UNIVERSITY *of York*

This is a repository copy of *Reliable Contrastive Learning for Semi-supervised Change Detection in Remote Sensing Images*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/194198/

Version: Accepted Version

Article:

Wang, Jia-Xin, Li, Teng, Chen, Si-Bao et al. (3 more authors) (2022) Reliable Contrastive Learning for Semi-supervised Change Detection in Remote Sensing Images. IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. 4416413. ISSN: 0196-2892

https://doi.org/10.1109/TGRS.2022.3228016

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Reliable Contrastive Learning for Semi-supervised Change Detection in Remote Sensing Images

Jia-Xin Wang, Teng Li*, Si-Bao Chen*, Jin Tang, Bin Luo and Richard C. Wilson

***Abstract*—With the development of deep learning in remote sensing image change detection, the dependence of change detection models on labeled data has become an important problem. To make better use of the comparatively resource-saving unlabeled data, the change detection method based on semi-supervised learning is worth further study. This paper proposes a reliable contrastive learning method for semi-supervised remote sensing image change detection. First, according to the task characteristics of change detection, we design the contrastive loss based on the changed areas to enhance the model's feature extraction ability for changed objects. Then, to improve the quality of pseudo labels in semi-supervised learning, we use the uncertainty of unlabeled data to select reliable pseudo labels for model training. Combining these methods, semi-supervised change detection models can make full use of unlabeled data. Extensive experiments on three widely used change detection datasets demonstrate the effectiveness of the proposed method. The results show that our semi-supervised approach has better performance than related methods. The code is available at https://github.com/VCISwang/RC-Change-Detection.**

***Index Terms*—Contrastive learning, change detection, semi-supervised learning, remote sensing, semantic segmentation.**

I. Introduction

With the development of remote sensing technology, a large number of remote sensing images can be obtained more conveniently, which contains rich ground information. In the research of remote sensing (RS) image processing, change detection (CD) methods play an important role in addressing the issue of identifying change information in bitemporal co-registered images. The detection of changes in remote sensing at different times has important applications in assessing natural disasters [1], analyzing building changes [2] and urban expansion [3].

Traditional change detection methods can be divided into two categories: pixel-based CD methods [4] and feature-based CD methods [5] [6]. Pixel-based methods mainly detect pixel changes through difference calculation or ratio calculation of pixels of different images, such as change vector analysis (CVA) [4]. These methods are simple and fast, but it is difficult to distinguish changed areas from irrelevant objects. The feature-based method extracts the feature data of the object, and then compares the features of images at different times to obtain the change information of the region. Deng *et al*. [5] used principal component analysis (PCA) to extract the features of the target, and then compared the images to obtain the changed areas. Multivariate alteration detection (MAD) [6] and slow feature analysis (SFA) [7] also analyzed image changes based on feature transformation. For the unsupervised change detection, Cui *et al*. [8] used stochastic subspace ensemble learning to detect the changed areas, and they mainly used clustering algorithms to analyze the object features. These conventional methods usually get relatively crude predictions due to the limitations of the algorithm.

In past decades, deep convolutional natural networks (C-NNs) have been successfully applied in RS images, and change detection methods [9] [10] based on deep learning models also have achieved better performance. These methods are divided into single-stream networks [11] [12] and double-steam networks [13] [14] [15]. In single-stream networks, image-pairs are usually directly merged and input into the network, and then the encode network and decode network extract features to obtain the prediction of the changed areas. Alcantarilla *et al*. [11] used deconvolutional network for change detection, this method provides coarsely registered image pairs to a deep deconvolution network and predicts the changed areas. Peng *et al*. [12] combined low-dimensional features and high-dimensional features extracted from remote sensing images, and then use the attention module to enhance the feature identification ability, to achieve more accurate predictions of the changed areas. These single-stream methods are usually simple and efficient and can use the neural network to extra image features to achieve end-to-end change detection. On the other hand, more methods try to use the double-stream networks for the change detection on RS images. These networks usually extract features from images at different times in the feature extraction stage, and then merge features in the decode network to predict the change areas. These double-stream methods [16] [17] [13] are usually composed of two feature extraction networks with shared weights. They extract features from the images before and after the change, and then the changed regions obtained by the prediction network after feature fusion. Fang *et al*. [14] used two encoders to extract features of bi-temporal images, and then feed them into the UNet++ to generate the mask of change detection. These siamese networks [15] [17] usually have more accurate prediction due to the feature fusion modules. However, accurate detection results based on neural network models

Manuscript received XXXX, 2022.

*Corresponding authors. This work was supported in part by NSFC Key Project of International (Regional) Cooperation and Exchanges (No.61860206004), NSFC Key Project of Joint Fund for Enterprise Innovation and Development (No.U20B2068) and National Natural Science Foundation of China (No.61976004).

J. Wang, T. Li, S. Chen, J. Tang and B. Luo are with IMIS Laboratory of Anhui Province, Anhui Provincial Key Lab of Multimodal Cognitive Computation, MOE Key Lab of ICSP, Anhui University, Hefei, China (e-mail: jxwang@ahu.edu.cn).

R. Wilson is with Department of Computer Science, University of York, York, U.K.

usually depend on a large number of labeled data. Due to the complex scenes of RS images, the annotation of change detection images requires the manual judgment of the changes in different areas and then obtain labels, which will consume a lot of resources.

To alleviate the dependence of deep learning model on labeled data, researchers have proposed some methods. Earlier, researchers focused on semi-supervised approaches to image classification tasks. Semi-supervised methods usually require small labeled data sets, and then combine a larger number of unlabeled data in the model training process to obtain models with significantly improved performance. Virtual adversarial training (VAT) [18] and mean teacher [19] used the consistency regularization to achieve meaningful performance. These methods demonstrate the potential of semi-supervised learning to solve the model's dependence on data. Recently, based on the consistency regularization and pseudo labels, some methods [20] [21] [22] introduced strong and weak perturbation to improve the constraint ability of consistency on the model. Recent semi-supervised segmentation methods [23] mainly improved training methods based on the consistency regularization and pseudo labels, and also explore the perturbation methods. CutMix [24] proved that random region mixing is an effective perturbation method. CPS [25] proposed a training method that dual networks generate pseudo labels to guide unlabeled images.

For change detection, some methods obtain CD models using unsupervised algorithms instead of using labeled data. These methods make use of the contrastive loss [26] and similarity calculation [4], but these methods easily detected more unchanged regions. Then, some researchers try to use weak labels to replace pixel-level labels, such as image-level labels [27] and bounding boxes. These methods effectively improve the detection model performance, but also require additional manual annotation of the data. Some semi-supervised methods apply generative adversarial networks (GANs) [28] to solve the model's dependence on annotated data. They use generative networks to obtain simulated distribution data and discriminators to distinguish between different images. Although these methods improve the robustness of the models, they do not make full use of unlabeled data. Another semi-supervised CD method [29] introduced consistency regularization to unlabeled data. They add strong perturbation to unlabeled data or their features, and then use consistency loss to keep different prediction results consistent. However, the method based on the consistency regularization usually sets the threshold for the prediction probability of each pixel to obtain the pseudo label, which only contains parts of the original RS image, thus affecting the feature integrity of the changed objects.

This paper proposes a semi-supervised change detection method based on the reliable pseudo label and contrastive learning, which we call reliable contrastive learning for change detection (RCCD). First, to make the object features of the pseudo label more complete, we use the consistency of prediction of the model at different stages to the unlabeled image to select reliable image from the unlabeled data set. Pseudo labels are obtained from the best pre-trained model. Then, in order to improve the feature identification ability of the change detection model for the changed areas and unchanging areas, we select positive and negative samples for different regions. Different from the general contrastive learning [30] [31], the proposed semi-supervised contrastive loss is a pixel-level contrastive learning instead of image-level contrastive learning. Image-level contrastive learning is mainly to choose positive and negative sample pairs in different images. The method in this paper uses the characteristics of partial regional changes to design pixel-level positive and negative samples for contrastive learning, so as to improve the recognition ability of the model for pixel-level features. In general, the pseudo label based on reliable sample selection makes full use of the unlabeled RS images to improve the model, and the contrastive learning based on the changed areas also improves the detection ability of the semi-supervised model. These methods effectively improve the performance of the semi-supervised change detection model.

In summary, the main contributions of this work are as follows:

- We propose a semi-supervised change detection method based on reliable contrastive learning, which can obtain satisfactory performance by combining few labeled images and extra unlabeled samples.
- We select reliable samples according to the prediction uncertainty of unlabeled images in different stages of the model, and then obtain corresponding reliable pseudo labels for the training process.
- We propose the contrastive loss based on the changed areas, and it effectively improves the model detection ability for the changing objects.
- Experiments show that the proposed method can improve the model performance of small-scale datasets by using large-scale unlabeled data. Besides, our approach has more efficient training time and less parameters.

The rest of this paper is organized as follows. Section II analyzes related change detection works in detail. The proposed semi-supervised change detection method is described in Section III. The results of the experiments and the discussion are shown in Section IV and Section V. Finally, Section VI draws the conclusions of this paper.

II. Related Work

In this section, we discuss related semi-supervised methods about the data dependence of change detection models.

Some earlier change detection methods used classifiers to identify the change areas in the image-pairs, The semi-supervised research on these change detection method is mainly used cluster ensemble model to optimize the pseudo labels. Roy *et al.* [32] used a multiple classifier system in semi-supervised method, then they used iterative learning to label the unlabeled images. The final detection result is determined by multiple classifiers. This semi-supervised method mainly uses the co-training method of different models to improve the utilization of unlabeled images. For the unsupervised change detection, Shao *et al.* [33] first selected areas with a high probability of change by selecting thresholding the difference image histogram. The pseudo labels are jointly exploited with the intensity levels and spatial information, then they proposed a robust semi-supervised fuzzy C-means clustering algorithm. This method is similar to most semi-supervised methods in that probability threshold is used to obtain relatively reliable pseudo labels. The method proposed in this paper is to use the uncertainty of image to multiple models to select reliable data.

Some recent semi-supervised change detection methods utilize pre-training of models to obtain more latent information from unlabeled images. Li *et al*. [34] proposed a deep nonsmooth nonnegative matrix factorization network for synthetic aperture radar image change detection. This method mainly includes two stages: pretraining stage and fine-tuning stage. In the fine tuning stage, the decomposed matrices layer by layer and the latter aims to reduce the total reconstruction error. To solve the problem of insufficient labels, Tu *et al*. [35] used low-resolution labels to generate high-resolution change maps, and then fused the prediction results of the two training epochs to obtain the refined change prediction.

In the research of semi-supervised algorithm on remote sensing images, Wang *et al*. [36] explored the applicability of algorithms based on consistency for semantic segmentation on RS images. RanPaste [37] combined the image mixing method and proposes a more effective random paste perturbation for semi-supervised segmentation. These methods usually use the threshold value to select pixels with a high probability of prediction and then generate pseudo labels. However, the edge information of objects in these pseudo labels is easy to lost. The reliable pseudo labels proposed in this paper effectively avoids this problem.

In the change detection of remote sensing images, Susmita *et al*. [38] used the membership values of its K nearest neighbors to generate soft class labels. They proposed a heuristic method to select some patterns from the unlabeled ones for training. In addition, some early semi-supervised approaches use metric learning to exploit unlabeled data. Yuan *et al*. [39] proved that metric learning can extract change information from hyperspectral features, and use semi-supervised Laplacian regularization metric learning to solve sample problems. With the wide application of GAN in image processing tasks, researchers have begun to pay more attention to the use of generative networks to alleviate the data dependence problem in change detection. GDCN [40] used GAN to generate fake data using random noise for change detection model training. Although this method reduces the model's dependence on labeled data, it does not make use of unlabeled data. To leverage the unlabeled RS image, Peng *et al*. [41] proposed the SemiCD that uses GAN to make the model better distinguish ground truths from pseudo labels. This method improves the quality of pseudo labels generated by the model and finally improves the model performance. Recently, Wele *et al*. [29] introduced different types of perturbations into the network middle layer of change detection based on the consistency regularization, and then train the consistency loss of different prediction outputs after adding the perturbation. The revisiting consistency regularization (RCR) used complex disturbances to improve the robustness of the detection model and feature extraction capability, but ignores the characteristics of change detection on remote sensing images.

In this paper, we first improve the generation method of pseudo labels. Reliable samples are selected by calculating the prediction uncertainty of unlabeled data in different epochs of pre-trained models. Then, since the input of the change detection model contains bi-temporal images, we design a contrastive learning method based on the changed areas. Different from the general contrastive learning methods, this reliable contrastive learning proposed in this paper is based on definite positive and negative sample pixels. Because of the improvement of pseudo-label generation method and the addition of the contrastive loss, the proposed semi-supervised change detection method effectively improves the model performance by using unlabeled remote sensing images. The detailed modules will be illustrated in the following sections.

III. Proposed Approach

Our goal is to improve the accuracy of the semi-supervised change detection model by combining a large number of unlabeled images with a few labeled annotated images. For general semi-supervised approaches, the quality of the pseudo label is crucial. The proposed method in this paper selects a more reliable subset of all unlabeled data based on the uncertainty of model prediction on unlabeled images. This method improves the quality of the pseudo label and improves the detection model performance.

Different from the general image processing methods which only focus on the feature recognition of the object, the characteristic of change detection is to identify the difference of the object between the image-pairs. So, based on the characteristics of change detection on remote sensing images, we design the contrastive learning loss in semi-supervised change detection. We input the changed images and the unchanged images into the model respectively, then calculate the loss of the predicted results. The pixels corresponding to the changed areas are selected as negative samples, and the other pixels are selected as positive samples. Through contrastive learning for change detection, the model strengthens the ability to identify the changed areas, and finally improves the detection accuracy.

The main framework of the proposed approach is shown in Fig. 1. We will introduce the application method of different modules in the following subsections.

A. Overall of Proposed Reliable Contrastive Learning for Semi-supervised CD

In semi-supervised change detection tasks, a labeled dataset $D^l = \{(x_A^i, x_B^i, y^i)\}_{i=1}^M$ with few samples. Where x_A is the pre-change image, x_B is the post-change image, and y is the corresponding label. Meanwhile, we usually have a remote sensing image data set $D^{ul} = \{(u_A^i, u_B^i)\}_{i=1}^N$ that has not been annotated. Where u_A and u_B are a pair of bi-temporal remote sensing images, and in most cases $N \gg M$.

The proposed method is to improve the performance of the detection model by extracting the latent information from unlabeled data set D^{ul}. As shown in Fig. 1, our method is divided into two stages. In the first stage, labeled data

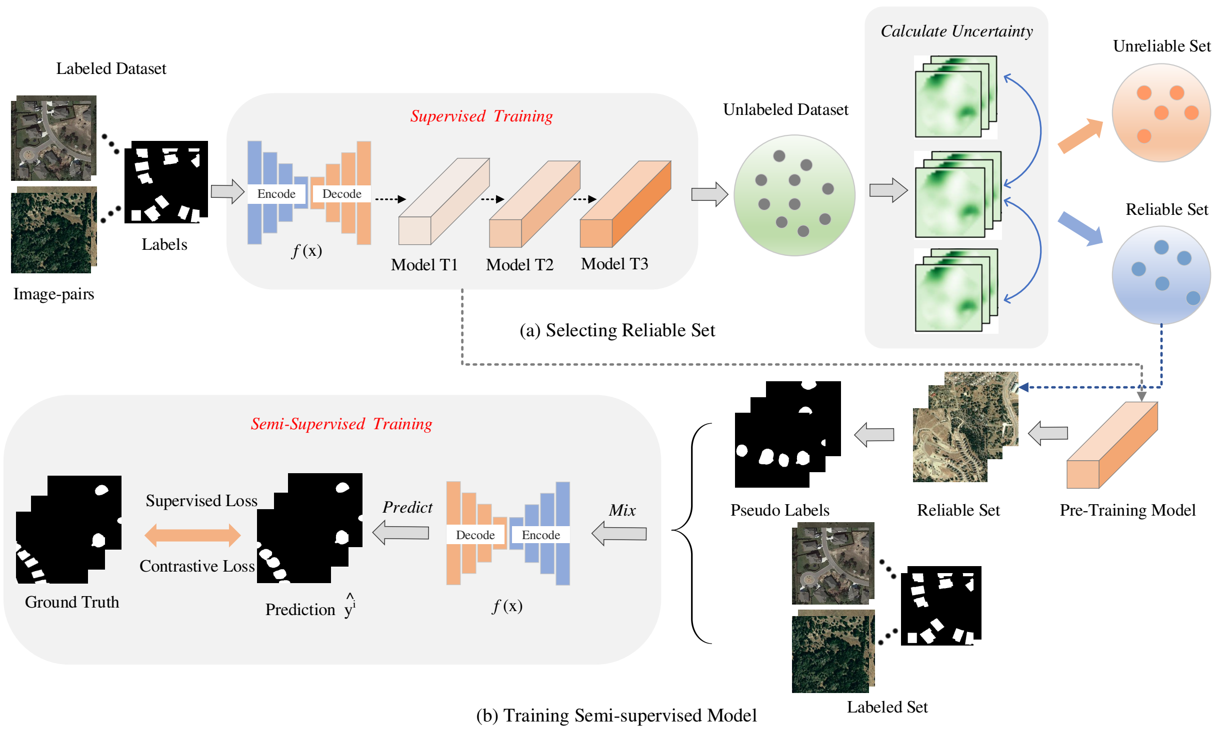

Fig. 1. Overview of the proposed reliable contrastive learning method. (a) Selecting reliable set. The uncertainty of unlabeled images is calculated by pre-training models in different epochs. Then, unlabeled data sets are divided into reliable set and unreliable set by their uncertainty. (b) Training semi-supervised model. We use the contrastive loss for unlabeled data to improve the performance of change detection model.

D^l is used to fine tune the pre-trained model, while models of different epochs are saved for uncertainty calculation. By sorting the uncertainty of unlabeled data, we divided different samples into the reliable data set and the unreliable data set.

Then, we use the pre-training model to obtain pseudo labels for reliable data sets and then mix these data with labeled samples. The detection network is reinitialized and then trained with this mixed data. In the semi-supervised training, we predicted the changed pixels and unchanged pixels respectively, and then calculated the contrastive loss. Finally, we obtain the semi-supervised change detection model. This method achieves better performance due to the improvement of pseudo label quality and better recognition ability of the changed areas.

B. Using Uncertainty to Obtain Reliable Set

An important module of the proposed semi-supervised method is to use uncertainty to select a reliable subset of unlabeled samples. As shown in Fig. 1-(a), in the pre-training stage, we use labeled data set D^l as the training set to train change detection model $f(x)$. In the proposed semi-supervised method, $f(x)$ uses the same architecture as the recent semi-supervised method [29], including the encode and decode. The encode uses a pre-trained ResNet50, image-pairs x_A and x_B are input the encode to get corresponding features. The decode is composed of the upsampling modules, and finally obtains the predicted change probability map $\hat{y^i}$.

$$\hat{y^i} = f(x_A^i, x_B^i). \tag{1}$$

For the predictions with labeled data, we utilize the Cross Entropy (CE) loss [42] as the supervised loss to train the detection model. The loss L_s is calculated as follows:

$$L_s = \frac{1}{M} \sum_{i \in D^l} CE(\hat{y^i}, y^i), \tag{2}$$

where y^i is the label of training data (x_A^i, x_B^i). In order to compute the uncertainty of the unlabeled data D^{ul}, we save three checkpoints of the model in different training epochs, which are model $T1$, $T2$ and $T3$. Then we utilize these models to predict the unlabeled data set D^{ul} respectively, and obtain predictions $f(u_A, u_B, \theta, T_j)$ of the changed areas. After argmax calculation of the prediction probability, the pseudo label y_{pj}^i predicted by different models on these data is obtained. j is the training epoch when training the supervised model.

$$y_{pj}^i = argmax\ f(u_A^i, u_B^i, \theta, T_j).\ j \in \{1, 2, 3\} \tag{3}$$

Since the supervised model accuracy is gradually improved in the training stage, some studies [43] have found that image-pairs that are relatively easy to identify produce accurate predictions earlier. In the experimental discussion section of this paper, we also visually compare the accuracy changes of reliable samples and unreliable samples in different training stages. The experimental results show that this theory can also be used in change detection. By mean Intersection over Union (meanIoU) calculation of the prediction results of unlabeled image-pairs at different epochs, we obtain the uncertainties

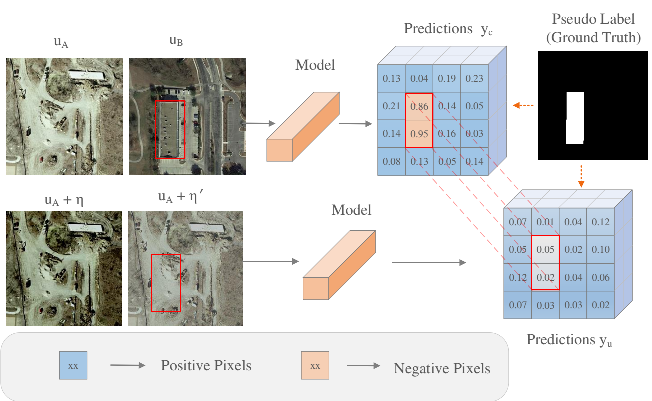

Fig. 2. Contrastive learning loss calculation of a pair of remote sensing images. The area in the red box is the changed objects that we need to detect.

uc_i of different image-pairs.

$$uc_i = \frac{1}{N}\sum_{i\in D^{ul}}\sum_{j=1}^{K-1} meanIoU(y_{pj}^i, y_{pK}^i), \tag{4}$$

where y_{pK}^i is the best prediction by pre-trained model T_k at epoch K, y_{pj}^i is predictions by other models at other epochs. In the proposed method, K is set as 3. Then, we sort the uncertainty of all unlabeled data. In the experimental part of this paper, we analyze the effect of different proportions of reliable data on the change detection model. The results show that the semi-supervised model has the best performance when half of the unlabeled data is selected as reliable data. These image-pairs with lower uncertainty as reliable subset, and the data with higher uncertainty as unreliable subset.

$$D^{ul} \rightarrow \begin{cases} D_r^{ul} = \{(u_A^i, u_B^i, y_p^i)\}_{i=1}^{N/2}, & Reliable\ set \\ D_u^{ul} = \{(u^i, u^{i'}, y_p^i)\}_{i=N/2}^{N}. & Unreliable\ set \end{cases} \tag{5}$$

After dividing these unlabeled data, we use the model T_j with the highest accuracy in the pre-training stage to obtain the pseudo labels y_p^i corresponding to the reliable data D_r^{ul}. For unreliable data D_u^{ul}, we make unchanged labels y_n by copying one of the images, so that their corresponding real labels y_p^i are the label with all unchanged areas.

C. Contrastive Learning for Change Detection

The change detection on remote sensing images is mainly to detect the changed areas in a image-pairs by the model, which is also the difference between them. We believe that the contrastive learning can enhance the model's ability to identify the changed areas, and improve the model's accuracy.

However, the general contrastive learning method is to select positive and negative samples in the training set to calculate the contrastive loss. In the semi-supervised change detection, we pay more attention to the target change information of the same geographic location in remote sensing image-pairs. Therefore, we design a contrastive learning loss based on positive and negative pixels according to these characteristics of change detection.

As shown in Fig. 1-(b), for labeled data and reliable subset, we use the model to generate their predictions $\hat{y^i}$ and $\hat{y_p^i}$ respectively. Meanwhile, when calculating the loss, the ground truth y^i is used for the labeled data, and the pseudo label y_p^i generated in the previous stage is used for the unlabeled data. For these RS images, we still use the Cross Entropy loss L_s training the semi-supervised model.

$$\hat{y^i} = f(x_A^i, x_B^i, \theta),\ \hat{y_p^i} = f(u_A^i, u_B^i, \theta). \tag{6}$$

$$L_s = \frac{1}{M}\sum_{i\in D^l} CE(\hat{y^i}, y^i) + \frac{1}{N}\sum_{i\in D^{ul}} CE(\hat{y_p^i}, y_p^i). \tag{7}$$

It is worth noting that based on the change detection on remote sensing images, the proposed contrastive learning method is shown in Fig. 2. We first input a image-pairs (u_A, u_B) into the detection model T_j with the best accuracy, which are images of the same scene. In addition, we use the contrastive loss for both labeled data and unlabeled data in semi-supervised experiments.

Then, to construct positive and negative sample pixels, we add two random strong perturbations η and η' to one of the image-pairs to generate a perturbed pair of unchanging images, and input them into the network. In the semi-supervised experiments, we use colorjitter, grayscale, blur, and Cutout [44] with random values filled to apply the strong data augmentations η and η'.

As shown in Fig 2, the original image-pair (u_A, u_B) has change areas, denoted as M_{ca}. In addition, since we added two random perturbations to image A, the new image-pair $(u_A + \eta, u_A + \eta')$ does not have any change areas. By comparing the two image-pairs of RS imges, it can be seen that these image-pairs should have opposite predictions in the changed areas M_{ca}. For the region outside the change area M_{ca}, it denoted as unchanged areas M_{ua}. For the original image-pair (u_A, u_B), the semi-supervised model obtains their change probability map y_c. For another unchanged image-pair $(u_A + \eta, u_A + \eta')$, the predicted change probability map is y_u. In the M_{ua}, two image-pairs should have consistent predictions about whether the area is changed or not. In the proposed contrastive loss L_c calculation, pixels with consistent predictions in region M_{ua} are taken as positive samples, and pixels with opposite predictions in region M_{ca} are taken as negative samples.

For the positive pixels in M_{ua}, their corresponding areas are not the changed areas that need to be recognized in bi-temporal images, so we calculate the loss L_p of these positive pixels:

$$L_p = MSE(y_c,\ y_u,\ M_{ua}). \tag{8}$$

To make the predictions of two image-pairs more similar, we use the mean square error (MSE) loss to train the semi-supervised model. In the changed areas, since the two groups of images should have opposite predictions, we divided the predicted probabilities of two image-pairs into the change probabilities y_c^0, y_u^0 and the unchanged probabilities y_c^1, y_u^1. When calculating the loss of negative samples, in order to maintain a steady decline in the negative loss, we make

Algorithm 1 Reliable Contrastive Learning.

Input: Labeled training set $D^l = \{(x_A^i, x_B^i, y^i)\}_{i=1}^M$, Unlabeled training set $D^{ul} = \{(u_A^i, u_B^i)\}_{i=1}^N$
Output: Semi-supervised change detection model $f(\theta)$
Train model $f(\theta)$ on D^l with L_s and L_{ct}
Save model $T1$, $T2$ and $T3$ on different training epochs
Compute the uncertainty for unlabeled set D^{ul}, and generate pseudo labels y_p
Select reliable samples to compose D_r^{ul}, and $D_u^{ul} = D^{ul} \setminus D_r^{ul}$ as unreliable samples
Train semi-supervised model $f(\theta)$ on $(D^l \cup D^{ul})$
for $t = 1 : maxIter$ **do**
  Select labeled image x_A, x_B, label y and unlabeled image u_A, u_B, pseudo label y_p
  Add strong perturbation η to reliable unlabeled RS images $\rightarrow \{(u_A, u_B)\ ,\ (u_A + \eta, u_A + \eta')\}$
  Set unreliable data $\{(u_A, u_B)\} \rightarrow \{(u_A, u_A), y_p\}$ or $\{(u_B, u_B), y_p\}$
  Train semi-supervised model with L_s on all samples
  if $\{u_A, u_B\} \in D_r^{ul}$ **then**
    Generate predictions y_c and predictions y_u for $\{(u_A, u_B)\ ,\ (u_A + \eta, u_A + \eta')\}$
    Train model with the contrastive loss L_{ct}
  else
    Generate predictions $f(u_A, u_A)$ and $f(u_B, u_B)$
    Use the label y_n for supervised training of unreliable samples
  end if
  Use L_s and L_{ct} to optimize the semi-supervised model
end for
return Change detection model $f(\theta)$

two predictions closer to each other in opposite prediction probability, so the model more accurately identify the changed areas. The negative loss L_n is calculated as follows:

$$L_n = MSE(y_c^0, y_u^1, M_{ca}) + MSE(y_c^1, y_u^0, M_{ca}). \quad (9)$$

Combined with the loss of positive and negative samples, the contrastive loss L_{ct} is calculated as follows:

$$L_{ct} = L_n + L_p. \quad (10)$$

In addition, since the changed area M_{ca} can be obtained by labels or pseudo labels, the proposed contrastive loss can be used in both labeled data and unlabeled data. We also add this loss in the pre-training stage of the model, and the experimental results show that it can significantly improve the model performance.

In order to better express the proposed semi-supervised method, Algorithm 1 shows the pseudocode of the reliable contrastive learning method. In the pre-training process, the model uses cross entropy loss L_s and contrastive loss L_{ct} to train the change detection model on labeled image-pairs D^l. Meanwhile, we saved several models T_i in different epochs. Then the reliable unlabeled image-pairs are selected by comparing the uncertainty of model's predictions. Finally, after screening the reliability of unlabeled data, the L_{all} of the proposed reliable contrastive learning method on the change detection model is:

$$L_{all} = L_s + \lambda L_{ct}, \quad (11)$$

where λ is the weight set by the semi-supervised loss, which is usually set as 1 in our experiments.

By improving the quality of pseudo labels and the model's ability to identify the changed areas, the proposed method significantly improves the accuracy of the detection model by utilizing a large number of unlabeled images when there are only few labeled images. We also verify the validity of the proposed method on different datasets, and the experimental results and discussions are described in the following sections.

IV. Experiments

A. Experimental Setup

Datasets. To verify the proposed semi-supervised method, we use three remote sensing image change detection datasets: SZTAKI airchange dataset [45], WuHan University (WHU) dataset [46] and LEarning, VIsion and Remote sensing (LEVIR)-CD dataset [2].

SZTAKI dataset contains 13 pairs of 952×640 aerial images with a spatial resolution of 1.5m. The objectives of the change mainly include: new build-up regions, building operations, planting, fresh plough-land and ground before building over. However, it should be noted that the label of this dataset only contains the changed areas, without corresponding semantic information. In the semi-supervised experiment, we crop each of the original images overlapping into 12 images of 256×256. Finally, we dropped some unchanged image-pairs, 122 remote sensing image-pairs were obtained. Then we have 98/12/12 image-pairs for training/validation/test, respectively.

WHU dataset mainly covers the area reconstructed after the earthquake. The bi-temporal images in this dataset consist of aerial images taken in 2012 and 2016, respectively. The change object to be detected is the buildings with large-scale changes. The original data is a remote sensing image with a large resolution, so researchers generally cut it into smaller patches for training. In our experiments, we first cut the original image into 256×256 images and found that there were many unchanged images, so we removed the unchanged data from this dataset. Finally, the WHU dataset has 1512/189/189 pairs of RS images for training/validation/test, respectively.

LEVIR-CD consists of 637 high-resolution remote sensing image patch pairs with a size of 1024×1024 pixels. These data mainly record the building changes in the same area, including warehouses, houses, buildings and so on. Following recent change detection methods [47] [29], we also cropped the original images into 256×256 non-overlapping patches. After processing the original data, we finally obtain 7120/1024/2048 pairs of RS images for training/validation/test, respectively.

As can be seen from Tabel I, semi-supervised change detection experiments use three different types of remote sensing image data sets. SZTAKI has a small amount of data, and this dataset is used to verify the performance of the method in this paper when large datasets are not available. It should also be noted that the data set contains more

TABLE I
COMPARISON OF THREE CHANGE DETECTION DATASETS.

Datasets	Image-pairs	Image size	Train / Val / Test	Resolution	Changes
SZTAKI [45]	13	952 × 640	98 / 12 / 12	1.5 m/pixel	building, planting, plough-land...
WHU [46]	1	15354 × 32507	1512 / 189 / 189	0.2 m/pixel	building
LEVIR-CD [2]	637	1024 × 1024	7120 / 1024 / 2048	0.5 m/pixel	building

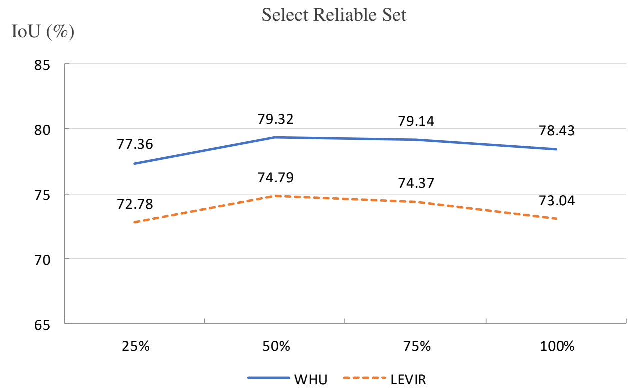

Fig. 3. Model performance comparison when different proportions of unlabeled data are selected as reliable data.

different changed objects, which is also conductive to verifying the applicability of the proposed semi-supervised method. In contrast, WHU dataset and LEVIR dataset have more remote sensing data available and can be fairly compared with other semi-supervised methods.

Implementation details. In the experiments of semi-supervised change detection, we usually focus on the influence of semi-supervised algorithms on model performance, so we use a widely-used change detection model to compare different semi-supervised methods. Besides, to compare different methods fairly, we maintain the same hyperparameters for all experiments. We set batch size to 8 for both supervised and semi-supervised models. In the semi-supervised experiments, each batch has 4 labeled samples and 4 unlabeled samples respectively. The basic learning rate lr of model training is 0.001, and the poly scheduling is used to decay the learning rate. The model all trained 80 epochs on WHU and LEVIR-CD datasets. For labeled images, we use weak data augmentations: random flipping, random crop, random re-scale and Gaussian blur. In order to improve the constraint ability of the consistency regularization on the model, we applied strong data augmentations for the unlabeled image, including color jitter, Cutout [44] and grayscale.

When comparing semi-supervised methods, 5%, 10%, 20% and 40% labeled data were randomly selected for model training, and others were used as unlabeled data to training the semi-supervised model. We use Intersection Over Union (IoU) as the main evaluation criterion when comparing the semi-supervised models, and the experimental results mainly compare the IoU of change class. In addition, we also applied the overall pixel accuracy (OA) to compare the model performance. Our method is applied on PyTorch, and the semi-supervised model is trained on an NVIDIA Quadro RTX 1080Ti GPU.

TABLE II
COMPARISON RESULTS (IOU, %) ON SZTAKI TEST SETS WITH DIFFERENT PERCENTAGES OF LABELED DATA.

Methods	10%	20%	40%	100%
Sup. only [47]	9.85	28.11	30.43	39.78
Ours. pre	12.37	29.56	32.72	41.25
Ours	12.68	32.57	34.38	-

Parameter Settings. In the proposed method, some images in the unlabeled dataset should be selected as a reliable subset, so we ranked their uncertainties. To analyze the influence of different proportions of reliable data on the semi-supervised model, we conducted comparative experiments on two datasets, and the results are shown in Fig 3. We selected different proportions of reliable data in the order of uncertainty from small to large to train the semi-supervised model. It can be seen from the results that when the proportion of reliable data is 50% or 75%, the model has better performance. In order to reduce the time-consuming of the method, we chose 50% of the unlabeled images as the reliable set in the experiments. At the same time, it can be seen that when we do not select reliable data, 100% of images are used as reliable images and added to training, and the accuracy of the model will decrease significantly.

B. Comparison with State-of-the-Art Methods

SZTAKI. On the SZTAKI dataset, we selected different proportions of labeled images to verify the proposed semi-supervised method. Due to the dataset contains only 98 image-pairs for training, the accuracy of the model fluctuates too much in the case of too little training data, which is not conductive to the comparison of method differences. Therefore, we choose 10%, 20% and 40% labeled data for training.

The results of the experiment are shown in TABLE II. When only 10% of the training data has labels, the IoU of detection model is only 9.85. The performance of change detection was significantly improved by using the pre-training process with the contrastive loss. However, when the semi-supervised model is further trained by the reliable pseudo-label method, the performance of the model is improved very little. After checking the quality of the pseudo labels, we found that when the number of labeled images was small, the predictions obtained by the model were very rough, so the pseudo labels' improvement on the semi-supervised method was also very limited. This problem can also be verified

TABLE III
COMPARISON RESULTS ON WHU TEST SETS WITH DIFFERENT PERCENTAGES OF LABELED DATA. THE TABLE LISTS THE MODEL PERFORMANCE AFTER DIFFERENT STAGES OF THE PROPOSED SEMI-SUPERVISED METHOD.

Method	5%		10%		20%		40%	
	IoU(%)	OA(%)	IoU(%)	OA(%)	IoU(%)	OA(%)	IoU(%)	OA(%)
Sup. only [47]	65.73	92.08	72.93	93.92	77.38	95.06	81.95	96.19
RCR [29]	76.65	95.2	79.10	95.70	83.87	96.68	84.66	96.84
Ours. pre	70.05	93.60	77.37	95.07	78.6	95.27	83.63	96.61
Ours	79.32	95.67	82.98	96.48	84.16	96.73	85.28	96.98

TABLE IV
COMPARISON RESULTS ON LEVIR-CD TEST SETS WITH DIFFERENT PERCENTAGES OF LABELED DATA. THE PROPOSED METHOD COMPARES MODEL ACCURACY OF DIFFERENT SSL METHODS IN CHANGE DETECTION.

Method	5%		10%		20%		40%	
	IoU(%)	OA(%)	IoU(%)	OA(%)	IoU(%)	OA(%)	IoU(%)	OA(%)
Sup. only [47]	61.0	97.60	66.8	98.13	72.3	98.44	74.9	98.60
AdvNet [48]	66.1	98.08	72.3	98.45	74.6	98.58	75.0	98.60
s4GAN [49]	64.0	97.89	67.0	98.11	73.4	98.51	75.4	98.62
SemiCDNet [41]	67.6	98.17	71.5	98.42	74.3	98.58	75.5	98.63
RCR [29]	72.5	98.47	75.5	98.63	76.2	98.68	77.2	98.72
Ours	74.79	98.78	76.7	98.83	77.01	98.87	77.10	98.89

by comparing experimental results with a high proportion of labeled data.

Our approach improve model performance by the contrastive loss when all data is labeled. At the same time, the training data did not need pseudo labels, so there was no corresponding results in the table. It also worth noting that the change targets of this dataset are not only buildings, but also other types of targets such as planting. The experimental results also prove that the proposed method can perform well when the ground changes are more complex. Therefore, from the overall results, the proposed semi-supervised change detection method can still have good performance on small datasets.

WHU. On the WHU dataset, when directly using these images to train the model, we found that the accuracy of model varies greatly, and the differences between different methods can not be well compared. Therefore, we screened some images in the dataset, removed the samples without any changes, and then compared the proposed methods in this subset. The experimental results are shown in TABLE III. We first train the supervised model using different numbers of labeled images as baseline. To compare with other semi-supervised change detection methods, we use the RCR to train semi-supervised models on different proportions of labeled data. This approach is also the state-of-the-art method with open source code, then we can train models on the WHU dataset. It can be seen from the table that when the number of labeled data decreases from 40% to 5%, the accuracy of the supervised model decreases by 16 percentages. These results demonstrate the dependence of the change detection model on labeled data. When using the proposed semi-supervised method, the accuracy of the pre-training model is improved compared with the supervised model. We believe that is mainly because we used the contrastive learning in the first stage. After combining the reliable samples, the semi-supervised model achieve 79.32 IoU when there are only 5% labeled images.

In Table 1, when only 5% and 10% labeled data were used, our method improve the IoU about 3 percents compared to other methods. However, it worth noting that the accuracy of semi-supervised methods has increased about 10 percents than the supervised model. Since different semi-supervised methods have been greatly improved compared with supervised models, our method actually has a significant performance improvement compared with other semi-supervised methods. In general, compared with other semi-supervised methods, our method obtains the best performance on both IoU and OA of the model.

LEVIR-CD. To better compare the differences between our method and different semi-supervised methods, our experiments on LEVIR-CD used the same setup as the RCR [29] method, and we repeated their experiments to obtain the semi-supervised models. In addition, we compared the results of AdvNet [48] and s4GAN [49], which are semi-supervised segmentation methods. In the TABLE IV, SemiCDNet [41] is the related semi-supervised change detection method. In the experiment, we still select different proportions of labeled images for semi-supervised change detection models. The results comparison of different methods is shown in TABLE IV.

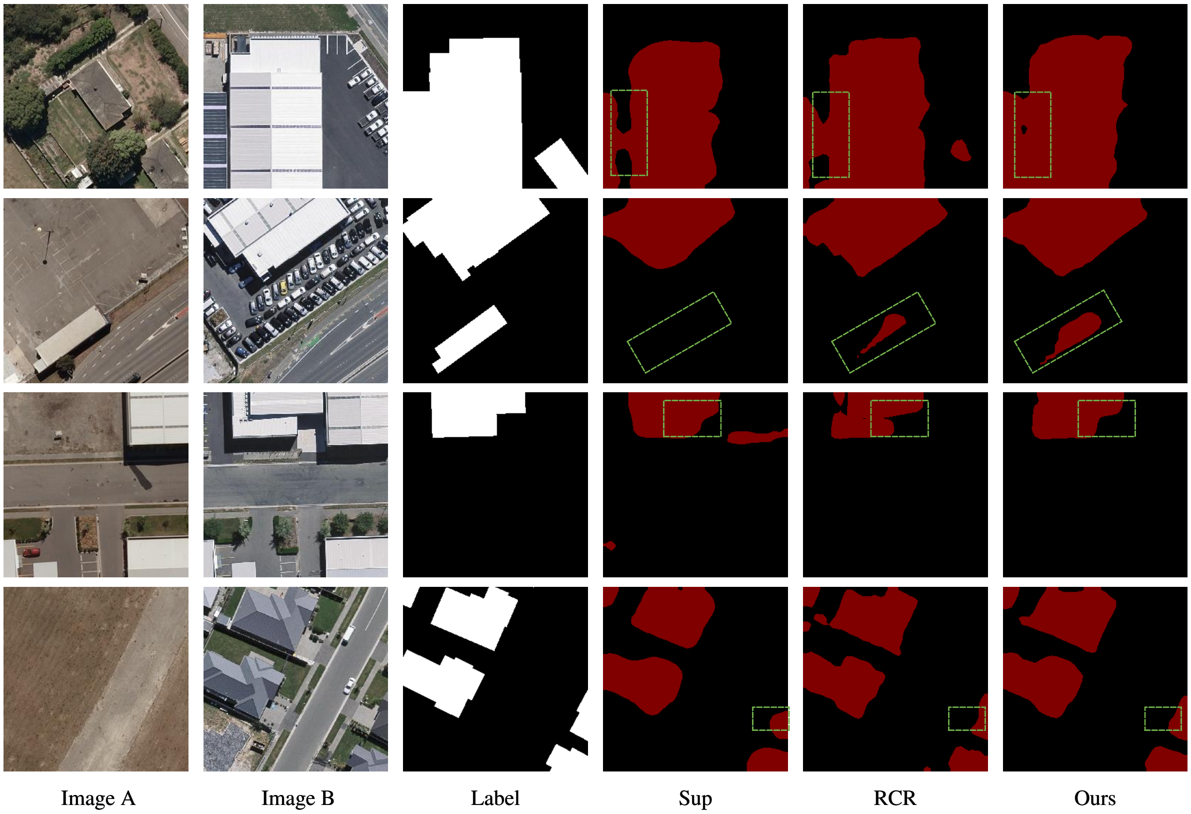

Fig. 4. Comparative examples of the proposed RCL on WHU dataset. Each line is an example. The first two columns are image pairs of CD and the third column is the ground truth. The right three columns are comparison of related methods (Sup [47] and RCR [29]) and our RCL method.

It can be seen from the table that the results of the semi-supervised method based on change detection are obviously better than other semi-supervised segmentation methods. This proves that semi-supervised learning has different characteristics in change detection, and semi-supervised change detection is worthy of further exploration. Compared with the latest semi-supervised change detection methods, the proposed semi-supervised method has better performance in different experiments, and the IoU increases by about 2 percentages when the labeled data is 5%. In addition, our method is slightly lower than RCR when the number of labeled data is at 40%. We believe that it is due to the differences between different semi-supervised methods decreasing as more labeled data is available. Meanwhile, semi-supervised methods usually focus on model performance when the number of labeled data is much less than the amount of unlabeled data.

Visualization. To compare the performance of different semi-supervised models in change detection, we also visualize the predictions on WHU and LEVIR-CD. As shown in Fig. 4, we conducted semi-supervised experiments with different proportions of labeled data on WHU. The results show that semi-supervised methods can obviously get more accurate predictions compared with the supervised training. Compared with the results of RCR, our model predictions have less error detection and can detect the changed areas more accurately. In addition, on the LEVIR dataset, we select a pair of images to compare the model predictions. The model prediction after training with different number of labeled images by different methods is shown in Fig. 5. As can be seen from the figure, the number of labeled images will significantly affect the performance of the model in the semi-supervised method, and the proposed method also has more accurate predictions in some edge areas. Based on the results of these datasets, we believe that the reliable contrastive learning effectively improve the recognition ability of the model to the changed areas.

V. Discussion

In this section, we discuss and analyze the innovations proposed in this paper, and demonstrate their influence in semi-supervised experiments. First, we analyzed the selection of reliable data, and then demonstrated the import role of contrastive loss through ablation experiments. Then, a series of semi-supervised experiments combining small-scale and large-scale data sets proves the generality of the proposed method. Finally, the differences in the time complexity between the proposed method and other methods are compared.

A. Effectiveness of the Reliable Data

Comparison of the image uncertainty. In the method section, we propose that more reliable images usually obtain accurate prediction earlier in the training process, and we also conducted visual comparison experiments to verify this method. We saved the model checkpoints obtained at different epochs on the WHU dataset. Then we randomly selected a reliable sample and and unreliable sample to obtain the predictions on these models respectively. The comparison of the prediction results of these unlabeled images is shown in the

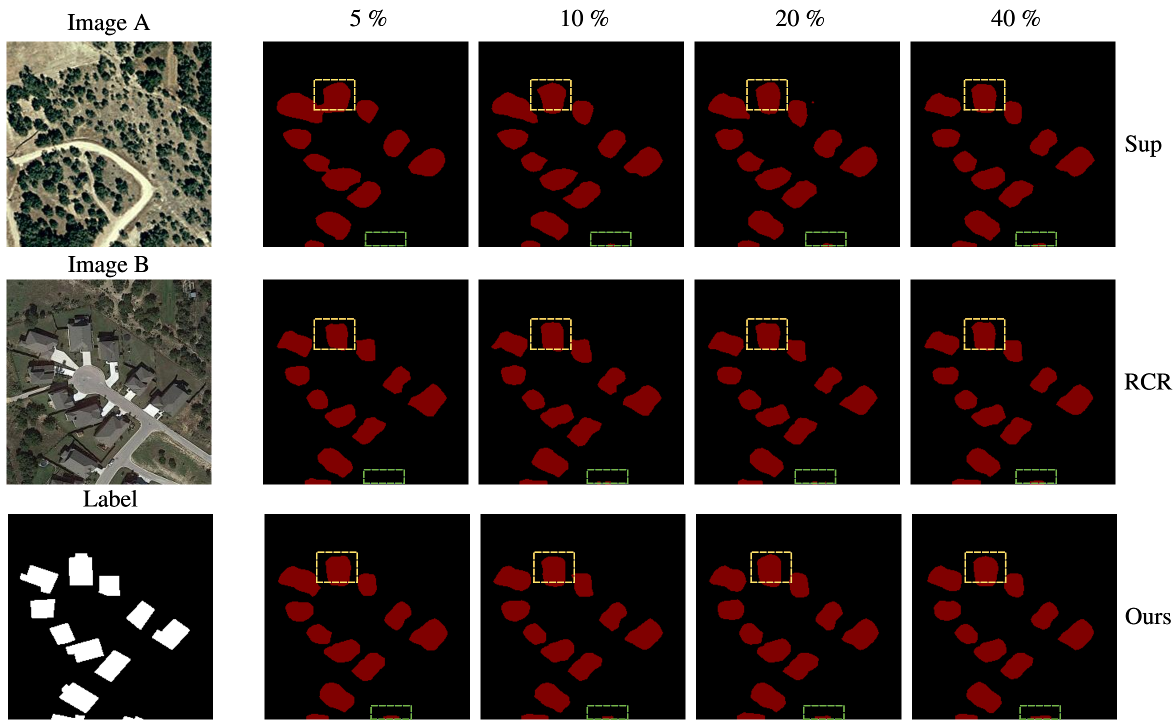

Fig. 5. Comparative examples on LEVIR-CD dataset. From left to right, the first column is the bi-temporal samples and ground truth, and the other columns are different model predictions trained with different proportions of labeled data.

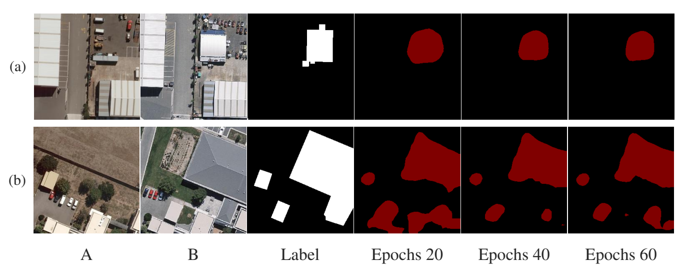

Fig. 6. Visual comparison of image uncertainty. (a) Reliable sample. (b) Unreliable sample.

Fig 6. As can be seen from the figure, the labels of reliable sample selected after calculating the uncertainty are usually easier to identify, and the early model can predict the accurate change area. Besides, the unreliable sample label is more complex change objects, and the model predictions in different epochs are quite different. These results also prove that we can select more reliable data as reliable subset by calculating the model uncertainty in different training epochs. In conclusion, our method can further improve the performance of semi-supervised models by selecting unlabeled images.

Comparison of pseudo label. More reliable data are selected in order to use their pseudo labels to train semi-supervised models and improve model performance. Therefore, we divide the unlabeled data into reliable set and unreliable set, and then generate pseudo labels for them respectively. So, we use the ground truths to calculate the accuracy of pseudo labels. In the experiment, we used models at different epochs in the pre-training to compare their predictions. The results of the comparison are shown in the Fig 7. As can be seen from the figure, in different training epochs, the accuracy

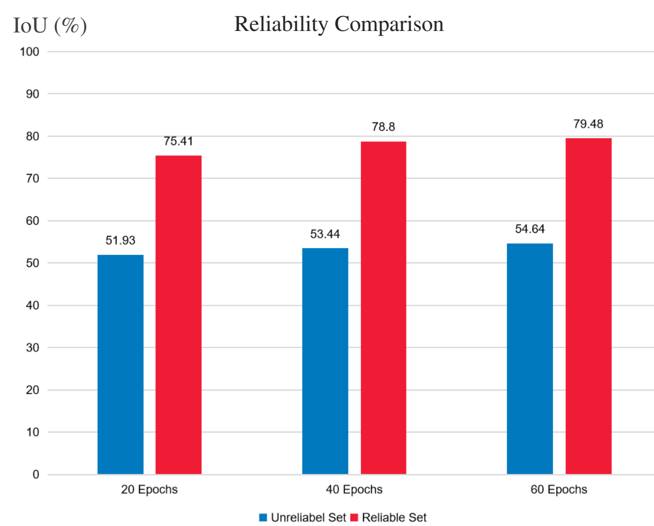

Fig. 7. Accuracy of pseudo labels generated by reliable set and unreliable set on WHU dataset.

of the reliable samples selected by the proposed method is significantly higher than the unreliable samples. This indicates that the proposed method selects samples with more accurate pseudo labels from unlabeled samples. In the training of the semi-supervised model, the information contained in the pseudo labels of these samples effectively improve the semi-supervised model performance.

By comparing the reliable data and the unreliable data in model training, the results show that the proposed pseudo label generation method can make full use of the unlabeled remote sensing image data combined with the change detection model.

TABLE VI
MODEL ACCURACY (cIoU, %) OF THE PROPOSED APPROACH WHEN COMBINING DIFFERENT DATASETS.

Methods	Labeled data / Unlabeled data	10%	20%	40%	100%
Sup. only [47]	SZTAKI / -	9.85	28.11	30.43	39.78
Ours. semi	SZTAKI / SZTAKI	12.68 (+2.83)	32.57 (+4.46)	34.38 (+3.95)	41.25 (+1.47)
Ours + ext (WHU)	SZTAKI (98) / WHU (7120)	16.66 (+6.81)	34.42 (+6.31)	38.75 (+8.32)	44.54 (+4.76)

TABLE V
EFFECTIVENESS OF DIFFERENT METHODS. EXPERIMENTS ARE CONDUCTED TO TRAIN SEMI-SUPERVISED CD MODEL USING 5% LABELED DATA ON WHU DATASET. THE RESULTS ARE COMPARED WITH THE IoU OF THE CHANGED AREAS.

Methods	Sup	Aug	CT_Loss	Reliable	Unreliable	cIoU(%)
(a)	✓					65.73
(b)	✓	✓	✓			70.05
(c)	✓			✓		77.33
(d)	✓	✓		✓		78.24
(e)	✓	✓	✓	✓		79.03
(f)	✓	✓	✓	✓	✓	79.32

B. Effectiveness of the Contrastive Loss

One of the contribution in the proposed semi-supervised change detection method is the contrastive loss of pairwise images. To demonstrate the effect of the proposed contrastive loss on the semi-supervised change detection model, we conducted a series of ablation experiments. The effects of contrastive loss on the model were compared in the experiments. The experiments compare the contributions of different techniques to semi-supervised methods, and the results are shown in TABLE V.

We first use 5% labeled data to train a supervised model on the original network as the baseline, and the experimental results are shown in (a) of TABLE III. In experiment (b), we still only used 5% labeled images, and then added the unchanged sample pairs and the contrastive loss. The results showed that the model IoU increased by 4.32 percentage points. We believe that when the number of labeled data is small, adding unlabeled samples effectively improves the generalization ability of the model, and the contrastive loss also further strengthens the feature extraction ability of the change detection model.

Besides, we selected reliable unlabeled images for training the semi-supervised model without the contrastive loss, and the results are shown in method (c). This proves the importance of reliable pseudo labels for semi-supervised models. After using the proposed method step by step, the model accuracy can be improved continuously. Finally, for the unreliable samples that are not selected, we create corresponding unchanged samples and add them to the training.

The comparison between experiments (e) and (f) also shows that this method makes full use of these data. It can be proved from experiments (b) and (e) in the table that the proposed contrastive loss can not only significantly improve the model's ability to identify the changed area in the pre-training process, but also effectively improve the model's performance when the semi-supervised experiment is conducted with reliable data. Combined with these experiments, we find that the proposed different techniques achieve different degrees of performance improvement in the semi-supervised change detection.

C. Model Generalizability and Time Complexity

Model Generalizability. It is very important for the semi-supervised change detection method to improve the model performance by combining different types of datasets. For general scenarios, small-scale labeled datasets are usually available, and a large number of different types of unlabeled remote sensing image data are also relatively easy to obtain. Therefore, whether the remote sensing image data sets of different modes can be used to improve the semi-supervised model performance is a problem worth studying. To verify the generality of the proposed method, we designed a set of semi-supervised experiments combining STAKI dataset and WHU dataset.

The experimental results are shown in TABLE VI. First, we use supervised models as the baseline, which only use partially labeled images as the training set. Then, similar to general semi-supervised experiments, we used partial labeled STAKI images and the remaining unlabeled STAKI images to train the semi-supervised change detection model. It can be seen from the table that when the semi-supervised model is trained with labeled images of different proportions, the model performance can be significantly improved by the unlabeled images. It should be noted that when all image-pairs are labeled, our semi-supervised approach also improves performance due to the contrastive loss.

In order to verify that the proposed semi-supervised method uses unlabeled large-scale datasets to improve the model of small datasets, we use part of the image-pairs in SZTAKI dataset as labeled data and 7120 image-pairs in WHU dataset as unlabeled data to train the semi-supervised model. The experimental results in TABLE VI show that the addition of WHU data can significantly improve the model performance when only 10%, 20% and 40% SZTAKI data are used. Since the pre-training model is more accurate when there has 40% labeled data, the pseudo-labels generated by WHU data can also be used more effectively. Finally, when all SZTAKI images are used as labeled images and 7120 pairs of WHU images are used as unlabeled images, the semi-supervised method can also increase the detection model IoU by 4.76 percentage points. These results prove that the proposed semi-supervised change detection method effectively use large-scale unlabeled remote sensing images to improve the change

TABLE VII
TIME COMPLEXITY ANALYSIS OF DIFFERENT SEMI-SUPERVISED METHODS.

Methods	Params	Train times	cIoU(%)	OA(%)
Sup [47]	46.85 M	0.5 h	65.73	92.08
RCR [29]	50.69 M	7.1 h	76.65	95.20
Ours. pre	46.85 M	1.0 h	70.05	93.60
Ours	46.85 M	6.0 h	79.32	95.67

detection model performance, which also has very important research significance.

Time Complexity. Although the proposed reliable contrastive learning has significantly improved the performance of the semi-supervised change detection model, we also need to further analyze the model parameter variation and time complexity. In order to directly compare the effects of different semi-supervised methods on the model, we used the same change detection network to train the model on the WHU dataset. All experiments were carried out on a 1080Ti GPU, and different models were trained the same epochs. The comparison of experimental results is shown in TABLE VII. As can be seen from the table, the supervised model has the least number of parameters and the fastest training time, but the model accuracy is also the lowest.

Combined with the proposed contrastive loss, the performance of our pre-trained model is significantly improved, and the training time is also slightly increased due to the contrastive loss and the addition of positive and negative samples. When comparing our semi-supervised model with related semi-supervised methods, it can be seen from the table that our method does not add additional model parameters, but is more efficient in training time and the model performance is improved more obviously. These results prove that the proposed semi-supervised method in this paper has a more efficient training process and the performance of the final model is also better than other related methods.

VI. CONCLUSION

Remote sensing image change detection methods usually need a large number of labeled images for model training, but the labeling of bi-temporal remote sensing images usually consumes huge resources. In order to make full use of unlabeled remote sensing image data, this paper proposes a reliable contrastive learning method for semi-supervised change detection. The contrastive loss combines the task characteristics of change detection, and the positive and negative pixels are designed according to the labels or pseudo labels. This loss effectively improves the feature identification ability of the model. In addition, selecting reliable data from unlabeled data to generate pseudo-labels, and then adding them to the training of the semi-supervised model can further improve the performance of the detection model. Extensive experimental results demonstrate the effectiveness of the proposed method.

In the future, we will try to combine different training stages to complete semi-supervised model training more efficiently through real-time reliability calculation and pseudo-label generation. In addition, in the selection of unlabeled data, we will further explore the applicability of different types of remote sensing images.

REFERENCES

[1] J. Z. Xu, W. Lu, Z. Li, P. Khaitan, and V. Zaytseva, "Building damage detection in satellite imagery using convolutional neural networks," *arXiv preprint arXiv:1910.06444*, 2019.

[2] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sensing*, vol. 12, no. 10, 2020.

[3] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Trans. Geosci. Remote. Sens.*, vol. 57, no. 1, pp. 574–586, 2019.

[4] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Transactions on Geoscience and Remote sensing*, vol. 38, no. 3, pp. 1171–1182, 2000.

[5] J. Deng, K. Wang, Y. Deng, and G. Qi, "Pca-based land-use change detection and analysis using multitemporal and multisensor satellite data," *International Journal of Remote Sensing*, vol. 29, no. 16, pp. 4823–4838, 2008.

[6] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (mad) and maf postprocessing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sensing of Environment*, vol. 64, no. 1, pp. 1–19, 1998.

[7] C. Wu, B. Du, and L. Zhang, "Slow feature analysis for change detection in multispectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2858–2874, 2013.

[8] B. Cui, Y. Zhang, L. Yan, J. Wei, and H. Wu, "An unsupervised sar change detection method based on stochastic subspace ensemble learning," *Remote Sensing*, vol. 11, no. 11, p. 1314, 2019.

[9] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional siamese networks for change detection," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 4063–4067.

[10] M. Gong, X. Niu, P. Zhang, and Z. Li, "Generative adversarial networks for change detection in multispectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2310–2314, 2017.

[11] P. F. Alcantarilla, S. Stent, G. Ros, R. Arroyo, and R. Gherardi, "Street-view change detection with deconvolutional networks," *Autonomous Robots*, vol. 42, no. 7, pp. 1301–1322, 2018.

[12] X. Peng, R. Zhong, Z. Li, and Q. Li, "Optical remote sensing image change detection based on attention mechanism and image difference," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 9, pp. 7296–7307, 2020.

[13] C. Zhang, P. Yue, D. Tapete, L. Jiang, B. Shangguan, L. Huang, and G. Liu, "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 183–200, 2020.

[14] S. Fang, K. Li, J. Shao, and Z. Li, "Snunet-cd: A densely connected siamese network for change detection of vhr images," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[15] J. Chen, Z. Yuan, J. Peng, L. Chen, H. Huang, J. Zhu, Y. Liu, and H. Li, "Dasnet: Dual attentive fully convolutional siamese networks for change detection in high-resolution satellite images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1194–1206, 2020.

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[17] Y. Zhang, L. Fu, Y. Li, and Y. Zhang, "Hdfnet: Hierarchical dynamic fusion network for change detection in optical aerial images," *Remote Sensing*, vol. 13, no. 8, p. 1440, 2021.

[18] T. Miyato, S.-i. Maeda, M. Koyama, and S. Ishii, "Virtual adversarial training: a regularization method for supervised and semi-supervised learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 8, pp. 1979–1993, 2018.

[19] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *Advances in neural information processing systems*, vol. 30, 2017.

[20] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," *Advances in neural information processing systems*, vol. 33, pp. 596–608, 2020.

[21] Q. Xie, Z. Dai, E. Hovy, T. Luong, and Q. Le, "Unsupervised data augmentation for consistency training," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6256–6268, 2020.

[22] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," *Advances in neural information processing systems*, vol. 32, 2019.

[23] L. Yang, W. Zhuo, L. Qi, Y. Shi, and Y. Gao, "St++: Make self-training work better for semi-supervised semantic segmentation," in *CVPR*, 2022.

[24] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6023–6032.

[25] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2613–2622.

[26] Y. Chen and L. Bruzzone, "Self-supervised sar-optical data fusion of sentinel-1/-2 images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.

[27] Y. Zhang, S. Zhang, Y. Li, and Y. Zhang, "Coarse-to-fine satellite images change detection framework via boundary-aware attentive network," *Sensors*, vol. 20, no. 23, p. 6735, 2020.

[28] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

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

[30] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 18 661–18 673, 2020.

[31] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.

[32] M. Roy, S. Ghosh, and A. Ghosh, "A novel approach for change detection of remotely sensed images using semi-supervised multiple classifier system," *Information Sciences*, vol. 269, pp. 35–47, 2014.

[33] P. Shao, W. Shi, P. He, M. Hao, and X. Zhang, "Novel approach to unsupervised change detection based on a robust semi-supervised fcm clustering algorithm," *Remote Sensing*, vol. 8, no. 3, p. 264, 2016.

[34] H.-C. Li, G. Yang, W. Yang, Q. Du, and W. J. Emery, "Deep nonsmooth nonnegative matrix factorization network with semi-supervised learning for sar image change detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 160, pp. 167–179, 2020.

[35] L. Tu, J. Li, and X. Huang, "High-resolution land cover change detection using low-resolution labels via a semi-supervised deep learning approach-2021 ieee data fusion contest track msd," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE, 2021, pp. 2058–2061.

[36] J. Wang, C. HQ Ding, S. Chen, C. He, and B. Luo, "Semi-supervised remote sensing image semantic segmentation via consistency regularization and average update of pseudo-label," *Remote Sensing*, vol. 12, no. 21, p. 3603, 2020.

[37] J.-X. Wang, S.-B. Chen, C. H. Ding, J. Tang, and B. Luo, "Ranpaste: Paste consistency and pseudo label for semisupervised remote sensing image semantic segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2021.

[38] S. Ghosh, M. Roy, and A. Ghosh, "Semi-supervised change detection using modified self-organizing feature map neural network," *Applied Soft Computing*, vol. 15, pp. 1–20, 2014.

[39] Y. Yuan, H. Lv, and X. Lu, "Semi-supervised change detection method for multi-temporal hyperspectral images," *Neurocomputing*, vol. 148, pp. 363–375, 2015.

[40] M. Gong, Y. Yang, T. Zhan, X. Niu, and S. Li, "A generative discriminatory classified network for change detection in multispectral imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 1, pp. 321–333, 2019.

[41] D. Peng, L. Bruzzone, Y. Zhang, H. Guan, H. Ding, and X. Huang, "Semicdnet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5891–5906, 2020.

[42] K. P. Murphy, *Machine learning - a probabilistic perspective*, ser. Adaptive computation and machine learning series. MIT Press, 2012.

[43] L. Yang, W. Zhuo, L. Qi, Y. Shi, and Y. Gao, "St++: Make self-training work better for semi-supervised semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 4268–4277.

[44] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *arXiv preprint arXiv:1708.04552*, 2017.

[45] C. Benedek and T. Szirányi, "Change detection in optical aerial images by a multilayer conditional mixed markov model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 10, pp. 3416–3430, 2009.

[46] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 574–586, 2018.

[47] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021.

[48] T.-H. Vu, H. Jain, M. Bucher, M. Cord, and P. Pérez, "Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2517–2526.

[49] S. Mittal, M. Tatarchenko, and T. Brox, "Semi-supervised semantic segmentation with high-and low-level consistency," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 4, pp. 1369–1379, 2019.