PAdaptCD: Progressive Adaptive Thresholding and Bitemporal Image Augmentation for Semisupervised Change Detection

Linlin Wang, Junping Zhang, *Senior Member, IEEE*, Dong Chen, *Student Member, IEEE*, and Lorenzo Bruzzone, *Fellow, IEEE*

***Abstract*—Change detection (CD) aims to identify pixel-level changes of interest in multitemporal remote sensing images (RSIs). Due to the high cost of pixel-level annotations, semisupervised approaches have gained attention by leveraging limited labeled data alongside abundant unlabeled data. However, most current semisupervised change detection (SSCD) methods ignore the severe imbalance between unchanged and changed pixels in CD tasks, which may lead to predictions biased toward the dominant category (unchanged pixels). In addition, there is a lack of data augmentation techniques specifically designed for SSCD, which consider the complexity of RSIs and the unique characteristics of bitemporal images. In this article, we propose a novel SSCD method, PAdaptCD, which incorporates a progressive adaptive dual-threshold (PADT) strategy designed for selecting relatively balanced and reliable classwise pseudolabels. The PADT strategy progressively adjusts thresholds for changed and unchanged classes in an online manner, selecting pseudolabels based on the training status. In addition, bitemporal image augmentation (BTIA) techniques are developed to better capture the specific properties of bitemporal data. BTIA first introduces a highly random intensity-based augmentation, which sets an appropriate degree of data augmentation. Subsequently, a simple yet effective approach called bitemporal image mixing augmentation is proposed to facilitate style interaction between different temporal images. Experiments are carried out on four public datasets, including two building CD datasets (LEVIE-CD and WHU-CD) and two datasets with natural and human-induced changes (JL1-CD and CDD). The proposed PAdaptCD achieves higher performance F1-scores with respect to literature methods with a small amount of labeled data. The results confirm the effectiveness of the proposed method.**

***Index Terms*—Data augmentation, progressive adaptive dual–threshold (PADT), pseudolabel selection, remote sensing, semisupervised change detection (SSCD).**

I. Introduction

REMOTE sensing image change detection (RSCD) is currently one of the focal topics in the remote sensing (RS) field. Its objective is to identify changes of interest in the same scene from multitemporal remote sensing images (RSIs). Change detection (CD) has found widespread application in a variety of fields, including urban development decision making [1], disaster assessment [2], and natural resource monitoring [3].

The task of CD has experienced rapid development in recent years. This can be attributed to advancements in Earth observation technology, which have provided extensive high-quality multitemporal RSIs. In addition, the surge in deep learning research has significantly improved the performance of CD. However, supervised change detection (SSCD) relies on delicate pixel-level manual labeling, which is time consuming and requires rich experience and sound geographic knowledge. This poses a challenge for CD models that have to effectively learn features of changes of interest also in pressure of a limited number of labeled images. Furthermore, CD datasets often exhibit highly unbalanced distributions between changed and unchanged pixels, with changed pixels typically being much less than unchanged ones. This issue becomes more pronounced in datasets with limited samples, where areas of change may be represented by extremely few labeled samples. Consequently, addressing the scarcity and imbalance of labeled samples remains a major challenge in RSCD.

Semisupervised learning has been a promising research direction, which consists in complementing a small number of labeled images with the information extracted from a large set of unlabeled images. Semisupervised learning methods have been introduced in the RS domain in different processing tasks including semantic segmentation [4], [5] and CD tasks [6], [7], [8], [9], [10]. These methods can be divided into three main categories: adversarial learning [4], [5], [6], consistency regularization [7], [8], and pseudolabeling [9], [10]. Adversarial learning methods either use a generative model to create synthetic images as additional input or rely solely on a segmentation network without a generative model. In the latter case, the discriminator distinguishes between segmentation maps generated by the network and real maps (i.e., ground truth). Consistency regularization methods introduce additional losses computed on unlabeled samples and enforce consistency in network outputs under input and/or model perturbations. Pseudolabeling methods perform training on labeled samples to generate pseudolabels for the

Received 12 March 2025; revised 2 June 2025; accepted 1 July 2025. Date of publication 4 July 2025; date of current version 25 July 2025. This work was supported by the National Natural Science Foundation of China under Grant 62271171. *(Corresponding author: Junping Zhang.)*

Linlin Wang, Junping Zhang, and Dong Chen are with the School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China (e-mail: wll_hit@163.com; zhangjp@hit.edu.cn; chendong_hit@stu.hit.edu.cn).

Lorenzo Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy (e-mail: lorenzo.bruzzone@unitn.it).

Digital Object Identifier 10.1109/JSTARS.2025.3586016

unlabeled samples. The generated pseudolabels are added to the labeled set for subsequent training.

The perturbation of data or model has a significant impact on model performance in consistency regularization. Various data augmentation techniques have been explored, e.g., in studies [11], [12], [13], [14], and [15], while works [16], [17], and [18] have primarily focused on different model perturbations in the field of computer vision (CV). TCFSL [19] also explores feature-level and distribution-level adjustments. Data from CD tasks is unique, as it consists of bitemporal images captured at different times. However, recent CD methods often overlook cross-temporal relationships. For instance, WACS-SemiCD [20] and CutMix-CD [21] apply the CutMix augmentation strategy to mix image patches within the same temporal phase, while I3PE [22] exchanges intra- and interimage patches within a single temporal phase. Therefore, it is crucial to design tailored data augmentation techniques for the SSCD task that incorporate temporal information.

Most pseudolabeling methods select high-confidence pseudolabels by setting a fixed threshold (e.g., 0.95 in UniMatch [12]), which is unsuitable for imbalanced data. In the context of CD, due to the imbalance between changed and unchanged categories, selecting highly confident pseudolabels to integrate into the labeled set results in lower recall scores for the changed category. This means that fewer changes are correctly detected, leading to a vicious cycle. Recently, some works [23], [24], [25]; have attempted to address the issue of unbalanced distribution in CD with limited labels. The study in [23] reduces the risk of class imbalance by synthesizing new CD samples through generative adversarial training and image blending techniques. The method in [24] introduces dynamic adaptive focal loss to dynamically and adaptively balance the ratio of changed and unchanged samples according to their quantity in the batch training data. However, the method in [23] is only applicable to building CD, while the method in [24] shows unsatisfactory performance when labeled samples are scarce. The study in [25] aims to alleviate this phenomenon by using a rebalanced consistency learning strategy that incorporates uncertainty-based class weighting. However, a challenge remains that is associated with the need to accurately measuring classwise uncertainty. Here, we argue that an unbalanced distribution is a significantly neglected problem in CD with limited labels, which affects further advancements in SSCD.

Inspired by UniMatch [12], in this article, we propose a novel semisupervised RSCD method, called PAdaptCD, with a progressive adaptive dual-threshold (PADT) strategy. In comparison to UniMatch [12], first, PAdaptCD is more sensitive to the change category in CD tasks due to the PADT strategy. This strategy can leverage the information from unlabeled data more effectively in the initial stages of the model, showing good performance with a minimal amount of labeled samples. Second, PAdaptCD incorporates a random intensity-based augmentation to avoid generating overdistorted samples. Random intensity-based augmentation selects a random number of different intensity-based augmentations. Excessive augmentation is not suitable for CD tasks, as bitemporal images naturally introduce variation through the time dimension, complementing artificial augmentations. Finally, considering the unique characteristics of bitemporal images that may exhibit significant differences caused by seasons and imaging conditions, a bitemporal images mixing (Bi-Mix) technique is introduced in the data strong augmentation process to mitigate the errors caused by these differences.

The main contributions of our work are summarized as follows.

1) We propose a novel semisupervised RSCD method, called PAdaptCD, to address the challenge of enabling CD models to effectively learn interest change features with limited labeled images and a large number of unlabeled images. This method achieves high-precision CD while minimizing labeling costs.
2) A PADT strategy is designed to alleviate class imbalance in bitemporal RS images. Classwise thresholds are dynamically adjusted based on the unlabeled data confidence, progressively selecting balanced and reliable pseudolabels to enhance training effectiveness.
3) A bitemporal image mixing approach is developed to facilitate the interaction of information across different temporal phases and alleviate the interference caused by representational differences.

The rest of this article is organized as follows. Section II reviews the related work about fully supervised CD, semisupervised learning, and SSCD. Section III introduces the proposed method in detail. Section IV presents the experimental results and discussions. Finally, Section V concludes this article.

II. Related Work

A. Fully Supervised CD

Fully supervised CD has been widely studied in the field of remote sensing. Due to the powerful discriminative ability of deep convolutional neural networks (CNNs), numerous CNN-based methods [26], [27], [28], [29], [30], [31] have been proposed for CD. These methods utilize multiscale spatial information [27], [28], employ dilated convolutions [29], or use attention modules [30], [31] to improve the CD model. Recently, to alleviate the limited receptive field caused by the nature of CNNs, transformer-based [32] and CNN-transformer-based [33] networks have been designed, demonstrating highly accurate performance. Very recently, building on the success of Vim [34] and VMamba [35], which have demonstrated superior performance across various CV tasks, ChangeMamba [36] has explored the potential of the Mamba architecture for CD.

The aforementioned fully supervised CD methods have proven effective in RSCD tasks. However, their main limitation lies in their reliance on large amounts of labeled data, resulting in reduced generalization when the labels are scarce.

B. Semisupervised Learning

As mentioned in Section I, designing effective perturbation techniques is crucial for consistency regularization. Data augmentations have played a major role in consistency

regularization-based semisupervised studies [12], [37], [38], [39], [40], [41]. FixMatch [37] employs separate weak and strong augmentations for consistency regularization, using the weak-to-strong approach to supervise strongly perturbed unlabeled images with predictions from their corresponding weakly perturbed versions. Different from FixMatch [37], UniMatch [12] uses separate weak and dual-stream strong augmentations for consistency regularization. ST++ [38] introduces strong data augmentations (SDAs) on unlabeled images to mitigate overfitting caused by noisy labels and to differentiate similar predictions between the teacher and student models. Moreover, CutMix [39] is a powerful augmentation technique in consistency regularization. It has been widely used in semisupervised semantic segmentation tasks [40], [41]. It is applied between unlabeled instances with a predefined probability. This technique cuts a random rectangular region from one image and pastes it onto another image, along with copying the corresponding pseudolabels to the other instance. The study in [40] presents an adaptive CutMix transformation, which generates adaptively sized cut boxes. ClassMix [41] further develops CutMix by dynamically generating binary masks that follow semantic boundaries. Although these data augmentation strategies perform well in many tasks, they do not incorporate temporal information, which is unique and crucial for CD tasks.

Ensuring the reliability of pseudolabels is essential in pseudolabeling. One intuitive idea is to select pseudolabels with high prediction confidence. This way is simple and efficient. However, it neglects the issue of class imbalance by treating each class equally, which leads to biased training. To address this problem, some methods attempt to use dynamic thresholds for pseudolabel selection. FlexMatch [42] introduces a curriculum learning strategy, using a predefined threshold to adaptively adjust the threshold for each class. FreeMatch [43] first obtains an adaptive global threshold and then calculates local thresholds (class-specific thresholds), introducing self-adaptive fairness to counteract the negative effects of imbalance. AdaptMatch [44] calculates adaptive thresholds for the foreground and background based on their historical predictions from both labeled and unlabeled data during the training stage. The study in [45] estimates the adaptive threshold using an exponential moving average (EMA) and sets the initial threshold as the inverse of the number of classes.

FlexMatch [42] and FreeMatch [43] compute the normalized estimated learning effect between each class and the predicted optimal class to adaptively adjust thresholds for pseudolabel selection. However, these methods are limited as they do not decouple the learning status of each class. While AdaptMatch [44] requires constructing memory banks to store historical predictions, which brings a storage burden. The threshold in [45] is heavily influenced by the preset prior value, hindering real-time adjustments during training.

C. Semi-Supervised CD (SSCD)

Research in SSCD is much less extensive than research in fully supervised CD due to the following substantial challenges.

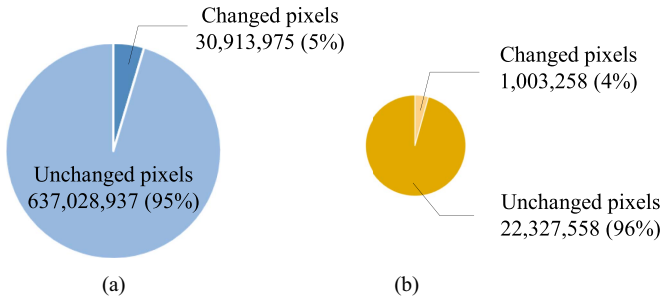

Fig. 1. Number of changed and unchanged pixels and their respective proportions, using the LEVIR-CD dataset [31] as an example. (a) Statistics for all samples in LEVIR-CD. (b) Statistics for the selected 5% labeled samples.

Fig. 2. Two samples of bitemporal image pairs showing different representations of the same object (the same buildings are highlighted in the red and blue boxes, respectively).

1) RSCD tasks suffer from severe data imbalance, and selecting only a few labeled samples exacerbates this imbalance (with fewer changed samples), as shown in Fig. 1.
2) Bitemporal RS images are affected by differences in acquisition times, leading to variations in illumination, seasons, and sensor characteristics, which can cause the same objects to exhibit different spectral properties, as shown in Fig. 2. This challenge highlights the need for a large number of labeled samples.

Semisupervised learning has gradually been introduced into CD tasks. In the literature, several SSCD methods have demonstrated good performance [6], [7], [10], [25], [46], [47], [48], [49], [50], [51]. SemiCDNet [6] uses a generative adversarial network for SSCD tasks. SemiCD [7] applies different random perturbations to a deep feature difference map; then it uses consistency regularization to drive the predicted change probability map from the CD network to be consistent under these perturbations. SemiSiROC [10] relies on half-sibling regression for CD as an unsupervised teacher model to generate pseudolabels. It pre-trains different student models by selecting the highest-confidence pseudolabels. ST-RCL [25] adopts a self-training approach, performing consistency learning on weakly augmented and unrotated and rotated strongly augmented predictions, while introducing a rebalanced consistency learning loss to alleviate the imbalanced distribution. SemiSANet [46] constructs Siamese Networks with graph attention to pretrain on the limited labeled data and utilizes consistency regularization on the high-confidence pseudolabels under different augmentations of the unlabeled data. FPA [47] introduces class-aware feature alignment and pixelwise prediction alignment to select

pseudolabels that exceed a fixed threshold for training on unlabeled samples. The study in [48] proposes SSCD method based on a graph convolutional network, which encodes multitemporal images as a graph through multiscale parcel segmentation. C2F-SemiCD [49] employs consistency regularization by minimizing the prediction difference between the student and the teacher models and designs a multiscale attention mechanism to extract change features from coarse to fine. The research in [50] proposes a dynamically updated cross-supervised SSCD network, which dynamically updates the pseudolabels during the retraining phase. RISL [51] integrate boundary trimming and instance reliability evaluation to generate pseudolabels. STCR-Net [52] selects pseudoimages at the image level during the self-training stage and applies consistency regularization to both the prediction results and feature dimensions. CT^2 RCDSS [53] leverages consistency loss to constrain the generation of pseudolabels from networks initialized with different parameters. S^3 CDRNet [54] defines a class confidence index to adjust the sampling rate in the data augmentation process. CutMix-CD [21] employs a change-aware CutMix operation to generate mixed samples, with a student–teacher framework designed to predict changes in both the original and mixed samples.

However, among aforementioned semisupervised methods, only ST-RCL [25], S^3 CDRNet [54], and CutMix-CD [21] attempt to alleviate the imbalanced distribution problem in the CD task. There is still ample room for further investigation into the imbalance issue in SSCD. In addition, there is little literature specifically focused on designing data augmentation strategies for the CD task. STCRNet [52] briefly mentions the use of the CutMix augmentation strategy on bitemporal images. However, this CutMix strategy is applied between preevent and postevent images of the same pair, lacking the incorporation of information mixing across image pairs. CutMix-CD [21] applies the CutMix operation between the same temporal images from two image pairs. Nevertheless, images from the same temporal phase often exhibit similar styles. Therefore, exploring CutMix strategies across different temporal images from different pairs is a promising direction.

III. Methodology

In SSCD, we are given a small amount of labeled samples and a large amount of unlabeled samples. Let $\mathcal{D}^l = \{(x^l_{1,i}, x^l_{2,i}, y^l_i)\}_{i=1}^{N_l}$ represent the N_l labeled bitemporal samples, and $\mathcal{D}^u = \{(x^u_{1,i}, x^u_{2,i})\}_{i=1}^{N_u}$ denote the unlabeled bitemporal samples, where $(x^l_{1,i}, x^l_{2,i})$ represents the ith labeled bitemporal image pair, $x^l_{1,i}$ and $x^l_{2,i}$ are preevent image and postevent image, respectively, and y^l_i is the labeled change map. Similarly, $(x^u_{1,i}, x^u_{2,i})$ is the ith unlabeled bitemporal image pair.

The proposed method is based on UniMatch [12], incorporating weak-to-strong consistency regularization and pseudolabeling techniques, as well as employing dual-stream perturbations. Differently from UniMatch [12], considering the unique aspects of the RSCD task, we design bitemporal images augmentation and PADT strategies, as detailed in Sections III-B and III-C.

TABLE I
List of Various Image Transformations in PAdaptCD

Weak Augmentation	
-Apply all	
Random Scale	Randomly resizes the image by [0.8, 1.2].
Random Crop	Horizontally flips the image with a probability of 0.5.
Random Flip	Randomly crops a region from the image.
Strong Augmentation	
Highly Random Intensity-Based image-level Augmentation	
-Apply k randomly	
Identity	Returns the original image.
Autocontrast	Maximizes (normalize) the image contrast.
Equalize	Equalizes the image histogram.
Gaussian blur	Blurs the image with a Gaussian kernel.
Contrast	Adjusts the contrast of the image by [0.05, 0.95].
Brightness	Adjusts the brightness of the image by [0.05, 0.95].
Color	Enhances the color balance of the image by [0.05, 0.95].
Sharpness	Adjusts the sharpness of the image by [0.05, 0.95].
Posterize	Reduces each pixel to [4,8] bits.
Solarize	Inverts image pixels above a threshold from [1,256).
Hue	Jitters the hue of the image by [0.0, 0.5].
A novel mixing mechanism	
Bi-Mix	Mixes areas across different temporal images from different pairs.

A. Network Architecture

The overall architecture of the proposed PAdaptCD is illustrated in Fig. 3. As previously mentioned, PAdaptCD employs a weak-to-strong consistency regularization to process unlabeled samples. First, each unlabeled image pair $(x^u_{1,i}, x^u_{2,i})$ is simultaneously perturbed by weak augmentation A_w and dual-stream strong augmentation branches A_{s_1} and A_{s_2} (details in Table I). Then, the labeled image pair $(x^l_{1,i}, x^l_{2,i})$ and the generated weak augmented unlabeled image pair $(x^{uw}_{1,i}, x^{uw}_{2,i})$, as well as the dual-stream strong augmented unlabeled image pairs $(x^{us_1}_{1,i}, x^{us_1}_{2,i})$ and $(x^{us_2}_{1,i}, x^{us_2}_{2,i})$ are all given as input to a shared weighted Siamese networks. The strong augmentation combines highly random intensity-based augmentation (HRIA) with a tailored bitemporal image mixing strategy. Both labeled and unlabeled samples are used simultaneously for model training. For unlabeled samples, pseudolabels with confidence exceeding threshold τ are selected to train the model. The progressive adaptive dual thresholds are dynamically updated in real-time based on feedback during training. Finally, the predicted change map is obtained.

The overall objective function $\mathcal{L}$ consists of supervised loss $\mathcal{L}_s$ and unsupervised loss $\mathcal{L}_u$ and is defined as

$$\mathcal{L} = \frac{1}{2}(\mathcal{L}_s + \mathcal{L}_u). \tag{1}$$

Specifically, the supervised loss $\mathcal{L}_s$ is the cross-entropy loss between model prediction map p^l and ground-truth map y^l. The unsupervised loss $\mathcal{L}_u$ is defined as

$$\mathcal{L}_u = \frac{1}{B_u}\sum \mathbb{1}(\max(\hat{p}^w \geq \tau)) \cdot \Bigg(\lambda H(\hat{y}^w, p^{fp}) + \frac{\mu}{2}(H(\hat{y}^w, p^{s_1}) + H(\hat{y}^w, p^{s_2}))\Bigg) \tag{2}$$

where B_u is the batch size for unlabeled data, $\hat{p}^w$ denotes the predicted confidence, and τ is a scalar hyperparameter representing the threshold beyond which pseudolabels are selected.

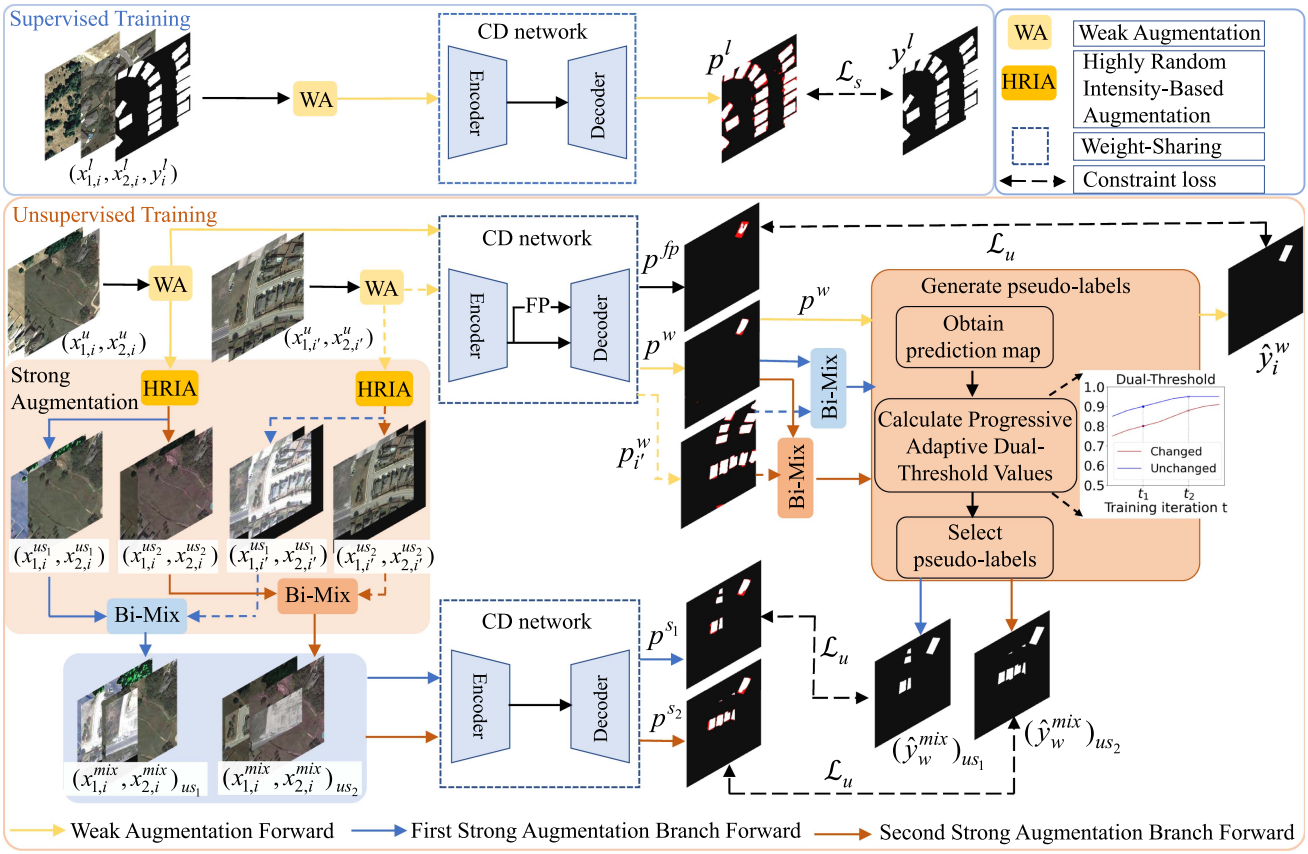

Fig. 3. Architecture of the proposed PAdaptCD. The labeled sample pairs $(x^l_{1,i}, x^l_{2,i}, y^l_i)$ and the unlabeled sample pairs $(x^u_{1,i}, x^u_{2,i})$ in the batch are simultaneously used for supervised and unsupervised training, respectively. In supervised training, standard learning is performed. In unsupervised training, data augmentation is first applied, including both weak and strong augmentations, where the strong augmentations consist of highly random intensity-based augmentations and bitemporal image mixing (Bi-Mix). The weakly augmented and strongly augmented data are then fed into CD networks to produce predicted change feature maps p^w, p^{s_1}, and p^{s_2}. Pseudolabels are subsequently selected using a PADT strategy. Finally, the model is optimized using a combination of supervised loss $\mathcal{L}_s$ and unsupervised loss $\mathcal{L}_u$.

$\hat{y}^w$ represents the pseudolabels selected from weakly perturbed images, p^{fp} is the output from feature perturbation, and p^{s_1} and p^{s_2} are the outputs from dual-stream perturbations. H minimizes the entropy between the selected pseudolabels and the predictions obtained from feature perturbation and dual-stream perturbations.

Shared weighted Siamese networks: Following the UniMatch [12], we employ a DeepLabv3+ model based on ResNet-50 for change detection in RS images. The Siamese DeepLabv3+ networks are shared between the labeled data sampled from $\mathcal{D}^l$ and the unlabeled data sampled from $\mathcal{D}^u$. The labeled data are used in the supervised training part in Fig. 3, helping the model to acquire the preliminary ability to predict changes. The single-branch DeepLabv3+ model consists of an encoder–decoder architecture, while PAdaptCD adopts a Siamese structure to extract features from preevent and postevent images derived from bitemporal images. The bitemporal images $(x_{1,i}, x_{2,i})$ are first sent to the encoder $\mathcal{E}$. The decoder $\mathcal{G}$ is utilized to integrate the information from bitemporal images and generate the predicted change feature map p

$$p = \mathcal{G}(\mathcal{E}(x_{1,i}, x_{2,i})). \tag{3}$$

B. Bitemporal Images Augmentation

1) Highly Random Intensity-Based Augmentation (HRIA): The key to semisupervised learning is how to effectively utilize unlabeled samples. Here, strong and weak perturbations are used for image-level augmentation to generate prediction discrepancies. The weak augmentation A_w as shown in Table I, includes standard resizing, cropping, and flipping operations

$$(x^{uw}_{1,i}, x^{uw}_{2,i}) = A_w((x^u_{1,i}, x^u_{2,i})). \tag{4}$$

The strong augmentations A_{s_1} and A_{s_2} in dual-stream perturbations are implemented by introducing HRIA [18] and a novel mixing mechanism called Bi-Mix. As discussed in [15], augmentations with excessive distortion can disrupt data distribution and degrade the performance of semisupervised semantic segmentation. Instead of using a fixed number, we apply a random number of augmentations from the augmentation pool, up to a maximum value of k. This augmentation strategy features a high degree of randomness, without significantly affecting the data distribution

$$x^{us_1}_{1,i} = A_{s_1}(x^{uw}_{1,i}), x^{us_1}_{2,i} = A_{s_1}(x^{uw}_{2,i})$$

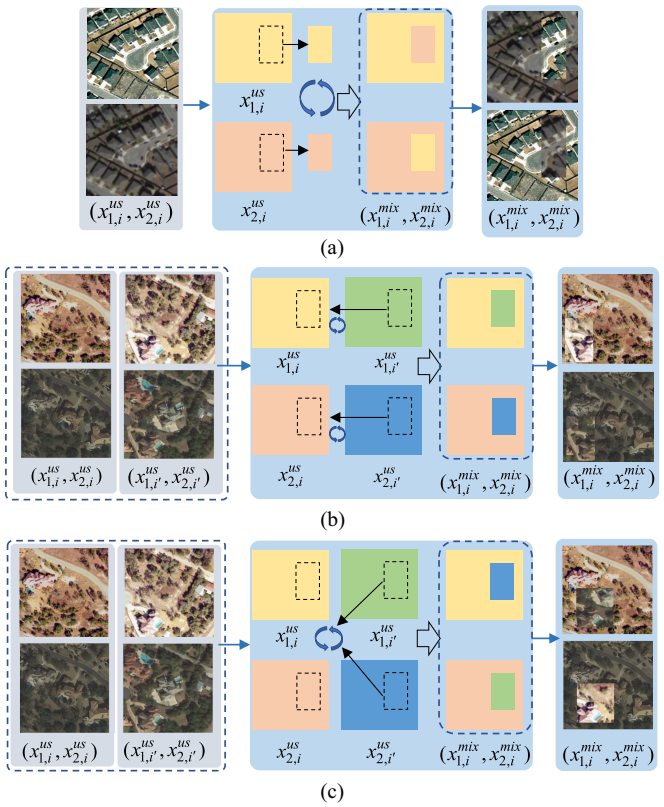

Fig. 4. Various CutMix strategies in bitemporal images for CD task. (a) CutMix-Same Pair: CutMix is applied within the same bitemporal image pair. (b) CutMix-Same Temporal: CutMix is applied to the same temporal images across different pairs. (c) CutMix-Dual Diff (Bi-Mix, used in PAdaptCD): CutMix is applied across different temporal images from different pairs.

$$x_{1,i}^{us_2} = A_{s_2}(x_{1,i}^{uw}), x_{2,i}^{us_2} = A_{s_2}(x_{2,i}^{uw}) \tag{5}$$

where A_{s_1} refers to the perturbation from the first strong augmentation branch, the A_{s_1} used for preevent and postevent images differs due to the application of HRIA. Similarly, A_{s_2} denotes the random perturbation from the second strong augmentation branch.

2) Bitemporal Images Mixing (Bi-Mix): Bi-Mix adopts a "CutMix-Dual Diff" strategy, which aims to mix the different temporal information across different RS bitemporal image pairs, which is applied as a strong augmentation, as shown in Fig. 4(c).

As already mentioned, in real-world problems, due to different acquisition times, illumination, and atmospheric conditions, land-cover objects in bitemporal images, whether changed or unchanged, exhibit differences in appearance, as illustrated in Fig. 2. Thus, the unchanged pixels may be easily confused with the changed ones. To alleviate this problem, we design a new CutMix strategy (named "Bi-Mix") for bitemporal images. The inputs to the Bi-Mix are two different pairs of bitemporal images that have already undergone HRIA. The process is implemented as follows.

1) Randomly generate a properly sized rectangular box based on the size of the input image, and generate a binary mask M, where the positions within the randomly generated rectangular box are set to 1, whereas the other values are 0.
2) Utilize the binary mask M where a value of 1 indicates the selection of the mixing pixels from $(x_{1,i}^{us}, x_{2,i}^{us})$ and another temporal image from another bitemporal image pair $(x_{1,i'}^{us}, x_{2,i'}^{us})$.
   The selected pixels are used to generate bitemporal mixed images; the generation process of mixed pseudolabels is similar.
3) Through Bi-Mix, $x_{1,i}^{us}$ and $x_{2,i}^{us}$ are generated as $x_{1,i}^{\text{mix}}$ and $x_{2,i}^{\text{mix}}$, respectively, for input into the dual stream.

The process of generating the bitemporal mixed images and the mixed pseudolabels $\hat{y}_w^{\text{mix}}$ can be written as

$$\begin{aligned} x_{1,i}^{\text{mix}} &= (1-M) \odot x_{1,i}^{us} + M \odot x_{2,i'}^{us} \\ x_{2,i}^{\text{mix}} &= (1-M) \odot x_{2,i}^{us} + M \odot x_{1,i'}^{us} \\ \hat{y}_w^{\text{mix}} &= (1-M) \odot \hat{y}_i^{w} + M \odot \hat{y}_{i'}^{w} \end{aligned} \tag{6}$$

where $\odot$ denotes elementwise multiplication, and $\hat{y}_i^w$ and $\hat{y}_{i'}^w$ are the pseudolabels generated from the two pairs of bitemporal images, respectively.

As shown in Fig. 4(c), the mixed bitemporal images contain information from the preevent and postevent images, as well as from another pair of images. It serves as a bridge to facilitate interaction between the styles of the bitemporal images, helping to mitigate interference caused by representational differences. Moreover it increases data diversity, improving the generalization ability of the model.

C. Progressive Adaptive Dual-Threshold (PADT)

Equation (2) provides the intuition that the utilization of unlabeled data is directly controlled by the confidence threshold τ. In the initial training phase, if the value of τ is too high, fewer unlabeled data will be utilized, failing to fully exploit the information embedded in the unlabeled samples, which leads to a slow convergence of the training process. However, if the value of τ is set too low, although more unlabeled data participate in training, a large number of inaccurate pseudolabels will affect the training of the model. As mentioned in [43], the key to determining the semisupervised threshold is that it should reflect the learning state. Hence, we propose a PADT strategy that automatically calculates the thresholds for the changed and unchanged categories based on the predictions obtained during each training iteration. To ensure the reliability of selected pseudolabels to a certain extent, we design a progressive increment adjustment approach to control the value of τ from being too low. Fig. 5 shows the illustration of PADT.

As shown in Fig. 3, the labeled image pair $(x_{1,i}^l, x_{2,i}^l)$, the weak augmented unlabeled image pair $(x_{1,i}^{uw}, x_{2,i}^{uw})$, and the dual-stream strong augmented unlabeled image pairs $(x_{1,i}^{us_1}, x_{2,i}^{us_1})$ and $(x_{1,i}^{us_2}, x_{2,i}^{us_2})$ are all trained at the same time in the proposed PAdaptCD. The weak augmentations yield relatively reliable results and are used for generating pseudolabels. Therefore, we calculate the adaptive thresholds based on the predictions of

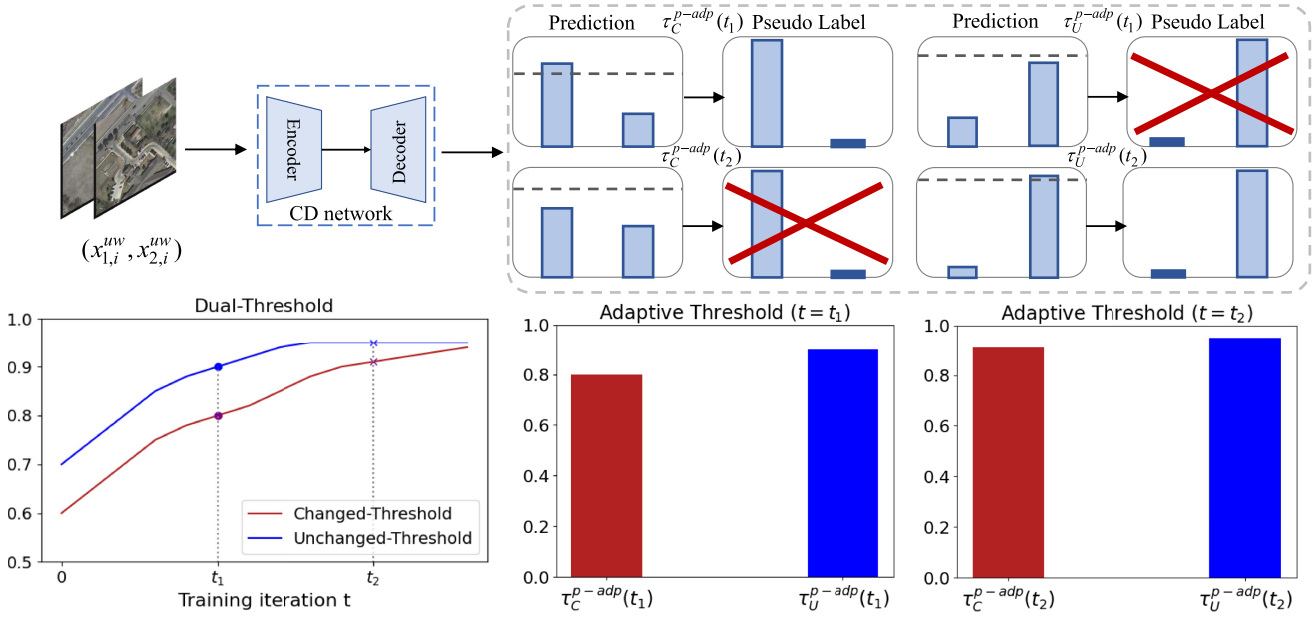

Fig. 5. Illustration of the PADT. Pseudolabels are selected using PADT, with threshold values that are calculated based on the EMA of prediction statistics and a progressive increment adjustment strategy. Filtered (discarded) samples are marked with red X.

the weak augmentations. After giving $(x_{1,i}^{uw}, x_{2,i}^{uw})$, $(x_{1,i}^{us_1}, x_{2,i}^{us_1})$ and $(x_{1,i}^{us_2}, x_{2,i}^{us_2})$ as input into the encoder–decoder network, the predicted change features can be obtained as

$$\begin{aligned} p^w &= \mathcal{G}(\mathcal{E}(x_{1,i}^{uw}, x_{2,i}^{uw})), p^{fp} = \mathcal{G}(\mathcal{P}(\mathcal{E}(x_{1,i}^{uw}, x_{2,i}^{uw}))) \\ p^{s_1} &= \mathcal{G}(\mathcal{E}(x_{1,i}^{us_1}, x_{2,i}^{us_1})), p^{s_2} = \mathcal{G}(\mathcal{E}(x_{1,i}^{us_2}, x_{2,i}^{us_2})) \end{aligned} \tag{7}$$

where $\mathcal{P}$ denotes feature perturbations (refer to [12] for more details).

To normalize each pixel prediction in p^w to the range [0, 1], the softmax function is applied to each pixel along the class dimension, i.e.,

$$\hat{p}^w = \text{Softmax}(p^w) \tag{8}$$

where $\hat{p}^w = \hat{p}_C^w \cup \hat{p}_U^w$, $\hat{p}_C^w$, and $\hat{p}_U^w$ represent the predicted probabilities of changed and unchanged pixels in $\hat{p}^w$, respectively.

Based on the prediction map $\hat{p}^w$, the pseudolabel map $\hat{y}^w \in \mathbb{R}^{H\times W}$ can be obtained for pseudosupervision training on the dual-stream strongly augmented unlabeled image. The $\hat{y}^w$ can be expressed as

$$\hat{y}^w(i,j) = \begin{cases} 1, & \text{if } \hat{p}_C^w(i,j) > 0.5 \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

The average confidence for each class can be displayed in real-time to indicate its convergence difficulty, thus having the capability to serve as the corresponding threshold. The adaptive dual-threshold can be calculated as

$$\begin{cases} \tau_C^{\text{adp}}(t) = \frac{\sum \hat{p}_C^w}{\text{Num}(\hat{p}_C^w)}, & \hat{p}_C^w(i,j) > 0.5 \\ \tau_U^{\text{adp}}(t) = \frac{\sum \hat{p}_U^w}{\text{Num}(\hat{p}_U^w)}, & \hat{p}_U^w(i,j) > 0.5 \end{cases} \tag{10}$$

where τ_C^{adp} and τ_U^{adp} represent average confidence for changed and unchanged pixels, respectively, t represents the tth time step (iteration), and $\text{Num}(\hat{p}_C^w)$ and $\text{Num}(\hat{p}_U^w)$ represent the number of changed and unchanged pixels in $\hat{p}_w$, respectively.

We use the EMA of the confidence at each training step to iteratively update the adaptive dual-threshold

$$\begin{cases} \tau_C^{\text{adp}}(t) \leftarrow m \cdot \tau_C^{\text{adp}}(t) + (1-m) \cdot \tau_C^{\text{adp}}(t-1) \\ \tau_U^{\text{adp}}(t) \leftarrow m \cdot \tau_U^{\text{adp}}(t) + (1-m) \cdot \tau_U^{\text{adp}}(t-1) \end{cases} \tag{11}$$

where $m \in (0,1)$ is a momentum coefficient of EMA. Here, we set $m = 0.9$.

As the training progresses, the adaptive dual-threshold should ideally steadily increase until it stabilizes. Therefore, we design a progressive increment to adjust the adaptive threshold, ensuring that more incorrect pseudolabels are discarded. Specifically, we adjust the progressive increments according to the training status and the iteration number, and set an upper bound τ_{upper} of 0.95 for the confidence threshold. The final PADT values $\tau_C^{p-\text{adp}}(t)$ and $\tau_U^{p-\text{adp}}(t)$ can be expressed as

$$\tau_C^{p-\text{adp}}(t) = \begin{cases} \tau_C^{\text{adp}}(t) + \frac{\tau_{\text{upper}} - \tau_C^{\text{adp}}(t)}{1+(T)^{-t}}, & \text{if } \tau_C^{\text{adp}}(t) < \tau_{\text{upper}} \\ \tau_{\text{upper}}, & \text{otherwise} \end{cases} \tag{12}$$

$$\tau_U^{p-\text{adp}}(t) = \begin{cases} \tau_U^{\text{adp}}(t) + \frac{\tau_{\text{upper}} - \tau_U^{\text{adp}}(t)}{1+(T)^{-t}}, & \text{if } \tau_U^{\text{adp}}(t) < \tau_{\text{upper}} \\ \tau_{\text{upper}}, & \text{otherwise} \end{cases}$$

where T is a hyperparameter.

Finally, the unsupervised loss $\mathcal{L}_u$ at the tth iteration is given by

$$\mathcal{L}_u = \frac{1}{B_u} \sum \mathbb{1}(\max(\hat{p}^w \geq \tau_{\text{argmax}(\hat{p}_w)}^{p-\text{adp}}(t))) \cdot \Big(\lambda H(\hat{y}_w, p^{fp}) + \frac{\mu}{2}(H((\hat{y}_w^{\text{mix}})_{us_1}, p^{s_1}) + H((\hat{y}_w^{\text{mix}})_{us_2}, p^{s_2})) \Big). \tag{13}$$

Algorithm 1: PAdaptCD Algorithm.

Input: Labeled set $\mathcal{D}^l = \{(x^l_{1,i}, x^l_{2,i}), y^l_i\}_{i=1}^{N_l}$, unlabeled set $\mathcal{D}^u = \{x^u_{1,i}, x^u_{2,i}\}_{i=1}^{N_u}$, CD networks $\mathcal{E}$ - $\mathcal{G}$, total epoch number N_{epoch}, batch size bs, weak augmentation $A_w(\cdot)$, HRIA $A_s(\cdot)$ and randomly generated binary mask M in Bi-Mix

Output: Updated CD network weights $\mathcal{E}$ - $\mathcal{G}$

for epoch $\leftarrow 1$ **to** N_{epoch} **do**
  for $iter \leftarrow 1$ **to** $\left\lfloor \frac{len(\mathcal{D}^u)}{bs} \right\rfloor$ **do**
    Sample labeled pair $\{(x^l_{1,i}, x^l_{2,i}, y^l_i)\}$, sample unlabeled pair $(x^u_{1,i}, x^u_{2,i})$
    Obtain weakly augmentation image pair $(x^{uw}_{1,i}, x^{uw}_{2,i}) \leftarrow A_w(\cdot) \leftarrow (x^u_{1,i}, x^u_{2,i})$
    Obtain HRIA $(x^{us}_{1,i}, x^{us}_{2,i}) \leftarrow A_s(\cdot) \leftarrow (x^{uw}_{1,i}, x^{uw}_{2,i})$
    Obtain bi temporal mixed images $x^{\text{mix}}_{1(2),i} = (1-M) \odot x^{us}_{1(2),i} + M \odot x^{us}_{2(1),i'}$
    Generate the predicted change feature maps $p^w, p^{fp}, p^{s_1}, p^{s_2} \leftarrow \mathcal{G}(\mathcal{E}(\cdot)), \mathcal{G}(\mathcal{P}(\mathcal{E}(\cdot))) \leftarrow (x^{uw}_{1,i}, x^{uw}_{2,i}), (x^{us_1}_{1,i}, x^{us_1}_{2,i}), (x^{us_2}_{1,i}, x^{us_2}_{2,i})$
    Generate the pseudo labels $\hat{y}^w \leftarrow \hat{p}^w$
    Generate the mixed pseudo labels $\hat{y}^{mix}_w = (1-M) \odot \hat{y}^w_i + M \odot \hat{y}^w_{i'}$
    Calculate current iter adaptive dual-threshold values $(\tau^{adp}_C(t), \tau^{adp}_U(t)) \leftarrow \hat{p}^w$
    Update the adaptive dual-threshold values $\tau^{\text{adp}}_{C(U)}(t) \leftarrow (\tau^{\text{adp}}_{C(U)}(t), \tau^{\text{adp}}_{C(U)}(t-1))$
    Obtain the final PADT values
    if $\tau^{\text{adp}}_{C(U)}(t) < \tau_{upper}$ **then**
      $\tau^{p-\text{adp}}_{C(U)}(t) = \tau^{\text{adp}}_{C(U)}(t) + \frac{\tau_{\text{upper}} - \tau^{\text{adp}}_{C(U)}(t)}{1 + (T)^{-t}}$;
    else
      $\tau^{p-\text{adp}}_{C(U)}(t) = \tau_{\text{upper}}$;
    end
    Calculate the unsupervised loss $\mathcal{L}_u \leftarrow (\hat{p}^w, \tau^{p-\text{adp}}_{\text{argmax}(\hat{p}_w)}(t), p^{fp}, p^{s_1}, p^{s_2}, \hat{y}_w, \hat{y}^{\text{mix}}_w)$
    Update CD network weights $\mathcal{E}$ - $\mathcal{G}$ by $\mathcal{L} = \mathcal{L}_s + \mathcal{L}_u$
  end
end

To better illustrate the update steps of the proposed algorithm, the detailed training process of the PAdaptCD is described in Algorithm 1.

IV. Experiments

A. Datasets

Four public high-resolution RSCD datasets, i.e., LEVIR-CD, WHU-CD, CDD, and JL1-CD datasets, are employed to assess the effectiveness of the proposed method.

The LEVIR-CD dataset [31] consists of 637 pairs of bitemporal images with a size of 1024 × 1024 pixels and a resolution of 0.5 m, obtained from Google Earth. It is a building CD dataset, including changes in various types of buildings such as villa residences, tall apartments, small garages, and large warehouses. The images were acquired between 2002 and 2018, thus containing interference from many irrelevant changes, such as seasonal and illumination variations. In the experiments, the images are cropped into 256 × 256 pixels. We adopt its default dataset split: 7120/1024/2048 for training/validation/test.

The WHU-CD dataset [55] consists of a pair of bitemporal images with a size of 32 507 × 15 354 pixels and a resolution of 0.3 m. It is also a building CD dataset, which covers the area in the region of Christchurch, New Zealand, which experienced a 6.3 magnitude earthquake in February 2011 and underwent reconstruction in the following years. The dataset was captured in 2012 and 2016. The WHU-CD dataset is challenging due to the highly heterogeneous distribution of change areas. Following [7] and[40], the images have been cropped into nonoverlapping patches with a size of 256 × 256 pixels: 5947/743/744 for training/validation/test.

The JL1-CD dataset [56] consists of 5000 pairs of bitemporal satellite images with a size of 512 × 512 pixels and submeter resolutions ranging from 0.5 to 0.75 m. The images were captured between early 2022 and the end of 2023 across various regions in China, including Shandong, Ningxia, Anhui, Hebei, and Hunan. The dataset covers diverse common human-induced and natural surface changes, such as piled Earth, buildings, roads, hardened surfaces, woodlands, grasslands, croplands, and water bodies. Following the default dataset split, the dataset is divided into 4000 pairs for training and 1000 pairs for testing.

The CDD dataset [57] contains 11 pairs of Google Earth images with a spatial resolution between 3 and 100 cm. It includes seven pairs of season-varying images with size 4725 × 2700 pixels. In addition, four season-varying image pairs with minimal changes and size of 1900 × 1000 pixels are considered to manually add additional objects. The dataset includes changes in buildings, roads, vehicles, etc., while excluding changes caused by seasonality and lighting. Following the default dataset split, the original images have been cropped and rotated to generate 16000 pairs of images with size 256 × 256 pixels: 10000/3000/3000 for training/validation/test.

B. Implementation Details and Evaluation Criteria

During training, each minibatch consists of eight labeled image pairs and eight unlabeled image pairs. The initial learning rate is set to 0.02, and the SGD optimizer is used. The model is trained for 80 epochs on LEVIR-CD, WHU-CD, and JL1-CD, and 120 epochs on CDD, using a polynomial learning rate scheduler. All the experiments are implemented in PyTorch 1.9.0 and conducted on an NVIDIA 4090 24 GB and an NVIDIA A100 80 GB GPU for training and testing.

To evaluate the performance of the proposed method, five commonly used CD accuracy indices are employed, including precision (Pre), recall (Rec), F1-score (F1), intersection over union (IoU), and overall accuracy (OA). Besides the OA, which is used to evaluate the overall performance for both the changed and unchanged classes, the other four indices are all used to

compute results for the changed class, as this task primarily focuses on the change areas. The aforementioned indices are expressed as follows:

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{14}$$

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

$$F1 = \frac{2 \times \text{Pre} \times \text{Rec}}{\text{Pre} + \text{Rec}} \tag{16}$$

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

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{18}$$

where TP, FP, FN, and TN represent true positive, false positive, false negative, and true negative, respectively.

C. Comparisons With SOTA Methods

To verify the effectiveness of the proposed method, several state-of-the-art semisupervised methods are selected as comparative techniques, including the basic Only-sup method, s4GAN [58] (which is an adversarial learning method), SemiCD [7] and C2FNet [49] (which are two consistency regularization methods), RCL [9] (which is a pseudolabeling method), SemiSANet [46], FPA [47], CutMix-CD [21], SemiCD-VL [59], and UniMatch [12] (which are five hybrid methods incorporating characteristics of consistency regularization and pseudolabeling).

Only-sup refers to the method that uses only labeled data for training the model.

s4GAN [58] uses a dual-branch approach to semisupervised semantic segmentation, consisting of an enhanced GAN-based model and a semisupervised multilabel classification branch.

SemiCD [7] introduces an unsupervised CD loss that ensures consistency in the output change probability map of unlabeled bitemporal image pairs when small random perturbations are applied to their deep feature difference map.

C2FNet [49] is an SSCD method based on consistency regularization, which includes a coarse-to-fine CD network and a semisupervised updating method using the mean teacher approach.

RCL [9] employs the prediction uncertainty of unlabeled images to select reliable pseudolabels for model training and combines features of changed areas to design pixel-level positive and negative samples for contrastive learning.

SemiSANet [46] is an SSCD method based on pseudolabeling and consistency regularization. It constructs a Siamese nested UNet with a graph attention mechanism to capture long-range feature dependencies.

FPA [47] is a model-free SSCD framework composed of two alignment strategies. The class-aware feature alignment aligns the area-level change/no-change features, while pixelwise prediction alignment aligns the pixel-level change predictions under different image augmentations.

SemiCD-VL [59] is a visual language model-guided SSCD method that leverages VLMs to generate pseudolabels, providing additional supervision for unlabeled data.

CutMix-CD [21] utilizes change-aware CutMix augmentation combined with consistency learning. It generates novel change samples from unlabeled data by synthesizing mixed change regions.

UniMatch [12] expands the perturbation space by introducing an auxiliary feature perturbation stream. In addition, a dual-stream perturbation technique is adopted to enable two strong augmentation views to be simultaneously guided by a common weak augmentation view. As mentioned earlier, we use UniMatch as our plain baseline.

Tables II compares the accuracy obtained by the proposed PAdaptCD with those of other methods on four CD datasets. On the LEVIR-CD test set, PAdaptCD achieves an F1-score of 90.01% on the 3% split, which is a 10.74% improvement over the Only-sup method. Moreover, compared to GAN-based, consistency-based and pseudolabeling methods, the proposed method maintains superior performance, e.g., it exceeds RCL by 7.74% in F1 and 11.96% in IoU^c. It also outperforms the baseline method, UniMatch, by 1.29% in F1. Similarly, on the WHU-CD test set, PAdaptCD brings significant improvements over the Only-sup method, increasing F1-scores by 33.83% and 19.09% with 178 and 297 labeled images, respectively. While SemiCD-VL achieves the highest F1-score at the 20% split, the proposed method achieves the best performance among all compared methods when trained with very few labeled samples (e.g., 3% and 5% splits), demonstrating its strong effectiveness under limited supervision.

For the two natural and human-induced CD datasets, JL1-CD and CDD, PAdaptCD also shows excellent results. On the JL1-CD test set, PAdaptCD achieves an F1-score 7.18% higher than Only-sup under the 5% label ratio. Compared to the recent CutMix-CD, PAdaptCD provides consistent advantages, improving F1-scores by 5.22% and 5.08% under the 3% and 5% splits, respectively. On the CDD test set, compared with Only-sup, the proposed method achieves performance gains of 10.03%, 11.71%, 9.30%, and 5.34% in F1 for 3%, 5%, 10%, and 20% labeled image splits, respectively, and yields the best results in all cases compared with literature methods.

To provide a more intuitive illustration of the performance differences among various methods, Fig. 6 illustrates the performance gap between PAdaptCD and other methods. The proposed method achieves superior performance, especially in scenarios with a very small number of labeled samples. Moreover, to evaluate model efficiency, Table III reports the number of parameters (Params.), floating-point operations (FLOPs), epoch time, and inference time on the LEVIR-CD dataset with 3% labeled data. Epoch time refers to the duration for one training epoch with a batch size of 8, and inference time is measured with a batch size of 1. The proposed PAdaptCD adopts the same DeepLabv3+ model (with ResNet-50 as the backbone) as UniMatch and s4GAN, resulting in an identical number of parameters and FLOPs. Minor differences in epoch time are due to varying training strategies. SemiCD, RCL, FPA, and CutMix-CD use a different ResNet-50-based model, resulting in

TABLE II
CHANGE DETECTION RESULTS ON THE LEVIR-CD, WHU-CD, JL1-CD AND CDD TEST SETS

Datasets	Method	3% (213)					5% (356)					10% (712)					20% (1424)				
		IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA
LEVIR-CD	Only-sup	65.66	79.27	91.69	69.81	98.14	69.87	82.27	89.56	76.07	98.33	75.93	86.32	91.38	81.78	98.68	78.75	88.11	92.92	83.77	98.85
	s4GAN	49.50	66.22	62.35	70.61	96.33	58.53	73.84	80.18	68.44	97.53	68.29	81.16	79.60	82.78	98.04	71.86	83.63	86.15	81.25	98.38
	SemiCD	70.12	82.43	85.46	79.62	97.89	74.88	85.63	89.57	82.03	98.60	76.82	86.89	90.66	83.42	98.72	77.45	87.29	91.33	83.60	98.76
	C2FNet	79.45	88.55	90.39	86.79	98.86	81.33	89.71	92.15	**87.39**	98.98	82.35	90.32	92.15	**88.57**	99.03	82.61	90.46	92.58	88.65	99.05
	RCL	69.88	82.27	86.44	78.48	98.28	71.52	83.40	86.58	80.45	98.37	74.93	85.67	87.92	83.53	98.58	75.66	86.14	87.48	84.84	98.61
	SemiSANet	71.37	83.30	85.83	80.91	98.35	71.63	83.47	84.28	82.68	98.33	77.11	87.07	89.01	85.22	98.71	78.69	88.07	90.52	85.76	98.82
	FPA	70.38	82.61	79.04	86.53	97.91	73.83	84.94	88.92	81.31	98.53	75.41	85.98	90.22	82.03	98.52	80.05	88.92	92.98	83.91	98.62
	CutMix-CD	69.24	81.82	85.29	78.63	98.22	77.77	87.49	89.40	85.67	98.75	79.50	88.58	89.48	87.69	98.85	81.05	89.54	90.96	88.15	98.95
	SemiCD-VL	80.65	89.29	90.87	87.76	98.93	80.72	89.33	92.50	86.38	98.95	82.30	90.29	92.43	88.25	99.03	82.41	90.36	91.60	**89.15**	99.03
	UniMatch | DeepLabv3+	79.73	88.72	**92.90**	84.90	98.90	80.00	88.89	**93.11**	85.03	98.92	81.72	89.94	**93.54**	86.61	99.01	81.65	89.90	**93.97**	86.17	99.01
	PAdaptCD(proposed)	**81.84**	**90.01**	92.05	**88.06**	**99.00**	**81.92**	**90.06**	92.91	87.38	**99.02**	**82.56**	**90.44**	93.13	87.91	**99.05**	**82.67**	**90.52**	92.72	88.42	**99.06**
WHU-CD	Method	3% (178)					5% (297)					10% (594)					20% (1189)				
		IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA
	Only-sup	38.99	56.10	55.73	56.48	96.49	55.08	71.03	68.50	73.76	97.61	61.88	76.46	72.87	80.41	98.04	47.23	64.16	55.23	76.54	96.61
	s4GAN	47.80	64.68	80.84	53.90	97.66	58.23	73.60	78.79	69.05	98.03	61.54	76.19	85.00	69.04	98.29	66.36	79.78	86.63	73.93	98.51
	SemiCD	54.67	70.70	77.64	64.89	97.87	62.64	77.03	76.65	77.40	98.17	70.44	82.66	81.93	83.40	98.61	76.43	86.64	87.80	85.52	98.95
	C2FNet	67.35	80.49	79.18	81.85	98.43	72.16	83.83	84.21	83.45	98.72	76.47	86.67	85.60	87.75	98.93	81.49	89.78	88.42	91.23	99.18
	RCL	49.59	66.30	80.31	56.45	97.72	51.14	67.67	77.76	59.91	97.73	63.12	77.39	82.50	72.87	98.31	73.70	84.86	86.56	83.23	98.82
	SemiSANet	52.71	69.03	73.10	65.39	97.67	52.06	68.48	79.03	60.41	97.79	60.58	75.45	78.94	72.25	98.13	67.59	80.66	81.15	80.18	98.47
	FPA	58.52	73.83	67.45	81.55	97.71	72.20	83.86	87.07	80.87	98.76	60.38	75.29	73.70	76.96	98.00	72.07	83.77	79.67	88.32	98.64
	CutMix-CD	60.61	75.48	76.84	74.16	98.09	61.40	76.08	77.89	74.35	98.15	68.79	81.51	84.93	78.35	98.59	76.80	86.88	82.06	92.30	98.89
	SemiCD-VL	80.97	89.48	91.71	**87.36**	99.19	81.30	89.69	95.08	84.87	99.23	81.65	89.90	92.37	87.56	99.22	**83.25**	**90.86**	**89.69**	92.06	**99.26**
	UniMatch | DeepLabv3+	80.83	89.40	**94.24**	85.03	99.20	80.67	89.30	**95.52**	83.84	99.20	81.74	89.95	91.97	88.02	99.22	81.81	90.00	88.21	91.85	99.19
	PAdaptCD(proposed)	**81.71**	**89.93**	93.75	86.41	**99.23**	**82.02**	**90.12**	93.36	**87.10**	**99.24**	**82.90**	**90.65**	**92.79**	**88.60**	**99.27**	82.05	90.14	87.94	**92.45**	99.20
JL1-CD	Method	3% (120)					5% (200)					10% (400)					20% (800)				
		IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA
	Only-sup	37.24	54.27	55.96	52.67	91.66	50.56	67.17	72.54	62.53	94.26	53.20	69.45	70.58	68.36	94.35	57.76	73.22	79.25	68.05	95.33
	s4GAN	41.87	59.03	74.06	49.07	93.60	46.78	63.75	76.64	54.57	94.17	50.65	67.24	75.67	60.50	94.46	55.21	71.14	79.04	64.69	95.07
	SemiCD	36.83	53.84	67.60	44.73	92.80	49.14	65.90	79.45	56.30	94.53	52.42	68.78	80.12	60.26	94.86	58.86	74.10	79.70	69.23	95.46
	C2FNet	39.95	57.09	73.93	46.50	93.44	53.71	69.89	76.45	64.37	94.79	57.34	72.88	80.00	66.93	95.32	61.02	75.79	79.98	72.02	95.68
	RCL	46.49	63.47	71.82	56.86	93.85	53.06	69.34	75.54	64.08	94.68	54.31	70.39	75.04	66.29	94.76	58.24	73.61	77.92	69.76	95.30
	SemiSANet	37.47	54.51	66.49	46.19	92.76	47.87	64.75	69.77	60.41	93.82	51.66	68.13	78.58	60.13	94.72	52.18	68.58	80.40	59.79	94.86
	FPA	45.78	62.80	68.69	57.85	93.57	53.90	70.04	75.99	64.96	94.78	56.42	72.14	81.35	64.80	95.30	57.50	73.02	75.77	70.46	95.11
	CutMix-CD	49.13	65.89	63.73	**68.20**	93.37	52.99	69.27	69.89	68.66	94.28	58.01	73.42	71.35	**75.62**	94.86	60.52	75.40	78.14	**72.85**	95.54
	SemiCD-VL	51.39	67.89	**86.01**	56.08	95.02	55.50	71.38	79.97	64.46	95.15	55.67	71.52	**86.65**	60.89	95.45	58.09	73.49	**86.49**	63.89	95.67
	UniMatch | DeepLabv3+	53.61	69.80	80.91	61.37	95.01	57.38	72.92	79.19	67.57	95.29	57.46	72.98	84.09	64.47	95.52	60.28	75.22	83.52	68.42	95.77
	PAdaptCD(proposed)	**55.17**	**71.11**	78.69	64.85	**95.05**	**59.17**	**74.35**	**80.33**	**69.20**	**95.52**	**59.43**	**74.55**	81.65	68.59	**95.60**	**61.46**	**76.13**	84.50	69.27	**95.92**
CDD	Method	3% (300)					5% (500)					10% (1000)					20% (2000)				
		IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA
	Only-sup	57.53	73.04	81.32	66.29	94.23	61.06	75.82	83.33	69.55	94.77	69.83	82.23	89.30	76.20	96.11	80.14	88.97	93.25	85.07	97.51
	s4GAN	53.10	69.59	71.96	67.03	93.46	58.94	74.16	76.61	71.88	94.67	66.73	80.05	86.93	74.17	95.64	73.67	84.84	89.10	80.97	96.59
	SemiCD	58.54	73.85	81.25	67.68	93.80	68.07	81.00	87.22	75.60	95.42	75.56	86.08	90.92	81.72	96.58	82.77	90.57	92.89	88.37	97.62
	C2FNet	53.35	69.71	72.22	66.92	93.54	59.97	74.98	85.12	67.00	94.72	70.71	82.84	87.97	78.28	96.17	80.30	89.08	91.22	87.03	97.48
	RCL	63.43	77.62	82.99	72.90	94.57	69.30	81.87	84.21	79.65	95.44	76.22	86.50	89.02	84.13	96.61	81.57	89.85	90.84	88.88	97.40
	SemiSANet	60.84	75.65	85.89	67.59	94.81	62.63	77.02	81.83	72.74	94.82	67.33	80.48	83.56	77.62	95.51	74.22	85.20	87.96	82.61	96.58
	FPA	64.00	78.05	87.63	70.35	94.88	70.23	82.51	91.45	75.16	95.88	77.55	87.36	91.34	83.71	96.87	81.64	89.89	93.54	86.51	97.48
	CutMix-CD	63.00	77.30	85.61	70.47	94.65	68.87	81.57	86.72	76.99	95.50	75.44	86.00	89.47	82.79	96.52	81.49	89.80	92.64	87.13	97.44
	SemiCD-VL	68.24	81.12	91.29	72.99	95.99	73.36	84.63	**93.61**	77.23	96.69	81.59	89.86	**95.28**	85.03	97.74	87.26	93.20	**96.82**	89.84	98.45
	UniMatch | DeepLabv3+	69.75	82.18	90.34	75.37	96.14	76.81	86.89	92.53	81.89	97.08	83.25	90.86	94.31	87.66	97.92	88.39	93.84	95.61	92.13	98.57
	PAdaptCD(proposed)	**71.04**	**83.07**	**91.74**	**75.90**	**96.35**	**77.83**	**87.53**	92.56	**83.02**	**97.21**	**84.38**	**91.53**	95.10	**88.22**	**98.07**	**89.23**	**94.31**	95.65	**93.00**	**98.68**

All results are described as percentages (%). The best scores are marked in Bold. The fractions N% and the following integers (m) denote the proportions and numbers of labeled images, respectively.

a higher number of parameters and greater computational overhead compared to PAdaptCD. These methods also require longer training durations. Although SemiCD-VL has fewer parameters and lower FLOPs than PAdaptCD, it has the longest training time among all methods, as it involves not only training the semisupervised model but also additional time and resources for generating pseudolabels. C2FNet has a relatively small number of parameters but exhibits higher computational complexity, and its inference time is the longest among all methods. SemiSANet has the lowest number of parameters and FLOPs, as well as the shortest epoch time, however, it converges more slowly and typically requires around 600 epochs to converge on the LEVIR-CD dataset with 3% labeled data. In summary, the proposed PAdaptCD achieves low total training time and moderate computational complexity (only higher than SemiSANet and SemiCD-VL). Although it does not lead in parameter size or inference speed, it consistently provides superior overall performance.

To further showcase the advantages of PAdaptCD, particularly with minimal labeled data, we present representative qualitative results on the LEVIR-CD and WHU-CD datasets with 3% labeled data. Fig. 7 shows the visual qualitative results on the LEVIR-CD. Overall, the proposed PAdaptCD detects more changes than other methods, thanks to the PADT strategy, which enhances the model focus on changes. Due to the disturbances caused by seasonal and illumination variations, as shown in rows (a) and (c) of Fig. 7, the CutMix-CD, SemiCD-VL, and UniMatch methods in row (a) exhibit missed detections. These methods apply CutMix between images from the same temporal phase of different pairs, which may limit their

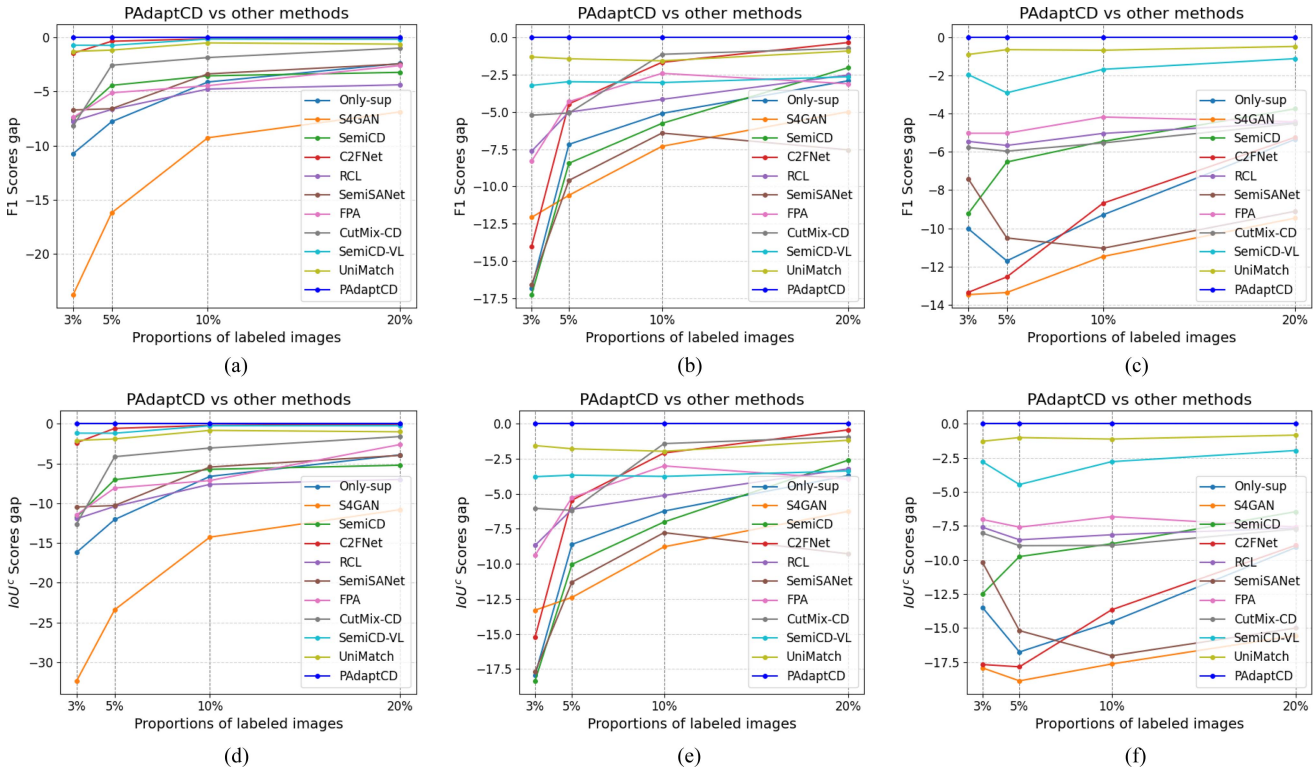

Fig. 6. Gap between the proposed PAdaptCD and other comparison methods in terms of F1 and IOU^c on the LEVIR-CD, JL1-CD, and CDD datasets. (a) and (d) LEVIR-CD dataset. (b) and (e) JL1-CD dataset. (c) and (f) CDD dataset.

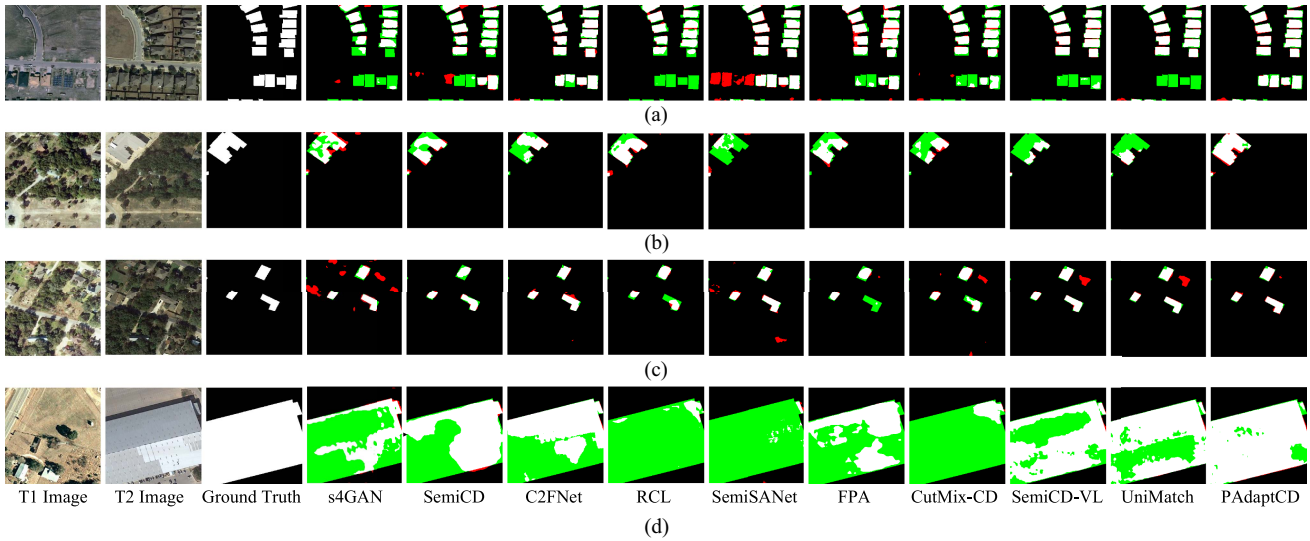

Fig. 7. Examples of qualitative results obtained by different methods on the LEVIR-CD dataset with 3% labeled data. (a)–(d) Four representative samples. For a better view, white, black, red, and green represent TP (true positive), TN (true negative), FP (false positive), and FN (false negative), respectively.

robustness to temporal style discrepancies. Similarly, in row (c) in Fig. 7, these methods produce false detections (top-right red region), possibly because seasonal vegetation variations cause the building in the T2 image to resemble the surrounding background. In contrast, PAdaptCD exhibits greater accuracy in detecting the changed buildings. This could be due to the proposed bitemporal image augmentation (BTIA), particularly the Bi-Mix strategy, which blends styles from both bitemporal images, helping to mitigate the large style gap between the two phases.

Fig. 8 shows the visual qualitative results on the WHU-CD. Similar to the findings on LEVIR-CD, the proposed method

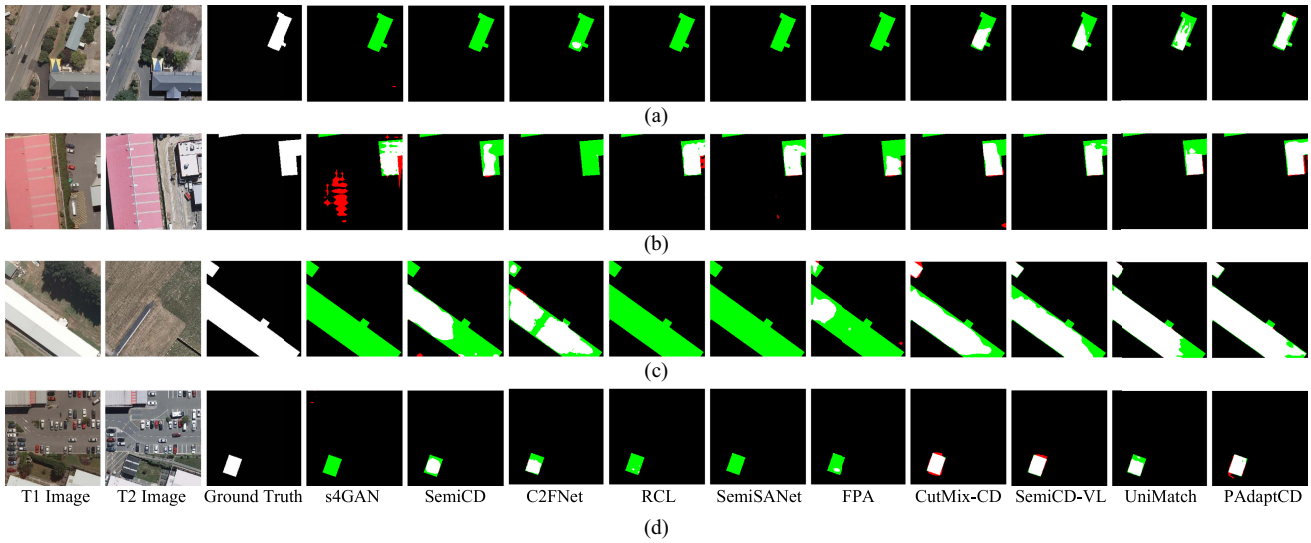

Fig. 8. Examples of qualitative results obtained by different methods on the WHU-CD dataset with 3% labeled data. (a)–(d) Four representative samples. For a better view, white, black, red, and green represent TP (true positive), TN (true negative), FP (false positive), and FN (false negative), respectively.

TABLE III
EFFICIENCY COMPARISON OF DIFFERENT MODELS ON THE LEVIR-CD WITH 3% LABELED DATA IN TERMS OF PARAMETERS (PARAMS.), FLOPS, EPOCH TIME (THE BATCH SIZE IS 8), AND INFERENCE TIME

Method	Params.(M)	FLOPs(G)	Epoch Time(s)	Inference Time(ms)
s4GAN	40.47	29.10	409.93	38.72
SemiCD	46.85	73.23	692.25	34.74
C2FNet	16.17	62.10	395.33	48.10
RCL	46.85	73.23	477.67	35.34
SemiSANet	1.61	4.67	51.45	28.42
FPA	46.85	73.23	588.67	35.25
CutMix-CD	46.85	73.23	745.95	35.94
SemiCD-VL	24.88	21.60	971.88	44.46
UniMatch | DeepLabv3+	40.47	29.10	386.42	41.97
PAdaptCD(proposed)	40.47	29.10	394.34	38.50

involves fewer missed changes than the other techniques. When facing large changes, the proposed method demonstrates higher detection integrity and compactness, as illustrated in row (c). In rows (a) and (d), where the appearance of the changed buildings closely resembles the surrounding background, such as colors similar to asphalt roads, PAdaptCD is still able to detect these changes. In contrast, most comparison methods fail to identify them.

To demonstrate the generalization capability of the proposed method PAdaptCD for detecting changes in more common objects, Fig. 9 presents the visual qualitative results on JL1-CD. In rows (a) and (b), which involve changes in roads and the widening of water bodies, respectively, PAdaptCD achieves high accuracy in identifying such narrow and elongated changes. In rows (c) and (d), which show changes in hardened surfaces and cropland, PAdaptCD detects the changed areas more completely than other methods.

Based on the visual comparison, the proposed PAdaptCD demonstrates superior CD performance with a small number of labeled samples. This may be attributed to the PADT strategy, which enables the model to select more pseudolabels for changed pixels during the initial training phase, making a more effective use of this minority class in the unlabeled data. In addition, the use of the strong augmentation BTIA technique specifically designed for change detection reduces the model sensitivity to disturbances such as seasonal or illumination changes between the two bitemporal images.

D. Pseudolabel Selection Reliability

To intuitively demonstrate the effectiveness of PAdaptCD, we visualize the selection results of pseudolabels during training, using the LEVIR-CD dataset with 3% labeled data. In Fig. 10, the red dashed line and the blue solid line represent the ratio of unchanged pixels to changed pixels in the pseudolabels selected by UniMatch [12] (with a fixed threshold of 0.95) and the proposed PAdaptCD (with PADTs), respectively.

As shown in Fig. 10, the overall trend of the pixel ratio obtained by UniMatch and PAdaptCD is similar. This is because both methods employ the same CD networks. However, at the early stages of training, the pixel ratio in PAdaptCD is substantially lower compared to UniMatch. For instance, in the first epoch, PAdaptCD achieves a pixel ratio of 62.87, while UniMatch's ratio stands at 142.75. This result aligns with our expectations: the adaptive threshold effectively mitigates the severe class imbalance in the initial part of training, leading to faster model convergence.

To further highlights the reliability of pseudolabel selection with PAdaptCD, Fig. 11 compares the number of accurate pseudolabels obtained from the UniMatch and PAdaptCD methods, respectively. It is evident that for the minority class of changed pixels, PAdaptCD consistently selects more accurate pseudolabels than UniMatch, especially in the early stages of training. While PAdaptCD produces slightly more incorrect pseudolabels than UniMatch, this can be overlooked considering

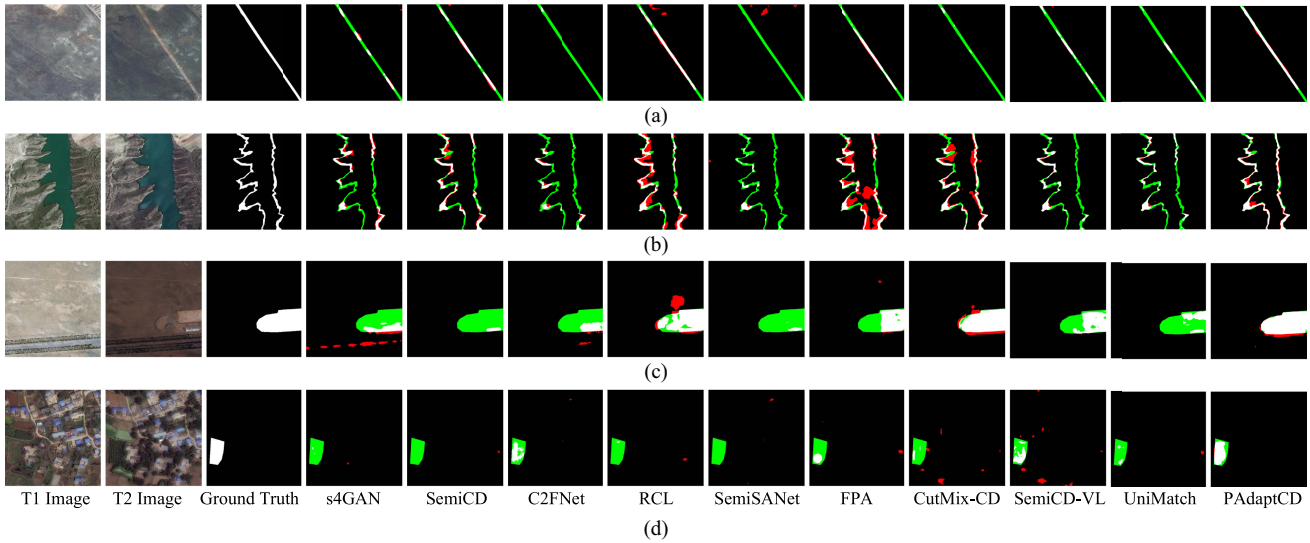

Fig. 9. Examples of qualitative results obtained by different methods on the JL1-CD dataset with 10% labeled data. (a)–(d) Four representative samples. For a better view, white, black, red, and green represent TP (true positive), TN (true negative), FP (false positive), and FN (false negative), respectively.

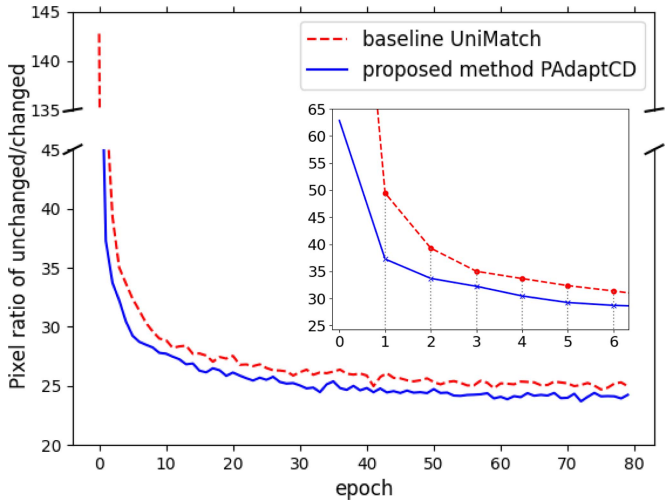

Fig. 10. Pixel ratio of predicted unchanged/changed pixels in unlabeled images from the training set of the LEVIR-CD dataset with 3% labeled data, using a fixed threshold (UniMatch) and PADTs (the proposed PAdaptCD), respectively. The red dash line and blue solid line represent the pixel ratios of unchanged/changed pixels in the pseudolabels selected by UniMatch and PAdaptCD, respectively.

the significant increase in correct pseudolabels. For unchanged pixels, both methods show little difference in the number of correct pseudolabels, while UniMatch tends to produce slightly more incorrect pseudolabels. This further illustrates the superiority of the proposed method in addressing class imbalance during pseudolabel selection, as it selects a greater number of accurate pseudolabels for the minority class, thereby enhancing the exploration of unlabeled data.

E. Ablation Studies

In this section, we conduct a series of experiments to investigate the effectiveness of our proposed PADT and (BTIA strategies.

1) Effectiveness of the PADT: The PADT aims to mitigate the impact of sample imbalance during the self-training by setting adaptive thresholds for changed and unchanged categories when selecting pseudolabels. It incorporates three key components: EMA, progressive increment adjustment, and dual-threshold strategies. To assess the effectiveness of PADT, we replaced it with a fixed threshold (0.95, as in UniMatch). As shown in Table IV, removing PADT (#2 in Table IV) results in a decrease in F1-score of 1.06% and 1.34% on the 3% and 5% labeled JL1-CD datasets, respectively. In addition, using PADT without EMA (#3 in Table IV) leads to worse performance. This may be because EMA maintains smoother threshold updates, reducing sensitivity to outlier data. In the PADT variant without progressive increment adjustment (#4 in Table IV), where dual-threshold values are directly obtained after the EMA step, performance also drops, indicating that this adjustment aligns with the training process and ensures model stability. Finally, testing PADT without dual-threshold (#5 in Table IV) using a single threshold $\tau_C^{p-\text{adp}}(t)$ for pseudolabel selection, results in poor performance. This highlights the critical role of dual-threshold values in PADT, as they help balance the number of changed and unchanged pixels, mitigating bias in pseudolabeling and preventing degradation in iterative self-training.

To further demonstrate the advantage of PADT, we conducted experiments using the adaptive threshold strategies from other methods, AdaptMatch [44] and FreeMatch [43], on three datasets: LEVIR-CD, JL1-CD, and CDD, which contain different types of changes. AdaptMatch calculates adaptive thresholds using the historical predicted probabilities from both the labeled branch and the weakly augmented unlabeled branch, stored in two memory banks. FreeMatch computes class-specific thresholds in two steps: it first estimates a global threshold via the EMA of model confidence on unlabeled data, and then, scales it with the normalized classwise EMA of predicted probabilities to obtain local thresholds. Unlike AdaptMatch, the proposed PADT derives classwise adaptive thresholds based solely on

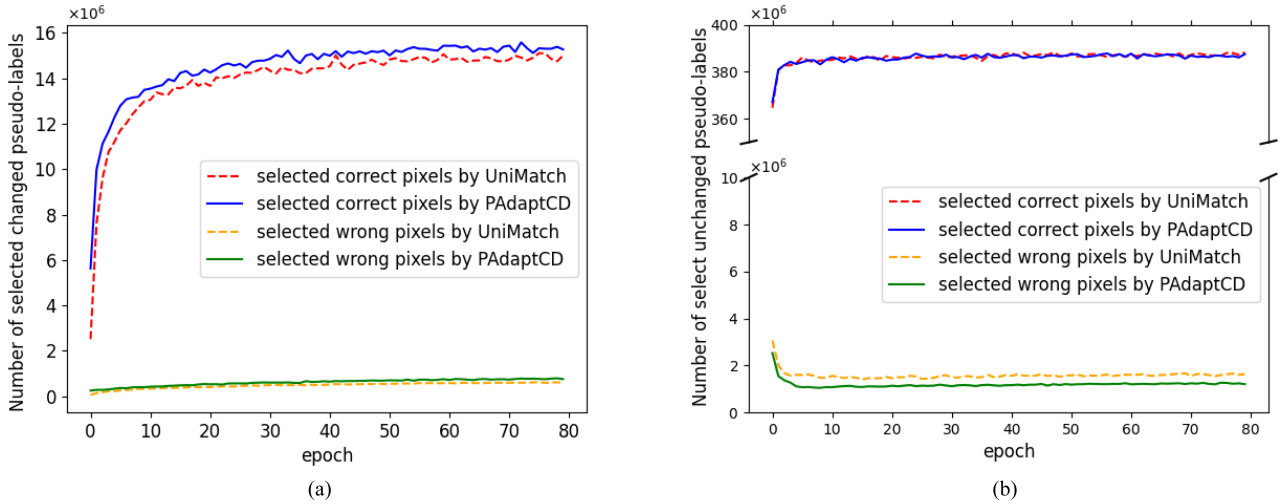

Fig. 11. Numbers of selected pseudolabels for changed/unchanged pixels in unlabeled images from the training sets of the LEVIR-CD dataset with 3% labeled data, using a fixed threshold (UniMatch) and PADTs (the proposed PAdaptCD), respectively. The red dash line and blue solid line represent the selected of correct changed/unchanged pixels by UniMatch/PAdaptCD. The orange dashed line and green solid line represent the selected of wrong changed/unchanged pixels by UniMatch/PAdaptCD. (a) Number of selected unlabeled examples with changed pseudolabels. (b) Number of selected unlabeled examples with unchanged pseudolabels.

TABLE IV
ABLATION STUDIES OF THE PROPOSED METHOD ON THE JL1-CD AND CDD DATASET WITH 3% /5% LABELED DATA

No.	Variants	JL1-CD				CDD			
		3% (120)		5% (200)		3% (300)		5% (500)	
		F1	IOU^c	F1	IOU^c	F1	IOU^c	F1	IOU^c
#1	PAdaptCD	71.11	55.17	74.35	59.17	83.07	71.04	87.53	77.83
(a) PADT									
#2	w/o PADT	70.05	53.91	73.01	57.49	82.55	70.28	86.01	75.46
#3	PADT (w/o EMA)	69.93	53.76	73.25	57.80	81.58	68.89	86.59	76.36
#4	PADT (w/o Progressive Increment Adjustment)	69.79	53.60	72.70	57.11	81.72	69.10	86.42	76.08
#5	PADT (w/o Dual-Threshold)	69.68	53.46	72.46	56.82	81.53	68.82	86.51	76.22
(b) BTIA									
#6	w/o BTIA	70.17	54.04	73.46	58.05	82.64	70.41	86.37	76.02
#7	BTIA(w/o HRIA)	69.85	53.66	73.12	57.63	81.62	68.95	86.55	76.29
#8	BTIA (w/o Bi-Mix)	69.77	53.58	72.77	57.20	81.62	68.94	86.15	75.67

All results are described as percentages (%).

unlabeled data, and does not rely on memory banks. Moreover, in FreeMatch, class-specific thresholds are derived by normalizing classwise prediction confidence, which makes them sensitive to class distribution. In contrast, PADT treats each class independently and computes thresholds based on the predicted probabilities of each class. Furthermore, a progressive increment mechanism is designed to adjust the thresholds dynamically according to the training status, thereby ensuring the reliability of the generated pseudolabels. In these experiments, we replaced the PADT module in the proposed PAdaptCD with the adaptive thresholding strategies from AdaptMatch and FreeMatch, respectively. The results are shown in Table V. It can be observed that the proposed PADT achieves the highest F1-scores across all datasets under different labeled image splits. Meanwhile, the adaptive thresholding strategies from FreeMatch achieve higher recall compared to PAdaptCD, whereas PAdaptCD shows better precision. This indicates that PADT selects more reliable pseudolabels, discarding a greater number of inaccurate ones.

2) Effectiveness of BTIA: The BTIA is specifically designed for the SSCD task. To verify the effectiveness of BTIA, we replaced it with the original perturbations used in UniMatch (#6 in Table IV). The results show a noticeable drop in performance can be observed. BTIA consists of two key modules: HRIA and Bi-Mix. HRIA applies highly random intensity-based augmentations and distortion strengths from a continuous space, aiming to mitigate overdistortion issues. To verify the effectiveness of HRIA, we replaced it with the original strong perturbations of brightness, contrast, saturation, and hue in UniMatch (#7 in

TABLE V
DIFFERENT ADAPTIVE THRESHOLD METHODS ON THE LEVIR-CD, JL1-CD AND CDD TEST SETS

Datasets	Method	3% (213)					5% (356)					10% (712)					20% (1424)				
		IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA
LEVIR-CD	*AdaptMatch	81.04	89.53	91.35	87.77	98.95	81.47	89.79	91.97	87.71	98.98	82.27	90.27	92.48	88.17	99.03	82.33	90.31	92.63	88.10	99.04
	*FreeMatch	79.88	88.81	87.63	**90.03**	98.84	81.46	89.78	88.65	**90.95**	98.95	81.85	90.02	89.16	**90.90**	98.97	81.90	90.05	88.60	**91.56**	98.97
	PAdaptCD(proposed)	**81.84**	**90.01**	**92.05**	88.06	**99.00**	**81.92**	**90.06**	**92.91**	87.38	**99.02**	**82.56**	**90.44**	**93.13**	87.91	**99.05**	**82.67**	**90.52**	**92.72**	88.42	**99.06**
JL1-CD	Method	3% (120)					5% (200)					10% (400)					20% (800)				
		IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA
	*AdaptMatch	54.29	70.37	77.93	64.15	94.93	58.03	73.44	**80.52**	67.51	95.42	58.98	74.20	76.46	72.06	95.29	59.80	74.84	77.48	72.38	95.43
	*FreeMatch	54.68	70.70	74.74	**67.08**	94.78	57.17	72.75	74.52	**71.07**	95.00	58.97	74.19	73.57	**74.82**	95.11	61.13	75.88	76.34	**75.42**	95.50
	PAdaptCD(proposed)	**55.17**	**71.11**	**78.69**	64.85	**95.05**	**59.17**	**74.35**	80.33	69.20	**95.52**	**59.43**	**74.55**	**81.65**	68.59	**95.60**	**61.46**	**76.13**	**84.50**	69.27	**95.92**
CDD	Method	3% (300)					5% (500)					10% (1000)					20% (2000)				
		IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA	IOU^c	F1	Pre	Rec	OA
	*AdaptMatch	69.65	82.11	91.39	74.54	96.17	76.79	86.87	**93.27**	81.29	97.10	82.76	90.57	**95.43**	86.18	97.88	88.35	93.81	95.14	92.53	98.56
	*FreeMatch	64.67	78.55	77.65	**79.46**	94.88	73.04	84.42	80.17	**89.15**	96.12	79.24	88.42	85.86	**91.12**	97.18	88.10	93.67	91.26	**96.22**	98.47
	PAdaptCD(proposed)	**71.04**	**83.07**	**91.74**	75.90	**96.35**	**77.83**	**87.53**	92.56	83.02	**97.21**	**84.38**	**91.53**	95.10	88.22	**98.07**	**89.23**	**94.31**	**95.65**	93.00	**98.68**

All results are described as percentages (%). The best scores are marked in bold. The fractions N% and the following integers (m) denote the proportions and numbers of labeled images, respectively. * indicates that the adaptive thresholding Strategy of the corresponding method is integrated into the proposed PAdaptCD.

TABLE VI
ABLATION STUDIES OF THE VARIOUS CUTMIX STRATEGIES ON THE LEVIR-CD WITH 3% /5% LABELED DATA

No.	CutMix Variants	3% (213)		5% (356)	
		F1	IOU^c	F1	IOU^c
#1	CutMix-Same Temporal	89.06	80.28	89.21	80.53
#2	CutMix-Same Pair	89.55	81.07	89.54	81.06
#3	CutMix-Dual Diff	90.01	81.84	90.06	81.92

All results are described as percentages (%).

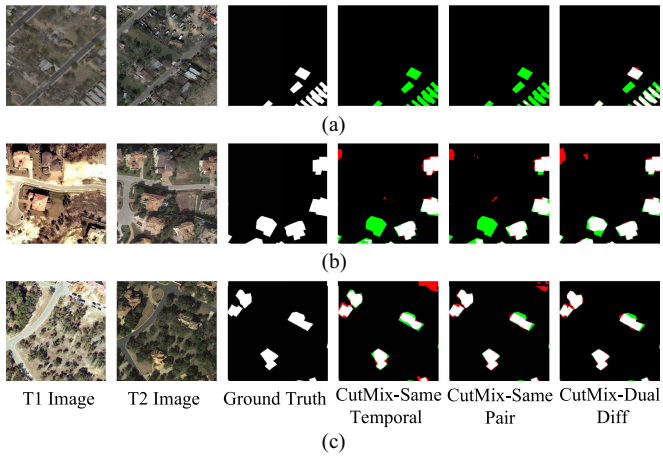

Fig. 12. Examples of qualitative results obtained by several variants of CutMix on the LEVIR-CD dataset with 3% labeled data. (a)–(c) Three representative samples. For a better view, white, black, red, and green represent TP (true positive), TN (true negative), FP (false positive), and FN (false negative), respectively.

Table IV). The experimental results demonstrate the advantages of highly random perturbations. The Bi-Mix mechanism "CutMix-Dual Diff" is designed to better exploit SSCD. When it is replaced with the original CutMix operation in UniMatch (#8 in Table IV), the F1-score drops by 1.45% /1.38% on the 3% /5% labeled CDD dataset.

To further verify that the Bi-Mix strategy is tailored for the SSCD task, we explore several variants of CutMix, including "CutMix-Same Temporal" (applying CutMix to the same temporal images across different pairs), "CutMix-Same Pair" (CutMix applied within the same bitemporal image pair across different temporal information), and our proposed "CutMix-Dual Diff" (CutMix applied across different temporal images from different pairs); see Fig. 4 for more details. Experimental results for these cases are presented in Table VI, and the analysis is as follows.

1) Applying CutMix to the same temporal images across different pairs (#1 in Table VI) is not a desirable approach. Temporal spectral variations, caused by differences in acquisition time, often affect the representation in bitemporal image pairs. Mixing within the same temporal images fails to mitigate these variations, leading to poor outcomes.
2) Applying CutMix within the same bitemporal image pair (#2 in Table VI) can effectively suppress differences in unchanged regions while emphasizing changes. However, this approach lacks information from other image pairs, reducing data diversity and leading to suboptimal performance.
3) The proposed Bi-Mix mechanism "CutMix-Dual Diff" (#3 in Table VI) achieves the best results, in line with our design objectives. It reduces the negative impact of domain differences between bitemporal images and increases data diversity, allowing the model to learn from a broader range of scenarios and enhancing its generalization ability.

Fig. 12 presents the visual qualitative results of different CutMix variants to intuitively demonstrate the advantages of the proposed Bi-Mix strategy. We selected several bitemporal image samples from the LEVIR-CD dataset. In rows (a) and (b) in Fig. 12, compared with the Bi-Mix mechanism ("CutMix-Dual Diff"), the other two CutMix variants result in a large number of missed detections. Seasonal and imaging condition variations make it challenging to identify the actual changed regions. In row (c) in Fig. 12, "CutMix-Same Temporal" results in false detections (top-right red region), where the apparent differences are caused by imaging variations rather than real changes, and the building in the T1 image blends with the surroundings. These examples demonstrate the effectiveness of the Bi-Mix strategy in mitigating domain discrepancies between bitemporal images, achieving the best results in cases with seasonal and imaging condition variations.

TABLE VII
ABLATION STUDIES OF THE MAXIMUM NUMBER OF SELECTED INTENSITY-BASED AUGMENTATIONS K ON THE WHU-CD WITH 3% /5% LABELED DATA

k	3% (178)		5% (297)	
	F1	IOU^c	F1	IOU^c
0	83.58	71.80	87.27	77.42
1	88.59	79.52	89.84	81.55
2	89.93	81.71	90.12	82.02
3	89.22	80.54	88.98	80.16
4	88.86	79.95	89.13	80.39
5	89.06	80.27	88.95	80.10
6	87.03	77.04	89.00	80.19

All results are described as percentages (%).

TABLE VIII
ABLATION STUDIES OF THE HYPER-PARAMETER T IN THE PADT ON THE JL1-CD WITH 5% /10% LABELED DATA

T	5% (200)		10% (400)	
	F1	IOU^c	F1	IOU^c
1.050	71.99	56.24	73.37	57.94
1.075	73.15	57.67	73.45	58.04
1.100	74.35	59.17	74.55	59.43
1.125	73.68	58.32	73.31	57.86
1.150	72.95	57.41	73.22	57.75

All results are described as percentages (%).

3) Hyperparameter Evaluations: To investigate the impact of the maximum number of selected intensity-based augmentations k, we set different values for k. In Table VII, $k = 0$ means that no HRIA is applied. This results in particularly poor accuracy, which demonstrates the relevance of strong augmentation. It can be observed that when $k = 1$, the results decline, which might be due to insufficient data augmentation. Proper data augmentation is crucial for contrastive learning-based training, as it introduces nonessential variations while preserving essential semantic features, thus enhancing the learning of representative and generalizable embeddings. By observation, a larger k also leads to degradation, as for the CD task, the bitemporal images provide an additional source of natural variation through the time dimension, which complements artificial image augmentations. However, excessive artificial augmentation can lead to information loss. We also examine the hyperparameter T in the PADT, as shown in Table VIII. It can be observed that when $T = 1.100$, the performance consistently outperforms others. In the results reported in this article, we have set $k = 2$ and $T = 1.100$ as the default for all runs.

The setting of the hyperparameter τ_{upper} in (12) is critical, as it significantly influences the performance of the proposed method. When τ_{upper} is set too low, a larger number of pseudolabels are incorporated into training, potentially introducing more noisy labels and degrading model accuracy. In contrast, an excessively high threshold may improve the accuracy of pseudolabels but can lead to an insufficient number being selected, particularly for changed samples, which are typically much fewer than unchanged samples in CD tasks. This may reduce the model's sensitivity to change regions and increase the risk of omission errors. To determine an appropriate value for τ_{upper}, we conduct experiments on the CDD dataset. As shown in Table IX, the best performance is achieved when $\tau_{\text{upper}} = 0.95$. Therefore, we adopt $\tau_{\text{upper}} = 0.95$ as the default setting throughout this article.

TABLE IX
ABLATION STUDIES OF THE HYPER-PARAMETER τ_{UPPER} ON THE CDD WITH 3% /10% LABELED DATA

τ_{upper}	3% (300)		10% (1000)	
	F1	IOU^c	F1	IOU^c
0.80	82.27	69.88	91.24	83.88
0.90	82.88	70.76	91.40	84.16
0.94	82.00	69.49	90.88	83.28
0.95	83.07	71.04	91.53	84.38
0.96	81.34	68.55	91.02	83.52
0.97	81.35	68.57	90.75	83.07

All results are described as percentages (%).

V. DISCUSSION

A. Generalization Ability Evaluation

In order to demonstrate the generalizability of the proposed semisupervised strategy, we name it the PAdapt strategy, which comprises both PADT and BTIA techniques. We evaluate its effectiveness by integrating it with various CD methods from the literature. Since PAdapt focuses on pseudolabel selection and data augmentation, we select corresponding methods for integration. Specifically, we incorporate PAdapt into FPA [47], SemiCD-VL [59], and UniMatch (with PSPNet) [12], and conduct experiments on three representative datasets, LEVIR-CD, JL1-CD, and CDD.

As shown in Table X, introducing the proposed strategy consistently improves performance. For example, under the 3% label ratio on the LEVIR-CD dataset, incorporating PAdapt into FPA and UniMatch improves the F1-scores by 3.16% and 3.50%, respectively. On the JL1-CD dataset with 3% labeled samples, integrating PAdapt into SemiCD-VL and UniMatch improves their F1-scores by 1.51% and 1.63%, respectively. These results demonstrate that the proposed semisupervised strategy can enhance the performance of various CD models, highlighting its strong generalizability.

B. Limitations and Prospects

The proposed method PAdaptCD has achieved excellent performance, particularly with minimal labeled data, as discussed in Section IV-C. However, there are still some limitations in our model. A considerable number of changed areas remain undetected. As shown in Table II, the recall is significantly lower than the precision in most cases, and this discrepancy becomes more pronounced under lower label ratios. This indicates that our pseudolabel selection process may overly prioritize accuracy while overlooking a broader coverage of change regions. Future work should aim to better balance the quantity and quality of selected pseudolabels. On the other hand, the current Bi-Mix strategy selects mixing masks randomly, which may, to

TABLE X
GENERALIZATION ABILITY EXPERIMENTAL RESULTS OF THE PROPOSED SEMISUPERVISED STRATEGY ON THE LEVIR-CD, JL1-CD AND CDD TEST SETS

Methods	LEVIR-CD				JL1-CD				CDD			
	3%(213)		5%(356)		3% (120)		5% (200)		3% (300)		5% (500)	
	F1	IOU^c	F1	IOU^c	F1	IOU^c	F1	IOU^c	F1	IOU^c	F1	IOU^c
FPA	82.61	70.38	84.94	73.83	62.80	45.78	70.04	53.90	78.05	64.00	82.51	70.23
+PAdapt	85.77+3.16	75.08+4.70	86.30+1.36	75.90+2.07	64.02+1.22	47.09+1.31	70.60+0.56	54.56 +0.66	79.15+1.10	65.49 +1.49	83.29+0.78	71.37+1.14
SemiCD-VL	89.29	80.65	89.33	80.72	67.89	51.39	71.38	55.50	81.12	68.24	84.63	73.36
+PAdapt	89.97+0.68	81.77+1.12	89.95+0.62	81.73 +1.01	69.40 +1.51	53.14 +1.75	72.38 +1.00	56.71 +1.21	81.71 +0.59	69.07 +0.83	85.75 +1.12	75.06+1.70
UniMatch | PSPNet	83.40	71.52	86.06	75.52	63.02	46.01	70.29	54.19	77.75	63.60	84.57	73.27
+PAdapt	86.90+3.50	76.84+5.32	87.07+1.01	77.10+1.58	64.65+1.63	47.77+1.76	71.46+1.17	55.60+1.41	78.87+1.12	65.11+1.51	85.72+1.15	75.01 +1.74

All results are described as percentages (%). The fractions n% and the following integers (m) denote the proportionsand numbers of labeled images, respectively. The (+number) denotes the absolute improvements brought by the PAdapt strategy.

some extent, limit further performance improvements. There are still promising directions to explore for Bi-Mix. For example, creating binary masks M based on model predictions could enable finer grained mixing. Ensuring the boundary integrity of change regions within the mixed areas might also be beneficial. Furthermore, it remains an open question whether mixing more changed or unchanged regions is more suitable for SSCD tasks, or whether this distinction has little effect, which deserves further exploration.

VI. CONCLUSION

In this article, we propose a novel SSCD method named PAdaptCD, which integrates consistency regularization and pseudolabeling techniques. By implementing PADT, more changed pixels are selected as pseudolabels with their accuracy maintained as much as possible, effectively alleviating the exacerbated sample imbalance problem in the RSCD task during semisupervised training. In addition, we present the BTIA technique tailored for semisupervised RSCD. Specifically, the HRIA generates appropriate strong augmentations, while the Bi-Mix mechanism enhances the interaction of information across different temporal phases and image pairs. Results on four benchmark datasets with very few labeled samples confirm that the proposed method significantly improves the selection of pseudolabels, leading to the detection of more changed pixels. Furthermore, it suppresses the impact of varying representations in bitemporal images. Extensive experiments under various labeled data amounts have shown that PAdaptCD exhibits good generalization and effective detection performance.

In the future, we will further extend the Bi-Mix technique, such as by creating binary masks M based on predictions to enable finer-grained mixing areas, enhancing its potential for RSCD tasks.

REFERENCES

[1] G. Xian, C. Homer, and J. Fry, "Updating the 2001 national land cover database land cover classification to 2006 by using landsat imagery change detection methods," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1133–1147, 2009.

[2] F. Bovolo and L. Bruzzone, "A split-based approach to unsupervised change detection in large-size multitemporal images: Application to tsunami-damage assessment," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1658–1670, Jun. 2007.

[3] J. Gao and Y. Liu, "Determination of land degradation causes in Tongyu county, Northeast China via land cover change detection," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 12, no. 1, pp. 9–16, Feb. 2010.

[4] H. Chen, Z. Li, J. Wu, W. Xiong, and C. Du, "SemiRoadExNet: A semisupervised network for road extraction from remote sensing imagery via adversarial learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 198, pp. 169–183, Apr. 2023.

[5] Y. Zheng et al., "Semi-supervised adversarial semantic segmentation network using transformer and multiscale convolution for high-resolution remote sensing imagery," *Remote Sens.*, vol. 14, no. 8, 2022, Art. no. 1786.

[6] 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.

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

[8] S. Hafner, Y. Ban, and A. Nascetti, "Urban change detection using a dual-task Siamese network and semi-supervised learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Kuala Lumpur, Malaysia, 2022, pp. 1071–1074.

[9] 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.

[10] L. Kondmann, S. Saha, and X. X. Zhu, "SemiSiROC: Semisupervised change detection with optical imagery and an unsupervised teacher model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3879–3891, 2023.

[11] Z. Ke, D. Qiu, K. Li, Q. Yan, and R. W. Lau, "Guided collaborative training for pixel-wise semi-supervised learning," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 429–445.

[12] L. Yang, L. Qi, L. Feng, W. Zhang, and Y. Shi, "Revisiting weak-to-strong consistency in semi-supervised semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 7236–7246.

[13] J. Yuan, Y. Liu, C. Shen, Z. Wang, and H. Li, "A simple baseline for semi-supervised semantic segmentation with strong data augmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Montreal, QC, Canada, 2021, pp. 8209–8218, doi: 10.1109/ICCV48922.2021.00812.

[14] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, "MixMatch: A holistic approach to semi-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 5049–5059.

[15] Z. Zhen, L. Yang, S. Long, J. Pi, L. Zhou, and J. Wang, "Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 11350–11359.

[16] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 2613–2622, doi: 10.1109/CVPR46437.2021.00264.

[17] Z. Feng et al., "DMT: Dynamic mutual training for semi-supervised learning," *Pattern Recognit.*, vol. 130, 2022, Art. no. 108777.

[18] A. Tarvainen and V. Harri, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1195–1204.

[19] S. Feng et al., "Transformer-based cross-domain few-shot learning for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5501716.

[20] D. Peng, M. Liu, and H. Guan, "Perturbation matters: A novel approach for semi-supervised remote sensing imagery change detection," *Remote Sens.*, vol. 17, no. 4, 2025, Art. no. 576.

[21] Q. Shu et al., "CutMix-CD: Advancing semi-supervised change detection via mixed sample consistency," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 4400915.

[22] H. Chen et al., "Exchange means change: An unsupervised single-temporal change detection framework based on intra-and inter-image patch exchange," *ISPRS J. Photogramm. Remote Sens.*, vol. 206, pp. 87–105, 2023.

[23] 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.

[24] Y. Xu et al., "Remote sensing change detection method based on dynamic adaptive focal loss," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5621515.

[25] X. Zhang, X. Huang, and J. Li, "Joint self-training and rebalanced consistency learning for semi-supervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5406613.

[26] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional Siamese networks for change detection," in *Proc. 25th IEEE Int. Conf. Image Process.*, Oct. 2018, pp. 4063–4067.

[27] X. Ou, L. Liu, B. Tu, L. Qing, G. Zhang, and Z. Liang, "CB-WMSSANet: A CNN framework with compact band weighting and multiscale spatial attention for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5507618, doi: 10.1109/TGRS.2023.3263563.

[28] Z. Lv, P. Zhong, W. Wang, Z. You, and N. Falco, "Multiscale attention network guided with change gradient image for land cover change detection using remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 2501805.

[29] Z. Lv, F. Wang, G. Cui, J. A. Benediktsson, T. Lei, and W. Sun, "Spatial–spectral attention network guided with change magnitude image for land cover change detection using remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4412712.

[30] Q. Shi, M. Liu, S. Li, X. Liu, F. Wang, and L. Zhang, "A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5604816.

[31] 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, May 2020, Art. no. 1662.

[32] C. Zhang, L. Wang, S. Cheng, and Y. Li, "SwinSUNet: Pure transformer network for remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5224713, doi: 10.1109/TGRS.2022.3160007.

[33] Y. Feng, H. Xu, J. Jiang, H. Liu, and J. Zheng, "ICIF-Net: Intra-scale cross-interaction and inter-scale feature fusion network for bitemporal remote sensing images change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4410213.

[34] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, "Vision Mamba: Efficient visual representation learning with bidirectional state space model," in *Proc. 41st Int. Conf. Mach. Learn.*, 2024, pp. 62429–62442.

[35] Y. Liu et al., "VMamba: Visual state space model," in *Proc. Adv. Neural Inf. Process. Syst.*, 2024, pp. 103031–103063.

[36] H. Chen, J. Song, C. Han, J. Xia, and N. Yokoya, "ChangeMamba: Remote sensing change detection with spatiotemporal state space model," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4409720.

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

[38] L. Yang, W. Zhuo, L. Qi, Y. Shi, and Y. Gao, "ST: Make self-training work better for semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 4268–4277.

[39] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6023–6032.

[40] L. Bai, H. Wang, X. Zhang, W. Qin, B. Liu, and S. Du, "AP-Semi: Improving the semi-supervised semantic segmentation for VHR images through adaptive data augmentation and prototypical sample guidance," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5631813.

[41] V. Olsson, W. Tranheden, J. Pinto, and L. Svensson, "ClassMix: Segmentation-based data augmentation for semi-supervised learning," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2021, pp. 1368–1377.

[42] B. Zhang et al., "FlexMatch: Boosting semi-supervised learning with curriculum pseudo labeling," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2021, pp. 18408–18419.

[43] Y. Wang, Y. Shen, Z. Li, H. Xu, Y. Bai, and Z. Liu, "FreeMatch: Self-adaptive thresholding for semi-supervised learning," in *Proc. Int. Conf. Learn. Representations*, 2023, doi: 10.48550/arXiv.2205.07246.

[44] W. Huang, Y. Shi, Z. Xiong, and X. X. Zhu, "AdaptMatch: Adaptive matching for semisupervised binary segmentation of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5625416.

[45] W. Zhang et al., "Semi-supervised change detection with data augmentation and adaptive thresholding for high-resolution remote sensing images," *Remote Sens.*, vol. 17, no. 2, 2025, Art. no. 178.

[46] C. Sun et al., "SemiSANet: A semi-supervised high-resolution remote sensing image change detection model using Siamese networks with graph attention," *Remote Sens.*, vol. 14, no. 12, 2022, Art. no. 2801.

[47] 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.

[48] S. Saha, L. Mou, X. X. Zhu, F. Bovolo, and L. Bruzzone, "Semisupervised change detection using graph convolutional network," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 4, pp. 607–611, Apr. 2021.

[49] C. Han, C. Wu, M. Hu, J. Li, and H. Chen, "C2F-SemiCD: A coarse-to-fine semi-supervised change detection method based on consistency regularization in high-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4702621.

[50] S. Yuan, R. Zhong, C. Yang, Q. Li, and Y. Dong, "Dynamically updated semi-supervised change detection network combining cross-supervision and screening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–14, 2024.

[51] Y. Zuo et al., "Robust instance-based semi-supervised learning change detection for remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4404815.

[52] L. Wang, M. Zhang, and W. Shi, "STCRNet: A semi-supervised network based on self-training and consistency regularization for change detection in VHR remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 2272–2282, 2024.

[53] P. Lv, M. Li, and Y. Zhong, "A semi-supervised pyramid cross-temporal attention transformer for change detection in high-resolution remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024, Art. no. 2503305, doi: 10.1109/LGRS.2024.3404645.

[54] P. Lv, P. Cheng, C. Ma, and Y. Zhong, "A semi-supervised semantic and spatial change detail retention network for semantic change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4416516, doi: 10.1109/TGRS.2024.3497983.

[55] 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. 2019.

[56] Z. Liu et al., "JL1-CD: A new benchmark for remote sensing change detection and a robust multi-teacher knowledge distillation framework," 2025, *arXiv:2502.13407*.

[57] 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," *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 42, no. 2, pp. 565–571, 2018.

[58] 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. 2021, doi: 10.1109/TPAMI.2019.2960224.

[59] K. Li et al., "SemiCD-VL: Visual-language model guidance makes better semi-supervised change detector," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5601613, doi: 10.1109/TGRS.2024.3512548.

Linlin Wang received the B.S. degree in electronic and information engineering, in 2017, from Harbin Engineering University, Harbin, China, and the M.S. degree in electronic and information engineering, in 2019, from the Harbin Institute of Technology, Harbin, where she is currently working toward the Ph.D. degree in information and communication engineering.

Her research interests include multitemporal image analysis and change detection and its applications.

Junping Zhang (Senior Member, IEEE) received the B.S. degree in biomedical engineering and instrument from Harbin Engineering University and Harbin Medical University, Harbin, China, in 1993, and the M.S. and Ph.D. degrees in signal and information processing from the Harbin Institute of Technology (HIT), Harbin, in 1998 and 2002, respectively.

She is currently a Professor with the Department of Information Engineering, School of Electronics and Information Engineering, HIT. Her research interests include hyperspectral data analysis and image processing, multisource information fusion, multitemporal change detection, and machine learning and its applications.

Dong Chen (Student Member, IEEE) received the B.S. degree in electronic information engineering from the Dalian University of Technology, Dalian, China, in 2019, and the M.S. degree in communication and information system from the Nanjing Research Institute of Electronic Technology, Nanjing, China, in 2022. He is currently working toward the Ph.D. degree in electronic information with the Harbin Institute of Technology, Harbin, China.

His research interests include hyperspectral unmixing, multitemporal hyperspectral image change detection, and their applications.

Lorenzo Bruzzone (Fellow, IEEE) received the Laurea (M.S.) degree in electronic engineering (summa cum laude) and the Ph.D. degree in telecommunications from the University of Genoa, Genoa, Italy, in 1993 and 1998, respectively.

He is currently a Full Professor of telecommunications with the University of Trento, Trento, Italy, where he teaches remote sensing, radar, and digital communications. He is also the Founder and the Director of the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento. He promotes and supervises research within the frameworks of many national and international projects related to Earth observation and planetary exploration. He is the author (or coauthor) of more than 420 scientific publications in referred international journals, more than 390 papers in conference proceedings, and 25 book chapters. He is editor/coeditor of 18 books/conference proceedings and two scientific books. His papers are highly cited, as proven from the total number of citations (more than 58000) and the value of the h-index (111) (source: Google Scholar). His current research interests include the areas of remote sensing, radar and SAR, signal processing, machine learning, and pattern recognition.

Dr. Bruzzone is the Principal Investigator of many research projects. Among the others, he is currently the Principal Investigator of the Radar for Icy Moon Exploration (RIME) and of the Subsurface Radar Sounder (SRS) instruments in the framework of the JUICE and the Envision missions, respectively, of the European Space Agency (ESA) and is the Science Lead for the High Resolution Land Cover project in the framework of the Climate Change Initiative of ESA. He was invited as a keynote speaker at more than 40 international conferences and workshops. He was a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) between 2009 and 2023, where since 2019, he had been the Vice-President for Professional Activities. He ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Seattle, July 1998. He was also the recipient of many international and national honors and awards, including the recent IEEE GRSS 2015 Outstanding Service Award, the 2017 and 2018 IEEE IGARSS Symposium Prize Paper Awards, the 2019 WHISPER Outstanding Paper Award, and the 2022 Letter Prize Paper Award for the best paper published in the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He was a Guest Coeditor of many Special Issues of international journals. He is the cofounder of the IEEE International Workshop on the Analysis of Multi-Temporal Remote-Sensing Images (MultiTemp) series and is currently a member of the Permanent Steering Committee of this series of workshops. Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. He has been the founder of the *IEEE Geoscience and Remote Sensing Magazine* for which he was an Editor-in-Chief between 2013 and 2017. He is currently an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society between 2012 and 2016.