Semantic Change Detection in Multi-Temporal Remote Sensing Images Using Deep Learning

Jing Zhang

October 26, 2022

Semantic Change Detection in Multi-temporal Remote Sensing Images Using Deep Neural Networks

Qualifying Exam

Jing Zhang
Remote Sensing Laboratory
University of Trento
jing.zhang-1@unitn.it

ABSTRACT

Semantic change detection (SCD) is derived from change detection (CD) and is highly valuable in remote sensing. The traditional change detection approach mainly focuses on identifying where changes have occurred in multitemporal remote sensing images. Convolutional Neural Network (CNN) for SCD can be three-branched, with one branch recording change information and two branches recording information at two different time. CNN-based semantic change detection supports finer-grained and three-dimensional change analysis and provides rich semantic information about the details on the Earth's surface before and after the transition. However, many problems occur in existing studies, including the convergence problem while training with limited change samples, the low accuracy of classifying the semantic classes, and the inconsistency between multi-temporal results. To address these issues, our research aims to advance the development of semantic change detection in remote-sensing images by using deep learning methods. To this end, the research we plan to conduct includes the following: 1) Developing semi-supervised SCD methods to improve the training under limited training samples to alleviate the 'data-hungry' problem; 2) Analyzing the coherence between semantic and change information. We will model the inherent mechanism in SCD tasks to reduce false detection and omission, thus improving the training stability; 3) Developing SCD methods for image time-series. We will analyze the temporal features to distinguish between seasonal change and semantic changes, thus modeling the change trend in the observed regions. The developed methodologies will be compared with state-of-the-art network models to test their performance.

KEYWORDS

Semantic Change Detection, Remote Sensing, Deep Learning

1 INTRODUCTION

Change detection (CD) in remote sensing data is a process to identify surface changes from the joint analysis of two (or more) images acquired in the same area and at different times [1] [2]. With the emergence and application of remote sensing technology, this is one of the earliest and most widely considered research fields [3]. Moreover, it plays an essential part in multi-industry and multi-discipline operations such as land and resource surveys, disaster monitoring and assessment, environmental agricultural forestry monitoring, and urban resource management.

Change detection compares the contents of two images in the same area and detects the difference between them [4]. While CD algorithms can monitor and analyze regions of interest in remote sensing images (RSI), they can only inform 'where' changes have occurred without specifying 'what' the detailed change types are. To overcome this limitation, the task of semantic change detection was proposed in recent literature, which makes the semantic information representation of land-cover/land-use (LCLU) richer and more complex. it not only provides change information but also provides detailed LCLU maps before and after the change [5]. Semantic change detection (SCD) extends binary change detection (BCD). SCD is beneficial in various remote sensing applications, such as urban management, environment monitoring, crop monitoring, and damage assessment.

Semantic Change Detection (SCD) [6] provides detailed "from-to" change information. In early papers, it was often called multi-class change detection or post-classes change detection [7]. Compared to binary change detection, semantic change detection requires semantic classes before and after the observation interval to be noted in the results, which can be expressed as a) a binary change map and two semantic classes of pre-temporal and post-temporal, or b) two semantic change maps that contain only information on change region classes [8]. For example, Fig.1 (a)(b) shows the differences between binary change detection and semantic change detection. Semantic change detection provides richer change information and therefore can address deeper research and application needs [9].

Deep neural networks include CNNs and RNNs. In general, CNNs are applied to analyze images, RNNs are used to analyze language association. However, in recent year, CNNs and RNNs have been widely used for change detection in remote sensing images (RSI). Current CNN-based change detection methods such as FCN [10], U-Net [11], and DeepLab [12], provide flexible techniques to deal with CD. However, SCD based on the integration between CNN and RNN has been rarely studied. SCD is more complex than BCD, and it includes two assignments: semantic segmentation and change detection. Semantic segmentation refers to the extraction of the bi-temporal semantic information in changed areas, where CD refers to identifying the changes. To represent the semantic information before and after the change, SCD generates one change map [13] and two semantic change maps [14]. Some CNN-based methods has been proposed for SCD in recently published articles [15][16]. However, the learning of change information is difficult, especially in form of representation of the transition of ground surface classes. Thus, the SCD tasks have higher requirements for CNN-based methods [17].

The planned Ph.D. activities aim at addressing the abstraction power

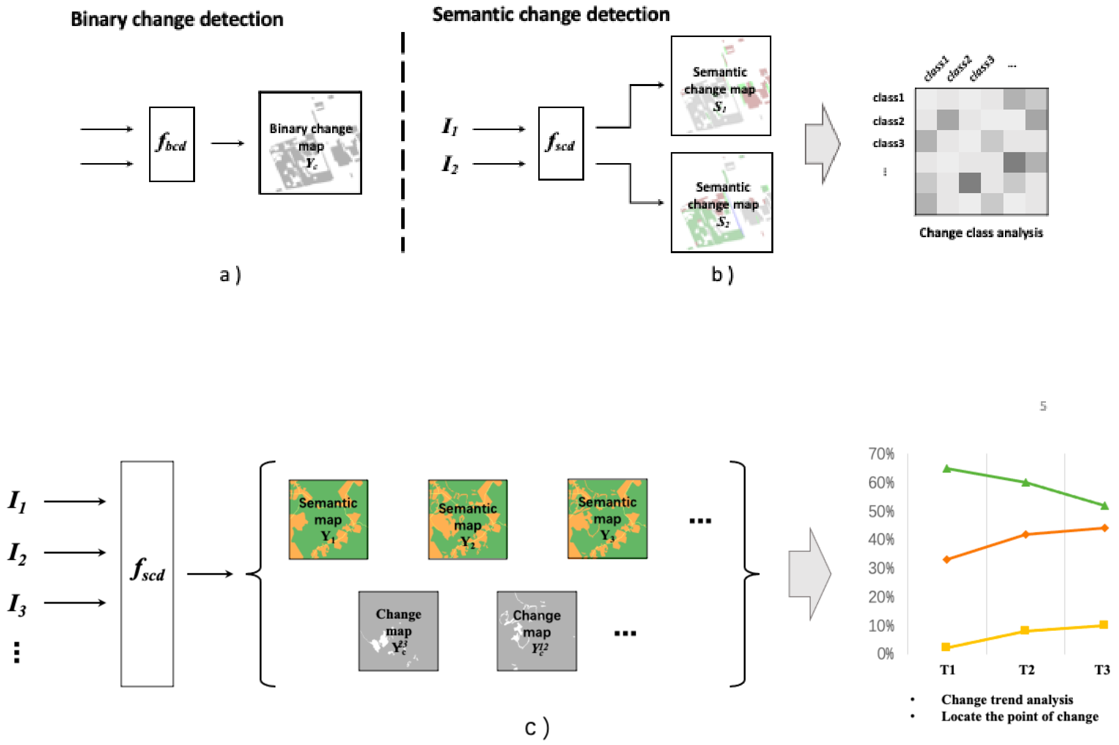


Figure 1: Illustration of different types of CD tasks: a) binary change detection, b) semantic change detection and c) Semantic change detection for time-series images.

of deep learning to boost the accuracy and efficiency of semantic change detection on RSIs. The specific objectives of the research proposal are,

- Analysing the coherence between semantic and change information by modeling their relations to reduce false detection and omission, thus improving the training stability.
- Developing semi-supervised SCD methods to improve the training under limited training samples to alleviate the 'data-hungry' problem.
- Developing SCD methods for the analysis of image time-series to distinguish between seasonal changes and semantic changes, thus modeling the changing trend in the observed regions.

The rest of the proposal organized is as follows. Section 2 is dedicated to introducing state-of-the-art on change detection, semantic change detection for time-series images, and semantic change detection for multi-temporal Image. The problem statement, motivation, goals, and proposed methodologies are described in section 3, 4. section 5 focusing on the preliminary results. Section 6 concludes the proposal.

2 RELATED WORK

2.1 Change Detection

Change detection (CD) is the task that finds areas that have changed in remotely sensed images (RSI) (RSI)[4]. Change detection from multi- temporal satellite imagery detects anthropogenic or natural spatial changes[18]. It has critical applications in environmental monitoring, and identifying changes in land-use and land-cover. Since the emergence of change detection, researchers have developed many methods for change detection,using spectral and textual features. Many PBCD(Pixel-Based Change Detection) algorithms have been proposed, such as methods based on image processing [19] [20], classification of images [7], and machine learning [21] [22]. Numerous deep learning approaches have been implemented for change detection since the emergence of these techniques. Some of them presented consider both the high spatial correlation between pixels in ultra-high spatial resolution (VHR) images and the differences in multi-sensor images [23]. Others combine CNN features with Super Pixel Segmentation [24].

2.2 Semantic Change Detection

Traditional semantic change detection focuses on the post-classification change detection method. This approach first classifies surface classes and produces semantic categories, generating semantic classification maps, and then calculates the interconversion statistics between each category [25]. Since the analysis of the change two time images is done inadequately this method overlooks their interconnection, and the error accumulation problem is more significant. In order to improve this method, Xian et al. [26] calculated the variation probabilities through change analysis and identified the areas with high variation probability [26]. Bruzzone et al. [27][28] optimized the process for change discrimination, using Bayes prior probability estimation to measure the transition probability for semantic categories to improve the classification accuracy on semantic changes. Wu et al. [7] combined both ideas to calculate change probabilities by slow feature analysis and synthesized the change types by joint probability estimation. In general, the post-classification change detection method accuracy depends on the semantic information extraction and change discrimination algorithm, which is influenced by the parameters (thresholds) and the cumulative error problem.

2.3 Semantic change detection in images time-series

In the existing literature, change detection in images time-series is strongly related to land use/cover mapping and other thematic mappings. A common strategy involves using change detection to filter and reduce classification areas, and then classification for semantic information is applied to each temporal phase. Peiman et al. [29] used principal component analysis to exclude areas with low change probability and then used post-change classification for land use/cover mapping and change detection for urban areas in the Pisa region of Italy. Zhu et al. [30] compared the latest Landsat images with historical data to detect changes for continuous updating of land-use maps. Demir et al. [31] used Landsat data to update land use maps by combining change detection and active learning to predict unknown temporal phases from semantic labeling of known temporal phases. Yan et al. [32] used change detection to improve the performance of land use/cover classification. Firstly, the change pixels were identified by analysis of the change curves in the time dimension, and then the long time series was divided into several short series to classify the MODIS data between 2000 and 2018. Lu et al. [33] extended the BFAST algorithm from 1D pixel value analysis to Spatio-temporal 3D array analysis to detect vegetation changes recorded in MODIS data. Ai et al. [34] applied an object-based classification algorithm using Landsat data to the dynamics of the Yangtse River estuary between 1985 and 2016.
Qiu et al. [35] improved the accuracy of land-cover classification using Sentinel-2 images by combining CNN and RNN to learn seasonal observation images. Sudi et al. [36] devised an unsupervised method for temporal image change detection by pre-training CNNs to extract features and using self-supervised training LSTM networks to reconstruct the temporal sequence.
In recent year, CNN-based semantic change detection studies have emerged. Compared with the post-classification change detection method, CNN-based semantic change detection extracts semantic features from two-temporal images directly and generates semantic change prediction maps "end-to-end", thus avoiding the cumulative error problem. Daudt et al. [6] presented a CNN architecture for two-temporal image semantic change detection and proposed a network structure with three CNN branches, including two branches to extract the semantic information in each temporal item and one branch to learn the change information. Yang et al. [8] exploited this 3-branch network architecture by weighting in semantic branches fusion of multilayer features in the semantic branches, and learning the semantic feature differences in the changing branches to improve the detection of which change. Peng et al. [16] learned change information by multi-layer difference features. They use a visual attention mechanism to introduce change information into semantic branches and directly output semantic change maps. Ding et al. [15]proposed a novel CNN structure that fuses high-level semantic features to detect changes. Compared to the Daudt et al. [6] method, this method shows advantages in accuracy and efficiency. Recurrent Neural Networks(RNNs) have shown excellent performance in the field of natural language processing. Ding et al. [15] proposed RNNs as a unit with multiclass in detecting temporal features in a recent paper.

3 MOTIVATION AND GOALS IN THIS THESIS

In this Ph.D. activity, our goal is to develop solutions to improve the accuracy and efficiency of semantic change detection of RSI using deep learning.

3.1 Problem Statement

The issues we mainly focus on are as follows:

1) Modeling the correlations between semantic and change information.
Semantic information and change information are correlated with each other in semantic change detection. Semantic feature similarity directly reflects whether surface classes change and change features can support reasoning on multi-temporal semantic classes. Most of the existing change detection methods extract change features by CNN directly, without analyzing the correlation between semantic information and change information. The key point to address in this thesis is how to model the learning of "semantics-change" correlations.
2) Joint extraction of semantic information and changes detection with limited label samples.
Semantic change detection requires identifying the kinds of surface classes before and after the change, so it is more complicated than BCD. This makes it difficult to obtain a sufficient number of training samples. Thus, a fundamental problem is how to exploit supervised information and unsupervised data with limited sample quantity to improve recognition accuracy.
3) Analysis of time-series image change features.
Surface information shows a stage change pattern during

the observation period. Analyzing the time series of observation data features can help to more robustly get semantic information on the surface, distinguishing between periodic and semantic changes, and locating the time point of change. However, there are few studies on change detection in time-series images. The central issue is how to extract historical memory information from time-series images and analyze changes in time-series features.

3.2 Motivation of the Thesis

As already mentioned, this thesis focuses on the semantic change detection problem of multi-temporal high-resolution images, which is urgently needed in remote sensing applications, It is proposed to overcome problems encountered by current methodologies such as high data dependency, poor detection stability, and difficulty in analyzing temporal changes. In order to enhance the capability of intelligent ground observation and analysis, and to enrich the theoretical techniques related to semantic change detection, we will promote the radicalization for change detection in remote sensing applications. For this reason, We plan to studying three aspects:

3.2.1 Enhance Stability in Semantic Change Detection. To optimize the distributions of semantic classes in the prediction results (to make it closer to the prior distribution), reduce false and missed detection, where learn correlations between semantic information, and change information on the landscape to improve the discrimination of minority classes and the consistency of multi-temporal prediction results.

3.2.2 To Improve the Training of Semantic Change Detection Techniques with a limited amount of Labels samples. To design training methods for semantic change detection to alleviation problems such as insufficient variation of samples and difficult semantic class identification.

3.2.3 Develop of SCD methods for time-series HR of RSIs. Through the joint analysis of spatio-temporal features, we will achieve end-to-end predictions of each temporal semantic map and adjacent temporal change maps. We will analyze the historical information from unlabeled observation images, to distinguish between periodic changes and semantic changes, and to improve the accuracy of semantic class discrimination.

3.3 Goals of the thesis

In order to address the above mentioned motivations. Firstly, we study semi-supervised training strategies to address issue of the limited number of samples. Secondly, We use classical theory and cutting-edge technology to analyze the correlation between learning semantic features and change features to address the technological difficulties. Finally, we consider the application requirements for time series observation, designing a semantic change detection algorithm for combining spatio-temporal information to expand application scenarios. In the three studies, the analysis of the inherent mechanisms of semantic changes relevant. Fig.2, we presents our research objectives, research content, experimental data, and range for this thesis.
There are three main goals in this thesis that we presented in detail in the next subsection.

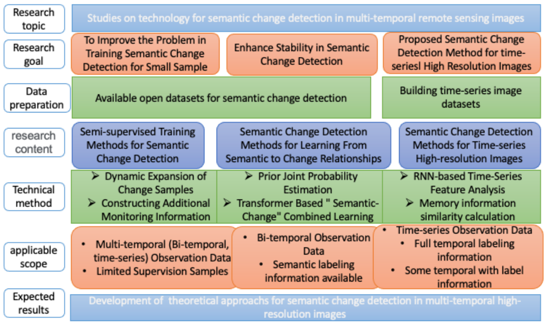

Research topic
Studies on technology for semantic change detection in multi-temporal remote sensing images
Research goal
To Improve the Problem in Training Semantic Change Detection for Small Sample
Enhance Stability in Semantic Change Detection
Proposed Semantic Change Detection Method for time-seriesl High Resolution Images
Data preparation
Available open datasets for semantic change detection
Building time-series image datasets
research content
Semi-supervised Training Methods for Semantic Change Detection
Semantic Change Detection Methods for Learning From Semantic to Change Relationships
Semantic Change Detection Methods for Time-series High-resolution Images
Technical method
➢ Dynamic Expansion of Change Samples
➢ Constructing Additional Monitoring Information
➢ Prior Joint Probability Estimation
➢ Transformer Based " Semantic-Change" Combined Learning
➢ RNN-based Time-Series Feature Analysis
➢ Memory information similarity calculation
applicable scope
• Multi-temporal (Bi-temporal, time-series) Observation Data
• Limited Supervision Samples
• Bi-temporal Observation Data
• Semantic labeling information available
• Time-series Observation Data
• Full temporal labeling information
• Some temporal with label information
Expected results
Development of theoretical approachs for semantic change detection in multi-temporal high-resolution images

Figure 2: Overview of the Ph.D. research program

3.3.1 Detection Method for Learning the Between Semantic and Change Relationships. Existing studies on semantic change detection mainly extend the CNN-based semantic change detection method. It mainly extracts different information directly via continuous convolutional transform, neglecting to learn the relevant relationships between specific semantic and change information. We focus on a specialized module to analyze the relationship between semantic and change information, and to optimize the feature classification process by prior semantic change probability. This will be done by develpoing a bi-directional RNN for semantic change detection of remote sensing images.

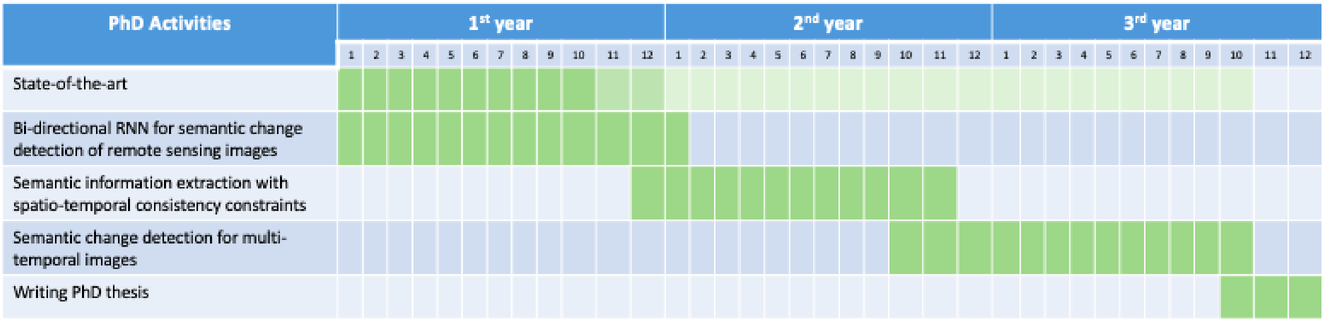
3.3.2 Semi-supervised Training Method for Semantic Change Detection. Currently, CNN-based semantic change detection models require large-scale data for training to achieve high accuracy and reliable data results. However, there are only a few semantic and change annotation available for model training in change detection applications. This makes the model insufficiently trained and easily under-fitted. To address this issue, in this thesis we propose to explore the use of semi-supervised approaches. On the one hand, we plan to expand the availability of labeled change sample, on other hand, we guide the training by constraint information. In this context we plan two main activities:

- Object-based change sample dynamic augmentation;
- Semantic information extraction with spatio-temporal consistency constraints.

3.3.3 Semantic Change Detection Method for Time-series High-resolution Images. There is growing demand to detect ongoing change information via time-series observations in remote sensing applications. Existing change detection methods are mainly designed for two images, thus it is difficult to have an end-to-end prediction of the semantic changes in time-series images. This section presents semantic change detection studies for time-series images to address practical application requirements. We consider two different application scenarios:

- Semantic change detection for learning historical image memory information;
- Learning spatio-temporal correlation for semantic change detection in time-series images.

In order to achieve the goals of the Ph.D., we plan to schedules activities according to the Gantt presented in Fig.3.



PhD Activities	1st year												2nd year												3rd year											
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
State-of-the-art	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	░	░	░	░	░	░	░	░	░	░	░	░	░	░	░	░	░	░	░	░	░	░	░	░		
Bi-directional RNN for semantic change detection of remote sensing images	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩																							
Semantic information extraction with spatio-temporal consistency constraints												🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩													
Semantic change detection for multi-temporal images																						🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
Writing PhD thesis																																		🟩	🟩	🟩

Figure 3: Gantt chart of the Ph.D. activities

4 PROPOSED METHODOLOGIES

Based on the issues mentioned in the Section 3, we developed new methodologies for semantic change detection. This study exploits open datasets to develop semantic change detection in two-temporal high-resolution images. Then we build time series of high-resolution image datasets to explore the semantic change detection approach in time-series of high-resolution data. Before introducing specific research methods, we mathematically define the semantic change detection tasks. We assume there are two pair of images (I_1, I_2) observed at different times from the same area, the (p_1, p_2) are corresponding pixels in (I_1, I_2). Semantic change detection can be denoted as a mapping function f_{scd} resulting in:

$$f_{scd}(p_1, p_2) = \begin{cases} (0,0), & c_m = c_n \\ (c_m, c_n), & c_m \neq c_n \end{cases} \tag{1}$$

where (c_m, c_n) is the semantic pair of classes corresponding to (p_1, p_2). After mapping all pixels, the generated results are two semantic change maps (S_1, S_2).

4.1 Bi-directional RNN for Semantic Change Detection in Remote Sensing Images

The traditional CNN-based change detection model ignores time series correlation. Each image of the time series is analyzed independent by a change detection branch, which is computationally intensive and not appropriate for modifying time series connection. RNN networks play a significant role in natural language processing, as they learn to analyze time-series signals, Their input and output formats can be flexible, making them good at Seq2Seq (sequence input to sequence output) tasks. We address the research topic of combining CNN and RNN deep network models to address the problem of semantic change detection problem.

Although CNNs have good performance in semantic change detection, they have some shortcomings in establishing a correlation between time series and change. CNNs lack comprehension of the context of the given input, thereby misjudging the change class. To address this issue, we propose a bi-directional RNN for semantic change detection of remote sensing images. RNNs adapt at handling time-series information learning from previous time steps to represent it. Therefore we use RNN to model time-series-change correlations in semantic change detection.

The network SSCDl [15] extracts features by two semantic branches, and change branch intended to learn change information. This has been experimentally tested in SCD networks. The weak connection between the semantic and change branches, leads to inconsistent predictions between the two temporal phases. This is a crucial problem. To tackle this issue, we introduce an RNN module to learn "semantic-temporal" correlations, which communicates information between the three branches, and propose a hybrid CNN and RNN network structure. As shown in Fig.4, encoder1 and encoder2 are used to obtain the two temporal features F_A and F_B. We present a Bi-directional RNN to build a correlation between the two times and learn the change features simultaneously. In the proposed BiRNN module, the neural units all take the two-temporal semantic features and memory features as input, and provides output the enhanced semantic features and memory features.

The first direction of change is computed as:

$$A_0 = f(\vec{U}^0 * S_0 + \vec{W}^0 * A_0 + \vec{b}^0) \tag{2}$$

while the second change direction is calculated as:

$$A_1 = f(\vec{U}^1 * S_1 + \vec{W}^1 * A_1 + \vec{b}^1) \tag{3}$$

where S denotes the memory information. S_0 indicates the initial value of the memory information (set to 0), S_1 indicates the memory (change) information in one direction and S_2 indicates memory (change) information in both directions. U, W, and b are model parameters, After calculations in both directions, A_0 and A_1 obtained information on the changes in the forward and backward directions, respectively. The last step is to utilize the weight matrix V to fuse A_1 and A_2 to generate the output feature F'_A at the time A as:

$$\mathbf{F}'_A = softmax(V) * [\vec{A}_1, \vec{A}_2](4)$$

Here A_1 and A_2 are connected together.

The output at the current time is determined by the memory and the input (F_A and F_B) of the current time.Thus, the output at each time takes into account the two-way change information, so that the semantic information at each time is not extracted in isolation, but is fused with the time-series information. Finally, a simple convolutional transformation is applied to the output memory S_2 to obtain the binary change map C. To sum up, the proposed can build better links between time semantic information and change information.

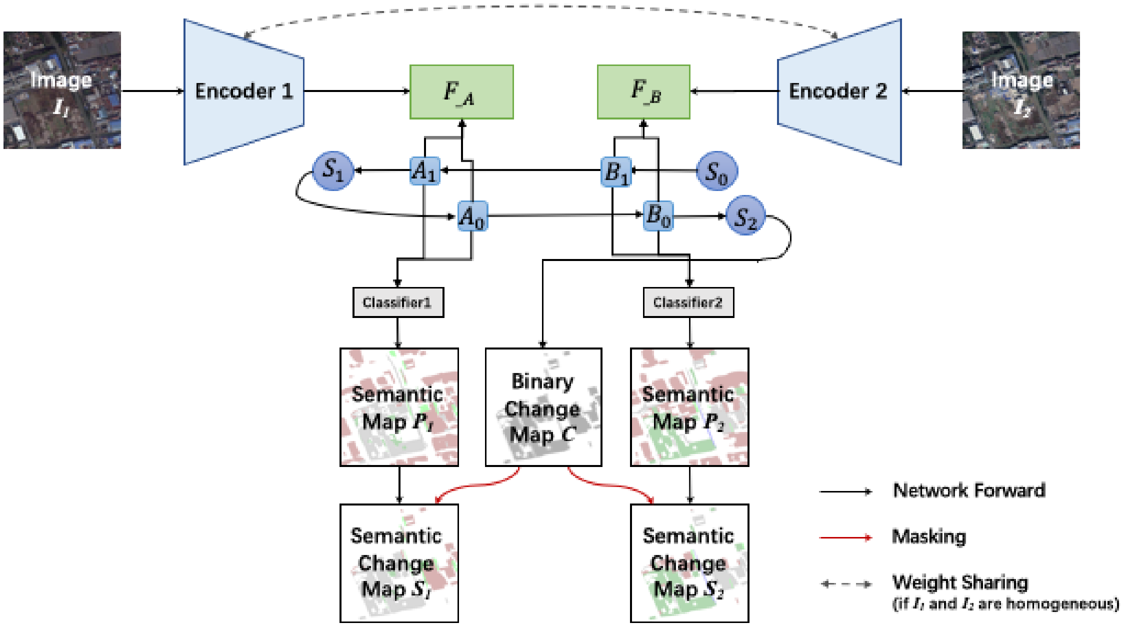


Figure 4: Bi-directional RNN for change analysis in time-series HR RSIs

4.2 Semantic Information Extraction with Spatio-Temporal Consistency Constraints

4.2.1 Object-based change sample dynamic augmentation. Surface semantic change classes. This results is insufficient amount of training samples for surface change. The training data should contain semantic annotations of each time period in a semantic change detection scenario. Therefore, we utilize the CutMix sample augmentation method in the semantic segmentation task to dynamically generate change samples using annotation information to alleviate the problem of insufficient valid samples.

The CutMix augmentation method is well recognized for its accuracy improvement in natural image semantic segmentation tasks. The methodology is basically based on the principle of cutting local regions of different images for patching and correspondingly patching their semantic annotations. We consider that remote sensing images contain fewer semantic classes than natural images, and the distribution of features is more concentrated (well-defined areas). In this section we extend CutMix to refine the expansion of semantic information.

We consider a pair of input images $\{I_1, I_2\}$ and its supervision information $\{L_1, L_2, L_c\}$, where L_1 and L_2 are the semantic information provided in L_1, L_2, to increase the number of change samples. The proposed technique is shown in Fig.5. First, each image I_i, is analyzed with its corresponding semantic annotation L_i, and the related feature objects are "cropped" by a mask calculation. This operation is performed on all training samples to generate a semantic sample library with semantic labels and feature patches. Afterwards, when the data is loading, a geometrically transformed feature object$\{O_j^I, O_j^L\}$ is randomly added to I_1 (or I_2) and L_1 (or L_2) by mask calculation, and the corresponding area $\{O_j^L\}$ is added to L correspondingly. This can be denoted as:

$$\mathbf{O}_j^M = O_j^I \geq 0 \tag{5}$$

$$\hat{\mathbf{L}}_c = L_c + O_j^M \tag{6}$$

$$\hat{\mathbf{I}}_1 = O_j^M \odot O_j^I + (1 - O_j^M) \odot I_1 \tag{7}$$

$$\hat{\mathbf{L}}_1 = O_j^M \odot O_j^L + (1 - O_j^M) \odot L_1 \tag{8}$$

where $\{O_j^M\}$ is $\{O_j^L\}$ of the mask, and $\odot$ means pixel-by-pixel multiplication operation. After this expansion operation, the number of non-zero regions is increased on L_c, which means more samples of variation.

This variation augmentation process is dynamic and regulates the selection of $\{O_j^I, O_j^L\}$ through the definition of random parameters and geometric modification. The generated $\{O_j^I, O_j^L\}$ class and location must be controlled so that they adhere to the prior class probabilities and disturb as little as possible the foreground/background and contextual spatial relationships on the remote sensing images so that the expanded change samples match the original data distribution. In this research activities we study how to control these space and class limitations to increase the accuracy of generated samples. Potential solutions include the introduction of unsupervised constraint methods, such as adversarial training.

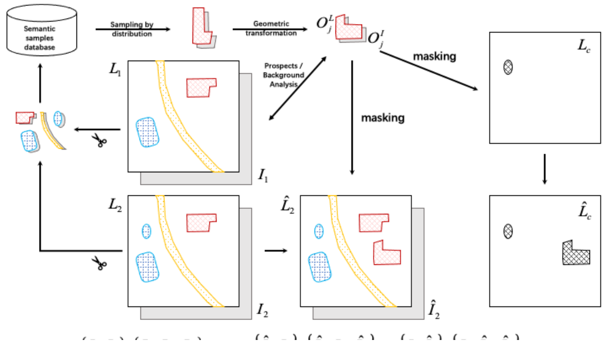


Figure 5: Dynamic augmentation to the semantic change samples.

4.2.2 Semantic information extraction with spatio-temporal consistency constraints. In the common settings of the SCD task, the changed areas are provided with semantic labels, whereas the unchanged areas are annotated with 'unchanged'. In other words, the number of semantic labels are very limited. This leads to the challenge of learning semantic information with limited samples. However, considering the internal mechanism in the SCD, it is possible to utilize the bi-temporal consistency constraints to improve the learning of semantic information. Let us consider the cases in 'changed' and 'unchanged' region, respectively. See Fig.6. In the following we discuss the two cases respectively. First, let us consider the 'changed' areas that are provided with semantic labels. For the input image $\{I_1, I_2\}$, via the network model $f_{\theta 1}(\cdot)$ and $f_{\theta 2}(\cdot)$ (where θ_1, θ_2 are the network parameters), prediction of the semantic classes corresponding to the map $\{Y_1, Y_2\}$ is obtained. General methods use $\{Y_1, Y_2\}$ and semantic labeling $\{\ell_1, \ell_2\}$ to calculate the loss function $\mathcal{L}_{sem}$. For example, the common cross-entropy losses can be calculated as:

$$\mathcal{L}_{sem}(Y_1) = -L_c log(Y_1) - (1 - L_1) log(1 - Y_1) \tag{9}$$

$$\mathcal{L}_{sem}(Y_2) = -L_c log(Y_2) - (1 - L_2) log(1 - Y_2) \tag{10}$$

According to characteristics of the semantic change detection task, the semantic information of Y_1 and Y_2 should have similarity in the unchanged region, but not in the changed region. Therefore, by using this factor as auxiliary information, a semantic consistency loss function L_{sc} can be constructed to guide the network training. For the point (p_1, p_2) and its corresponding prediction category (c_m, c_n), let the change corresponding to this position be labeled as y_c (1 denotes *changed*, 0 denotes *unchanged*). $\mathcal{L}_{sc}$ is calculated as:

$$\mathcal{L}_{sc} = \begin{cases} 1 - sim(c_m, c_m), & y_c = 1 \\ sim(x_1, x_2), & y_c = 0 \end{cases} \tag{11}$$

where *sim*() represents a similarity function calculated on the vectors. The loss objectives can be represented as follow:

$$\mathcal{L}_p = \mathcal{L}_{sem} + \mathcal{L}_{sc} \tag{12}$$

By adding $\mathcal{L}_{sc}$, two temporal semantic information is jointly considered, which improves the discrimination of critical areas.

For *unchanged* areas that are not provided with semantic labels, their semantic categories can be inferred with the bi-temporal predictions. We adopt the pseudo-labeling method that is commonly used in semi-supervised semantic segmentation tasks, generating a class label from areas with high confidence in the model prediction results. Most surface areas show no change and the semantic information remains unchanged in the semantic change detection. Therefore, we can jointly consider the multi-temporal predictions to discriminate the surface type. It is considered that high confidence levels can be achieved in regions where class predictions on Y_1 and Y_2 are close, generating pseudo-semantic labeling L. Calculation at p is given by:

$$\mathbf{L}_p = \begin{cases} argmax(Y_p^1), & sim(Y_p^1, Y_p^1) \geq T_{sim} \\ 0, & sim(Y_p^1, Y_p^1) < T_{sim} \end{cases} \tag{13}$$

where T_{sim} is the similarity threshold. A mark 0 denotes this pixel is not involved in training loss calculation. Semantic loss function is the same as in equation (9,10), except that the multi-temporal semantic loss calculation occurs via L is calculated.

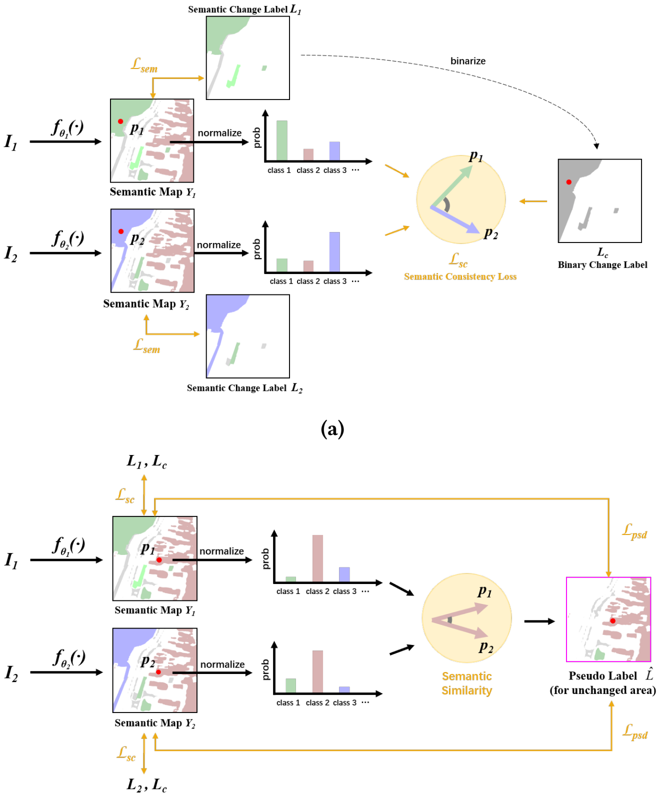


Figure 6: The objectives to learn semantic information in (a) 'changed' areas and (b) 'unchanged' areas.

4.3 Semantic Change Detection Method for Time-series of High-resolution Images

4.3.1 Semantic change detection for learning historical image memory information. Multiple historical images from different time periods of observation are often available. However, limited images have corresponding manual annotations in time-series monitoring applications. In this case, the memory information from the historical images can be extracted for training, and the accuracy and stability of the class prediction can be improved.

In Fig.7 we present a semantic change detection network based on historical image similarity analysis of us consider on existing recent observed image $I_t$with a variable length historical image sequence $\{I_{t-1}, I_{t-2}...I_1\}$ $(T \geq 3)$. Moreover, let us consider that only a part of the historical images (more than one) has annotation information. For presentation purposes, let us assume that the image at time $t-1$ has label L_{t-1} available. This section presents research that aims to use historical images to predict semantic class map Y_t at the time t, and the relative semantic change map Y_c at the time $t-1$. Firstly, the CNN network is used for feature extraction of $\{I_{t-1}, I_{t-2}...I_1\}$, I_t and L_{t-1} to obtain three sets of features: 1) memory features $K_m \in \mathbb{R}^{t \times c \times hw}$ (t is the time duration, c is the feature length, h and w are the spatial dimension size), 2) query features $K_q \in \mathbb{R}^{c \times hw}$, 3) memory value features $V_m \in \mathbb{R}^{c \times hw}$, and then make a similarity computation between K_m and K_q to evaluate the consistency between the current image and the historical image.

$$\mathbf{S}^k = sim(K_m, K_q) \tag{14}$$

where: Sim $\in \mathbb{R}^c \times \mathbb{R}^c \rightarrow \mathbb{R}$ is a position-by-position similarity function. $S^k \in \mathbb{R}^{t \times hw \times hw}$ is the calculated similarity matrix. Then we pool s^k along the time dimension. After normalization, a similarity weight matrix $W^k \in \mathbb{R}^{hw \times hw}$ is obtained. This calculation is performed point-by-point and can be expressed as:

$$\hat{\mathbf{S}}^k_{jk} = avg(S^k_{ijk}), \mathbf{W}^k_{jk} = \frac{expS^k_{jk}}{\sum_j expS^k_{jk}} \tag{15}$$

The query value feature $V_q \in \mathbb{R}^{c \times hw}$ is calculated by multiplying V_m with W^k:

$$\mathbf{V}^q = V_m W^k \tag{16}$$

Finally, V_q is processed to convolution and up-sampling, to obtain the predicted results $Y_t$and Y_c. Compared to the bi-temporal semantic change detection, this method can search the similarity values in the memory information to assist in the discrimination, thus improving the stability of prediction.

4.3.2 Learning spatio-temporal correlation for semantic change detection in time-series images. In this section, we will build a multi-temporal dataset. In section 4.1, we discussed RNNs and CNNs, we plan to extend this method to multi-temporal semantic change detection. Let us consider a collection of observation images $\{I_1, I_2...I_T\}$ $(T \geq 3)$. For time-series image semantic change tasks, it is necessary to get the class prediction map for each pixel in the time series and the change detection results for each neighboring time series. Semantic change diagram for each time-series $\{S_1, S_2...S_T\}$ can be further calculated by masking:

$$\mathbf{S}_t = P_t \bigodot Y^t_c \tag{17}$$

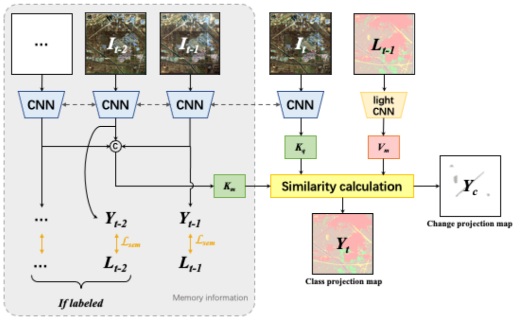


Figure 7: Similarity analysis based on historical images for semantic change detection

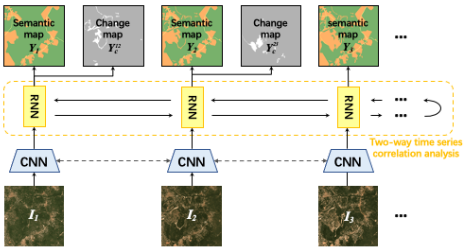


Figure 8: RNN based on CNN Semantic Change Detection for Multi-temporal Images

where t denotes the t time series) Y^t_c is obtained. A simple illustration of this model follows in Fig.8. The model firstly uses several lightweight CNN networks with shared weights to extract spatial features and get the semantic information of each time step; Secondly, it analyses the temporal correlation between each temporal feature via a series of bi-directional RNN modules. Finally, it learns the temporal "memory" information.

Let $f_\theta(\cdot)$ be the feature transformation function for the CNN network counterpart. For each image (I_1), features are extracted by the CNN network and expanded into a vector signal along the spatial dimension$f_\theta(I_1)$. In this model, all CNN networks share the same weights to reduce the number of network parameters and reduce the dependence on samples. To lower dimensional spatial-temporal space and interaction with temporal memory information. We apply the following computation:

$$\mathbf{O}^t = W_i X^t + W_s r^{t-1} \tag{18}$$

$$\mathbf{r}^t = \tanh o^t \tag{19}$$

where o^t is the output feature in time phase t, r^{t-1} and r^t are temporal memory information for the former and current time phases, respectively. W_i and W_x both represent the transformation matrix of the fully connected layer. In order to make the change information interact with the semantic information, we define r^0 as the change feature generated by convolution and dimensional transformation from all temporal features $X = \{x_1, x_2 ... x_T\}$. The RNN network processes time-series signals in both chronological and inverse directions, learning time-series correlations of semantic features. Then it transforms output of each temporal phase into a semantic prediction map, and transforms hidden features o^t that memorize time-series states into a binary change map between adjacent time steps. In this model, spatial semantic information extraction includes time-series signals, while change information comes from time-series state transformation directly, thus achieving joint learning for time-domain and space-domain information.

5 PRELIMINARY RESULTS

In this section we present the results obtained by the network described in section 4.1 and provide preliminary evaluation on its performance compared to several state-of-the-art models in the computer vision community.

5.1 Dataset and Evaluation Metrics

5.1.1 Datasets Description. The experiments were carried out in the SEmantic Change detection Dataset (SECOND) [14], a benchmark dataset for the SCD. The SECOND is constructed with bi-temporal HR optical images (containing RGB channels) collected by several aerial platforms and sensors. The observed regions include several cities in China, including Hangzhou, Chengdu and Shanghai. Each image has the spatial size of 512 × 512 pixels. The spatial resolution varies from 0.5m to 3m per pixel. The semantic labels were annotated by a professional annotation team. The LC categories before and after the change events are provided. In each GT semantic change map, one change class and six LC classes are annotated, including *unchanged*, *non-vegetated ground surface*, *tree*, *low vegetation*, *water*, *buildings* and *playgrounds*. These LC classes are selected considering the commonly interesting LC classes and the frequent geographical changes [37]. The bi-temporal LC transitions raise a total of 30 LC change types. The changed pixels account for 19.87% of the total image pixels. Among the 4662 pairs of temporal images, 2968 ones are openly available. We further split them into a training set and a test set with the numeric ratio of 4 : 1 (i.e., 2375 image pairs for training and 593 ones for testing).

5.1.2 Evaluation Metrics. In this study, 3 evaluation metrics are adopted to evaluate the SCD accuracy, including: overall accuracy (OA), mean Intersection over Union (mIoU) and Separated Kappa (SeK) coefficient. OA has been commonly adopted in both semantic segmentation [38]and CD [6] tasks. Let us denote $Q = \{q_{i,j}\}$ as the confusion matrix where $q_{i,j}$ represents time the number of pixels that are classified into class i while their index is j ($i, j \in \{0, 1, ..., N\}$, (represents *unchanged*). OA is calculated as:

$$OA = \sum_{i=0}^{N} q_{ii} / \sum_{i=0}^{N} \sum_{j=0}^{N} q_{ij}. \tag{20}$$

Since the *unchanged* pixels are the majority, *OA* cannot well-describe the identification of semantic categories. Therefore, in [8] mIoU and SeK are introduced to assess the discrimination of *changed/ unchanged* regions and the segmentation of LC classes, respectively.

mIoU is the mean value of the *IoU* of *unchanged* regions (IoU_{nc}) and that of the changed regions (IoU_c):

$$mIoU = (IoU_{nc} + IoU_c)/2, \tag{21}$$

$$IoU_{nc} = q_{00} / (\sum_{i=0}^{N} q_{i0} + \sum_{j=0}^{N} q_{0j} - q_{00}), \tag{22}$$

$$IoU_c = \sum_{i=1}^{N} \sum_{j=1}^{N} q_{ij} / (\sum_{i=0}^{N} \sum_{j=0}^{N} q_{ij} - q_{00}), \tag{23}$$

The SeK coefficient is calculated based on the confusion matrix $\hat{Q} = \{\hat{q}_{ij}\}$, where $\hat{q}_{ij} = q_{ij}$ except that $\hat{q}_{00} = 0$. This is to exclude the true positive *unchanged* pixels, whose number is dominant. The calculations are as follows:

$$\rho = \sum_{i=0}^{N} \hat{q}_{ii} / \sum_{i=0}^{N} \sum_{j=0}^{N} \hat{q}_{ij}, \tag{24}$$

$$\eta = \sum_{i=0}^{N} (\sum_{j=0}^{N} \hat{q}_{ij} * \sum_{j=0}^{N} \hat{q}_{ji}) / (\sum_{i=0}^{N} \sum_{j=0}^{N} \hat{q}_{ij})^2, \tag{25}$$

$$SeK = e^{IoU_c - 1} \cdot (\rho - \eta)/(1 - \eta). \tag{26}$$

The mIoU and SeK directly evaluate the sub-tasks in SCD, i.e., the CD and the SS of LCLU classes, respectively. Additionally, to evaluate more intuitively the segmentation of LCLU classes in changed areas, the F_{scd} is introduced in [15], which is calculated as:

$$P_{scd} = \sum_{i=1}^{N} q_{ii} / \sum_{i=1}^{N} \sum_{j=0}^{N} q_{ij}, \tag{27}$$

$$R_{scd} = \sum_{i=1}^{N} q_{ii} / \sum_{i=0}^{N} \sum_{j=1}^{N} q_{ij}, \tag{28}$$

$$F_{scd} = \frac{2 * P_{scd} * R_{scd}}{P_{scd} + R_{scd}} \tag{29}$$

where P_{scd} and R_{scd} are variants of the *Precision* and *Recall* [38] , respectively, that are calculated in the changed areas only.

5.2 Experimental settings

The experiments are conducted on servers with NVIDIA RTX3090 GPUs. The methods are implemented with Pytorch. The same training parameters are set, including batch size (8), running epochs (50) and initial learning (0.1) The gradient descent optimization method is Stochastic Gradient Descent (SGD) with Nesterov momentum. The augmentation strategy includes random flipping and rotating while loading the image pairs. For simplicity, no test-time augmentation operation is applied.

5.3 Results

5.3.1 Quantitative Results. As introduced in Sec.4, we introduced task-specific loss objectives ($\mathcal{L}_{sc}$ and $\mathcal{L}_{psd}$) to guide the learning of semantic information, and introduce RNN modules to improve the learning of semantic-change correlations. We perform an ablation study on the basis of the SCD framework SSCDl [15] to evaluate these proposed methods, and report the results in Table.1.

Table 1: Quantitative results of the ablation study.

Methods	Proposed Techniques				Accuracy			
	RNN	Bidirectional RNN	$\mathcal{L}_{psd}$	$\mathcal{L}_{sc}$	mIoU(%)	Sek(%)	OA(%)	F_{scd}(%)
SSCDl [15]					72.60	21.86	87.19	61.22
SSCDl*				√	73.06	22.68	87.48	61.98
SSCDl**			√	√	73.17	22.97	87.50	62.26
SSCDl-RNN	√			√	73.13	23.34	87.47	62.84
SSCDl-BiRNN		√		√	73.14	23.26	87.77	62.91
SSCDl-BiRNN*		√	√	√	**73.27**	**23.65**	**87.87**	**63.32**

First let us assess the improvements brought by the semantic learning objectives. The SSCDl* and SSCDl** refer to the methods of training the SSCDl baseline with different semantic learning objectives, i.e. using only $\mathcal{L}_{sc}$ and using both $\mathcal{L}_{sc}$ and $\mathcal{L}_{psd}$, respectively. One can observe that SSCDl* has advantages of around 0.76% in F_{scd} and 0.29% in OA compared to the baseline SSCDl. The SSCDl** further outperforms the SSCDl* by around o.3% in both Sek and F_{scd}.

Then we further assess the accuracy of using RNN modules. First we add the RNN module on the basis of SSCDl*. This improves the learning of semantic classes, which improves F_{Sek} by 0.66% and F_{scd} by 0.86%. Second, we replace the RNN to BiRNN. This brings marginal improvements of around 0.07% in F_{scd} and 0.3% in OA (compared to the results of SSCDl-RNN). Finally, we use the SSCDl-BiRNN and adopt the semantic learning objective $\mathcal{L}_{psd}$. This further improves the accuracy by around 0.4% in Sek and F_{scd}.

To conclude, the RNN modules improves the discrimination of semantic classes. The semantic learning objectives improve not only the learning of semantic information but also the detection of changes.

5.3.2 Effects of the Bidirectional RNN module. In Fig.9 we present the quantitative results before and after the use of RNN modules. One can observe that each of our proposed approaches has effective detection for *building* and *low vegetation* (Fig.9(a)). The RNN modules also help to reduce the false alarms (see in Fig.9(b)). The RNN method is much better than the SSCDl method in detecting change. The results of SSCDl-RNN and SSCDl-BiRNN are pretty close, while the latter is slightly better in discriminating the semantic classes.

5.3.3 Visualization of the pseudo labels. The $\mathcal{L}_{psd}$ is calculated with pseudo labels, which are generated with bi-temporal semantic predictions. To visually assess the quality of generated pseudo labels, we present some examples in Fig.10. The pseudo labels cover the pixels with no semantic labels, thus their correctness should be visually compared with the images. One can observe that the generated labels are generally correct, and different semantic categories are included. This supervision function improves the learning of semantic information in unchanged regions where the bi-temporal prediction confidence is high.

5.3.4 Effects of the semantic learning objectives. To qualitatively assess the performance of the proposed techniques, in Fig.11 we present the results obtained before and after the use of $\mathcal{L}_{sc}$ and $\mathcal{L}_{psd}$. One can observe that the semantic learning objectives improve the recognition of certain semantic categories, such as the discrimination between *vegetation* and *ground* in Fig.11(a1). They also help to better detect the change from *ground* in Fig.11(b1) to *playground* in Fig.11(b2).

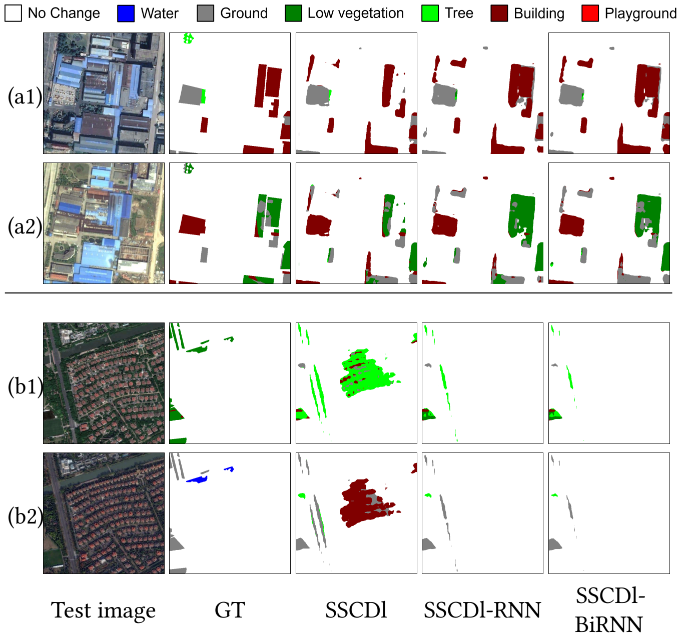


Figure 9: Example of results provided by different methods in the comparative experiments.

6 CONCLUSIONS

Semantic change detection in remote sensing images has been studied in this doctoral thesis. We found state-of-the-art studies mostly based on CNNs for semantic change detection, in which the information sharing between temporal and change branches is insufficient. Thus, we propose to employ Bi-directional RNN for semantic change detection in HR remote sensing images. RNN is placed on the basis of the SSCDl network architecture. The proposed

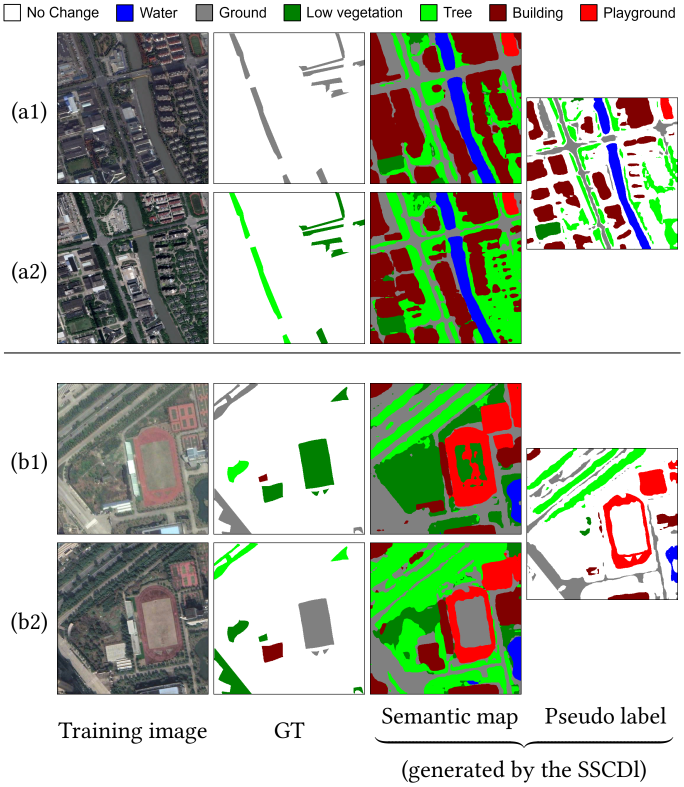


Figure 10: The generation of pseudo labels.

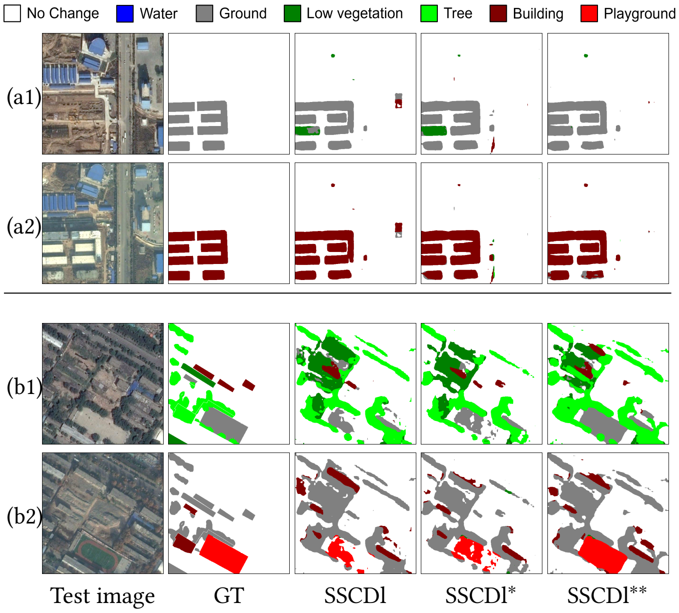


Figure 11: Comparison of results obtained by different methods in the ablation study (before and after the use of semantic learning objectives).

network employs a hybrid CNN-RNN structure, where the CNN captures the spatial features, while the RNN models the temporal correlations in terms of the semantic information. This allows the joint spatio-temporal analysis on the change features. Ablation studies have been conducted to assess this method. According to the preliminary results, this improves greatly the detection of changes and the discrimination of semantic classes.

Furthermore, we propose to improve the learning of semantic features through two task-specific objectives. A semantic consistence loss is proposed to improve the consistence of bi-temporal results, while a pseudo supervision technique is introduced to enhance the learning of semantic information in unchanged areas. Ablation study have demonstrated the effectiveness of proposed techniques.

In the near feature, we will also investigate to generate dynamic change samples to improve the learning of semantic changes. We will also extend the SCD to time-series images to continuously capture changes and semantic classes. Promising research outcomes can be expected in this Ph.D. activity, which will benefit the use of RS techniques in solving real-world problems.

REFERENCES

[1] Lorenzo Bruzzone and Francesca Bovolo. A novel framework for the design of change-detection systems for very-high-resolution remote sensing images. *Proceedings of the IEEE*, 101(3):609–630, 2012.

[2] Xi Guo, Qiqi Zhu, Weihuan Deng, and Qingfeng Guan. A siamese global learning framework for multi-class change detection. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 4348–4351. IEEE, 2021.

[3] Tong Li, Lizhen Cui, Zhihong Xu, Ronghai Hu, Pawan K Joshi, Xiufang Song, Li Tang, Anquan Xia, Yanfen Wang, Da Guo, et al. Quantitative analysis of the research trends and areas in grassland remote sensing: a scientometrics analysis of web of science from 1980 to 2020. *Remote Sensing*, 13(7):1279, 2021.

[4] Francesca Bovolo and Lorenzo Bruzzone. The time variable in data fusion: A change detection perspective. *IEEE Geoscience and Remote Sensing Magazine*, 3(3):8–26, 2015.

[5] Emily Hoffhine Wilson, James D Hurd, Daniel L Civco, Michael P Prisloe, and Chester Arnold. Development of a geospatial model to quantify, describe and map urban growth. *Remote sensing of environment*, 86(3):275–285, 2003.

[6] Rodrigo Caye Daudt, Bertrand Le Saux, Alexandre Boulch, and Yann Gousseau. Multitask learning for large-scale semantic change detection. *Computer Vision and Image Understanding*, 187:102783, 2019.

[7] Chen Wu, Bo Du, Xiaohui Cui, and Liangpei Zhang. A post-classification change detection method based on iterative slow feature analysis and bayesian soft fusion. *Remote Sensing of Environment*, 199:241–255, 2017.

[8] Kunping Yang, Gui-Song Xia, Zicheng Liu, Bo Du, Wen Yang, Marcello Pelillo, and Liangpei Zhang. Asymmetric siamese networks for semantic change detection in aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–18, 2021.

[9] Fei Yuan, Kali E Sawaya, Brian C Loeffelholz, and Marvin E Bauer. Land cover classification and change analysis of the twin cities (minnesota) metropolitan area by multitemporal landsat remote sensing. *Remote sensing of Environment*, 98(2-3):317–328, 2005.

[10] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[12] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2018.

[13] Hirokatsu Kataoka, Soma Shirakabe, Yudai Miyashita, Akio Nakamura, Kenji Iwata, and Yutaka Satoh. Semantic change detection with hypermaps. *arXiv preprint arXiv:1604.07513*, 2(4), 2016.

[14] Kunping Yang, Gui-Song Xia, Zicheng Liu, Bo Du, Wen Yang, Marcello Pelillo, and Liangpei Zhang. Semantic change detection with asymmetric siamese networks. *arXiv preprint arXiv:2010.05687*, 2020.

[15] Lei Ding, Haitao Guo, Sicong Liu, Lichao Mou, Jing Zhang, and Lorenzo Bruzzone. Bi-temporal semantic reasoning for the semantic change detection in hr remote

sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.

[16] Daifeng Peng, Lorenzo Bruzzone, Yongjun Zhang, Haiyan Guan, and Pengfei He. Scdnet: A novel convolutional network for semantic change detection in high resolution optical remote sensing imagery. *International Journal of Applied Earth Observation and Geoinformation*, 103:102465, 2021.

[17] Yi Zhang, Damon M Chandler, and Xuanqin Mou. Quality assessment of screen content images via convolutional-neural-network-based synthetic/natural segmentation. *IEEE transactions on image processing*, 27(10):5113–5128, 2018.

[18] Anju Asokan and JJESI Anitha. Change detection techniques for remote sensing applications: a survey. *Earth Science Informatics*, 12(2):143–160, 2019.

[19] Turgay Celik. Unsupervised change detection in satellite images using principal component analysis and k-means clustering. *IEEE geoscience and remote sensing letters*, 6(4):772–776, 2009.

[20] JS Deng, K Wang, YH Deng, and GJ Qi. Pca-based land-use change detection and analysis using multitemporal and multisensor satellite data. *International Journal of Remote Sensing*, 29(16):4823–4838, 2008.

[21] Chengquan Huang, Kuan Song, Sunghee Kim, John RG Townshend, Paul Davis, Jeffrey G Masek, and Samuel N Goward. Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote sensing of environment*, 112(3):970–985, 2008.

[22] Michele Volpi, Devis Tuia, Francesca Bovolo, Mikhail Kanevski, and Lorenzo Bruzzone. Supervised change detection in vhr images using contextual information and support vector machines. *International Journal of Applied Earth Observation and Geoinformation*, 20:77–85, 2013.

[23] Sudipan Saha, Francesca Bovolo, and Lorenzo Bruzzone. Unsupervised multiple-change detection in vhr multisensor images via deep-learning based adaptation. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 5033–5036. IEEE, 2019.

[24] CREST JST. Change detection from a street image pair using cnn features and superpixel segmentation. 2015.

[25] Ashbindu Singh. Review article digital change detection techniques using remotely-sensed data. *International journal of remote sensing*, 10(6):989–1003, 1989.

[26] George Xian, Collin Homer, and Joyce Fry. Updating the 2001 national land cover database land cover classification to 2006 by using landsat imagery change detection methods. *Remote Sensing of Environment*, 113(6):1133–1147, 2009.

[27] Lorenzo Bruzzone and Sebastiano B Serpico. An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images. *IEEE transactions on geoscience and remote sensing*, 35(4):858–867, 1997.

[28] Lorenzo Bruzzone, Diego F Prieto, and Sebastiano B Serpico. A neural-statistical approach to multitemporal and multisource remote-sensing image classification. *IEEE Transactions on Geoscience and remote Sensing*, 37(3):1350–1359, 1999.

[29] Reihaneh Peiman. Pre-classification and post-classification change-detection techniques to monitor land-cover and land-use change using multi-temporal landsat imagery: a case study on pisa province in italy. *International journal of remote sensing*, 32(15):4365–4381, 2011.

[30] Zhe Zhu and Curtis E Woodcock. Continuous change detection and classification of land cover using all available landsat data. *Remote sensing of Environment*, 144:152–171, 2014.

[31] Begüm Demir, Francesca Bovolo, and Lorenzo Bruzzone. Updating land-cover maps by classification of image time series: A novel change-detection-driven transfer learning approach. *IEEE Transactions on Geoscience and Remote Sensing*, 51(1):300–312, 2012.

[32] Jining Yan, Lizhe Wang, Weijing Song, Yunliang Chen, Xiaodao Chen, and Ze Deng. A time-series classification approach based on change detection for rapid land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158:249–262, 2019.

[33] Meng Lu, Edzer Pebesma, Alber Sanchez, and Jan Verbesselt. Spatio-temporal change detection from multidimensional arrays: Detecting deforestation from modis time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117:227–236, 2016.

[34] Jinquan Ai, Chao Zhang, Lijuan Chen, and Dajun Li. Mapping annual land use and land cover changes in the yangtze estuary region using an object-based classification framework and landsat time series data. *Sustainability*, 12(2):659, 2020.

[35] Chunping Qiu, Lichao Mou, Michael Schmitt, and Xiao Xiang Zhu. Local climate zone-based urban land cover classification from multi-seasonal sentinel-2 images with a recurrent residual network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 154:151–162, 2019.

[36] Sudipan Saha, Francesca Bovolo, and Lorenzo Bruzzone. Change detection in image time-series using unsupervised lstm. *IEEE Geoscience and Remote Sensing Letters*, 2020.

[37] Xin-Yi Tong, Gui-Song Xia, Qikai Lu, Huanfeng Shen, Shengyang Li, Shucheng You, and Liangpei Zhang. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sensing of Environment*, 237:111322, 2020.

[38] Lei Ding, Hao Tang, and Lorenzo Bruzzone. Lanet: Local attention embedding to improve the semantic segmentation of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(1):426–435, 2020.