An Object-Oriented Semi-Supervised Land-Use/Land-Cover Change Detection Method Based on Siamese Autoencoder Graph Attention Network

Yifan Wu, Zhiwei Xie, Zaiyang Ma, Min Chen, Fengyuan Zhang, Zhenguo Shi, Wengang Li, and Shuaizhi Zhai

***Abstract*—Change detection via remote sensing data is a popular method for monitoring land cover/land use. Graph attention (GAT) network is a method that can improve the change detection performance of land-cover/land-use monitoring by enhancing the feature representation of remote sensing images. However, the shortcomings of connection sparsity and insufficient sample feature mining of the GAT affect the application of these methods in change detection. This article proposes a Siamese autoencoder GAT network for object-oriented land-cover/land-use change detection via high-resolution remote sensing, which is useful for semi-supervised problem methods with poor simplicity. First, we reduce the pressure of graph network model operations and obtain multidimensional features via the growing multilevel segmentation strategy. The adjacency matrix is established by adding strongly connected edges via the weighted difference similarity matching method with image objects as nodes. Second, we use the Siamese autoencoder to pretrain the node features and migrate the weight parameters to the GAT feature extraction layer. Finally, a small number of samples are selected to train the GAT and predict all nodes. The experimental results show that the average overall accuracy of the proposed method is 95.41%, and the average $F1$-score is 89.35%, which are at least 5.04% and 10.84% better than those of other typical methods, such as the graph convolutional network. In particular, detecting roads and bare soil is significantly better than that of other methods.**

***Index Terms*—Graph attention (GAT) networks, growing multilevel segmentation, Siamese autoencoder (AE), similarity matching.**

I. Introduction

IDENTIFYING and recording Earth's surface changes are critical topics for geographic and environmental research. Land-use/ land-cover change detection is an important method for Earth surface change monitoring. Moreover, remote images from different times within the same landscape can be identified, and these images have wide applications in ecological environment monitoring and land resource management [1], [2], [3]. Remote sensing image change detection methods can be divided into two categories based on the units of analysis: pixel-based methods [4] and object-oriented methods [5]. Pixel-based change detection methods are mostly applied to medium- and low-resolution remote sensing images; these images have low-pixel purity, and it is difficult to form pixel sets with high homogeneity [6]. The object-oriented change detection method can combine spatial, spectral, and textural information from images. With the rapid development of Earth observation technology, the spatial resolution of remote sensing images is increasing. Compared with conventional remote sensing images, high-resolution remote sensing images have richer spatial feature information. A higher spatial resolution highlights the correlation between pixels, which increases "salt and pepper" noise [7]. Therefore, compared with pixel-based methods, object-oriented change detection methods have more advantages in terms of efficiency and feature dimension capacity [8], [9].

Deep learning models, such as convolutional neural networks (CNNs), are not suitable for object-oriented image processing due to the non-Euclidean data structure of image objects [10], [11], [12]. A graph convolutional network (GCN) is a machine

Manuscript received 22 August 2023; revised 1 November 2023, 22 December 2023, 17 January 2024, and 14 February 2024; accepted 25 February 2024. Date of publication 8 March 2024; date of current version 18 March 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42101353, in part by the Humanities and Social Sciences Foundation of the Ministry of Education of China (General Program) under Grant 21YJC790129, and in part by the Basic Research Programs of the Educational Department of Liaoning Province of China under Grant LJKMZ20220946 and Grant LJKMZ20222128. *(Corresponding author: Zhiwei Xie.)*

Yifan Wu, Zhenguo Shi, Wengang Li, and Shuaizhi Zhai are with the School of Transportation and Geomatics Engineering, Shenyang Jianzhu University, Shenyang 110168, China (e-mail: q984641286@gmail.com; szgaq77@163.com; 664271710@qq.com; zsz2152@sina.com).

Zhiwei Xie is with the School of Transportation and Geomatics Engineering, Shenyang Jianzhu University, Shenyang 110168, China, also with the Key Laboratory of Virtual Geographic Environment (Ministry of Education of PRC), Nanjing Normal University, Nanjing 210023, China, also with the State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing Normal University, Nanjing 210023, China, and also with the Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing Normal University, Nanjing 210023, China (e-mail: zwxrs@sjzu.edu.cn).

Zaiyang Ma and Min Chen are with the Key Laboratory of Virtual Geographic Environment (Ministry of Education of PRC), Nanjing Normal University, Nanjing 210023, China, also with the State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing Normal University, Nanjing 210023, China, and also with the Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing Normal University, Nanjing 210023, China (e-mail: mazaiyang@njnu.edu.cn; chenmin0902@163.com).

Fengyuan Zhang is with the School of Environment, Nanjing Normal University, Nanjing 210023, China (e-mail: zhangfengyuan@nnu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3373039

learning method that has advantages in image processing because of the non-Euclidean data structure of image objects [13], [14]. To date, there are few GCN applications in remote sensing image change detection. GCN-based change detection methods can effectively improve change detection accuracy through end-to-end learning of node features and structural information [15], [16], [17]. However, the adjacency weights of the above methods need to be manually customized, which limits their application. A graph attention (GAT) network can aggregate neighboring nodes via attention. It achieves adaptive assignment of adjacency weights and can extract richer feature information than a GCN [18]. GATs are designed for application in traditional networks, and their standardized graph propagation structure is suitable for traditional network data with rich connectivity relationships, such as social network data [19], [20], [21] and traffic network data [22], [23], [24]. The sparse connection relations of remote sensing images cannot meet the GAT network density requirements. Moreover, the dimensionality of the feature space of remote sensing data is much greater than that of traditional network data. This will put considerable pressure on feature selection and fusion in the GAT. Effective feature depth mining methods are urgently needed to maintain a constant sample size.

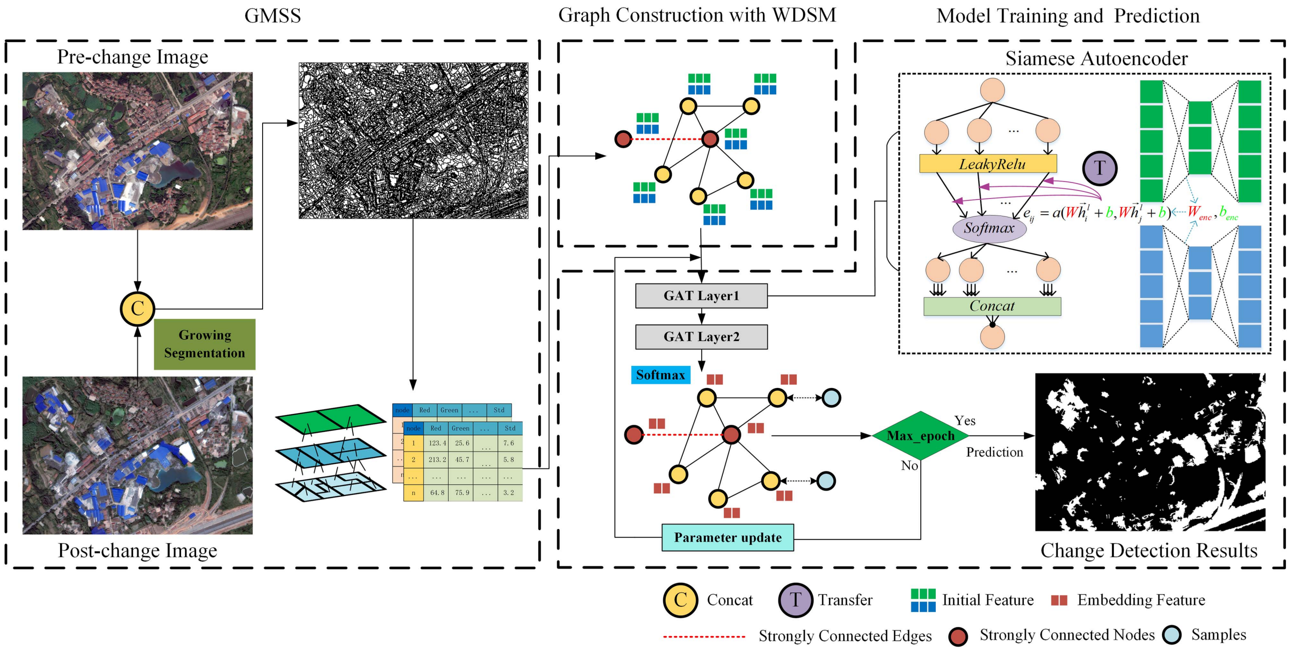
Training sample size control is another challenge for change detection methods. Deep learning models can be classified as supervised or unsupervised according to their dependence on samples. The supervised models can determine the land-cover change classes. It is robust for different atmospheric and lighting conditions, and a large number of high-quality training samples are needed to improve its accuracy [25], [26], [27]. The unsupervised model does not require samples, and it can automatically fit the classification function through the feature distribution of the difference images. Moreover, these methods lack data-based modeling capabilities and guidance on output results, which may lead to failure in practice. Semi-supervised models have been developed to combine the advantages of supervised and unsupervised models. Semi-supervised models use a small number of labeled samples for learning, continuously mine information from a large amount of unlabeled data during the learning process, and update the labeled samples with this information [28]. The semi-supervised method solves the problem of an insufficient number of labeled samples and has better applications in change detection [29]. In addition, graph transductive models, represented by GCN and GAT, are feasible for semi-supervised learning. It assumes that nearby nodes tend to have the same label, thus propagating the information from labeled to unlabeled nodes [30].

A semi-supervised GAT uses a small number of labeled samples to achieve node feature representation. A small number of labeled samples reduces the cost of sample labeling but may give rise to deficiencies in feature representation. The main solutions to this deficiency include enriching the feature space of the labeled samples and learning knowledge from unlabeled samples. Multiscale features, temporal features, pixel-level features, and image object-level features are applied to expand the feature space. Shuai et al. [31] proposed a superpixel-based multiscale Siamese graph attention network to increase feature expression and generalization through multiscale node features. The authors in [32] and [33] designed temporal–spatial joint GAT (TSJGAT), a GAT network capable of fusing pixel-level and image object-level features, to correct the accumulation of node feature errors due to image segmentation using pixel-level features and to improve the deficiencies of image object-level feature representation. Moreover, the TSJGAT utilizes the temporal–spatial joint correlations reflected in the multitemporal images and obtains a temporal–spatial affinity matrix, allowing the graph nodes to aggregate multiple effective features according to the correlation matrix, further enhancing the feature representation capability. Unlabeled samples can be obtained by analyzing their similarity to labeled samples. Jia et al. [34] took advantage of GATs ability to ingest knowledge related to unlabeled regions from unlabeled image datasets with a small number of labeled samples. They explored the regions of consistency between multitemporal images in the latent spatial domain and utilized the spatial information of the opposite images to obtain enhanced feature representations. Sun et al. [35] implemented the Siamese nested UNet (SANet) using a GAT network. Pseudolabels with high confidence are obtained with the help of pixel-level threshold filtering. The accuracy of the pseudolabel is improved by comparing the pseudolabel with the enhanced original image. The above methods enhance the feature expression ability of the GAT, but there is still much room for improvement.

To effectively leverage the performance advantages of GAT feature expression for land-use/land-cover remote sensing monitoring, the problems of connection sparsity and insufficient sample feature mining need to be addressed. This article proposed a Siamese autoencoder GAT (SA-GAT) network for semi-supervised land-use/land-cover change detection via high-resolution remote sensing. First, we used a growing multilevel segmentation strategy (GMSS) to reduce the computing pressure and constructed multidimensional and multilevel feature spaces. Second, with the image objects as nodes, an adjacent matrix was established to construct a multitemporal image network. Then, we used the chi-square transformation (CST) to calculate the weighted difference of the multidimensional features, and the weighted difference similarity matching (WDSM) method was used to improve the graph connection density. Moreover, the Siamese autoencoder (AE) was used to deeply mine the high-dimensional features of node attributes, and the obtained weight parameters were subsequently transferred to the GAT. Finally, the model was trained using a small number of samples to obtain a classifier for change detection. A flowchart of the proposed change detection method is shown in Fig. 1.

The major contributions of this study are given as follows.

1) A GAT change detection network based on a Siamese AE is proposed. The traditional GAT relies on only labeled samples to obtain a high-dimensional representation of nodes during network training. The proposed network uses a Siamese AE for unsupervised training and obtains a high-dimensional representation of feature vectors by continuously minimizing the reconstruction errors between the input and output.



node	Red	Green	...	Std
1	123.4	25.6	...	7.6
2	213.2	45.7	...	5.8
...	...	...	...	...
n	64.8	75.9	...	3.2

Fig. 1. Flowchart of the proposed change detection method.

2) We propose a graph connection density method based on similarity matching. Traditional graph construction methods do not consider the connectivity of similar nodes during network propagation. By using the similarity matching method, virtual connection edges are added to realize information sharing among similar nodes.

It is beneficial to break out of the information jam in the process of node aggregation and to spread network supervised information.

3) A growing multilevel image segmentation strategy is proposed. Small-scale image segmentation results were taken as constrained thematic maps, and the subsequent model operation efficiency and multilevel feature extraction were optimized by combining growth segmentation and multilevel segmentation models.

The rest of this article is organized as follows. Section II introduces related works, such as attention mechanism, multimodel remote sensing change detection, and semi-supervised learning. In Section III, the GMSS, similarity matching graph connection density enhancement method, and SA-GAT change detection method are introduced. Sections IV and V present the experimental details, ablation experiment, and parameter discussions, respectively. Finally, Section VI concludes this article.

II. Related Work

A. Attention Mechanism

Attention mechanisms improve the ability of neural networks to perceive meaningful features. In 2014, attention mechanisms were successfully applied to image classification and natural language processing based on recurrent neural networks (RNNs) [36], [37]. Two years later, Yin et al. [38] proposed the use of attention mechanisms in CNNs and applied them to statement modeling. Subsequently, Google came up with a text representation using a self-attention mechanism, which forms the core module of transformer [39]. The spatial attention mechanism determines the location information of interest by learning all the bands [40]. The spatial weight matrix obtained by the spatial attention mechanism is one of the main methods for enhancing CNN encoders to capture global contextual features [41], [42], [43]. The spectral attention mechanism establishes correlations between different bands and can enhance important spectral features [44], [45], [46]. In addition, the spatial–spectral attention mechanism combines the advantages of spectral and spatial features to optimize the feature representation of images in local regions [47], [48]. In response to the temporal features of multitemporal remote sensing, encoders with temporal attention mechanisms focus on spatial regions characterized by continuous temporal changes. The main frameworks used in this mechanism are neural networks, such as fully convolutional networks [40], RNNs, and long short-term memory (LSTM) networks [49]. The GAT can improve the deficiency of CNNs that cannot fuse long-range feature dependencies [35], [50]. For the sample self-training problem, the GAT mines information related to the labeled regions from the unlabeled image dataset [34]. It has a prominent role in capturing the global spatial dependence of image pixels or image objects [51].

B. Multimodel Remote Sensing Change Detection

The spectral, spatial, and temporal features of multitemporal remote sensing data, and multisource remote sensing data and the combined use of all the above information can help to improve the change detection performance. In general, multimodel change detection methods can be categorized into multimodel-integrated frameworks and multimodel data fusion frameworks

[52]. The multimodel-integrated framework has a hybrid structure, similar to a double-stream design, that combines the advantages of feature extraction from different models [53]. The objective was to highlight the characteristics of real changes between multitemporal images and to suppress the interference of spectral differences within the same class of land cover. Kou et al. [54] combined convolutional LSTM with the conditional generative adversarial network (GAN) to form a domain translation framework. The framework reconstructs the prechange image using temporal features and ensures that it has spectral consistency with the postchange image. Li et al. [55] combined the deep features of a deep belief network and the low-level features of a support vector machine to analyze multitemporal ZiYuan-3 images and verified that the multimodal features help to improve detection accuracy. The multimodal data fusion framework improves detection accuracy by integrating information from multiple sources of data. Zhang et al. [56] introduced the use of a multiscale morphological gradient in a nonsubsampled shearlet transform pulse-coupled neural network for edge detection. This method improved the utilization of edge features and achieved the fusion of SAR and optical remote sensing data. Liu et al. [57] subsequently addressed the difference in resolution of the input data in the different-resolution change detection task from the perspective of feature alignment. Zhang et al. [58] proposed a multispatial resolution image change detection framework that combines high-level deep learning features and low-level mapping change features. All of the above multimodel remote sensing change detection methods are used to enrich and refine the feature space through the cross application of multiple models or multisource data. Multidimensional features play a key role in encoding and decoding changes via deep learning and will always constitute a hot research topic.

C. Semi-Supervised Learning

The semi-supervised learning method is one of the main methods for overcoming the problem of insufficient training samples. A pixel-level semi-supervised change detection method was proposed by Bandara and Patel [59]. The method utilizes unlabeled samples to randomly perturb the difference map with hidden features and maintains the consistency of the predicted change probability map in the change detection network. To fully utilize the potential of unlabeled data, Peng et al. [60] used two discriminators to enhance the consistency of feature distributions in segmentation and entropy maps between labeled and unlabeled data. Liu et al. [61] proposed a semi-supervised remote sensing change detection method based on a GAN and a graph model. First, the multitemporal remote sensing change detection problem is transformed into a semi-supervised learning problem on a graph, which includes most unlabeled nodes and a few labeled nodes. Then, GANs are used to generate samples in a competitive manner and help improve classification accuracy. The above semi-supervised change detection methods are all pixel based. Integrating the characteristics of semi-supervised methods into object-oriented deep learning change detection and establishing an object-oriented semi-supervised deep model are still problems that need to be further explored.

D. Parameters of Fractal Net Evolution Approach (FNEA)

FNEA segmentation is widely used for remote sensing image segmentation and is the core algorithm of eCognition. The main parameters of FNEA include the shape factor w_{shape}, the compactness scale w_{compact}, and the segmentation scale ζ_{scale}. Inappropriate settings of these parameters may result in both over- and undersegmentation [62]. Usually, the default values for the first two parameters are 0.1 and 0.5. According to several related studies, the default values for these two parameters are generally acceptable [63], [64], [65], [66], [67], [68], [69], [70], [71], [72], [73]. The parameter ζ_{scale} is a user-defined threshold for controlling average object size. More segmented objects are typically produced by smaller values of ζ_{scale}, and vice-versa [74]. Currently, there are two methods available ζ_{scale}: automatic [75], [76], [77] and manual methods [70], [78], [79], [80]. The estimation of scale parameters of eCognition introduced the local variance (LV) index to reflect the homogeneity of the segmentation results. The LV index will stop increasing when the segmentation scale reaches a particular range, and this scale may be the optimal segmentation scale [76]. The existing automatic determination methods mostly provide reference candidates for the optimal segmentation scale, but manual or segmentation strategies need to be used in conjunction to obtain the final results. The manual selection of segmentation scale methods usually has better applicability because the parameters are determined according to the specific application. The characteristics of the data and ground objects are important factors that affect the manual selection of segmentation scales.

III. Materials and Methods

This study comprised three primary stages, as illustrated in Fig. 1. First, we used GMSS to segment the multitemporal images and construct the multidimensional feature space of the image objects. Second, we constructed the graph using image object nodes and a first-order adjacency matrix and increased the connection density via the WDSM. Third, we used the Siamese AE to express multidimensional features and trained the GAT with a few training samples to obtain a change detection classifier.

A. Image Object Acquisition With GMSS

We designed a new segmentation strategy, named GMSS, to obtain multitemporal image segmentation objects with the same segmentation boundary. In this article, an image object is a segmentation result, not a single building. The GMSS fully considers the spectral and spatial features of an image and integrates image features at different levels through multilevel segmentation. This approach has the potential to enhance the adjacency relationship between image objects at a small scale, thereby significantly improving both local and global sensing field features in subsequent processing.

The traditional overlay segmentation strategy involves first segmenting the images and then overlaying them to ensure consistent segmentation boundaries of multitemporal remote sensing images [81]. Due to the influence of multitemporal

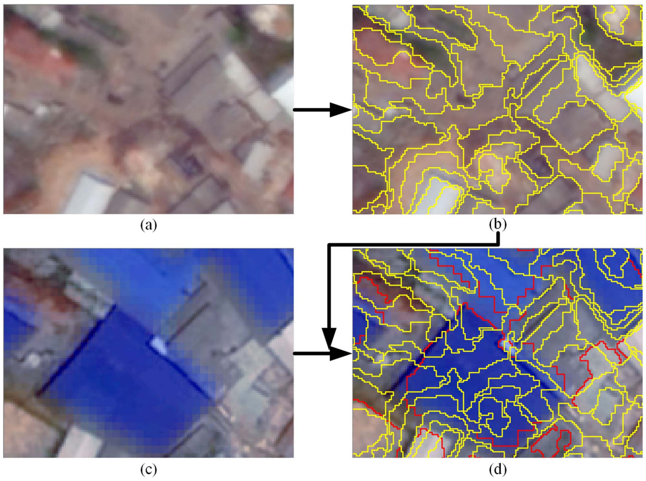


Fig. 2. Schematic of the growing segmentation strategy. (a)–(d) are prechange image, prechange image segmentation, postchange image, and postchange image segmentation. This is one of the all levels of segmentation. The yellow lines signify the boundaries extracted from the prechange image, while the red lines represent the newly derived boundaries from the postchange image. The arrows indicate the direction of influence, symbolizing how the segmentation boundaries of the prechange image guide the segmentation process of the postchange image. The partial images are derived from Data A.

remote sensing image registration and spectral differences, this strategy generated too many broken boundaries and noisy image objects. Moreover, a single segmentation scale was inadequate for accurately delineating the diverse ground object boundaries. A GMSS schematic diagram is shown in Figs. 2 and 3. In Fig. 2(d), the yellow lines represent the segmentation boundaries of the prechange image, and the red lines represent the new segmentation boundaries of the postchange image. The segmentation boundaries of the prechange image were used as the thematic map to constrain the postchange image segmentation, and the segmentation results of the postchange image were used as the final segmentation boundaries. Moreover, a multilevel segmentation model was established. L segmentation layers were set up, the minimum-scale image objects in the first layer were obtained by using growth segmentation, and the multilevel segmentation objects were obtained step by regionally merging the lower level image objects based on the heterogeneity of spectral and shape. The multilevel segmentation scale is given as follows:

$$\text{level}n = \text{level}_{\text{ini}} + n * T_{\text{scale}} \tag{1}$$

where $\text{level}_{\text{ini}}$ represents the starting segmentation scale, T_{scale} is the fixed step size, n denotes the level, and $0 \leq n \leq L$. The segmentation scale at the first level is level0, and the pixels are taken as the merging objects. We first obtained the prechange image segmentation boundary preB0 at scale level0; then, using the boundary preB0 as the thematic map, we derive the postchange image segmentation boundary postB0 at the same scale level0. The final segmentation results of the first level of segmentation are the segmentation boundary postB0 and the image object sets Object0. Moving on to the second level, the merged objects were the image object sets Object0, and other operations were similar to the first level of segmentation. The segmentation process for the other levels was consistent with that for the second level. We generally set $L = 4$, $T_{\text{scale}} = 15$, and $\text{level}_{\text{ini}}$ according to the image and feature characteristics. We used FNEA to obtain the segmentation boundaries.

Multidimensional feature combinations have been proven to be suitable for land-use/land-cover change detection, but multilevel feature representations have been less considered. After multilayer segmentation, each layer of the image object feature space was constructed by spectral, textural, geometric, indexing, and other features. We used eCognition to extract the features of the image objects, including spectral, textural, indexing, and geometric features. The spectral features included the average, maximum, and minimum gray values of the blue, green, red, and near-infrared (NIR) bands, and the standard deviation of gray values of the NIR band. The texture features included the mean and entropy of the grayscale cogeneration matrix of the NIR band. The index features included the normalized difference water index and normalized difference vegetation index. The geometric features were the area of the image object at each segmentation level.

B. Graph Connection Density Enhancement Method Based on WDSM

The graph structure can describe the relationships between nodes. A good graph structure is a prerequisite and key to fully utilizing the performance of a graph neural network (GNN). We used the image objects as nodes and the relationships between image objects as edges. Change image objects often have excessive variability with surrounding neighboring nodes, which affects the connectivity of the graph. This approach made it difficult to effectively transfer supervised information. Theoretically, two nodes with close differentiation had similar supervised information. Thus, a strong graph connection density enhancement method based on similarity matching was proposed. After constructing the first-order adjacency matrix using the location information of nodes, we used the CST to calculate the weighted difference of the multidimensional features [82]. The virtual connected edges were added by the WDSM to enhance the connection density of the graph.

The CST used the variance in the difference bands as the weight of the fusion and compressed the multidimensional features into 1-D features. The CST is given as follows:

$$d_n = \sum_{i=1}^{M} \left(\frac{F_i^2 - F_i^1}{\sigma_i} \right)^2, i = 1, 2, \ldots, M \tag{2}$$

where d_n is the weighted difference of the image object n, and $D = \{d_1, \ldots, d_n \ldots, d_N\}$. N is the total number of image objects, $t \in \{1, 2\}$ represents the prechange and postchange times, $F_t = \{F_{t1}, \ldots, F_{ti}, \ldots, F_{tM}\}$ is the M-dimensional feature vector, and σ_i is the standard deviation of the difference band i. The image objects with similar degrees of difference had similar supervised information. With this in mind, this article designed a graph connectivity enhancement method based on difference similarity. Iterating through all the elements in D, we calculated the absolute value of the difference between d_n and the other elements in D, and sorted the results from smallest

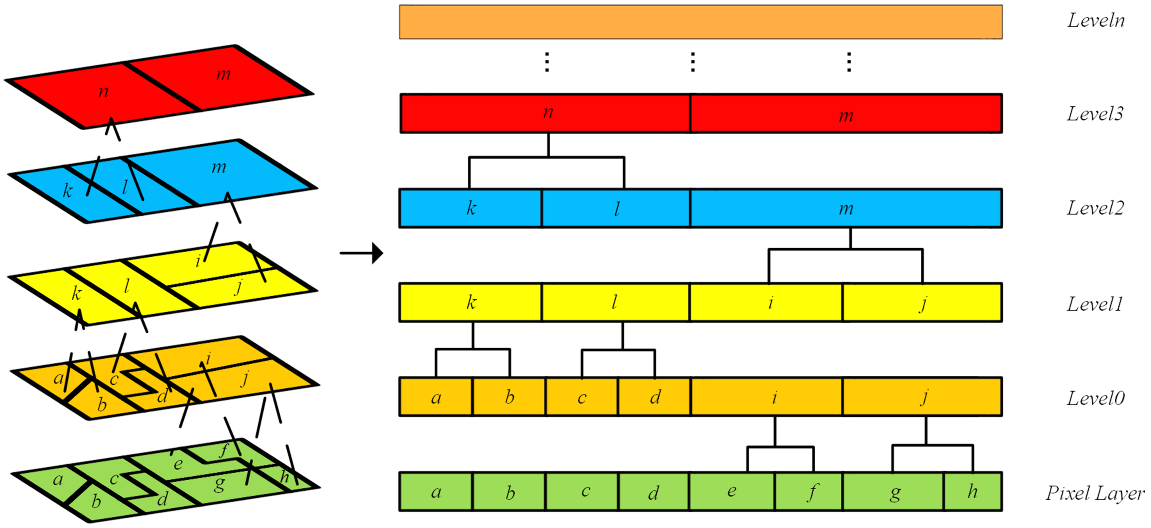


Fig. 3. Schematic diagram of the GMSS. GMSS takes the image objects obtained from the smallest segmentation scale as the basic units and establishes the correlation of the segmentation results at different scales by merging the basic units level-by-level.

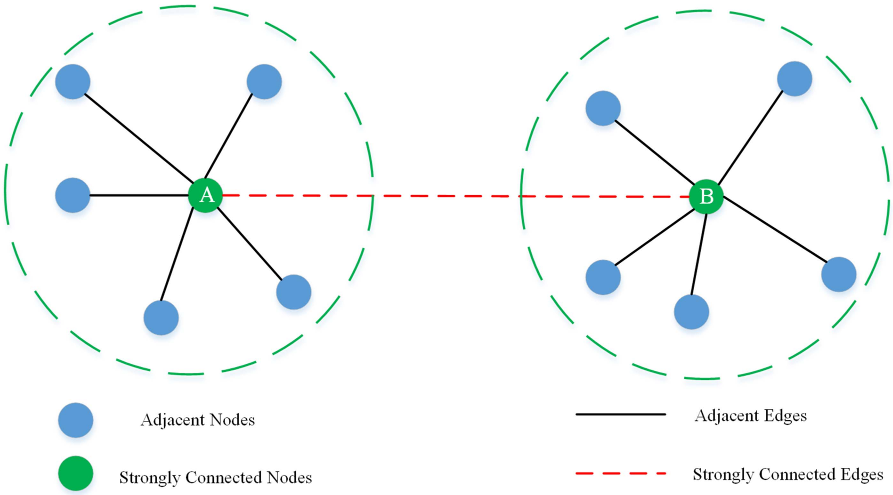


Fig. 4. Schematic diagram of the WDSM. Each node represents an image object. The nodes within each green circle represent the set of nodes that have a first-order adjacency relationship with node A or B, respectively.

to largest. Then, we set the similarity step s, match the first s nodes with the smallest absolute value of the difference with d_n as strongly connected nodes, and establish the connection relationship between nodes belonging to d_n. A schematic diagram of the WDSM is shown in Fig. 4. The red edge connecting nodes A and B between subgraphs is a virtual connected edge constructed by the WDSM, which can improve the connectivity density of the graph.

The article constructed the difference graph by the above steps. The graph was an undirected graph of $G = (V, E)$, where $V = \{v_1, \ldots, v_i, \ldots, v_N\}$ was the set of nodes and E was the set of edges. N was the number of nodes. Since an image object corresponded to a node, the number of nodes in the graph corresponded to the number of image objects. A was defined as the strongly connected adjacency matrix of G. The element a_{ij} in A represents the adjacency relationship between v_i and v_j. Each node represents the physical meaning of each image object, while the graph representation constructs the spatial, spectral, and topological relationships of image objects in a graph structure, providing a comprehensive understanding of the semantic information of image objects from various feature dimensions.

C. Change Detection Based on SA-GAT

1) GAT Network: The GAT achieved the adaptive assignment of different neighbor weights by aggregating neighbor nodes through an attention mechanism. This approach greatly improved the expressive power of the GNN. We assumed that a graph has N nodes and that the multidimensional feature set of nodes can be represented as $h = \{\vec{h}_1, \ldots, \vec{h}_i, \ldots, \vec{h}_N\}$. Here, the nodes were derived from the graph constructed from the image objects, and the number of nodes was consistent with the number of image objects. The purpose of the attention layer was to output the new node feature set $h' = \{\vec{h}'_1, \ldots, \vec{h}'_i, \ldots, \vec{h}'_N\}$.

The method applied the linear transformation W, from $\vec{h}_i$ to a new feature $\vec{h}'_i$. W was the shared weight matrix, which transformed the input into higher level features. The self-attention mechanism was shared along the connecting edges of adjacent nodes to calculate the attention coefficient between the target node and the neighboring nodes

$$e_{ij} = \alpha\left(W\vec{h}^l_i, W\vec{h}^l_j\right) \tag{3}$$

where e_{ij} is the calculated attentional coefficient and α is a shared attentional mechanism. We computed e_{ij} from the feature vector $\vec{h}^l_{\mathrm{i}}$ of node i in the GAT layer l and the feature vector $\vec{h}^l_j$ of node j. The above attention calculation considered any two nodes in the graph, i.e., the influence of each node in the graph on the target node was considered so that the graph structure information was lost. For the target node i, only the correlation of the nodes in its neighborhood to the target node (including its own influence) was calculated. The softmax function normalizes the attention coefficient to an easily comparable form. The spatial relationship matrix A was used as the basis, and the coefficients at nonzero positions in A were retained. Finally, the final normalized attention coefficients α_{ij} were obtained by applying the LeakyReLU activation function as follows:

$$\alpha_{ij} = \mathrm{Softmax}_j\left(\mathrm{LeakyReLU}\left(e_{ij}\right)\right) = \frac{\exp(e_{ij})}{\sum_{v_p \in \tilde{N}(v_i)} \exp(e_{ip})} \tag{4}$$

where $\tilde{N}(V_i)$ is a neighbor of v_i and v_p is a neighbor node. The LeakyReLU activation function gives a nonzero slope to all negative variables, which improves the vanishing gradients problem. Following the idea of weighted summation of attention mechanisms, the new feature vector of node v_i is given as follows:

$$\vec{h}^{l+1}_i = \sigma\left(\sum_{v_j \in \tilde{N}(v_i)} \alpha_{ij} W \vec{h}^l_j\right) \tag{5}$$

where σ is the sigmoid activation function. The sigmoid activation function maps the variables to the interval 0–1. In addition, to stabilize the learning process of self-attention, a group of independent attention mechanisms was applied to the top and the output results were spliced according to the following equation:

$$\vec{h}^{l+1}_i = \|^K_{k=1}\ \sigma\left(\sum_{v_j \in \tilde{N}(v_i)} \alpha^k_{ij} W^k \vec{h}^l_j\right) \tag{6}$$

where $\|$ is the concatenation operation, α^k_{ij} is the weight coefficient computed by the k th attention mechanism, and K is the total number of attention mechanisms. W^k is the input linear transformation matrix. In addition, to avoid the problem of oversmoothing the model due to too many GAT layers, we designed the GAT using two GAT layers. The number of GAT layers should not be too high; otherwise, it will cause oversmoothing.

2) SA-GAT and Change Detection: The GAT is a semi-supervised learning model and has the advantage of fully combining GAT propagation characteristics with topological structure [83]; however, the following problems remain. First, the GAT optimizes network features by sharing a weight matrix to train and extract features through labeled samples. The effectiveness of the GAT in extracting features would be affected if there were insufficient training samples for semi-supervised change detection or if the samples deviated significantly from the overall data in terms of distribution [84]. Second, conventional change detection methods superimpose feature matrices of multitemporal images and then import them into deep learning networks. This process allowed the variance detection to start from the first layer, resulting in features belonging to different layers interacting with each other. It was difficult to maintain the high-dimensional features of the original image [85]. To improve the above deficiencies, we optimized the training and feature extraction of the GAT using the Siamese AE.

Pretraining of node features via the Siamese AE involves identifying optimal high-dimensional features. The encoder transforms the original low-dimensional features $F_t = \{f_{t1}, \ldots, f_{ti}, \ldots, f_{tM}\}$ of the multitemporal nodes into high-dimensional features by an encoding matrix W_{enc}. The decoder retransforms the high-dimensional features into temporary low-dimensional features $F_t{}' = \{f'_{t1}, \ldots, f'_{ti}, \ldots, f'_{tM}\}$ by decoding the matrix W_{dec} to obtain the optimal high-dimensional features. The training is performed by the gradient descent algorithm and iterative operation to compare the difference between the low-dimensional original feature F_t and the newly generated low-dimensional feature $F_t{}'$. When the difference does not change, the encoder obtains the optimal parameters W_{enc} and b_{enc}. In addition, the process of obtaining high-dimensional features via multiple temporal nodes is synchronized, and the parameters W_{enc} and b_{enc} are shared among them. To migrate the weight parameters of the Siamese AE, we replaced the weight feature extraction layer of the GAT with a fully connected (FC) layer. That is, the weight parameters of the original GAT feature extraction layer were replaced with the weight parameter W_{enc} and the bias parameter b_{enc} of the Siamese AE. This allowed the parameter form of the feature extraction layer of the GAT to be adjusted to ensure consistency with the form of the weight parameters of the Siamese AE. A schematic diagram of the SA-GAT is shown in Fig. 5.

An AE was able to learn the implicit features of input data in an unsupervised manner and continuously minimized the reconstruction error between the input and output through training. The change detection task is mostly dominated by a multitemporal data space, which was used to construct a Siamese AE [86]. The Siamese network was used to obtain high-dimensional features of input multitemporal nodes through unsupervised pretraining. These high-dimensional features were directly bought into the subsequent GAT node representations by the network as parameters, allowing for the extraction of more feature information even with limited samples. In the Siamese network structure, feature vectors of multitemporal images were used as double inputs, and the parameters of the network layer were shared to ensure that the multitemporal images had the

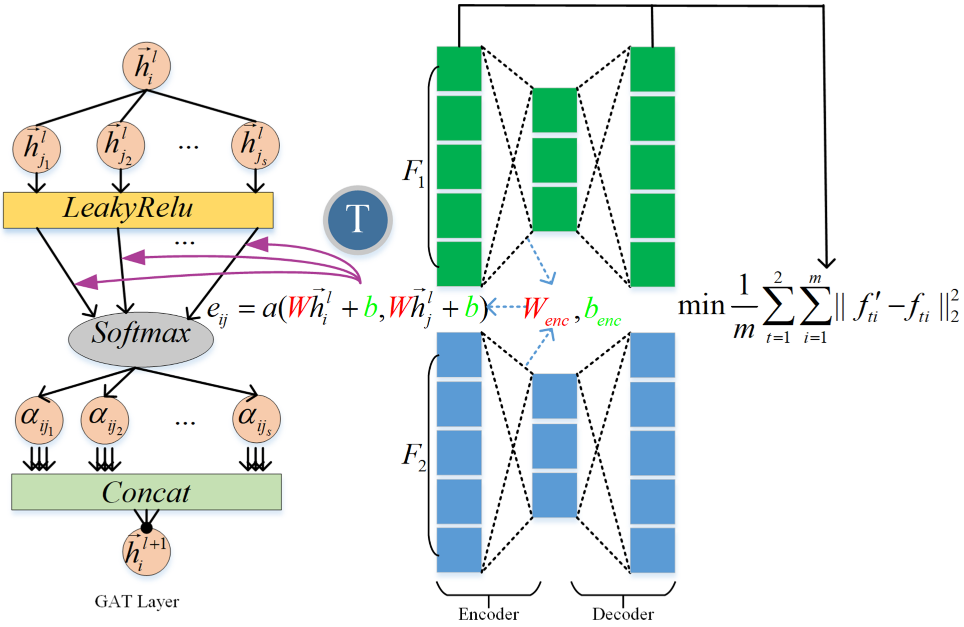


Fig. 5. Schematic diagram of the SA-GAT. The *T* in the circle stands for transfer, i.e., transferring the parameters trained by the AE directly to the layer of the GAT.

same high-dimensional features. Given a multitemporal input feature vector F_t, the change weight matrix from the input layer to the hidden layer was W_{enc}, and the encoder encoded the input as h_t. The decoder mapped the encoded features h_t to the input space, and the reconstructed input $F_t^{'}$ was obtained. The method assumed that the encoding matrix from the hidden layer to the output layer was W_{dec}. The AE is a three-layer neural network that includes an input layer, a hidden layer, and an output layer. The AE compresses the number of input features from 144 to 32. The reconstructed input is shown in the following equation, where σ is the sigmoid activation function:

$$h_t = \sigma\left(W_{\text{enc}} F_t + b_{\text{enc}}\right) \tag{7}$$

$$F_t' = \sigma\left(W_{\text{dec}} h_t + b_{\text{dec}}\right). \tag{8}$$

The root-mean-square error Er was used as the reconstruction error function [87], and the model was trained by continuously minimizing the reconstruction error between the input and output. The parameters were continuously optimized using the gradient descent algorithm, and the model was made to converge by the iterative algorithm

$$\text{Er} = \frac{1}{m}\sum_{t=1}^{2}\sum_{i=1}^{m}\| f'_{ti} - f_{ti} \|_2^2. \tag{9}$$

To ensure the consistency of the model parameters during the migration from the Siamese AE parameters to the GAT network, we replaced the feature extraction layer of the first layer of the GAT with FC. The change was that the bias parameter b_{enc} was added to enhance the fit of the network. The above-noted coefficient exhibited the following changes:

$$e_{ij} = \alpha\left(W_{\text{enc}}\vec{h}_i^l + b_{\text{enc}}, W_{\text{enc}}\vec{h}_j^l + b_{\text{enc}}\right). \tag{10}$$

The weight parameter W_{enc} and bias parameters b_{enc} of the encoder obtained from the Siamese AE training were saved as the initialization parameters of the GAT FC layer. We calculated the loss $\mathcal{L}_q$ of the labeled sample using the cross-entropy loss function $\mathcal{L}$, and samples were labeled by visual interpretation

$$\mathcal{L} = \sum_{q=1}^{N_S} \mathcal{L}_q \tag{11}$$

where N_S is the number of labeled nodes. We used the gradient descent algorithm for supervised training.

After running the model for several epochs, a softmax classifier was used to complete the binary classification of all node predictions and generate a change result. The final task of this article was to predict whether each pixel belongs to a changed or unchanged class. By using this method, we achieved binary classification of all nodes. Since each node corresponds to an object in the image and all the image objects contain all the pixels, we have accomplished the prediction of the classes of all the pixels.

IV. Result

A. Dataset and Implementation Details

WorldView-2 images from November 2012 and Pleiades images from August 2013 from the East Lake New Technology Development Zone in Wuhan, China, were used for the experimental data. WorldView-2 was launched by the United States in October 2009. It stays in a sun-synchronous orbit, approximately 770 km above the Earth. The French Pleiades satellites consist of two identical satellites, Pleiades 1 and Pleiades 2. These satellites were successfully launched into their own sun-synchronous orbits in 2011 and 2012. The above data included red, green, blue,

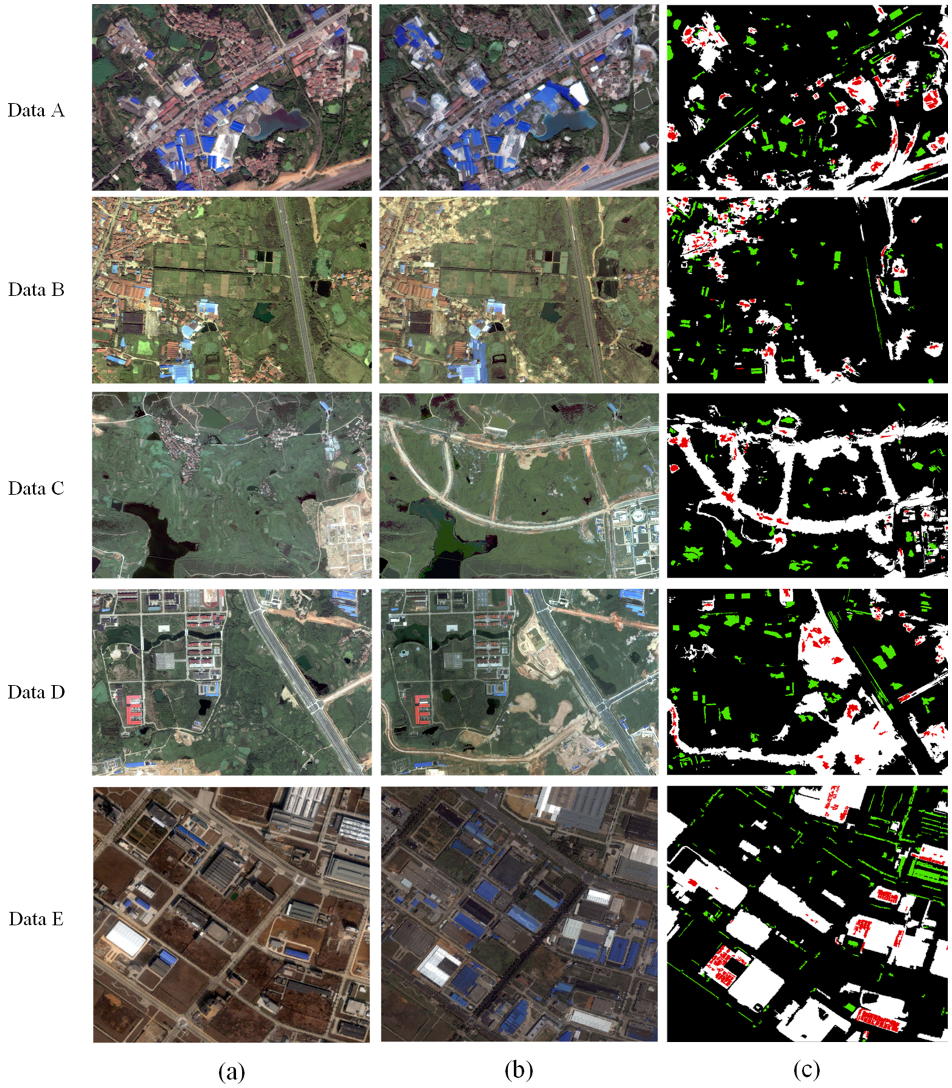


Fig. 6. From top to bottom are four datasets A, B, C, D, and E. (a)–(c) are prechange image, postchange image, and reference change map generated by visual interpretation. The red and green pixels are the changed and unchanged samples in the training set in column (c).

near-infrared, and panchromatic bands. The spatial resolutions of the panchromatic and multispectral bands were 0.5 m and 2 m, respectively. The Pansharp algorithm fused multispectral and panchromatic bands, and the polynomial model was used to register the 2012 and 2013 images. The article selected four sets of experimental data, named Data A, B, C, and D. The sizes of Data A, B, C, and D are 2232 × 1524, 2158 × 1517, 3476 × 2456, and 3020 × 1994 pixels, respectively. To verify the validity of the segmentation scale for different data sources, we added Data E from the multitemp scenes Wuhan dataset (MtS-WH) [22]. The MtS-WH dataset was obtained by IKONOS sensors in February 2002 and June 2009. It includes red, green, blue, and near-infrared bands with a spatial resolution of 1 m. The size of Data E is 1000 × 1000 pixels. The experimental data and reference data are shown in Fig. 6.

We compared the prechange and postchange image objects at corresponding positions via manual interpretation and used the set of changed image objects as reference data. The ratios of changed to unchanged pixels in the reference data for Data A, B, C, D, and E are 0.22, 0.14, 0.30, 0.29, and 0.34, respectively. In addition, similar to some remote sensing image analysis methods based on GNNs, we selected some of the nodes as the training set and the rest as the test set [16], [88], [89], [90]. For Data A, B, C, D, and E, the division ratios of the training set and test set were 300:10 579, 200:7806, 3001: 2871, 300:9805, and 400:17 589, respectively. Our experiments were based on the Windows 10 operating system and an Intel(R) Core (TM) i5-6200U CPU @ 2.30 GHz processor with 8 GB of running memory. The programming language used was Python, and the deep learning framework used was PyTorch.

We used the GMSS to segment multitemporal remote sensing images. The segmentation scale of the bottom layer of the prechange image was set to 25 or 20, the compactness factor was 0.5, and the shape factor was 0.1. The segmentation results are shown in Fig. 7, and the segmentation parameters are shown in Table I. We manually selected the changed objects as positive samples and the unchanged objects as negative samples and kept the ratio of positive to negative samples consistent. Our

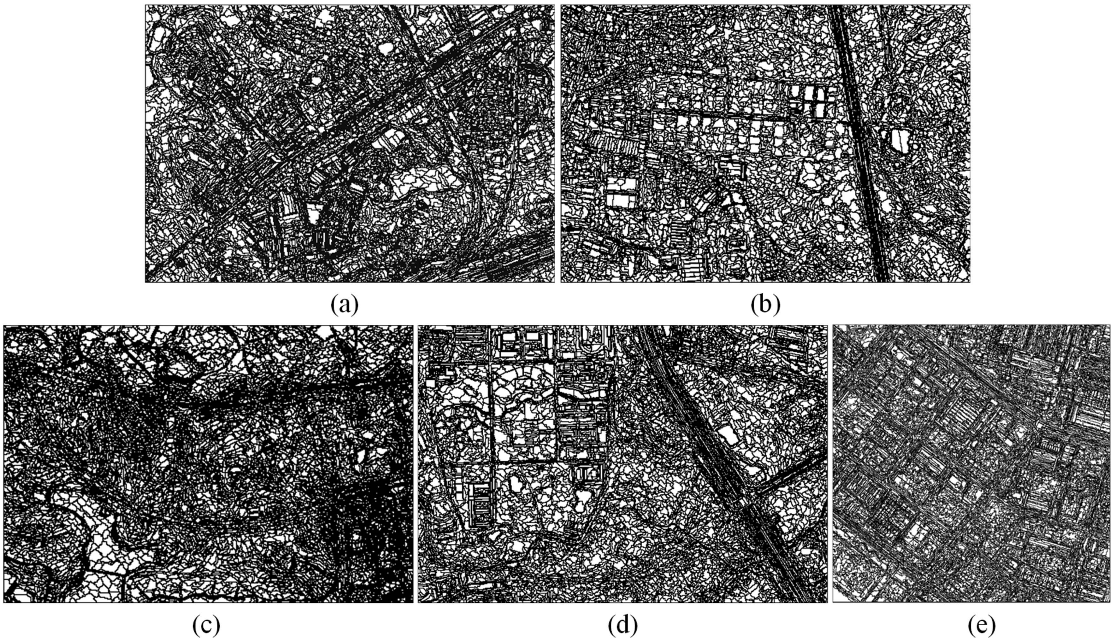


Fig. 7. Segmentation results. (a)–(e) are data A, B, C, D, and E. The black lines represent the boundaries between segmented image objects.

TABLE I
SEGMENTATION PARAMETERS

Data	Number of pixels	Segmentation scale	Shape	Compactness	Layers	Objects	Samples
A	3401568	25,40,55,70	0.1	0.5	4	10579	300
B	3273686	20,35,50,65	0.1	0.5	4	7806	200
C	8537056	25,40,55,70	0.1	0.5	4	12871	300
D	6021880	25,40,55,70	0.1	0.5	4	9805	300
E	1000000	25,40,55,70	0.1	0.5	4	17589	400

acceptable minimum area of a changed object was 50 m^2. After change detection, we merged adjacent change image objects and removed image objects that were smaller than 50 m^2. The proportions of samples to the total for the four datasets were 2.84%, 2.56%, 1.55%, and 3.06%, respectively. The above proportions ensured that the number of samples was much smaller than the number of unlabeled samples, which was conducive to an objective description of the effectiveness and usefulness of the algorithm. The numbers of changed image objects annotated for Data A, B, C, D, and E are 150, 100, 150, 150, and 200, respectively. The numbers of unchanged image objects annotated for Data A, B, C, D, and E were the same as those of the changed image objects.

Consistent with most related references [75], [79], [91], the changed and unchanged image objects of our reference data were also obtained by manual visual interpretation. We formed comparison groups of multitemporal image objects and visually compared the differences among the groups one by one. If we found a significant difference, we labeled it a changed image object. Significant differences were generated due to the conversion among vegetation, built-up land, and water, and the changed image objects are shown in Fig. 8. A portion of the changed image objects were labeled training samples, and image groups that did not change significantly were labeled unchanged image objects; the unchanged images are shown in Fig. 9. The multitemporal image objects are combinations of vegetation, water, and built-up land, as shown in Fig. 9(a), (b), and (c), respectively.

B. Comparison Methods

We used the proposed SA-GAT for change detection on all the experimental data. In addition, the authors compared the results of our method with those of similar methods, including the unsupervised change detection method DCST [92], the semi-supervised machine learning method TSVM [93], the semi-supervised machine learning method TSVM with deep features (TSVM-VGG) [94], the standard GAT method without Siamese AE optimization [18], the GCN semi-supervised change detection method [16], and the graph sample and aggregate-attention (SAGE-A) change detection method [95]. DCST uses CST to construct a 1-D difference feature space of the image object and obtains the binary segmentation threshold through the maximum expectation and Bayesian minimum error. The TSVM computes the absolute difference of each feature and implements binary classification in a multidimensional feature space via hyperplanes. In TSVM-VGG, a second layer of depth features extracted by the classical visual geometry group (VGG) is added

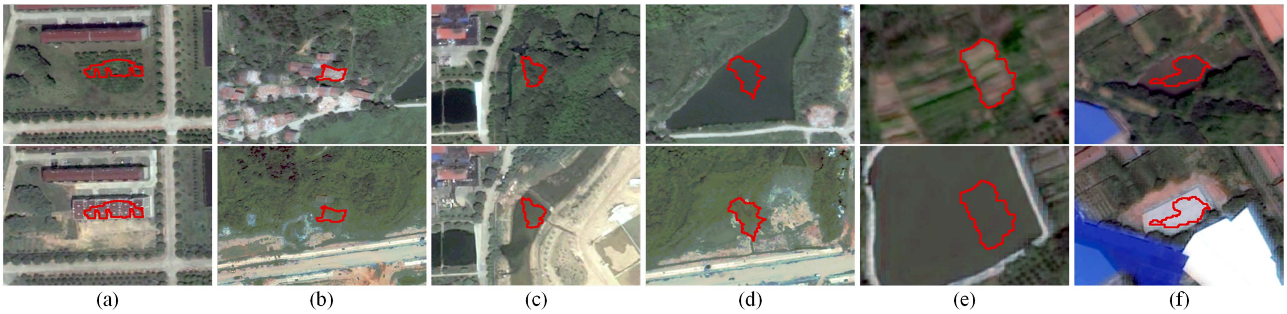


Fig. 8. Changed image objects. (a)–(f) are vegetation to built-up area, vegetation to built-up area, vegetation to built-up area, vegetation to built-up area, built-up area to vegetation, vegetation to water, water to vegetation, built-up area to water, and water to built-up area. The red polygons are the regions of interest.

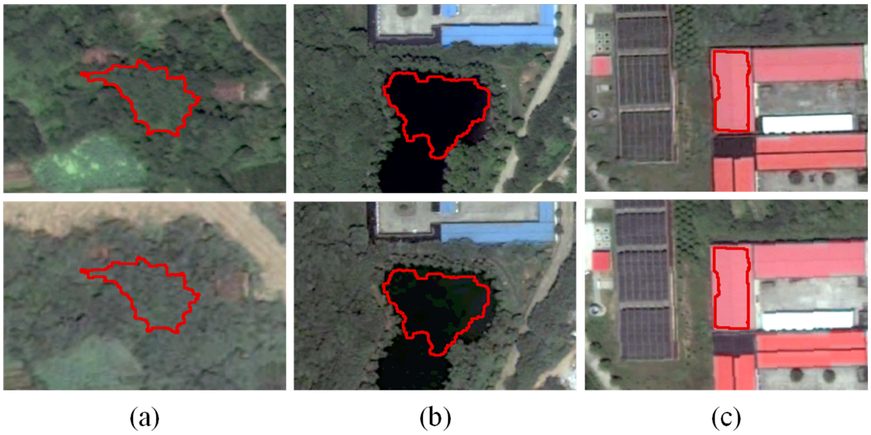


Fig. 9. Unchanged image objects. (a)–(c) are unchanged vegetation, water, and built-up areas, respectively. The red polygons are the regions of interest.

to the feature space, and the remaining steps are the same as those in TSVM. GAT is the same method as ours, except that Siamese AE optimization is not used. Conversely, for GCN and SAGE-A, the corresponding GNNs are used to replace the GAT of the proposed methods; otherwise, the GCNs and SAGE-A do not differ from each other. The DCST, TSVM, and TSVM-VGG are traditional machine learning methods and are fundamentally different from the proposed methods in terms of feature representation. In addition, the same labeled samples were used for all methods to ensure the fairness of the experiments.

To quantitatively evaluate the performance, precision, recall, overall accuracy (OA), and $F1$-score were used as accuracy analysis metrics. They were calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{12}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{13}$$

$$\text{OA} = \frac{(\text{TP}+\text{TN})}{(\text{TP}+\text{TN}+\text{FP}+\text{FN})} \tag{14}$$

$$F1 = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}. \tag{15}$$

TP represents the true positive, which refers to the part correctly predicted as changed areas. FP represents the false positive, which refers to the part wrongly predicted as changed areas. TN represents the true negative, which refers to the part correctly predicted as unchanged areas. FN represents the false negative, which refers to the part wrongly predicted as unchanged areas.

C. Results and Analysis

We accomplished change detection for Data A, B, C, D, and E using DCST, TSVM, TSVM-VGG, GAT, GCN, SAGE-A, and SA-GAT. The similarity step $s = 1$, the number of hidden layer units $d = 16$, and the number of training epochs $i = 400$. The change detection results for DCST, TSVM, TSVM-VGG, GAT, GCN, SAGE-A, and SA-GAT are shown in Fig. 10. The results of the DCST included more trivial image objects, as shown in the red box in Fig. 10(a). The integrity of the detection results of the TSVM and TSVM-VGG methods was better than that of the DCST method, but the change detection results in the first dataset were poorer, and the false alarms were more serious, as shown in the red box in Fig. 10(b) and (c). GAT had serious false detections in Data B and C. Specifically, a large amount of unchanged vegetation adjacent to roads and bare soil was incorrectly classified as changed, as shown in the red box in

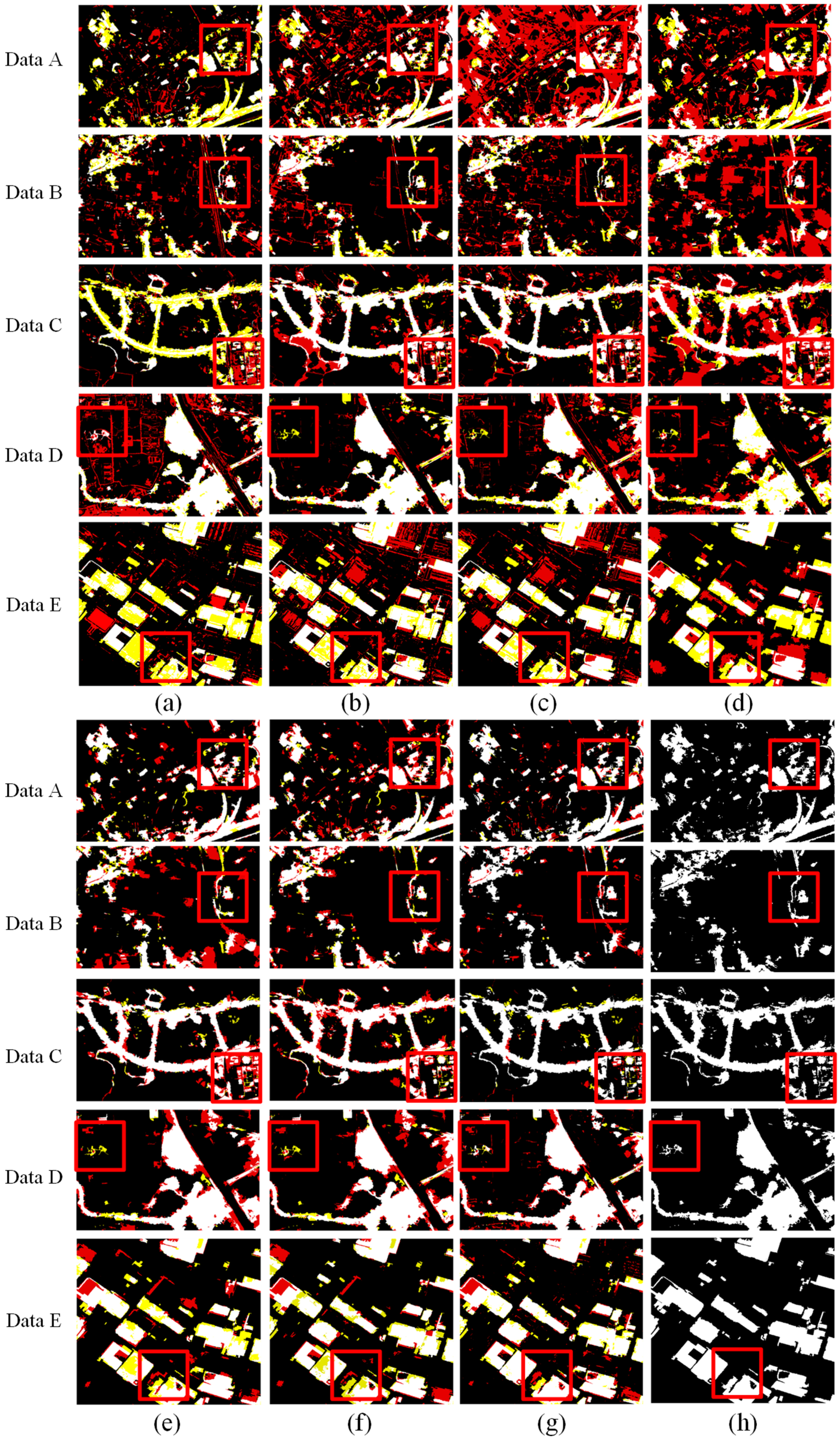


Fig. 10. Change detection results. (a)–(h) DCST, TSVM, TSVM-VGG, GAT, GCN, SAGE-A, SA-GAT, and reference data. Red boxes highlight interesting areas where models diverge. These regions of interest point out the specialties of the proposed method. The errors of models are highlighted with different colors, the red for false positive and the yellow for false negative.

Fig. 10(d). GCN and SAGE-A had good detection results for all four datasets. The accuracy of the road and building boundary extraction was slightly worse than that of the proposed method, as shown in the red boxes in Fig. 10(e), (f), (g), and (h). SA-GAT could effectively identify changes in houses, vegetation, bare land, and water, especially in roads and bare soil. Overall, the proposed method could achieve high-precision change region extraction for all four datasets, which proved its effectiveness and applicability for change detection in high-resolution remote sensing images.

TABLE II
ACCURACY EVALUATION RESULTS

Data	Method	Supervision	Precision (%)	Recall (%)	OA (%)	F1(%)
A	DCST	No	65.18	54.12	86.04	59.14
	TSVM	Semi	45.23	76.97	78.34	56.98
	TSVM+CNN	Semi	27.93	84.36	57.30	41.97
	GAT	Semi	41.46	58.90	76.81	48.67
	GCN	Semi	72.67	86.52	91.41	78.99
	SAGE-A	Semi	67.40	89.95	90.20	77.06
	SA-GAT	Semi	84.23	96.57	95.99	89.98
B	DCST	No	46.06	66.53	86.65	54.43
	TSVM	Semi	52.14	71.88	88.72	60.44
	TSVM+CNN	Semi	45.62	71.65	86.37	55.75
	GAT	Semi	26.33	72.50	72.39	38.63
	GCN	Semi	48.25	81.96	87.31	60.74
	SAGE-A	Semi	67.49	87.41	93.45	76.17
	SA-GAT	Semi	77.46	97.54	96.30	86.35
C	DCST	No	67.23	52.60	83.17	59.02
	TSVM	Semi	66.30	88.76	87.01	75.90
	TSVM+CNN	Semi	68.74	90.78	88.36	78.24
	GAT	Semi	42.59	76.66	70.81	54.76
	GCN	Semi	76.70	93.40	91.94	84.23
	SAGE-A	Semi	73.45	93.09	90.65	82.11
	SA-GAT	Semi	94.28	92.07	96.89	93.16
D	DCST	No	70.56	75.95	87.59	73.16
	TSVM	Semi	79.90	86.10	92.08	82.88
	TSVM+CNN	Semi	61.79	84.08	84.88	71.23
	GAT	Semi	64.00	73.18	84.86	68.28
	GCN	Semi	70.17	94.41	89.82	80.51
	SAGE-A	Semi	71.36	93.06	90.14	80.78
	SA-GAT	Semi	78.46	95.97	93.24	86.34
E	DCST	No	66.60	50.16	80.80	57.22
	TSVM	Semi	60.04	72.16	80.58	65.54
	TSVM+CNN	Semi	63.48	72.28	82.26	67.60
	GAT	Semi	60.02	59.57	79.50	59.79
	GCN	Semi	74.05	74.28	86.75	74.16
	SAGE-A	Semi	77.18	74.97	87.92	76.06
	SA-GAT	Semi	80.92	84.35	90.90	82.60

TABLE III
AVERAGE OF ACCURACY EVALUATION RESULTS

Method	Supervision	Precision (%)	Recall (%)	OA (%)	F1(%)
DCST	No	No	66.60	50.16	80.80
TSVM	Semi	Semi	60.04	72.16	80.58
TSVM+CNN	Semi	Semi	63.48	72.28	82.26
GAT	Semi	Semi	60.02	59.57	79.50
GCN	Semi	Semi	74.05	74.28	86.75
SAGE-A	Semi	Semi	77.18	74.97	87.92
SA-GAT	Semi	Semi	80.92	84.35	90.90

We used precision, recall, OA, and $F1$-scores to quantitatively evaluate the accuracy of the change detection results for DCST, TSVM, TSVM-VGG, GAT, GCN, SAGE-A, and SA-GAT. The accuracies of DCST, TSVM, TSVM-VGG, GAT, GCN, SAGE-A, and SA-GAT are shown in Table II. The average accuracy is shown in Table III. The average precision, recall, OA, and $F1$-scores of SA-GAT were 85.38%, 94.27%, 95.41%, and 89.35%, respectively. Compared with those of DCST, TSVM, TSVM-VGG, GAT, GCN, and SAGE-A, the average OA of SA-GAT improved by 10.20%, 8.39%, 13.31%, 19.25%, 5.04%, and 4.68%, respectively. The average $F1$-score improved by 27.24%, 17.20%, 22.35%, 34.01%, 10.84%, and 9.55%, respectively. The DCSTs and TSVMs were easy to implement. However, they could not make full use of the topological features of image objects, resulting in poor detection results. The GAT results were also unsatisfactory because the GAT failed to take full advantage of deep features. The detection results of GCN and SAGE-A were better than those of traditional DCST, TSVM, and TSVM-VGG, but the accuracy was not as good as that of semi-supervised SA-GAT.

Among them, DCST, TSVM, and TSVM-VGG were simpler to extract than the other methods, but they failed to make full use of the topological relationships of image objects, resulting in poor detection. The results of the GAT directly applied to the four experimental datasets were also unsatisfactory because the GAT could not be effectively trained with a few samples. GCN and SAGE-A outperformed traditional DCST and TSVM, but the extraction accuracy was not as high as that of SA-GAT.

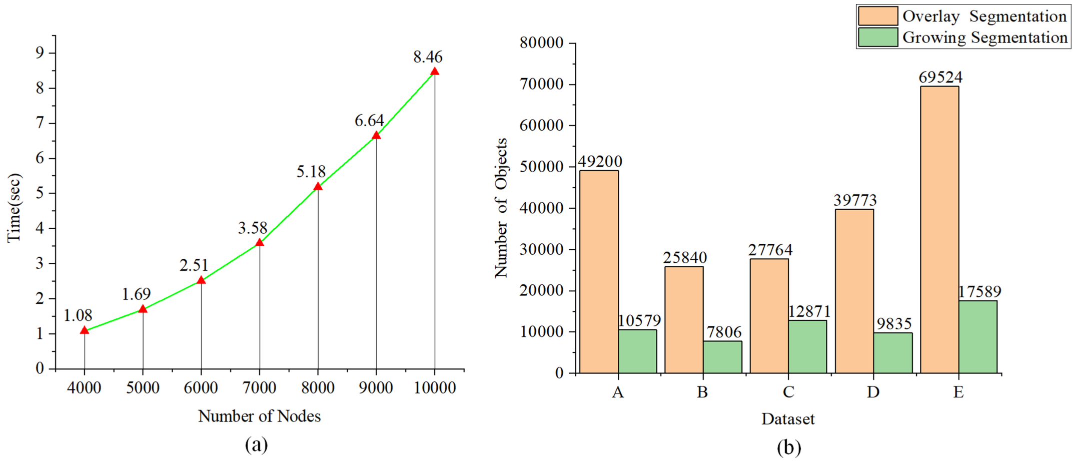


Fig. 11. Experimental results using different network nodes. (a) Comparison of the running times of single epochs with different numbers of nodes. (b) Comparison of image objects of the overlay segmentation and growing segmentation.

Compared with GCN SAGE-A, SA-GAT could better assign weights to adjacent image objects, and the feature vectors pretrained by the Siamese AE were also more conducive to deep feature extraction.

V. Discussion

A. Effect of Image Segmentation Scale

The GAT is a type of direct push learning [40]. The larger the quantity of input data is, the more computing power is digested during GAT training. Therefore, the number of image objects affects the operational efficiency of the GAT. The experimental results using different network nodes are shown in Fig. 11(a). We compared the changes in the running times of single epochs as the number of network nodes increased from 4000 to 10 000. The single-epoch running time is positively correlated with the number of network nodes. For this reason, our segmentation strategy effectively controlled the number of image objects. Moreover, ensuring the effectiveness of the segmentation algorithm improved the operation efficiency of the model.

Specifically, we introduced GMSS to construct multilevel segmentation object layers. First, the optimal segmentation parameters were obtained by the heuristic method [96], and the segmentation scale of the bottom layer of the prechange image was determined to be 25, the compactness factor was 0.5, and the shape factor was 0.1. The postchange image was segmented by the same parameters, and the segmentation result of the prechange image was taken as the base image. The segmentation of the postchange image was constrained by the results of the previous temporal segmentation, and a new object was generated when and only when the postchange image was significantly different from the prechange image. Then, based on the segmentation results at the initial scale, the postchange segmentation layer scale was determined by increasing the amplitude to 15. The number of segmentation layers N depended on the spatial resolution of the images and the size of the ground objects. If N was too small, it was insufficient to obtain the global receptive field feature. A too-large N will cause data redundancy. A total of four object segmentation layers, level0, 1, 2, and 3, were ultimately determined. A comparison of the numbers of image objects in the growing segmentation and overlay segmentation is shown in Fig. 11(b). The increase in segmentation greatly reduced the number of image objects compared with that of the conventional overlay segmentation. A dataset was used to demonstrate the results of growing overlay segmentation and was compared with the methods that did not use the segmentation strategy; the results are shown in Fig. 12. A comparison of the segmentation results, as delineated by the red boxes in Fig. 12(c) and (d), shows that the growing segmentation reduced the incorrect segmentation caused by image registration and spectral differences and that the segmentation results had greater integrity.

According to the eCognition software manual and related references, it is common to set the parameter shape factor w_{shape} to 0.1 and the compactness factor w_{compact} to 0.5 (the default values). This is the basis for setting our parameters w_{shape} and w_{compact}. At the segmentation scale ζ_{scale}, both automatic and manual methods require human involvement, which is why we set the segmentation parameters based on experience. We set the bottom layer of the prechange image to 15, 20, 25, 30, and 35 and used SA-GAT for change detection at different segmentation scales. Using Data A and E as examples, Fig. 13 shows the impact of the segmentation scale, and setting the bottom layer of the prechange image to 25 achieved better change detection accuracy for Data A and E.

The experimental results verify that the segmentation scale has an important impact on the object-oriented change detection method. In addition, the segmentation scale of 25 in this article has good adaptability for high-resolution remote sensing images with spatial resolutions of less than 1 m. How to obtain the optimal segmentation scale automatically is still worth paying attention to.

B. Effect of the Similarity Step

The similarity step s determines the number of virtual connected edges of the nodes, which affects the connectivity of the graph. The details of the edges obtained without the WDSM and

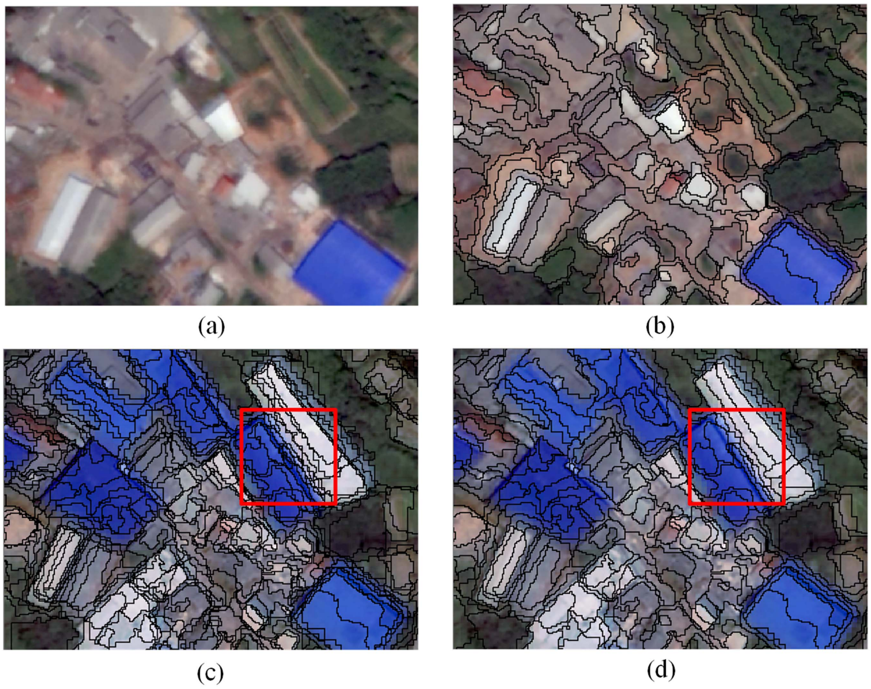


Fig. 12. Comparison of the overlay segmentation and growth segmentation results (a)–(d) are prechange image, prechange segmentation, postchange overlay segmentation, and postchange growing segmentation. The partial images are derived from data A.

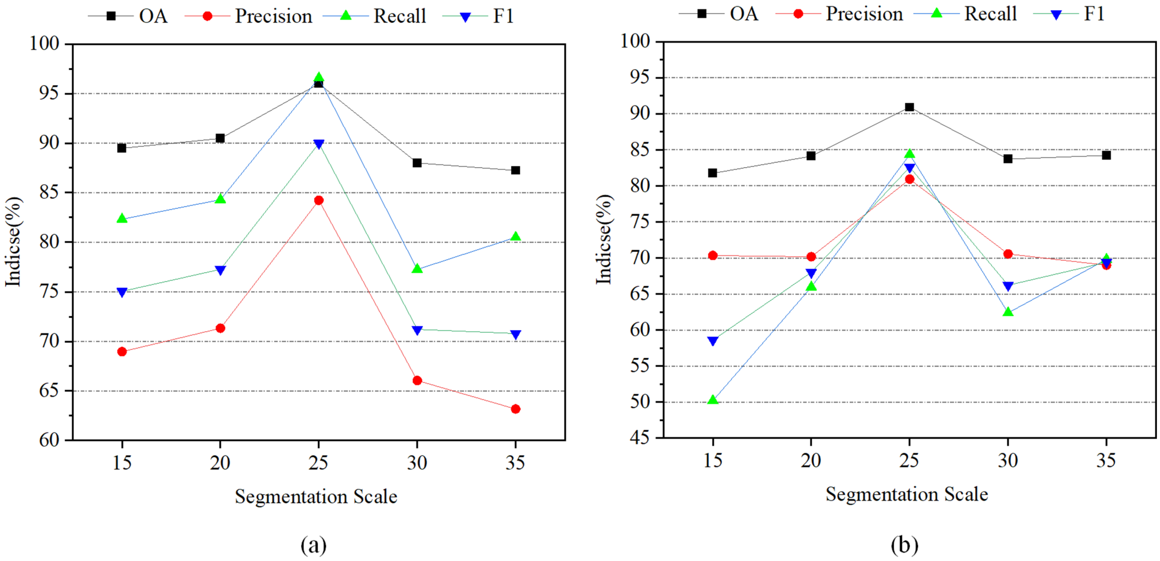


Fig. 13. Impact of the segmentation scale. (a) and (b) Data A and E.

the edges obtained with the WDSM are shown in Fig. 14. As Fig. 14 shows, the edges obtained with the WDSM have significantly better connection density than those obtained without the WDSM. An increase in connection density can improve the feature expression ability of the GAT [18]. The similarity step s of the connection density enhancement method determines the number of strongly connected edges. The other parameters were controlled to be the same, and the change detection results using different similarity steps are shown in Fig. 15. When $s = 0$, no virtual connection edge was established. A few obvious change areas that were not identified could be observed. The reason was that the differences between this node and the surrounding neighboring nodes were too large, which made it difficult to effectively transmit supervised information. When $s = 1$, as shown in the red boxes of Fig. 15(a) and (b), most of the obvious change regions could be effectively extracted, and the precision and recall rates increased, among which the recall rate significantly improved, which proved the effectiveness of the proposed method. A comparison of the change detection accuracy for different numbers of similarity steps is shown in Fig. 16. The

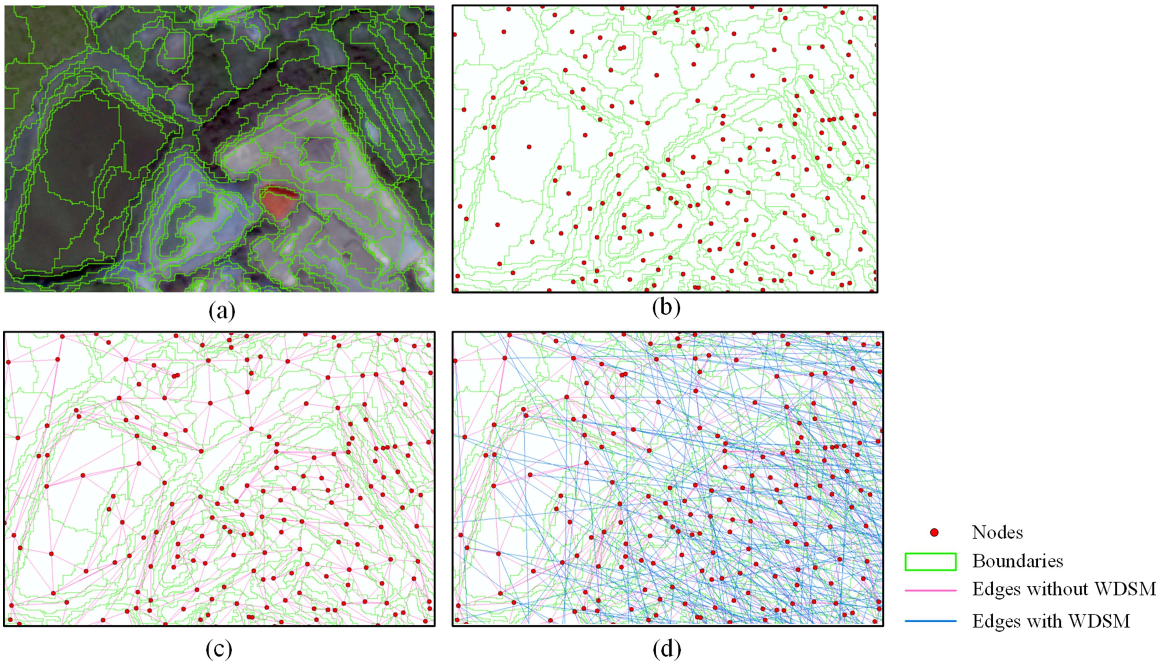


Fig. 14. Details of the edges obtained without WDSM and the edges obtained with WDSM. (a)–(c) are the segmentation result, image object boundaries, connected edges without WDSM, and connected edges with WDSM.

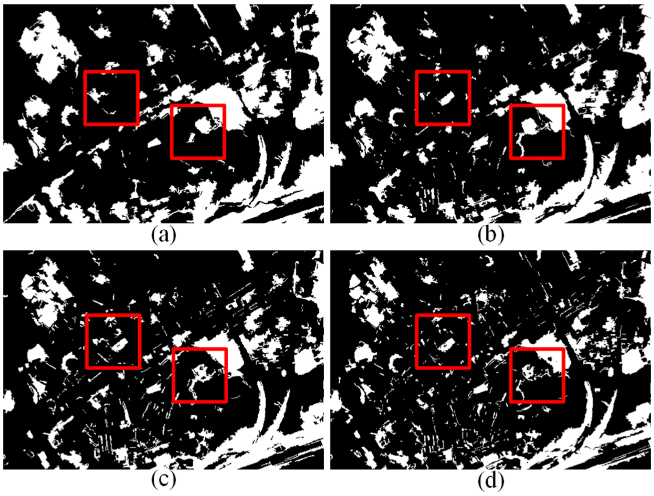


Fig. 15. Change detection results using different similarity steps. (a)–(d) are $s = 0$, 1, 2, and 3, respectively. The s is the similarity step. The partial images are derived from data A.

results showed that the change detection accuracy reached its highest when $s = 1$. When the value continued to increase, as shown in Fig. 15(c) and (d), the number of noisy change image objects increased, and the change detection accuracy decreased, indicating that the value should not be too large.

C. Parameter Analysis of GAT

We used Data A to conduct an experimental analysis on the GAT parameters. The experimental parameters included training epochs i, dropout of notes dr, the number of hidden layer units d, and the number of training samples tr. We first tested the effect of different training epochs on our methods. We fixed $\mathrm{dr} = 0.2$, $d = 16$, $\mathrm{tr} = 200$, and the step size was 100. We changed the training epochs i from 100 to 500. The experimental results using different training epochs are shown in Fig. 17(a). As i increased, the experimental accuracy increased rapidly at the beginning and then stabilized. We tested the effect of different dropout durations on our method. We fixed $i = 400$, $d = 16$, $\mathrm{tr} = 300$, and the step size was 0.1. We changed the dropout from 0 to 0.3. The experimental results using different dropouts are shown in Fig. 17(b). When $\mathrm{dr} = 0.1$ and $\mathrm{dr} = 0.2$, the experiment achieved a more desirable accuracy, with a 5%–10% improvement in both accuracy and recall compared to $\mathrm{dr} = 0$. However, the dropout should not be too large, and the precision will drop rapidly if it exceeds 0.2. We tested the effect of different numbers of hidden layers on the experimental results. We fixed $i = 400$, $\mathrm{dr} = 0.2$, $\mathrm{tr} = 300$, and the multiplicative step size $= 2$, and changed d from 4 to 32. The experimental results using different numbers of hidden layers are shown in Fig. 17(c). With the increase in the number of hidden layers, the accuracy and recall rate of the data gradually increased, which indicated that the model's difference feature extraction ability was enhanced. At $d = 16$, the accuracy of change detection peaked and then showed a downward trend. The reason was that d determined how much the model compressed the multidimensional features. When d was too small, too much information was lost in the input data. A too-large d led to more training parameters, which was not conducive to model training.

In addition, we tested the effect of sample size on the accuracy of the model. The experimental results using different numbers of samples are shown in Fig. 17(d). We fixed $i = 400$, $d = 16$, $\mathrm{dr} = 0.2$, and the step size $= 100$, and changed the number of samples from 100 to 400. In the initial stage, the accuracy

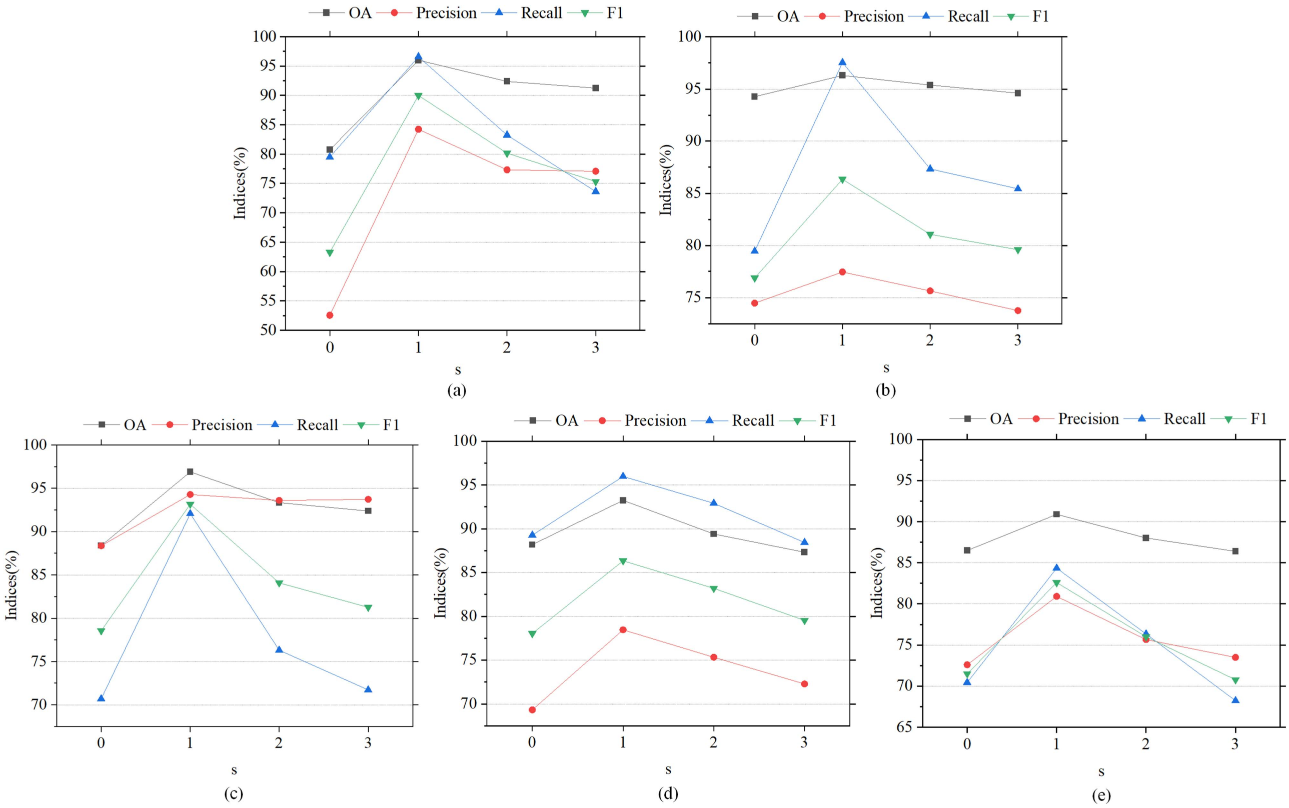


Fig. 16. Change detection accuracy comparison for similarity steps. (a)–(e) are data A, B, C, D, and E.

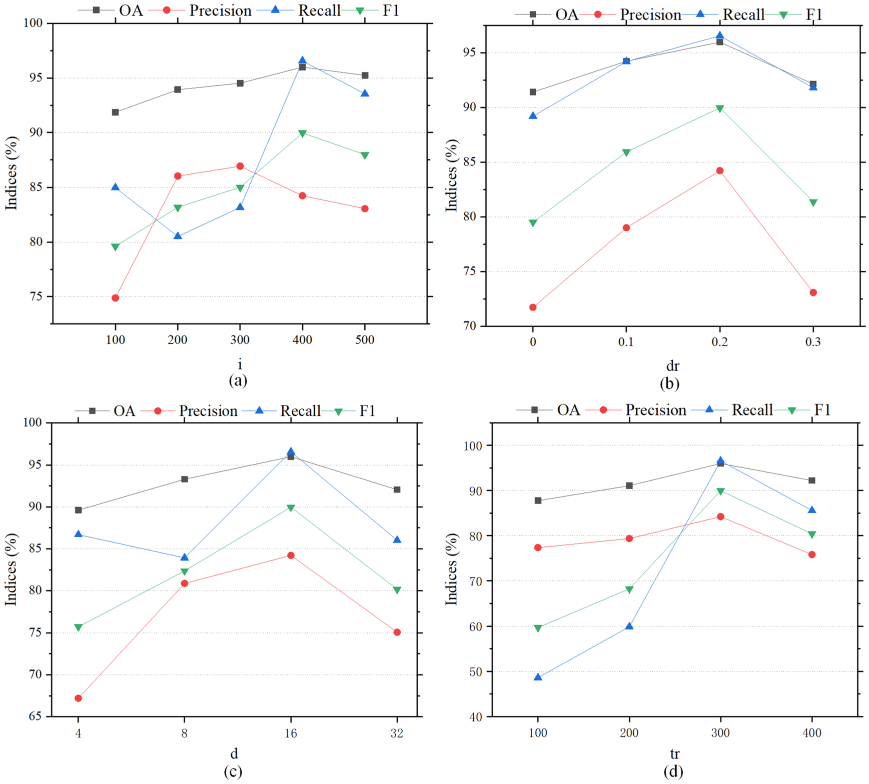


Fig. 17. Parameter analysis of the model training. (a)–(d) are different training epochs, different dropouts, different numbers of hidden layers, and different numbers of samples.

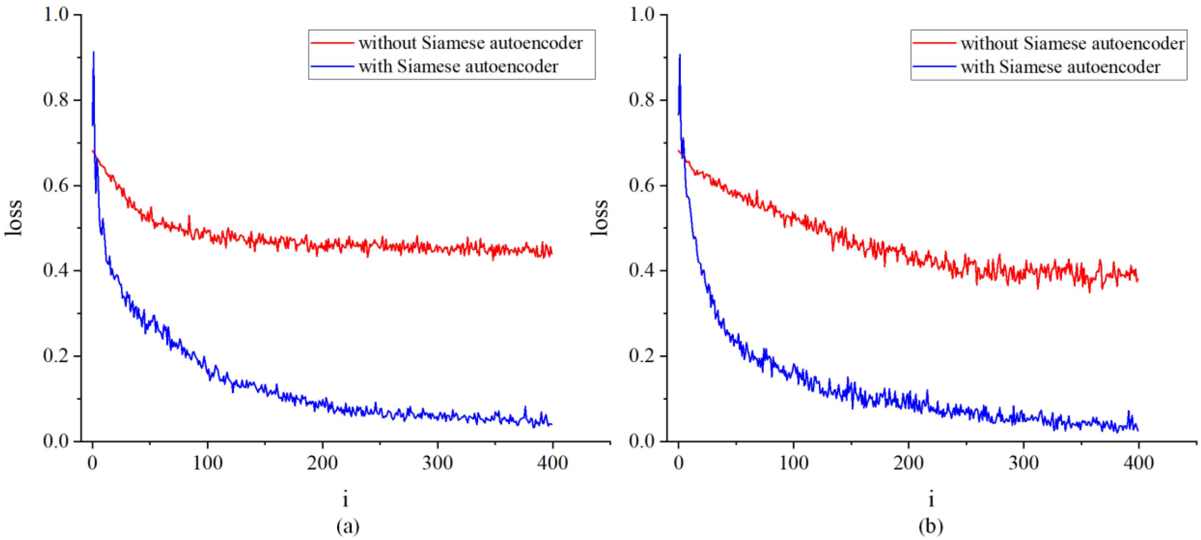


Fig. 18. Analysis of the model training gradient loss value. (a) and (b) are data A and B, respectively.

increased sharply, peaked at tr $=$ 300, and then showed an up-and-down trend. The analysis showed that too many manual samples would lead to an increase in the possibility of incorrect samples, which would affect the experimental accuracy. Therefore, while the number of samples satisfied the model training requirements, the sample quality should also be guaranteed.

D. Impact of Siamese AE Pretraining

We compared the impact of Siamese AE pretraining on the performance of the GAT using Data A and B as examples. In the model training, the training gradient loss represents the error between the prediction result of the training set in the model and the real result, and its trend reflects the model training status and completion. The curve change in the two methods with the number of training epochs i during the model training is shown in Fig. 18. With the increase in training iterations, the loss of the GAT pretrained without the Siamese AE decreased slowly, and at $i = 400$, the loss still fluctuated around approximately 0.4. This result indicated that the model was not trained efficiently and failed to learn effectively. The loss of GAT signal using the Siamese AE during pretraining decreased rapidly; at $i = 200$, loss decreased to less than 0.1, and there was still a small decrease in the interval [200, 400]. The training feature vector after the Siamese AE was involved in pretraining fits the GAT adequately, which helped speed up the training process and the model was learned effectively.

E. Ablation Experiment

We designed ablation experiments for the proposed method using four datasets. To improve the change detection performance of the GAT, we proposed two improvements in this article.

The WDSM method was proposed to enhance the connection strength of nodes, and the Siamese AE was proposed to obtain high-dimensional features. The ablation experiments analyzed

TABLE IV
ACCURACY OF THE ABLATION EXPERIMENT

Data	WDSM	Siamese autoencoder	OA (%)	F1 (%)
A	–	–	76.81	48.67
	–	√	80.75	63.27
	√	–	87.72	72.32
	√	√	95.99	89.98
B	–	–	72.39	38.63
	–	√	94.27	76.88
	√	–	92.60	73.77
	√	√	96.30	86.35
C	–	–	70.81	54.76
	–	√	88.38	78.55
	√	–	95.33	89.53
	√	√	96.89	93.16
D	–	–	84.86	68.28
	–	√	88.19	78.04
	√	–	94.58	88.42
	√	√	93.24	86.34
E	–	–	79.50	59.79
	–	√	86.50	71.51
	√	–	85.09	70.74
	√	√	90.90	82.60
Average	–	–	77.08	57.68
	–	√	87.72	76.08
	√	–	93.76	85.4
	√	√	95.62	89.67

TABLE V
RESULTS OF THE GENERALIZABILITY EXPERIMENT

Data		A	B	C	D	E
A	OA (%)	95.99	91.94	92.00	91.65	83.10
	F1 (%)	89.98	76.19	80.97	83.30	63.34
E	OA (%)	85.92	85.02	80.75	88.13	90.90
	F1 (%)	70.78	71.95	68.68	74.76	82.60

TABLE VI
Time Complexity (s) and GPU Memory (GB) Complexity

	SA-GAT			TSVM+CNN		
	Data A	**Data B**	**Data E**	**Data A**	**Data B**	**Data E**
Segmentation	260.27	243.77	236.64	260.27	243.77	236.64
Graph construction	68.19	61.45	83.66	×	×	×
Training and prediction	3762.82	1953.27	8362.38	105.79	101.33	94.39
Total	4091.28	2258.49	8682.68	260.27	243.77	236.64
GPU memory	2.24	1.68	4.46	3.30	3.21	2.13

the changes in accuracy by controlling for how the improvement measures were used. The experiments included the methods using WDSM and Siamese encoders, the methods without WDSM and Siamese encoders, and the methods using WDSM or Siamese encoders. The accuracy of the ablation experiment is shown in Table IV.

The OA and $F1$-scores of change detection methods using only the WDSM or Siamese encoder were higher than those of the methods not using them. The OA and $F1$-scores of the change detection method using the WDSM and Siamese encoders were higher than those of the methods using only one of them. The average OA and $F1$-scores of our method were 95.41% and 89.35%, respectively, which were 19.25% and 34.01% greater than those of the methods that did not use any of the improvements. The accuracy of the ablation experiment demonstrates that our use of the WDSM and Siamese encoder to improve change detection using GAT was effective.

F. Generalizability Analysis

In order to verify the generalizability of the method, we apply the model of data A trained by SA-GAT to the prediction of the other data. Considering the differences in data sources, Data E was subjected to the same experiments. The results of the generalizability experiment are shown in Table V. The model for Data A was applied to the other data and the prediction accuracies all decreased. Data E and Data A belong to different sensors from each other, and the decrease in prediction accuracy is even more pronounced. The same conclusion can be reached for the experiment with Data E. It can be seen that our method is generalizable to data from the same sensor and less generalizable to data from nonsimilar sensors.

G. Complexity Analysis

Complexity analysis can evaluate the efficiency of each module of the algorithm. The time complexity analysis records the running time of each module. The graphics processing unit (GPU) memory complexity analysis records the running memory of the neural network [97], [98], [99]. Data A, B, and E are chosen as the experimental data. Data E has a different data source compared to Data A and B, which can increase the generalizability. TSVM+CNN was used as a comparison experiment without graph construction. The training and prediction time of TSVM+CNN includes the time to construct deep features of CNN and the training and prediction time of TSVM. The GPU memory of TSVM+CNN is occupied by CNN. The time complexity (s) and GPU memory (GB) complexity are shown in Table VI. Among them, the time to construct deep spatial features of CNN is 26.52 s, 28.52 s, and 22.53 s, respectively. The time complexity of SA-GAT is much higher than that of TSVM+CNN. The running time of segmentation and graph construction are on average 7% and 2% of the total. Graph construction increases the time complexity, but the training and prediction consume more time. The experiments on Data E also illustrate that the training and prediction time is proportional to the number of nodes. The GPU memory usage is similar. The construction of CNN deep features requires more GPU memory.

H. Limitation of the Proposed Approach

Compared with that of other methods, the superiority of this method was proven, but there were still some shortcomings.

1) The proposed method requires manual selection of samples, and samples cannot be shared among different datasets, which means that different data need to be reselected for different samples. This was because, such as most GNNs, the proposed method is essentially a kind of transductive learning. That is, learning from the observed training dataset and then predicting the label of the test dataset was not conducive to batch data processing. This problem was similarly challenged in other article change detection papers using GNNs [15], [16]. An unsupervised model can also mine its own supervised information from large-scale unsupervised data. It trains the network with this constructed supervised information and can learn representations that are valuable for downstream tasks [100]. In the future, we will discuss the possibility of automatic sample selection. It is used to implement an unsupervised change detection model to improve the efficiency of the application.
2) This method is suitable for high-resolution, multiband satellite images and does not work well for RGB images with low resolution and lack of an NIR band. In addition, the proposed method expands the receptive field of image objects through multilevel incremental segmentation and controls the number of image objects to remain within an operable range. However, the optimal number of segmentation layers and segmentation scale parameters are

determined by heuristic methods, and no standard determination method is available [101]. This is because the number of segmentation layers and the segmentation scale depended on the spatial resolution of the image and the size of the ground object.

VI. Conclusion

This article proposed a semi-supervised land-use/land-cover object-oriented change detection method based on an SA-GAT network via high-resolution remote sensing to solve the problems of connection sparsity and inadequate sample feature mining in remote sensing change detection via GAT. We experimentally verified the effectiveness of the proposed method and obtained the following conclusions.

1) Too many nodes could lead to memory overrun of the GNN. The growing multilevel image segmentation strategy proposed in this article reduced the number of objects while ensuring segmentation accuracy and extracting multilevel features, which was conducive to improving operational efficiency and accuracy.
2) When using GNNs for change detection problems, it is important to consider whether the supervised information of the model can be effectively propagated. The similarity matching method was used to increase the virtual connection edge, which could improve the connection density of the graph, which was conducive to the dissemination of supervised information.
3) Through network layer replacement, Siamese AE pretraining, and fine-tuning with a few labeled samples, the GAT quickly extracts the high-dimensional features of nodes, which helped improve the model training efficiency and accuracy. Through experiments, this article proved that the proposed method was effective for semi-supervised change detection in high-resolution remote sensing images using a few samples.

The proposed method can serve the natural resources management and provide them with a reference technology for land-use/land-cover change monitoring. At the same time, the optimization of pretraining and graph connection density enhancement provides a new idea for how to effectively apply GNNs to remote sensing image change detection. This is undoubtedly beneficial to researchers of GNNs. In the future, we will develop a method for changing region classification that can directly use the output data from this article. Finally, the code and data for our project, after being organized, will be made available on the web.

References

[1] P. Du and S. Liu, "Change detection from multi-temporal remote sensing images by integrating multiple features," *J. Remote Sens.*, vol. 16, no. 4, pp. 663–677, Oct. 2012.

[2] M. Jiang, X. Zhang, Y. Sun, W. Feng, Q. Gan, and Y. Ruan, "AFSNet: Attention-guided full-scale feature aggregation network for high-resolution remote sensing image change detection," *GIScience Remote Sens.*, vol. 59, no. 1, pp. 1882–1900, Nov. 2022.

[3] Z. Wang et al., "Object-based change detection for vegetation disturbance and recovery using Landsat time series," *GIScience Remote Sens.*, vol. 59, no. 1, pp. 1706–1721, Oct. 2022.

[4] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Change detection techniques," *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2401, Jan. 2004.

[5] G. Chen, G. J. Hay, L. M. Carvalho, and M. A. Wulder, "Object-based change detection," *Int. J. Remote Sens.*, vol. 33, no. 14, pp. 4434–4457, Jul. 2012.

[6] Z. Shengyin, A. Ru, and Z. Meiru, "Urban change detection by aerial remote sensing using combining features of pixel-depth-object," *Acta Geodaetica et Cartographica Sinica*, vol. 48, no. 11, pp. 1452–1463, Nov. 2019.

[7] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu, "A review of supervised object-based land-cover image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 277–293, Aug. 2017.

[8] X. Tong Yang, H. Liu, and X. Gao, "Land cover changed object detection in remote sensing data with medium spatial resolution," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 38, pp. 129–137, Jun. 2015.

[9] M. Janalipour and M. Taleai, "Building change detection after earthquake using multi-criteria decision analysis based on extracted information from high spatial resolution satellite images," *Int. J. Remote Sens.*, vol. 38, no. 1, pp. 82–99, Apr. 2017.

[10] S. Aykent and T. Xia, "Gbpnet: Universal geometric representation learning on protein structures," in *Proc. 28th ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2022, pp. 4–14.

[11] J. Sieg, F. Flachsenberg, and M. Rarey, "In need of bias control: Evaluating chemical data for machine learning in structure-based virtual screening," *J. Chem. Inf. Model.*, vol. 59, no. 3, pp. 947–961, Mar. 2019.

[12] F. Zhang, B. Zhou, C. Ratti, and Y. Liu, "Discovering place-informative scenes and objects using social media photos," *Roy. Soc. Open Sci.*, vol. 6, no. 3, Mar. 2019, Art. no. 181375.

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

[14] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. V. D. Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *Proc. Eur. Semantic Web Conf.*, 2018, pp. 593–607.

[15] J. Wu et al., "A multiscale graph convolutional network for change detection in homogeneous and heterogeneous remote sensing images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, Feb. 2021, Art. no. 102615.

[16] S. Saha, L. Mou, X. X. Zhu, F. Bovolo, and L. Bruzzone, "Semisupervised change detection using graph convolutional network," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 4, pp. 607–611, Apr. 2021.

[17] J. Qu, Y. Xu, W. Dong, Y. Li, and Q. Du, "Dual-branch difference amplification graph convolutional network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2022, Art. no. 5519912.

[18] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.

[19] S. Min, Z. Gao, J. Peng, L. Wang, K. Qin, and B. Fang, "STGSN—A spatial–temporal graph neural network framework for time-evolving social networks," *Knowl. Based Syst.*, vol. 214, Jan. 2021, Art. no. 106746.

[20] N. Mu, D. Zha, Y. He, and Z. Tang, "Graph attention networks for neural social recommendation," in *Proc. IEEE 31st Int. Conf. Tools Artif. Intell.*, 2019, pp. 1320–1327.

[21] Y. Liu, K. Zeng, H. Wang, X. Song, and B. Zhou, "Content matters: A GNN-based model combined with text semantics for social network cascade prediction," in *Proc. Pacific-Asia Conf. Knowl. Discov. Data Mining*, 2021, pp. 728–740.

[22] A. Roy, K. K. Roy, A. Ahsan Ali, M. A. Amin, and A. Rahman, "SST-GNN: Simplified spatio-temporal traffic forecasting model using graph neural network," in *Proc. Pacific-Asia Conf. Knowl. Discov. Data Mining*, 2021, pp. 90–102.

[23] C. Zhang, J. J. Q. James, and Y. Liu, "Spatial-temporal graph attention networks: A deep learning approach for traffic forecasting," *IEEE Access*, vol. 7, pp. 166246–166256, 2019.

[24] C. Zhang, S. Zhang, J. J. Q. James, and S. Yu, "FASTGNN: A topological information protected federated learning approach for traffic speed forecasting," *IEEE Trans. Ind. Inform.*, vol. 17, no. 12, pp. 8464–8474, Dec. 2021.

[25] D. Peng, Y. Zhang, and H. Guan, "End-to-end change detection for high resolution satellite images using improved UNet++," *Remote Sens.*, vol. 11, no. 11, Jun. 2019, Art. no. 1382.

[26] M. Wang, H. Zhang, W. Sun, S. Li, F. Wang, and G. Yang, "A coarse-to-fine deep learning based land use change detection method for high-resolution remote sensing images," *Remote Sens.*, vol. 12, no. 12, Jun. 2020, Art. no. 1933.

[27] M. Zhang and W. Shi, “A feature difference convolutional neural network-based change detection method,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7232–7246, Oct. 2020.

[28] A. Oliver, A. Odena, C. A. Raffel, E. D. Cubuk, and I. Goodfellow, “Realistic evaluation of deep semi-supervised learning algorithms,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 1–12.

[29] S. Ghosh, M. Roy, and A. Ghosh, “Semi-supervised change detection using modified self-organizing feature map neural network,” *Appl. Soft Comput.*, vol. 15, pp. 1–20, Feb. 2014.

[30] Z. Yang, W. Cohen, and R. Salakhudinov, “Revisiting semi-supervised learning with graph embeddings,” in *Proc. 33rd Int. Conf. Mach. Learn.*, 2016, pp. 40–48.

[31] W. Shuai, F. Jiang, H. Zheng, and J. Li, “MSGATN: A superpixel-based multi-scale Siamese graph attention network for change detection in remote sensing images,” *Appl. Sci.*, vol. 12, no. 10, May 2022, Art. no. 5158.

[32] X. Wang, K. Zhao, X. Zhao, and S. Li, “CSDBF: Dual-branch framework based on temporal–spatial joint graph attention with complement strategy for hyperspectral image change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2022, Art. no. 5540118.

[33] T.-H. Lin and C.-H. Lin, “Hyperspectral change detection using semi-supervised graph neural network and convex deep learning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Jun. 2023, Art. no. 5515818.

[34] M. Jia, C. Zhang, Z. Zhao, and L. Wang, “Bipartite graph attention autoencoders for unsupervised change detection using VHR remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jul. 2022, Art. no. 5626215.

[35] C. Sun, J. Wu, H. Chen, and C. Du, “SemiSANet: A semi-supervised high-resolution remote sensing image change detection model using Siamese networks with graph attention,” *Remote Sens.*, vol. 14, no. 12, May 2022, Art. no. 2801.

[36] V. Mnih, N. Heess, and A. Graves, “Recurrent models of visual attention,” in *Proc. Neural Inf. Process. Syst.*, 2014, pp. 1–12.

[37] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” 2014, *arXiv:1409.0473*.

[38] W. Yin, H. Schütze, B. Xiang, and B. Zhou, “ABCNN: Attention-based convolutional neural network for modeling sentence pairs,” *Trans. Assoc. Comput. Linguistics*, vol. 4, pp. 259–272, Jun. 2016.

[39] A. Vaswani et al., “Attention is all you need,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–15.

[40] Z. Li, X. Ye, F. Xiong, J. Lu, J. Zhou, and Y. Qian, “Spectral-spatial-temporal attention network for hyperspectral tracking,” in *Proc. 11th Workshop Hyperspectral Imag. Signal Process., Evol. Remote Sens.*, 2021, pp. 1–5.

[41] M. Liu, Z. Chai, H. Deng, and R. Liu, “A CNN-transformer network with multiscale context aggregation for fine-grained cropland change detection,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4297–4306, May 2022.

[42] Q. Li, R. Zhong, X. Du, and Y. Du, “TransUNetCD: A hybrid transformer network for change detection in optical remote-sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Apr. 2022, Art. no. 5622519.

[43] Q. Yuan and N. Wang, “Buildings change detection using high-resolution remote sensing images with self-attention knowledge distillation and multiscale change-aware module,” *ISPRS Arch.*, vol. 46, pp. 225–231, Aug. 2022.

[44] M. Hu, C. Wu, and L. Zhang, “GlobalMind: Global multi-head interactive self-attention network for hyperspectral change detection,” 2023, *arXiv:2304.08687*.

[45] L. Wang, Liguo Wang, Q. Wang, and P. M. Atkinson, “SSA-SiamNet: Spectral–spatial-wise attention-based Siamese network for hyperspectral image change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jul. 2022, Art. no. 5510018.

[46] M. Zhang, Z. Liu, J. Feng, L. Liu, and L. Jiao, “Remote sensing image change detection based on deep multi-scale multi-attention Siamese transformer network,” *Remote Sens.*, vol. 15, no. 3, Feb. 2023, Art. no. 842.

[47] Y. Li et al., “CBANet: An end-to-end cross-band 2-D attention network for hyperspectral change detection in remote sensing,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, May 2023, Art. no. 5513011.

[48] Z. Li, X. Cui, L. Wang, H. Zhang, X. Zhu, and Y. Zhang, “Spectral and spatial global context attention for hyperspectral image classification,” *Remote Sens.*, vol. 13, no. 4, Feb. 2021, Art. no. 771.

[49] H. Chen and Z. Shi, “A spatial-temporal attention-based method and a new dataset for remote sensing image change detection,” *Remote Sens.*, vol. 12, no. 10, May 2020, Art. no. 1662.

[50] Y. Dong, Q. Liu, B. Du, and L. Zhang, “Weighted feature fusion of convolutional neural network and graph attention network for hyperspectral image classification,” *IEEE Trans. Image Process.*, vol. 31, pp. 1559–1572, Jan. 2022.

[51] W. Zi, W. Xiong, H. Chen, J. Li, and N. Jing, “SGA-Net: Self-constructing graph attention neural network for semantic segmentation of remote sensing images,” *Remote Sens.*, vol. 13, no. 21, Oct. 2021, Art. no. 4201.

[52] W. Shi, M. Zhang, R. Zhang, S. Chen, and Z. Zhan, “Change detection based on artificial intelligence: State-of-the-art and challenges,” *Remote Sens.*, vol. 12, no. 10, May 2020, Art. no. 1688.

[53] Q. Zhang, Y. Lu, S. Shao, L. Shen, F. Wang, and X. Zhang, “MFNet: Mutual feature-aware networks for remote sensing change detection,” *Remote Sens.*, vol. 15, no. 8, Apr. 2023, Art. no. 2145.

[54] R. Kou, B. Fang, G. Chen, and L. Wang, “Progressive domain adaptation for change detection using season-varying remote sensing images,” *Remote Sens.*, vol. 12, no. 22, Nov. 2020, Art. no. 3815.

[55] X. Li, Z. Tang, W. Chen, and L. Wang, “Multimodal and multi-model deep fusion for fine classification of regional complex landscape areas using ZiYuan-3 imagery,” *Remote Sens.*, vol. 11, no. 22, Nov. 2019, Art. no. 2716.

[56] X. Zhang et al., “ADHR-CDNet: Attentive differential high-resolution change detection network for remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Nov. 2022, Art. no. 5634013.

[57] M. Liu, Q. Shi, J. Li, and Z. Chai, “Learning token-aligned representations with multimodel transformers for different-resolution change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 2022, Art. no. 4413013.

[58] P. Zhang, M. Gong, L. Su, J. Liu, and Z. Li, “Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images,” *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 24–41, Jun. 2016.

[59] W. G. C. Bandara and V. M. Patel, “Revisiting consistency regularization for semi-supervised change detection in remote sensing images,” 2022, *arXiv:2204.08454*.

[60] D. Peng, L. Bruzzone, Y. Zhang, H. Guan, H. Ding, and X. Huang, “SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5891–5906, Jul. 2021.

[61] J. Liu et al., “Semi-supervised change detection based on graphs with generative adversarial networks,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 74–77.

[62] G. J. Hay, T. Blaschke, D. J. Marceau, and A. Bouchard, “A comparison of three image-object methods for the multiscale analysis of landscape structure,” *ISPRS J. Photogramm. Remote Sens.*, vol. 57, no. 5/6, pp. 327–345, Feb. 2003.

[63] B. Schultz, M. Immitzer, A. Roberto Formaggio, I. Del’Arco Sanches, A. José Barreto Luiz, and C. Atzberger, “Self-guided segmentation and classification of multi-temporal Landsat 8 images for crop type mapping in Southeastern Brazil,” *Remote Sens.*, vol. 7, no. 11, pp. 14482–14508, Oct. 2015.

[64] T. Su, “Efficient paddy field mapping using Landsat-8 imagery and object-based image analysis based on advanced fractel net evolution approach,” *GIScience Remote Sens.*, vol. 54, no. 3, pp. 354–380, Dec. 2017.

[65] J. Zhao, Y. Zhong, H. Shu, and L. Zhang, “High-resolution image classification integrating spectral-spatial-location cues by conditional random fields,” *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4033–4045, Sep. 2016.

[66] M. Li, W. Bijker, and A. Stein, “Use of binary partition tree and energy minimization for object-based classification of urban land cover,” *ISPRS J. Photogramm. Remote Sens.*, vol. 102, pp. 48–61, Feb. 2015.

[67] O. Brovkina, E. Cienciala, P. Surový, and P. Janata, “Unmanned aerial vehicles (UAV) for assessment of qualitative classification of Norway spruce in temperate forest stands,” *Geo-spatial Inf. Sci.*, vol. 21, no. 1, pp. 12–20, Jan. 2018.

[68] A. Räsänen, M. Kuitunen, E. Tomppo, and A. Lensu, “Coupling high-resolution satellite imagery with ALS-based canopy height model and digital elevation model in object-based boreal forest habitat type classification,” *ISPRS J. Photogramm. Remote Sens.*, vol. 94, pp. 169–182, Aug. 2014.

[69] T. Su, H. Li, S. Zhang, and Y. Li, “Image segmentation using mean shift for extracting croplands from high-resolution remote sensing imagery,” *Remote Sens. Lett.*, vol. 6, no. 12, pp. 952–961, Sep. 2015.

[70] Y. Zhong, J. Zhao, and L. Zhang, “A hybrid object-oriented conditional random field classification framework for high spatial resolution remote sensing imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7023–7037, Nov. 2014.

[71] 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. Photogramm. Remote Sens.*, vol. 150, pp. 115–134, Apr. 2019.

[72] B. Zhao, D. Zhang, R. Zhang, Z. Li, P. Tang, and H. Wan, "Delineation and analysis of regional geochemical anomaly using the object-oriented paradigm and deep graph learning—A case study in southeastern inner Mongolia, North China," *Appl. Sci.*, vol. 12, no. 19, Oct. 2022, Art. no. 10029.

[73] S. Cai and D. Liu, "A comparison of object-based and contextual pixel-based classifications using high and medium spatial resolution images," *Remote Sens. Lett.*, vol. 4, no. 10, pp. 998–1007, Aug. 2013.

[74] M. Baatz and A. Schape, "Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation," in *Proc. Beiträge zum AGIT-Symp.*, 2000, pp. 12–23.

[75] J. Wu, B. Li, W. Ni, W. Yan, and H. Zhang, "Optimal segmentation scale selection for object-based change detection in remote sensing images using Kullback–Leibler divergence," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 7, pp. 1124–1128, Jul. 2020.

[76] L. Drăguţ, D. Tiede, and S. R. Levick, "ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data," *Int. J. Geographical Inf. Sci.*, vol. 24, no. 6, pp. 859–871, Mar. 2010.

[77] T. Feng, H. Ma, X. Cheng, and H. Zhang, "Calculation of the optimal segmentation scale in object-based multiresolution segmentation based on the scene complexity of high-resolution remote sensing images," *J. Appl. Remote Sens.*, vol. 12, no. 2, May 2018, Art. no. 025006.

[78] Y. Tang, X. Huang, and L. Zhang, "Fault-tolerant building change detection from urban high-resolution remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1060–1064, Sep. 2013.

[79] Z. Y. Lv, W. Shi, X. Zhang, and J. A. Benediktsson, "Landslide inventory mapping from bitemporal high-resolution remote sensing images using change detection and multiscale segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1520–1532, May 2018.

[80] J. Wu et al., "A dual neighborhood hypergraph neural network for change detection in VHR remote sensing images," *Remote Sens.*, vol. 15, no. 3, Jan. 2023, Art. no. 694.

[81] C. Huo, Z. Zhou, H. Lu, C. Pan, and K. Chen, "Fast object-level change detection for VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 118–122, Jan. 2010.

[82] A. D'Addabbo, G. Satalino, G. Pasquariello, and P. Blonda, "Three different unsupervised methods for change detection: An application," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2004, pp. 1980–1983.

[83] W. Cui et al., "Knowledge and spatial pyramid distance-based gated graph attention network for remote sensing semantic segmentation," *Remote Sens.*, vol. 13, no. 7, Mar. 2021, Art. no. 1312.

[84] S. Khodadadeh, L. Boloni, and M. Shah, "Unsupervised meta-learning for few-shot image classification," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 10132–10142.

[85] C. Zhang et al., "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 183–200, Aug. 2020.

[86] Y. Wang, H. Yao, and S. Zhao, "Auto-encoder based dimensionality reduction," *Neurocomputing*, vol. 184, pp. 232–242, Apr. 2016.

[87] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)—Arguments against avoiding RMSE in the literature," *Geosci. Model Develop.*, vol. 7, no. 1, pp. 1247–1250, Jan. 2014.

[88] H. Su, X. Zhang, Y. Luo, C. Zhang, X. Zhou, and P. M. Atkinson, "Nonlocal feature learning based on a variational graph auto-encoder network for small area change detection using SAR imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 193, pp. 137–149, Nov. 2022.

[89] D. A. Kumar, S. K. Meher, and K. P. Kumari, "Knowledge-based progressive granular neural networks for remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5201–5212, Dec. 2017.

[90] N. S. Kothari, S. K. Meher, and G. Panda, "Improved spatial information based semisupervised classification of remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 329–340, Jan. 2020.

[91] J. Wu, B. Li, Y. Qin, W. Ni, and H. Zhang, "An object-based graph model for unsupervised change detection in high resolution remote sensing images," *Int. J. Remote Sens.*, vol. 42, no. 16, pp. 6209–6227, Jun. 2021.

[92] Y. Jia, Z. Xie, Z. Lv, and M. Zhu, "A new change detection method of remote sensing image," *Geomatics Inf. Sci. Wuhan Univ.*, vol. 41, no. 8, pp. 1001–1006, Aug. 2016.

[93] T. Joachims, "Transductive inference for text classification using support vector machines," in *Proc. 16th Int. Conf. Mach. Learn.*, 1999, pp. 200–209.

[94] D. Wen et al., "Change detection from very-high-spatial-resolution optical remote sensing images: Methods, applications, and future directions," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 68–101, Dec. 2021.

[95] Y. Ding, X. Zhao, Z. Zhang, W. Cai, and N. Yang, "Graph sample and aggregate-attention network for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Mar. 2022, Art. no. 5504205.

[96] K. Sörensen, M. Sevaux, and F. Glover, "A history of metaheuristics," in *Handbook of Heuristics*. Berlin, Germany: Springer, 2018, pp. 791–808.

[97] Y. Zhang, X. Wang, X. Jiang, and Y. Zhou, "Robust dual graph self-representation for unsupervised hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 2022, Art. no. 5538513.

[98] X. Li et al., "Superpixel segmentation based on anisotropic diffusion model for object-oriented remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, to be published, doi: 10.1109/JSTARS.2023.3324770.

[99] D. Xiu, Z. Pan, Y. Wu, and Y. Hu, "MAGE: Multisource attention network with discriminative graph and informative entities for classification of hyperspectral and LiDAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2022, Art. no. 5539714.

[100] A. V. D. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018, *arXiv:1807.03748*.

[101] M. Kim, M. Madden, and T. Warner, "Estimation of optimal image object size for the segmentation of forest stands with multispectral IKONOS imagery," in *Object-Based Image Analysis*. Berlin, Germany: Springer, 2008, pp. 291–307.

Yifan Wu received the the B.E. degree in surveying and mapping engineering from Shenyang Jianzhu University, Shenyang, China, in 2020.

She is a graduate student with the School of Transportation and Geomatics Engineering, Shenyang Jianzhu University, Shenyang, China. She is a student of Zhiwei Xie. Her main research interest focuses on spatial data analysis.

Zhiwei Xie received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2016.

He is currently a Postdoctoral Fellow with Nanjing Normal University, Nanjing, China. He is an Associate Professor with the School of Transportation and Geomatics Engineering, Shenyang Jianzhu University, Shenyang, China. His research interests include urban remote sensing, remote sensing image intelligent processing, and spatial data analysis.

Zaiyang Ma received the Ph.D. degree in cartography and geographic information system from Nanjing Normal University, Nanjing, China, in 2022.

He is currently a Postdoctor with Nanjing Normal University, Nanjing, China, and a member of OpenGMS Research Group. His research interests include the areas of collaborative geographic modeling and geographic information systems.

Min Chen received the Ph.D. degree in cartography and geographic information system from Nanjing Normal University, Nanjing, China, in 2009.

He is currently a Professor with the School of Geography, Nanjing Normal University, Nanjing, China, and a member of OpenGMS Research Group. His research interests include virtual geographic environment, geographic modeling and simulation, and geographic information system.

Fengyuan Zhang received the Ph.D. degree in cartography and geographic information system from Nanjing Normal University, Nanjing, China, in 2021.

He is currently a Postdoctor with the School of Environment, Nanjing Normal University, Nanjing, China. He is a member of OpenGMS Research Group. His research interests include model sharing and reusing for geographical simulations on the web, and HD map data fusion.

Zhenguo Shi received the B.E. degree in surveying and mapping engineering from Shandong University of Science and Technology, Qingdao, China, in 2020.

He studied with the School of Transportation and Surveying Engineering, Shenyang Jianzhu University, Shenyang, China. He studied under Professor Zhiwei Xie, mainly focusing on high-resolution image change detection.

Wengang Li received the bachelor's degree in 2019 from Shenyang Jianzhu University, Shenyang, China, where he is currently working toward the master's degree with the School of Transportation and Geomatics Engineering.

He studies under teacher Zhiwei Xie. His main research direction is remote sensing image processing.

Shuaizhi Zhai received the undergraduate degree in 2022 from Shenyang Jianzhu University, Shenyang, China, where he is currently working toward the master's degree with the School of Transportation and Geomatics Engineering.

He studies under teacher Zhiwei Xie. His main research direction is remote sensing image processing.