# HRMS-SCD：A High-Resolution Multi-Scene Satellite Imagery Dataset for Comprehensive Land-Cover Semantic Change Detection

Peixin Guo [1,2], Siyu Yang [1], Hanchao Zhang [1,*], Xiao Huang [3], Xiaogang Ning [1], Yilong Han [2], Ruiqian Zhang [1], Minghui Hao [1]

[1] Institute of Photogrammetry and Remote Sensing, Chinese Academy of Surveying and Mapping, Beijing, China
- 17xiao1223@gmail.com(P.G.); - ysy2000yrqs@outlook.com(S.Y.); - zhanghc@casm.ac.cn(H.Z.); - ningxg@casm.ac.cn(X.N.); - zhangrq@casm.ac.cn(R.Z.); - haomh@casm.ac.cn(M.H.)
[2] College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao, Shandong, China
- 17xiao1223@gmail.com(P.G.); - hanyl@sdust.edu.cn(Y.H.)
[3] Department of Environmental Sciences, Emory University, Atlanta, GA, USA - xiao.huang2@emory.edu

**Keywords:** Semantic Change Detection, Remote Sensing, High Resolution, Multi-Scene, Satellite Imagery Dataset.

## Abstract

Semantic change detection (SCD) focuses on identifying changes in surface coverage while simultaneously classifying the types of changes. This approach provides detailed information valuable for urban planning, environmental monitoring, and other applications, making it a key area of interest in remote sensing research. Despite recent advances, existing SCD studies are hindered by the lack of high-resolution satellite imagery datasets and insufficiently comprehensive semantic label coverage in publicly available datasets. To address these limitations, we have developed a large-scale high-resolution remote sensing dataset consisting of 11,587 satellite image pairs, each with 1-meter spatial resolution and a size of 512 × 512 pixels, representing land cover changes across Beijing between 2017 and 2018. This dataset encompasses diverse land surface scenes with comprehensive semantic annotations. Furthermore, it includes full-coverage semantic segmentation labels from pre-change phases and a larger sample size of 2048 × 2048 pixels to support future research on multi-class and large-format change detection. We benchmark eight state-of-the-art SCD algorithms using this dataset, providing critical performance metrics that serve as valuable references for subsequent research. This dataset not only addresses existing gaps but also establishes a robust foundation for advancing deep learning-based semantic change detection, enabling more accurate and comprehensive analysis of complex and diverse land cover changes. More information about the project can be found at https://github.com/17x-osborn/HRMS-SCD.

## 1. Introduction

Change detection involves recognizing differences in the state of an object or phenomenon by observing it at different periods (Cao et al., 2023). This process is of significant importance in fields such as urban planning and environmental protection. As a crucial branch of remote sensing image analysis, semantic change detection extends beyond locating the range of change, as in binary change detection, to identifying the types of changes, making it a focal point in current research (Tollerud et al., 2023).

In recent years, the application of deep learning in change detection has led to significant breakthroughs in remote sensing image analysis. Unlike traditional feature engineering-based methods, deep learning can automatically learn change features in images by building end-to-end models, exhibiting strong advantages, especially in handling complex scenes and multi-scale changes. Notably, classical network architectures such as Fully Convolutional Networks (Wu et al., 2023), Convolutional Neural Networks (Ding et al., 2022; Feng et al., 2023; Guo et al., 2021; Huang et al., 2024), and CycleGAN (Huang and Zhang, 2024; Zhang et al., 2022), as well as more advanced models like Transformers (Chen et al., 2021; Lin et al., 2024; Zheng et al., 2022) and CNN–Transformer networks (Jiang et al., 2024; Li et al., 2023a; Niu et al., 2023), have achieved remarkable results in change detection tasks. Being data-driven methods, the performance of deep learning models heavily depends on the quality and scale of the datasets used. Therefore, high-quality, diverse, and high-resolution change detection datasets are critical for effective model development.

To evaluate the effectiveness of change detection methods across different applications, researchers have released datasets such as HRSCD(Daudt et al., 2019), SECOND(Yang et al., 2020), Landsat-SCD(Yuan et al., 2022). While these existing semantic change detection datasets in remote sensing have significantly advanced the field, they have several notable limitations. Most are derived from aerial imagery or medium- to low-resolution satellite images. Although aerial imagery provides high spatial resolution, it is costly, covers limited areas, and is affected by atmospheric conditions and changing illumination. Medium- to low-resolution satellite images lack the spatial detail necessary to detect subtle or small-scale changes. Additionally, many datasets are limited to specific scenes, reducing their generalizability and applicability. The absence of multi-scene datasets restricts models' ability to generalize across diverse environments, which is crucial for handling real-world complexities. Finally, many datasets have sparse or incomplete label categories, limiting studies on common land use and cover transitions and thus constraining the research scope and understanding of a broader range of semantic changes.

To address these shortcomings, we introduce the HRMS-SCD dataset—a large-scale, high-resolution, multi-scene satellite imagery dataset for semantic change detection. This dataset consists of 11,587 image pairs with a spatial resolution of 1 meter, each sized at 512 × 512 pixels, collected over Beijing from 2017 to 2018. The HRMS-SCD dataset exhibits several key features. First, it covers diverse land surface types, including seven common categories such as planted land, forest and grass cover, and buildings, with comprehensive


https://doi.org/10.5194/isprs-annals-X-G-2025-323-2025  

semantic annotations. Second, it provides two image sizes (512 × 512 pixels and 2048 × 2048 pixels) to accommodate different computational resources and task requirements. The larger images preserve global context information, while the smaller images offer memory efficiency, making them suitable for routine training and rapid experimentation. Lastly, it spans multiple geographic environments—including urban, rural, mountainous, forest, cropland, and river scenes—enhancing model adaptability and robustness across complex scenarios. This dataset provides a robust foundation for advancing semantic change detection research.

To assess the effectiveness of our proposed HRMS-SCD dataset, we compare it against existing general semantic change detection datasets, presenting their differences in Table 1. In our experiments, we apply multiple baseline detection algorithms, including seven baseline detectors on the HRMS-SCD dataset, to thoroughly analyze their performance, advantages, challenges, and limitations. This comprehensive evaluation not only assesses the effectiveness of these algorithms on the proposed dataset but also uncovers potential directions for future research. The key contributions of this work are as follows:

| Dataset | Resolution | Image count | Image size (Pixels) | Classes | Data Source | Regional Distribution |
|---|---|---|---|---|---|---|
| HRSCD (Daudt et al., 2019) | 0.5m | 291 | 10000×10000 | 5 | Aerial dataset | Rennes and Caen. |
| SECOND (Yang et al., 2020) | 0.5 m | 4662 | 512×512 | 6 | Aerial images | Hangzhou, Chengdu, and Shanghai. |
| Landsat-SCD (Yuan et al., 2022) | 30 m | 8468 | 416×416 | 4 | Mosaic images | Tumushuke |
| Hi-UCD (mini) (Tian et al., 2020) | 0.1 m | 1293 | 1024×1024 | 9 | Aerial images | Tallinn |
| Hi-UCD (Tian et al., 2022) | 0.1 m | 40800 | 512×12 | 9 | Aerial images | Tallinn |
| xDB (Gupta et al., 2019) | <0.8 m | 11034 | 1024×1024 | 4 | Globe Open Data | Global |
| HRMS-SCD （Ours） | 1 m | 11587 | 512×512/ 2048×2048 | 7 | Mosaic images | Beijing |

Table 1. Comparison of different SCD datasets

(1) **High-Resolution Satellite Imagery**: The dataset is derived from high-resolution satellite remote sensing images, providing clearer and more detailed surface information compared to aerial images or low- and medium-resolution satellite images. This enables models to learn finer surface features and effectively supports the improvement of deep learning models' detection performance.
(2) **Comprehensive Semantic Labels**: The dataset includes semantic labels of seven common features—such as planted land, forest and grass cover, and buildings—offering a unique advantage in change type diversity. The full-coverage semantic labels effectively address the category imbalance problem, enhance the model's generalization ability during training, and improve the classification accuracy in change detection.
(3) **Multi-Scene Coverage**: Our dataset encompasses different types of geographic environments, including mountains, forests, croplands, grasslands, and rivers. This compensates for the lack of complex scenario coverage in existing datasets, helping models cope with more intricate environmental changes and improving adaptability and robustness in multiple scenarios.
(4) **Flexible Image Sizes**: To meet different computational resources and task requirements, the dataset provides two image sizes: 512 × 512 pixels and 2048 × 2048 pixels. The larger images better retain global context information, while the smaller images have advantages in memory consumption and computational efficiency, making them suitable for routine training tasks and quick experiments.

## 2. Related Work

### 2.1. Semantic Change Detection Dataset

In recent years, several datasets have significantly advanced research in semantic change detection within the field of remote sensing. Each dataset offers unique features; however, they also present certain limitations that impact their applicability across various scenarios. Below is an overview of some key semantic change detection datasets:

HRSCD (Daudt et al., 2019) is one of the earliest semantic change detection datasets, composed of high-resolution aerial imagery with annotated surface types. It is particularly useful for monitoring urban and agricultural land use changes. Nevertheless, the labels are derived from Urban Atlas vector maps, which are not always precisely aligned with the imagery, potentially leading to classification errors.

SECOND (Yang et al., 2020) provides 4,662 pairs of annotated aerial images covering six surface categories in cities such as Hangzhou, Chengdu, and Shanghai. While it is well-annotated, the limited number of categories may not meet the needs of more complex change detection tasks. Additionally, the reliance on aerial imagery can affect data consistency due to acquisition constraints.

Landsat-SCD (Yuan et al., 2022) is a multi-temporal dataset built on Landsat satellite imagery, offering long-term land cover change data with high temporal resolution. However, its relatively low spatial resolution of 30 meters limits its applicability for detecting fine-scale changes.

Hi-UCD (mini) (Tian et al., 2020) is an ultra-high-resolution urban change detection dataset with a spatial resolution of 0.1 meters, providing detailed semantic labels for urban features. However, the increased resolution introduces challenges such as shadows and occlusions, which can affect change detection accuracy.

Hi-UCD (Tian et al., 2022) expands upon the Hi-UCD (mini) dataset, covering 102 square kilometers in Tallinn with 40,800 image pairs, supporting multi-temporal semantic segmentation and change detection. Despite its comprehensive coverage, certain fine-grained changes—such as farmland transitioning to grassland—may not be effectively captured, affecting its utility in specific scenarios.

xBD (Gupta et al., 2019) is a large-scale dataset focused on post-disaster building damage assessment, providing pre- and post-disaster imagery with annotations for building damage levels. While valuable for disaster-related change detection, its narrow focus limits its application to other types of change detection tasks.

While these existing datasets each contribute significantly to the field, they also exhibit limitations: aerial imagery is costly, offers limited coverage, and is susceptible to weather conditions; low- and medium-resolution satellite images struggle to capture subtle changes; and datasets limited to single scenes with fewer label classifications constrain the model's generalization ability. To address these issues, the new dataset proposed in this paper utilizes high-resolution satellite images rich in detailed information, covers multiple scenes, and provides a more comprehensive label categorization. This facilitates in-depth study of multiple change types and promotes the advancement of semantic change detection research.

### 2.2. Semantic Change Detection Algorithm

In recent years, the application of deep learning to remote sensing image change detection has made remarkable progress. Early models predominantly relied on Convolutional Neural Networks (CNNs), which excelled in computer vision tasks and were thus widely adopted in semantic change detection. For example, the High-Resolution Semantic Change Detection (HRSCD) algorithm is primarily based on deep learning and CNNs(Ding et al., 2022). This algorithm approaches the problem of semantic change detection as two interrelated tasks: binary change detection and classification of change types.

To achieve this, HRSCD proposed four distinct strategies(Cheng et al., 2024). First, it involves directly comparing land cover maps by training a land cover mapping network and comparing the predicted labels; however, the accuracy of this approach depends on the quality of the land cover maps. Second, it treats changes as independent labels and performs direct semantic change detection. Although intuitive, this method results in the number of categories growing quadratically with the number of land cover categories, leading to class imbalance issues. Third, it trains two separate networks: one for binary change detection and another for land cover mapping, which simplifies category prediction and optimizes performance. Fourth, it integrates these two networks into a multi-task network that utilizes land cover information for change detection. By inputting two co-registered images and outputting three maps, this approach enables information sharing and improves detection accuracy.

Among these methods, multi-task semantic change detection networks demonstrate greater potential and have become a focal point of research. Building on this foundation, the Asymmetric Siamese Network (ASN) (Yang et al., 2020) introduced a novel architecture to capture asymmetric changes between different time periods. It employs an asynchronous spatial pyramid to reduce computation while focusing on varying spatial regions and integrates features through deep connections to enhance change detection. To address some limitations of prior methods, SSCD-l extends previous work by using two separate CNN encoders to extract temporal semantic features and merging them through a deep change detection unit, optimizing for changes over time. Expanding on the concept of temporal correlation, Bi-SRNet adds semantic reasoning modules to capture temporal correlations and introduces semantic consistency loss to enhance accuracy in unchanged regions (Ding et al., 2022). The latest CdSC network (Wang et al., 2024a) enhances the modeling of complex changes by exploring the interaction of features across time.

Recently, Vision Transformers (ViTs), known for modeling long-range dependencies with self-attention, have gained popularity in computer vision and have demonstrated superior performance in remote sensing image change detection (Dubey and Singh, 2024). ChangeMask (Zheng et al., 2022) decouples Semantic Change Detection (SCD) into temporal segmentation and binary change detection tasks, leveraging semantic causality and temporal symmetry to improve detection efficiency and accuracy. Extending this approach, MTSCD-Net adopts a multi-task learning framework with Swin Transformer-based multi-scale feature extraction and a feature aggregation module to integrate low- and high-level features (Cui and Jiang, 2023), effectively balancing task correlation and model performance. Moreover, SCanNet (Ding et al., 2024) considers spatio-temporal dependencies to improve the accuracy of SCD and significantly outperforms baseline methods in detecting critical semantic changes and maintaining semantic consistency in the obtained bitemporal results. CTST (Wang et al., 2024b) designs a unique feature-integrated encoding model combining CNN and Transformer architectures, which enhances the model's understanding of global dependencies and improves the extraction of local features, outperforming mainstream and state-of-the-art methods on three datasets

Through these innovative methods, semantic change detection is advancing toward higher accuracy and robustness. However, the success of deep learning algorithms relies not only on model design but also on the availability of large-scale, high-quality annotated datasets. Such datasets enhance model adaptability to diverse and complex scenarios and improve the ability to capture subtle changes. Therefore, constructing diverse, high-quality datasets is essential for advancing deep learning in change detection. High-quality datasets are not only catalysts for technological breakthroughs but also the foundation for achieving precise and reliable change detection.

## 3. Design of the Dataset

Originating from land cover surveys in China, the dataset has been collected and quality-checked multiple times by experienced remote sensing teams, ensuring high data accuracy and reliability. This dataset is designed for both research purposes and practical land use monitoring and management.

### 3.1. Basic Image Information

The dataset consists of 11,587 images with a size of 512 × 512 pixels and a resolution of 1 meter, which were acquired via the Resource 1, Resource 2, and Beijing 2 satellites, and cover the whole land category changes in Beijing during 2017-2018. In addition, the dataset includes large-size images of 2048 × 2048 pixels, which helps to improve the accuracy and finesse of change detection.In order to visualize the image changes more, we provide the change legend in Figure 1.

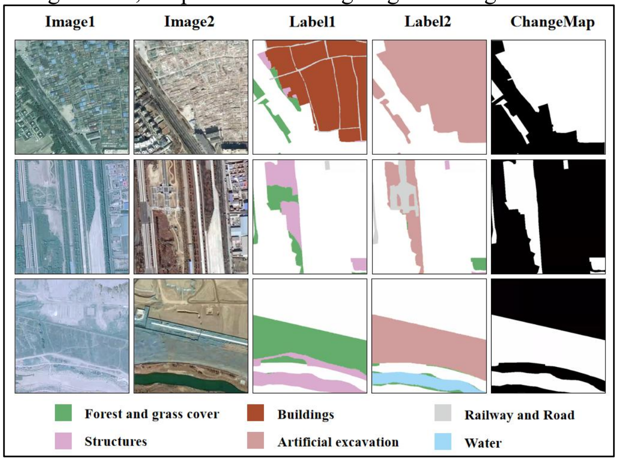



Figure 1. Anterior and posterior time-image changes legend.

### 3.2. Full Coverage Labeling

In this paper, reference is made to the Content and Indicators of China Geographic National Information Census and other related documents to ensure that the labeling system of the dataset is scientific and reasonable. In order to simplify the classification system and meet the practical application requirements, this paper combines cropland and plantation into "Planting land", reflecting the land use type of agricultural and horticultural production. Meanwhile, forest land and grassland, which have similar ecological functions, are combined into "forest and grass cover" to simplify the categorization while maintaining an accurate reflection of the natural environment. In the end, the feature classification system consists of seven categories: planting land, forest and grass cover, buildings, railway and roads, structures, artificial excavation, and waters. The classification provides comprehensive coverage of both urban and natural surfaces to support change detection and land use analysis, and detailed definitions and illustrations of each category are presented in Table 2.

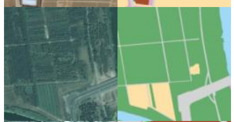
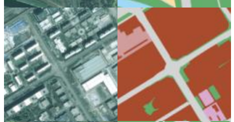
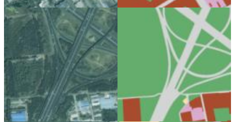
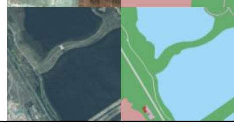
| Types | Examples | color | Image Examples |
|---|---|---|---|
| Planting land | Paddy fields, dry land, orchards, tea gardens , etc. | Light Yellow |  |
| Forest and grass cover | Grassland, tree forest, shrub forest, bamboo forest, green forest land, etc. | Grass Green |  |
| Buildings | Multi story building area, low rise building area, abandoned building area, etc. | Brick Red |  |
| Railway and Road | Track road surface, trackless road surface, etc. | Light Gray |  |
| Structures | Hardened surface, hydraulic facilities, city walls, greenhouses, greenhouses, solidification pools, etc. | Lavender |  |
| Artificial excavation | Open pit mining sites, stacking materials, construction sites, etc. | Light Pink |  |
| Water | Water surface, water channels, glaciers, and perennial snow accumulation, etc. | Sky Blue |  |

Table 2. Labeling Categories and Legends

### 3.3. Multi-scene Variations

The dataset covers various typical landscape scenarios, including urban areas, rural areas, mountains, forests, croplands, grasslands, rivers, and lakes, as shown in Figure 2. Including urban and rural areas enhances the model's ability to capture human activity, while mountains and forests improve its recognition of changes in natural ecosystems. Farmland and grassland scenarios relate to agricultural production, helping monitor crop growth and land use changes. Changes in rivers and lakes provide key information for waterbody monitoring, enhancing the model's robustness and broad applicability in change detection. This diverse combination of scenarios improves the model's performance and accuracy across various geographical environments, both natural and man-made.

Figure 2. Different scene examples in the dataset.

### 3.4. Double-sized Image

For large or longer objects, their geometric features and semantic correlations cannot be well rendered in a local window, so contextual information should be modeled over a larger image scale(Ding et al., 2021). GLNet (Chen et al., 2019) significantly improves segmentation by processing the whole image and local pixel blocks at the same time. MFVNet(Li et al., 2023b) is proposed, with pyramid sampling and scale alignment, solves the multi-scale information fusion problem and achieves leading performance on multiple datasets.

Therefore, this paper also proposes a large-size image of 2048 × 2048 pixels to preserve both global and local information to achieve more accurate semantic segmentation results when dealing with complex geographic scenes.

### 3.5. Statistical Characterization

We analyzed Table 3 and Figure 3 for changes in surface cover in the before and after time phases, and the results show that planted land decreased from 1.85% to 0.98%, buildings decreased from 2.98% to 0.40%, and manual excavation increased from 1.01% to 4.79%. The decrease in cultivated land and the decline in buildings are associated with urban development and the renovation of old urban areas, while the increase in manual excavation indicates the intensification of regional infrastructure development. These changes have not only affected land-use patterns, but have also had a profound impact on the ecosystem and regional economic development.

| Label | T1 | T2 |
|---|---|---|
| Planting land | 1.85% | 0.98% |
| Forest and grass cover | 3.00% | 3.26% |
| Buildings | 2.98% | 0.40% |
| Railway/Road | 0.24% | 0.27% |
| Structures | 2.11% | 1.38% |
| Artificial excavation | 1.01% | 4.79% |
| Water | 0.11% | 0.21% |

Table 3. Proportion of different land cover types.

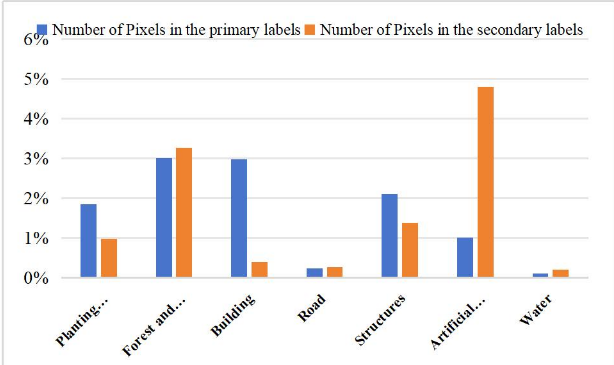


Figure 3.The number of pixels of different types by region.

## 4. Evaluation Results

### 4.1. Baseline Algorithms

To evaluate the performance of the dataset, we selected eight algorithms that perform well in semantic change detection tasks for comparison. These include HRSCD-str4, which takes into account temporal correlation through differential hopping connections; SSESN(Zhao et al., 2022), which utilizes spatial and semantic feature aggregation modules to improve accuracy; SSCD-l, which extracts bitemporal image features; Bi-SRNet, which introduces semantic inference blocks and consistency loss; MTSCD-Net, which combines multiscale features with spatial enhancement module; SCanNet(Ding et al., 2024), modeling semantic changes using (CSWin)Transformer; CdSC, exploring feature interactions in conjunction with 3D convolution; and DEFO-MTLSCD(Li et al., 2024), boosting performance with dual- and triple-branch decoder architectures.

### 4.2. Experiment Details

The experiments were run on a desktop workstation equipped with an NVIDIA GeForce RTX 3090 GPU with 24G of memory, and all programs were implemented based on the PyTorch platform. The input image size was 512 × 512 pixels, and the training set was normalized and data-enhanced with random Gaussian noise, random flipping and rotation. The batch size is 4 and the model is trained for 50 epochs. HRSCD-str4, SSESN, SSCD-l, Bi-SRNet, SCanNet, and DEFO-MTLSCD use the SGD optimizer with an initial learning rate of 0.1; MTSCD-Net is learned at a rate of 0.00015 using AdamW(Loshchilov and Hutter, 2017)optimization and warm-up strategy; CdSC also uses AdamW with a learning rate of 1e-4 and linear decay.

### 4.3. Evaluation Metrics

In this paper, we use four widely adopted evaluation metrics to assess the accuracy of the semantic change detection task, including Overall Accuracy (OA), Mean Intersection and Union Ratio (mIoU), Separate Kappa Coefficients SeK(Yang et al., 2020), and F_scd scores(Ding et al., 2022). The confusion matrix $Q = \{q_{i,j}\}$ is computed from the prediction results and the truth labels, where $q_{i,j}$ denotes the number of pixels that are categorized as class $i$ while the truth category is $j (i, j \in \{0, 1, \dots, N\})$ (0 means no change) in the number of pixels. OA denotes the percentage of pixels with correct category prediction to the total pixels of the image and is calculated as shown below:

$$OA = \sum_{i=0}^{N} q_{ii} / \sum_{i=0}^{N}\sum_{j=0}^{N} q_{ij} \tag{1}$$

In the SCD task the unchanged class occupies the majority, there is a class imbalance problem, and the recognition accuracy of semantic categories cannot be accurately assessed by OA alone, so mIoU and $SeK$ are introduced to assess the performance of the two subtasks of CD and SS, respectively. Where mIoU is the average of the invariant region $\mathrm{IoU}_1$ and the changing region $\mathrm{IoU}_2$, and the calculation formula is shown below:

$$\mathrm{IoU}_1 = q_{00} / \left( \sum_{i=0}^{N} q_{i0} + \sum_{i=0}^{N} q_{0j} - q_{00} \right) \tag{2}$$

$$\mathrm{IoU}_2 = \sum_{i=1}^{N}\sum_{i=1}^{N} q_{ij} / \left( \sum_{i=0}^{N}\sum_{j=0}^{N} q_{ij} - q_{00} \right) \tag{3}$$

$$\mathrm{mIoU} = (\mathrm{IoU}_1 + \mathrm{IoU}_2)/2 \tag{4}$$

Pixels with zero category truth value and zero prediction will be ignored in the $SeK$ calculation, and only the categories in the changing region will be considered to evaluate the classification performance, thus mitigating the effect of label imbalance.The $SeK$ index is defined as follows:

$$\rho = \sum_{i=1}^{N} q_{ii} / \left( \sum_{i=0}^{N}\sum_{j=1}^{N} q_{ij} - q_{11} \right) \tag{5}$$

$$\eta = \left( \sum_{i=1}^{N} \left( \sum_{j=0}^{N} q_{ij} * \sum_{j=0}^{N} q_{ji} \right) + \sum_{j=1}^{N} q_{0j} * \sum_{j=1}^{N} q_{j0} \right) / \left( \sum_{i=0}^{N}\sum_{j=0}^{N} q_{ij} - q_{00} \right)^2 \tag{6}$$

$$SeK = e^{IoU_1 - 1} * (\rho - \eta)/(1 - \eta) \tag{7}$$

In addition, the $F_{scd}$ metric was used to focus on assessing the precision of land cover classification within the change area, which is based on the same principle as the F1 score, and was calculated by the precision $R_{scd}$ and the recall $R_{scd}$ labeled as change area:

$$P_{scd} = \sum_{i=1}^{N} q_{ii} / \sum_{i=1}^{N}\sum_{j=0}^{N} q_{ij} \tag{8}$$

$$R_{scd} = \sum_{i=1}^{N} q_{ii} / \sum_{i=0}^{N}\sum_{j=1}^{N} q_{ij} \tag{9}$$

$$F_{scd} = \frac{2 * P_{scd} * R_{scd}}{P_{scd} + R_{scd}} \tag{10}$$

With the help of the above four evaluation metrics, this paper is able to provide a comprehensive and detailed evaluation of the performance of the semantic change detection task, thus providing insight into the performance of the proposed dataset in various aspects.

## 4.4. Results and Analysis

### 4.4.1. Analysis of quantitative results

Table 4 demonstrates that the CdSC and MTSCD methods outperform others, achieving Fscd metrics of 44.22% and 46.85%, respectively—over 8% higher than the competing methods. The superiority of CdSC stems from its innovative use of three-dimensional convolution, which explores interactions between inter-temporal features and intrinsic depth differences. This approach enables more effective modeling of complex topography and varied changes in bi-temporal remote sensing imagery, resulting in the highest scores in the OA metrics. MTSCD, on the other hand, combines the spatial attention weight map of the change detection task with the location priori information from the semantic segmentation task. By fully exploiting the correlation between these two subtasks, it effectively addresses the class imbalance problem in the dataset, demonstrating excellent performance across all three metrics: mIoU, SeK, and Fscd. Conversely, the traditional deep learning method HRSCD4 performs poorly due to its underutilization of semantic information and limited connections between subtasks, making it ill-equipped to handle diverse semantic changes.

Overall, the detection accuracies of existing algorithms on the new dataset range from 18% to 46%, indicating significant room for optimization. Future research on feature extraction and inter-subtask linkage needs to address the challenges posed by large-scale scenes and diverse types of semantic changes.

| Methods | OA(%) | mIoU(%) | SeK(%) | $F_{scd}$(%) |
|---|---|---|---|---|
| HRSCD4 | 80.14 | 51.98 | -4.93 | 18.46 |
| SSESN | 82.85 | 55.34 | 4.46 | 35.48 |
| SSCD-l | 81.19 | 55.61 | 2.29 | 35.96 |
| Bi-SRNet | 81.87 | 56.37 | 2.78 | 36.13 |
| SCanNet | 80.21 | 54.62 | 0.70 | 33.79 |
| DEFO-MTLSCD | 81.98 | 56.64 | 2.14 | 34.73 |
| CdSC | 84.29 | 59.80 | 8.75 | 44.22 |
| MTSCD | 83.65 | 60.57 | 9.43 | 46.85 |

Table 4. Results of change detection in the dataset

To provide a detailed comparison, we evaluated the F1 scores of each SCD model across different semantic change categories within the change region. As shown in Table 5, change detection for forest and grass cover, house structures, and man-made heaps and diggings performed stably, achieving more than 60% accuracy—likely due to the large number of samples. In contrast, detection of planted land and structures varied more, leading to misclassifications caused by similar geometry and texture. Railroad roads were poorly detected, possibly because of the low number of tags. Despite the small number of samples, water bodies maintained high detection accuracy for SSESN, CdSC, and MTSCD due to their unique characteristics.

| Methods | F1-Score(%) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Planting land | Forest and grass cover | Building | Railway and Road | Structures | Artificial excavation | Water | Change |
| HRSCD4 (Caye Daudt et al., 2019) | 25.48 | 45.59 | 63.32 | 0.00 | 22.14 | 49.48 | 0.00 | 40.32 |
| SSESN (Zhao et al., 2022) | 52.70 | 64.77 | 90.73 | 33.60 | 54.65 | 87.68 | 75.82 | 45.12 |
| SSCD-l (Ding et al., 2022) | 46.66 | 64.52 | 86.90 | 15.20 | 51.75 | 84.01 | 29.28 | 48.31 |
| Bi-SRNet (Ding et al., 2022) | 53.03 | 62.32 | 87.99 | 21.00 | 46.45 | 83.26 | 17.95 | 49.17 |
| SCanNet (Ding et al., 2024) | 37.97 | 60.16 | 86.83 | 21.73 | 39.83 | 82.98 | 23.65 | 47.26 |
| DEFO-MTLSCD (Li et al., 2024) | 38.70 | 58.14 | 86.70 | 12.78 | 48.04 | 80.82 | 19.66 | 49.71 |
| CdSC (Wang et al., 2024) | 77.10 | 74.63 | 90.15 | 34.50 | 66.41 | 89.09 | 81.91 | 53.91 |
| MTSCD (Cui and Jiang, 2023) | 77.43 | 78.11 | 91.23 | 27.18 | 69.58 | 89.89 | 78.96 | 56.56 |

Table 5. Performance of the SCD model for each change type on the dataset.

Combining the results from Table 4 and Table 5, we observe that this dataset exhibits rich and complex change patterns, underscoring the potential of semantic change detection algorithms in complex scene understanding. Although some algorithms perform well in specific categories, existing techniques still have limitations in multi-scale feature extraction, global modeling, and dual-task correlation, and do not fully utilize the category feature information of the dataset. Future research should focus on developing finer local-global modeling architectures and tighter dual-task complementary strategies, as well as enhancing category distinguishability through full-coverage semantic labeling to address class imbalance.

#### 4.4.2. Analysis of qualitative results

To qualitatively assess the experimental results, we analyzed randomly selected samples presented in Figure 4. In examining the outcomes of groups a and d, we found that existing algorithms more accurately recognized changes between manually excavated land and house structures. This heightened accuracy is likely due to the distinct characteristics and significant morphological differences of these objects. However, the algorithms' accuracy decreased markedly when detecting changes between forest and grass cover or between planted land and structures. In groups c and d, some algorithms struggled to recognize these changes or misclassified the categories, mainly because of the ambiguous feature differences and visual similarities among these types in the images. The algorithms exhibited significant limitations in handling subtle variations between complex surface types, underscoring the challenge of processing complex and diverse scenes. Future work should focus on developing more robust models and refining algorithm designs to enhance the accuracy and robustness of change detection.

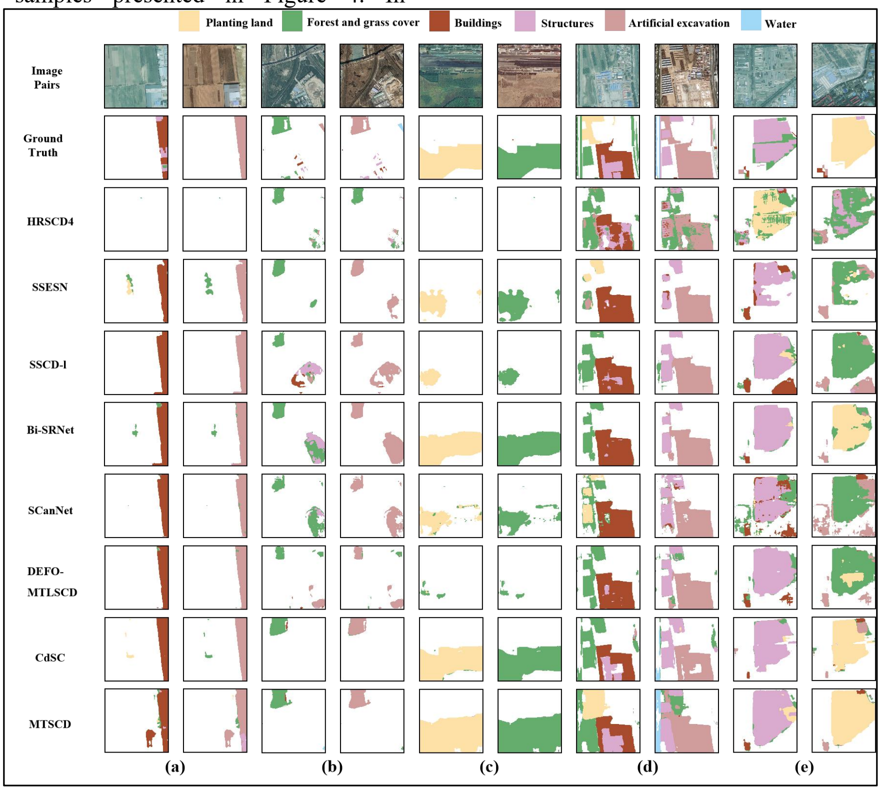



Figure 4. Comparison of label prediction results using different SCD algorithms

## 5. Conclusion

In this study, we introduce the HRMS-SCD dataset—a high-resolution, multi-scene satellite imagery dataset designed for comprehensive land-cover SCD to advance remote sensing research. HRMS-SCD comprises 11,587 pairs of images with a 1-meter spatial resolution, capturing land-cover changes in Beijing between 2017 and 2018. The dataset stands out due to its coverage of diverse land surface types and the inclusion of detailed semantic annotations, making it suitable for identifying and classifying changes across various scenes. Additionally, it offers two image sizes—512×512 pixels and larger 2048×2048 pixels—allowing researchers to study both fine details and broader contexts. Furthermore, the dataset facilitates benchmarking of SCD algorithms, providing a valuable resource for advancing deep learning models in the analysis of complex and dynamic land-cover changes. Experimental results demonstrate that while current methods perform well for certain change types, they face limitations when detecting subtle changes and handling diverse scenarios. This underscores the potential for further research into more refined local-global modeling architectures, improved dual-task integration, and enhanced semantic labeling to address class imbalance and improve robustness in detecting subtle changes.

## Acknowledgements

This work was supported by the National Key Research and Development Program of China [grant number 2023YFB3907602], the National Natural Science Foundation of China [grant number 42401500, 42201440], the Science and Technology Association's Young Talent Nurturing Program [grant number YESS20220173].

## References

Cao, Y., Huang, X., Weng, Q., 2023. A multi-scale weakly supervised learning method with adaptive online noise correction for high-resolution change detection of built-up areas. *Remote Sensing of Environment* 297, 113779.

Chen, H., Qi, Z., Shi, Z., 2021. Remote sensing image change detection with transformers. *IEEE Transactions on Geoscience and Remote Sensing* 60, 1-14.

Chen, W., Jiang, Z., Wang, Z., Cui, K., Qian, X., 2019. Collaborative Global-Local Networks for Memory-Efficient Segmentation of Ultra-High Resolution Images, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8924-8933.

Cheng, G., Huang, Y., Li, X., Lyu, S., Xu, Z., Zhao, H., Zhao, Q., Xiang, S., 2024. Change detection methods for remote sensing in the last decade: A comprehensive review. *Remote Sensing* 16, 2355.

Cui, F., Jiang, J., 2023. MTSCD-Net: A network based on multi-task learning for semantic change detection of bitemporal remote sensing images. *International Journal of Applied Earth Observation and Geoinformation* 118, 103294.

Daudt, R.C., Saux, B.L., Boulch, A., Gousseau, Y., 2019. Multitask learning for large-scale semantic change detection. *Computer Vision and Image Understanding* 187, 102783.

Ding, L., Guo, H., Liu, S., Mou, L., Zhang, J., Bruzzone, L., 2022. Bi-Temporal Semantic Reasoning for the Semantic Change Detection in HR Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing* 60, 1-14.

Ding, L., Lin, D., Lin, S., Zhang, J., Cui, X., Wang, Y., Tang, H., Bruzzone, L., 2021. Looking Outside the Window: Wide-Context Transformer for the Semantic Segmentation of High-Resolution Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing* 60, 1-13.

Ding, L., Zhang, J., Guo, H., Zhang, K., Liu, B., Bruzzone, L., 2024. Joint Spatio-Temporal Modeling for Semantic Change Detection in Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing* 62, 1-14.

Feng, Y., Jiang, J., Xu, H., Zheng, J., 2023. Change detection on remote sensing images using dual-branch multilevel intertemporal network. *IEEE Transactions on Geoscience and Remote Sensing* 61, 1-15.

Guo, Q., Zhang, J., Zhu, S., Zhong, C., Zhang, Y., 2021. Deep multiscale Siamese network with parallel convolutional structure and self-attention for change detection. *IEEE Transactions on Geoscience and Remote Sensing* 60, 1-12.

Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., Heim, E., Choset, H., Gaston, M., 2019. xBD: A Dataset for Assessing Building Damage from Satellite Imagery. *arXiv preprint*, 1-9.

Huang, Y., Li, X., Du, Z., Shen, H., 2024. Spatiotemporal Enhancement and Interlevel Fusion Network for Remote Sensing Images Change Detection. *IEEE Transactions on Geoscience and Remote Sensing* 62, 1-14.

Huang, Y., Zhang, P., 2024. CSDACD: Domain-adaptive Change Detection Network for Cross-seasonal Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters* 21, 1-5.

Jiang, M., Chen, Y., Dong, Z., Liu, X., Zhang, X., Zhang, H., 2024. Multiscale Fusion CNN-Transformer Network for High-Resolution Remote Sensing Image Change Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17, 5280-5293.

Li, W., Xue, L., Wang, X., Li, G., 2023a. ConvTransNet: A CNN–transformer network for change detection with multiscale global–local representations. *IEEE Transactions on Geoscience and Remote Sensing* 61, 1-15.

Li, Y., Chen, W., Huang, X., Gao, Z., Li, S., He, T., Zhang, Y., 2023b. MFVNet:a deep adaptive fusion network with multiple field-of-views for remote sensing image semantic segmentation. *Science China Information Sciences* 66, 140305.

Li, Z., Wang, X., Fang, S., Zhao, J., Yang, S., Li, W., 2024. A Decoder-Focused Multitask Network for Semantic Change Detection. *IEEE Transactions on Geoscience and Remote Sensing* 62, 1-15.

Lin, H., Hang, R., Wang, S., Liu, Q., 2024. DiFormer: A Difference Transformer Network for Remote Sensing Change Detection. *IEEE Geoscience and Remote Sensing Letters* 21, 1-5.

Loshchilov, I., Hutter, F., 2017. Fixing Weight Decay Regularization in Adam. *arXiv preprint arXiv:1711.05101* 5, 1-5.

Niu, Y., Guo, H., Lu, J., Ding, L., Yu, D., 2023. SMNet: Symmetric Multi-Task Network for Semantic Change Detection in Remote Sensing Images Based on CNN and Transformer. *Remote Sensing* 15, 949.

Tian, S., Zheng, Z., Ma, A., Zhong, Y.J.A., 2020. Hi-UCD: A Large-scale Dataset for Urban Semantic Change Detection in Remote Sensing Imagery. *arXiv preprint* abs/2011.03247, 1-6.

Tian, S., Zhong, Y., Zheng, Z., Ma, A., Tan, X., Zhang, L., 2022. Large-scale deep learning based binary and semantic change detection in ultra high resolution remote sensing imagery: From benchmark datasets to urban application. *ISPRS Journal of Photogrammetry and Remote Sensing* 193, 164-186.

Tollerud, H.J., Zhu, Z., Smith, K., Wellington, D.F., Hussain, R.A., Viola, D., 2023. Toward consistent change detection across irregular remote sensing time series observations. *Remote Sensing of Environment* 285, 113372.

Wang, Q., Jing, W., Chi, K., Yuan, Y., 2024a. Cross-Difference Semantic Consistency Network for Semantic Change Detection. *IEEE Transactions on Geoscience and Remote Sensing* 62, 1-12.

Wang, S., Wu, W., Zheng, Z., Li, J., 2024b. CTST: CNN and Transformer-Based Spatio-Temporally Synchronized Network for Remote Sensing Change Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17, 16272-16288.

Wu, C., Du, B., Zhang, L., 2023. Fully convolutional change detection framework with generative adversarial network for unsupervised, weakly supervised and regional supervised change detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 9774-9788.

Yang, K., Xia, G.S., Liu, Z., Du, B., Pelillo, M., 2020. Asymmetric Siamese Networks for Semantic Change Detection. *IEEE Transactions on Geoscience and Remote Sensing* 60, 1-18.

Yuan, P., Zhao, Q., Zhao, X., Wang, X., Long, X., Zheng, Y., 2022. A transformer-based Siamese network and an open optical dataset for semantic change detection of remote sensing images. *International Journal of Digital Earth* 15, 1506-1525.

Zhang, C., Feng, Y., Hu, L., Tapete, D., Pan, L., Liang, Z., Cigna, F., Yue, P., 2022. A domain adaptation neural network for change detection with heterogeneous optical and SAR remote sensing images. *International Journal of Applied Earth Observation and Geoinformation* 109, 102769.

Zhao, M., Zhao, Z., Gong, S., Liu, Y., Yang, J., Xiong, X., Li, S., 2022. Spatially and Semantically Enhanced Siamese Network for Semantic Change Detection in High-Resolution Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15, 2563-2573.

Zheng, Z., Zhong, Y., Tian, S., Ma, A., Zhang, L., 2022. ChangeMask: Deep multi-task encoder-transformer-decoder architecture for semantic change detection. *ISPRS Journal of Photogrammetry and Remote Sensing* 183, 228-239.