Dec. 2018, Art. no. 1946.

[22] V. S. Martins, A. L. Kaleita, B. K. Gelder, H. L. F. da Silveira, and C. A. Abe, “Exploring multiscale object-based convolutional neural network (multi-OCNN) for remote sensing image classification at high spatial resolution,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 168, pp. 56–73, Oct. 2020.

[23] W. Zhao, S. Du, and W. J. Emery, “Object-based convolutional neural network for high-resolution imagery classification,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 7, pp. 3386–3396, Mar. 2017.

[24] S. Zhang et al., “EMMCNN: An ETPS-based multi-scale and multi-feature method using CNN for high spatial resolution image land-cover classification,” *Remote Sens.*, vol. 12, no. 1, Dec. 2020, Art. no. 66.

[25] W. R. Tobler, “A computer movie simulating urban growth in the Detroit region,” *Econ. Geogr.*, vol. 46, pp. 234–240, Jun. 1970.

[26] O. Song and Y. Li, “Combining deep semantic segmentation network and graph convolutional neural network for semantic segmentation of remote sensing imagery,” *Remote Sens.*, vol. 13, no. 1, Dec. 2021, Art. no. 119.

[27] W. Wang, F. Liu, W. Liao, and L. Xiao, “Cross-modal graph knowledge representation and distillation learning for land cover classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5520318.

[28] I. Kotaridis and M. Lazaridou, “Cnns in land cover mapping with remote sensing imagery: A review and meta-analysis,” *Int. J. Remote Sens.*, vol. 44, no. 19, pp. 5896–5935, Oct. 2023.

[29] F. Ma, F. Gao, J. Sun, H. Zhou, and A. Hussain, “Attention graph convolution network for image segmentation in big SAR imagery data,” *Remote Sens.*, vol. 11, no. 21, Nov. 2019, Art. no. 2586.

[30] Q. Liu, L. Xiao, J. Yang, and Z. Wei, “CNN-enhanced graph convolutional network with pixel- and superpixel-level feature fusion for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8657–8671, Nov. 2021.

[31] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “SLIC superpixels,” EPFL, Lausanne, Switzerland, Tech. Rep. 149300, Jun. 2010.

[32] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, May 2012.

[33] D. Ming, T. Ci, H. Cai, L. Li, C. Qiao, and J. Du, “Semivariogram-based spatial bandwidth selection for remote sensing image segmentation with mean-shift algorithm,” *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 5, pp. 813–817, Feb. 2012.

[34] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” 2016, *arXiv:1609.02907*, .

[35] C. A. Ramezan, T. A. Warner, A. E. Maxwell, and B. S. Price, “Effects of training set size on supervised machine-learning land-cover classification of large-area high-resolution remotely sensed data,” *Remote Sens.*, vol. 13, no. 3, Jan. 2021, Art. no. 368.

Zhongyi Huang received the B.S. degree in surveying and mapping engineering from Xinjiang University, Wulumuqi, China, in 2020, and the M.S. degree in surveying and mapping engineering from Henan Polytechnic University, Jiaozuo, China, in 2023.

His research interests include land cover/land use, remote sensing, and deep learning.

Jiehai Cheng received the B.S. degree in land planning and utilization and M.S. degree in cartography and geography information engineering from Henan Polytechnic University, Jiaozuo, China, in 2002 and 2005, respectively, and the Ph.D. degree in cartography and geography information system from Beijing Normal University, Beijing, China, in 2013.

He is currently a Professor with the School of Surveying & Land Information Engineering, Henan Polytechnic University, Jiaozuo, China. His research interests include high-resolution remote sensing, nighttime light remote sensing, and deep learning.

Guoqing Wei is currently working toward the graduate degree in geographic information science in Henan Polytechnic University, Jiaozuo, China.

His research interests include image processing.

Xiang Hua received the B.S. degree in surveying and mapping engineering from Henan Polytechnic University, Jiaozuo, China, in 2021.

His main research interests include nighttime light remote sensing and deep learning.

Yuyao Wang received the B.S. degree in geographic information science from Henan Polytechnic University, Jiaozuo, China, in 2022.

His research interests include remote sensing image classification.