Joint Self-Training and Rebalanced Consistency Learning for Semi-Supervised Change Detection

Xueting Zhang, Xin Huang, *Senior Member, IEEE*, and Jiayi Li, *Senior Member, IEEE*

***Abstract*— Change detection (CD) is an important Earth observation task that can monitor change areas at two times from the view of space. However, fully supervised CD has a heavy dependence on numerous manually labeled data, limiting their applications in practice. Beyond the fully supervised setting, semi-supervised CD (SSCD), which uses a few labeled data to guide the unsupervised learning of dominant unlabeled data, has attracted increasing attention for its significant advantage in alleviating the demand for annotations. To this end, in this article we propose a joint self-training and rebalanced consistency learning (ST-RCL) framework for SSCD, which consists of a basic supervised branch for the labeled data and a novel unsupervised branch for the unlabeled data. To make full use of the unlabeled data, the unsupervised branch generates pseudolabels from weakly augmented unlabeled remote sensing image (RSI) pairs to supervise the CD of two strongly augmented counterparts, including an unrotated version and a rotated version. On one hand, the unrotated unlabeled RSI pairs are pseudosupervised with the pseudolabels by confidence-based self-training (ST). On the other hand, to further enhance model robustness to rotation nonequivariance and imbalanced distribution, the predictions of rotated unlabeled RSI pairs are aligned to the pseudolabels by a well-designed rebalanced consistency learning (RCL) strategy based on uncertainty-based class weighting. Extensive experiments are performed on four widely used CD datasets, and the proposed ST-RCL yields new state-of-the-art results on all these datasets in comparison to some other SSCD methods, demonstrating its effectiveness and generalization. Our code will be available at https://github.com/zxt9/STRCL-SSCD.**

***Index Terms*— Change detection (CD), rebalanced consistency learning (RCL), remote sensing, self-training (ST), semi-supervised learning.**

I. Introduction

CHANGE detection (CD) is an important Earth observation task of quantitatively analyzing and identifying the change areas within the same land surface at two different times between a pair of bitemporal remote sensing images (RSIs). CD can further support a wide range of downstream tasks, such as urban development analysis [1], [2], [3], disaster detection [4], [5], [6], and environment monitoring [7], [8], [9].

Manuscript received 22 May 2023; revised 29 July 2023; accepted 27 August 2023. Date of publication 12 September 2023; date of current version 25 September 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 41971295, Grant 42271328, and Grant 42071311; and in part by the Special Fund of Hubei Luojia Laboratory under Grant 220100031. *(Corresponding author: Xin Huang.)*

Xueting Zhang and Xin Huang are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: zhangxueting@whu.edu.cn; xhuang@whu.edu.cn).

Jiayi Li is with the School of Remote Sensing and Information Engineering and the Luojia Laboratory, Wuhan University, Wuhan 430079, China (e-mail: zjjerica@whu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2023.3314452

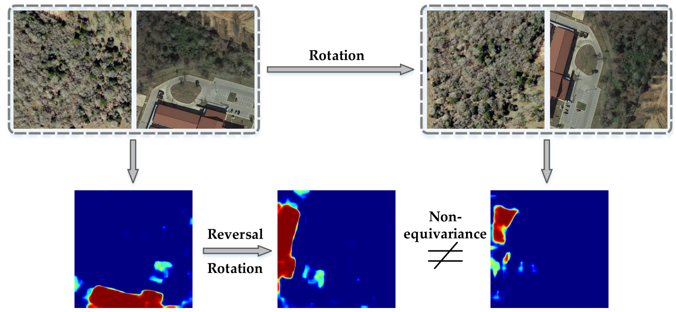


Fig. 1. CD inconsistency between unrotated and rotated RSI pairs.

The state-of-the-art CD methods are based on fully supervised deep neural networks, especially convolutional neural networks (CNNs), which have a huge demand for manually labeled change annotations at the pixel level between bitemporal images. However, it is hard and even impossible to obtain numerous change labels in practice, limiting its fast applications in real scenarios. To mitigate the demand for annotations, a promising task of semi-supervised change detection (SSCD) has attracted increasing attention, which can make full use of abundant unlabeled data for model training under the guidance of only a few labeled data.

The current SSCD methods attempt to leverage the unlabeled data for training roughly from: 1) adversarial training [10] that reduces the difference in change feature representation between the labeled and unlabeled data, which focuses on image-level alignment in an implicit manner and 2) self-training (ST)/consistency learning of the unlabeled data on model optimization that makes consistent predictions between clean unlabeled data and a perturbed version in an explicit manner [11], [12], [13].

Although the existing SSCD methods have achieved significant progress, some problems still exist with the unsupervised learning of the unlabeled data. In this article, we identify two underlying problems of SSCD: 1) rotation nonequivariance between the unrotated bitemporal RSI pair and its rotated version as shown in Fig. 1, which inherits from the characteristics of deep neural networks, especially CNNs [14], [15], [16]. The nonequivariance between unrotated and rotated bitemporal RSI pairs is harmful to model stability in angle-unfixed change predictions and 2) imbalanced distribution of change and nonchange, which generates imbalanced pseudolabels and their pseudosupervised training on the CD model exacerbates the imbalanced distribution in return. Unfortunately, such progress is normally irreversible because of a common phenomenon in semi-supervised learning, i.e., confirmation bias [17].

Aiming at these two problems, we propose a joint ST and rebalanced consistency learning (ST-RCL) framework for SSCD. The proposed ST-RCL generally consists of two branches, one basic supervised branch and one novel unsupervised branch, which share an encoder–decoder-based CD model. The supervised branch pays attention to the prediction of labeled RSI pairs, which can help the model learn the preliminary CD ability. Furthermore, built upon a strong-to-weak paradigm of FixMatch [18], the unsupervised branch contains an effective unsupervised dual-path workflow of unlabeled data on the CD model. In detail, the pseudolabels are generated from a weakly augmented unlabeled RSI pair, and then they are used to supervise the change prediction of two strongly augmented counterparts of the same unlabeled RSI pair, including an unrotated RSI pair and a rotated RSI pair. On one hand, the unrotated RSI pair is pseudosupervised with the pseudolabel-based confidence-based ST, which only selects the high-confidence pixels for training. As a result, ST can prompt the model to learn augmentation-invariant feature representation and change prediction, which are beneficial for model generalization. On the other hand, to enhance model robustness to the above-mentioned first problem of rotation nonequivariance, the predictions of the rotated RSI pair are kept consistent with the pseudolabels from the unrotated weakly augmented RSI, which can reduce rotation inconsistency of the CD model. Taking into consideration the second problem of imbalanced distribution, we design a corresponding rebalanced consistency learning (RCL) strategy to assign adaptive classwise weights to the consistency learning loss between the unrotated and rotated unlabeled RSI pairs, which are calculated based on class-aware uncertainty during training.

To evaluate the effectiveness and robustness of the proposed ST-RCL framework to SSCD, extensive ablation studies, comparison experiments, and visualization experiments are performed on four commonly used CD datasets in the semi-supervised setting, including CDD, WHU-CD, LEVIR-CD, and GZ-CD under a wide range of labeled ratios of 5%, 10%, 20%, and 40%. The proposed ST-RCL outperforms other SSCD methods and yields most of the state-of-the-art results on all the datasets.

Overall, our contributions can be summarized as follows.

1) We identify two underlying problems of SSCD, rotation nonequivariance and imbalanced distribution, which limit the further performance improvement of this field. To solve the two problems, we propose a novel joint ST-RCL framework for SSCD.
2) The proposed ST-RCL consists of one supervised branch for labeled data and one unsupervised branch for unlabeled data. In particular, the unsupervised branch simultaneously involves consistency learning of rotated unlabeled data for reducing the rotation inconsistency and uncertainty-based classwise weighting for alleviating the imbalanced distribution.
3) In comparison to other SSCD methods, the proposed ST-RCL method achieves most of the new state-of-the-art results in four commonly used CD datasets under various labeled ratios, demonstrating its effectiveness and robustness.

II. Related Work

A. Semi-Supervised Change Detection

Fully supervised CD [19], [20], [21] typically requires numerous labeled data for model training, which is laborious and time-consuming. To alleviate this high dependence on annotations, recently some semi-supervised work has been developed [11], [22], [23], [24]. For example, by integrating extreme learning machine (ELM), [25] presented a deep nonsmooth nonnegative matrix factorization (nsNMF) network for synthetic aperture radar (SAR) image CD in the semi-supervised setting. The learning process of nsNMF model contains two stages, i.e., pretraining and fine-tuning, where the labeled and unlabeled data can be used for training. Peng et al. [10] proposed a semi-supervised CD network (SemiCDNet) based on a UNet++ segmentation network. By adopting entropy adversarial learning (EAL) and segmentation adversarial learning, two discriminators are designed to decrease uncertain predictions of the change maps of the unlabeled samples and encourage the consistency of the segmentation-predicted feature distribution, respectively. Bandara and Patel [11] developed a consistency-based regularization CD model. In this method, different random perturbations are added to the feature difference map, and then the corresponding predicted change probability maps are constrained to be consistent with the one without any perturbation. In [12], a Siamese nested UNet with graph attention mechanism (SANet) was designed for SSCD, where strong augmentation and consistency regularization are introduced to achieve the consistency between predictions of distorted images and the generated pseudolabels with high confidence. Shu et al. [26] devised a multitask consistency network (MTCNet) for building SSCD, where the building segmentation information is introduced for features constraint in the encoder. Using extra T1 labels, MTCNet is able to keep the consistency of the predictions between the original T1 branch and the Diff-T1 branch, and thus improving the generalization ability of the whole framework. Wang et al. [13] proposed a reliable contrastive learning-based method for SSCD. In their case, the uncertainty of unlabeled data is used for the selection of reliable pseudolabels to train the model. Then a contrastive loss is designed for feature identification ability of the model, where the positive and negative samples are selected for the changed and unchanged regions. It is worth noting that most SSCD methods pay more attention to the design of the training optimization strategy instead of the architecture of the model, which allows flexible embedding into various segmentation models.

B. Rotation Nonequivariance

Rotation nonequivariance of CNNs is a challenging issue in the image process field, especially in the era of deep learning [27], [28], [29]. Recent advanced studies mainly focus on devising rotation-invariant CNNs to acquire the ability of rotation-equivariant learning, such as group equivariant CNNs (G-CNNs) [14], [30], [31], steerable CNNs [15], [32], [33], and harmonic CNNs [34], [35]. For instance, Dieleman et al. [36] introduced four operations of layers into neural network models to build up an architecture of building rotation equivariance. At the same time, Cohen and Welling [14] first proposed G-CNNs to achieve equivariant representations for large groups of symmetries, the functions of which are feature maps obtained from the transformed filters. Worrall et al. [34] designed harmonic networks (H-Nets) for CNN equivariance to rotation and patchwise translation. Particularly, regular CNN filters are replaced by circular harmonics, and thus, a maximal response and orientation can be acquired for receptive field patches. In [15], a general architecture of steerable CNNs termed as E(2)-equivariant convolutions was provided to build up a general theory of equivariance. On the basis of Geodesic ICOsahedral Pixelation (GICOPix), Yang et al. [37] developed a spherical graph convolutional network (SGCN) containing the transition layer and hierarchical pooling operator. Then, a rotation-invariant aerial object detection network (RINet) was presented by [16] in a weakly supervised manner. This method aims to solve the problems of instance missing and rotation sensitivity from the perspective of multi-instance mining and rotation-invariant learning.

C. Imbalanced Distribution in Remote Sensing

In the past several decades, various remote sensing algorithms have been designed to acquire the ability of the feature representation from the data samples on an even distribution assumption [38], [39], [40], [41]. Nevertheless, in a real-world setting, a small number of classes typically have numerous samples while there is only a small portion of samples belonging to the other classes [42], [43], [44]. Consequently, this class-imbalanced phenomenon would in return exacerbate model bias to the majority classes during training [45], [46], which is known as imbalanced distribution. As a common yet practical problem, imbalanced distribution can be easily observed in different remote sensing fields, such as scene classification [47], [48], [49], CD [50], [51], [52], semantic segmentation [53], [54], and object detection [55], [56]. Recently, some attempts have been made to mitigate this problem. For example, a multigranularity decoupling network called MGDNet is presented by Miao et al. [48] for the class-imbalanced scene classification of RSI. In this work, a Gaussian mixture model (GMM) is fit by the per-sample loss distribution of imbalanced data for the high confidence pseudolabels' selection. Normally, the change area in the real scene is rarely and sparsely distributed between bitemporal image pairs. Chen et al. [52] aim to mitigate this phenomenon by generating new synthesized CD samples using generative adversarial network (GAN)-based and image blending techniques. According to the number of effective samples, Zhou et al. [54] propose a dynamic effective class-balance (DECB) weighting method for the class-imbalance problem in remote sensing semantic segmentation. In [55], an adaptive balanced network (ABNet) is developed for remote sensing object detection by introducing some additional components on the basis of Faster RCNN [57].

In this study, we concentrate on alleviating the imbalanced distribution problem of the CD task in the semi-supervised setting from the perspective of optimization without model parameter increase at the test phase.

III. Methodology

This section starts with the introduction of some notations and settings of SSCD, then describes the encoder–decoder model shared among all kinds of RSI pairs, and finally formulates the supervised branch and unsupervised branch of the proposed ST-RCL architecture with the summary of the whole training procedure. The overall ST-RCL workflow is depicted in Fig. 2.

A. SSCD Settings

To better describe SSCD, some definitions and notations are first given in this section. At the training stage, the whole training set is made up of two subsets, including a labeled training set and an unlabeled training set. Here, the labeled trained set is represented as $\mathcal{S}_l = \{\mathbf{x}_a^l, \mathbf{x}_b^l, \mathbf{y}_i^l\}_{i=1}^M$, where $\{\mathbf{x}_a^l, \mathbf{x}_b^l\}$ is the ith labeled RSI pair consisting of a pretemporal image $\mathbf{x}_a^l$ and a posttemporal image $\mathbf{x}_b^l$, and $\mathbf{y}_i^l$ is their change/nonchange labels at the pixel level. M denotes the total number of labeled RSI pairs. Besides, the unlabeled data are represented as $\mathcal{S}_u = \{\mathbf{x}_a^u, \mathbf{x}_b^u\}_{i=1}^N$, where $\{\mathbf{x}_a^u, \mathbf{x}_b^u\}$ represents the ith unlabeled RSI pair with no labels of the unlabeled training set and N denotes the number of the RSI pairs of this set. In the SSCD setting, M is smaller than N.

B. Shared Encoder–Decoder Model

The widely used encoder–decoder architecture is used as the CD model. In this study, it consists of a shared encoder denoted as $\mathcal{E}$, a pyramid pooling module (PPM in abbreviation), and a decoder denoted as $\mathcal{G}$. The encoder extracts a cross-temporal change feature map from an RSI pair, the PPM fuses the multiscale bitemporal features, and the decoder generates a corresponding pixelwise change prediction map.

First, a bitemporal RSI pair $\{\mathbf{x}_a, \mathbf{x}_b\}$ in the same dimension of $\mathbb{R}^{H\times W\times 3}$ (H and W are, respectively, spatial height and width) is sent into the encoder $\mathcal{E}$. The Siamese architecture ([11], [58]) (i.e., a shared CNN backbone) is used as the encoder to obtain the respective feature maps from $\{\mathbf{x}_a, \mathbf{x}_b\}$ as

$$\mathbf{f}_a = \mathcal{E}(\mathbf{x}_a), \quad \mathbf{f}_b = \mathcal{E}(\mathbf{x}_b) \tag{1}$$

where $\mathbf{f}_a \in \mathbb{R}^{H/s\times W/s\times C}$ and $\mathbf{f}_b \in \mathbb{R}^{H/s\times W/s\times C}$ are the prechange feature map from $\mathbf{x}_a$ and the postchange feature map from $\mathbf{x}_b$, respectively. Here, C denotes the feature dimension, and s is the spatial scaling ratio, which depends on the used backbones.

Next, the absolute difference map of $\mathbf{f}_a$ and $\mathbf{f}_b$ is calculated for avoiding the interference of the time order of $\mathbf{x}_a$ and $\mathbf{x}_b$. On top of it, PPM further transfers the absolute difference map into a corresponding change feature map $\mathbf{f}$, which contains

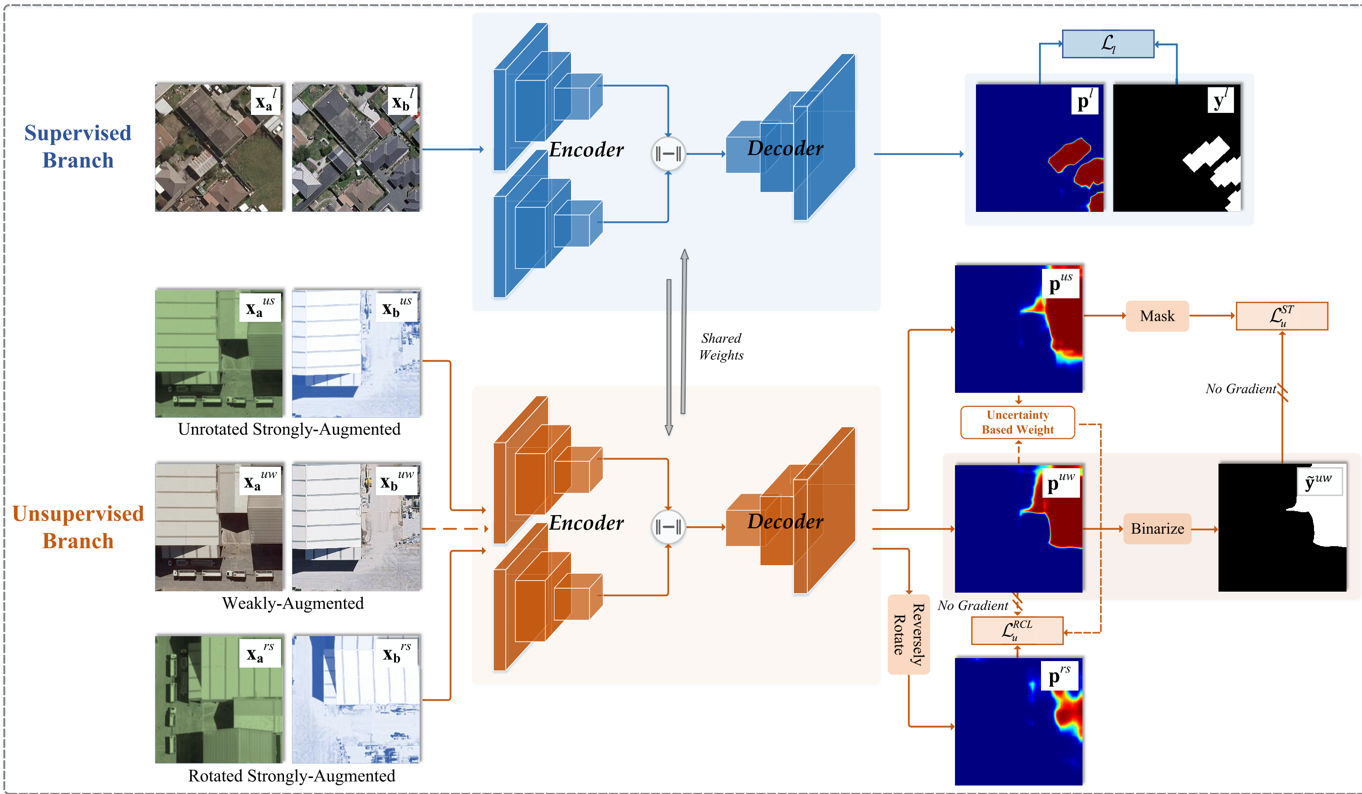


Fig. 2. Illustration of the proposed ST-RCL for SSCD. All the types of RSI data share the same encoder–decoder-based CD model.

both high-level semantic features and low-level texture/color features. The above operations are formulated as

$$\mathbf{f} = \mathrm{PPM}(|\mathbf{f}_a - \mathbf{f}_b|). \tag{2}$$

Finally, a convolutional upsampling module [59] is used as the decoder $\mathcal{G}$ to decode and upsample the bitemporal feature map $\mathbf{f}$ into a pixel-level change probability map, which is denoted as $\mathbf{p} \in \mathbb{R}^{H\times W\times 2}$. Here, 2 means the two classes of "change" and "nonchange." It is formulated as

$$\mathbf{p} = \mathcal{G}(\mathbf{f}). \tag{3}$$

The probability sum of the prediction at location $[i, j]$, $\mathbf{p}_{ij}$, is scaled to 1 by the softmax function as

$$\mathbf{p}_{ij} = \frac{e^{\mathbf{p}_{ijk}}}{\sum_{k=1}^{2} e^{\mathbf{p}_{ijk}}} \tag{4}$$

where k represents the class index.

At the test stage, the prediction results are obtained from (1) to (4) sequentially.

C. Supervised Branch

In ST-RCL, as shown in the "Supervised Branch" part of Fig. 2, the labeled RSI pairs are directly used for supervised training that can help the shared encoder–decoder model acquire the preliminary CD ability. For an RSI pair of the labeled training set $\{\mathbf{x}_a^l, \mathbf{x}_b^l, \mathbf{y}^l\}$, the weakly augmented bitemporal RSI pair $\{\mathbf{x}_a^l, \mathbf{x}_b^l\}$ is fed into the encoder–decoder-based CD model, which generates a pixelwise change prediction map $\mathbf{p}^l$ as (1)–(3). The cross-entropy (CE) loss serves as the supervision loss $\mathcal{L}_l$ to reduce the gap of the prediction map $\mathbf{p}^l$ and the associated label map $\mathbf{y}^l$ as

$$\mathcal{L}_l = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\ell_{\mathrm{ce}}\left(\mathbf{p}_{ij}^l, \mathbf{y}_{ij}^l\right). \tag{5}$$

D. Unsupervised Branch

As shown in the middle-down part of Fig. 2, the "Unsupervised Branch" further achieves two kinds of unsupervised learning of unlabeled RSI pairs, including the ST of unrotated RSI pairs and uncertainty-based RCL of rotated RSI pairs.

1) ST of Unrotated RSI Pairs: In the unlabeled branch, first, an unlabeled RSI pair is randomly sampled from the unlabeled training set and weakly augmented by weak augmentations given in Section IV-C, which is denoted as $\{\mathbf{x}_a^{uw}, \mathbf{x}_b^{uw}\}$. It is worth noting that $\{\mathbf{x}_a^{uw}, \mathbf{x}_b^{uw}\}$ share the same weak augmentations to ensure their spatial and semantic consistency. On top of the weak augmentations, the RSI pair is further augmented to a strongly augmented pair $\{\mathbf{x}_a^{us}, \mathbf{x}_b^{us}\}$, formulated as

$$\mathbf{x}_a^{us} = \mathcal{A}\left(\mathbf{x}_a^{uw}\right), \quad \mathbf{x}_b^{us} = \mathcal{A}\left(\mathbf{x}_b^{uw}\right) \tag{6}$$

where $\mathcal{A}$ represents two connected spatial-irrelevant strong augmentations randomly selected from a predefined augmentation list of RandAugment [60]. Unlike weak augmentations, each of $\{\mathbf{x}_a^{us}, \mathbf{x}_b^{us}\}$ is individually strongly-augmented to increase their diversity.

The weakly augmented prediction map $\mathbf{p}^{uw}$ is extracted from $\{\mathbf{x}_a{}^{uw}, \mathbf{x}_b{}^{uw}\}$ by (1)–(4). Similarly, the corresponding strongly augmented prediction map $\mathbf{p}^{us}$ is obtained from $\{\mathbf{x}_a^{us}, \mathbf{x}_b^{us}\}$ also by (1)–(4).

To promote the encoder–decoder-based CD model to learn effective feature representation from the unlabeled data, the weakly augmented unlabeled RSI pairs are used to supervise the change prediction of the strongly augmented counterparts, i.e., ST. A pseudolabel map $\tilde{\mathbf{y}}^{uw} \in \mathbb{R}^{H\times W}$ is generated from the weakly augmented prediction map $\mathbf{p}^{wu}$ as

$$\tilde{\mathbf{y}}_{ij}^{uw} = \arg\max_{k=\{0,1\}} \ \mathbf{p}_{ijk}^{uw} \tag{7}$$

where $\tilde{\mathbf{y}}^{uw}$ denotes the maximum-activated class index (i.e., pseudolabel) at the location $[i, j]$.

The ST loss between the weakly augmented unlabeled RSI pair and the (unrotated) strongly augmented unlabeled RSI pair is calculated by confidence-based CE loss, which is formulated as

$$\mathcal{L}_u^{\mathrm{ST}} = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} \ell_{\mathrm{ce}}\left(\mathbf{p}_{ij}^{us}, \tilde{\mathbf{y}}_{ij}^{uw}\right)\cdot \mathbb{1}\left(\max\left(\mathbf{p}_{ij}^{uw}\right) > \tau\right) \tag{8}$$

where $\mathbb{1}(\cdot)$ is an indicator function that only allows the high-confidence pixels for ST to reduce the impact of noisy pseudolabels. For the pixel of the position $[i, j]$, the value of $\mathbb{1}(\max(\mathbf{p}_{ij}^{uw}) > \tau)$ is equal to 1 when the probability of the max-activated class at this position is bigger than a threshold τ, otherwise, it would be 0.

2) Uncertainty-Based RCL of Rotated RSI Pairs: Then a rotated strongly augmented RSI pair $\{\mathbf{x}_a^{rs}, \mathbf{x}_b^{rs}\}$ is obtained from the unrotated strongly augmented RSI pair $\{\mathbf{x}_a^{us}, \mathbf{x}_b^{us}\}$ as

$$\mathbf{x}_a^{rs} = \mathcal{R}\left(\mathbf{x}_a^{us}\right), \quad \mathbf{x}_b^{rs} = \mathcal{R}\left(\mathbf{x}_b^{us}\right) \tag{9}$$

where $\mathcal{R}$ is the random rotation operation with an angle r randomly selected from a list of [90°, 180°, 270°, and 360°]. Here, 360° equals 0°, that is, no rotation. Furthermore, by (1)–(4), the rotated prediction map $\mathbf{p}^{rs}$ can be obtained from $\{\mathbf{x}_a^{rs}, \mathbf{x}_b^{rs}\}$.

To keep consistent predictions of unrotated and rotated RSI pairs, inspired by the idea of consistency learning and strong-to-weak alignment, we design the first version of the consistency learning loss between the rotated strongly augmented prediction map $\mathbf{p}^{rs}$ and the unrotated weakly augmented prediction map $\mathbf{p}^{wu}$ as

$$\mathcal{L}_u^{\mathrm{CL}} = \sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{k=1}^{2}\left|\mathbf{p}_{ijk}^{uw} - \hat{\mathcal{R}}\left(\mathbf{p}_{ijk}^{rs}\right)\right| \tag{10}$$

where $\hat{\mathcal{R}}$ is the reversed rotation of $\mathcal{R}$, which applies the rotation operation with a reverse angle, $360 - r$. Here, L1 loss is used to measure the pixelwise distance. It is worth noting that there is no gradient backward for $\mathbf{p}^{uw}$, which means $\mathbf{p}^{uw}$ serves as the pseudo-ground-truth of $\mathbf{p}^{rs}$.

To further take into consideration the imbalanced distribution between "change" and "non-change," an uncertainty-based classwise weighting strategy is devised to rebalance the training process. We assume that the uncertainty of a class during training indicates its convergence difficulty, and the higher the uncertainty, the harder the convergence difficulty. Based on this assumption, the principle of RCL is to give a higher weight to the class with higher uncertainty during the training stage, which probably is "change."

The classwise uncertainty is denoted as $u \in \mathbb{R}^2$, and it is calculated every epoch to get the dataset-level value that is more robust. The class-k uncertainty is calculated as

$$u_k = \frac{\sum_{l=1}^{N}\sum_{i=1}^{H}\sum_{j=1}^{W}\left|\mathbf{p}_{lijk}^{uw} - \mathbf{p}_{lijk}^{us}\right|\cdot \mathbb{1}\left(\tilde{\mathbf{y}}_{lij}^{u} = k\right)}{\sum_{l=1}^{N}\sum_{i=1}^{H}\sum_{j=1}^{W}\mathbb{1}\left(\tilde{\mathbf{y}}_{lij}^{u} = k\right)} \tag{11}$$

which means the class-k uncertainty is the dataset-level average value of the absolute difference between weakly augmented predictions $\mathbf{p}^{uw}$ and unrotated strongly augmented predictions $\mathbf{p}^{us}$ when their pseudolabels are k. For the first epoch, the initial value of u would be set to 0. Besides, it is worth mentioning that there could be other uncertainty calculation methods as long as they can reasonably measure the class convergence difficulty.

Then, the uncertainty-based class weight, denoted as $w \in \mathbb{R}^2$, can be calculated as

$$w_k = 1 + \lambda u_k \tag{12}$$

where λ is the weight coefficient of w. Corresponding to u, w is calculated every epoch.

Finally, given $\mathbf{p}^{uw}$, $\mathbf{p}^{rs}$, and w, the RCL loss is formulated as

$$\mathcal{L}_u^{\mathrm{RCL}} = \sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{k=1}^{2}\left|\mathbf{p}_{ijk}^{uw} - \hat{\mathcal{R}}\left(\mathbf{p}_{ijk}^{rs}\right)\right|\cdot w_k. \tag{13}$$

Algorithm 1 Training Procedure of ST-RCL

Input: labeled set $\mathcal{S}_l = \{\mathbf{x}_a^l, \mathbf{x}_b^l, \mathbf{y}_i^l\}_{i=1}^{M}$, unlabeled set $\mathcal{S}_u = \{\mathbf{x}_a^u, \mathbf{x}_b^u\}_{i=1}^{N}$, shared encoder–decoder $\mathcal{E} - \mathcal{G}$, epoch number N_E, iteration number N_I and batch size B, class uncertainty u and class weight w

for *epoch* $\leftarrow 1$ **to** N_E **do**
- **for** *iteration* $\leftarrow 1$ **to** N_I **do**
  - **Data Processing:** sample and transform four RSI pairs $\{\mathbf{x}_a^l, \mathbf{x}_b^l, \mathbf{y}_i^l\}_{i=1}^B$, $\{\mathbf{x}_a^{uw}, \mathbf{x}_b^{uw}\}_{i=1}^B$, $\{\mathbf{x}_a^{us}, \mathbf{x}_b^{us}\}_{i=1}^B$, and $\{\mathbf{x}_a^{rs}, \mathbf{x}_b^{rs}\}_{i=1}^B$, via (6) and (9);
  - **Supervised Learning:** optimize $\mathcal{E}$-$\mathcal{G}$ by the supervised loss L_l calculated from $\{\mathbf{x}_a^l, \mathbf{x}_b^l, \mathbf{y}_i^l\}_{i=1}^B$ via (1)–(5);
  - **Self-Training:** optimize $\mathcal{E}$-$\mathcal{G}$ by the ST loss L_u^{ST} calculated between $\{\mathbf{x}_a^{uw}, \mathbf{x}_b^{uw}\}_{i=1}^B$ and $\{\mathbf{x}_a^{us}, \mathbf{x}_b^{us}\}_{i=1}^B$ via (1)–(4) and (7)–(8);
  - **Rebalanced Consistency Learning:** optimize $\mathcal{E}$-$\mathcal{G}$ by the RCL loss L_u^{RCL} calculated between $\{\mathbf{x}_a^{uw}, \mathbf{x}_b^{uw}\}_{i=1}^B$ and $\{\mathbf{x}_a^{rs}, \mathbf{x}_b^{rs}\}_{i=1}^B$ with the class weight w, via (1)–(4) and (11)–(13);
  - **Accumulation of** u: accumulate the numerator and denominator terms of (11) for u.
- **end**
- **Calculating** u **and** w: calculate u and w via (11) and (12)

end

Output: optimized $\mathcal{E}$-$\mathcal{G}$

E. Overall Loss and Procedure of ST-RCL

The overall loss of the ST-RCL framework, denoted as $\mathcal{L}$, consists of the supervised loss $\mathcal{L}_l$, the ST loss $\mathcal{L}_u^{\mathrm{ST}}$, and the

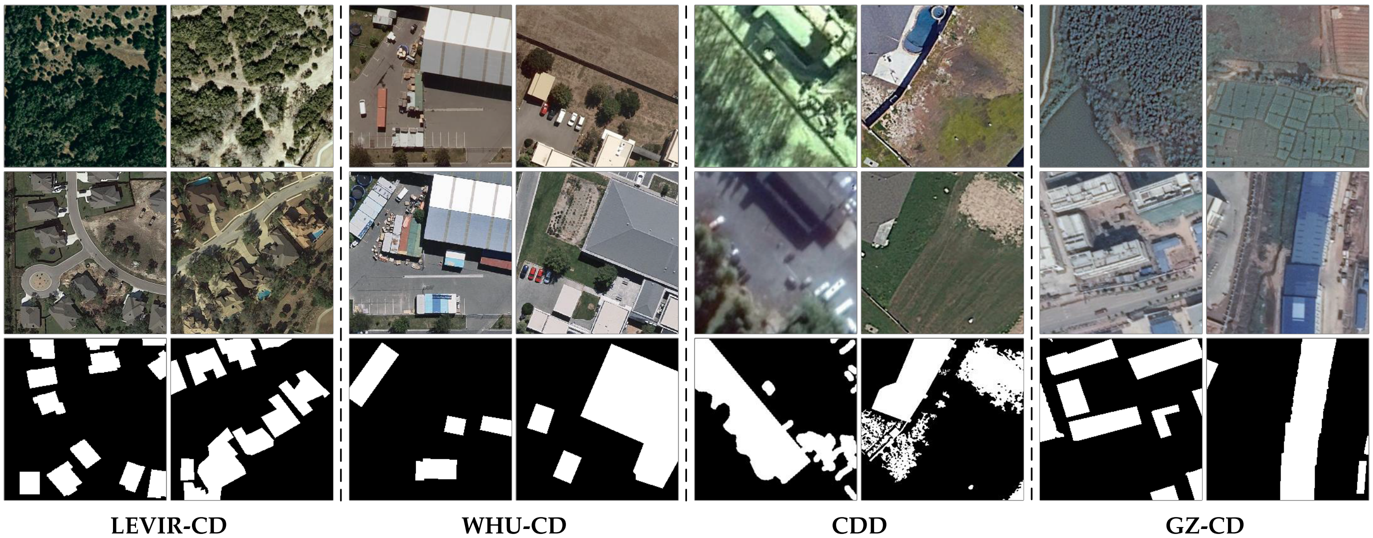


Fig. 3. Some samples of the LEVIR-CD, WHU-CD, CDD, and GZ-CD datasets. The pretemporal images, the posttemporal images, and their ground truths are, respectively, shown in the first row, the second row, and the third row.

RCL loss $\mathcal{L}_u^{\mathrm{RCL}}$. It is formulated as

$$\mathcal{L} = \mathcal{L}_l + \mathcal{L}_u^{\mathrm{ST}} + \mathcal{L}_u^{\mathrm{RCL}}. \tag{14}$$

For a clear understanding of the workflow of ST-RCL, its training procedure is summarized in Algorithm 1.

IV. Experiments

This section begins with the experimental settings, including datasets and implementation details. Then, several ablation studies are conducted to verify the necessity and effectiveness of each part of the proposed ST-RCL framework. Finally, comparison experiments on ST-RCL and other state-of-the-art methods are performed with some visualization samples for intuitive comparison.

A. Datasets

We perform the experiments on four widely used CD datasets, i.e., the Learning, Vision, and Remote Sensing Laboratory (LEVIR-CD), Wuhan University building CD (WHU-CD), CDD, and Guangzhou (GZ-CD).

1) LEVIR-CD: The LEVIR-CD dataset [51] is made up of 637 bitemporal image pairs obtained from Google Earth of different cities in Texas of USA between 2002 and 2018. These images have a spatial resolution of 0.5 m measuring 1024 × 1024 pixels, containing over 31 000 independently labeled change instances. Following [11], they are cropped into 256 × 256 patches without overlapping with a total number of 10 192 for the experiments, and 7120, 1024, and 2048 pairs of RSI patches are, respectively, used for training, validation, and testing.

2) Wuhan University Building CD: There are a pair of bitemporal RSIs with a spatial resolution of 0.075 m measuring 32 507 × 15 354 pixels in the WHU-CD dataset [61], which were, respectively, captured in Christchurch of New Zealand in 2012 and 2016. Similar to the LEVIR-CD dataset, these RSIs are also cropped into patches of 256 × 256 without overlapping. Following the common splitting, 80%, 10%, and 10% RSI pairs are, respectively, adopted as the training, validation, and test sets. Accordingly, there are 5947, 743, and 744 patches for training, validation, and testing, respectively.

3) CDD: The CDD dataset [62] is cropped from Google Earth (DigitalGlobe) of seven pairs of images with season variation measuring 4725 × 2700 pixels. It comprises 16 000 images with a spatial resolution from 0.03 to 1 m of 256 × 256 pixels. Specifically, 10 000, 3000, and 3000 pairs of RSI patches are, respectively, used for training, validation, and testing.

4) GZ-CD: This dataset [10] consists of 19 pairs of season-varying images captured in Guangzhou City, China, between 2006 and 2019. All the images measure from 1006 × 1168 pixels to 4936 × 5224 pixels with a pixel resolution of 0.55 m. For the convenience of comparison, they are similarly cropped into 256 × 256 RSI patches without overlapping. Consequently, there are, respectively, 2882, 360, and 361 pairs of RSI patches for training, validation, and testing.

Fig. 3 shows some examples of the above-mentioned datasets.

B. Evaluation Metrics

We evaluate our approach using five common CD evaluation metrics, including recall (Rec.), precision (Pre.), intersection over union (IoU), F1-score (F1), and Kappa coefficient (Kappa). The values of IoU, F1, recall, and precision are in the scope from 0 to 100%, and that of Kappa is from −1 to 1. They are formulated as

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{15}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{16}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{17}$$

$$\text{F1} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{18}$$

TABLE I
ABLATION STUDY RESULTS OF ST-RCL ON THE LEVIR-CD DATASET. THE BEST SCORES ARE MARKED IN **BOLD**

$\mathcal{L}_l$	$\mathcal{L}_u^{ST}$	$\mathcal{L}_u^{CL}$	$\mathcal{L}_u^{RCL}$	5%					10%					20%					40%				
				IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.
✓				65.55	79.19	0.7822	75.89	82.79	72.13	83.81	0.8306	80.60	87.29	75.97	86.35	0.8571	82.93	90.05	77.33	87.22	0.8662	84.45	90.17
✓	✓			75.10	85.78	0.8513	81.11	91.02	77.60	87.39	0.8680	84.72	90.23	79.19	88.39	0.8782	86.08	90.83	80.74	89.35	0.8884	87.07	91.74
✓	✓	✓		76.88	86.93	0.8633	82.93	91.33	79.37	88.50	0.8796	85.20	92.07	79.77	88.75	0.8821	87.15	90.40	81.16	89.60	0.8911	87.21	92.14
✓	✓		✓	**78.09**	**87.70**	**0.8711**	86.67	88.75	**79.77**	**88.75**	**0.8821**	86.23	91.42	**80.93**	**89.46**	**0.8896**	87.91	91.06	**81.43**	**89.77**	**0.8928**	88.11	91.49

$$\mathrm{OA} = \frac{\mathrm{TN}+\mathrm{TP}}{\mathrm{TN}+\mathrm{FP}+\mathrm{TP}+\mathrm{FN}} \tag{19}$$

$$\mathrm{PRE} = \frac{(\mathrm{FP}+\mathrm{TP})\times(\mathrm{FN}+\mathrm{TP})}{(\mathrm{TN}+\mathrm{FP}+\mathrm{TP}+\mathrm{FN})^2} + \frac{(\mathrm{TN}+\mathrm{FP})\times(\mathrm{FN}+\mathrm{TN})}{(\mathrm{TN}+\mathrm{FP}+\mathrm{TP}+\mathrm{FN})^2} \tag{20}$$

$$\mathrm{Kappa} = \frac{\mathrm{OA}-\mathrm{PRE}}{1-\mathrm{PRE}} \tag{21}$$

where TN and TP stand for the number of unchanged pixels and changed pixels that are correctly classified, respectively. In contrast, FN is the number of changed pixels that are wrongly classified as unchanged pixels, and FP is the total number of unchanged pixels that are wrongly identified as changed pixels. It is worth noting that as the focus is more on the area of change, in this work only the change class of all the used metrics is calculated and given.

C. Experimental Settings

To make a fair comparison among different methods, all the training settings are set the same as RCR [11]. In detail, the optimizer of stochastic gradient descent (SGD) is adopted for the optimization of the supervised and unsupervised losses, with its initial learning rate set to 0.01 under a weight decay of 1e−4 and a momentum of 0.9; The models of all kinds of methods are trained for 80 epochs, with the mini-batch size set to 8 for both the labeled and the unlabeled training sets; all the methods use Dilated ResNet50 + PPM as the encoder–decoder CD model; some universal weak data augmentations, including random rescaling between [0.5, 2.0], random vertical flipping, random horizontal flipping, and random cropping, are applied to each RSI pair of both labeled data for supervised learning and unlabeled data for unsupervised learning. For FixMatch [18], UniMatch [63], and the proposed ST-RCL, a strong augmentation list derived from RandAugment [60] is further used to perturb the weakly augmented unlabeled RSI pairs, containing `Identity`, `Contrast`, `Autocontrast`, `Equalize`, `Brightness`, `Color`, `Posterize`, `Sharpness`, and `Solarize`; for each weakly augmented RSI pair, only two strong augmentations randomly selected from the list are applied to avoid excessive damage to its semantic content. For ST-RCL, the rebalancing weight coefficient λ in (12) is set to 10 for LEVIR-CD, GZ-CD, and CDD, and 1 for WHU-CD because of their different degrees of imbalanced distribution; in (8), for all the four datasets the confidence threshold τ is set to 0.95 by default. All the experiments are implemented on PyTorch 1.9.0 on two GeForce RTX 2080Ti.

D. Ablation Study of ST-RCL

To evaluate the individual role of each component of the proposed ST-RCL framework on SSCD, its ablation study is conducted on LEVIR-CD under all the labeled ratios of 5%, 10%, 20%, and 40%. Table I shows the ablation experimental results. Here, $\mathcal{L}_l$ stands for the "Only-sup" approach that only the labeled data of the supervised branch are used for model training, $\mathcal{L}_l + \mathcal{L}_u^{\mathrm{ST}}$ represents the method involving both the supervised training of the labeled data and the ST of the unrotated strongly augmented unlabeled data, $\mathcal{L}_l + \mathcal{L}_u^{\mathrm{ST}} + \mathcal{L}_u^{\mathrm{CL}}$ denotes the method further using the consistency learning of rotated strongly labeled data based on $\mathcal{L}_l + \mathcal{L}_u^{\mathrm{ST}}$, and $\mathcal{L}_l + \mathcal{L}_u^{\mathrm{ST}} + \mathcal{L}_u^{\mathrm{RCL}}$ is the method further applying uncertainty-based weighting to the consistency learning of rotated strongly labeled data.

Compared with the baseline method of $\mathcal{L}_l$, $\mathcal{L}_l + \mathcal{L}_u^{\mathrm{ST}}$ obtains overall improvement of all the metrics, verifying the significant value of the ST of unlabeled data in SSCD. Based on $\mathcal{L}_l + \mathcal{L}_u^{\mathrm{ST}}$, the introduction of rotation consistency learning, i.e., $\mathcal{L}_l + \mathcal{L}_u^{\mathrm{ST}} + \mathcal{L}_u^{\mathrm{CL}}$, boosts the model performance with a considerable gain, such as the IoU^c increase from 75.10 to 76.88 by 1.78 at the labeled ratio of 5%. It demonstrates the effectiveness of cross-view prediction alignment. After performing the uncertainty-based classwise weighting, $\mathcal{L}_u^{\mathrm{RCL}}$ increases the recall rate of "change" at all the ratios, especially the small ratios, compared with $\mathcal{L}_l + \mathcal{L}_u^{\mathrm{ST}} + \mathcal{L}_u^{\mathrm{CL}}$. It indicates the positive influence of the RCL module on rebalancing the imbalanced distribution between "change" and "nonchange." As a result, taking IoU^c as an example, the complete $\mathcal{L}_l + \mathcal{L}_u^{\mathrm{ST}} + \mathcal{L}_u^{\mathrm{RCL}}$ achieves notable and robust improvements of SSCD at various ratios with the performance advantage of more than 4.1 over the baseline method of OnlySup and 0.69–2.99 over $\mathcal{L}_l + \mathcal{L}_u^{\mathrm{ST}}$.

Overall, the ablation results indicate that each component of ST-RCL can contribute to performance improvement, verifying their respective effectiveness.

E. Rotation Nonequivariance and Necessity of Rotation Consistency Learning

Some prediction examples of unrotated RSI pairs and rotated RSI pairs are given and compared in Fig. 4. These examples intuitively show the existence of rotation nonequivariance in SSCD, leading to unstable CD performance at unfixed rotation angles. Based on the fact that the angles of RSIs are highly variable because of the uncertainty of overhead shooting, it is necessary to make the alignment between the unrotated and rotated RSI pairs to achieve rotation-invariant

TABLE II

ABLATION STUDY RESULTS ON THE LEVIR-CD DATASET IN DIFFERENT ROTATION SETTINGS. BASE1: ROTATING LABELED + UNLABELED RSI PAIRS, BASE2: ONLY ROTATING LABELED RSI PAIRS, BASE3: ONLY ROTATING UNLABELED RSI PAIRS, AND BASE4: NO ROTATING. THE BEST AND SECOND SCORES ARE, RESPECTIVELY, MARKED IN **BOLD** AND IN UNDERLINE

Method	5%					10%					20%					40%				
	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.
Base1	75.13	85.80	0.8515	81.53	90.55	76.76	86.85	0.8625	83.05	91.03	77.90	87.58	0.8700	84.30	91.13	78.57	88.00	0.8743	85.31	90.85
Base2	**75.87**	**86.28**	**0.8564**	82.51	90.41	77.52	87.34	0.8675	83.68	91.33	78.43	87.91	0.8733	86.50	89.37	79.12	88.35	0.8780	85.29	91.62
Base3	74.69	85.51	0.8485	81.29	90.20	77.29	87.19	0.8659	84.11	90.50	78.51	87.96	0.8739	86.96	88.99	80.55	89.23	0.8872	86.83	91.77
Base4	75.10	85.78	0.8513	81.11	91.02	**77.60**	**87.39**	**0.8680**	84.72	90.23	**79.19**	**88.39**	**0.8782**	86.08	90.83	**80.74**	**89.35**	**0.8884**	87.07	91.74

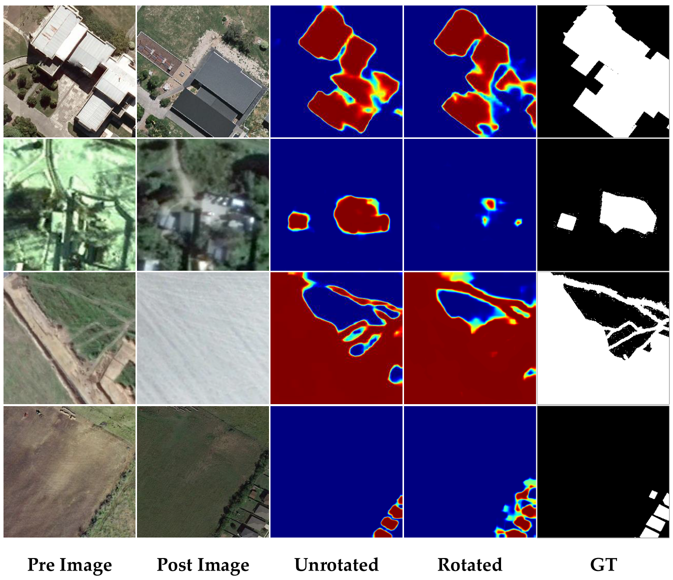


Fig. 4. Some prediction examples of unrotated RSI pairs and 90° clockwise rotated RSI pairs.

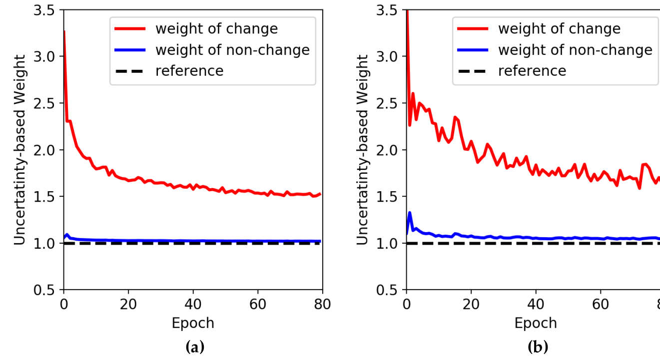


Fig. 5. Uncertainty-based class-wise weight of nonchange/change of 5% labeled (a) LEVIR-CD and (b) GZ-CD.

and stable CD, especially in the situation that there is not enough labeled data for supervised training in the semi-supervised setting.

To further demonstrate the importance and necessity of consistency learning among unrotated and rotated RSI pairs, as shown in Table II, we make the ablation study of four types of rotation augmentation on input data of the LEVIR-CD dataset. They are: Base1 that applies the random rotation within the angles of [90°, 180°, 270°, 360°] to both the labeled and unlabeled RSI pairs, Base2 that only applies random rotation to the labeled RSI pairs, Base3 that only applies to random rotation to the unlabeled RSI pairs, and Base4 that applies random rotation to neither labeled nor unlabeled RSI pairs. All of them are performed based on $\mathcal{L}_l + \mathcal{L}_u^{\mathrm{ST}}$. It could be found that Base4 achieves the best performance without any rotation augmentation operation. Such results indicate that applying random rotation directly to input data, including both labeled and unlabeled data, does not always improve performance and is even harmful in some cases. There are three possible reasons for it. First, the random flipping operation can achieve partial functions of random rotation from the aspect of spatial augmentation, and the further use of random rotation increases the difficulty of model training. Second, in the semi-supervised setting, stable ST of unlabeled data is very important. When there are limited labeled data, the rotation operation may reduce the prediction stability of rotated data and degrade the performance. Third, some datasets have some dominant shooting angles for both labeled and unlabeled data, while the random rotation operation may increase the prediction difficulty at the dominant angles. In contrast, the proposed rotation consistency learning optimized by $\mathcal{L}_u^{\mathrm{ST}}$ can effectively boost the CD performance as shown in Table I, demonstrating its effectiveness and necessity.

F. Effect of Uncertainty-Based Class Weighting of RCL

To clearly show the ability of the uncertainty-based class-wise weighting operation to adapt to convergence difficulty, the weights of "change" and "nonchange" during training are plotted in Fig. 5 except the initial weight [1, 1]. The higher the weight is, the greater importance its corresponding class is given. It can be found that the weights of both "change" and "nonchange" decrease with epochs, which reveals the two classes converge as training. The reason is that both the classes' uncertainties, u_k in (11), shrink with training, resulting in the decrease in the corresponding weights, w_k in (12). It can be seen that the minority class of "change" has a higher weight than the majority class of "nonchange" at all times, which verifies RCL's rebalancing effect of assigning more weight to "change." On the flip side, as a result of greater weight, "change" has a faster convergence speed than "nonchange," which means "change" receives higher attention during training.

To further verify the effectiveness of the uncertainty-based weighting on CL, i.e., RCL, we make a comparison between CL ($\mathcal{L}_l + \mathcal{L}_u^{\mathrm{ST}} + \mathcal{L}_u^{\mathrm{CL}}$) and RCL ($\mathcal{L}_l + \mathcal{L}_u^{\mathrm{ST}} + \mathcal{L}_u^{\mathrm{RCL}}$) based on 5% labeled LEVIR-CD and 5% labeled GZ-CD. Here,

TABLE III
COMPARISON EXPERIMENTAL RESULTS ON THE LEVIR-CD DATASET. THE BEST AND SECOND SCORES ARE, RESPECTIVELY, MARKED IN **BOLD** AND IN UNDERLINE

Method	5%					10%					20%					40%				
	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.
Only-sup	65.55	79.19	0.7822	75.89	82.79	72.13	83.81	0.8306	80.60	87.29	75.97	86.35	0.8571	82.93	90.05	77.33	87.22	0.8662	84.45	90.17
AdvEnt [64]	71.74	83.54	0.8279	79.42	88.12	75.98	86.35	0.8572	81.97	91.22	78.59	88.01	0.8745	84.53	91.79	80.35	89.10	0.8859	87.11	81.19
s4GAN [65]	55.86	71.68	0.7060	59.33	90.46	65.20	78.93	0.7794	76.82	81.16	76.16	86.47	0.8586	80.88	92.89	80.50	89.19	0.8869	86.29	92.30
CPS [66]	72.53	84.08	0.8334	79.95	88.67	77.34	87.22	0.8663	83.26	91.59	79.29	88.45	0.8790	86.12	90.99	<u>81.00</u>	<u>89.50</u>	<u>0.8901</u>	87.27	91.86
RCR [11]	<u>75.75</u>	<u>86.20</u>	<u>0.8554</u>	85.40	87.01	<u>77.97</u>	<u>87.62</u>	<u>0.8703</u>	86.16	89.13	<u>80.00</u>	<u>88.89</u>	<u>0.8836</u>	87.18	90.67	80.61	89.26	0.8875	87.53	91.07
FixMatch [18]	75.10	85.78	0.8513	81.11	91.02	77.60	87.39	0.8680	84.72	90.23	79.19	88.39	0.8782	86.08	90.83	80.74	89.35	0.8884	87.07	91.74
UniMatch [63]	74.75	85.55	0.8489	81.10	90.52	77.50	87.33	0.8673	84.58	90.26	78.57	88.00	0.8743	85.28	90.90	80.17	88.99	0.8848	86.15	92.03
Our ST-RCL	**78.09**	**87.70**	**0.8711**	86.67	88.75	**79.77**	**88.75**	**0.8821**	86.23	91.42	**80.93**	**89.46**	**0.8896**	87.91	91.06	**81.43**	**89.77**	**0.8928**	88.11	91.49
Fully-sup	IoU^c=82.87, F1=90.63, Kappa=0.9019, Rec.=88.12, Rre.=93.29 (Oracle)																			

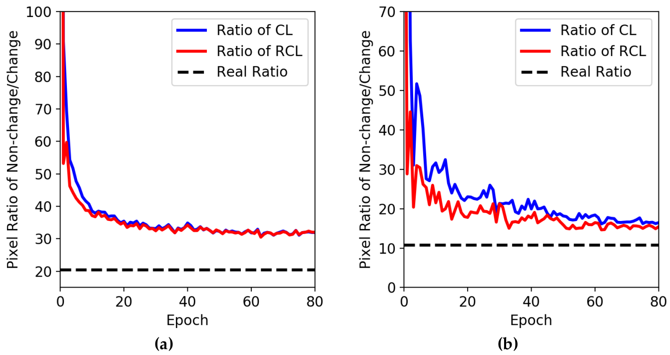


Fig. 6. Pixel ratio of nonchange/change of 5% labeled (a) LEVIR-CD and (b) GZ-CD.

we use the ratio of the pixels predicted as nonchange to the pixels predicted as "change" to evaluate the imbalanced degree during training as given in Fig. 6. It is worth noting that the dashed lines are the real ratios of "nonchange" to change used as a reference. The results show that RCL has a faster convergence speed of the change class for LEVIR-CD and GZ-CD and a higher percentage of it for GZ-CD. It reveals that the uncertainty-based weighting can make the consistency learning more stable and relieve the imbalanced distribution biased to the majority class of the nonchange. For LEVIR-CD, although RCL does not increase the relative ratio of "nonchange" to "change," it gives more weight to "change" and increases its absolute recall as given in Table I, contributing to the improvement of IoU^c as well.

Overall, the uncertainty-based weighting operation of RCL can speed up the convergence of the minority class of change and improve its precision, alleviating the negative impact of the imbalanced distribution.

G. Comparison Experiments

To verify the effectiveness and robustness of the proposed ST-RCL approach, the comparison experiments are performed among it, two baseline methods, and six state-of-the-art semi-supervised methods reproduced for CD. The baseline methods include the following. 1) Only-sup that only uses the limited labeled RSI pairs for model training. It can evaluate the performance gains obtained from SSCD methods. 2) The Oracle fully supervised method (Fully-sup) that uses all the labeled RSI pairs for model training. It can serve as a reference to evaluate the gap between SSCD methods and provide the upper bound of performance in the fully supervised setting.

The reproduced SSCD methods consist of AdvEnt [64], s4GAN [65], CPS [66], RCR [11], FixMatch [18], and UniMatch [63]. As adversarial-learning-based approaches, s4GAN and AdvEnt are transferred from the field of semi-supervised semantic segmentation (SSS). RCR and CPS are based on consistency learning, and RCR was especially proposed for SSCD while CPS is transferred from SSS. FixMatch originally proposed for semi-supervised classification uses the high-confidence pseudolabels of weakly augmented RSI pairs to supervise the change prediction of strongly augmented RSI pairs, and UniMatch further introduced the dual strongly augmented branches and a feature perturbation branch. For SSCD methods, four types of ratios of labeled training data, 5%, 10%, 20%, and 40%, are used for supervised training; accordingly, the remaining 95%, 90%, 80%, and 60% unlabeled training data are used for unsupervised training. To compare these methods fairly, all the methods have the same encoder–decoder model of Dilated ResNet50 + PPM.

The comparison experimental results on LEVIR-CD, WHU-CD, CDD, and GZ-CD are, respectively, listed in Tables III–VI. When it comes to datasets, the four datasets have a wide variety of image styles, building types, spatial resolutions, and sensor types, as well as universal cross-temporal noise of the background. The proposed ST-RCL gets most of the best performance on all the four datasets at different ratios with a considerable and robust advantage when compared with other approaches. The advantage of our approach shows its robustness of CD ability when facing various real factors.

In terms of SSCD methods, it could be found that FixMatch and UniMatch are in the first tile except for our ST-RCL. It verifies the stability of ST-based strategies in SSCD. RCR falling behind FixMatch and UniMatch is in the second group, which is mainly because its progressive consistency learning has a lower convergence speed than the pseudolabel-based ST. AdvEnt and s4GAN have the worst performance with unstable metrics from small to large labeled ratios, which reveals the relative instability of adversarial training in SSCD.

TABLE IV

COMPARISON EXPERIMENTAL RESULTS ON THE WHU-CD DATASET. THE BEST AND SECOND SCORES ARE, RESPECTIVELY, MARKED IN **BOLD** AND IN UNDERLINE

Method	5%					10%					20%					40%				
	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.
Only-sup	66.51	79.89	0.7914	73.16	87.98	72.46	84.03	0.834	80.94	87.36	77.20	87.13	0.8661	86.41	87.87	81.47	89.79	0.8938	87.45	92.26
AdvEnt [64]	71.24	83.21	0.8259	75.93	92.02	78.17	87.75	0.8727	83.98	91.88	79.99	88.88	0.8844	86.58	91.31	82.84	90.62	0.9024	89.50	91.76
s4GAN [65]	53.24	69.48	0.6855	55.83	91.99	71.48	83.37	0.8269	82.42	84.34	76.99	87.00	0.8258	86.18	87.83	81.76	89.96	0.8956	88.05	91.97
CPS [66]	73.14	84.49	0.8388	80.78	88.56	76.99	87.00	0.8649	83.04	91.35	79.63	88.66	0.8822	84.83	92.86	83.35	90.92	0.9056	88.61	93.36
RCR [11]	73.68	84.85	0.8428	78.40	92.46	72.54	84.09	0.8348	79.43	89.33	80.62	89.27	0.8884	87.05	91.61	80.86	89.42	0.8900	86.28	92.79
FixMatch [18]	74.49	85.38	0.8482	81.27	89.94	76.29	86.55	0.8602	83.84	89.45	**81.17**	**89.61**	**0.8919**	88.05	91.22	83.68	91.12	0.9076	89.53	92.76
UniMatch [63]	76.57	86.73	0.8622	82.52	91.40	73.47	84.70	0.8409	83.31	86.14	80.97	89.49	0.8906	88.58	90.41	82.91	90.65	0.9028	89.30	92.05
Our ST-RCL	**78.25**	**87.80**	**0.8732**	83.80	92.20	**78.57**	**88.00**	**0.8752**	84.79	91.45	80.65	89.29	0.8886	86.82	91.90	**83.84**	**91.21**	**0.9086**	89.56	92.93
Fully-sup	IoU^c=91.70, F1=95.67, Kappa=0.9550, Rec.=95.32, Pre.=96.03 (Oracle)																			

TABLE V

COMPARISON EXPERIMENTAL RESULTS ON THE CDD DATASET. THE BEST AND SECOND SCORES ARE, RESPECTIVELY, MARKED IN **BOLD** AND IN UNDERLINE

Method	5%					10%					20%					40%				
	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.
Only-sup	59.28	74.43	0.7105	67.96	82.27	69.10	81.73	0.7922	76.67	87.51	76.84	86.90	0.8507	82.77	91.47	81.03	89.52	0.8801	87.03	92.16
AdvEnt [64]	63.06	77.35	0.7437	70.28	86.00	72.98	84.38	0.8230	77.33	92.83	81.08	89.55	0.8808	85.96	93.46	83.69	91.12	0.8985	88.28	85.83
s4GAN [65]	46.10	63.11	0.5905	51.20	82.25	67.44	80.55	0.7802	72.77	90.21	79.56	88.62	0.8702	84.67	92.95	83.75	91.16	0.8988	89.62	92.75
CPS [66]	62.35	76.81	0.7373	70.36	84.56	72.94	84.36	0.8225	77.86	92.03	80.63	89.27	0.8775	86.22	92.56	**83.93**	**91.28**	**0.9002**	90.14	92.45
RCR [11]	66.67	80.00	0.7739	72.37	89.44	74.64	85.48	0.8349	80.21	91.49	80.91	89.45	0.8794	86.47	92.64	81.95	90.08	0.8864	88.17	92.07
FixMatch [18]	67.06	80.28	0.7754	76.01	85.06	76.53	86.70	0.8486	81.91	92.08	81.92	90.06	0.8866	86.58	83.84	82.86	90.62	0.8929	87.56	93.92
UniMatch [63]	66.07	79.57	0.7672	75.58	84.00	75.31	85.92	0.8400	80.31	92.37	81.79	89.98	0.8856	86.47	93.80	83.58	91.06	0.8979	89.58	92.59
Our ST-RCL	**68.76**	**81.49**	**0.7884**	78.84	84.32	**79.62**	**88.65**	**0.8704**	85.45	92.10	**83.50**	**91.01**	**0.8970**	89.50	92.56	83.81	91.19	0.8991	89.73	92.70
Fully-sup	IoU^c=87.80, F1=93.50, Kappa=0.9255, Rec.=92.99, Pre.=94.02 (Oracle)																			

TABLE VI

COMPARISON EXPERIMENTAL RESULTS ON THE GZ-CD DATASET. THE BEST AND SECOND SCORES ARE, RESPECTIVELY, MARKED IN **BOLD** AND IN UNDERLINE

Method	5%					10%					20%					40%				
	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.	IoU^c	F1	Kappa	Rec.	Pre.
Only-sup	48.74	65.54	0.6253	65.27	65.81	52.56	68.90	0.6617	68.89	68.92	60.94	75.73	0.7375	71.04	81.09	56.58	72.27	0.7026	62.16	86.31
AdvEnt [64]	51.05	67.59	0.6494	63.39	72.38	54.57	70.61	0.6809	68.96	72.34	63.91	77.98	0.7617	73.50	83.04	66.73	80.04	0.7841	75.12	85.65
s4GAN [65]	41.14	58.30	0.5516	51.46	67.23	44.31	61.41	0.5892	49.18	81.47	61.32	76.02	0.7397	74.19	77.94	67.02	80.26	0.7862	76.12	84.87
CPS [66]	51.86	68.30	0.6564	65.61	71.22	56.82	72.46	0.7027	66.68	79.34	62.45	76.88	0.7508	69.71	85.96	69.05	81.69	0.8019	76.72	87.35
RCR [11]	46.34	63.33	0.6039	58.57	68.92	58.16	73.54	0.7134	70.25	77.71	64.34	78.30	0.7646	76.02	80.73	68.78	81.50	0.7995	78.32	84.96
FixMatch [18]	52.97	69.25	0.6696	60.62	80.75	64.19	78.19	0.7640	73.65	83.32	66.18	79.65	0.7795	76.00	83.67	69.04	81.68	0.8016	77.84	85.92
UniMatch [63]	56.42	72.14	0.6990	67.02	78.11	62.77	77.13	0.7528	71.90	83.17	65.23	78.95	0.7718	76.08	82.06	68.41	81.25	0.7973	75.56	87.86
Our ST-RCL	**62.51**	**76.93**	**0.7498**	74.26	79.80	**65.89**	**79.44**	**0.7762**	79.71	79.16	**67.28**	**80.44**	**0.7875**	79.42	81.48	**70.40**	**82.63**	**0.8115**	80.23	85.17
Fully-sup	IoU^c=72.35, F1=83.96, Kappa=0.8258, TPR=82.17, TNR=85.83 (Oracle)																			

As for the labeled training ratios, our ST-RCL achieves superior results with all the four labeled training ratios, especially with a few labeled training samples like 5%. When referring to the metric of IoU^c, there are, respectively, increases of 12.54, 11.74, 9.48, and 13.74 compared with the basic method of "Only-sup" on LEVIR-CD, WHU-CD, CDD, and GZ-CD. Our method gets improvements of IoU^c from 75.75, 76.57, 67.06, and 56.42 to 78.09, 78.25, 68.76, and 62.51 when comparing with the suboptimal methods, with absolute gains of 2.34, 1.68, 1.7, and 6.09 on LEVIR-CD, WHU-CD, CDD, and GZ-CD, respectively.

From the aspect of evaluation metrics, recall can describe the model's capacity to catch the real change pixels, while precision can evaluate the model's ability to make accurate predictions. They are usually conflicted with each other and there is a tradeoff between recall and precision, that is, the increase in the one leads to the decrease in the other. Fortunately, IoU, F1, and Kappa are able to evaluate the detection performance of all the approaches on change regions comprehensively from both recall and precision, and they have a similar tendency. The proposed ST-RCL achieves most of the best results of IoU, F1, and Kappa across datasets, especially at the small labeled training ratios.

Furthermore, to intuitively show the effectiveness and generalization of our ST-RCL, some visualization examples are illustrated in Fig. 7 compared with other SSCD methods.

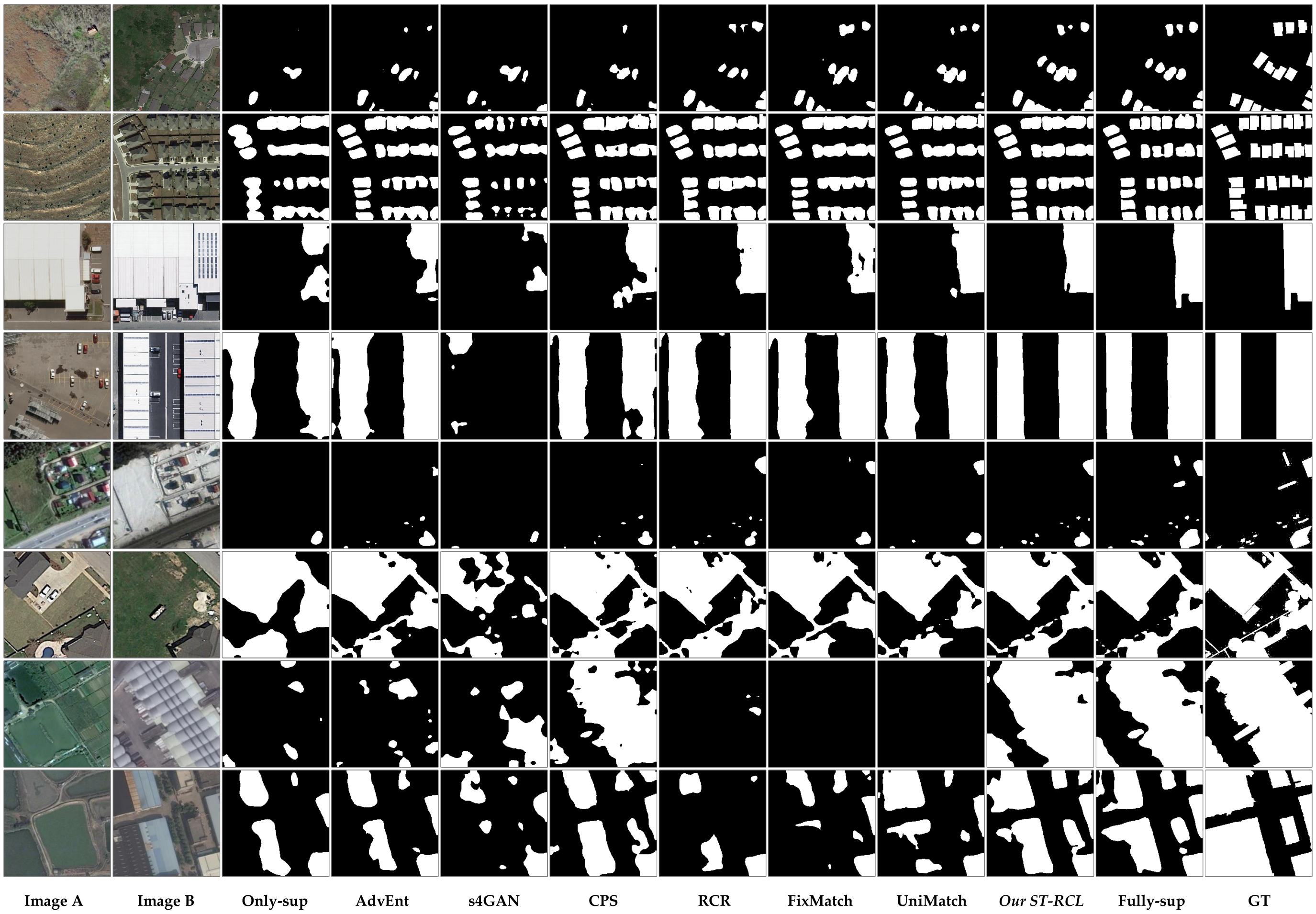


Fig. 7. Some sample visualizations obtained from different comparison methods including our ST-RCL. The 1 and 2 rows, 3 and 4 rows, 5 and 6 rows, and 7 and 8 rows, respectively, show the RSI pairs of LEVIR-CD, WHU-CD, CDD, and GZ-CD at the 5% labeled ratio.

As shown in the figure, the proposed ST-RCL is able to infer the change regions with more accurate and clearer boundaries, indicating that it can effectively and accurately detect the structures and shapes of the change regions.

In general, the proposed ST-RCL outperforms all the comparison SSCD approaches and sets new state-of-the-art performance in most of the scenes across different commonly used CD datasets in the SSL setting. It demonstrates the robustness of the superiority of our approach.

V. Conclusion

In this article, we identify two problems of rotation non-equivariance and imbalanced distribution in SSCD. Accordingly, a novel ST-RCL framework is proposed for SSCD. The proposed ST-RCL consists of a supervised branch that directly uses the labeled data for training and a novel unsupervised branch that uses the unlabeled data for training. The unlabeled branch involves the ST of unrotated RSI pairs and uncertainty-based RCL of rotated RSI pairs. As a result, our ST-RCL can effectively reduce the rotation nonequivariance and relieve the imbalanced distribution in SSCD during unsupervised learning of unlabeled RSI pairs. Compared with other latest SSCD methods, the proposed ST-RCL sets the new state-of-the-art results on four commonly used CD datasets, including LEVIR-CD, WHU-CD, CDD, and GZ-CD in the semi-supervised setting.

In future work, the proposed ST-RCL could be improved from the following two sides: 1) more accurate pseudolabels could be fused from the unrotated unlabeled RSI pairs and rotated unlabeled RSI pairs, which is better for the ST module and 2) a more fine-grained weighting at the pixel level could be designed for the RCL module.

Acknowledgment

The authors would like to thank the editors and anonymous reviewers for the insightful suggestions, which significantly improved the quality of this article.

References

[1] H. Nemmour and Y. Chibani, “Multiple support vector machines for land cover change detection: An application for mapping urban extensions,” *ISPRS J. Photogramm. Remote Sens.*, vol. 61, no. 2, pp. 125–133, Nov. 2006.

[2] I. R. Hegazy and M. R. Kaloop, “Monitoring urban growth and land use change detection with GIS and remote sensing techniques in daqahlia governorate Egypt,” *Int. J. Sustain. Built Environ.*, vol. 4, no. 1, pp. 117–124, Jun. 2015.

[3] S. Das and D. P. Angadi, “Land use land cover change detection and monitoring of urban growth using remote sensing and GIS techniques: A micro-level study,” *GeoJournal*, vol. 87, no. 3, pp. 2101–2123, Jun. 2022.

[4] L. Giustarini, R. Hostache, P. Matgen, G. J.-P. Schumann, P. D. Bates, and D. C. Mason, "A change detection approach to flood mapping in urban areas using TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2417–2430, Apr. 2013.

[5] R. Anniballe et al., "Earthquake damage mapping: An overall assessment of ground surveys and VHR image change detection after L'Aquila 2009 earthquake," *Remote Sens. Environ.*, vol. 210, pp. 166–178, Jun. 2018.

[6] Y. Qing et al., "Operational earthquake-induced building damage assessment using CNN-based direct remote sensing change detection on superpixel level," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, Aug. 2022, Art. no. 102899.

[7] A. A. Alesheikh, A. Ghorbanali, and N. Nouri, "Coastline change detection using remote sensing," *Int. J. Environ. Sci. Technol.*, vol. 4, no. 1, pp. 61–66, Dec. 2007.

[8] R. Liu, M. Kuffer, and C. Persello, "The temporal dynamics of slums employing a CNN-based change detection approach," *Remote Sens.*, vol. 11, no. 23, p. 2844, Nov. 2019.

[9] M. Decuyper et al., "Continuous monitoring of forest change dynamics with satellite time series," *Remote Sens. Environ.*, vol. 269, Feb. 2022, Art. no. 112829.

[10] 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 Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5891–5906, Jul. 2021.

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

[12] C. Sun, J. Wu, H. Chen, and C. Du, "SemiSANet: A semi-supervised high-resolution remote sensing image change detection model using Siamese networks with graph attention," *Remote Sens.*, vol. 14, no. 12, p. 2801, Jun. 2022.

[13] J.-X. Wang, T. Li, S.-B. Chen, J. Tang, B. Luo, and R. C. Wilson, "Reliable contrastive learning for semi-supervised change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4416413.

[14] T. Cohen and M. Welling, "Group equivariant convolutional networks," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 2990–2999.

[15] M. Weiler and G. Cesa, "General E(2)-equivariant steerable CNNs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 14334–14345.

[16] X. Feng, X. Yao, G. Cheng, and J. Han, "Weakly supervised rotation-invariant aerial object detection network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 14146–14155.

[17] E. Arazo, D. Ortego, P. Albert, N. E. O'Connor, and K. McGuinness, "Pseudo-labeling and confirmation bias in deep semi-supervised learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.

[18] K. Sohn et al., "FixMatch: Simplifying semi-supervised learning with consistency and confidence," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 596–608.

[19] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogramm. Remote Sens.*, vol. 80, pp. 91–106, Jun. 2013.

[20] Z. Zhou, L. Ma, T. Fu, G. Zhang, M. Yao, and M. Li, "Change detection in coral reef environment using high-resolution images: Comparison of object-based and pixel-based paradigms," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 11, p. 441, Nov. 2018.

[21] C. Zhang et al., "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 183–200, Aug. 2020.

[22] 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 *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2021, pp. 2058–2061.

[23] L. Liu, D. Hong, L. Ni, and L. Gao, "Multilayer cascade screening strategy for semi-supervised change detection in hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1926–1940, 2022.

[24] X. Zhang, X. Huang, and J. Li, "Semisupervised change detection with feature-prediction alignment," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5401016.

[25] 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 J. Photogramm. Remote Sens.*, vol. 160, pp. 167–179, Feb. 2020.

[26] Q. Shu, J. Pan, Z. Zhang, and M. Wang, "MTCNet: Multitask consistency network with single temporal supervision for semi-supervised building change detection," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 115, Dec. 2022, Art. no. 103110.

[27] Y. Duan, J. Lu, J. Feng, and J. Zhou, "Learning rotation-invariant local binary descriptor," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3636–3651, Aug. 2017.

[28] M. Weiler, F. A. Hamprecht, and M. Storath, "Learning steerable filters for rotation equivariant CNNs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 849–858.

[29] J. Han, J. Ding, N. Xue, and G. Xia, "ReDet: A rotation-equivariant detector for aerial object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2786–2795.

[30] D. Romero, E. Bekkers, J. Tomczak, and M. Hoogendoorn, "Attentive group equivariant convolutional networks," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 8188–8199.

[31] B. Smets, J. Portegies, E. J. Bekkers, and R. Duits, "PDE-based group equivariant convolutional neural networks," *J. Math. Imag. Vis.*, vol. 65, pp. 1–31, Jul. 2022.

[32] T. S. Cohen and M. Welling, "Steerable CNNs," 2016, *arXiv:1612.08498*.

[33] M. Weiler, M. Geiger, M. Welling, W. Boomsma, and T. Cohen, "3D steerable CNNs: Learning rotationally equivariant features in volumetric data," in *Proc. Conf. Workshop Neural Inf. Process. Syst.*, 2018, pp. 10381–10392.

[34] D. E. Worrall, S. J. Garbin, D. Turmukhambetov, and G. J. Brostow, "Harmonic networks: Deep translation and rotation equivariance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5028–5037.

[35] A. Poulenard, M. Rakotosaona, Y. Ponty, and M. Ovsjanikov, "Effective rotation-invariant point CNN with spherical harmonics kernels," in *Proc. Int. Conf. 3D Vis. (3DV)*, Sep. 2019, pp. 47–56.

[36] S. Dieleman, J. De Fauw, and K. Kavukcuoglu, "Exploiting cyclic symmetry in convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1889–1898.

[37] Q. Yang, C. Li, W. Dai, J. Zou, G.-J. Qi, and H. Xiong, "Rotation equivariant graph convolutional network for spherical image classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4302–4311.

[38] Q. Wang, W. Huang, Z. Xiong, and X. Li, "Looking closer at the scene: Multiscale representation learning for remote sensing image scene classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 4, pp. 1414–1428, Apr. 2022.

[39] Z. Xiong, W. Huang, J. Hu, Y. Shi, Q. Wang, and X. X. Zhu, "THE benchmark: Transferable representation learning for monocular height estimation," 2021, *arXiv:2112.14985*.

[40] Q. Wang, W. Huang, X. Zhang, and X. Li, "GLCM: Global–local captioning model for remote sensing image captioning," *IEEE Trans. Cybern.*, early access, Nov. 29, 2022, doi: 10.1109/TCYB.2022.3222606.

[41] W. Huang, Y. Shi, Z. Xiong, Q. Wang, and X. X. Zhu, "Semi-supervised bidirectional alignment for remote sensing cross-domain scene classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 195, pp. 192–203, Jan. 2023.

[42] W. Jiang, "A correlation coefficient for belief functions," *Int. J. Approx. Reasoning*, vol. 103, pp. 94–106, Dec. 2018.

[43] B. Zhou, Q. Cui, X.-S. Wei, and Z.-M. Chen, "BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9719–9728.

[44] Y. Zhang, B. Kang, B. Hooi, S. Yan, and J. Feng, "Deep long-tailed learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 10795–10816, Sep. 2023.

[45] Z. Deng, H. Liu, Y. Wang, C. Wang, Z. Yu, and X. Sun, "PML: Progressive margin loss for long-tailed age classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10498–10507.

[46] W. Jiang, C. Xie, M. Zhuang, and Y. Tang, "Failure mode and effects analysis based on a novel fuzzy evidential method," *Appl. Soft Comput.*, vol. 57, pp. 672–683, Aug. 2017.

[47] J. Guan, J. Liu, J. Sun, P. Feng, T. Shuai, and W. Wang, "Meta metric learning for highly imbalanced aerial scene classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 4047–4051.

[48] W. Miao, J. Geng, and W. Jiang, "Multigranularity decoupling network with pseudolabel selection for remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5603813.

[49] C. Jing et al., "Interclass similarity transfer for imbalanced aerial scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[50] T. Leichtle, C. Geiß, T. Lakes, and H. Taubenböck, "Class imbalance in unsupervised change detection—A diagnostic analysis from urban remote sensing," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 60, pp. 83–98, Aug. 2017.

[51] H. Chen and Z. Shi, "A spatial–temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sens.*, vol. 12, no. 10, p. 1662, 2020.

[52] H. Chen, W. Li, and Z. Shi, "Adversarial instance augmentation for building change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5603216.

[53] B. Bischke, P. Helber, D. Borth, and A. Dengel, "Segmentation of imbalanced classes in satellite imagery using adaptive uncertainty weighted class loss," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 6191–6194.

[54] Z. Zhou et al., "A dynamic effective class balanced approach for remote sensing imagery semantic segmentation of imbalanced data," *Remote Sens.*, vol. 15, no. 7, p. 1768, Mar. 2023.

[55] Y. Liu, Q. Li, Y. Yuan, Q. Du, and Q. Wang, "ABNet: Adaptive balanced network for multiscale object detection in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5614914.

[56] Y. Wu, K. Zhang, J. Wang, Y. Wang, Q. Wang, and X. Li, "GCWNet: A global context-weaving network for object detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5619912.

[57] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[58] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a 'Siamese' time delay neural network," in *Proc. Adv. Neural Inf. Process. Syst.*, 1993, pp. 737–744.

[59] W. Shi et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.

[60] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "Randaugment: Practical automated data augmentation with a reduced search space," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 702–703.

[61] 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, Jan. 2018.

[62] M. A. Lebedev, Y. V. Vizilter, O. V. Vygolov, V. A. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 565–571, May 2018.

[63] L. Yang, L. Qi, L. Feng, W. Zhang, and Y. Shi, "Revisiting weak-to-strong consistency in semi-supervised semantic segmentation," 2022, *arXiv:2208.09910*.

[64] T.-H. Vu, H. Jain, M. Bucher, M. Cord, and P. Pérez, "Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2517–2526.

[65] S. Mittal, M. Tatarchenko, and T. Brox, "Semi-supervised semantic segmentation with high-and low-level consistency," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 4, pp. 1369–1379, Apr. 2019.

[66] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2613–2622.

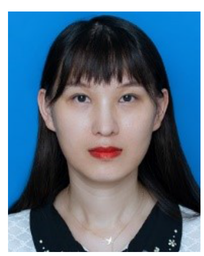
Xueting Zhang received the B.E. degree in measurement and control technology and instrument and the M.E. degree in computer science and technology from Northwestern Polytechnical University, Xi'an, China, in 2018 and 2021, respectively. She is currently pursuing the Ph.D. degree in photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China.

Her research interests include deep learning and remote sensing image processing.

Xin Huang (Senior Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009.

He is working with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, where he is currently a Full Professor, and teaches remote sensing, image interpretation, and so on. He is the Head of the Institute of Remote Sensing Information Processing (IRSIP), School of Remote Sensing and Information Engineering, Wuhan University. He has been supported by The National Program for Support of Top-Notch Young Professionals in 2017, the China National Science Fund for Excellent Young Scholars in 2015, and the New Century Excellent Talents in University from the Ministry of Education of China in 2011. He has published more than 200 peer-reviewed articles (Science Citation Index (SCI) articles) in the international journals. His research interests include remote sensing image processing methods and applications.

Prof. Huang was a recipient of the Boeing Award for the Best Paper in Image Analysis and Interpretation from the American Society for Photogrammetry and Remote Sensing (ASPRS) in 2010, the John I. Davidson President's Award from ASPRS in 2018, and the National Excellent Doctoral Dissertation Award of China in 2012. In 2011, he was recognized by the IEEE Geoscience and Remote Sensing Society (GRSS) as the best reviewer of IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He was the winner of the IEEE GRSS Data Fusion Contest in the years of 2014 and 2021. He was an Associate Editor of the *Photogrammetric Engineering and Remote Sensing* from 2016 to 2019, an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS from 2014 to 2020, an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING from 2018 to 2022, and now serves as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING since 2022. He is also an Editorial Board Member of the *Remote Sensing of Environment* since 2019.



Jiayi Li (Senior Member, IEEE) received the B.S. degree from Central South University, Changsha, China, in 2011, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2016.

She is an Associate Professor with the School of Remote Sensing and Information Engineering, Wuhan University. She has authored more than 60 peer-reviewed articles (Science Citation Index (SCI) articles) in international journals. Her research interests include hyperspectral imagery, sparse representation, computation vision and pattern recognition, and remote sensing images.

Dr. Li is a reviewer for more than 30 international journals, including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS), IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP), IEEE TRANSACTIONS ON CYBERNETICS (TYCB), *Remote Sensing of Environment* (RSE), and *International Society for Photogrammetry and Remote Sensing* (ISPRS-J). She is the Young Editorial Board Member of the *Geospatial-Information Science* (GSIS), the Guest Editor of the *Remote Sensing* (an open access journal from MDPI), and *Sustainability* (an open access journal from MDPI).