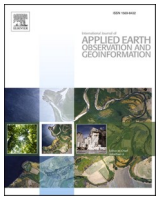
International Journal of Applied Earth Observation and Geoinformation 130 (2024) 103897

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Deep Siamese Network for annual change detection in Beijing using Landsat satellite data

Hanqing Bao [a,*], Vinzenz H.D. Zerres [a], Lukas W. Lehnert [a]

[a] Department of Geography, Ludwig-Maximilians-Universität, 80333 Munich, Germany

ARTICLE INFO

Keywords:
Spatio-temporal change
Remote sensing
Deep Learning
Object-level semantic information
DF-GCN

ABSTRACT

The detection of spatiotemporal changes in land use/land cover (LULC-SC) plays a paramount role in the analysis of smart cities, it can describe complex urban distribution, functions, and patterns. Chinas capital city has been developing rapidly in the past two decades, however, there are only few long-term studies on an annual scale of LULC-SC. To fill this research gap, we propose a remote sensing parallel framework for the detection of LULC-SC based on the combination of the Deep Siamese Network and long time series, which focuses on the spatial semantic information at the object level. A Landsat time series from 2002 to 2022 serves as input satellite data. First, we focus on building graph constructions at the object level and then use an autonomously constructed deeper-feature graph convolutional network (DF-GCN) to mine deeper features, spatial semantic, and relationships at the object level. Finally, the Siamese Network recognizes the changes in the spatial semantic tensors of long time series of Landsat images and quickly maps LULC-SC. The results prove that the proposed spatiotemporal change detection framework is effective in LULC-SC in Beijing. Compared with other networks, the optimal accuracy of semantic mining based on DF-GCN can reach about 90%. Over the past two decades, the LULC-SC of Beijing has changed in a complex way, with urbanization occurring primarily through the replacement of farmland. Consequently, the proposed framework can generate accurate LULC change maps at high temporal frequencies, which can contribute to a better comprehension of sustainable urban development and planning.

1. Introduction

Faced with the increasingly pressing issues of global population growth, dwindling resources, and environmental degradation, spatiotemporal changes in land use/land cover (LULC-SC) have become a vital topic in the field of global change (Junfu et al., 2022). LULC-SC represents the spatiotemporal mapping of human activities and the natural environment, reflecting the direct interactions between socio-economic activities and natural ecological processes. Therefore, mapping and monitoring changes in LULC has significant implications for agricultural production, urban planning and development, environmental protection, and sustainable development (Baohui and Peijun, 2023; Zheng et al., 2022).

With the booming development of remote sensing technology, satellite images are a comprehensive and timely data source, which can detect spatiotemporal changes across large areas of land surface, and have thus been increasingly applied to map LULC changes (Cetin et al., 2021). The most widely used satellite data to detect LULC changes are from the Landsat program because of the long time series and the sufficiently high spatial resolution (30 m), wide spatial coverage, and rich spectral information (Xuexian et al., 2022).

To detect LULC changes using remotely sensed imagery, LULC must be first classified using at least two images acquired at two different dates. Traditional image classification mainly uses low-level features, which include image bottom-level features and shallow visual features. Image bottom-level features such as spectrum, shape, and texture or widely used indices such as the normalized difference vegetation index (NDVI) and the normalized difference water index (NDWI) were used to directly classify LULC (Meng et al., 2023; Vincent and Irene, 2022). However, these traditional bottom-level features and indices are insufficient to characterize and explain complex land cover types. (Bao et al., 2020) compared the Shenzhen classification map of multiple methods to reveal the limitations of low-level features. Subsequently, some shallow visual features (LBP, SIFT, HOG) were developed. Although they have achieved some success in LULC classification, these visual-features-based works are mainly applicable to simple situation (Fotso Kamga Guy et al., 2018; Mohan and Kapil Dev, 2021). The bottom-level features and shallow visual features mentioned in the above methods are all

* Corresponding author.
E-mail addresses: Hanqing.Bao@campus.lmu.de (H. Bao), V.Zerres@campus.lmu.de (V.H.D. Zerres), lehnert.lu@lmu.de (L.W. Lehnert).

https://doi.org/10.1016/j.jag.2024.103897
Received 19 December 2023; Received in revised form 2 May 2024; Accepted 5 May 2024
Available online 8 May 2024

guided by fixed expert knowledge, thus limiting the transferability between different regions and datasets. More importantly, the accuracy of results is often affected by these inaccurate low-level features.

Recently, it has been demonstrated that Deep Learning (DL) has high potential to capture spatial features and semantics. Especially in the field of remote sensing, DL can solve problems of big data and complexity of geographic object features and their distribution (Xiangyu et al., 2023). Initially, Convolutional Neural Networks (CNN) and Fully Convolutional Networks (FCN) were used for geographical object classification and simple scene classification. However, contours and boundaries of objects were poor (Zhao et al., 2019; Zhenshi et al., 2022). While the U-net model makes up for the deficiency of CNN to extract detailed contour features of objects, it is only suitable for scenes with fewer objects (David and Ce, 2022; Huanxue et al., 2021); DeepLabv3 and the network based on the idea of attention mechanism (for example, the development of Transformer) can not only describe the detailed outline of geographic objects but also complete complex scene classification based on object levels (Jiaqi et al., 2023; Zhimin et al., 2022). However, all these methods that only use the features of the objects themselves and do not consider spatial knowledge such as spatial correlation between objects, as well as deep contextual information (Wang et al., 2023; Zhou et al., 2023).

Graph Convolutional Networks (GCN) have obvious advantages in mining spatial semantics from irregular data through graph convolution and is commonly applied in LULC classification research (Hong et al., 2020; Liang et al., 2020). The core idea of a GCN is the use of spatial semantic connections to represent node information of ground objects and the characterization of new nodes (Yongyang et al., 2022; Zhang et al., 2021). However, GCNs can only interpret spatial relationships and cannot mine depth features (Yao et al., 2022). This conflict raises the question of how we can express spatial information while considering depth features. Consequently, a combination of CNN and GCN (CNN + GCN) could overcome the disadvantages of both methods, because CNN can consider deep features while GCN accounts for spatial relationships (Ding et al., 2022; Liu et al., 2020). Previous studies from scene recognition fields have proven the effectiveness of this method, but highlighted several limitations, such as mining capabilities of CNN are insufficient (Liang et al., 2020). Furthermore, as a result of carriers that are too small (pixel level) or too large (geographical scene), the spatial relationships of ground objects might be unclear (Jafarzadeh et al., 2022). In addition, the combined CNN and GCN method has yet to be applied to land use and land use changes in complex large cities, as it is only suitable for simple and basic scene recognition (Li et al., 2020).

Objects function as important bridges between pixels and regions. For downscaling purposes, objects are collections of pixels that reveal the shape and structure of ground objects; for the purpose of upscaling, they are a composition of geographical units that carry spatial semantic information and structural relationships. Therefore, focusing on the object level can provide a more accurate description of the real boundaries and shapes of ground objects, which helps to better understand and explain surface phenomena. LULC refers to geographical units within a certain region with different geographical objects, spatial relationships, and functions. The irregular distribution of objects and the connections between types of objects are not simply linear. As shown in Fig. 1, the geographical objects can be simplified as nodes. Connecting edges between nodes represent the spatial relationships between objects, such as adjacency, intersection, and separation (Xie et al., 2022; Zhang et al., 2019).

The Siamese Network is a machine learning model that is akin to human insight and memory, enabling it to quickly find differences in a short time (Hongyang et al., 2023; Qiqi et al., 2022). Siamese Networks have shown great potential when facing long time series of data. Due to the tremendous changes over the past decades, Chinas capital Beijing serves as a good example to show the potential of Siamese Networks in detecting LULC-SC over time. Previous research on LULC-SC in Beijing focused on 5-year or 10-year time intervals, while annual changes have not been investigated so far (Huabing et al., 2017; Jiang et al., 2020) When assessed annually, changes in all types of LULC can not only be described in more detail, but they can also help better understand LULC-SC trajectories over a long time series. This is beneficial, as the detailed changes are often neglected and overlooked in a 5-year or 10-year time interval (Zhang et al., 2022).

To further improve the mapping of annual LULC changes in Beijing, we propose a LULC-SC detection framework based on the deep Siamese Network that focuses on semantic information and spatial relationship mining as well as fusion at the object level. The detection framework used an autonomously constructed Deeper-Feature GCN Network (DF-GCN) and a Siamese Network to monitor spatiotemporal changes in long time series of remote sensing images.

In contrast to previous studies, the monitoring framework proposed fully utilizes the object-level multispectral information and spatial semantics by automatically mapping LULC-SC based on the Siamese Network. In summary, the main aims and hypotheses of this paper are as follows:

1. We hypothesize that deeper feature and semantic information at the object level are important for evaluation. The strength of DF-GCN is fully utilized at the adaptive object level
2. The framework proposed in this paper uses an autonomously constructed DF-GCN network. We tested if the DF-GCN network can better handle complex Landsat data at the deeper features and semantic information levels. This results in a higher accuracy of DF-GCN networks compared to other methods
3. To map the LULC-SC of Beijing at an annual scale, a framework based on deep Siamese network is proposed, which can perceive spatial semantic tensors' changes in long time series of images and quickly map LULC-SC
4. By applying the novel method at an annual-scale, the transformation of LULC in Beijing over the past 20 years is mapped, which provides support for future urban planning and sustainable development.

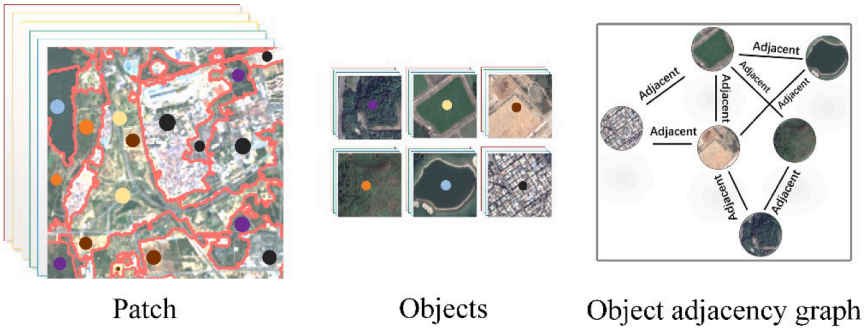


Fig. 1. Objects spatial adjacency graph. (Adjacency refers to being closely connected; intersection refers to cross connection; separation means independence from each other and no connection.).

2. Study area and data sets

2.1. Study area

This study focuses on the area of Beijing, located in the northern part of the North China Plain (Fig. 2). As the political, cultural, and international exchange center of China, it is becoming a more international metropolis since the reform and opening-up. the population and economic growth has experienced rapid, as well as significant changes in its urban landscape.

2.2. Satellite data

Landsat data has been widely used in LULC studies. After downloading all available land surface reflectance data from 2002 to 2022 for summer months that cover the study area, the data was visually screened and preprocessed. Finally, six spectral bands of the Landsat imagery were selected as input data, including the blue, green, red, near-infrared, SWIR-1, and SWIR-2 bands. To obtain suitable LULC classification and change monitoring results for Beijing, the atmospheric and illumination effects of the spectral reflectance from ground objects was removed during preprocessing. This was done by applying a radiometric and atmospheric correction to the Landsat imagery.

2.3. Sample and training parameters

The selection of samples is crucial to explore deeper features and spatial relationships. In this experiment, the labeled samples consisted of square patches attributed to land cover classes. Land use and cover types in Beijing were classified into seven categories: bare land, forest, shrubland, grassland, water body, urban, urban-green, and farmland. All samples were manually selected through visual interpretation to ensure representation of all classes and an appropriate and accurate number of samples for each category (Lv et al., 2018).

To reinforce the robustness of the training network, we manually selected nearly 2,400 sample points, 80 % of them were randomly selected as input samples. These are then divided into 80 % for training samples and 20 % for the verification dataset. The validation dataset was used to test the accuracy of the model. Fig. 2 and Table 1 presents the spatial distribution and the actual numbers of samples used for training and internal model validation (Bao et al., 2020). The test data set is generated by image segmentation and object centering (Section 4.1).

Furthermore, the input image has 7 bands, consisting of 6 remotely sensed spectral bands and one deep edge feature map. To ensure adaptability, the scale of the input patches matches the scale of the patches used to build the graph constructions (Section 5.2).

2.4. Ground truth samples

To reflect the accuracy of the LULC classification results more objectively and scientifically, therefore ground truth points are randomly generated to verify the accuracy of the model output results. Shown in Table 1, these 4500 ground truth points based on Land-use classes. Classes of each point have been manually labeled using the original image and Google Earth. Consequently, these ground truth points are independent and differ from the samples during the model training process. Finally, we constructed a confusion matrix to check the accuracy of model predictions against validation data from visual interpretation.

2.5. Existing products

GlobeLand30 is a 30 m resolution global land cover dataset, which is mainly derived from remote sensing satellite data (Landsat and China Environmental Disaster Reduction Satellite), aerial photography images, and ground survey data (Chen et al., 2017).

The Copernicus Global Land Service (CGLS) provides a range of biogeophysical products on surface conditions and evolution at the global scale. Global land cover maps are available at 100 m spatial resolution.

The European Space Agency (ESA) provides global land cover maps with a resolution of 10 m based on Sentinel data.

For this study, the GlobeLand30 data sets in 2010 and 2020, the Copernicus Global Land Cover data sets in 2018 and 2019, and the ESA data in 2020 and 2021 were downloaded and selected. The Beijing area was selected as a reference to compare and verify the proposed method.

3. Methods

The analysis has been performed on a Windows 10 operating system using a CPU (3.4 GHz core i7-6700), RAM (16 GB), and GPU (NVIDIA

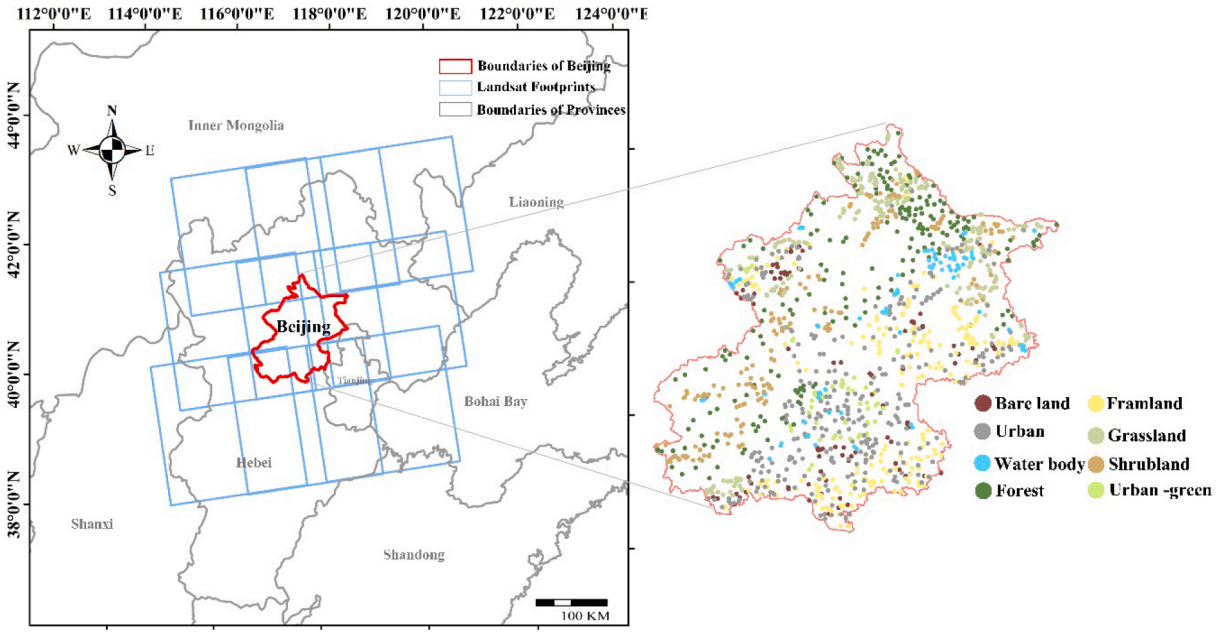


Fig. 2. Geographical location of Beijing and the training sample sets.

Table 1
The training samples, internal model validation samples and ground truth samples for validation of the LULC maps.

Category	Urban	Bare land	Urban-green	Forest	Grassland	Shrubland	Waterbody	Farmland
Training	398	156	89	218	208	189	60	200
Validation	113	34	22	53	53	41	16	48
Ground Truth	1200	450	270	600	600	600	180	600

RTX-A4000). TensorFlow2.3 was selected as the deep learning framework.

The overall process framework of LULC-SC mapping is shown in Figs. 3 and 4. As main input to Deep Siamese Networks, images acquired at two different time steps are used to generate semantic tensors, and the similarity of the two semantic tensors is calculated under the framework of the Siamese Network to implement change detection. Its main components are DF-GCN network and similarity calculation.

As the core component, DF-GCN network mainly consists of three parts (Fig. 4): the blue part symbolizes the segmentation of Landsat images by scale-adaptive hierarchies, which establishes the object-level spatial adjacency graph (OSAG). In the green part, DF-CNN is used to mine deeper features, while finally the exploration of spatial relationships and information using GCN is performed symbolized by the pink part.

Finally, semantic tensor is generated through fusion and input to the Siamese Network to calculate the similarity to map LULC-SC in Beijing region at an annual scale (see Appendix for the algorithm).

3.1. Adaptive scale estimation strategy in hierarchical segmentation

To solve the problem of large data volume and insufficient single scale to express complex objects, adaptive-scale hierarchical segmentation strategy was adopted. As shown in Fig. 5, the entire image is first segmented into several large regions using multi-texture calculation. Afterwards, the spatial scale is estimated for each sub-region, allowing the geographical objects to obtain a better scale representation. Among them, the edge feature comes from texture calculation, which contains rich boundary information, so it can effectively limit the boundary to approximate to the real world.

This paper used the algorithm of adaptive multiscale estimation, by computing the average local variance of different windows in the image to verify the spatial scale transformation of geographical objects (Drăguţ et al., 2014). The method iteratively segments the image in a bottom-up manner, and when the average local variance of a scale is equal to or lower than the previous scale, the iteration, and the next level of scale estimation begins.

This method has good potential in the objectivity and adaptability of scale selection, and can meet the precise expression of different object scales. It therefore provides possibilities for subsequent research (Xu et al., 2019).

3.2. DF-GCN: The organic fusion of DF-CNN and MB-GCN

3.2.1. DF-CNN: Deeper features capturer

Inspired by the inception and attention modules, this study autonomously constructs a deeper-feature convolutional neural network (Deeper-feature CNN, DF-CNN) based on the attention mechanism. The DF-CNN consists of five convolutional modules and a Squeeze-and-Excitation (SE) module, along with pooling layers, and a global average pooling layer (Fig. 6).

Instead of traditional convolutional layers, we adopt a deeper convolutional module (D-Conv), which expands the perception field and fully utilizes contextual information. (D-Conv) includes a 1x1 convolutional filter, two 3x3 convolutional filters, and a 3x3 deep separable convolution. Deep separable convolution can consider spatial information first and then spectral information, accelerating the efficiency of convolutional computations.

The SE module represents the channel attention mechanism module, where S and E denote the squeeze and excitation operations, respectively. It can focus on bands with more information to improve feature representation capabilities. As shown in Fig. 6, it is composed of global pooling, two fully connected layers, ReLU, and sigmoid activation function.

Global average pooling is used as the squeeze operation, where the 2D feature space (composed of spectral bands and textures) is compressed into a scalar sequence along the spatial dimension. It mixes local information and has a global receptive field. The computation formula is as follows:

$$f_c = F_{sq}(X_c)\frac{1}{W^*H}\sum_{i=1}^{W}\sum_{j=i}^{H}X_c(i,j) \quad (1)$$

F_{sq} is the Squeeze operation function, X_c is the feature map of size

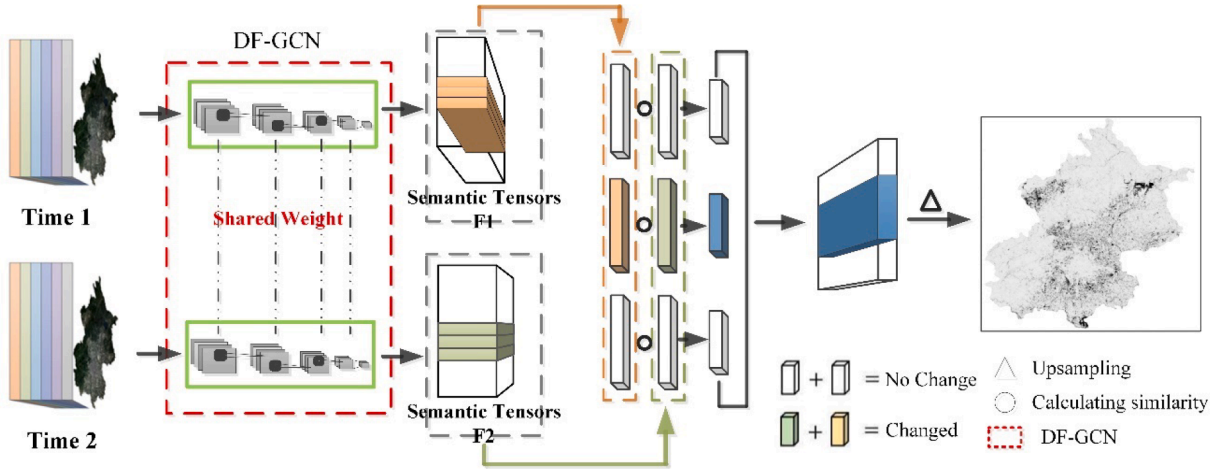


Fig. 3. Overview of the workflow of the proposed LULC-SC detection framework based on Deep Siamese Network. The red dashed box symbolizes the DF-GCN (mining semantic tensors) and comparator (calculate the semantic tensor similarity of the same points in different times), which is further illustrated in Fig. 4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

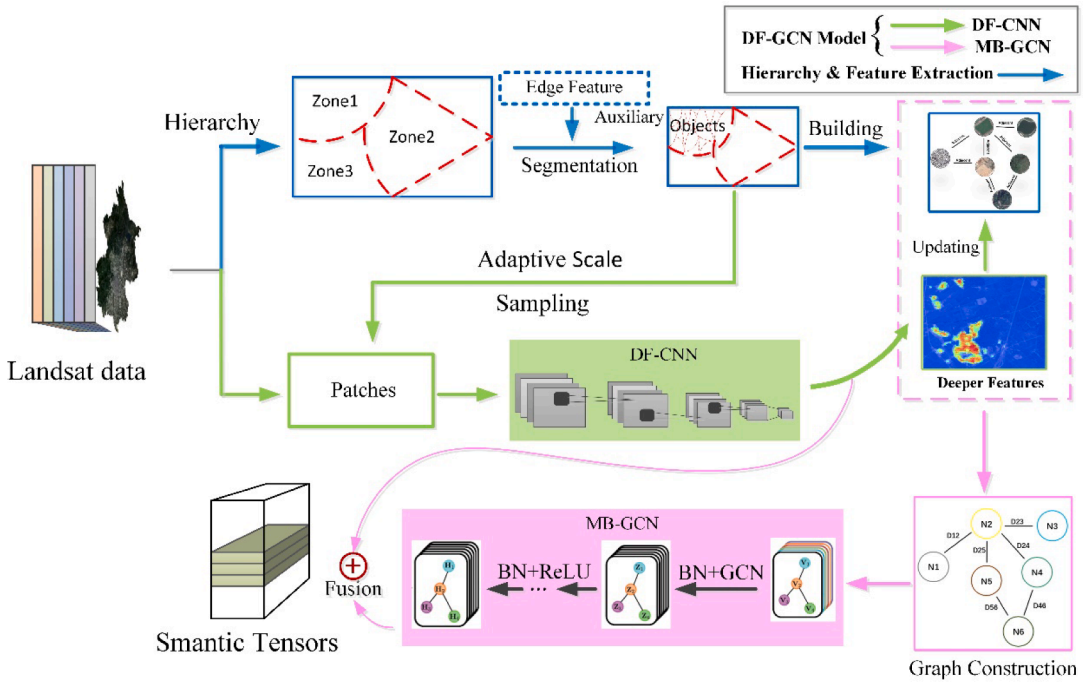


Fig. 4. DF-GCN framework (illustrated as red rectangle in Fig. 3). The blue part is hierarchical segmentation (see Fig. 5 for further details); the green part is the deep feature miner (see Fig. 6 for details): DF-CNN, and the pink part is the spatial relationship detector—Mini-GCN (see Fig. 7 for details). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

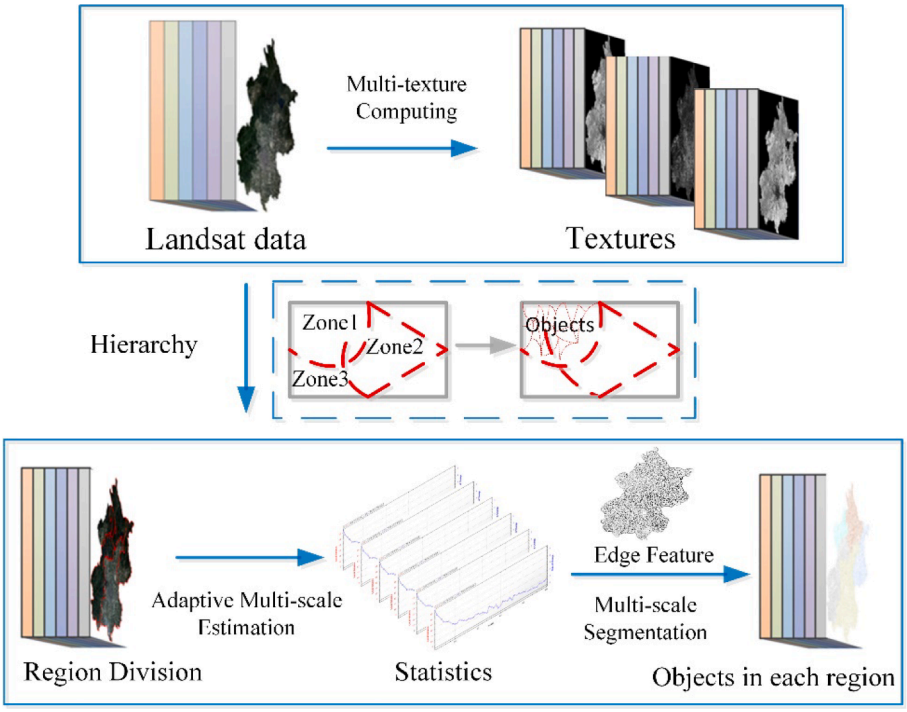


Fig. 5. Flowchart of the Adaptive Scale Estimation Strategy in Hierarchical Segmentation. From top to bottom, the segmentation from regional to object scale is shown. (Edge Feature Map is used to constrain segmentation boundaries).

W*H for the C-th feature space, and f_c is the compressed feature vector after squeezing.

The excitation operation generates a weight factor k for each feature space using the parameter w. The parameter w is used to learn the correlations between feature space.

$$k = F_{ex}(f_c, w) = sigmoid(g(f, w)) = sigmoid(w_2 ReUL(w_1, f)) \quad (2)$$

F_{ex} is the excitation operation function and f_c is the result of the squeeze operation. w_1 and w_2 are the dimension reduction and expansion parameters, $w_1 \in R^{\frac{C}{R}*C}$, $w_2 \in R^{C*\frac{C}{R}}$, where C is the number of feature space and R is the reduction ratio. The activation functions *sigmoid* and *ReUL* are used.

The Reweight operation involves using the output weights, k, from

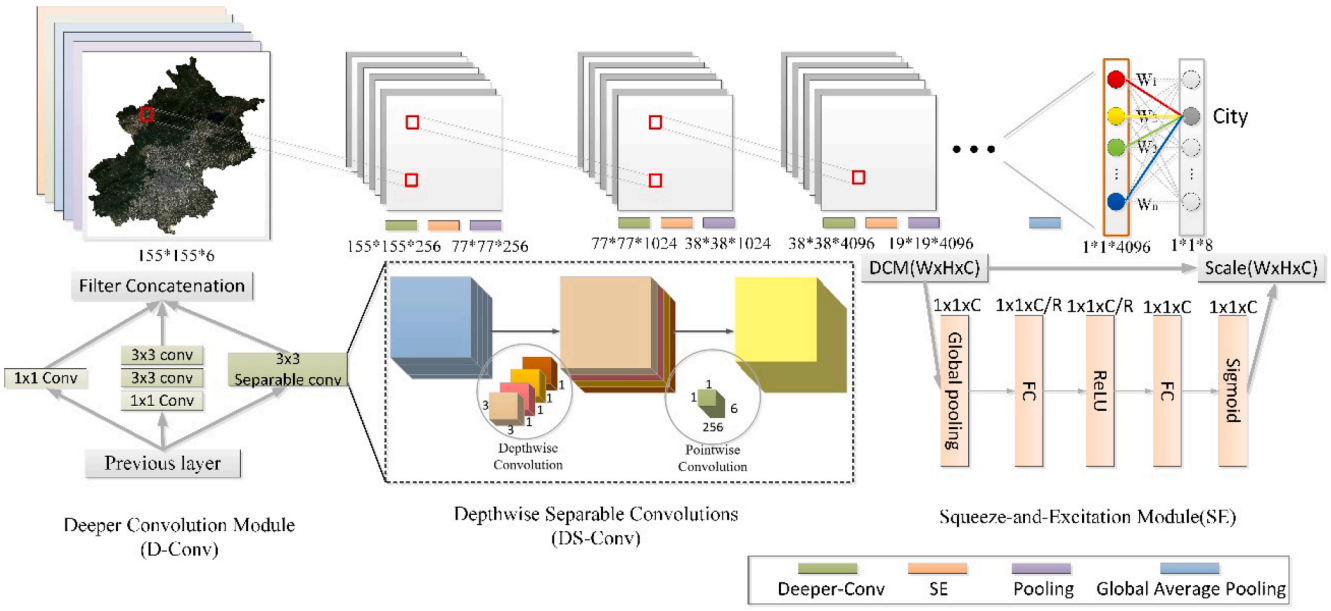


Fig. 6. Configuration of DF-CNN framework. The network has 5 layers, each layer is mainly composed of deeper convolution module, SE module and pooling layer. For an overall view on the CNN-GCN model, see Fig. 4 (green box). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the excitation operation as the weights for each feature space. These weights are the applied for each element through multiplication to the previous features, resulting in the recalibration of the original features along the feature-space dimension.

$$\widetilde{X}_C = F_{scale}(X_C, k_C) = k_C \bullet X_C \tag{3}$$

F_{scale} is the element-wise multiplication function. X_c represents the feature map of size W*H for the C-th feature space, and k_c represents the weight of the C-th feature space.

Furthermore, a global average pooling layer is used at the end to lessen the number of parameters and prevent overfitting, thus ensuring the representation of deep-level features of LULC.

3.2.2. GCN: Spatial relationship detector

3.2.2.1. Graph construction. In this paper, the nodes of the graph are generated from the geometric center points of the objects. A topological graph is constructed using the K-nearest neighbors (KNN) method. Furthermore, the similarity of object classes is used as the edge weight, with higher weights assigned to connections between nodes of the same class. This approach aims to construct a more accurate adjacency matrix.

The feature matrix of the object nodes is denoted as $X \in R^{N*C}$, where N and C represent the number of object nodes and features in the feature vector. The calculation of the Euclidean distance $D_{i,j}$ between two nodes N_i and N_j is given by the following formula:

$$d_{i,j} = \left(\sum_{c=1}^{C} |N_{ic} - N_{jc}|^2 \right)^{\frac{1}{2}} \tag{4}$$

The distance measurements between all pairs of nodes can be represented as a symmetric distance matrix $D = d_{i,j} \in R^{N*N}$. For instance, the element $d_{i,j}$ in matrix D represents the Euclidean distance between the i-th and j-th objects.

To facilitate graph constructions and reduce redundant information, we choose the K nearest objects to each sample node N_i as its adjacent nodes and connect them with edges. This process is used to build graph constructions as samples.

3.2.2.2. Minibatch graph convolutional network (MB-GCN). To focus on local information, we adopt mini-batch learning in GCN, which can reduce training complexity and accelerate computation. As graph convolutional networks can only directly operate on graph constructions, it is crucial to construct the Landsat imagery as a topological graph structure (Hong et al., 2020). Therefore, building the graph construction involves three steps: 1. Determining nodes and their own features; 2. Determining adjacency between nodes; 3. Computing edge weights for the nodes.

We represent an undirected graph as $G = (N, E, A)$, where N represents the set of nodes and $N \vee n$ indicates that there are n nodes in the input patch. E represents the set of edges, and the adjacency matrix $A \in R^{n*n}$ records the adjacency relationships between nodes. The degree matrix $D_{i,i} = \sum_i A_{i,j}$ stores the number of connections associated with node N_i.

To address the challenge of convolution definition caused by the lack of translation invariance in graph data, the convolution on graphs can be defined in the spectral domain after Fourier transformation using the following an equation:

$$g_\theta \bullet x = U g_\theta U^T x \tag{5}$$

Here, x represents the signal on the nodes, $g_\theta = diag(\theta)$ is a parameterized diagonal matrix, U is a matrix composed of the eigenvectors of the normalized Laplacian matrix, $U^T x$ represents the graph Fourier transform of x.

The calculation formula for the graph Laplacian matrix is as follows:

$$L = I_n - D^{\frac{-1}{2}} A D^{\frac{-1}{2}} = U \Lambda U^T \tag{6}$$

L represents the graph Laplacian matrix, Λ is the diagonal matrix composed of the eigenvalues of L.

The convolution operation defined in equation (5) requires the eigen-decomposition of the Laplacian operator. Considering the number of nodes and features, the equation is approximated using a Chebyshev polynomial expansion up to the k-th order.

$$g_\theta \bullet x \approx \sum_{k=0}^{K} \theta_k T_k\left(\widetilde{L}\right) x \tag{7}$$

$\widetilde{L} = \frac{2}{\lambda_{max}} L - I_n$, λ_{max} is the largest eigenvalue of L; $\theta \in R^k$ is the Chebyshev coefficient vector. To lessen parameters and computational complexity and prevent overfitting, we set $k = 1$ and $\lambda_{max} = 2$,

simplifying equation (6) to:

$$g_\theta \bullet x \approx \theta\left(I_n + D^{\frac{-1}{2}}AD^{\frac{-1}{2}}\right)x \tag{8}$$

Then, we introduce "renormalization" by denoting $I_n + D^{\frac{-1}{2}}AD^{\frac{-1}{2}}$ as $\widetilde{D}^{\frac{-1}{2}}\widetilde{A}\widetilde{D}^{\frac{-1}{2}}$, where $\widetilde{A} = A + I_n$, $\widetilde{D}_{i,i} = \sum_i \widetilde{A}_{i,j}$. By stacking multiple convolutional layers with the above definition, we obtain the GCN model. The expression for the $(l+1)$ th layer is:

$$H^{l+1} = \sigma\left(\widetilde{D}^{\frac{-1}{2}}\widetilde{A}\widetilde{D}^{\frac{-1}{2}}H^l W^l\right) \tag{9}$$

Here, W represents the weight matrix, l denotes the layer number in the convolutional network, and σ is the activation function.

Considering the minibatch GCN, the graph convolution for the j-th batch can be expressed as shown in equation (9):

$$H^{l+1} = \sigma\left(D_j^{\frac{-1}{2}}\widetilde{A}_j D_j^{\frac{-1}{2}}H_j^l W_j^l\right) \tag{10}$$

3.3. Feature fusion

We have designed a simple feature tensors fusion module. The feature descriptors extracted from two branches (DF-CNN and MB-GCN) are integrated and passed to the comparator of the deep Siamese Network. The description of feature fusion is as follows:

$$H_{fusion}^{l+1} = H_{DCNN}^{l} \oplus H_{Mb-GCN}^{l} \tag{11}$$

Where the operator $\oplus$ represents element-wise addition. H_{DCNN}^{l} and H_{GCN}^{l} respectively represent the features extracted from DF-CNN and MB-GCN at the $l-$ th layer.

3.4. Siamese network construction

We use images of two different years for the deep Siamese Network (Fig. 3), which shared weights in parallel network. By comparing the similarity of the semantic tensor's pairs to achieve image change detection.

The formula for calculating similarity is as follows:

$$D_w(T_1, T_2) = \|(S_w(T_2)) - (S_w(T_1))\| \tag{12}$$

where D_w is the distance between two feature tensors, and the similarity is calculated through the distance of tensor. T refers to different times. S_w represents the network model, and w is the shared parameter weight.

The essence of the Siamese Network is to train a metric function that outputs larger values in areas with significant changes and smaller values otherwise. Therefore, we adopt a contrastive loss function, which is expressed as follows:

$$L = \frac{1}{2N}\sum_{n=1}^{N} yd^2 + (1-y)max(margin - d, 0)^2 \tag{13}$$

Here, $d = \vee f_{t1} - f_{t2} \vee_2$ represents the Euclidean distance between different temporal semantic tensors, and y is the label indicating whether the feature tensor pairs are matching: $y = 1$ indicates that the semantic tensors are similar or matching, while $y = 0$ indicates that they are not matching. Margin refers to the threshold set for comparison.

3.5. Comparison with other machine learning classifiers

3.5.1. Comparison with other machine learning classifiers

To compare the performance of DF-GCN method, we used the same dataset of satellite and training data to train other widely used classifiers such as random forest (RF), CNN (AlexNet), DF-CNN, GCN and the combination of CNN + GCN. Afterward, the differences and overall accuracy between the different methods were compared qualitatively and quantitatively via confusion matrices.

3.5.2. Comparison with other change detection methods

To verify the effectiveness and superiority of our proposed change detection framework based on deep Siamese network, three popular change detection methods are selected as comparison methods in this study.

1. LandTrendr is a time segmentation algorithm used to capture long-term, gradual, or short-term drastic changes in time series. It can monitor each pixel to determine whether it has changed
2. Mspsnet builds a parallel convolution strategy. Different convolutions achieve feature aggregation and improve the receptive field, while the self-attention module makes it easier to detect changing areas
3. BiT is a transformer-based network. Transformer encoders are used to model the spatiotemporal context of compact pixel information. Changes can then be monitored by refining the raw features through a transformer decoder

3.6. Evaluation metrics

This study uses several evaluation metrics commonly used in change detection, including the overall accuracy (OA), precision (Pre), recall (Rec). Among these metrics, TP, TN, and FN indicate true positive, true negative, false positive, false negative, respectively. The calculations of these five indicators are formulated as follows.

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

$$Pre = \frac{TP}{TP + FP} \tag{15}$$

$$Rec = \frac{TP}{TP + FN} \tag{16}$$

4. Results

4.1. The results of the adaptive-scale hierarchical segmentation

Based on the multi-texture computing and the strategy of scale estimation, Beijing city was divided into 5 sub-zones. The segmentation parameters were adjusted continuously through empirical refinement. Subsequently, spatial scale estimation was carried out for each sub-zone. The estimated scale parameters are shown in Table 2 for the year 2002 as an example. In this paper, we adopt an adaptive multi-scale segmentation method. The hierarchical scale estimation results are shown in Fig. 8.

4.2. Ablation experiments

The designed ablation experiments are shown in Table 7, where every two adjacent experiments form a group of control experiments. For example, using RSI and EFM (Exp.2) has better results than using RSI alone (Exp.1) because the Overall is 0.041 higher. By combining hierarchical segmentation to focus on the object level (Exp.3), Overall can be further improved from 0.787 to 0.832. In addition, the contributions of deep feature-based models and spatial relationship-based models to the classification were tested separately in experiments (Exp.3–6). When only considering deep features, Overall improves from 0.832 (Exp.3) to 0.899 (Exp.4) respectively. In addition, through the joint use of deep features and spatial relationships, Overall is further improved to 0.921 (Exp.6). In Exp.3 and 5 and Exp.4 and 6, with the support of spatial relationships, the accuracy increased by 0.037 and 0.022 respectively.

Ablation experiments show that all strategies used in the proposed framework improve classification accuracy to a certain extent, with DF-

Table 2
The parameters of Hierarchical Segmentation.

	Zone 1			Zone 2			Zone 3			Zone 4			Zone 5		
Level	**L1**	**L2**	**L3**	**L1**	**L2**	**L3**	**L1**	**L2**	**L3**	**L1**	**L2**	**L3**	**L1**	**L2**	**L3**
SP	48	68	108	37	67	117	47	73	133	66	126	193	48	67	148
Num	4174	1822	641	3716	1072	367	4589	1811	471	6234	1381	541	5525	2552	506

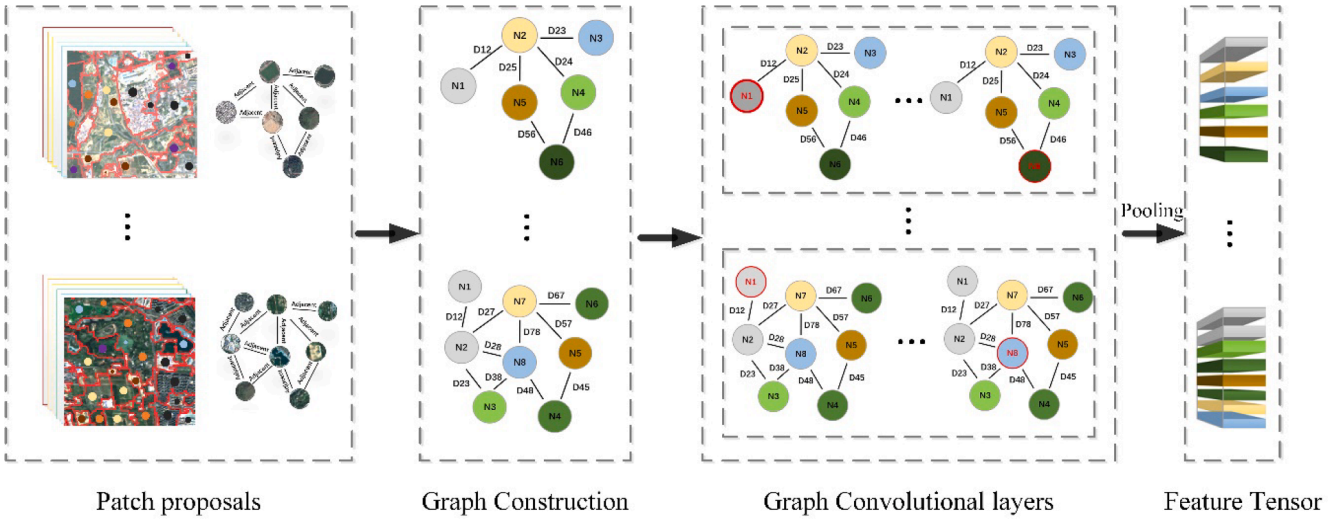


Fig. 7. The workflow of the Mb-GCN module. The node refers to the geometric center point of the object. Different colors represent different class labels. The red node is the node being calculated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

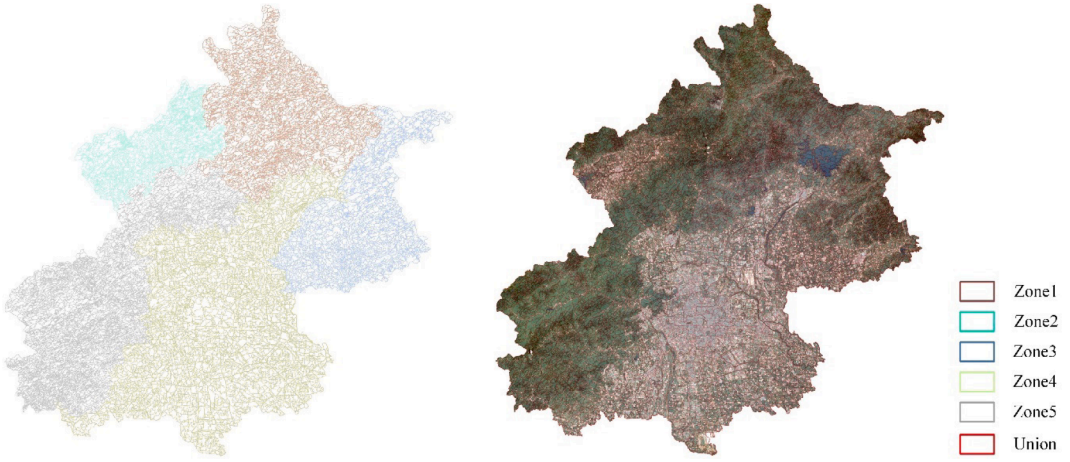


Fig. 8. Results of the adaptive-scale hierarchical segmentation left part and corresponding satellite image. Each zone is symbolized by different colors and contain adaptive segmentation objects.

CNN and GCN making the most significant contributions. Therefore, in the following sections, we focus on the contribution of using DF-CNN and GCN to the impact, including scale, the synergistic mechanism of DF-CNN and GCN and their advantages compared with other methods.

4.3. Results of deep Siamese network change detection

Fig. 9 shows in detail the LULC-SC in Beijing from 2002 to 2022 detected by the deep Siamese Network. Between 2002 and 2010, LULC in Beijing underwent major changes, especially around 2008; thereafter, the changes slowed down, with some apparent changes from 2016 to 2018; there was almost no change in LULC between 2020 and 2022.

Since there is currently no authoritative data on the change of LULC in Beijing, we validated the change maps using random samples. The results are shown in Appendix for Table B, which proves that the detection framework of the deep Siamese network is robust and effective.

5. Discussion

5.1. Effectiveness of the adaptive-scale hierarchical segmentation

To validate the availability of the adaptive scale hierarchical segmentation, the entire image of Beijing in 2002 was input into the network without hierarchy. According to Fig. 10 (a), it is evident that the overall accuracy of the non-hierarchy and fixed-scale approach is lower than that of the strategy we proposed. The optimal accuracy is achieved at a network training scale of 65 (discussed in 5.2), with a

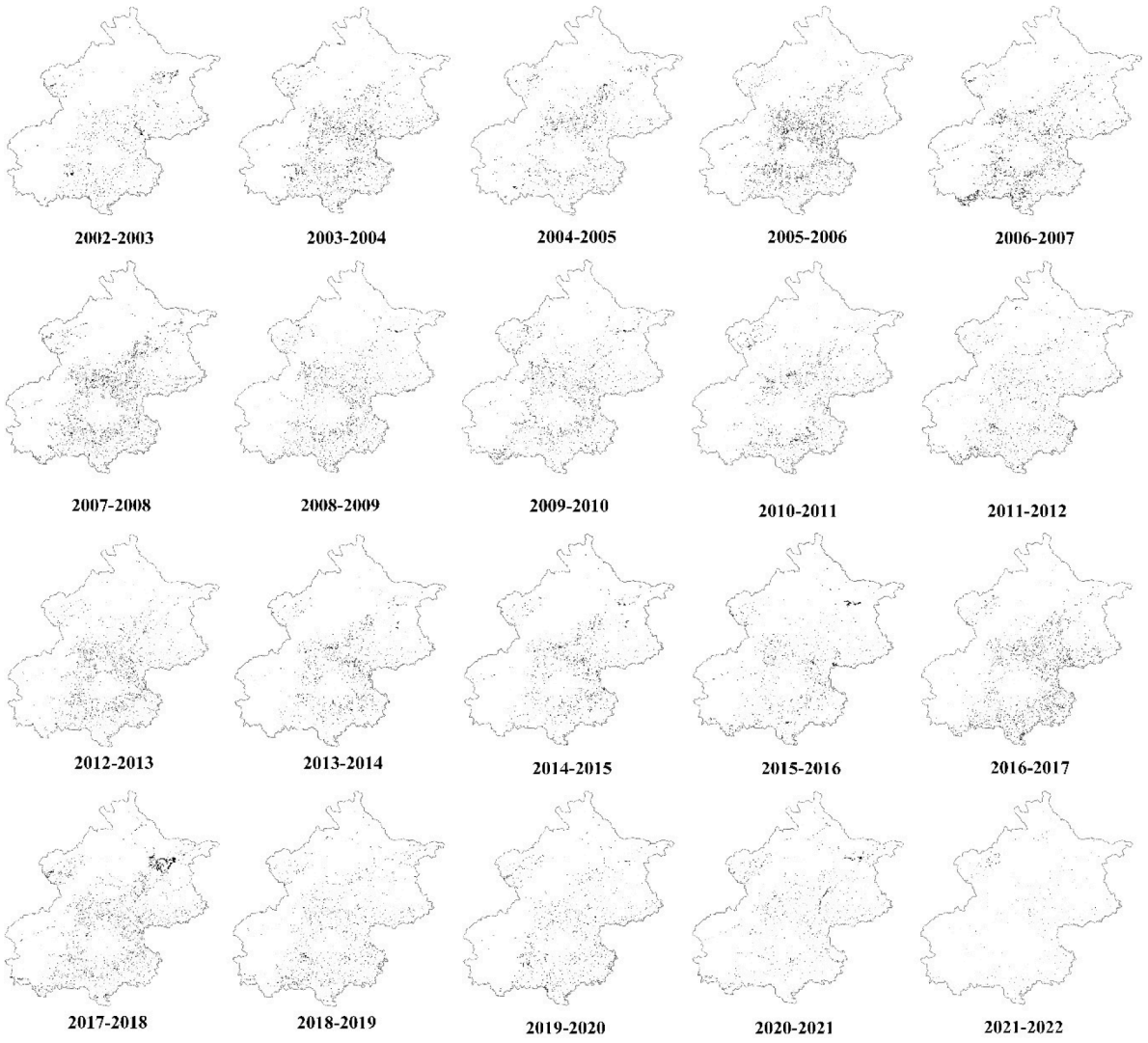


Fig. 9. The annual LULC changes in Beijing. In the maps, changes between years are symbolized by black color.

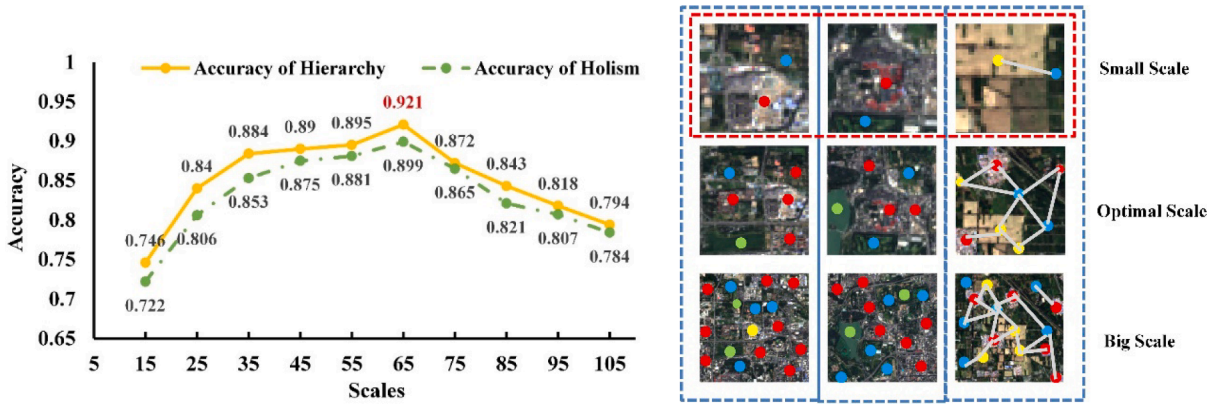


Fig. 10. **(a)** The accuracy of the adaptive-scale hierarchical strategy (Beijing 2002). **(b)** The information content of different training scales.

classification accuracy of 92.1 %. Hierarchy provides a better adaptive environment for segmentation while reducing the computation time. Scale Estimation can find the optimal segmentation scale to improve the efficiency of the experiments (Ming et al., 2015).

5.2. *Training scales effect*

The appropriate scale can capture both macro and micro features of the image, enabling a comprehensive analysis and interpretation of different geographical phenomena and patterns (Zhou et al., 2020). As can be seen from Fig. 10 (b), if the patch is too small, there will be fewer

connections on the adjacent edges, and the relative use of spatial information is insufficient; if the patch is too large, the spatial information may be redundant, and key information cannot be collected.

A separate classifier can realize the output of the intermediate product-classification map. Fig. 11 presents the classification results at different training scales, highlighting the presence of the salt-and-pepper phenomenon in small-scale results. As the scale increases, the phenomenon gradually diminishes, and the accuracy of LULC classification initially improves and then decreases (as shown by the orange line in Fig. 10 (a)). The spatial relationship of vegetation is relatively simple and the proportion of the same type is relatively large. When the scale is large, it cannot focus on the connection of a small number of different types, which results in the large-scale distribution of classified vegetation. The best classification accuracy rate in Beijing in 2002 was scale 65, reaching 92.1 %. Table 3 clarifies the specific accuracy of various scales in Beijing in 2002.

5.3. Advantages of DF-GCN

5.3.1. Feature visualization

To better demonstrate the effectiveness of DF-GCN, we visualized and analyzed the convolution kernel and intermediate features (Wang et al. 2020). Fig. 12 (a) shows a partial visualization of the convolution kernel of the convolution module: at first, it is a simple color filter, which horizontal, vertical and density looks relatively regular. As the number of convolution layers increases, the convolution kernel begins to show a water ripple texture, the density changes continuously and the shape gradually becomes abstract. Finally, the convolution kernel at layer 5–1 abstracts into very complex fossil-like textures.

Fig. 12 (b) indirectly explains the process of deep feature mining: as the layers are gradually deepened, the textures gradually change from fine and detailed to coarse and abstract. Further analysis shows that the different feature texture maps seem to highlight features at different locations of the image, which also clarifies that different abstract convolution kernels match different features of objects at different locations.

5.3.2. Ablation experiment based on DF-GCN

To demonstrate the superiority and robustness of DF-GCN in this paper, we conducted experiments on the same Landsat data (Beijing 2002) using RF, CNN (AlexNet), DF-CNN, GCN and the same combinations with GCN, respectively. The comparison of classification accuracy among the methods is shown in Table 4. Random forest performed worst, while deep learning approaches led to more promising results. Among the previously applied deep learning approaches, AlexNet + GCN had the highest accuracy of 90.03 %. Our new approach outperformed all other approaches and led to an accuracy of 92.10 %.

Table 3
The ablation experiment results in Beijing (2002 Year). RSI: remote sensing imagery. EFM: edge feature map. HS: Hierarchical segmentation Based on object. DFCNN: Deep convolution module and SE module. GCN: local spatial relationship and semantics modeling.

Exp.	Strategy					Overall
	RSI	EFM	HS	DFCNN	GCN	
1	√					0.746
2	√	√				0.787
3	√	√	√			0.832
4	√	√	√	√		0.899
5	√	√	√		√	0.869
6	√	√	√	√	√	0.921

Fig. 13 presents the LULC maps of different combination methods. The Random Forest method performs relatively poorly from a visual perspective, such as misclassifying some forests and grasslands as shrublands. The GCN method, which only considers adjacency relationships, also has mediocre performance. The use of CNN and DF-CNN improves the classification of forests, grasslands, and shrublands. Further improved CNN + GCN and DF-GCN achieve relatively better results in the classification of urban areas, urban green spaces, and farmlands with close relationships. According to the red block in Fig. 13, we can observe the disadvantage of the random forest method in distinguishing vegetation categories, while the introduction of neural networks improves the classification of vegetation. Under the influence of deep-level features and spatial relationships, DF-GCN demonstrates excellent performance in LULC classification. For example, in the enlarged image within the black box, we can observe that the classification result of GCN is relatively chaotic, and CNN has some misclassification of farmland and bare land. Random forest and CNN achieve relatively accurate classification, but relying solely on feature prediction seems to have limitations. Due to the complexity and proximity of LULC types and various interference from adjacency relationships, some misclassifications and false classifications still occur. The two hybrid

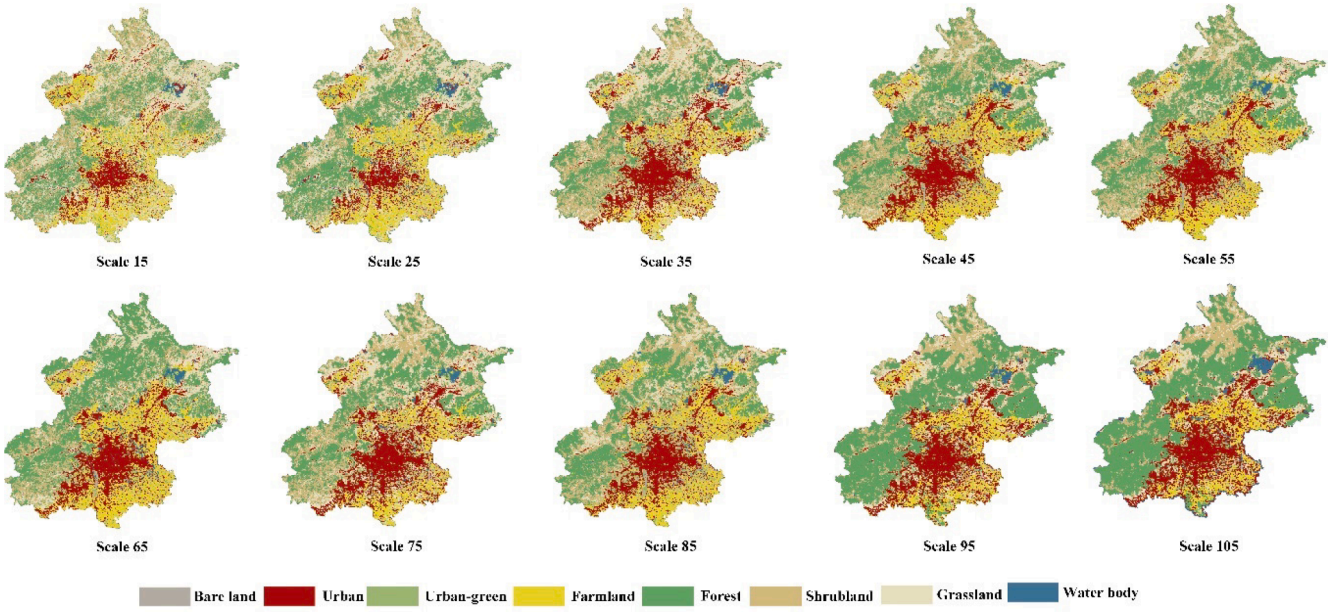


Fig. 11. Classification results for different training scales (Beijing 2002).

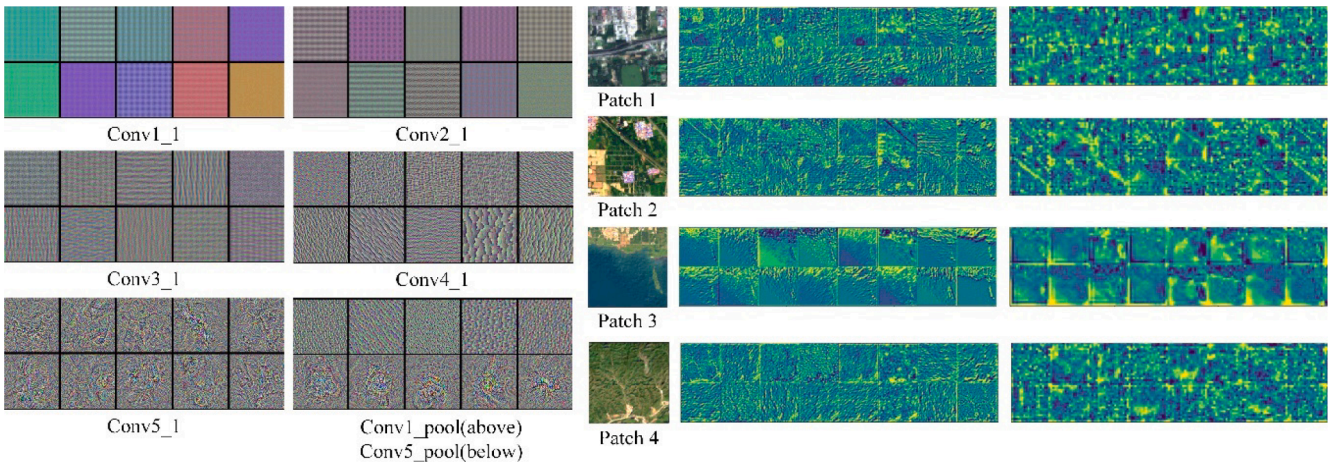


Fig. 12. Visualization of Convolution Kernel (left panel) and visualization of intermediate features (right panel).

Table 4
Overall accuracy of the different scales (year 2022). (T: training accuracy; V: verification accuracy; GT: ground truth accuracy).

Scales	15	25	35	45	55	65	75	85	95	105
T (%)	95.1	94.6	95.5	95.8	96.2	95.6	94.8	94.7	95.6	96.6
V (%)	83.0	88.2	90.1	93.6	93.1	94.5	92.7	90.5	88.8	85.4
GT (%)	76.4	84.0	88.4	89.0	89.5	92.1	87.2	84.3	81.8	79.4

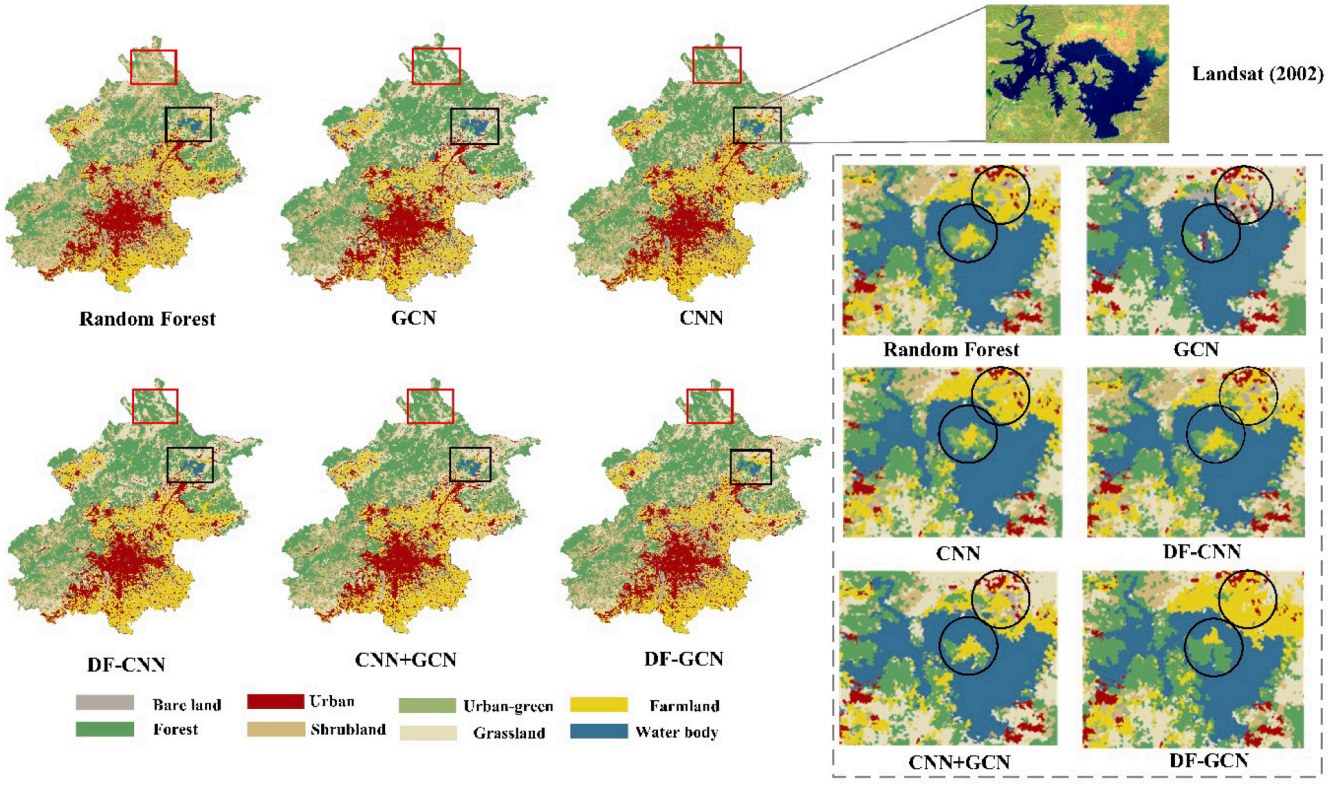


Fig. 13. Fig. 13. Comparison of classification results of the different methods for year 2002. (Red box: macroscopic differences. Black Box: Comparison of details.). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

networks exhibit relatively good extraction results in this area, with DF-GCN outperforming the others, further proving the advantages of DF-GCN in exploring and integrating deep-level features and spatial relationships.

Therefore, conclusions can be drawn based on quantitative and qualitative analysis: RF only considers shallow information. The introduction of AlexNet and DF-CNN addresses the issue of insufficient exploration of deeper features and improves the classification accuracy of LULC. GCN lacks the support of deeper features and does not achieve satisfactory accuracy. The novel DF-GCN model combines deeper

features and spatial information, fully realizing the expression of land features, and provides the possibility for better LULC classification in Beijing (Ding et al., 2022; Yao et al., 2022).

5.4. Advantages of change detection framework based on deep Siamese network

5.4.1. Comparison between feature fusion and decision fusion strategy

Table 5 shows the response time and detection accuracy of the connection fusion method and the additive fusion method (2002). Connection fusion is the merging of feature sets. Although all the information of the two feature maps is retained and the number of features is significantly increased, the information redundancy leads to an increase in the amount of calculation and may even lead to errors. In contrast, additive fusion can enhance important features, weaken noise, speed up response time, and help improve the stability and generalization ability of the network.

5.4.2. Comparison with the State-of-the-Art method of change detection

Since there is no change true value map available, we used the classification map of DF-GCN, which has the highest accuracy. The difference between 2002 and 2003 were calculated to approximate the true value. Table 6 shows the numerical results of different change detection methods. Our proposed deep Siamese-based change detection framework achieves relatively good results.

The change maps obtained by LandTrendr have salt-and-pepper spots and contain false detections. BiT has a Transformer module that can emphasize the connection between high-level semantic features, and the detection results are relatively complete. However, there are also a few discontinuities and pieces. The above methods are all due to pixel-level detection, which results in ground objects being granular and partially blurred. Compared with other methods, our proposed detection framework focuses on the object level, and the detection results are closer to the real situation of ground objects, with fewer missed and false detections. At the same time, compared with MSPSNet, our method also focuses on object spatial relationships and semantics, thus showing good detection results. Regardless of qualitative or quantitative analysis, our proposed deep Siamese-based change detection framework is effective in object-level multi-scale feature extraction and spatial semantic modeling.

5.5. Comparison with GlobeLand30

To demonstrate the validity of the method under consideration of the time span and availability of the data, we selected three existing products for comparison. Table 7 shows the accuracy of different products. Combined with the results in Table A in the Appendix, the overall accuracy of our results is higher than existing products. We also provided several detailed views to explore the differences in local area classification. The LULC maps presented in this paper showed better coherence compared to GlobeLand30.This is because features need to be expressed at an appropriate spatial scale rather than the original pixel, and GlobeLand30 focuses more on global coverage, which limits its ability to capture local detail.

Furthermore, the changes over the 10-year period differed as well. In Fig. 15, the change detection results in this study were more detailed and smoother compared to GlobeLand30, which can be attributed to the consideration of both feature-level and spatial relationship

Table 5
Comparison of overall accuracy of the different methods.

Methods	RF	AlexNet	DF-CNN	GCN	AlexNet + GCN	DF-GCN
Overall accuracy (%)	80.67	85.87	88.65	82.37	90.03	92.10

Table 6
The Overall accuracy of the different methods.

Method	Connection fusion	Additive fusion
Pre	93.1	92.5
Rec	95.5	95.3
OA	96.9	96.6
Response time	33 min	22 min

Table 7
Numerical results of different change detection methods. (2002–2003).

Method	Landtrendr	MSPSNet	BiT	Our
Pre	86.24	90.11	88.67	92.5
Rec	88.37	91.59	90.37	95.3
OA	90.9	93.5	91.4	96.6

contributions at the object level (Chen et al., 2019).

As shown in Fig. 16, CGLC also has classification coherence problem, for example near water bodies (Fig. 16 (A1), this is due to too low resolution and fuzzy boundaries at the pixel level. Likewise, airport is mistakenly classified as bare-land. In addition, EAS WorldCover achieved better classification results with the help of high resolution. The boundaries of ground objects were clear, but there were some deviations in the division of forests, shrubland and grasslands Fig. 16 (A2).

Since the focus is laid on the object level, the proposed DF-GCN-based construction shows better p erformance in the classification (Table 8) of the Beijing area.

5.6. 20 years of LULC change analysis

With the development of industrialization and urbanization in Beijing, land resources are facing increasing spatial and environmental pressures. Beijing has experienced one of the highest rates of LULC change over the past two decades, mainly driven by urban expansion and a decrease in farmland.

Based on the analysis of changes in Beijing from 2002 to 2022, it is obvious that the urban area shows a growing trend, while the cultivated land area is relatively reduced, and other land use categories have relatively little change (Fig. 17). This is mainly due to the stimulation of the urban economy, and the population inflow into the city leads to the formation of an expansion-development cycle (Mahtta et al., 2022). Therefore, the urban area is continuing to expand, growing by approximately 11.58 % (Fig. 17 (C)). Consequently, a large amount of farmland and other vegetation land types have been cut down and developed into urban areas. Among them, farmland, which decreased by approximately 825.6 square kilometers (Fig. 17 (D)). Especially around 2008, hosting the Beijing Olympics stimulated economic growth, employment needs, and accelerated the urbanization process. However, excessive urbanization has caused exponential population growth. As of 2015, Beijing's population was 21.705 million, more than double that of 2002. The exponential growth of population has led to Beijing's urban congestion, increased emissions and pollution, urban heat island effect, and increased disaster risks. Simultaneously, the quality of people's living environment has declined, the proportion of urban green space has continued to decrease, and there are irregularities in the water's quality.

For the sustainable development of cities, China implemented the "Thirteenth Five-Year Plan" in 2016, focusing on ecological protection and environmentally sustainable development. Therefore, the urbanization process slowed down from 2018 to 2020, focusing on protecting and maintaining the ecological environment around the city; the proportion of urban green spaces and water bodies increased relatively (Fig. 17 (B1 and B2)), which improved the use of resources and alleviated the pressure on the city. However, due to the COVID-19 epidemic, Beijing's city closure policy has caused urbanization to stagnate until 2022 when the city opened more intensively. The increase in water

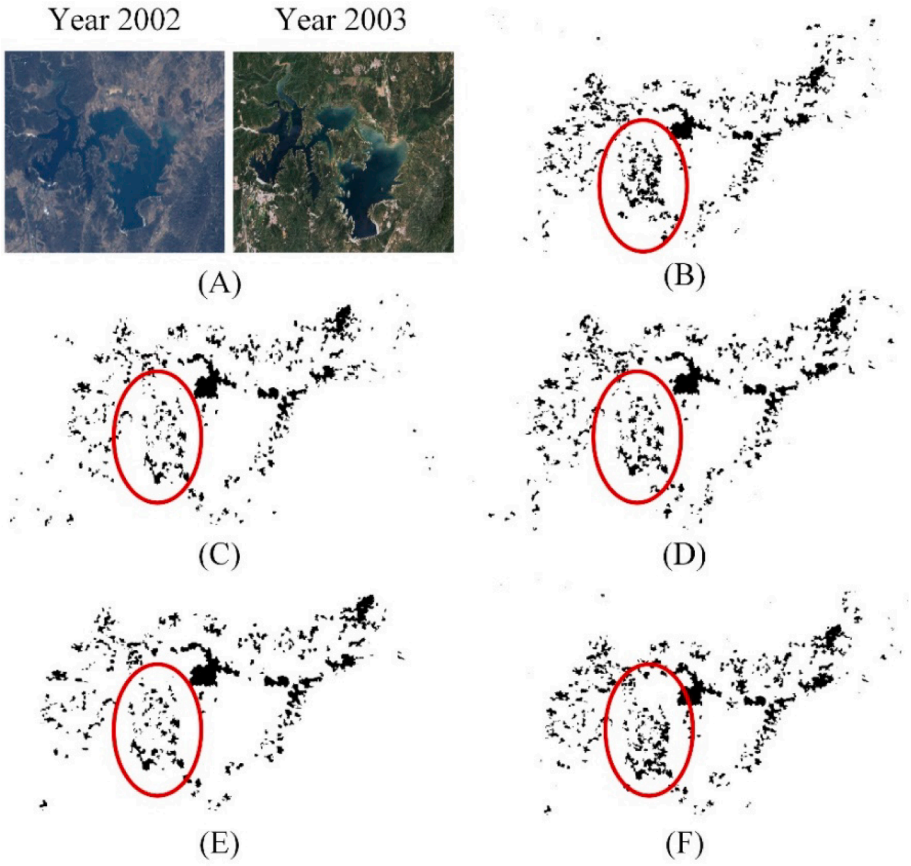


Fig. 14. Change detection results of different methods (for years 2002 to 2003). (A) The left is the 2002 image and the right is the 2003 image. (B) Difference between 2002 and 2003 classification maps (DF-GCN). (C) Change maps of LandTrendr (D) BiT (E) MSMPNet and (F) our method.

bodies in Beijing since 2015 is mainly caused by the water storage of the Miyun Reservoir in the northeast of Beijing. In addition, the construction of numerous artificial lakes in new residential areas also contributed to increase.

The change detection improves the readability and interpretability of land change trends and policies in the past 20 years. At the same time, the regularities and patterns can help planners better understand the development trajectory of cities and formulate more effective urban planning and policies to cope with urban growth and changes.

Furthermore, to respond to the sustainable development of cities, LULC-SC helps to assess the degree of sustainability, improve the efficiency of resource utilization, and improve the ecological health of the city.

5.7. Limitations of the study

The quantity and quality of the samples in this study may introduce some uncertainty (Lv et al., 2018). The construction of the graph should consider more spatial distances and relationship weights. Additionally, the validation of long-term continuous LULC-SC maps is an onerous task as obtaining multi-year reference data can be difficult. Moreover, ground truth data is obtained through manual visual interpretation, which also carries a certain level of uncertainty. In fact, the low temporal frequency may sometimes result in uncertainties due to the unavailability of high-resolution imagery for every year. Furthermore, this study only focuses on Beijing as an example, and the urban diversity and complexity is relatively limited. In the future, this method will be applied to multiple cities to evaluate its applicability.

6. Conclusions

In this study, we proposed a Siamese-based DF-GCN spatiotemporal change detection framework for mapping land cover in Beijing using Landsat time series data from 2002 to 2022. We generated annual, multi-class land cover maps and produced LULC-SC maps for Beijing. Compared to various classification algorithms, the proposed DF-GCN algorithm effectively integrates deep-level features and spatial adjacency relationships, achieving accurate classification of LULC in Beijing with classification accuracies exceeding 90 %. Moreover, the deep Siamese Network enables rapid comparison of spatial semantic tensors in time series of remote sensing images, facilitating the generation of annual-scale LULC maps for Beijing. Therefore, the method proposed in this study allows for the effective mapping of LULC over long time series, and the results can be used to planning and optimize LULC patterns, providing comprehensive knowledge to promote and coordinate regional sustainable development. (Gong et al., 2022).

CRediT authorship contribution statement

Hanqing Bao: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Vinzenz H.D. Zerres:** Writing – review & editing, Supervision. **Lukas W. Lehnert:** Writing – review & editing, Supervision, Resources, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

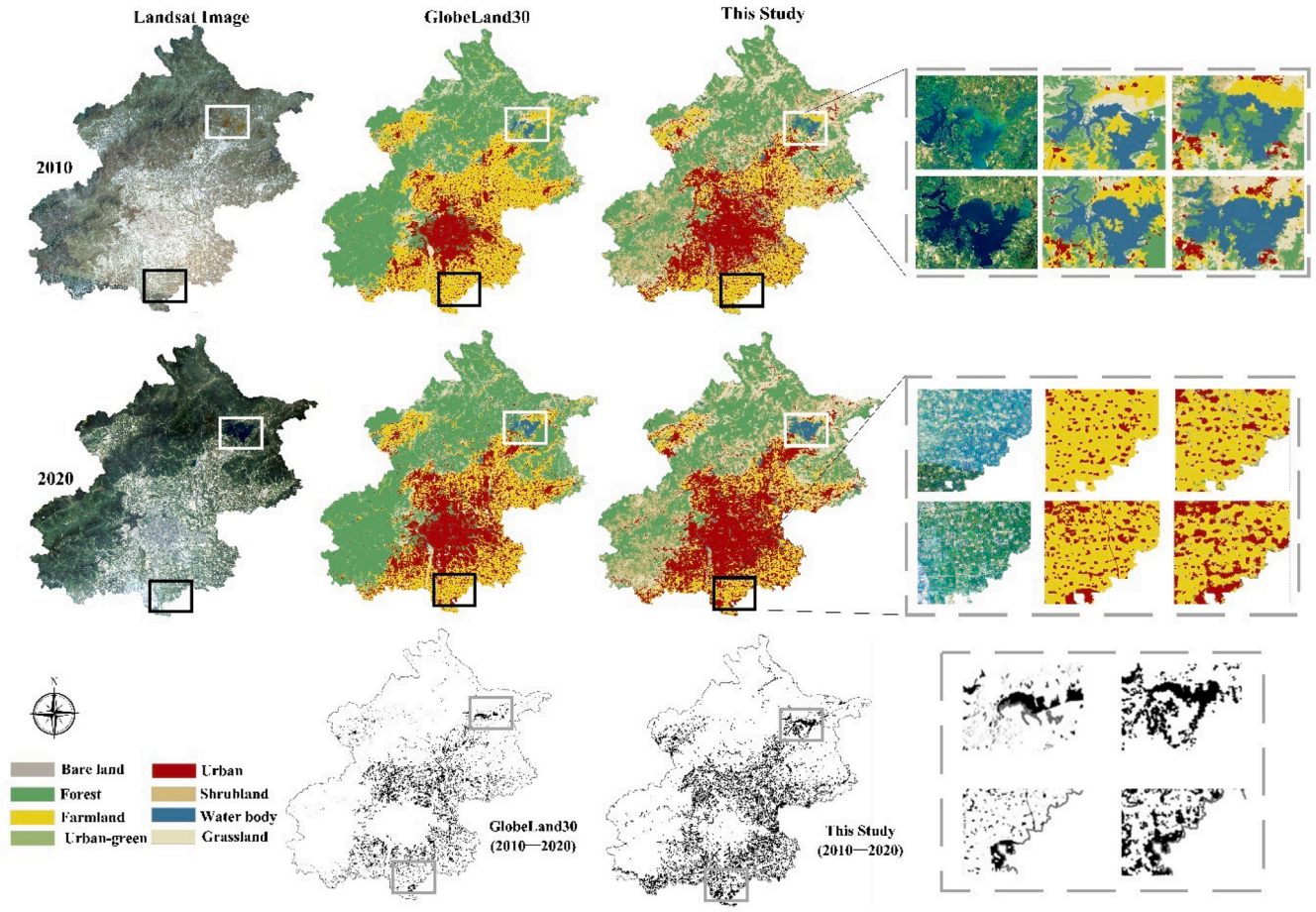


Fig. 15. Comparison between results of our method and existing GlobeLand30 data. (White box: Detailed view near the reservoir; Black box: detail of the suburbs; Gray box: 10 years of change).

Data availability

Data will be made available on request.

Appendix

Table A1. The Overall accuracy of each year in Beijing. (T: training accuracy; V: verification accuracy; GT: ground truth accuracy).

Year	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Optimal-Scale	65	65	55	65	55	55	55	65	65	65	65
T (%)	95.6	96.6	96.3	97.0	97.0	96.4	95.4	97.1	96.9	97.2	97.5
V (%)	94.5	96.4	95.7	95.6	96.3	96.0	95.6	96.8	95.8	95.9	96.5
GT (%)	**92.1**	**93.2**	**94.0**	**93.5**	**92.9**	**95.3**	**93.6**	**94.5**	**93.9**	**94.7**	**94.2**

Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Optimal-Scale	65	65	55	65	55	65	55	55	65	65
T (%)	97.6	97.5	96.8	95.6	96.8	96.9	96.8	97.4	97.3	96.7
V (%)	96.1	96.4	95.4	95.1	96.1	96.3	95.9	96.4	96.0	95.3
GT (%)	**93.6**	**92.9**	**94.2**	**93.2**	**94.6**	**94.1**	**94.0**	**94.3**	**93.8**	**93.5**

Table B1. Accuracy verification of change maps.

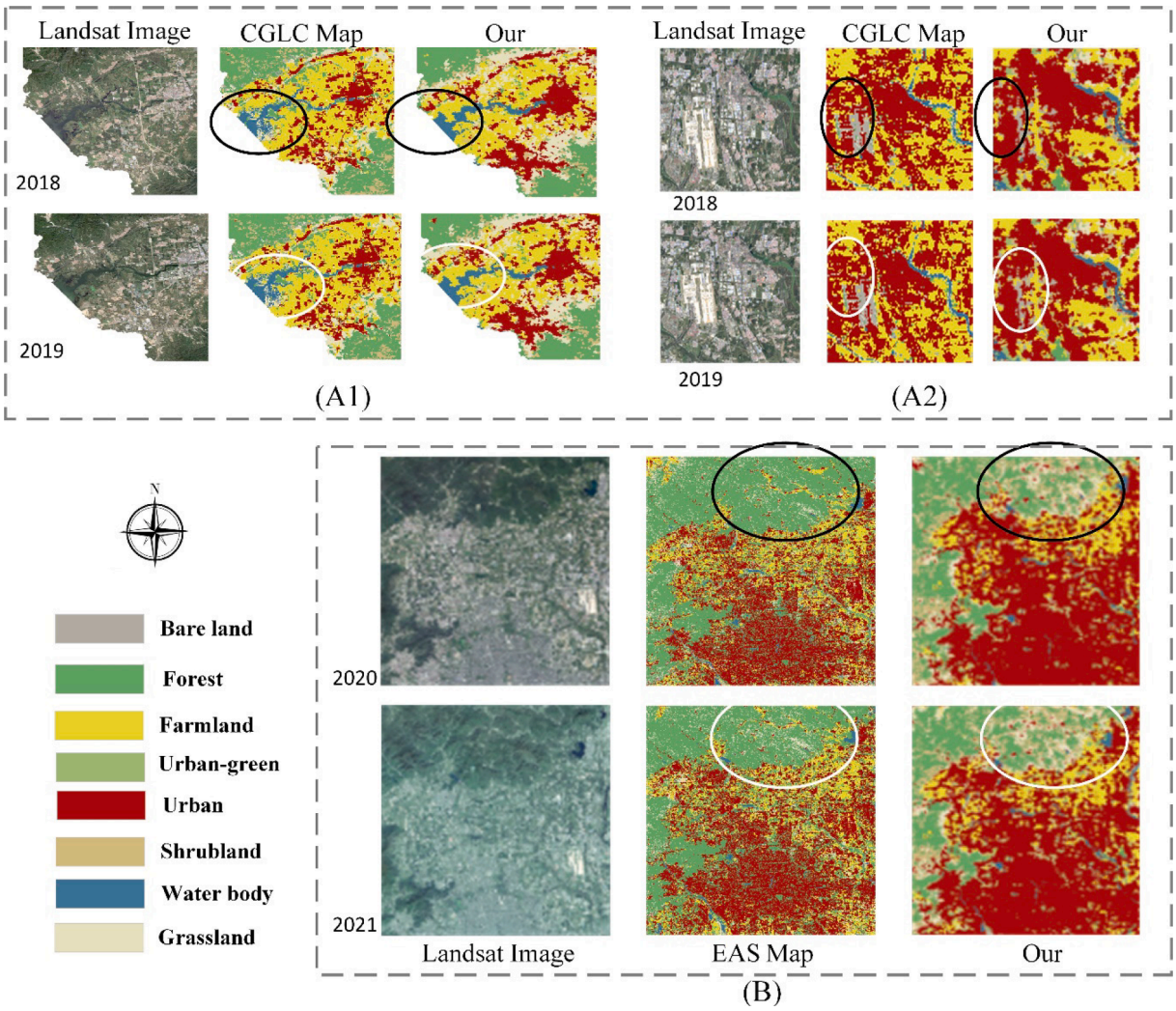


Fig. 16. Comparison between results of our method and existing other data produces. (A) Comparison with Copernicus Global Land Cover Map. (B) Comparison with EAS WorldCover Map. (Ellipses show details).

Table 8
The Overall accuracy of the different datasets.

Produces	GlobeLand30		Copernicus Global Land Cover		EAS WorldCover	
Year	2010	2020	2018	2019	2020	2021
Overall Accuracy (%)	81.6	82.9	78.3	80.7	88.6	90.1

Year	2002–2003	2003–2004	2004–2005	2005–2006	2006–2007
Precision (%)	92.5	92.3	93.3	93.0	94.0
Recall (%)	95.3	94.4	95.7	94.4	95.7
Accuracy (%)	96.6	96.2	96.0	95.5	96.9

Year	2007–2008	2008–2009	2009–2010	2010–2011	2011–2012
Precision (%)	92.4	93.4	93.1	92.9	91.2
Recall (%)	93.0	95.6	94.8	93.8	93.9
Accuracy (%)	94.3	96.6	95.5	95.9	95.7

Year	2012–2013	2013–2014	2014–2015	2015–2016	2016–2017
Precision (%)	91.5	91.6	93.5	92.8	92.6
Recall (%)	94.5	93.1	95.4	95.4	95.1
Accuracy (%)	95.2	95.6	96.9	97.5	97.2

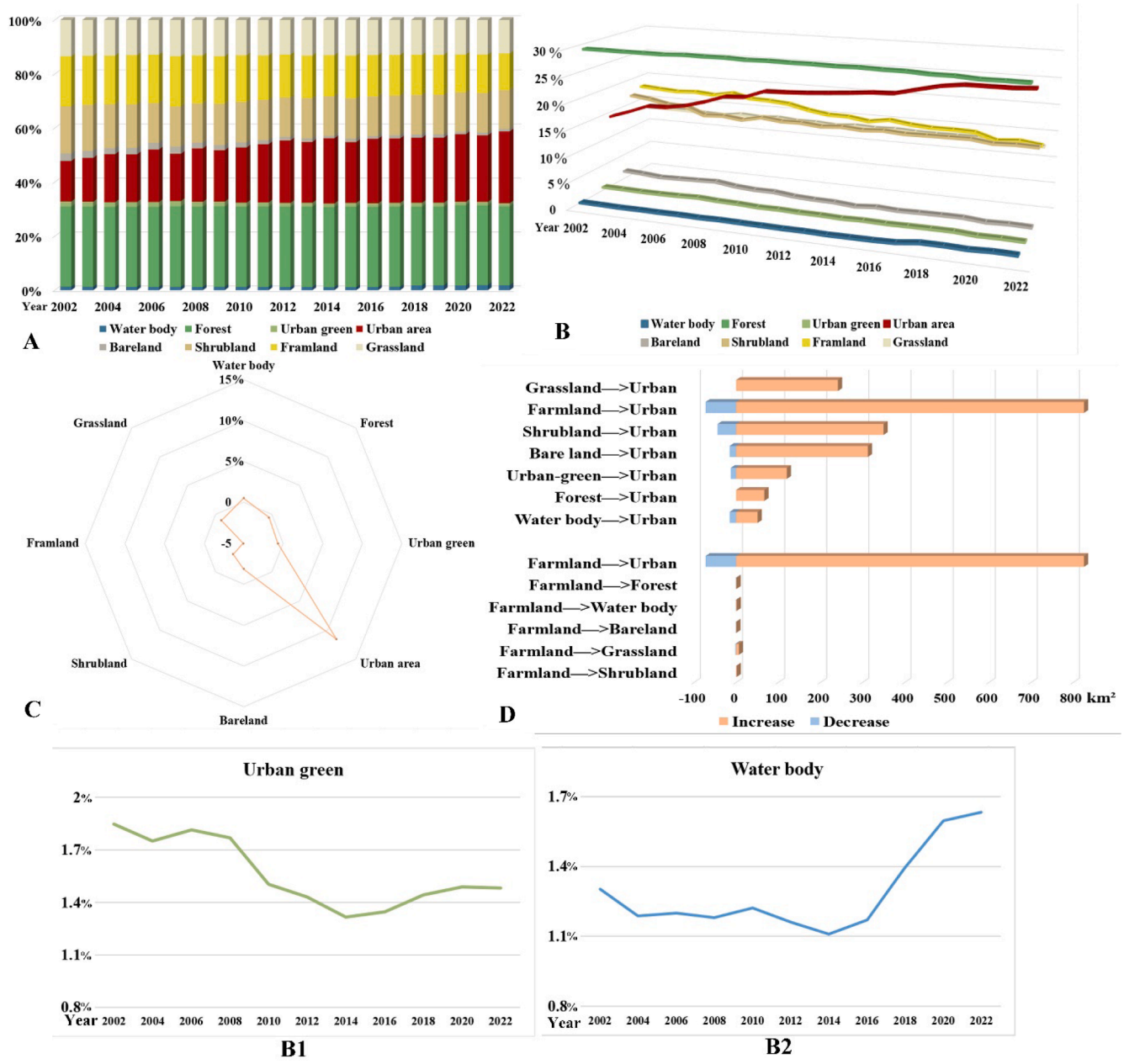


Fig. 17. Analysis of changes from 2002 to 2022. **(A)** Land-use proportion. **(B)** Land-use change curve. **(C)** Land-use change radar chart 2002–2022. **(D)** Conversion statistics chart (2002–2022, Urban and Farmland).

Year	2017–2018	2018–2019	2019–2020	2020–2021	2021–2022
Precision (%)	91.8	92.9	91.8	93.4	93.3
Recall (%)	93.6	94.3	93.9	96.4	95.0
Accuracy (%)	95.6	96.1	96.0	97.3	96.8

Fig. A1. Classification results for each year from 2002 to 2022 in Beijing.

References

Bao, H., Ming, D., Guo, Y., Zhang, K., Zhou, K., Du, S., 2020. DFCNN-Based Semantic Recognition of Urban Functional Zones by Integrating Remote Sensing Data and POI Data. Remote Sens. (Basel) 12, 1088.

Baohui, C., Peijun, L., 2023. An ensemble method for monitoring land cover changes in urban areas using dense Landsat time series data. ISPRS J. Photogramm. Remote Sens. 195, 29–42.

Cetin, M., Aksoy, T., Cabuk, S.N., Kurkcuoglu, M.A.S., Cabuk, A., 2021. Employing remote sensing technique to monitor the influence of newly established universities in creating an urban development process on the respective cities. Land Use Policy 109, 105705.

Chen, J., Cao, X., Peng, S., Ren, H., 2017. Analysis and Applications of GlobeLand30: A Review. ISPRS Int. J. Geo Inf. 6, 230.

Chen, Y., Ming, D., Lv, X., 2019. Superpixel based land cover classification of VHR satellite image combining multi-scale CNN and scale parameter estimation. Earth Sci. Inf. 12, 341–363.

David, J., Ce, Z., 2022. An attention-based U-Net for detecting deforestation within satellite sensor imagery. Int. J. Appl. Earth Obs. Geoinf. 107, 102685.

Ding, Y., Zhang, Z., Zhao, X., Hong, D., Cai, W., Yu, C., Yang, N., Cai, W., 2022. Multi-feature fusion: Graph neural network and CNN combining for hyperspectral image classification. Neurocomputing 501, 246–257.
Drăguţ, L., Csillik, O., Eisank, C., Tiede, D., 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. ISPRS J. Photogramm. Remote Sens. 88, 119–127.
Gong, Y., Cai, M., Yao, L., Cheng, L., Hao, C., Zhao, Z., 2022. Assessing Changes in the Ecosystem Services Value in Response to Land-Use/Land-Cover Dynamics in Shanghai from 2000 to 2020. Int. J. Environ. Res. Public Health 19, 12080.
Fotso Kamga Guy, A., Tallha, A., Bitjoka, L., Syed Rameez, N., Mengue Mbom, A., & Nazeer, M. (2018). A deep heterogeneous feature fusion approach for automatic land-use classification. Informat. Sci., 467, 199-218.
Hong, D., Gao, L., Yao, J., Zhang, B., Plaza, A., Chanussot, J., 2020. Graph convolutional networks for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 59, 5966–5978.
Hongyang, Y., Liguo, W., Yan, L., Min, X., Kai, H., Haifeng, L., Ming, Q., 2023. Attention-guided siamese networks for change detection in high resolution remote sensing images. Int. J. Appl. Earth Obs. Geoinf. 117, 103206.
Huabing, H., Yanlei, C., Nicholas, C., Jie, W., Xiaoyi, W., Caixia, L., Peng, G., Jun, Y., Yuqi, B., Yaomin, Z., Zhiliang, Z., 2017. Mapping major land cover dynamics in Beijing using all Landsat images in Google Earth Engine. Remote Sens. Environ. 202, 166–176.
Huanxue, Z., Mingxu, L., Yuji, W., Jiali, S., Xiangliang, L., Bin, L., Aiqi, S., Qiangzi, L., 2021. Automated delineation of agricultural field boundaries from Sentinel-2 images using recurrent residual U-Net. Int. J. Appl. Earth Obs. Geoinf. 105, 102557.
Jafarzadeh, H., Mahdianpari, M., Gill, E.W., 2022. Wet-GC: A Novel Multimodel Graph Convolutional Approach for Wetland Classification Using Sentinel-1 and 2 Imagery With Limited Training Samples. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 5303–5316.
Jiang, H., Hu, X., Li, K., Zhang, J., Gong, J., Zhang, M., 2020. PGA-SiamNet: Pyramid feature-based attention-guided Siamese network for remote sensing orthoimagery building change detection. Remote Sens. (Basel) 12, 484.
Jiaqi, Y., Bo, D., Liangpei, Z., 2023. From center to surrounding: An interactive learning framework for hyperspectral image classification. ISPRS J. Photogramm. Remote Sens. 197, 145–166.
Junfu, F., Qingyun, L., Zhoupeng, R., Zheng, C., Wenqiang, L., Yong, Y., Yuke, Z., 2022. Nighttime luminosity transitions are tightly spatiotemporally correlated with land use changes: A pixelwise case study in Beijing. China. *Ecological Indicators* 145, 109649.
Li, Y., Chen, R., Zhang, Y., Li, H., 2020. In: A CNN-GCN Framework for Multi-Label Aerial Image Scene Classification. IEEE, pp. 1353–1356.
Liang, J., Deng, Y., Zeng, D., 2020. A deep neural network combined CNN and GCN for remote sensing scene classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 4325–4338.
Liu, Q., Xiao, L., Yang, J., Wei, Z., 2020. CNN-enhanced graph convolutional network with pixel-and superpixel-level feature fusion for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 59, 8657–8671.
Lv, X., Ming, D., Lu, T., Zhou, K., Wang, M., Bao, H., 2018. A new method for region-based majority voting CNNs for very high resolution image classification. Remote Sens. (Basel) 10, 1946.
Mahtta, R., Fragkias, M., Güneralp, B., Mahendra, A., Reba, M., Wentz, E.A., Seto, K.C., 2022. Urban land expansion: the role of population and economic growth for 300+ cities. Npj Urban Sustainability 2, 5.
Meng, Z., Huaiqing, Z., Bo, Y., Hui, L., Xuexian, A., Yang, L., 2023. Spatiotemporal changes of wetlands in China during 2000–2015 using Landsat imagery. J. Hydrol. 621, 129590.
Ming, D., Li, J., Wang, J., Zhang, M., 2015. Scale parameter selection by spatial statistics for GeOBIA: Using mean-shift based multi-scale segmentation as an example. ISPRS J. Photogramm. Remote Sens. 106, 28–41.
Mohan, S., Kapil Dev, T., 2021. Pixel based classification for Landsat 8 OLI multispectral satellite images using deep learning neural network. Remote Sens. Appl.: Soc. Environ. 24, 100645.
Qiqi, Z., Xi, G., Weihuan, D., Sunan, S., Qingfeng, G., Yanfei, Z., Liangpei, Z., Deren, L., 2022. Land-Use/Land-Cover change detection based on a Siamese global learning framework for high spatial resolution remote sensing imagery. ISPRS J. Photogramm. Remote Sens. 184, 63–78.
Vincent, B.V., Irene, C.D., 2022. Annual satellite-based NDVI-derived land cover of Europe for 2001–2019. J. Environ. Manage. 302, 113917.
Wang, H., Gao, K., Zhang, X., Wang, J., Hu, Z., Yang, Z., Mao, Y., Liu, Y., 2023. In: Pixel- and Patch-Wise Context-Aware Learning with CNN and GCN Collaboration for Hyperspectral Image Classification. IEEE, pp. 7555–7558.
Xiangyu, Z., Jingliang, H., Lichao, M., Zhitong, X., Xiao Xiang, Z., 2023. Cross-city Landuse classification of remote sensing images via deep transfer learning. Int. J. Appl. Earth Obs. Geoinf. 122, 103358.
Xie, S., Liu, L., Zhang, X., Yang, J., 2022. Mapping the annual dynamics of land cover in Beijing from 2001 to 2020 using Landsat dense time series stack. ISPRS J. Photogramm. Remote Sens. 185, 201–218.
Xu, L., Ming, D., Zhou, W., Bao, H., Chen, Y., Ling, X., 2019. Farmland extraction from high spatial resolution remote sensing images based on stratified scale pre-estimation. Remote Sens. (Basel) 11, 108.
Xuexian, A., Wenping, J., Huaiqing, Z., Yang, L., Meng, Z., 2022. Analysis of long-term wetland variations in China using land use/land cover dataset derived from Landsat images. Ecol. Ind. 145, 109689.
Yao, D., Zhili, Z., Xiaofeng, Z., Danfeng, H., Wei, C., Chengguo, Y., Nengjun, Y., Weiwei, C., 2022. Multi-feature fusion: Graph neural network and CNN combining for hyperspectral image classification. Neurocomputing 501, 246–257.
Yongyang, X., Bo, Z., Shuai, J., Xuejing, X., Zhanlong, C., Sheng, H., Nan, H., 2022. A framework for urban land use classification by integrating the spatial context of points of interest and graph convolutional neural network method. Comput. Environ. Urban Syst. 95, 101807.
Zhang, X., Ge, Y., Ling, F., Chen, J., Chen, Y., Jia, Y., 2021. Graph convolutional networks-based super-resolution land cover mapping. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 7667–7681.
Zhang, S., Tong, H., Xu, J., Maciejewski, R., 2019. Graph convolutional networks: a comprehensive review. Computational Social Networks 6, 1–23.
Zhang, S., Zhang, Y., Yu, J., Fan, Q., Si, J., Zhu, W., Song, M., 2022. Interpretation of the spatiotemporal evolution characteristics of land deformation in beijing during 2003–2020 using sentinel, ENVISAT, and Landsat data. Remote Sens. (Basel) 14, 2242.
Zhao, W., Bo, Y., Chen, J., Tiede, D., Blaschke, T., Emery, W.J., 2019. Exploring semantic elements for urban scene recognition: Deep integration of high-resolution imagery and OpenStreetMap (OSM). ISPRS J. Photogramm. Remote Sens. 151, 237–250.
Zheng, Y., He, Y., Zhou, Q., Wang, H., 2022. Quantitative evaluation of urban expansion using NPP-VIIRS nighttime light and landsat spectral data. Sustain. Cities Soc. 76, 103338.
Zhenshi, L., Xueliang, Z., Pengfeng, X., 2022. Spectral index-driven FCN model training for water extraction from multispectral imagery. ISPRS J. Photogramm. Remote Sens. 192, 344–360.
Zhimin, W., Jiasheng, W., Kun, Y., Limeng, W., Fanjie, S., Xinya, C., 2022. Semantic segmentation of high-resolution remote sensing images based on a class feature attention mechanism fused with Deeplabv3+. Comput. Geosci. 158, 104969.
Zhou, H., Luo, F., Zhuang, H., Weng, Z., Gong, X., Lin, Z., 2023. Attention Multi-hop Graph and Multi-scale Convolutional Fusion Network for Hyperspectral Image Classification. IEEE Trans. Geosci. Remote Sens.
Zhou, W., Ming, D., Lv, X., Zhou, K., Bao, H., Hong, Z., 2020. SO–CNN based urban functional zone fine division with VHR remote sensing image. Remote Sens. Environ. 236, 111458.