



Consistency-guided lightweight network for semi-supervised binary change detection of buildings in remote sensing images

Qing Ding, Zhenfeng Shao, Xiao Huang, Xiaoxiao Feng, Orhan Altan & Bin Hu

To cite this article: Qing Ding, Zhenfeng Shao, Xiao Huang, Xiaoxiao Feng, Orhan Altan & Bin Hu (2023) Consistency-guided lightweight network for semi-supervised binary change detection of buildings in remote sensing images, GIScience & Remote Sensing, 60:1, 2257980, DOI: [10.1080/15481603.2023.2257980](https://doi.org/10.1080/15481603.2023.2257980)

To link to this article: <https://doi.org/10.1080/15481603.2023.2257980>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 19 Sep 2023.



Submit your article to this journal [↗](#)



Article views: 568



View related articles [↗](#)



View Crossmark data [↗](#)

Consistency-guided lightweight network for semi-supervised binary change detection of buildings in remote sensing images

Qing Ding^a, Zhenfeng Shao^a, Xiao Huang^b, Xiaoxiao Feng^c, Orhan Altan^d and Bin Hu^a

^aState Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China; ^bDepartment of Geosciences, University of Arkansas, Fayetteville, AR, USA; ^cSchool of Computer Science, Hubei University of Technology, Wuhan, China; ^dDepartment of Geomatics Engineering, Istanbul Technical University, Istanbul, Turkey

ABSTRACT

Precise identification of binary building changes through remote sensing observations plays a crucial role in sustainable urban development. However, many supervised change detection (CD) methods overly rely on labeled samples, thus limiting their generalizability. In addition, existing semi-supervised CD methods suffer from instability, complexity, and limited applicability. To overcome these challenges and fully utilize unlabeled samples, we proposed a consistency-guided lightweight semi-supervised binary change detection method (Semi-LCD). We designed a lightweight dual-branch CD network to extract image features while reducing model size and complexity. Semi-LCD fully exploits unlabeled samples by data augmentation, consistency regularization, and pseudo-labeling, thereby enhancing its detection performance and generalization capability. To validate the effectiveness and superior performance of Semi-LCD, we conducted experiments on three building CD datasets. Detection results indicate that Semi-LCD outperforms competing methods, quantitatively and qualitatively, achieving the optimal balance between performance and model size. Furthermore, ablation experiments validate the robustness and advantages of the Semi-LCD in effectively utilizing unlabeled samples.

ARTICLE HISTORY

Received 9 March 2023
Accepted 7 September 2023

KEYWORDS

Building change detection;
remote sensing images;
semi-supervised learning;
consistency regularization;
lightweight network

1. Introduction

Change detection (CD) is a frontier research field in remote sensing that identifies and analyzes differences between various images to detect and quantify changes (Zhu, Qiu, and Ye 2022; Mahmoodzadeh 2007). Macro-scale CD can provide essential support for various applications, including forest protection (Xiao et al. 2021) and water resource investigation (Rokni et al. 2015), to name a few. In addition, micro-scale building change detection (BCD) can also contribute to applications such as urban planning and reconstruction after disasters (Cao and Huang 2023; Guo and Du 2017; Park and Song 2023). With the increasing capability of space-based Earth observation, high-resolution satellites such as WorldView, GeoEye, QuickBird, Zi-Yuan 3, and Gaofen have been widely adopted and provide abundant data for CD (Chen et al. 2020; Huang, Tang, and Qin 2022). However, the huge data volume and complex images pose new challenges to the BCD task (Li, Shi, and Zhu 2022; Ding et al. 2021). In this context, there is an urgent need to develop an accurate and efficient binary BCD method.

Considering the differences in training paradigms, CD methods can be categorized as unsupervised, supervised, and semi-supervised methods (Shu et al. 2022). Among these, traditional unsupervised methods, mainly including algebra-based and transformation-based methods, extract change information without priori knowledge. Algebra-based methods typically perform operations on images such as difference (Bruzzone and Prieto 2000) and ratio (Liu, Xin, and Chen 2014), then employ a threshold or classifier to identify changed areas. Most transformation-based methods rely on principal component analysis (Celik 2009), tassel cap transformation (Han et al. 2007), or multivariate alteration detection (Nielsen, Conradsen, and Simpson 1998) to optimize the spectral feature space, aiming to improve the CD results. Although traditional unsupervised CD methods are generally easy to implement and do not require labeled samples, they suffer from low accuracy and poor generalization due to the low adaptability of hand-designed features. These shortcomings make traditional

CONTACT Zhenfeng Shao  shaozhenfeng@whu.edu.cn

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

unsupervised methods challenging to meet the demand of high-resolution remote sensing-based CD tasks.

In contrast, supervised CD methods introduce ground truths to optimize classifier parameters, thereby enhancing the accuracy of CD results. In the earlier stages, machine learning methods such as support vector machine (Habib et al. 2009) and random forest (Li, Im, and Beier 2013) were employed for supervised CD. Although these methods outperform unsupervised methods, shortcomings in big data utilization and high-dimensional nonlinear modeling have limited their development. With the continuous increase in computing power, deep learning methods with more efficient feature extraction and more robust modeling capabilities have further contributed to the rapid progress of CD (Hou et al. 2020; Shafique et al. 2022). According to the target size, CD methods based on deep learning can be categorized into feature-based, patch-based, and image-based methods (Peng, Zhang, and Guan 2019). Most feature-based methods utilize model-generated deep features to derive a difference map and adopt threshold segmentation to achieve CD (Abdi and Jabari 2021; Hou, Wang, and Liu 2017; Saha, Bovolo, and Bruzzone 2019). However, errors in generating the difference map can accumulate in the final CD results. Patch-based methods use image patches as the model input units to judge if the central pixel has changed (Gong et al. 2017; Zhang et al. 2016). Although patch-based methods mitigate the impact of error propagation, determining the optimal patch size can be challenging and may lead to data redundancy and poor performance (Lei et al. 2019). Fully convolutional networks have significantly advanced image-based CD methods (Long, Shelhamer, and Darrell 2015). The end-to-end processing approach effectively addresses the previously mentioned issues while increasing efficiency and accuracy (Zhang, Ma, and Zhang 2022; Daudt, Saux, and Boulch 2018; Peng, Zhang, and Guan 2019; Zheng et al. 2021). Combining fully convolutional networks and transformers has recently achieved impressive CD results (Deng et al. 2023; Zhang et al. 2022).

Nonetheless, most supervised CD methods have a substantial positive correlation between the number of labeled samples and the detection accuracy. However, labeled samples are mainly obtained by manual visual interpretation, and obtaining massive high-quality samples is labor-intensive (Ding et al.

2022). In practical situations, labeled training samples often only account for a small proportion of the total samples. Limited labeled samples are not effectively represent the data characteristics, negatively impacting the accuracy of CD results. Several approaches can be tried to address this problem, including data augmentation, weakly supervised learning, and semi-supervised learning (Sun et al. 2022). Although data augmentation can slightly improve model performance, it does not effectively utilize unlabeled samples. Additionally, the complex spectra and textures of high-resolution images pose incredible difficulties in manually obtaining coarse annotation information such as image-level labels, scribbles, or bounding boxes, considerably increasing the difficulty of applying weakly supervised methods for CD tasks. Entering the “Big Earth Data” era with an excessive number of remote sensing images acquired by various satellites, how to take advantage of unlabeled samples remains a challenge, especially for CD tasks. Compared to data augmentation and weakly supervised learning, semi-supervised learning has been recognized as a preferred approach for addressing the insufficiency of labeled data considering its capability of utilizing unlabeled samples (Van Engelen and Hoos 2020).

To effectively leverage both limited labeled samples and abundant unlabeled samples while achieving BCD concisely and efficiently, we proposed a consistency-guided semi-supervised binary BCD method and developed a lightweight CD network (LCD-Net) with fewer parameters and better performance. The primary contributions of this study include:

- (1) We developed a lightweight dual-branch neural network to ensure CD performance while reducing model size for easy deployment on lightweight devices;
- (2) We proposed a semi-supervised binary BCD method that integrates sample perturbation, pseudo-labeling, and consistency regularization to leverage unlabeled samples for further improving detection accuracy;
- (3) We conducted experiments on three BCD datasets to verify the advantages and practicality of the Semi-LCD. Compared to other methods, Semi-LCD shows improved detection performance and generalization ability, achieving a better balance between performance and model size.

Sections 2 and 3 outline the related work and the BCD datasets used in this study, respectively. Section 4 provides a detailed explanation of Semi-LCD. Section 5 presents the comparative results and model complexity. Section 6 discusses the method's parameter settings, generalization ability, and advantages and weaknesses. The findings of this study are summarized in Section 7.

2. Related work

Recently, a growing focus has been on utilizing lightweight models and reducing reliance on labeled data. This section briefly overviews the research on lightweight neural networks and semi-supervised learning for semantic segmentation. Additionally, this section highlights the limitations of existing semi-supervised CD methods, thereby clarifying the motivation of this study.

2.1. Lightweight neural networks

To enhance models' performance, numerous studies have focused on designing neural networks with deeper or wider structures. Attention mechanisms (Wang and Sertel 2023; Zhang et al. 2023), multi-scale feature extraction modules (Ye et al. 2022; Zhu et al. 2023), and deep supervision mechanisms (Zhang et al. 2020) have been incorporated into these networks to improve prediction accuracy. However, the downside of overly complex networks is the increased number of model parameters, resulting in higher computational costs and memory consumption. Consequently, deploying such models on resource-constrained devices like satellites, drones, and smartphones becomes challenging.

Currently, there is an increasing emphasis on developing lightweight and efficient neural networks that achieve superior performance while maintaining broader applicability. Lightweight networks such as MobileNet (Sandler et al. 2018), ShuffleNet (Zhang et al. 2018), and EfficientNet (Tan and Le 2019) are widely adopted in semantic segmentation, target detection, and image retrieval tasks (Wang et al. 2023; Wieland et al. 2023; Zhang et al. 2022). These networks are favored for their remarkable efficiency and superior modeling capabilities. Among them, MobileNet stands out by employing depth separable convolution instead of standard convolution, which

ensures effective feature extraction while significantly reducing the model size. In addition, it greatly inspires subsequent research on lightweight neural networks (Wang et al. 2022; Yang et al. 2022; Yin et al. 2023).

2.2. Semi-supervised learning for semantic segmentation

Semantic segmentation aims to assign independent categories for each pixel in an image, which typically requires a large number of labeled samples. However, labeling samples in real scenarios is complex and costly, often yielding a limited number of labeled samples. Semi-supervised learning can effectively improve model performance by using additional unlabeled samples. In this context, semi-supervised learning develops rapidly and plays a significant role in semantic segmentation tasks, which can be broadly classified into three categories: pseudo-labeling, generative adversarial-based, and consistency learning methods (Shu et al. 2022).

Pseudo-labeling methods train models with labeled samples, then make predictions on unlabeled samples. These predictions are incorporated into the training dataset as pseudo-labels, expanding the available labeled samples (Chen et al. 2022; Zou et al. 2018). Generative adversarial-based methods train both networks concurrently, with the generator typically generating prediction results and the discriminator assessing the reliability or truthfulness of these predictions (Hung et al. 2019; Mittal, Tatarchenko, and Brox 2021). Consistency learning methods improve segmentation results while enhancing model robustness by minimizing the differences in prediction results among different perturbed versions of the same unlabeled sample (Li et al. 2022; Ouali, Hudelot, and Tami 2020).

2.3. Semi-supervised change detection

Binary CD aims to classify pixels into two categories: changed and unchanged, which can also be seen as a specialized form of semantic segmentation. Inspired by the semi-supervised methods in semantic segmentation tasks, a few studies have explored integrating semi-supervised learning and deep learning to achieve accurate CD. For instance, Peng et al. (2021) established two different discriminators to enhance the consistency of feature distribution between labeled and unlabeled samples, thereby improving

detection accuracy. The unstable convergence of generative adversarial networks affects its practical application. Sun et al. (2022) developed a semi-supervised CD siamese network using graph attention, which combines weak augmentation, strong augmentation, and consistency comparison. The augmentation approaches employed in this method are relatively cumbersome, and the network with many parameters adds complexity. Wele and Patel (2022) attempted to add perturbations to the feature space of unlabeled samples, thereby minimizing the differences in output results of multiple decoders to achieve consistency constraints. However, the effectiveness of this method partly relies on pre-trained parameters, and complex perturbation methods and network constraints its application in scenarios with severe shortage of labeled samples. Shu et al. (2022) introduced single temporal supervision as a complement to improve BCD results, but the acquisition of sufficient single temporal labels remains challenging, which restricts the universality of this method in reality.

Although these semi-supervised CD methods have shown promising performance, room for improvement remains, as existing methods are relatively unstable, complex, and less ungeneralizable. In light of these limitations, this study proposed a lightweight semi-supervised binary BCD method that integrates consistency regularization and pseudo-labeling strategies from the perspectives of conciseness and effectiveness.

3. Building change detection datasets

In this study, three binary BCD datasets were selected to validate the superior performance and practicality of the Semi-LCD. These datasets are the Multi-temporal Scene WuHan (MtS-WH) dataset, the WHU Building dataset, and the high-resolution complex urban scene BCD (HRCUS-CD) dataset. Due to memory limitations of the computing device, the raw images in these datasets with large sizes need to be cropped before they can be used as input data for the neural networks. To ensure smooth image downsampling in the neural networks, the edge length of the cropped samples is usually set to an integer power of 2, which helps extract shallow and semantic features more effectively. For each dataset, the samples were randomly divided into three parts for model training, validation, and final accuracy testing.

3.1. MtS-WH dataset

The MtS-WH dataset (Wu, Zhang, and Du 2017) contains two IKONOS images covering the Hanyang District of Wuhan City, captured in 2002 and 2009, respectively. As seen in Figure 1, the images have a size of 7200×6000 pixels and a resolution of 1 m. The original annotation of the dataset focuses on the scene changes. Zhang, Ma, and Zhang (2022) supplemented the MtS-WH dataset with pixel-level annotations of building changes, encompassing both new and demolished buildings. Small patches (128×128 pixels) were cropped from the MtS-WH dataset to form the training (1,500 pairs), validation (500 pairs), and test (500 pairs) sets.

3.2. WHU building dataset

The WHU Building dataset (Ji, Wei, and Lu 2019) encompasses two images covering an area in New Zealand, captured in 2012 and 2016, respectively. This dataset reflects the reconstruction process of the buildings after an earthquake. As shown in Figure 2, patches (128×128 pixels) with a resolution of 0.2 m were cropped from the WHU Building dataset. The training set contains 4,000 pairs of patches, and both the validation set and test set contain 1,600 pairs of patches.

3.3. HRCUS-CD dataset

The HRCUS-CD dataset (Zhang et al. 2023) covering Zhuhai city comprises 11,388 pairs of image patches (256×256 pixels) with a resolution of 0.5 m. Representative change samples in the HRCUS-CD dataset are shown in Figure 3. These image patches cover the built-up area in 2019 and 2022, as well as the rapid development area in 2010 and 2018. Changes within the built-up area encompass the demolition of buildings, the construction of new residential areas, the establishment of commercial zones and several warehouses. In the earlier images, the rapid development area was mainly dominated by farmland, hills, and sparsely distributed buildings. Subsequent urbanization resulted in a significant increase in new buildings and the demolition of some earlier buildings. The model training, validation, and accuracy testing on the HRCUS-CD dataset employed 7,974, 2,276, and 1,138 pairs of patches, respectively.

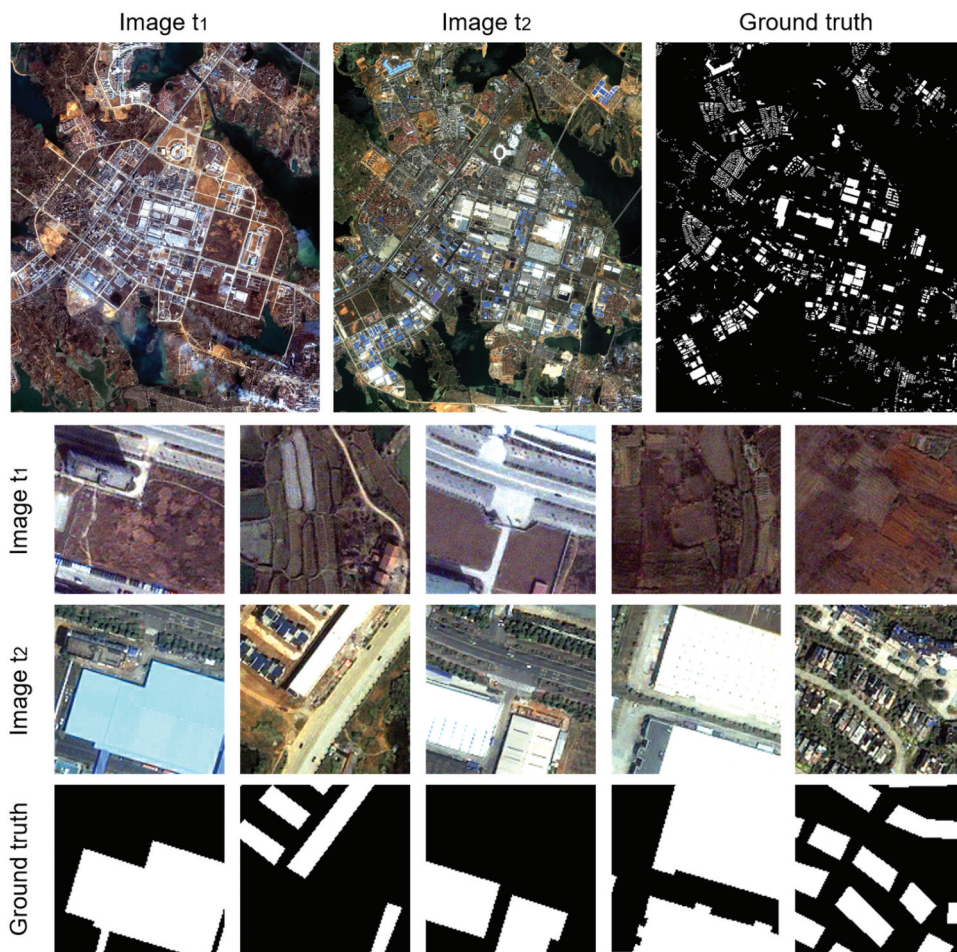


Figure 1. MtS-WH dataset and typical building change samples. Image t_1 and t_2 represent bi-temporal images covering the same area. Ground truth represents building change labels of bi-temporal images, with white indicating the changed area and black indicating the unchanged area.

3.4. Summary of different datasets

The details of the BCD datasets used in this study are further summarized. From Table 1, it is evident that the three datasets have significant differences in imaging time, resolution, sample size, quantity, and proportion of changed pixels.

The MtS-WH and WHU Building datasets are publicly available and exhibit high annotation quality. Comparative and ablation experiments on these two datasets more fairly demonstrate the superior performance of the Semi-LCD. However, given the relatively singular change scenarios and the limited samples in these two datasets, the applicability of the Semi-LCD in complex scenes is difficult to validate. To tackle this challenge, we utilized the HRCUS-CD dataset constructed through visual interpretation. This dataset covers both the built-up and rapid development areas, capturing various changes related to building

demolition and construction. Comparative experiments conducted on the HRCUS-CD dataset confirm the superior performance and practicality of Semi-LCD in detecting building changes in complex scenes. Additionally, we calculated the proportion of changed pixels in different datasets. Different subsets of the same dataset have similar changed proportions, confirming the randomness and rationality of subset division. Meanwhile, the HRCUS-CD dataset has the lowest percentage of changed pixels, approximately 2%, which poses a severe challenge for accurate CD and can more effectively validate the practicality of methods in real scenarios.

4. Methodology

To effectively enhance BCD performance with limited labeled samples, we proposed Semi-LCD method that leverages sample perturbation, consistency regularization, and pseudo-labeling. This section

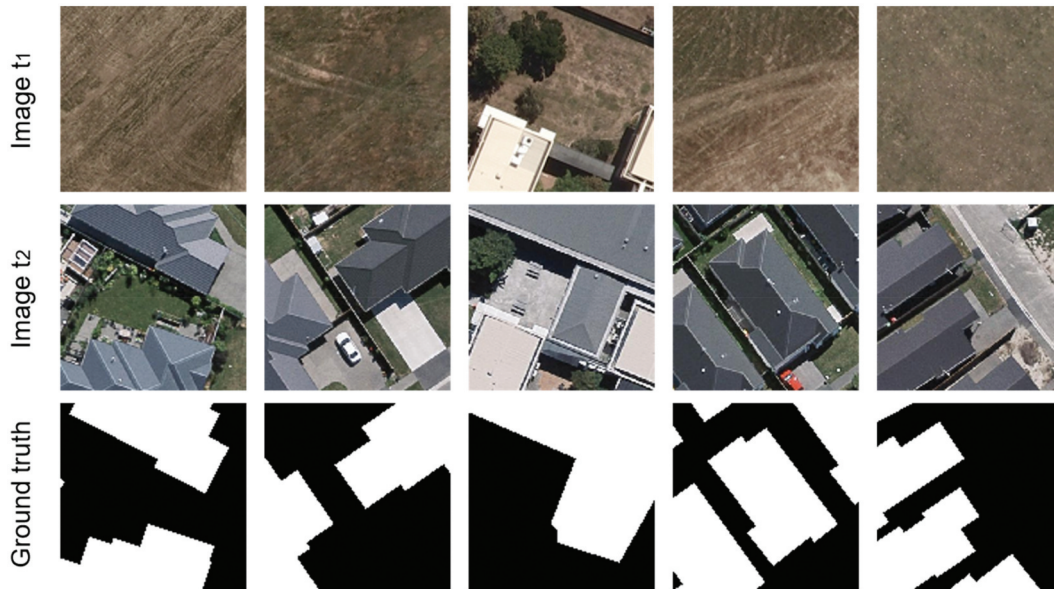


Figure 2. Representative change samples in the WHU Building dataset. Image t_1 and t_2 represent bi-temporal images covering the same area. Ground truth represents building change labels of bi-temporal images, with white indicating the changed area and black indicating the unchanged area.

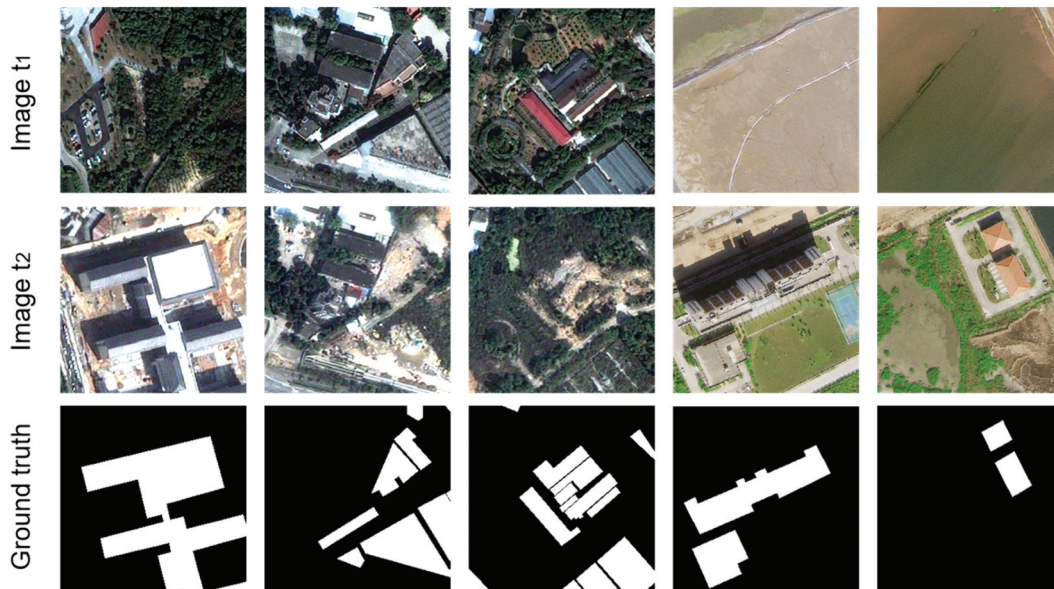


Figure 3. Representative change samples in the HRCUS-CD dataset. Image t_1 and t_2 represent bi-temporal images covering the same area. Ground truth represents building change labels of bi-temporal images, with white indicating the changed area and black indicating the unchanged area.

comprehensively describes the Semi-LCD, including its framework, network architecture, loss function, and performance evaluation.

4.1. Basic framework of semi-LCD

The proposed Semi-LCD consists of two modules: a supervised module that optimizes network parameters

using labeled samples, and an unsupervised module that improves model generalization using unlabeled samples. During model training, both the supervised and unsupervised modules update network parameters and enhance the BCD performance. Figure 4 demonstrates the basic framework of Semi-LCD.

In the supervised module, the LCD-Net takes image pairs that cover the same areas and have

Table 1. Imaging information and subset partitioning of three different BCD datasets.

Dataset	MtS-WH	WHU Building	HRCUS-CD
Acquisition time 1	2002	2012	2010, 2019
Acquisition time 2	2009	2016	2018, 2022
Resolution (m)	1	0.2	0.5
Sample size (pixel)	128×128	128×128	256×256
Training set	1,500	4,000	7,974
Validation set	500	1,600	2,276
Test set	500	1,600	1,138
Changed proportion (training)	8.70%	4.90%	2.16%
Changed proportion (validation)	9.11%	5.02%	2.23%
Changed proportion (test)	9.39%	4.53%	1.87%

corresponding ground truths as input data. The BCD results predicted by the model are then compared to the ground truths to calculate the supervised loss. The network parameters are updated using the backpropagation procedure and optimization algorithm.

The unsupervised module uses image pairs that cover the same areas but lack ground truths. These image pairs are fed into the LCD-Net to obtain the original BCD results. Subsequently, category confidence thresholds are set for changed and unchanged pixels, respectively. Pixel-level prediction results with higher confidence are retained, and the corresponding category pseudo-labels are obtained. Furthermore, the unlabeled image pairs and pseudo-labels are perturbed based on the same augmentation, ensuring the consistency between the corresponding pixels. The augmented unlabeled image pairs are input into the LCD-Net to obtain prediction results after perturbation. The unsupervised module employs several perturbation methods based on Albumentations library (Buslaev et al. 2020), including vertical flip, horizontal flip, random rotate 90°,

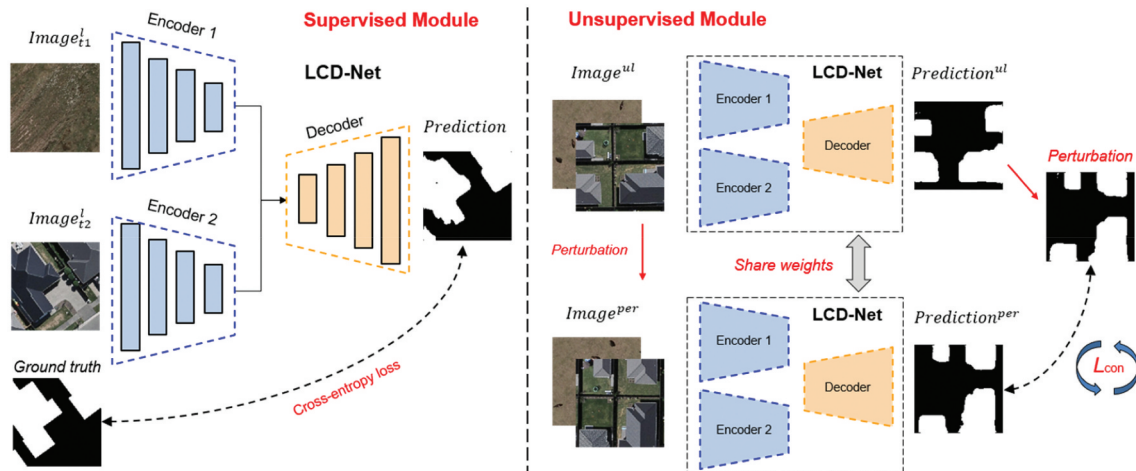
transpose, and random grid shuffle, as shown in Figure 5. For each pair of unlabeled samples, only one of the aforementioned perturbation methods is randomly applied at a time.

The principle of consistency regularization indicates that perturbations on input samples do not significantly affect output results (Fan et al. 2023). Therefore, the BCD results obtained by the model should remain consistent for the same object before and after perturbation. Applying this constraint, we compare the BCD results obtained after the perturbation with the pseudo-labels at the corresponding pixels. Finally, the consistency loss is obtained using unlabeled samples, which aids in updating the model parameters effectively, as shown below:

$$P = f(im_{t_1}^{ul}, im_{t_2}^{ul}; \theta) \quad (1)$$

$$P_{PER} = f(PER(im_{t_1}^{ul}), PER(im_{t_2}^{ul}); \theta) \quad (2)$$

$$L_{con} = F_{con}(CF(PER(P)), P_{PER}) \quad (3)$$

**Figure 4.** Schematic diagram of the proposed Semi-LCD method for binary BCD.

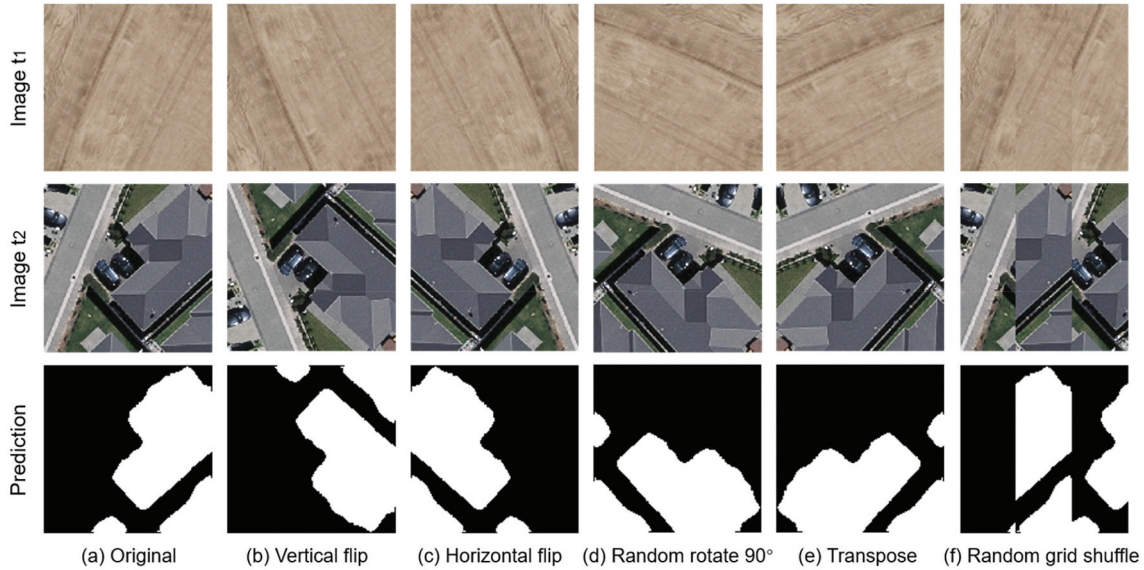


Figure 5. Schematic diagram of the sample perturbation methods used in the unsupervised module.

where im_{t1}^{ul} and im_{t2}^{ul} refer to bi-temporal images that lack ground truths, θ denotes the parameters of network f , PER represents the sample perturbation, and P and P_{PER} denote the predicted probabilities of the same sample obtained before and after adding perturbation, respectively. CF represents the confidence filter for screening high-quality pixels from unlabeled samples, and the predicted probabilities are converted into pseudo labels. F_{con} denotes the function used to calculate the consistency loss L_{con} of different results.

During the testing phase, the unsupervised module is not involved. Instead, we feed image pairs into the trained LCD-Net model with fixed parameters to obtain the change probabilities. Binarization of the change probabilities is performed to obtain the final BCD results, which can be compared with the ground truths to evaluate the model performance.

4.2. Lightweight change detection network

Existing state-of-the-art CD methods often exhibit considerable network complexity, which limits their application on lightweight devices. To address this problem, our study proposed LCD-Net, a high-resolution image CD network specifically designed to mitigate model complexity and size while

maintaining superior performance. LCD-Net introduces a dual-branch structure in which images of different phases are fed into different encoders with independent weights. Features extracted from encoders are combined and fed into the decoder, which employs an end-to-end approach to generate the final CD results. The overall architecture of the proposed LCD-Net is shown in Figure 6.

The two encoders in LCD-Net have the same structure. Each encoder consists of four convolutional units that extract multi-scale features from input images. Max pooling is employed between the convolutional units to downsample the features. Specifically, the first two convolutional units yield output features with 64 and 128 channels, respectively, each consisting of two consecutive convolutional layers. The latter two convolutional units increase the output features to 256 and 512 channels, respectively. We introduce lightweight high-dimensional feature extraction units, which use two consecutive depth separable convolutions to reduce the parameter count for extracting high-dimensional features. The architecture of the lightweight high-dimensional feature extraction unit is shown in Figure 7. Furthermore, these units also utilize a 1×1 convolutional operation to transform the input features in the channel dimension, enabling the effective fusion of features from different branches through a shortcut connection.

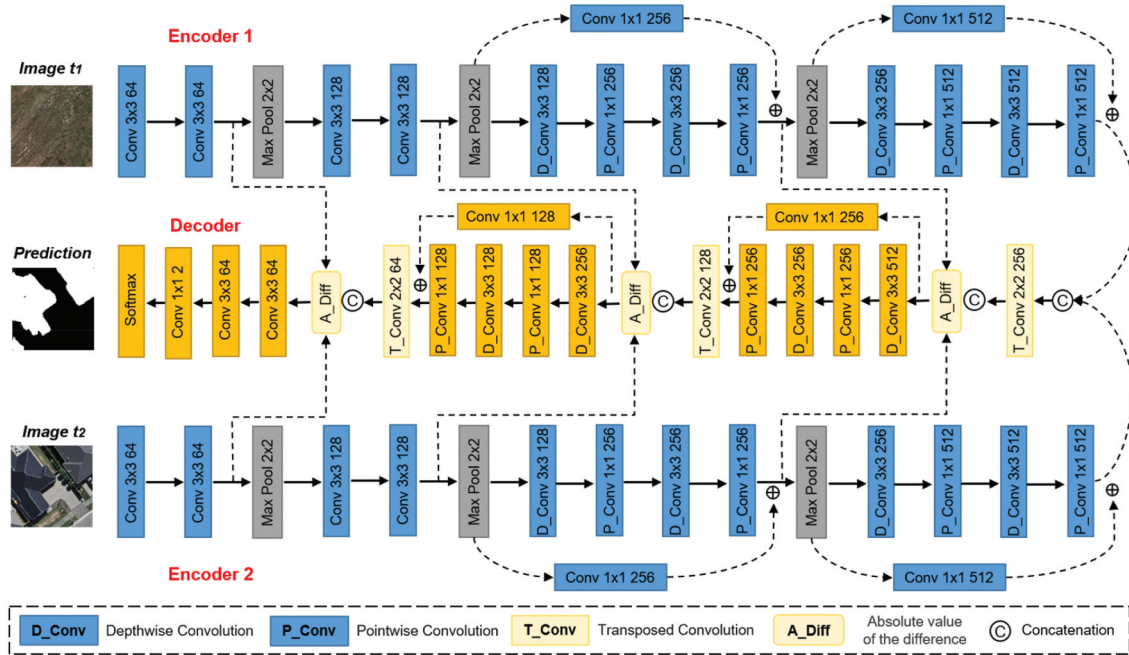


Figure 6. Overall architecture of the proposed LCD-Net.

These shortcut connections improve network performance while avoiding problems such as gradient vanishing and exploding.

The input of the decoder is the combination of features extracted by the fourth convolutional units in different encoders. During the decoding stage, we employ transposed convolution (Yang et al. 2023) to upsample the combined features while reducing the

number of feature channels, aiming to reduce the model parameters required for subsequent operations. The decoder comprises three convolutional units for extracting the difference features. The first two units also are lightweight high-dimensional feature extraction units. These convolutional units are separated by transposed convolution and skip connection. We process the features from different

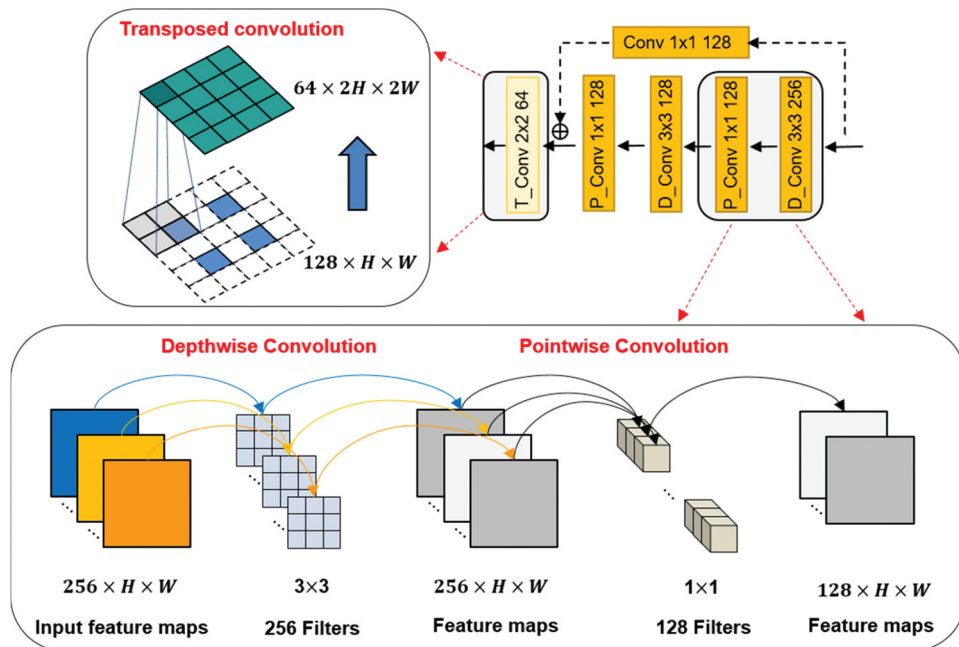


Figure 7. Architecture of the transposed convolution and the lightweight high-dimensional feature extraction unit.

encoders to obtain the absolute value of the differences. Additionally, skip connections combine features with the same scale between the encoders and the decoder. This approach leverages the shallow features from the encoders to enhance the diversity of difference features in the decoder while reducing information loss and improving model performance. In the proposed LCD-Net, except for the output layer, other conventional convolutions and depth separable convolutions are followed by feature batch normalization and ReLU activation to improve model accuracy and enhance convergence ability.

We use depth separable convolutions in the lightweight high-dimensional feature extraction unit to ensure feature dimensionality while reducing model size (Chollet 2017). Depth separable convolution is a two-step process involving depthwise convolution and pointwise convolution. First, depthwise convolution is performed in the two-dimensional plane using the same number of convolutional kernels as the input feature channels. Then, in pointwise convolution, features are weighted and summed over the channel dimension.

4.3. Semi-supervised loss function

During the experiments, the loss function of the Semi-LCD comprises both supervised and unsupervised losses, which can be expressed as:

$$L_{total} = L_{sup} + \lambda L_{unsup} \quad (4)$$

where the L_{sup} and L_{unsup} denote the supervised loss and unsupervised loss, respectively, and λ denotes the weight of unsupervised loss. The supervised loss used to optimize model parameters is the cross-entropy loss L_{ce} , as follows:

$$L_{ce} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^M y_{ic} \log(p_{ic}) \quad (5)$$

where N denotes the number of pixels, M denotes the number of predictable categories. For pixel i , y_{ic} and p_{ic} denote the ground truth label and predicted probability of the c -th category, respectively. The smaller the cross-entropy loss L_{ce} , the closer the predicted results are to the ground truths.

The unsupervised module obtains two results for the same sample before and after perturbation. It utilizes a consistency loss to reduce the difference

between these two results, thereby enhancing the model's generalization. When implementing the consistency constraints using pseudo-labels, the L_{ce} is also used as the unsupervised loss.

In our ablation experiments, the mean absolute error L_{MAE} and the mean square error L_{MSE} , that directly minimize the difference between results are also tested, denoted as:

$$L_{MAE} = \frac{1}{N} \sum_{i=1}^N |p_i - p'_i| \quad (6)$$

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N (p_i - p'_i)^2 \quad (7)$$

where p_i and p'_i are the predicted probabilities of same pixel before and after the perturbation, respectively. When L_{MAE} or L_{MSE} is used, the pixels of unlabeled samples are not filtered by confidence thresholds, and all pixels are involved in the calculation of unsupervised loss.

4.4. Competing methods

To prove the reliability and superior performance of the Semi-LCD, we chose several advanced semantic segmentation and CD methods for comparative experiments, including:

- (1) Random forest (RF) (Breiman 2001) is an ensemble learning algorithm that combines multiple decision trees to enhance prediction accuracy.
- (2) FC-Siam-conc (Daudt, Saux, and Boulch 2018) comprises two weight-sharing encoders that extract features from bi-temporal images. The extracted features from different encoders are concatenated and then connected with the features extracted by the decoder.
- (3) FC-Siam-diff (Daudt, Saux, and Boulch 2018) first calculates the absolute value of the difference between the features extracted by two encoders, then connects this absolute value with the features from the decoder.
- (4) PSPNet (Zhao et al. 2017) leverages atrous convolutions to increase the perceptual field and enable feature extraction. It also incorporates a pyramid pooling module for acquiring multi-scale global features.

- (5) SegNet (Badrinarayanan, Kendall, and Cipolla 2017) is a classic neural network for image segmentation featuring an encoder-decoder structure. During the feature upsampling process, the decoder references corresponding max pooling indices from the encoder.
- (6) U-Net (Ronneberger, Fischer, and Brox 2015) uses skip connections to transfer features from the encoder to the decoder, improving the retention of low-level features in predictions.
- (7) UNet++ (Zhou et al. 2018) without a deep supervision module is obtained by improving U-Net, which introduces dense connections to reduce feature heterogeneity and model training difficulty.
- (8) SNUNet (Fang et al. 2022) introduces two weight-sharing encoders based on UNet++ to extract deep features from different images. An ensemble channel attention module is proposed for fusing features to enhance detection performance.
- (9) s4GAN (Mittal, Tatarchenko, and Brox 2021) is a generative adversarial network for semi-supervised image segmentation. It improves model performance by minimizing the difference between the predicted results and ground truths through feature matching and pseudo-labeling.
- (10) SemiCD (Wele and Patel 2022) builds a complex network containing two weight-sharing encoders (pre-trained ResNet50) and multiple decoders. The auxiliary decoders introduce different perturbations to the feature space of unlabeled samples, further improving model robustness by minimizing the output results of different decoders.

4.5. Performance assessment

Similar to previous studies on CD (Chen et al. 2022; Kalantar et al. 2020; Zhang et al. 2023), the intersection over union (IoU), F1-score (F1), and Kappa coefficient (Kappa) are employed for quantitatively comparing BCD results obtained by different methods. These metrics can be calculated based on the confusion matrix (Foody 2002) as follows:

$$IoU = \frac{TP}{TP + FP + FN} \quad (8)$$

$$P = \frac{TP}{TP + FP} \quad (9)$$

$$R = \frac{TP}{TP + FN} \quad (10)$$

$$F1 = \frac{2PR}{P + R} \quad (11)$$

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

$$PRE = \frac{(TP + FN) \times (TP + FP) + (TN + FP) \times (TN + FN)}{(TP + TN + FP + FN)^2} \quad (13)$$

$$Kappa = \frac{OA - PRE}{1 - PRE} \quad (14)$$

where true positive (TP) and true negative (TN) refer to the number of correctly detected changed and unchanged pixels, respectively, and false negative (FN) and false positive (FP) refer to incorrectly detected changed and unchanged pixels, respectively (Foody 2010). P and R denote the precision and recall of changed pixels, respectively (Sokolova and Lapalme 2009). Overall accuracy (OA) and PRE are intermediate variables for calculating the Kappa coefficient (Foody 2002). The values of IoU, F1 and Kappa indicate how closely the CD results match the ground truths. Generally, higher values for these metrics correspond to more accurate CD results.

4.6. Experimental setup

The CD methods are implemented using the PyTorch framework and trained on a workstation equipped with 12th Intel(R) Core(TM) i9-12900K cores and NVIDIA GeForce RTX 3090 Ti GPU with 24GB memory. To balance the impact of the labeled and unlabeled samples, the weight λ of unsupervised loss is set to 0.5. To reduce unnecessary computational costs and avoid the risk of network overfitting, the models are trained for 60 epochs on all datasets. The order of labeled and unlabeled samples within each epoch is randomly shuffled. The batch size is set to 16 for labeled samples and 32 for unlabeled samples. Notably, when executing SemiCD method on the HRCUS-CD dataset, the

batch sizes of labeled and unlabeled samples are set to 8 and 16, respectively, due to the GPU memory limitation. Although increasing the batch size of unlabeled samples may allow more unlabeled samples to participate in model training, doing so comes at the cost of increased computational and run-time requirements and the risk of insufficient memory. The Adam optimizer with an initial learning rate of 0.0001 and a weight decay of 0.0001 is employed. To facilitate model convergence and maintain stability, the learning rate is multiplied by 0.8 every five epochs. For a fair comparison, all models save the parameters corresponding to the epoch with the lowest loss on the validation set, then evaluate the detection performance on the test set.

5. Results

5.1. Performance comparison on the MtS-WH dataset

To confirm the validity of the Semi-LCD in dealing with a limited number of labeled samples, we first selected the MtS-WH dataset for comparative experiments. Specifically, we trained the models using 50, 200, and 500 labeled samples, respectively. Table 2 presents the quantitative BCD results of various methods on the MtS-WH dataset.

We notice that Semi-LCD achieves the optimal BCD performance under different settings, exhibiting a lower percentage of missed and false changed pixels. The F1 scores obtained by Semi-LCD reach 0.5802, 0.7203, and 0.7461 for 50, 200, and 500 labeled samples, respectively. Among the competing methods, FC-Siam-diff and PSPNet exhibit inadequate performance and lack stability in their detection accuracy.

RF has a performance advantage when the number of labeled samples is 50. However, as the sample size increases, its detection accuracy is surpassed by most competing methods. The semi-supervised methods, s4GAN and SemiCD, outperform some supervised methods, with their accuracy progressively improving with an increasing number of labeled samples. Specifically, the F1 scores of BCD results obtained by SemiCD are 0.2614, 0.5511, and 0.6109 for 50, 200, and 500 labeled samples, respectively.

Precision measures the ratio of correctly detected changed pixels to the total number of changed pixels in the results. In contrast, recall measures the ratio of correctly identified changed pixels to the total number of changed pixels in the ground truths. As seen in Figure 8, influenced by factors like the network structure and model size, there are significant differences in the precision and recall between different methods under the same conditions. Among competing methods, PSPNet and SNUNet demonstrate comparatively lower recall. This suggests their limitation in extracting features of changed pixels from limited labeled samples, consequently leading to more substantial missed detections. FC-Siam-diff exhibits a low precision because it mistakenly detects many unchanged pixels as changed pixels. Conversely, Semi-LCD consistently maintains high precision, and its recall has significantly improved compared to other methods, resulting in better overall detection performance.

Moreover, the qualitative comparison of the BCD results demonstrates that Semi-LCD outperforms all other methods on the MtS-WH dataset. As seen in Figure 9, when utilizing only 50 labeled training samples, all competing methods face challenges in accurately detecting the changed regions. Speckle noises (RF, FC-Siam-conc, FC-Siam-diff, and SNUNet),

Table 2. Quantitative BCD results of different methods on the MtS-WH dataset.

Methods	Number of labeled training samples								
	50			200			500		
	IoU	F1	Kappa	IoU	F1	Kappa	IoU	F1	Kappa
RF	0.3279	0.4939	0.4534	0.3603	0.5297	0.4912	0.3605	0.5300	0.4920
FC-Siam-conc	0.3084	0.4714	0.4097	0.3784	0.5491	0.5162	0.3504	0.5190	0.4876
FC-Siam-diff	0.1810	0.3065	0.1908	0.1310	0.2316	0.1799	0.1393	0.2446	0.2080
PSPNet	0.1481	0.2580	0.2198	0.1236	0.2200	0.1993	0.2985	0.4598	0.4299
SegNet	0.2351	0.3807	0.3463	0.3911	0.5623	0.5291	0.3801	0.5508	0.5179
U-Net	0.3153	0.4794	0.4460	0.5120	0.6773	0.6497	0.5224	0.6863	0.6587
UNet++	0.3457	0.5137	0.4745	0.4217	0.5933	0.5638	0.4527	0.6232	0.5949
SNUNet	0.0866	0.1594	0.0486	0.2663	0.4206	0.3769	0.3043	0.4666	0.4273
s4GAN	0.2034	0.3380	0.2928	0.2958	0.4565	0.4240	0.4084	0.5800	0.5488
SemiCD	0.1504	0.2614	0.2275	0.3804	0.5511	0.5162	0.4398	0.6109	0.5772
Semi-LCD	0.4087	0.5802	0.5375	0.5628	0.7203	0.6942	0.5950	0.7461	0.7231

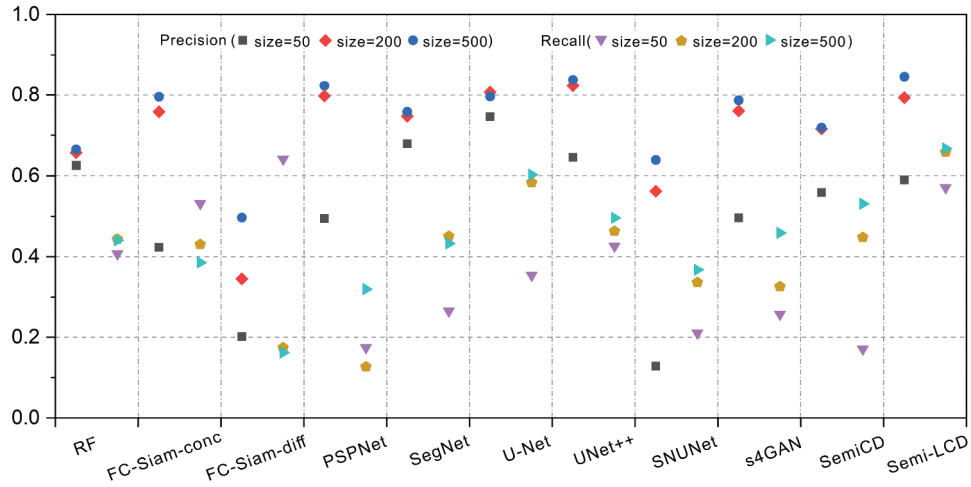


Figure 8. Precision and recall of BCD results obtained by different methods on the MtS-WH dataset. The size represents the number of labeled training samples used in the experiments.

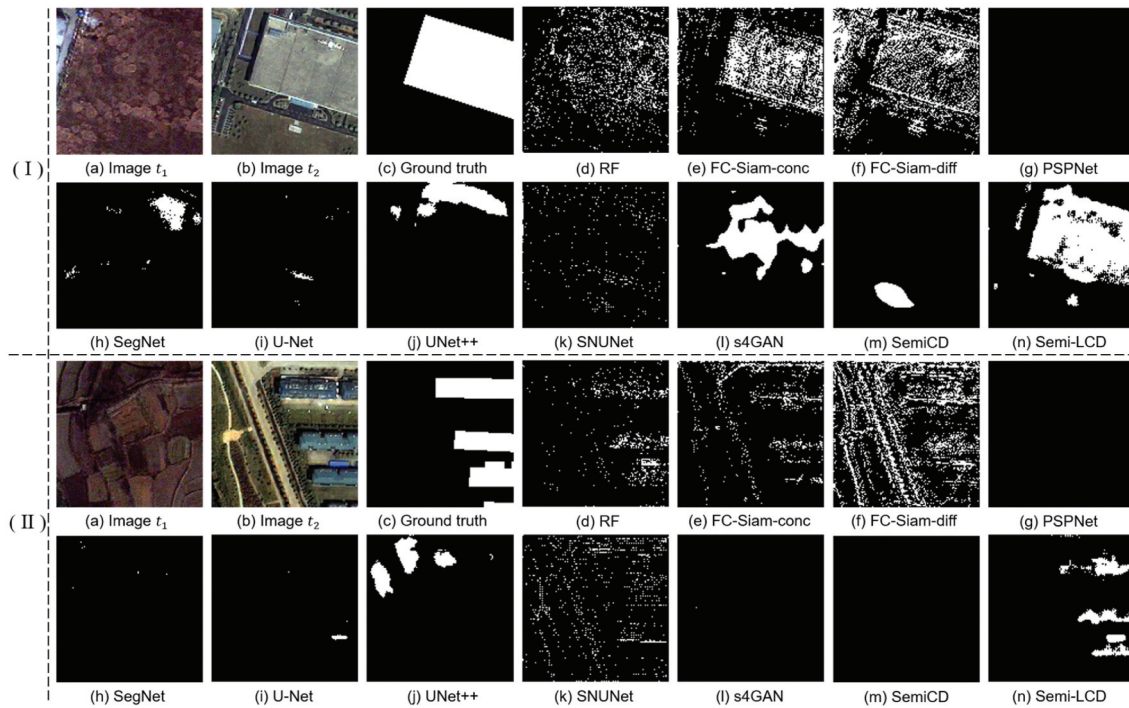


Figure 9. Visualized BCD results of different methods using 50 labeled samples on the MtS-WH dataset. Different image pairs are shown in (I) and (II).

significant omissions (PSPNet, SegNet, and U-Net), and severe false detections (UNet++ and SemiCD) are present in their BCD results. When dealing with small-size objects, s4GAN and SemiCD struggle to identify changed pixels. In contrast, Semi-LCD reliably identifies the locations and contours of the changed buildings even with limited labeled samples, effectively reducing the risk of missed and false detections.

5.2. Performance comparison on the WHU building dataset

To assess the detection capability of the Semi-LCD across diverse scenes, we further selected the WHU Building dataset for experiments and retrained the model parameters accordingly. We conducted experiments using 50, 200, 500, and 1,000 labeled training samples, respectively. The comparison experiments

conducted on the WHU Building dataset yield quantitative BCD results, as presented in Table 3.

According to Table 3, when only 50 labeled samples are available, most methods fail to accurately capture the overall distribution of the dataset, resulting in unsatisfactory BCD results. PSPNet, SegNet, and SemiCD fail to converge, and their models cannot effectively predict building changes. In contrast, Semi-LCD is less affected by the number of labeled samples and achieves higher values in evaluation metrics. The performance of most methods gradually improves with an increasing number of labeled samples. However, FC-Siam-conc and FC-Siam-diff exhibit unstable BCD performance due to their relatively simple network architecture. Under different conditions, Semi-LCD performs more stably and achieves superior BCD results than competing methods. When the number of labeled samples is 1,000, the IoU, F1 scores, and Kappa coefficient of Semi-LCD are 0.6967, 0.8212, and 0.8133, respectively. Among competing methods, U-Net

exhibits the highest accuracy, while Semi-LCD compared to it improves 0.0520, 0.0372, and 0.0389 in the IoU, F1 scores, and Kappa coefficient, respectively.

It is seen from Figure 10 that FC-Siam-conc and FC-Siam-diff exhibit inadequate detection performance, with precision or recall being less than 0.2 in many cases. Although U-Net demonstrates relatively high precision, its recall is low due to missed detections, consequently impacting the comprehensive accuracy of the detection results. Conversely, Semi-LCD displays superior detection performance with significantly higher recall than other methods, striking a better balance between precision and recall.

The qualitative results in Figure 11 also indicate that the Semi-LCD exhibits superior CD performance compared to other methods. Specifically, the CD results obtained by FC-Siam-diff show the lowest consistency with the ground truths, failing to capture the

Table 3. Quantitative BCD results of different methods on the WHU Building dataset.

Methods	Number of labeled training samples											
	50			200			500			1000		
	IoU	F1	Kappa	IoU	F1	Kappa	IoU	F1	Kappa	IoU	F1	Kappa
RF	0.1290	0.2286	0.2106	0.1775	0.3015	0.2847	0.2443	0.3927	0.3759	0.3249	0.4904	0.4722
FC-Siam-conc	0.0238	0.0465	0.0128	0.0004	0.0008	0.0005	0.0627	0.1179	0.1123	0.3950	0.5663	0.5483
FC-Siam-diff	0.0764	0.1419	0.0674	0.0044	0.0087	0.0054	0.0036	0.0071	0.0057	0.1093	0.1971	0.1849
PSPNet	fails to detect changes			0.2785	0.4357	0.4151	0.3501	0.5186	0.4967	0.4795	0.6482	0.6324
SegNet	fails to detect changes			0.2135	0.3519	0.3370	0.2953	0.4559	0.4416	0.4820	0.6505	0.6366
U-Net	0.1407	0.2466	0.2316	0.3474	0.5157	0.4980	0.5271	0.6903	0.6772	0.6447	0.7840	0.7744
UNet++	0.1030	0.1868	0.1714	0.3425	0.5103	0.4937	0.5210	0.6851	0.6724	0.6017	0.7514	0.7401
SNUNet	0.0377	0.0727	0.0237	0.0849	0.1565	0.1306	0.2581	0.4103	0.3824	0.3349	0.5017	0.4758
s4GAN	0.0273	0.0532	0.0474	0.1290	0.2285	0.2163	0.4061	0.5777	0.5621	0.5085	0.6741	0.6604
SemiCD	fails to detect changes			0.0983	0.1790	0.1665	0.3295	0.4957	0.4745	0.5891	0.7414	0.7297
Semi-LCD	0.2950	0.4557	0.4193	0.5035	0.6698	0.6547	0.6034	0.7526	0.7426	0.6967	0.8212	0.8133

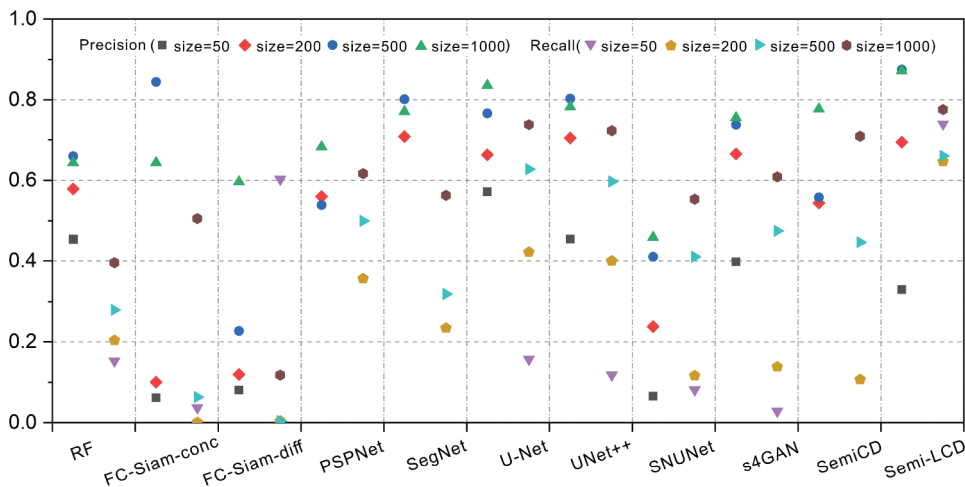


Figure 10. Precision and recall of BCD results obtained by different methods on the WHU Building dataset. The size represents the number of labeled training samples used in the experiments.

locations and contours of the changed buildings accurately. In contrast, FC-Siam-conc, PSPNet, SegNet, U-Net, and UNet++ demonstrate improved detection results, yet they suffer from numerous missed and false detections. Due to the lack of consideration for pixel neighborhood information, RF yields prediction results that are discontinuous in space, appearing as discrete points. On the other hand, s4GAN and SemiCD display relatively good stability in different scenes, but they exhibit lower boundary reducibility for changed buildings. Overall, the proposed Semi-LCD presents great consistency between its detection results and the ground truths.

5.3. Performance comparison on the HRCUS-CD dataset

We also selected the HRCUS-CD dataset, which features complex urban scenes, for further experimentation. For this dataset, we set the number of labeled samples to 300, 500, 1,200, and 2,000, respectively. Quantitative BCD results obtained by different methods on the HRCUS-CD dataset are presented in Table 4.

In complex urban scenes, the proposed Semi-LCD still exhibits the best BCD performance with varying numbers of labeled samples. Specifically, when utilizing 300, 500, 1,200, and 2,000 labeled samples, the F1 scores of Semi-LCD results are 0.4436, 0.4870, 0.5751, and 0.6263, respectively. Since the complex and diverse change scenes in the HRCUS-CD dataset, neither FC-Siam-conc nor FC-Siam-diff obtains satisfactory detection results. Compared to the MtS-WH and WHU Building datasets, the HRCUS-CD dataset encompasses a larger size of samples, and a greater number of labeled samples are used in the experiments. In this context, SemiCD with many model parameters achieves higher detection accuracy than other competing methods.

From Figure 12, it can also be found that the Semi-LCD pays attention to both the precision and recall of the detection results, dramatically reduces the missed detections and false detections, and then reflects higher comprehensive accuracy than other methods.

Figure 13 illustrates the comparison of BCD results obtained from two different scenes. The detection results of FC-Siam-diff with the lowest accuracy are not visualized. In the first scene, several buildings were demolished between two imaging

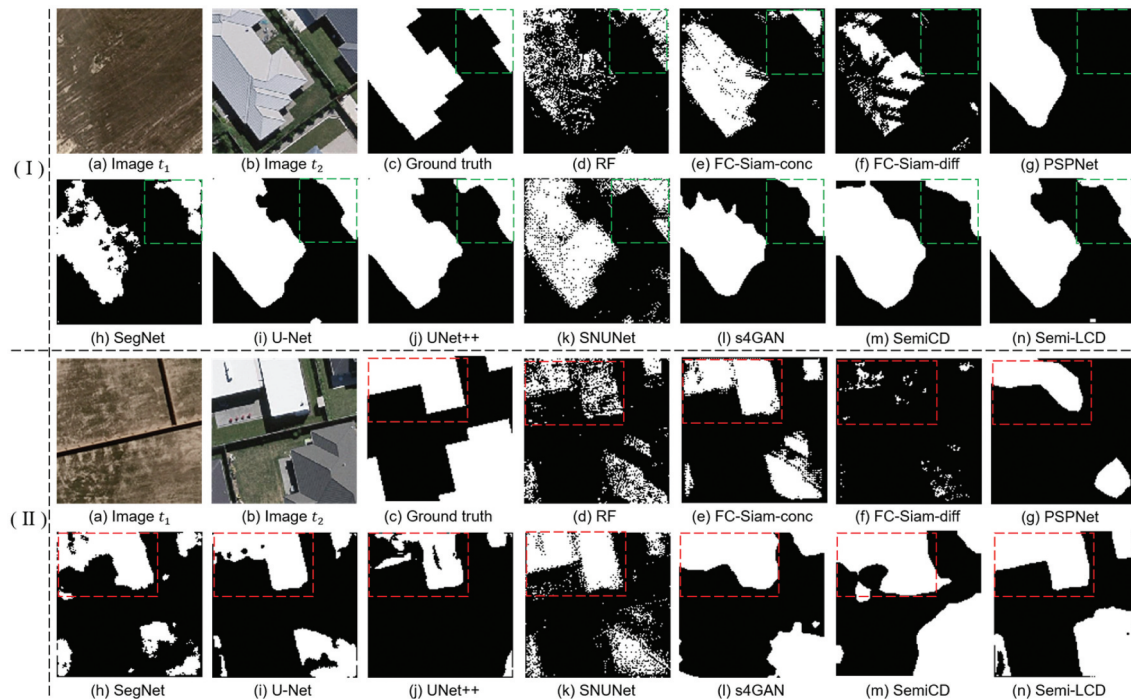
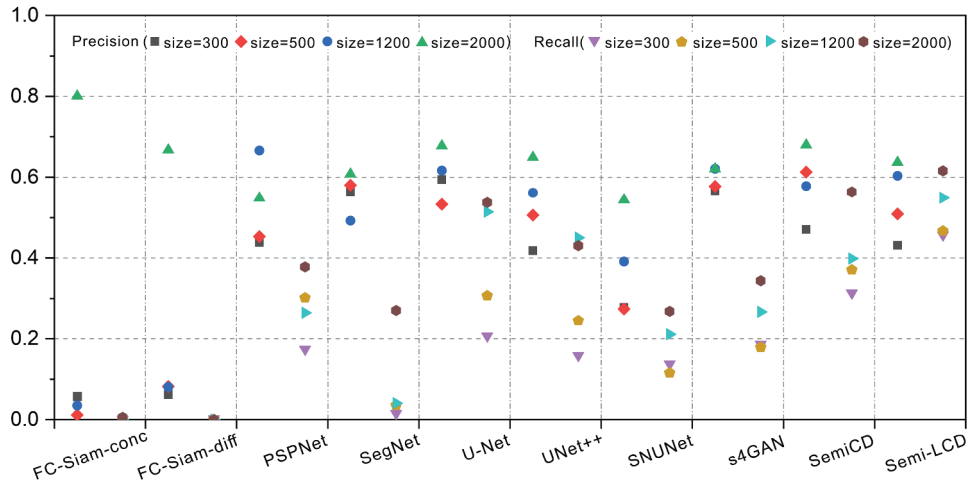


Figure 11. Visualized BCD results of different methods using 1,000 labeled samples on the WHU Building dataset. Different image pairs are shown in (I) and (II). Green and red rectangles represent interesting areas in different image pairs, respectively.

Table 4. Quantitative BCD results of different methods on the HRCUS-CD dataset.

Methods	Number of labeled training samples											
	300			500			1200			2000		
	IoU	F1	Kappa	IoU	F1	Kappa	IoU	F1	Kappa	IoU	F1	Kappa
FC-Siam-conc	0.0004	0.0008	0.0006	0.00001	0.00002	-0.00001	0.00004	0.0001	0.00003	0.0057	0.0112	0.0110
FC-Siam-diff	0.0022	0.0045	0.0031	0.0007	0.0014	0.0011	0.00001	0.00002	0.00001	0.0003	0.0006	0.0006
PSPNet	0.1421	0.2488	0.2407	0.2213	0.3623	0.3526	0.2339	0.3791	0.3724	0.2883	0.4476	0.4390
SegNet	0.0154	0.0304	0.0294	0.0348	0.0672	0.0652	0.0388	0.0747	0.0721	0.2304	0.3745	0.3672
U-Net	0.1810	0.3065	0.2997	0.2419	0.3896	0.3811	0.3897	0.5608	0.5532	0.4283	0.5997	0.5929
UNet++	0.1299	0.2299	0.2219	0.1979	0.3305	0.3222	0.3331	0.4998	0.4913	0.3492	0.5177	0.5103
SNUNet	0.1013	0.1839	0.1737	0.0883	0.1623	0.1529	0.1590	0.2743	0.2646	0.2189	0.3592	0.3512
s4GAN	0.1632	0.2806	0.2738	0.1584	0.2735	0.2670	0.2297	0.3736	0.3665	0.2846	0.4431	0.4355
SemiCD	0.2314	0.3759	0.3664	0.3006	0.4622	0.4545	0.3085	0.4715	0.4633	0.4456	0.6165	0.6098
Semi-LCD	0.2850	0.4436	0.4327	0.3219	0.4870	0.4777	0.4036	0.5751	0.5674	0.4559	0.6263	0.6193

**Figure 12.** Precision and recall of BCD results obtained by different methods on the HRCUS-CD dataset. The size represents the number of labeled training samples used in the experiments.

sessions. Except for U-Net, UNet++, SemiCD, and Semi-LCD, other methods have not effectively detected building changes. Notably, Semi-LCD performs better in identifying demolished buildings, with fewer missed and false detections. In the second scene, densely distributed buildings were constructed between two imaging sessions. FC-Siam-conc fails to detect building changes. Detection results of SNUNet exhibit noticeable speckle noises with unclear boundaries for the changed buildings. PSPNet, SegNet, s4GAN, and SemiCD obtain detection results with severe adhesion between adjacent buildings and unclear boundaries, resulting in poor quality. Although U-Net and UNet++ achieve relatively reliable BCD results, the detected buildings are incomplete and featured significant omissions and misclassifications. In comparison, Semi-LCD effectively distinguishes adjacent changed buildings with clear boundaries, and its results more closely align with the ground truths.

5.4. BCD performance and model parameter

In addition to quantitatively and qualitatively comparing the BCD results, we also selected the number of model parameters as an indicator to reflect the advantages of Semi-LCD over competing methods. Figure 14 illustrates the detection performance of each method and the corresponding number of model parameters in millions (M).

Noticeably, the LCD-Net used in the proposed Semi-LCD has fewer parameters (about 3.62 million). Compared to FC-Siam-conc and FC-Siam-diff with smaller parameters, Semi-LCD has great improvements in BCD performance. SemiCD has the most parameters (about 50.69 million) among all the models. Although U-Net achieves satisfactory BCD results, its parameter number (31.04 million) is 8.6 times higher than that of the LCD-Net. These results indicate that Semi-LCD can leverage unlabeled samples and owns an outstanding balance between performance and model size.

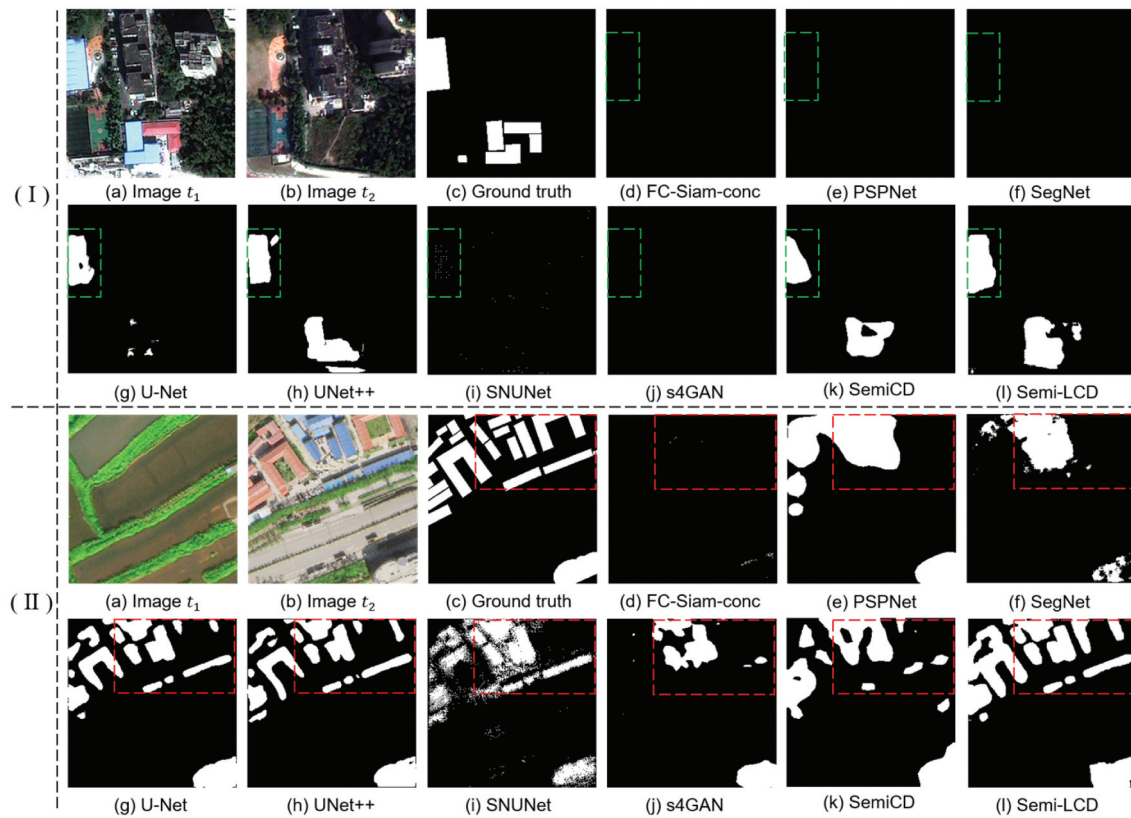


Figure 13. Visualized BCD results of different methods using 2,000 labeled samples on the HRCUS-CD dataset. Different image pairs are shown in (I) and (II). Green and red rectangles represent interesting areas in different image pairs, respectively.

6. Discussion

6.1. Confidence thresholds for selecting pseudo-labeled pixels

The proposed Semi-LCD is built upon the pseudo-labeling principle, which exploits reliable pixels from unlabeled samples to enhance the model's generalization ability by increasing the diversity of training samples. In CD tasks, there is typically an imbalanced proportion between changed and unchanged pixels, as well as varying difficulty levels in detecting pixels from different categories. Hence, we introduced different confidence thresholds for each pixel category to ensure the effective selection of reliable pseudo-labeled pixels. Low thresholds usually result in a large number of pseudo-labeled pixels, but the quality and accuracy of labels are difficult to guarantee. Conversely, high thresholds yield better quality but fewer pseudo-labeled pixels. This study employed a grid search strategy to determine the confidence thresholds t_0 and t_1 for unchanged and changed

pixels, respectively. The thresholds were selected from $\{0.6, 0.7, 0.8, 0.9\}$. Experimental results on different BCD datasets are illustrated in Figure 15.

The confidence thresholds significantly impact the accuracy of the BCD results obtained by Semi-LCD. Generally, a combination of higher t_0 and relatively lower t_1 yields better BCD accuracy. This is primarily due to the following reasons: (1) detecting unchanged pixels is relatively easier, and reliable unchanged pixels in the BCD results usually have high confidence. Setting the confidence threshold t_0 too low may increase uncertainty and errors in the pseudo-labels; (2) detecting changed pixels is comparatively challenging, and the confidence of many changed pixels in the BCD results is slightly higher than the segmentation threshold. Using a relatively low threshold t_1 can increase the number of changed pixels in the pseudo-labels while maintaining accuracy; (3) setting different confidence thresholds for unchanged and changed pixels can alleviate the imbalance between different

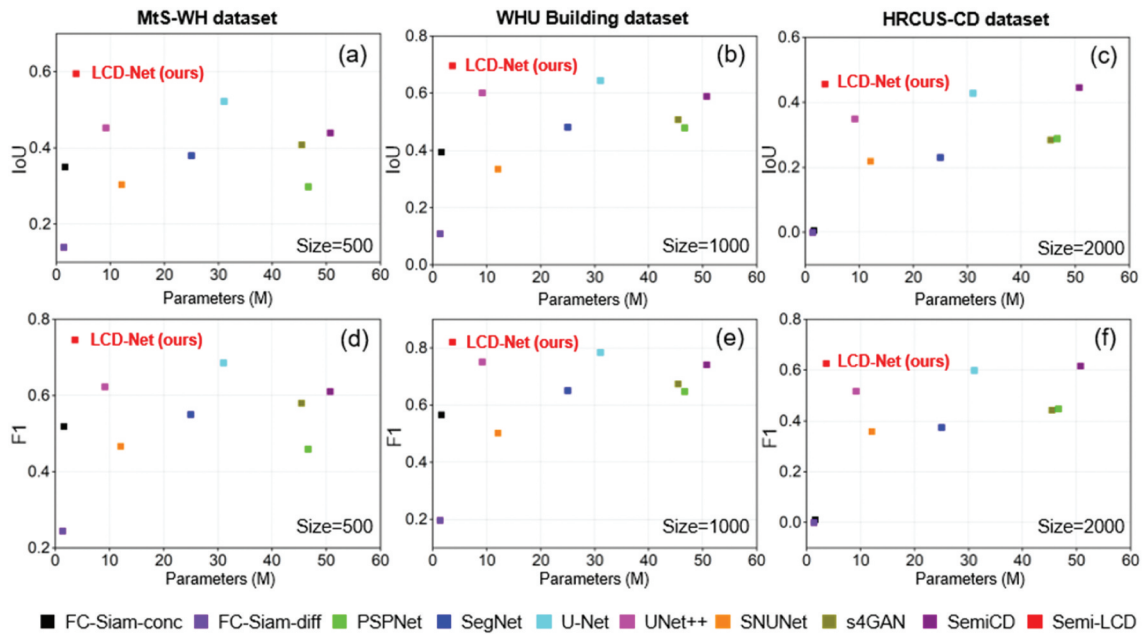


Figure 14. Comparison of performance metric IoU on (a) MtS-WH dataset, (b) WHU Building dataset, (c) HRCUS-CD dataset and model parameter of different methods. Comparison of performance metric F1 score on (d) MtS-WH dataset, (e) WHU Building dataset, (f) HRCUS-CD dataset and model parameter of different methods. The size represents the number of labeled training samples used in the experiments.

pixel categories and improve the model’s convergence speed.

Figure 15 shows that Semi-LCD exhibits superior BCD performance on the MtS-WH and WHU Building datasets when the confidence thresholds t_0 and t_1 are set to 0.8 and 0.6, respectively. On the HRCUS-CD dataset, Semi-LCD achieves better BCD results when the confidence thresholds t_0 and t_1 are set to 0.9 and 0.6, respectively. Despite minor differences in the optimal confidence thresholds among different datasets, the overall trend of the thresholds remains consistent. These observations demonstrate the feasibility and necessity of setting confidence thresholds independently for different pixel categories when the Semi-LCD is deployed.

6.2. Robustness of the semi-supervised BCD method

To assess the robustness of the proposed semi-supervised BCD method, we compared the performance of U-Net both before and after integrating the semi-supervised BCD method on the MtS-WH and WHU Building datasets.

From Figure 16, we notice that integrating the proposed semi-supervised BCD method with U-Net

effectively enhances the accuracy of BCD results on different datasets. By leveraging the principles of consistency regularization and pseudo-labeling, this method effectively expands the training set by selecting pseudo-labeled pixels with high confidence from unlabeled samples. Consequently, there are remarkable improvements in the BCD accuracy and generalization ability. These findings demonstrate the robustness and transferability of the proposed semi-supervised BCD method. In addition, compared to the results in Section 5, it is shown that the proposed LCD-Net achieves relatively better performance than U-Net in the semi-supervised mode.

The comparative results in Figure 17 demonstrate the improved detection performance achieved by integrating the proposed semi-supervised BCD method with U-Net across different datasets. In the semi-supervised mode, Semi-U-Net identifies changed buildings with higher integrity and significantly reduces missed and false detections.

6.3. Advantages and weaknesses of semi-LCD

To validate the advantages of the proposed LCD-Net over other supervised methods, we only utilized labeled samples to optimize model parameters.

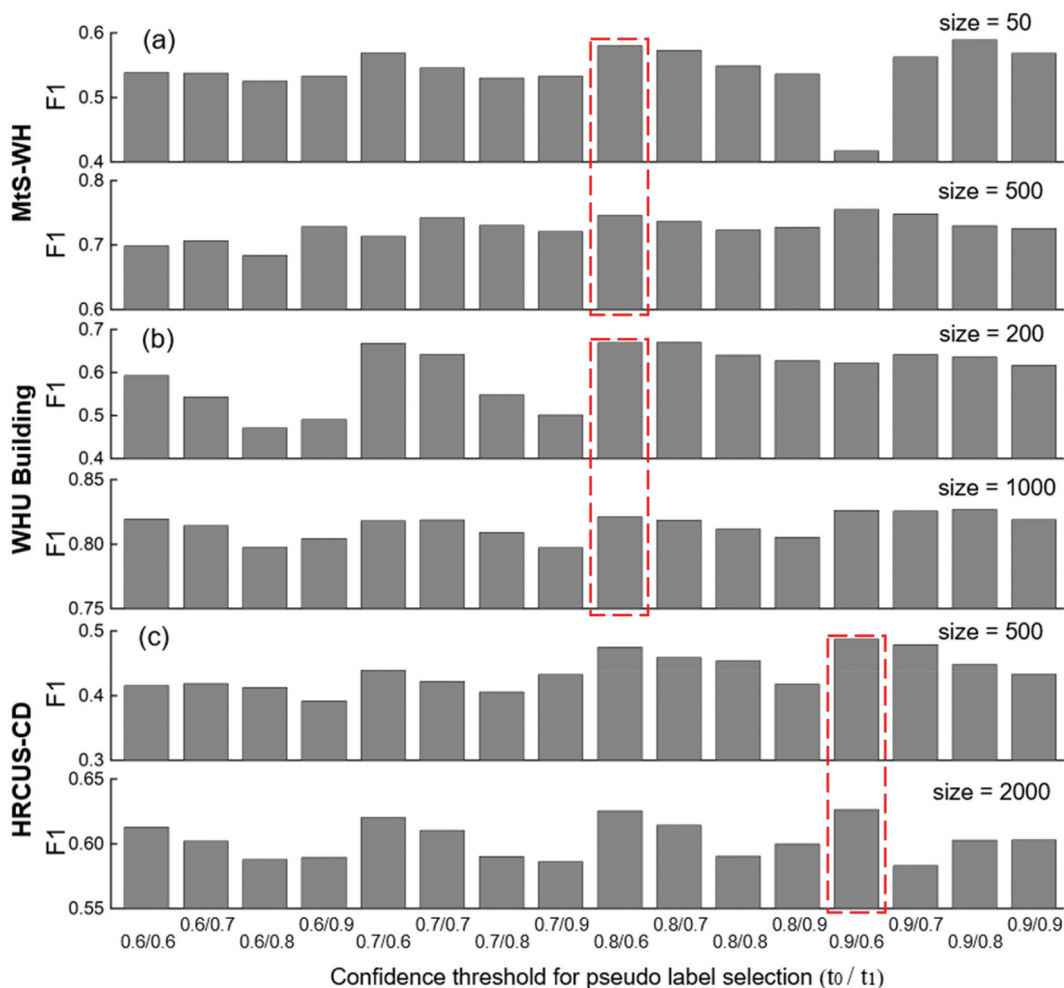


Figure 15. Grid search results of confidence thresholds for pseudo-labeled pixel selection on (a) MtS-WH dataset, (b) WHU Building dataset, (c) HRCUS-CD dataset. The size represents the number of labeled training samples used in the experiments. Red rectangles represent the confidence thresholds corresponding to the optimal CD results on different datasets.

Moreover, sample perturbation is necessary for achieving consistency constraints and subsequently semi-supervised BCD in this study. To validate the role of sample perturbation in the Semi-LCD, we also employed pseudo-labels (PL) to supervise the original predictions of the unlabeled samples. Concurrently, to demonstrate the effectiveness of the Semi-LCD based on the principles of consistent regularization and pseudo-labeling, mean absolute error (MAE) and mean square error (MSE) losses were also assessed to achieve consistency constraints in the semi-supervised mode. The configurations related to these experiments are presented in Table 5.

Specifically, the LCD only utilizes labeled training samples and updates model parameters based on cross-entropy loss. In contrast, other methods utilize both labeled and unlabeled samples for model

training. Excluding the LCD and LCD+PL, the remaining methods incorporate consistency constraints supported by sample perturbation. The Semi-LCD further introduces the pseudo-labeling strategy to facilitate screening high-quality pixels from unlabeled samples, resulting in more accurate and reliable consistency constraints.

In conjunction with the findings presented in Section 5, it is evident that the LCD-Net outperforms competing methods when exclusively utilizing labeled samples for model training (as seen in Tables 6 and 7). By incorporating unlabeled samples into the training set, Semi-LCD exhibits substantial accuracy improvements compared to the fully supervised LCD-Net. This observation highlights the efficacy of the consistency constraints and pseudo-labeling strategies. In addition, the detection accuracy of the Semi-LCD consistently improves with an

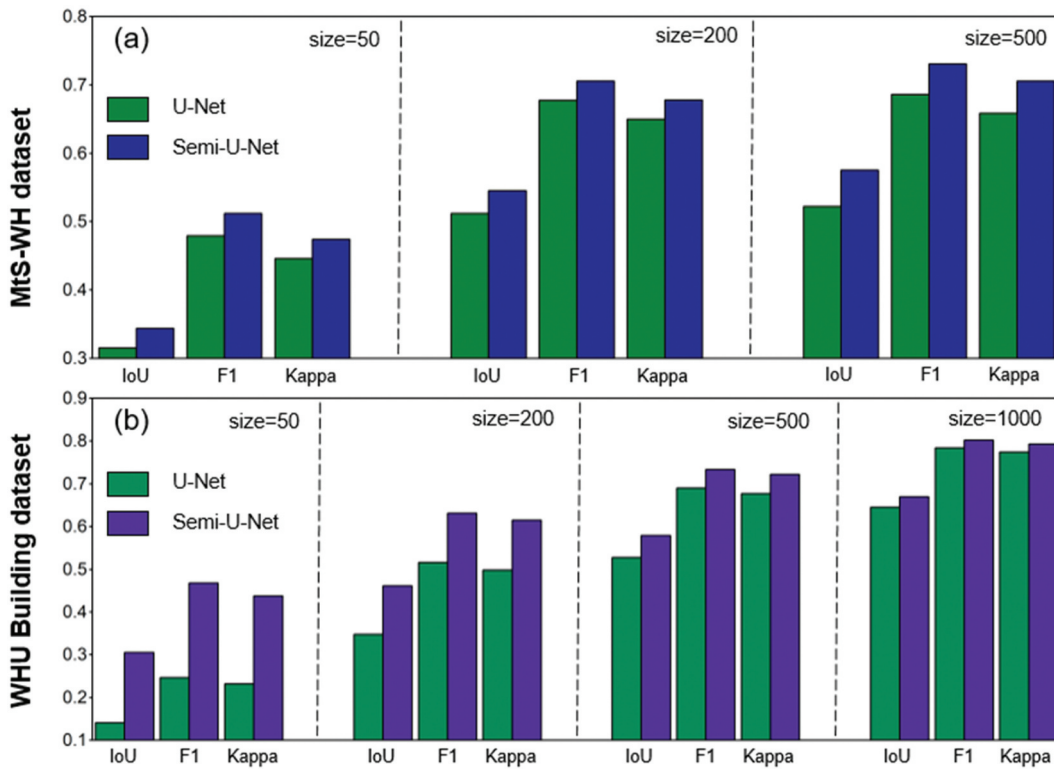


Figure 16. Robustness verification of the proposed semi-supervised BCD method on (a) MtS-WH dataset and (b) WHU Building dataset. The size represents the number of labeled training samples used in the experiments.

increase in the number of labeled samples used on different datasets. These improvements can mainly be attributed to the fact that more labeled samples can provide richer priori knowledge for model training, aiding the convergence of network parameters in a more accurate direction.

Compared to the LCD+PL, the proposed Semi-LCD, which incorporates the consistency constraints supported by sample perturbation, demonstrates performance enhancements in almost all cases. Notably, on the WHU Building dataset, where only 50 labeled samples are available, the fully supervised LCD-Net achieves relatively low detection accuracy, with the F1 only being 0.2823. This result is primarily due to the lack of representative changed pixels in the limited labeled samples. In such a situation, LCD-Net also struggles to extract global features effectively from limited labeled samples in semi-supervised mode, and the model's anti-interference ability is insufficient. Therefore, introducing sample perturbation to achieve consistency constraints has a certain impact on model convergence in semi-supervised mode, resulting in slightly lower performance of Semi-LCD than LCD+PL. In addition to this special case,

combining consistency regularization and pseudo-labeling strategies effectively improves the model's detection accuracy. Next, our study will focus on enhancing the detection accuracy and anti-interference ability of the model in scenarios with severe shortage of labeled samples.

In addition, utilizing MAE or MSE loss as the consistency constraints can also improve the detection accuracy in some cases. However, it is essential to acknowledge that MAE or MSE loss only reduces the variance of different results and lacks the reliability evaluation for pixel-level BCD results. For instance, when many incorrect predictions exist within the unlabeled samples, MAE and MSE losses may cause incorrect model convergence, leading to low accuracy and unstable detection results. In contrast, the proposed Semi-LCD screens reliable pixels from unlabeled samples based on confidence thresholds, which ensures the validity and reliability of the consistency constraints.

From Figure 18, it can be observed that the proposed Semi-LCD can more accurately detect changes in buildings. It exhibits fewer missed and false detections and more regular building boundaries. Although

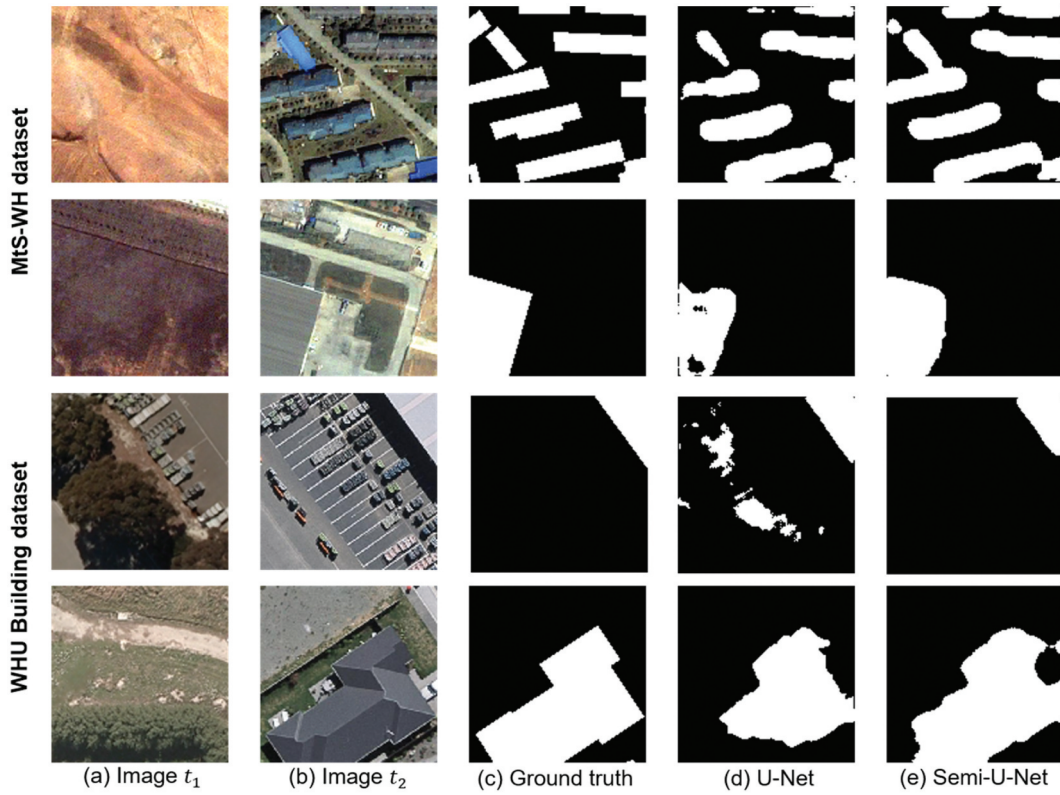


Figure 17. Visualized BCD results obtained using 500 labeled samples on the MtS-WH dataset (rows 1 and 2) and visualized BCD results obtained using 1,000 labeled samples on the WHU Building dataset (rows 3 and 4).

Table 5. Configuration differences between different ablation experiments.

Methods	LCD	LCD+PL	LCD+MAE	LCD+MSE	Semi-LCD
Labeled samples	✓	✓	✓	✓	✓
Supervised loss	Cross-entropy	Cross-entropy	Cross-entropy	Cross-entropy	Cross-entropy
Unlabeled samples	×	✓	✓	✓	✓
Sample perturbation	×	×	✓	✓	✓
Consistency regularization	×	×	✓	✓	✓
Pseudo-labeling	×	✓	×	×	✓
Unsupervised loss	×	Cross-entropy	Mean absolute error	Mean square error	Cross-entropy

Table 6. Quantitative BCD results of ablation experiments on the MtS-WH dataset.

Methods	Number of labeled training samples								
	50			200			500		
	IoU	F1	Kappa	IoU	F1	Kappa	IoU	F1	Kappa
LCD	0.3542	0.5231	0.4913	0.5472	0.7074	0.6814	0.5604	0.7183	0.6931
LCD+PL	0.4019	0.5734	0.5309	0.5516	0.7110	0.6833	0.5858	0.7388	0.7145
LCD+MAE	0.3528	0.5216	0.4888	0.5223	0.6862	0.6599	0.5377	0.6994	0.6742
LCD+MSE	0.3554	0.5245	0.4910	0.5494	0.7092	0.6827	0.5515	0.7110	0.6866
Semi-LCD	0.4087	0.5802	0.5375	0.5628	0.7203	0.6942	0.5950	0.7461	0.7231

Table 7. Quantitative BCD results of ablation experiments on the WHU Building dataset.

Methods	Number of labeled training samples											
	50			200			500			1000		
	IoU	F1	Kappa	IoU	F1	Kappa	IoU	F1	Kappa	IoU	F1	Kappa
LCD	0.1643	0.2823	0.2638	0.4826	0.6510	0.6367	0.5353	0.6973	0.6859	0.6603	0.7954	0.7865
LCD+PL	0.3153	0.4795	0.4458	0.5017	0.6682	0.6529	0.5794	0.7337	0.7226	0.6964	0.8210	0.8129
LCD+MAE	0.1299	0.2300	0.2149	0.3498	0.5183	0.5020	0.5476	0.7077	0.6967	0.6904	0.8169	0.8090
LCD+MSE	0.1741	0.2966	0.2804	0.3746	0.5450	0.5291	0.5655	0.7225	0.7112	0.6765	0.8071	0.7988
Semi-LCD	0.2950	0.4557	0.4193	0.5035	0.6698	0.6547	0.6034	0.7526	0.7426	0.6967	0.8212	0.8133

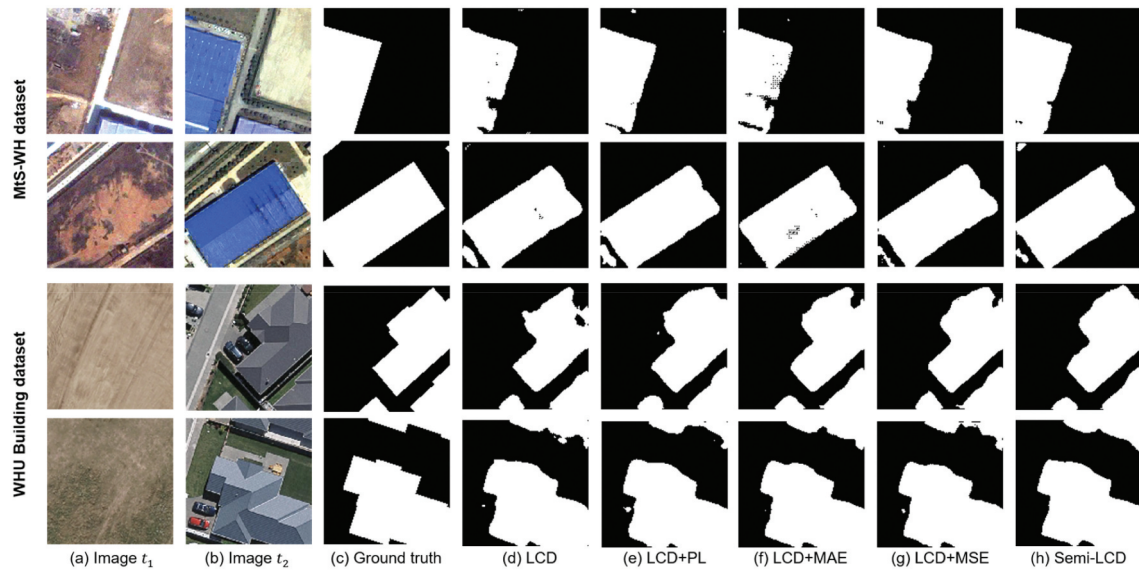


Figure 18. Visualized BCD results obtained using 500 labeled samples on the MtS-WH dataset (rows 1 and 2) and visualized BCD results obtained using 1,000 labeled samples on the WHU Building dataset (rows 3 and 4).

the Semi-LCD achieves better detection results on different datasets, its design primarily emphasizes conciseness and effectiveness. Future study will further explore more effective perturbation methods to improve CD accuracy. Additionally, we will study the relationships between different perturbation methods, perturbation intensities, and the number of labeled samples, aiming to gain further insights into their interactions.

7. Conclusions

To address the issues of convergence difficulty and low detection accuracy in remote sensing-based binary BCD tasks with limited labeled samples, we proposed Semi-LCD based on data augmentation, consistency regularization, and pseudo-labeling. To enhance the applicability of Semi-LCD, we proposed a lightweight dual-branch CD network. The lightweight high-dimensional feature extraction unit is specifically designed to effectively extract image features while minimizing the model size. Additionally, Semi-LCD employs different confidence thresholds for various pixels, effectively screening pseudo-labeled pixels from unlabeled samples. Further, the consistency constraints are applied to reduce the differences in the BCD results of unlabeled samples before and after perturbation, improving the

detection accuracy and generalization capability of the model.

Experimental results confirm the superior performance and practicality of the proposed Semi-LCD, indicating its better balance between model size and performance. Furthermore, multiple confidence thresholds facilitate fully screening high-quality pseudo-labeled pixels, thus enhancing binary BCD performance. Integrating the proposed method with classical neural network also improve BCD accuracy, proving its robustness and adaptability. Consistency constraints based on the pseudo-labeling can fully exploit unlabeled samples and demonstrates better performance and stability than that based on MAE and MSE losses. In future studies, we intend to optimize the network performance and explore the impact on detection results of varying perturbation methods, proportion, and intensities.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the National Natural Science Foundation of China [42090012]; 03 special research and 5G project of Jiangxi Province in China [20212ABC03A09]; Hubei

key R&D plan [2022BAA048]; Key R&D project of Sichuan science and technology plan [2022YFN0031]; Zhuhai industry university research cooperation project of China [ZH22017001210098PWC].

CRediT authorship contribution statement

Qing Ding: Methodology, Software, Writing – original draft. Zhenfeng Shao: Methodology, Data curation, Visualization. Xiao Huang: Writing – review & editing. Xiaoxiao Feng: Validation, Visualization. Orhan Altan: Validation, Conceptualization. Bin Hu: Data curation.

Data availability statement

The MtS-WH dataset is available at http://sigma.whu.edu.cn/newspage.php?q=2019_03_26_ENG; the WHU Building dataset is available at http://gpcv.whu.edu.cn/data/building_data_set.html; the HRCUS-CD dataset is available from the corresponding author upon reasonable request.

References

- Abdi, G., and S. Jabari. 2021. "A Multi-Feature Fusion Using Deep Transfer Learning for Earthquake Building Damage Detection." *Canadian Journal of Remote Sensing* 47 (2): 337–352. <https://doi.org/10.1080/07038992.2021.1925530>.
- Badrinarayanan, V., A. Kendall, and R. Cipolla. 2017. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (12): 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>.
- Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Bruzzone, L., and D. Prieto. 2000. "Automatic Analysis of the Difference Image for Unsupervised Change Detection." *IEEE Transactions on Geoscience and Remote Sensing* 38 (3): 1171–1182. <https://doi.org/10.1109/36.843009>.
- Buslaev, A., V. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. Kalinin. 2020. "Albumentations: Fast and Flexible Image Augmentations." *Information* 11 (2): 125. <https://doi.org/10.3390/info11020125>.
- Cao, Y., and X. Huang. 2023. "A Full-Level Fused Cross-Task Transfer Learning Method for Building Change Detection Using Noise-Robust Pretrained Networks on Crowdsourced Labels." *Remote Sensing of Environment* 284:113371. <https://doi.org/10.1016/j.rse.2022.113371>.
- Celik, T. 2009. "Unsupervised Change Detection in Satellite Images Using Principal Component Analysis and K-Means Clustering." *IEEE Geoscience and Remote Sensing Letters* 6 (4): 772–776. <https://doi.org/10.1109/LGRS.2009.2025059>.
- Chen, J., B. Sun, L. Wang, B. Fang, Y. Chang, Y. Li, J. Zhang, X. Lyu, and G. Chen. 2022. "Semi-Supervised Semantic Segmentation Framework with Pseudo Supervisions for Land-Use/Land-Cover Mapping in Coastal Areas." *International Journal of Applied Earth Observation and Geoinformation* 112:102881. <https://doi.org/10.1016/j.jag.2022.102881>.
- Chen, H., C. Wu, B. Du, L. Zhang, and L. Wang. 2020. "Change Detection in Multisource VHR Images via Deep Siamese Convolutional Multiple-Layers Recurrent Neural Network." *IEEE Transactions on Geoscience and Remote Sensing* 58 (4): 2848–2864. <https://doi.org/10.1109/TGRS.2019.2956756>.
- Chen, Z., Y. Zhou, B. Wang, X. Xu, N. He, S. Jin, and S. Jin. 2022. "EGDE-Net: A Building Change Detection Method for High-Resolution Remote Sensing Imagery Based on Edge Guidance and Differential Enhancement." *ISPRS Journal of Photogrammetry and Remote Sensing* 191:203–222. <https://doi.org/10.1016/j.isprsjprs.2022.07.016>.
- Chollet, F. 2017. "Xception: Deep Learning with Depthwise Separable Convolutions." 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, United states, 1800–1807.
- Daudt, R. C., B. L. Saux, and A. Boulch. 2018. "Fully Convolutional Siamese Networks for Change Detection." 25th IEEE International Conference on Image Processing (ICIP 2018), Athens, Greece, 4063–4067.
- Deng, Y., Y. Meng, J. Chen, A. Yue, D. Liu, and J. Chen. 2023. "TChange: A Hybrid Transformer-CNN Change Detection Network." *Remote Sensing* 15 (5): 1219. <https://doi.org/10.3390/rs15051219>.
- Ding, Q., Z. Shao, X. Huang, and O. Altan. 2021. "DSA-Net: A Novel Deeply Supervised Attention-Guided Network for Building Change Detection in High-Resolution Remote Sensing Images." *International Journal of Applied Earth Observation and Geoinformation* 105:102591. <https://doi.org/10.1016/j.jag.2021.102591>.
- Ding, Q., Z. Shao, X. Huang, O. Altan, and B. Hu. 2022. "Time-Series Land Cover Mapping and Urban Expansion Analysis Using OpenStreetmap Data and Remote Sensing Big Data: A Case Study of Guangdong-Hong Kong-Macao Greater Bay Area, China." *International Journal of Applied Earth Observation and Geoinformation* 113:113001. <https://doi.org/10.1016/j.jag.2022.103001>.
- Fang, S., K. Li, J. Shao, and Z. Li. 2022. "SNUNet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images." *IEEE Geoscience and Remote Sensing Letters* 19:1–5. <https://doi.org/10.1109/LGRS.2021.3056416>.
- Fan, Y., A. Kukleva, D. Dai, and B. Schiele. 2023. "Revisiting Consistency Regularization for Semi-Supervised Learning." *International Journal of Computer Vision* 131 (3): 626–643. <https://doi.org/10.1007/s11263-022-01723-4>.
- Foody, G. 2002. "Status of Land Cover Classification Accuracy Assessment." *Remote Sensing of Environment* 80 (1): 185–201. [https://doi.org/10.1016/S0034-4257\(01\)00295-4](https://doi.org/10.1016/S0034-4257(01)00295-4).
- Foody, G. 2010. "Assessing the Accuracy of Land Cover Change with Imperfect Ground Reference Data." *Remote Sensing of Environment* 114 (10): 2271–2285. <https://doi.org/10.1016/j.rse.2010.05.003>.
- Gong, M., T. Zhan, P. Zhang, and Q. Miao. 2017. "Superpixel-Based Difference Representation Learning for Change Detection in Multispectral Remote Sensing Images." *IEEE*

- Transactions on Geoscience and Remote Sensing* 55 (5): 2658–2673. <https://doi.org/10.1109/TGRS.2017.2650198>.
- Guo, Z., and S. Du. 2017. "Mining Parameter Information for Building Extraction and Change Detection with Very High-Resolution Imagery and GIS Data." *GIScience and Remote Sensing* 54 (1): 38–63. <https://doi.org/10.1080/15481603.2016.1250328>.
- Habib, T., J. Inglada, G. Mercier, and J. Chanussot. 2009. "Support Vector Reduction in SVM Algorithm for Abrupt Change Detection in Remote Sensing." *IEEE Geoscience and Remote Sensing Letters* 6 (3): 606–610. <https://doi.org/10.1109/LGRS.2009.2020306>.
- Han, T., M. Wulder, J. White, N. Coops, M. Alvarez, and C. Butson. 2007. "An Efficient Protocol to Process Landsat Images for Change Detection with Tasselled Cap Transformation." *IEEE Geoscience and Remote Sensing Letters* 4 (1): 147–151. <https://doi.org/10.1109/LGRS.2006.887066>.
- Hou, B., Q. Liu, H. Wang, and Y. Wang. 2020. "From W-Net to CDGAN: Bitemporal Change Detection via Deep Learning Techniques." *IEEE Transactions on Geoscience and Remote Sensing* 58 (3): 1790–1802. <https://doi.org/10.1109/TGRS.2019.2948659>.
- Hou, B., Y. Wang, and Q. Liu. 2017. "Change Detection Based on Deep Features and Low Rank." *IEEE Geoscience and Remote Sensing Letters* 14 (12): 2418–2422. <https://doi.org/10.1109/LGRS.2017.2766840>.
- Huang, D., Y. Tang, and R. Qin. 2022. "An Evaluation of PlanetScope Images for 3D Reconstruction and Change Detection - Experimental Validations with Case Studies." *GIScience and Remote Sensing* 59 (1): 744–761. <https://doi.org/10.1080/15481603.2022.2060595>.
- Hung, W., Y. Tsai, Y. Liou, Y. Lin, and M. Yang. 2019. "Adversarial Learning for Semi-Supervised Semantic Segmentation." 29th British Machine Vision Conference (BMVC 2018), Newcastle, United Kingdom.
- Ji, S., S. Wei, and M. Lu. 2019. "Fully Convolutional Networks for Multisource Building Extraction from an Open Aerial and Satellite Imagery Data Set." *IEEE Transactions on Geoscience and Remote Sensing* 57 (1): 574–586. <https://doi.org/10.1109/TGRS.2018.2858817>.
- Kalantar, B., N. Ueda, H. Al-Najjar, and A. Halin. 2020. "Assessment of Convolutional Neural Network Architectures for Earthquake-Induced Building Damage Detection Based on Pre-And Post-Event Orthophoto Images." *Remote Sensing* 12 (21): 1–20. <https://doi.org/10.3390/rs12213529>.
- Lei, Y., X. Liu, J. Shi, C. Lei, and J. Wang. 2019. "Multiscale Superpixel Segmentation with Deep Features for Change Detection." *IEEE Access* 7:36600–36616. <https://doi.org/10.1109/ACCESS.2019.2902613>.
- Li, J., X. Huang, L. Tu, T. Zhang, and L. Wang. 2022. "A Review of Building Detection from Very High Resolution Optical Remote Sensing Images." *GIScience and Remote Sensing* 59 (1): 1199–1225. <https://doi.org/10.1080/15481603.2022.2101727>.
- Li, M., J. Im, and C. Beier. 2013. "Machine Learning Approaches for Forest Classification and Change Analysis Using Multi-Temporal Landsat TM Images Over Huntington Wildlife Forest." *GIScience and Remote Sensing* 50 (4): 361–384. <https://doi.org/10.1080/15481603.2013.819161>.
- Li, Q., Y. Shi, and X. Zhu. 2022. "Semi-Supervised Building Footprint Generation with Feature and Output Consistency Training." *IEEE Transactions on Geoscience and Remote Sensing* 60:1–17. <https://doi.org/10.1109/TGRS.2022.3174636>.
- Liu, Y., W. Xin, and Z. Chen. 2014. "Remote Sensing Change Detection Study Based on Adaptive Threshold in Pixel Ratio Method." Land Surface Remote Sensing II, Beijing, China, 9260.
- Long, J., E. Shelhamer, and T. Darrell. 2015. "Fully Convolutional Networks for Semantic Segmentation." IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), Boston, MA, United States, 431–440.
- Mahmoodzadeh, H. 2007. "Digital Change Detection Using Remotely Sensed Data for Monitoring Green Space Destruction in Tabriz." *International Journal of Environmental Research* 1 (1): 35–41.
- Mittal, S., M. Tatarchenko, and T. Brox. 2021. "Semi-Supervised Semantic Segmentation with High- and Low-Level Consistency." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43 (4): 1369–1379. <https://doi.org/10.1109/TPAMI.2019.2960224>.
- Nielsen, A., K. Conradsen, and J. Simpson. 1998. "Multivariate Alteration Detection (MAD) and MAF Postprocessing in Multispectral, Bitemporal Image Data: New Approaches to Change Detection Studies." *Remote Sensing of Environment* 64 (1): 1–19. [https://doi.org/10.1016/S0034-4257\(97\)00162-4](https://doi.org/10.1016/S0034-4257(97)00162-4).
- Ouali, Y., C. Hudelot, and M. Tami. 2020. "Semi-Supervised Semantic Segmentation with Cross-Consistency Training." 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020), Virtual, Online, United States, 12671–12681.
- Park, S., and A. Song. 2023. "Hybrid Approach Using Deep Learning and Graph Comparison for Building Change Detection." *GIScience and Remote Sensing* 60 (1). <https://doi.org/10.1080/15481603.2023.2220525>.
- Peng, D., L. Bruzzone, Y. Zhang, H. Guan, H. Ding, and X. Huang. 2021. "SemiCdnet: A Semisupervised Convolutional Neural Network for Change Detection in High Resolution Remote-Sensing Images." *IEEE Transactions on Geoscience and Remote Sensing* 59 (7): 5891–5906. <https://doi.org/10.1109/TGRS.2020.3011913>.
- Peng, D., Y. Zhang, and H. Guan. 2019. "End-To-End Change Detection for High Resolution Satellite Images Using Improved Unet++." *Remote Sensing* 11 (11): 1382. <https://doi.org/10.3390/rs11111382>.
- Rokni, K., A. Ahmad, K. Solaimani, and S. Hazini. 2015. "A New Approach for Surface Water Change Detection: Integration of Pixel Level Image Fusion and Image Classification Techniques." *International Journal of Applied Earth*

- Observation and Geoinformation* 34 (1): 226–234. <https://doi.org/10.1016/j.jag.2014.08.014>.
- Ronneberger, O., P. Fischer, and T. Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), Munich, Germany, 234–241.
- Saha, S., F. Bovolo, and L. Bruzzone. 2019. "Unsupervised Deep Change Vector Analysis for Multiple-Change Detection in VHR Images." *IEEE Transactions on Geoscience and Remote Sensing* 57 (6): 3677–3693. <https://doi.org/10.1109/TGRS.2018.2886643>.
- Sandler, M., A. Howard, M. Zhu, A. Zhmoginov, and L. Chen. 2018. "MobileNetV2: Inverted Residuals and Linear Bottlenecks." 31th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018), Salt Lake City, UT, United states, 4510–4520.
- Shafique, A., G. Cao, Z. Khan, M. Asad, and M. Aslam. 2022. "Deep Learning-Based Change Detection in Remote Sensing Images: A Review." *Remote Sensing* 14 (4): 871. <https://doi.org/10.3390/rs14040871>.
- Shu, Q., J. Pan, Z. Zhang, and M. Wang. 2022. "MTCNet: Multitask Consistency Network with Single Temporal Supervision for Semi-Supervised Building Change Detection." *International Journal of Applied Earth Observation and Geoinformation* 115:103110. <https://doi.org/10.1016/j.jag.2022.103110>.
- Sokolova, M., and G. Lapalme. 2009. "A Systematic Analysis of Performance Measures for Classification Tasks." *Information Processing & Management* 45 (4): 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>.
- Sun, C., J. Wu, H. Chen, and C. Du. 2022. "SemiSanet: A Semi-Supervised High-Resolution Remote Sensing Image Change Detection Model Using Siamese Networks with Graph Attention." *Remote Sensing* 14 (12): 2801. <https://doi.org/10.3390/rs14122801>.
- Tan, M., and Q. Le. 2019. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." 36th International Conference on Machine Learning (ICML 2019), Long Beach, CA, United states, 10691–10700.
- Van Engelen, J., and H. Hoos. 2020. "A Survey on Semi-Supervised Learning." *Machine Learning* 109 (2): 373–440. <https://doi.org/10.1007/s10994-019-05855-6>.
- Wang, B., P. Jiang, J. Gao, W. Huo, Z. Yang, and Y. Liao. 2023. "A Lightweight Few-Shot Marine Object Detection Network for Unmanned Surface Vehicles." *Ocean Engineering* 277:114329. <https://doi.org/10.1016/j.oceaneng.2023.114329>.
- Wang, P., and E. Sertel. 2023. "Multi-Frame Super-Resolution of Remote Sensing Images Using Attention-Based GAN Models." *Knowledge-Based Systems* 266:110387. <https://doi.org/10.1016/j.knosys.2023.110387>.
- Wang, X., Q. Zhao, P. Jiang, Y. Zheng, L. Yuan, and P. Yuan. 2022. "LDS-YOLO: A Lightweight Small Object Detection Method for Dead Trees from Shelter Forest." *Computers and Electronics in Agriculture* 198:107035. <https://doi.org/10.1016/j.compag.2022.107035>.
- Wele, G., and V. Patel. 2022. "Revisiting Consistency Regularization for Semi-Supervised Change Detection in Remote Sensing Images." *arXiv: 2204.08454*.
- Wieland, M., S. Martinis, R. Kiefl, and V. Gstaiger. 2023. "Semantic Segmentation of Water Bodies in Very High-Resolution Satellite and Aerial Images." *Remote Sensing of Environment* 287:113452. <https://doi.org/10.1016/j.rse.2023.113452>.
- Wu, C., L. Zhang, and B. Du. 2017. "Kernel Slow Feature Analysis for Scene Change Detection." *IEEE Transactions on Geoscience and Remote Sensing* 55 (4): 2367–2384. <https://doi.org/10.1109/TGRS.2016.2642125>.
- Xiao, P., G. Sheng, X. Zhang, H. Liu, and R. Guo. 2021. "Direction-Dominated Change Vector Analysis for Forest Change Detection." *International Journal of Applied Earth Observation and Geoinformation* 103:102492. <https://doi.org/10.1016/j.jag.2021.102492>.
- Yang, H., Y. Chen, W. Wu, S. Pu, X. Wu, Q. Wan, and W. Dong. 2023. "A Lightweight Siamese Neural Network for Building Change Detection Using Remote Sensing Images." *Remote Sensing* 15 (4): 928. <https://doi.org/10.3390/rs15040928>.
- Yang, R., S. Wang, X. Wu, T. Liu, and X. Liu. 2022. "Using Lightweight Convolutional Neural Network to Track Vibration Displacement in Rotating Body Video." *Mechanical Systems and Signal Processing* 177:109137. <https://doi.org/10.1016/j.ymssp.2022.109137>.
- Ye, W., J. Lao, Y. Liu, C. Chang, Z. Zhang, H. Li, and H. Zhou. 2022. "Pine Pest Detection Using Remote Sensing Satellite Images Combined with a Multi-Scale Attention-UNet Model." *Ecological Informatics* 72:101906. <https://doi.org/10.1016/j.ecoinf.2022.101906>.
- Yin, M., S. He, T. Soomro, and H. Yuan. 2023. "Efficient Skeleton-Based Action Recognition via Multi-Stream Depthwise Separable Convolutional Neural Network." *Expert Systems with Applications* 226:120080. <https://doi.org/10.1016/j.eswa.2023.120080>.
- Zhang, P., M. Gong, L. Su, J. Liu, and Z. Li. 2016. "Change Detection Based on Deep Feature Representation and Mapping Transformation for Multi-Spatial-Resolution Remote Sensing Images." *ISPRS Journal of Photogrammetry and Remote Sensing* 116:24–41. <https://doi.org/10.1016/j.isprsjprs.2016.02.013>.
- Zhang, H., G. Ma, and Y. Zhang. 2022. "Intelligent-BCD: A Novel Knowledge-Transfer Building Change Detection Framework for High-Resolution Remote Sensing Imagery." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15:5065–5075. <https://doi.org/10.1109/JSTARS.2022.3184298>.
- Zhang, H., G. Ma, Y. Zhang, B. Wang, H. Li, and L. Fan. 2023. "MCHA-Net: A Multi-End Composite Higher-Order Attention Network Guided with Hierarchical Supervised Signal for High-Resolution Remote Sensing Image Change Detection." *ISPRS Journal of Photogrammetry and Remote Sensing* 202:40–68. <https://doi.org/10.1016/j.isprsjprs.2023.05.033>.
- Zhang, J., Z. Shao, Q. Ding, X. Huang, Y. Wang, X. Zhou, and D. Li. 2023. "AERNet: An Attention-Guided Edge Refinement Network and a Dataset for Remote Sensing Building Change Detection."

- IEEE Transactions on Geoscience and Remote Sensing* 61:1–16. <https://doi.org/10.1109/TGRS.2023.3308936>.
- Zhang, C., L. Wang, S. Cheng, and Y. Li. 2022. "SwinSunet: Pure Transformer Network for Remote Sensing Image Change Detection." *IEEE Transactions on Geoscience and Remote Sensing* 60:1–13. <https://doi.org/10.1109/TGRS.2022.3160007>.
- Zhang, Y., F. Xie, X. Song, H. Zhou, Y. Yang, H. Zhang, and J. Liu. 2022. "A Rotation Meanout Network with Invariance for Dermoscopy Image Classification and Retrieval." *Computers in Biology and Medicine* 151 (Pt A): 106272. <https://doi.org/10.1016/j.combiomed.2022.106272>.
- Zhang, C., P. Yue, D. Tapete, L. Jiang, B. Shangguan, L. Huang, and G. Liu. 2020. "A Deeply Supervised Image Fusion Network for Change Detection in High Resolution Bi-Temporal Remote Sensing Images." *ISPRS Journal of Photogrammetry and Remote Sensing* 166:183–200. <https://doi.org/10.1016/j.isprsjprs.2020.06.003>.
- Zhang, R., H. Zhang, X. Ning, X. Huang, J. Wang, and W. Cui. 2023. "Global-Aware Siamese Network for Change Detection on Remote Sensing Images." *ISPRS Journal of Photogrammetry and Remote Sensing* 199:61–72. <https://doi.org/10.1016/j.isprsjprs.2023.04.001>.
- Zhang, X., X. Zhou, M. Lin, and J. Sun. 2018. "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices." 31th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018), Salt Lake City, UT, United states, 6848–6856.
- Zhao, H., J. Shi, X. Qi, X. Wang, and J. Jia. 2017. "Pyramid Scene Parsing Network." 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, United states, 6230–6239.
- Zheng, Z., Y. Wan, Y. Zhang, S. Xiang, D. Peng, and B. Zhang. 2021. "CLNet: Cross-Layer Convolutional Neural Network for Change Detection in Optical Remote Sensing Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 175:247–267. <https://doi.org/10.1016/j.isprsjprs.2021.03.005>.
- Zhou, Z., M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang. 2018. "UNet++: A Nested U-Net Architecture for Medical Image Segmentation." *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Granada, Spain, 11045:3–11.
- Zhu, Z., S. Qiu, and S. Ye. 2022. "Remote Sensing of Land Change: A Multifaceted Perspective." *Remote Sensing of Environment* 282:113266. <https://doi.org/10.1016/j.rse.2022.113266>.
- Zhu, F., C. Wang, B. Zhu, C. Sun, and C. Qi. 2023. "An Improved Generative Adversarial Networks for Remote Sensing Image Super-Resolution Reconstruction via Multi-Scale Residual Block." *The Egyptian Journal of Remote Sensing and Space Sciences* 26 (1): 151–160. <https://doi.org/10.1016/j.ejrs.2022.12.008>.
- Zou, Y., Z. Yu, B. Kumar, and J. Wang. 2018. "Unsupervised Domain Adaptation for Semantic Segmentation via Class-Balanced Self-Training." 15th European Conference on Computer Vision (ECCV 2018), Munich, Germany, 297–313.