AdaSemiCD: An Adaptive Semi-supervised Change Detection Method Based on Pseudo Label Evaluation

Lingyan Ran, Dongcheng Wen, Tao Zhuo, Shizhou Zhang, Xiuwei Zhang, Yanning Zhang, *Fellow IEEE*

***Abstract*—Change detection(CD) is an essential field in remote sensing, with a primary focus on identifying areas of change in bitemporal image pairs captured at varying intervals of the same region. The data annotation process for CD tasks is both time-consuming and labor-intensive. To better utilize the scarce labeled data and abundant unlabeled data, we introduce an adaptive semi-supervised learning method, AdaSemiCD, to improve pseudo-label usage and optimize the training process. Initially, due to the extreme class imbalance inherent in CD, the model is more inclined to focus on the background class, and it is easy to confuse the boundary of the target object. Considering these two points, we develop a measurable evaluation metric for pseudo-labels that enhances the representation of information entropy by class rebalancing and amplification of ambiguous areas, assigning greater weights to prospective change objects. Subsequently, to enhance the reliability of samplewise pseudo-labels, we introduce the AdaFusion module, to dynamically identifying the most uncertain region and substituting it with more trustworthy content. Lastly, to ensure better training stability, we introduce the AdaEMA module, which updates the teacher model using only batches of trusted samples. Experimental results on ten public CD datasets validate the efficacy and generalizability of our proposed adaptive training framework.**

***Index Terms*—Pseudo label, Semi-supervised Learning, Change Detection, Mean Teacher, Adaptive Learning**

I. Introduction

CHANGE detection (CD) has emerged as a significant research focus within the field of remote sensing. Its objective is to identify regions of interest that have experienced alterations in bi-temporal image pairs captured at varying times of the same geographical area. This method plays a crucial role in remote sensing data analysis and is particularly important in various civilian sectors, such as urban planning [1], [2], rural land management [3], [4], and disaster assessment [5], [6].

Given that the process of accurately annotating masks for change detection tasks is notably labor-intensive, direct application of traditional supervised learning approaches, such as convolutional neural networks (CNN) [7], [8] and Transformers [9], [10], to a limited set of labeled data often results in limited performance. In response to these challenges, researchers have explored a range of approaches such as weakly supervised change detection (WSCD) [11], [12], unsupervised change detection (USCD) [13], [14], and semi-supervised change detection (SSCD) [15], [16]. Although WSCD is cost-efficient, it relies on incomplete or inaccurate labels, which can introduce significant errors and unpredictable noise. USCD, on the other hand, does not require labeled data and leverages the intrinsic patterns present in the data; however, it often faces challenges when tackling specific tasks like classification or detection. Some methods would adopt sample generation strategies [17]–[20], which include data augmentation [19], generative adversarial networks (GAN) [18], and diffusion models [20], frequently necessitate the simulation or synthesis of additional data. However, when dealing with limited available samples, these methods may encounter constraints due to insufficient diversity in the generated data, which can diminish the model's ability to generalize. As a result, SSCD [21]–[23] emerges as a potentially more effective solution. The paradigm of semi-supervised learning (SSL) [24]–[26] aims to enhance CD performance by leveraging the limited available labeled data and the large volume of unlabeled samples. Typically, researchers generate pseudo-labels for the unlabeled data to act as guidance during training. These pseudo-labels are often temporary predictions with higher probabilities. The most widely used approach is the mean-teacher (MT) [27] framework, which employs a teacher model to generate pseudo-labels that serve as guidance for the student model during the training process. The teacher model is subsequently updated using the exponential moving average (EMA) [28] of the student model. The student model benefits from training on a mix of limited labeled data along with ample pseudo-labeled data, enabling it to identify more important features and resulting in marked enhancements in performance.

While these methods produce acceptable outcomes, significant problems persist: the model indiscriminately treats all samples, irrespective of their quality, and the training process lacks flexibility. Firstly, it is evident that unlabeled samples may not always function as efficient 'teachers'. Models frequently encounter difficulties in generating reliable high-quality pseudo-labels for intricate samples, which in turn

This work is supported in part by the National Natural Science Foundation of China (62476226), Natural Science Basic Research Program of Shaanxi (2024JC-YBQN-0719), Natural Science Foundation of NingBo (2023J262). *(Corresponding author: Tao Zhuo)*

Lingyan Ran, Dongcheng Wen, Shizhou Zhang, Xiuwei Zhang, and Yanning Zhang are with the School of Computer Science, Shaanxi Provincial Key Laboratory of Speech and Image Information Processing, and the National Engineering Laboratory for Integrated Aerospace-GroundOcean Big Data Application Technology, the Ningbo Institute of Northwestern Polytechnical University, Northwestern Polytechnical University, Xi'an 710072, China. Tao Zhuo is in the College of Information Engineering, Northwest A&F University, Yangling, 712100, China.

Manuscript received April 19, 2021; revised August 16, 2021.

introduces extra noise that can mislead the model's training. Subsequently, the EMA updating process does not take into account the quality of samples. Given that training batches may be biased or contain noise [29], dynamically determining the timing of training updates could contribute to the stability of the training process. These factors underscore the need for a more precise supervisory approach, failing which it could negatively impact the model's training.

In this study, we introduce an adaptive learning strategy, AdaSemiCD, designed to improve the accuracy of pseudo-labels and streamline the training process. Our framework builds upon the traditional semi-supervised training approach, augmented by two innovative functional modules, AdaFusion and AdaEMA. Initially, AdaFusion is employed to suppress noise at the individual sample level, thereby enhancing the accuracy of pseudo-labels. Contrary to previous methods like Augseg [30] or CutMix [31] that relied on entirely random fusion regions, our AdaFusion technique proactively identifies the most uncertain region and substitutes them with reliable content from either labeled datasets or unlabeled datasets of superior quality. Following this, we dynamically adjust the rate of parameter updates in the teacher-student model via AdaEMA to ensure improved stability. Although the traditional EMA effectively mitigates fluctuations in model parameters, thereby boosting stability, it persists in uniformly updating after each training iteration, neglecting the model's varying learning outcomes across different iterations when handling a range of training samples. If unlabeled samples contain excessive erroneous information, it can misdirect the model's training. Therefore, our AdaEMA introduces an adaptive selection process for model-level parameter updates, allowing the model to fully integrate superior parameters.

The main contributions of this paper are as follows:

- We propose an adaptive SSCD framework named AdaSemiCD, which dynamically improves the pseudo-labels as well as adjusts the training procedure with pseudo label quality assessment.
- We propose an AdaFusion strategy to enhance unreliable unlabeled samples. The fusion region and the trusted contents are selectively chosen with the uncertainty map.
- We propose an AdaEMA parameter update strategy, which updates the teacher model with a batch-wise pseudo-labels improving assessment.
- Experimental results on ten publicly available datasets demonstrate the effectiveness of our method.

II. Related Work

A. Semi-supervised Learning

Semi-supervised learning involves applying supervised learning on a limited amount of labeled data while employing unsupervised learning on a vast set of unlabeled data. SSL is typically divided into three strategies: consistent regularization (CR), self-training, and holistic methods, with the latter integrating the first two strategies within an SSL framework.

CR techniques are grounded in the concept of perturbed consistency, which utilizes the coherence between the model's output after varying degrees of perturbed input data as a training constraint. The three typical consistent regularization frameworks consist of the Π-model [32], the Temporal-ensembling model [32], and the mean-teacher model [28]. The Π-model's double-branch network shares the weight; the Temporal-ensembling model amalgamates all the outputs in the time series, with each image's pseudo-labels being the EMA of the previously generated results; the MT model carries out this smoothing operation at the model parameters level. This model has found application in subsequent semi-supervised research across various domains, such as Active-Teacher for semi-supervised object detection [33], [34]–[36] for semi-supervised general semantic segmentation, [37] for image classification, and [38], [39] for semi-supervised medical image segmentation. There has also been explored in the area of perturbation design, with [40] and [21] examining the image-level and feature-level perturbation of CR respectively.

In the realm of self-training methods, the authenticity of pseudo-labels is of paramount importance. This has led to extensive research into the effective selection of high-quality pseudo-labels for supervised learning. Feng et al. [41] proposed a method that incrementally adds labeled instances, reduces class bias via Synthetic minority over-sampling technique, and adaptively selects the optimal number of instances to enhance classifier performance. [42] employs a set probability threshold as a selection standard. ST++ [43] has developed a multi-tier self-training structure, where labels of high confidence are used for self-training repeatedly until all unlabeled samples have been utilized. [44] uses a constant entropy value as the filtering limit.

CR and self-training are commonly employed together rather than in isolation, creating a holistic strategy for semi-supervised learning. This is illustrated in [30], [34], [36], [37] and [43], as partially described in previous sections.

B. Semi-supervised Change Detection

Since annotating a large number of images for CD is time-consuming, recent methods mainly focus on the SSCD. In the realm of CR techniques, the incorporation of the mean-teacher model in CD was first introduced by Bousias et al. [45]. However, the initial outcomes did not show considerable potential, as this SSCD method fell short when compared to a benchmark that exclusively used a restricted quantity of labeled data for entirely supervised learning. Despite increasing labeled data, this disparity continues to expand. Using this as a basis, Mao et al. [46] implemented minor and major improvements to the inputs of the teacher and student models, respectively. Furthermore, they formulated an extra teacher-virtual adversararial training component to further reduce the harmful effects of the pseudo label noise.

Additionally, other semi-supervised methods employ either a single model or a two-branch model with shared weights. Such as Sun et al. [47] introduce a siamese network. They incorporated additional self-training based on pseudo-labels, employing threshold filtering to eliminate low-quality pseudo-labels. The rationale behind this filtering lies in the potential noise introduced by pseudo-labels with low confidence, which could adversely affect SSL training. Hafner et al. [48] propose

a dual-task SSCD framework that combines building segmentation and change detection, two closely related downstream tasks. They devised a novel consistency constraint between the two change detection masks produced by the siamese segmentation network and the CD network. Bandara et al. [21] explored feature-based perturbations of the regularization term, applying various data perturbations at the feature level to expand the distribution space of consistency constraints. This approach fully leverages the information embedded in unlabeled samples. In recent work, Zhang et al. [23] imposed two constraints of class consistency and feature consistency on unlabeled datasets. By aligning the feature representations of unlabeled samples on varying and invariant classes, the model could learn from a feature space that is closer to the real distribution, this contributed to the renewal of the best performance record at that time.

Some approaches primarily employ generative adversarial networks to produce data samples and understand feature distributions that approximate real labeled data [22], [49], [50]. Although these efforts have shown notable achievements in SSCD, the notoriously erratic nature of GAN training presents difficulties with hyperparameter tuning. Furthermore, the occurrence of gradient vanishing is a common challenge during training. Moreover, the discriminator's robust ability to differentiate can cause an imbalance between the generator and discriminator's performance within the GAN unless additional training strategies are applied. As a result, reaching an ideal balance is demanding, complicating the practical application of this method. Therefore, we continue the semi-supervised framework leveraging consistency and self-training techniques.

III. METHODOLOGY

Fig. 1 provides a comprehensive summary of our AdaSemiCD framework, aiming to enhance the SSCD performance by leveraging scarce labeled data and a vast quantity of unlabeled samples. We commence with a general introduction of the framework, followed by an detailed explanation of the uncertainty map used to assess our pseudo-labels, and finally present the specifics of AdaFusion and AdaEMA.

A. Overview of the AdaSemiCD Framework

The process of semi-supervised change detection is broadly outlined as follows: We are given a labeled dataset, denoted as $D_l = \left\{\left\{x^l_{a,i}, x^l_{b,i}\right\}, y^l_i\right\}_{i=1}^{m}$, and an unlabeled dataset $D_u = \left\{x^u_{a,j}, x^u_{b,j}\right\}_{j=1}^{n}$. Here, $\left\{\left\{x^l_{a,i}, x^l_{b,i}\right\}, y^l_i\right\}$ illustrates the i-th pair of labeled images alongside their corresponding true labels, while $\left\{x^u_{a,j}, x^u_{b,j}\right\}$ refers to the j-th pair of unlabeled images. The subscript a signifies the image from the 'Pre'-event period, and b signifies the 'Post'-event period. Importantly, the quantity of samples with labels and those without are m and n, with n considerably larger than m. The aim is for model M to not only derive key insights from D_l but also to enhance its feature extraction capability using the extensive collection of unlabeled samples in D_u, thus boosting the model's generalization potential. Ordinarily, samples are subject to either weak augmentation $A_w(\cdot)$ or strong augmentation $A_s(\cdot)$ prior to being passed into the network to guarantee superior generalization performance.

Model architecture: In this study, we employ the widely used MT framework for SSCD tasks. Two integral parts make up the network: a student model, denoted as M_{stu}, and a teacher model, represented as M_{tea}. Both components possess an identical architecture. The student model is trained to extract significant features from a small number of labeled samples and a large volume of unlabeled samples, with the aid of optimization via gradient descent methods. Conversely, the teacher model M_{tea} generates pseudo-labels to guide the student in assimilating unlabeled data, and it is updated using the EMA method.

Objectives: The objective is to minimize the supervised loss $\mathcal{L}_s$ on D_l while ensuring consistency on the disturbed D_u with a minimal $\mathcal{L}_u$. During training, samples are fed into the network in randomly shuffled batches, $\mathcal{B}_l$ and $\mathcal{B}_u$.

The loss $\mathcal{L}_s$ for labeled samples within a batch $\mathcal{B}_l$ is calculated as the cross entropy(CE) of ground truth y^l_i and its prediction p^l_i:

$$\mathcal{L}_s = \frac{1}{|\mathcal{B}_l|}\sum_{i=1}^{|\mathcal{B}_l|}\mathrm{CE}\left(p^l_i, y^l_i\right), \tag{1}$$

where $|\mathcal{B}_l|$ represents the mini-batch size, p^l_i represents the change detection result for the i-th pair of images.

The loss $\mathcal{L}_u$ for unlabeled samples within a batch B_u is quite similar. Here, we use the pseudo-labels from M_{tea} to supervise the predictions from M_{stu}. The loss $\mathcal{L}_u$ is calculated as follows:

$$\mathcal{L}_u = \frac{1}{|\mathcal{B}_u|}\sum_{j=1}^{|\mathcal{B}_u|}\mathrm{CE}\left(p^u_{s,j}, p^u_{w,j}\right), \tag{2}$$

where $|\mathcal{B}_u|$ represents the size of the unlabeled image mini-batch. $p^u_{w,j} = M^{\theta}_{tea}(w^u_{a,j}, w^u_{b,j})$ and $p^u_{s,j} = M^{\theta}_{stu}(s^u_{a,j}, s^u_{b,j})$ denote the change detection outcomes from the teacher model for the j-th pair of weak and strong augmentations of unlabeled images, respectively.

To summarize, the overall loss associated with the AdaSemiCD training process is defined as:

$$\mathcal{L} = \mathcal{L}_s + \lambda(\cdot)\mathcal{L}_u. \tag{3}$$

λ is the weight of unsupervised loss, which is typically set to a constant value [37], [38]. For SSCD, we contend that using a constant λ could potentially disrupt the training procedure. In the early stages of training, the pseudo-labels for unlabeled data tend to be highly unreliable, and excessive dependence on unsupervised training at this point can inject substantial noise. Conversely, as training progresses into the intermediate and final phases, the model enhances its ability to generate high-quality pseudo labels, diminishing the importance of the limited labeled dataset. This shift warrants a gradual decrease in the emphasis on supervised training compared to unsupervised training over time. To prevent overfitting and refine the feature space, it is crucial to have a systematic approach that dynamically modifies the balance between these two elements in the loss function. We employ a ramp-up

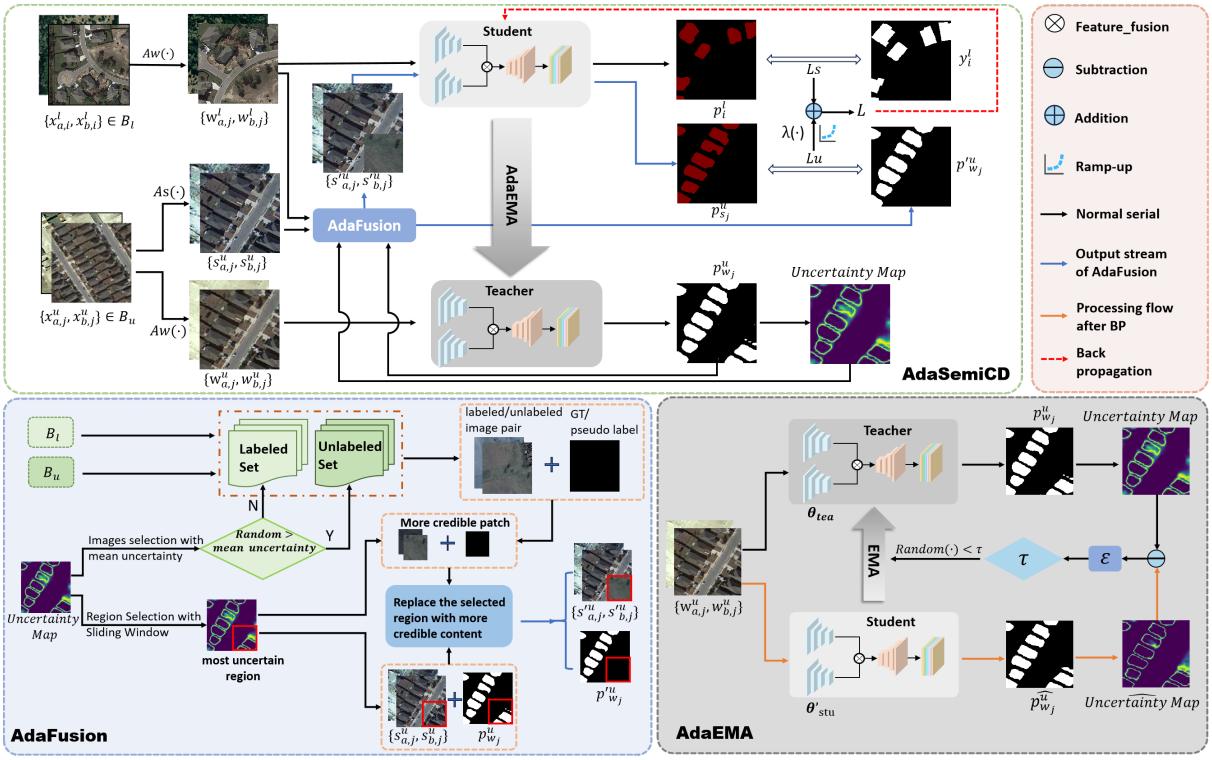

Fig. 1. The overview of the proposed AdaSemiCD framework and details of two adaptive modules. Our approach is based on the common MT training pipeline for SSCD, and we suggest utilizing the AdaFusion module to produce samples with increased reliability and the AdaEMA module to enhance the efficiency of EMA update.

approach, defining $\lambda(\cdot)$ as a function that adapts through the course of training, thereby dynamically modulating the balance between supervised and unsupervised training throughout various phases.

$$\lambda(\cdot) = w_{max} \times e^{-\phi \times (1 - iter_{cur}/iter_{max})^2} \tag{4}$$

$$iter_{max} = \gamma \times iter_total \tag{5}$$

Here, w_{max} represents the maximum weight value of the unsupervised loss, and ϕ controls the severity of the ramp-up. $iter_{cur}$ represents the current iteration cycles; $iter_{max}$ is the total number of ramp-up cycles, calculated by multiplying γ (where $0 < \gamma < 1$) by the total number of training iterations, as shown in 5. After the ramp-up process, the weight of the unsupervised loss stabilizes at w_{max} and no longer changes. In the early stages of training, this weight is relatively low, so unsupervised training plays a negligible role, but in the middle and later stages of training, this weight gradually increases and the unsupervised loss after weighting exceeds the supervised loss, making unsupervised training the dominant factor.

Training strategy: To reduce the total loss $\mathcal{L}$, the parameters of the student network, denoted as θ_{stu}, are refined via Stochastic Gradient Descent (SGD). Concurrently, the teacher network updates its parameters θ_{tea} using an exponential moving average calculated from the student network's parameters θ_{stu} over a time sequence:

$$\theta_{tea} = \beta \theta_{tea} + (1 - \beta)\theta_{stu} \tag{6}$$

The hyperparameter β acts as a momentum factor, where a higher β value leads to a broader moving average window. Generally, β is selected to be near 1.0; in this study, for instance, β is set at 0.996.

Proposed modules: The effectiveness of semi-supervised learning depends largely on the quality of pseudo-labels. Nevertheless, it is clear that the previously mentioned procedure does not take into account the varying impact of individual samples on training. This paper concentrates on two critical aspects that are directly related to the generation of pseudo-labels: the initial pair of unlabeled images, and the efficiency of the pseudo-label-generation network, aka the teacher model, in identifying changes. To offer more reliable supervisory signals to unlabeled information and reduce training uncertainty, we propose an adaptive training strategy to tackle these two key issues. Our initial proposal is to develop a metric that quantifies the uncertainty of pseudo-labels, serving as the basis for adaptive modifications. Following that, we recommend applying adaptive modifications at the image level to the unlabeled training samples and suitably integrating reliable contents. Additionally, we apply adaptive and selective EMA updates to the teacher network during the training phase to minimize variations, ensuring more consistent and higher-quality pseudo-labels. The details of these methods are discussed in the following sections.

B. Pseudo-label Qualification Metric

To enhance the effectiveness of pseudo-labels by accurately gauging their quality, a crucial step is the implementation

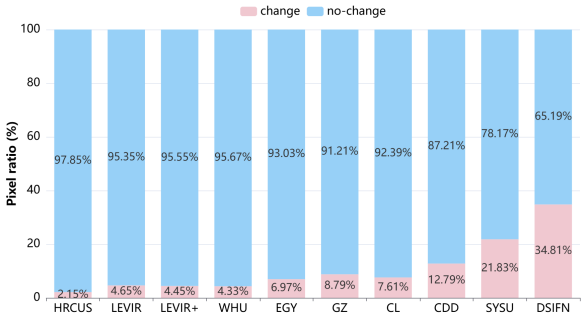

Fig. 2. Class statistics of common CD datasets. The proportion of the changed/unchanged categories is extremely unbalanced.

of an evaluation metric. This metric plays a key role in identifying reliable labels and determining the value of each training sample. Unlike labeled image pairs, where true labels serve as a benchmark for assessing the model's performance, pseudo-labels lack such reference points and can only be compared to themselves. To evaluate the quality of pseudo-labels, a common technique is the computation of information entropy [51],

$$E(x_i) = -P\left(x_i\right) \log_2 P\left(x_i\right), \tag{7}$$

where $P\left(x_i\right)$ is the output probability of a trained model on sample x_i. We propose that reduced information entropy in the predicted values implies greater reliability of the prediction. In contrast, increased information entropy points to a prediction with more uncertainty, indicating a more balanced probability distribution across the pixels and resulting in decreased confidence in the model's predictions.

In the CD task, directly applying entropy may not yield the optimal results due to the significant challenge posed by class imbalance, as illustrated in Fig. 2. The ratios of changed to unchanged categories are highly skewed. This imbalance may lead our model to learn the target categories inadequately during training while disproportionately capturing feature distributions from the background categories. As a result, the model tends to classify pixels as background during inference. To reduce the influence of this class imbalance when assessing the quality of pseudo-labels, we assign different weights to the two categories when computing information entropy,

$$E'\left(X\right) = w_1 \times E\left(x_i\right)[0] + w_0 \times E\left(x_i\right)[1], \tag{8}$$

where w_0 and w_1 represent the proportions of pixels in the current batch that belong to the unchanged and changed categories, respectively. The formulas for their calculation are:

$$w_0 = \frac{\sum_{i=1}^{|Bu|} \sum_{k=1}^{H \times W} \left(P(x_i) == 0\right)}{|Bu| \times H \times W} \tag{9}$$

$$w_1 = \frac{\sum_{i=1}^{|Bu|} \sum_{k=1}^{H \times W} \left(P(x_i) == 1\right)}{|Bu| \times H \times W} \tag{10}$$

Furthermore, pixels from different regions play distinct roles. Pixels that are more crucial, like those found on edges, targets, or areas resembling the background—frequently predicted with higher uncertainty—should be prioritized. To achieve this, we start by enhancing the significance of these key regions by calculating the absolute difference between the prediction probabilities of the two classes,

$$D\left(x_i\right) = abs(P\left(x_i\right)[1] - P\left(x_i\right)[0]) \tag{11}$$

In this context, abs represents the absolute value function, employed to prevent inconsistencies that could occur due to differing changes before and after a phase. By carrying out a pixel-wise multiplication with the information entropy, we derive the uncertainty map for the image x_i,

$$U\left(x_i\right) = 1 - D\left(x_i\right) \cdot E'\left(x_i\right) \tag{12}$$

In areas characterized by low information entropy, this procedure is unlikely to cause significant changes. However, in regions where information entropy is high, the operation enhances the effect.

To summarize, aiming to assess the quality of pseudo-labels in change detection, we propose a quantifiable metric U to evaluate uncertainty. This metric improves overall information entropy by considering elements such as class imbalance and regions of confusion.

C. AdaFusion: Adaptive Sample Fusion

Image fusion is widely used to augment samples and enhance generalization, with CutMix [31] and MixUp [52] being typical examples. In this study, our aim is to apply image fusion techniques to exclude unreliable areas from the training samples. This process consists of two steps: region selection, which determines the location for the operation, and image selection, which specifies the content to be used.

Adaptive selection of fusion region. Contrary to the conventional CutMix technique, which arbitrarily selects blending regions, our approach is more refined. First, we initialize a bounding box of random size, then slide the window to identify the area with the highest total uncertainty using Eq. (12), as the region to be fused. These regions typically include boundaries or complex areas where target identification is difficult. The selection approach guarantees a varied sample and, at the same time, decreases unsupervised noise.

Adaptive selection of fusion contents. For the region of maximum uncertainty, we can select a substitute from either the corresponding samples in the labeled set B_l or the unlabeled set B_u with higher reliability. This strategy helps prevent overreliance on labeled samples and further mitigates the risk of overfitting. The selection of fusion content is guided by an adaptive threshold that determines whether to fuse with a labeled image pair. We directly use the computed uncertainty as the threshold, and if the randomly generated probability exceeds the total uncertainty of the sample, labeled images in the training batch are selected as the fusion content. Otherwise, other unlabeled image pairs of higher quality are chosen. It is clear that the higher the reliability of a sample, the better the quality of the pseudo-label. If a sample is considered reliable enough, we avoid fusing it with limited labeled images, which could increase the risk of overfitting. Samples with higher uncertainty contain more noise, and incorporating new content is expected to reduce the noise effectively.

D. AdaEMA: Adaptive EMA Update Strategy

Within the MT framework, the teacher model's update process involves incorporating the student's exponential moving average over time. Our goal is to reach a state of optimal coevolution, where the teacher acts as an aggregate reflection of the student model's progress. A pivotal factor in this coevolution is whether the student model advances or declines with each update of the teacher. Evaluating the student model's state is closely tied to the training process's validation phase. Typically, after several training epochs, we evaluate the model by using samples from the validation set, performing inference, and comparing the output against true labels to determine accuracy. Conducting this validation after every iteration incurs a significant computational cost and extends training duration. One may address this by reducing the validation set size to just a few pairs; however, this risks insufficiently assessing model performance if the sample is too small. The challenge lies in striking a balance between these considerations.

During each training phase, we commence by updating the student model, denoted as M_{stu}^{θ}, according to the strategy outlined in section III-A, which results in the updated model $M_{stu}^{\theta'}$. Subsequently, we assess both the modified student model, $M_{stu}^{\theta'}$, and the teacher model, M_{tea}^{θ} on the current set of unlabeled training samples, $\mathcal{B}_u$. Using Eq. (12), we calculate the corresponding uncertainty maps, referred to as U_{tea} and U_{stu}. The changes in the student model's development are captured through fluctuations in uncertainty,

$$\varepsilon = \frac{\sum U_{stu} - \sum U_{tea}}{|B_u|}. \tag{13}$$

Additionally, we introduce a probability τ to regulate model updates according to the reliability of the student model, defined as

$$\tau = \begin{cases} \frac{1}{iter^2+\epsilon} & , \quad \varepsilon \leq 0 \\ 1.0 & , \quad \varepsilon > 0 \end{cases}, \tag{14}$$

Here, $iter$ denotes the current iteration count and $\epsilon = 1e-5$. When ε is greater than zero, the student model progresses, allowing a straightforward update of M_{tea}^{θ}. In contrast, if ε is zero or less, the student model encounters either regression or fluctuation. To gauge the probability of altering M_{tea}^{θ}, we incorporate some randomness. The adaptive updating details are provided in Algorithm 1.

Algorithm 1 The AdaEMA algorithm.

Input:
Student model M_{stu}^{θ}, Teacher model M_{tea}^{θ}
The set of training samples for the current batch, $\mathcal{B} = \{\mathcal{B}_l, \mathcal{B}_u\}$
Output:
Updated Teacher model, $M_{tea}^{\theta'}$;

Calculate the supervised loss $\mathcal{L}_s$ on labeled samples $\mathcal{B}_l$ using Eq. (1);
Calculate the unsupervised loss $\mathcal{L}_u$ on unlabeled samples $\mathcal{B}_u$ using Eq. (2);
Update the student model M_{stu}^{θ} to $M_{stu}^{\theta'}$ using SGD to minimize the total loss, as described in Eq. (3).
Calculate the uncertainty U_{tea} of the pseudo-labels generated on $\mathcal{B}_u$ by the teacher model M_{tea}^{θ} as described in Eq. (12);
Calculate the uncertainty U_{stu} of the pseudo-labels generated on $\mathcal{B}_u$ by updated student model $M_{stu}^{\theta'}$ as described in Eq. (12);
Calculate the upper bound of the update probability τ according to Eq. (13) and Eq. (14);
if $random(\cdot) < \tau$ **then**
    Update the teacher model to $M_{tea}^{\theta'}$ by EMA;
else
    $M_{tea}^{\theta'} = M_{tea}^{\theta}$;
end if
return $M_{tea}^{\theta'}$.

IV. Experiment

A. Experimental Setup

1) Datasets: Our method is empirically tested on ten benchmark datasets, namely the LEVIR-CD [53], LEVIR-CD+ [53], WHU-CD [54], EGY-CD [55], HRCUS-CD [56], CDD [57], GZ-CD [22], DSIFN-CD [58], SYSU-CD [59], and CL-CD [60]. As summarized in Table I, these datasets cover different resolutions (0.03m-2.0m), different data sizes (1-20000 pairs), different annotation categories (binary or multiclass), and different time spans (2-16 years). Fig. 3 shows some classic sample images for each dataset.

We employ an identical configuration across all datasets, using 5%, 10%, 20%, and 40% as the proportions of labeled samples. For LEVIR-CD and WHU-CD, we adopted the semi-supervised partitioning as described in [21]–[23]. For the remaining eight datasets, we utilized random partitioning.

2) Evaluation Metrics: For easy comparison with the most advanced techniques, we utilized overall accuracy (OA) to assess general performance. Due to the significant imbalance in CD categories and our main focus on the altered area, we applied the intersection over union for the changed category IoU^c. The calculation formulas are provided below:

$$IoU^c = TP/(TP + FP + FN) \tag{15}$$

$$OA = (TP + TN)/(TP + FP + FN + TN) \tag{16}$$

Where TP represents the positive sample correctly predicted (the correct changing pixel), TN refers to the negative sample correctly predicted (the correct unchanged pixel), and FP denotes the positive sample wrongly predicted (the unchanged pixel wrongly detected), FN represents the negative sample wrongly predicted (the pixel missed as the unchanged pixel). For both metrics, the larger the value, the better the change detection performance of the model.

3) Implementation Details: We employ ResNet50+PPM as our change detection framework, as referenced in [21] and [23]. The learning rate starts at 0.01 and decreases linearly to 1e-4, with momentum kept at 0.9. Training of all competing approaches is conducted using the SGD optimizer over 80 epochs. For both labeled and unlabeled data, the mini-batch size is set at 8. Moreover, the augmentations applied are the

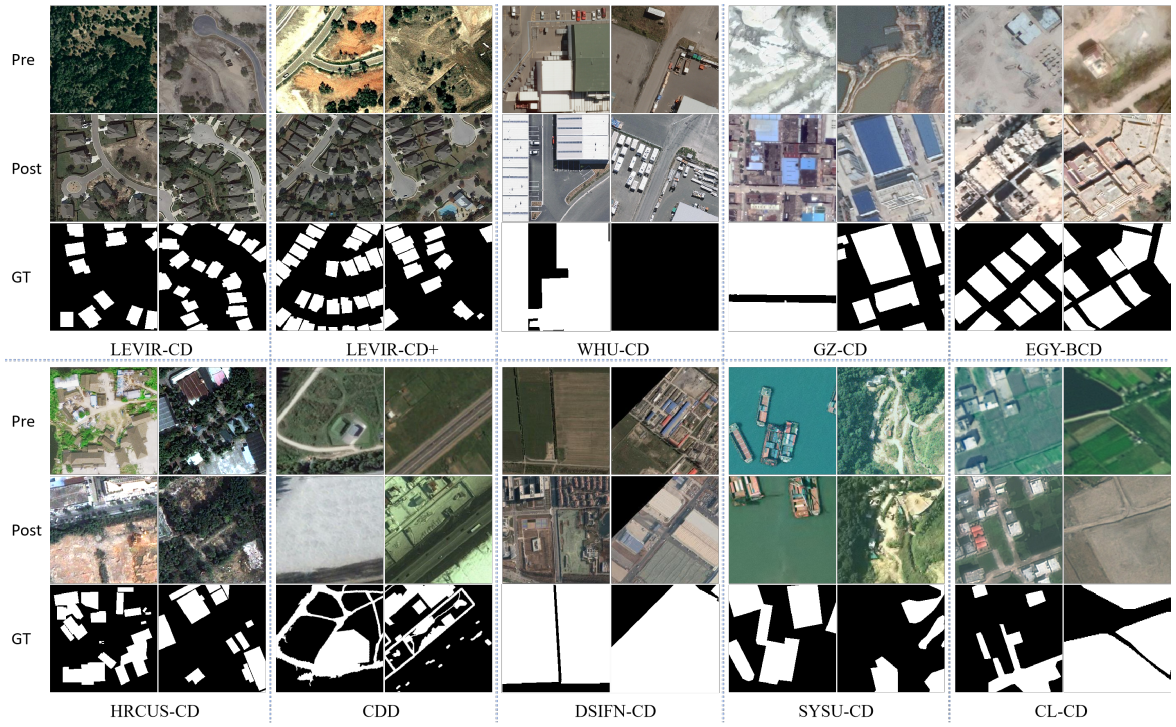

Fig. 3. Typical examples of ten change detection datasets we use, with Pre-event image, Post-event image, and Ground Truth.

TABLE I
TEN PUBLICLY AVAILABLE CHANGE DETECTION DATASETS USED IN THE EXPERIMENT.

Category	Datesets	Spatial Resolution	Size	Annotated Samples	Time Spans	Download
Binary	LEVIR-CD [53]	0.5m	1024 × 1024	637	5 to 14 years	Link
	LEVIR-CD+ [53]	0.5m	1024 × 1024	985	5 to 14 years	Link
	WHU-CD [54]	0.2m	15354×32507	1	2012 to 2016	Link
	GZ-CD [22]	0.55m	Varying	19	2006 to 2019	Link
	EGY-BCD [55]	0.25m	256 × 256	6091	2015 to 2022	Link
	HRCUS-CD [56]	0.5m	256 × 256	11388	Varing	Link
Multiclass	CDD [57]	0.03m-1.0m	256 × 256	16000	Varing	Link
	DSIFN-CD [58]	Unknown	512 × 512	3940	Unknown	Link
	SYSU-CD [59]	0.5m	256 × 256	20000	2007 to 2014	Link
	CL-CD [60]	0.5-2.0m	512 × 512	600	2017 to 2019	Link

same as those in FPA [23], which consist of weak augmentations such as random flip, random resizing from 0.5 to 2.0, and random cropping, along with nine robust augmentations [61]. A pseudo label threshold of 0.95 is adopted for all models in our studies. When calculating loss, ϕ is fixed at 5. The entire set of experiments was executed using PyTorch on four NVIDIA GeForce RTX 3090 GPUs.

B. Compare with the SOTA Methods

To demonstrate the advantages of our proposed method, we have conducted a comparison with a number of leading semi-supervised change detection techniques [21]–[23], [35], [62].

1) Quantitative Results: Tables II and III display our experimental outcomes for both the Building CD datasets and the multi-class CD datasets. The 'Sup. only' results denote supervised training outcomes from a restricted portion of the labeled dataset, whereas 'Oracle' represents full supervision results using the entire training dataset. It's notable that our approach achieves SOTA performance in nearly every partition configuration across these datasets. Remarkably, in most scenarios, all semi-supervised CD approaches outperformed the supervised ones within the same partition setting, confirming the efficacy of semi-supervised methods in leveraging numerous unlabeled training samples. Additionally, the superiority of our method highlights the practical importance of our adaptive strategy for learning more effectively from unlabeled data.

Building CD Datasets: As presented in Table II, the proposed AdaSemiCD framework demonstrates outstanding performance across nearly all building CD datasets. Notably, on LEVIR-CD, LEVIR-CD+, and WHU-CD datasets, AdaSemiCD achieves a significant improvement in IoU^c, with gains of 3.1, 1.3, and 1.7 percentage points, respectively. On EGY-CD and HRCUS-CD datasets, AdaSemiCD maintains an overall leading position with average improvements of 0.75 and 0.4 points, respectively. However, performance on certain configurations is marginally lower than SOTA methods. This

limitation can be attributed to the unique challenges posed by these datasets: the EGY-CD dataset suffers from overexposure, which often blends buildings with the background, increasing the difficulty of accurate discrimination. Similarly, the HRCUS-CD dataset is affected by vegetation occluding the changing areas of some buildings. These challenges hinder the model's ability to generate reliable pseudo-labels when trained with a limited number of real labels.

It is evident that AdaSemiCD encounters notable challenges on the GZ-CD dataset, underperforming compared to the FPA method across most experimental configurations, except for maintaining a lead in the 5% labeled datasets. We attribute this to both the dataset's intrinsic characteristics and the inherent limitations of our approach. Firstly, the GZ-CD dataset has the lowest spatial resolution among the datasets evaluated. This limitation, coupled with the prevalence of small buildings, results in objects that are difficult to discern visually at such resolutions, even for human observers. Secondly, the manual annotations of this dataset are notably coarse, likely due to these resolution constraints. For regions with multiple targets, annotators often delineated broad areas that encompass significant portions of the background, as illustrated in Fig. 3. This introduces substantial noise in the labeled data, especially for building change detection tasks. Lastly, the dataset's suburban setting introduces additional challenges. Urban development and expansion over the dataset's temporal span of over five years result in dramatic differences between the two images, beyond the annotated building change areas. These variations in the background—often so significant that it becomes difficult to ascertain whether the images correspond to the same location—introduce strong non-interesting changes that interfere with the model's ability to focus on the relevant change areas. However, unlike AdaSemiCD, FPA and RCR do not rely solely on one-hot hard label. Instead, they leverage consistency constraints at the feature level, which effectively mitigates this issue to a significant extent.

Furthermore, as indicated by the experimental results, the overall accuracy of all methods on these binary change detection datasets is notably high, primarily due to the dominance of the background class. OA often correlates closely with IoU^c, where even a slight improvement in OA translates into a relatively significant improvement in IoU^c. Our AdaSemiCD consistently enhances OA across nearly all datasets, as it effectively reduces noise and minimizes the influence of false signals, leading to more accurate predictions.

Multiclass CD Datasets: As shown in Table III, our proposed AdaSemiCD demonstrates optimal performance in the majority of scenarios, despite the increased complexity of multi-category change detection tasks compared to binary building change detection tasks. Notably, optimal results were achieved across all experimental settings on the CDD dataset, with IoU^c and OA increasing by an average of 1.8 percentage points and 0.25 percentage points respectively, which can be attributed to its exceptionally accurate annotations, as illustrated in Fig. 3. This high-quality labeling is particularly advantageous for pseudo-label-based semi-supervised methods. Furthermore, AdaSemiCD exhibits overall performance improvements on the SYSU-CD and CL-CD datasets. However, its performance is slightly below the SOTA results in certain experimental configurations. This can be attributed to the inherent challenges in identifying some mountain and land changes present in these datasets. These changes are often large-scale, and misclassification in such regions may result in fluctuations in performance metrics.

TABLE II
THE AVERAGE QUANTITATIVE METRICS OF DIFFERENT CD METHODS ON BUILDING CHANGE DETECTION DATASETS. THE HIGHLIGHTED PARTS IN BLUE ARE THE BEST RESULTS, AND THE UNDERLINED ONES ARE THE SECOND BEST RESULTS.

Dataset	Method	5%		10%		20%		40%	
		IoU^c	OA	IoU^c	OA	IoU^c	OA	IoU^c	OA
LEVIR-CD	Sup. only	61.0	97.60	66.8	98.13	72.3	98.44	74.9	98.60
	AdvEnt [62]	66.1	98.08	72.3	98.45	74.6	98.58	75.0	98.60
	s4GAN [35]	64.0	97.89	67.0	98.11	73.4	98.51	75.4	98.62
	SemiCDNet [22]	67.6	98.17	71.5	98.42	74.3	98.58	75.5	98.63
	RCR [21]	72.5	98.47	75.5	98.63	76.2	98.68	77.2	98.72
	FPA [23]	73.7	98.57	76.6	98.72	77.4	98.75	77.0	98.74
	AdaSemiCD	**77.7**	**98.78**	**79.4**	**98.87**	**80.3**	**98.92**	**80.6**	**98.93**
	Oracle	IoU^c=**77.9** and OA=**98.77**							
LEVIR-CD+	Sup. only	52.0	97.72	58.4	98.06	66.1	98.31	66.2	98.42
	AdvEnt [62]	52.2	97.68	59.9	98.11	65.9	98.37	68.0	98.51
	s4GAN [35]	46.5	97.25	51.4	97.66	62.8	98.18	67.2	98.46
	SemiCDNet [22]	52.6	97.66	60.7	98.24	64.8	98.37	66.1	98.38
	RCR [21]	64.9	98.25	67.5	98.45	68.5	98.52	68.4	98.51
	FPA [23]	64.6	98.30	67.3	98.40	70.3	**98.64**	69.0	98.59
	AdaSemiCD	**66.7**	**98.49**	**68.8**	**98.51**	**70.6**	98.63	**70.9**	**98.64**
	Oracle	IoU^c=**70.5** and OA=**98.63**							
WHU-CD	Sup. only	50.0	97.48	55.7	97.53	65.4	98.20	76.1	98.94
	AdvEnt [62]	55.1	97.90	61.6	98.11	73.8	98.80	76.6	98.94
	s4GAN [35]	18.3	96.69	62.6	98.15	70.8	98.60	76.4	98.96
	SemiCDNet [22]	51.7	97.71	62.0	98.16	66.7	98.28	75.9	98.93
	RCR [21]	65.8	98.37	68.1	98.47	**74.8**	98.84	77.2	98.96
	FPA [23]	66.3	98.45	57.4	97.69	62.5	98.48	73.1	98.69
	AdaSemiCD	**67.8**	**98.62**	**70.8**	**98.70**	74.7	**98.86**	**79.6**	**99.13**
	Oracle	IoU^c=**85.5** and OA=**99.38**							
GZ-CD	Sup. only	47.5	93.56	51.4	94.26	58.0	95.65	66.3	96.62
	AdvEnt [62]	48.6	94.39	50.9	94.89	60.2	95.79	66.2	96.58
	s4GAN [35]	50.8	94.38	52.4	94.98	60.8	95.94	64.2	96.39
	SemiCDNet [22]	48.4	93.58	49.7	94.79	59.0	95.66	66.3	96.57
	RCR [21]	50.8	93.82	50.8	94.69	62.5	96.07	67.8	96.61
	FPA [23]	51.2	93.92	**58.9**	**95.78**	**63.1**	**96.26**	**68.2**	**96.82**
	AdaSemiCD	**51.6**	**94.56**	57.1	95.57	62.4	96.21	68.0	96.75
	Oracle	IoU^c=**69.0** and OA=**96.93**							
EGY-CD	Sup. only	49.8	95.73	54.6	96.38	61.4	96.83	65.1	97.25
	AdvEnt [62]	52.7	96.01	57.8	96.58	62.6	96.86	64.0	97.19
	s4GAN [35]	52.9	95.94	58.6	96.50	64.7	97.09	64.9	97.27
	SemiCDNet [22]	52.4	96.00	57.9	96.31	62.8	96.95	63.8	97.19
	RCR [21]	58.1	96.50	59.9	96.77	63.9	97.08	64.2	97.18
	FPA [23]	57.5	96.52	60.1	**96.86**	**65.2**	**97.25**	65.7	97.34
	AdaSemiCD	**59.0**	**96.55**	**60.5**	96.80	65.0	97.20	**67.4**	**97.39**
	Oracle	IoU^c=**67.6** and OA=**97.54**							
HRCUS-CD	Sup. only	29.5	98.11	36.0	98.45	43.4	98.68	48.9	98.84
	AdvEnt [62]	29.1	98.11	36.9	98.40	42.5	98.61	48.8	98.71
	s4GAN [35]	25.0	97.86	28.2	98.24	40.1	98.62	50.3	98.85
	SemiCDNet [22]	28.4	98.00	34.7	98.44	44.1	98.68	48.5	98.74
	RCR [21]	36.1	98.36	42.1	98.69	45.3	98.76	49.6	98.66
	FPA [23]	35.2	98.37	**43.7**	98.65	46.7	98.82	**51.2**	**98.81**
	AdaSemiCD	**37.8**	**98.59**	42.6	**98.70**	**48.1**	**98.84**	50.8	98.87
	Oracle	IoU^c=**59.0** and OA=**99.06**							

An additional significant observation is that, when the overall accuracy declines, the correlation between OA and IoU^c becomes less straightforward. In particular, there are instances where IoU^c reaches its maximum value, while OA does not, underscoring their different focuses. OA is mainly concerned with overall accuracy, highlighting the detection of background categories, whereas IoU^c prioritizes the precise detection of target classes, which is more vital for change detection tasks than identifying background categories. This difference is apparent in the DSIFN-CD dataset, where our method demonstrates strong performance on the IoU^c metric.

Among the evaluated datasets, our method demonstrates the poorest performance on DSIFN-CD. While it performs well on the critical IoU^c metric, its performance on OA is suboptimal. This can be attributed to similarities between DSIFN-CD and GZ-CD, as both datasets suffer from coarse manual annotations and low spatial resolution, as evident in Fig. 3.

TABLE III
THE AVERAGE QUANTITATIVE METRICS OF DIFFERENT CD METHODS ON MULTICLASS CHANGE DETECTION DATASETS. THE HIGHLIGHTED PARTS IN BLUE ARE THE BEST RESULTS, AND THE UNDERLINED ONES ARE THE SECOND BEST RESULTS.

Dataset	Method	5%		10%		20%		40%	
		IoU^c	OA	IoU^c	OA	IoU^c	OA	IoU^c	OA
CDD-CD	Sup. only	60.4	94.25	67.9	95.46	75.6	96.59	82.3	97.56
	AdvEnt [62]	63.3	94.65	71.2	96.01	79.3	97.14	82.9	97.66
	s4GAN [35]	62.3	94.69	71.0	95.94	79.0	97.10	82.8	97.63
	SemiCDNet [22]	63.5	94.68	71.2	95.99	79.1	97.13	82.8	97.63
	RCR [21]	67.6	95.40	75.5	96.57	80.2	97.26	82.7	97.61
	FPA [23]	68.9	95.66	74.9	96.55	79.7	97.20	81.1	97.37
	AdaSemiCD	**70.1**	**95.89**	**77.3**	**96.89**	**82.1**	**97.56**	**83.9**	**97.80**
	Oracle	IoU^c=**87.8** and OA=**98.10**							
DSIFN-CD	Sup. only	34.8	78.34	38.9	83.41	40.2	87.00	39.6	87.00
	AdvEnt [62]	31.8	77.83	36.3	83.86	40.8	85.92	37.4	86.31
	s4GAN [35]	36.6	84.10	34.8	86.87	37.9	**87.69**	40.1	86.52
	SemiCDNet [22]	33.6	78.60	37.9	84.18	39.1	86.77	39.1	87.05
	RCR [21]	26.7	83.78	32.9	86.05	40.8	86.70	36.7	86.08
	FPA [23]	**39.2**	**84.27**	38.5	**87.12**	36.0	87.41	35.8	86.50
	AdaSemiCD	36.9	80.46	**39.2**	82.94	**41.1**	85.45	**45.1**	**87.12**
	Oracle	IoU^c=**58.1** and OA=**90.82**							
SYSU-CD	Sup. only	62.9	89.57	64.4	90.18	66.0	90.82	66.4	90.93
	AdvEnt [62]	61.2	89.36	64.5	90.18	65.7	90.35	68.3	91.24
	s4GAN [35]	64.4	90.02	66.5	90.48	66.9	90.26	68.2	91.51
	SemiCDNet [22]	61.7	89.32	64.8	90.25	66.7	90.97	67.0	91.08
	RCR [21]	62.5	89.76	66.0	90.75	64.1	90.22	65.3	90.56
	FPA [23]	**67.7**	90.95	68.3	91.09	70.1	92.01	69.3	**91.97**
	AdaSemiCD	67.5	**91.16**	**68.7**	**91.59**	**70.1**	**92.03**	**69.9**	91.90
	Oracle	IoU^c=**68.2** and OA=**91.64**							
CL-CD	Sup. only	18.1	91.90	31.4	92.42	37.2	93.32	45.9	94.98
	AdvEnt [62]	24.3	92.13	33.2	93.01	37.6	93.59	42.9	94.06
	s4GAN [35]	22.1	92.00	26.6	93.09	37.4	93.59	43.4	93.87
	SemiCDNet [22]	24.0	92.20	28.3	**93.42**	36.2	92.41	45.3	94.22
	RCR [21]	27.1	91.63	32.8	92.99	36.4	93.07	48.5	94.94
	FPA [23]	29.0	91.00	**38.2**	93.37	39.6	93.88	43.1	94.15
	AdaSemiCD	**30.6**	**92.52**	33.5	92.40	**41.6**	**94.21**	**49.1**	**95.85**
	Oracle	IoU^c=**50.1** and OA=**95.66**							

Moreover, DSIFN-CD exhibits only a slight degree of class imbalance, with the proportion of change classes reaching as high as 34.81%. Consequently, the category rebalancing strategy we designed has limited effectiveness, leading to the misclassification of many background pixels. This, in turn, contributes to the observed reduction in OA.

2) Qualitative Results: Fig. 4 and Fig. 5 showcase some examples of the visualizations on the test sets of building and mutilclass CD datasets respectively, in which the area selected in the box is the error-prone area. It is apparent that on those datasets, our approach has notably mitigated the common issues of missed and false detections. In challenging scenarios, our method could still effectively identify the areas of change that were of interest to us.

Some failure cases of the detected changes are unrelated to the buildings of interest. The absence of adequate supervision information makes it challenging to mitigate such interference, leading to decreased model performance. Additionally, the task is further complicated by the detection of small and densely changing areas, which proves to be difficult for the model.

In complex scenarios, alternative semi-supervised models may encounter guidance issues, leading them to inadvertently amplify errors during training. As a result, semi-supervised methods may perform worse than supervised methods trained exclusively on labeled data in such situations. Our model, however, adaptively mitigates these noises during the training phase by dynamically excluding them and integrating parameters from a progressively refined model throughout the training process. In these intricate cases, the quality of pseudo-labels is incrementally improved until they reach a reliable standard, thereby providing an accurate signal for model training. Consequently, our model demonstrates superior performance in these challenging regions, as highlighted by the boxes in Fig. 4 and Fig. 5. This adaptive approach forms the cornerstone of our proposed method.

C. Ablation Study

Effectiveness of proposed modules: Due to the differences in model architecture between the current semi-supervised change detection methods referred to and ours, we did not rush to verify the superiority of our method at first. Instead, we conducted model architecture experiments first, using the classical Mean-Teacher architecture. In addition to setting hyperparameters for it to control the weight of unsupervised losses, the rest of the data augmentations and CD network remained the same as [23] and [21]. The gain of this semi-supervised framework compared with the single model and the two-branch network with shared weights is very obvious, and it can almost approach the previous optimal performance. This also demonstrates the validity of the principle of perturbed consistency and the parameter integration, on the basis of which we explore the adaptive training mechanism. Therefore, we separately integrated our proposed AdaEMA and AdaFusion into the MT framework and achieved average improvement of 1.2 and 5.1 on IoU^c respectively, as shown in Table IV. * in the table indicates that in adapt, an adaptive judgment operation is performed to determine whether the fusion is performed, and a random selection strategy is used when selecting the fusion region, and a huge gap between the two is evident. Finally, the two adaptive modules contain the complete method, and better results are obtained on the basis of individual modules, which shows that the two modules proposed by us are decoupled, and the model architecture is reasonable.

Pseudo-label qualification metric: As described in Section III-B, the pseudo-label evaluation metric we proposed is based on information entropy, enhanced by class rebalancing and confusion region amplification. To demonstrate the effectiveness, we first evaluated our AdaSemiCD using information entropy as the sole metric, then incrementally incorporated class rebalancing and confusion region amplification, and finally evaluated the full model, which integrates all components.

The experimental results are presented in Table V. When only information entropy is used as the evaluation metric for implementing the adaptive training mechanism, significant performance improvements are achieved, attributed to the AdaFusion and AdaEMA modules we designed. Incorporating class rebalancing and confusion region amplification based on information entropy further enhances performance, demonstrating that these improvements enable more accurate identification of pseudo-labels during training, thereby facilitating more precise adaptive operations. Notably, the gain from class rebalancing exceeds that from confusion region amplification. This is because, with class rebalancing, the evaluation of pseudo labels places greater emphasis on the foreground, somewhat neglecting background identification, thus allowing AdaFusion to focus more on the uncertainty region in the foreground. In addition, confusion region amplification proves more beneficial as more labeled data becomes

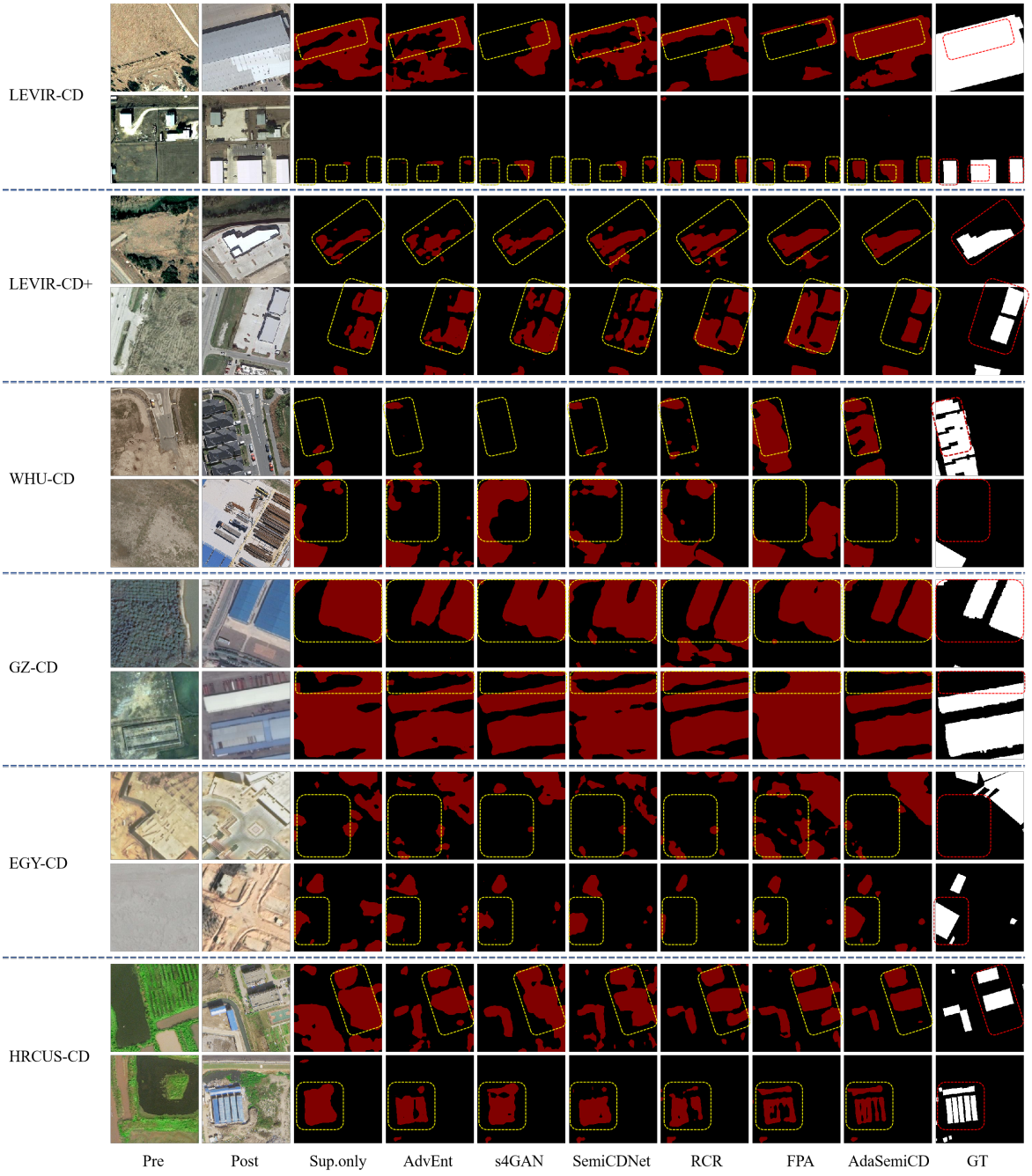

Fig. 4. Visualizations of different models on six building change detection datasets at the 5% labeled training ratio.

available. Finally, using uncertainty, an overall assessment that includes both components, leads to further performance gains, suggesting that these two improvements are independent and can be effectively combined for more accurate assessments.

We have stored the pseudo-labels generated during the training process, and Fig. 6 illustrates the IoU^c between the pseudo-labels and the corresponding Ground Truth throughout training, both with and without class rebalancing in the pseudo-label evaluation (note: the Ground Truth used here for unlabeled samples are solely for calculating the IoU^c with the pseudo-labels and are not involved in any other aspect of the training process). It is evident that, although class rebalancing does not directly improve the pseudo-labels, it enables a more accurate evaluation of the pseudo-labels. As a result, our adaptive training mechanism enhances the quality of them, bringing them closer to the real labels, which leads to a higher IoU^c between them.

D. Complexity Analysis

Since we utilize the same architecture with FPA and RCR, the number of training parameters (46.85M) and computational amount (585.85 GFLOPs) were the same. The variations in parameters, FLOPs, and the time required for training and inference are presented in Table VII. Our AdaSemiCD is

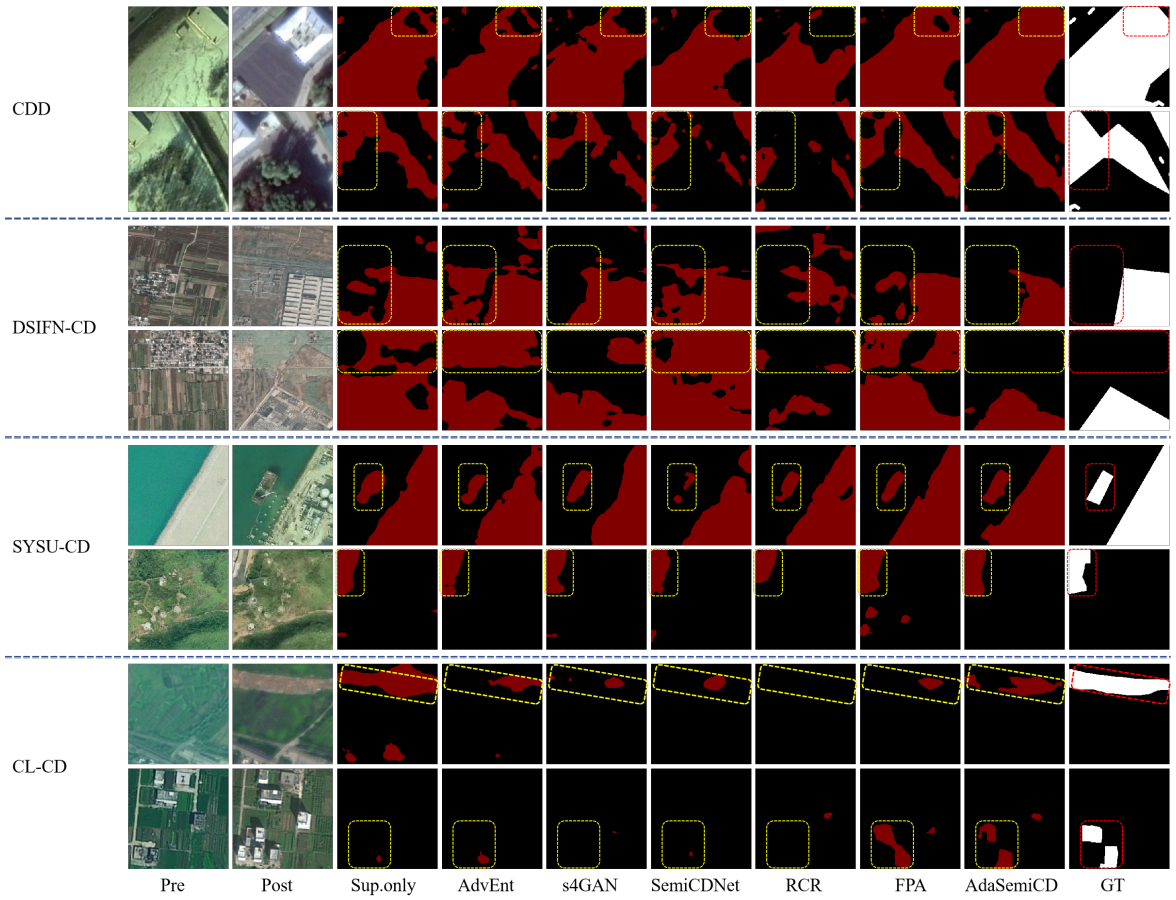

Fig. 5. Visualizations of different models on four mutil-class change detection datasets at the 5% labeled training ratio.

TABLE IV
ABLATION STUDY OF OUR PROPOSED ADASEMICD ON LEVIR-CD DATASET. AEMA AND AF DENOTES OUR ADAEMA, ADAFUSION MODULE RESPECTIVELY. AND AF* REPRESENTS THE FUSION REGION ARE RANDOMLY SELECTED.

Method	5%		10%		20%		40%	
	IoU^c	OA	IoU^c	OA	IoU^c	OA	IoU^c	OA
Sup. only	61.0	97.60	66.8	98.13	72.3	98.44	74.9	98.60
MT-EMA	67.1 (+6.1)	98.14 (+0.54)	75.0 (+8.2)	98.63 (+0.50)	76.6 (+4.3)	98.71 (+0.27)	77.0 (+2.1)	98.73 (+0.13)
MT-AEMA	68.9 (+7.9)	98.23 (+0.63)	76.1 (+9.3)	98.66 (+0.53)	77.7 (+5.4)	98.78 (+0.34)	77.8 (+2.9)	98.78 (+0.18)
(MT-EMA)+AF*	72.0 (+11.0)	98.43 (+0.83)	76.8 (+10.0)	98.72 (+0.59)	77.5 (+5.2)	98.74 (+0.30)	78.5 (+3.6)	98.80 (+0.20)
(MT-EMA)+AF	77.0 (+16.0)	98.72 (+1.12)	78.8 (+12.0)	98.83 (+0.70)	80.4 (+8.1)	98.91 (+0.47)	80.0 (+5.1)	98.90 (+0.30)
AdaSemiCD	77.7 (+16.7)	98.78 (+1.18)	79.4 (+12.6)	98.87 (+0.74)	80.3 (+8.0)	98.92 (+0.48)	80.6 (+5.7)	98.93 (+0.33)

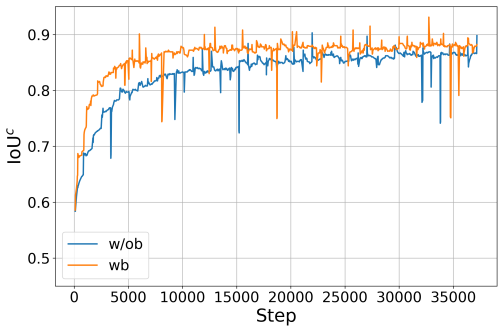

Fig. 6. The influence of class rebalancing on the quality of pseudo-labels generated for unlabeled samples during training.

comparable to other methods with reference count, FLOPs, and inference time. The main reason for the longer training time was that it took about 0.3s for each iteration to generate and evaluate pseudo-labels twice, while it only took about 0.006s and 0.03s for fusion and EMA parameters updating, respectively. Moreover, our model balances performance and time consumption, offering a notable performance benefit with only a slight increase in time. **Hyperparameters in Ramp-up:** The Ramp-up process has a significant influence on the performance of our AdaSemiCD on SSCD. Therefore, we conduct experiments on the selection of two hyperparameters (γ and w_{max}) that control the Ramp-up process. As shown in Table VI, our method achieves the best performance on the three datasets under the combination of parameters (0.1, 10), (0.1, 0.1), and (0.1, 1.0) respectively. Moreover, our method is sensitive to this hyperparameter, and inappropriate

TABLE V
ABLATION STUDY OF PSEUDO-LABEL QUALIFICATION METRIC ON LEVIR-CD DATASET.

Method	5%		10%		20%		40%	
	IoU^c	OA	IoU^c	OA	IoU^c	OA	IoU^c	OA
Sup. only	61.0	97.60	66.8	98.13	72.3	98.44	74.9	98.60
Entropy	73.2 (+12.2)	98.50 (+0.90)	77.4 (+10.6)	98.75 (+0.62)	78.6 (+6.3)	98.81 (+0.37)	79.2 (+4.3)	98.89 (+0.29)
Entropy+rebalance	76.0 (+15.0)	98.63 (+1.03)	78.7 (+11.9)	98.82 (+0.69)	79.6 (+7.3)	98.85 (+0.41)	79.6 (+4.7)	98.86 (+0.26)
Entropy+confusion	74.6 (+13.6)	98.61 (+1.01)	78.0 (+11.2)	98.79 (+0.66)	79.2 (+6.9)	98.82 (+0.38)	79.8 (+4.9)	98.87 (+0.27)
Uncertainty	77.7 (+16.7)	98.78 (+1.18)	79.4 (+12.6)	98.87 (+0.74)	80.3 (+8.0)	98.92 (+0.48)	80.6 (+5.7)	98.93 (+0.33)

TABLE VI
SENSITIVITY ANALYSIS OF RAMP-UP HYPERPARAMETERS WITH 10% LABELED DATA ON THE LEVIR-CD, WHU-CD, AND CDD DATASETS.

γ	w_{max}	LEVIR-CD		WHU-CD		CDD	
		IoU^c	OA	IoU^c	OA	IoU^c	OA
0	0	66.8	98.13	55.7	97.53	67.9	95.46
0.05	1.0	67.2	98.17	53.8	97.02	74.4	96.35
0.1	1.0	71.8	98.32	61.0	98.10	**77.3**	**96.89**
0.3	1.0	69.9	98.26	59.4	98.03	76.2	96.56
0.5	1.0	68.7	98.15	60.9	98.25	76.3	96.58
1.0	1.0	67.3	98.13	60.5	98.18	72.3	95.98
0.1	0.1	65.2	97.60	**70.8**	**98.70**	71.6	95.77
0.1	0.5	68.3	98.14	66.9	98.54	75.8	96.67
0.1	5.0	73.9	98.75	60.1	98.00	69.1	95.51
0.1	10.0	**79.4**	**98.87**	52.4	97.40	68.2	95.50
0.1	30.0	71.9	98.42	50.34	97.12	65.4	95.20

TABLE VII
COMPARISON OF PARAMETERS, COMPUTING COMPLEXITY, AND TRAINING TIME OF DIFFERENT SSCD METHODS ON 5% LABELED LEVIR-CD DATASET.

Method	Params(M)	FLOPs(G)	Training Time(s)	Inference Time(ms)	IoU^c
Sup.Only	46.85	585.85	77	56	61.0
AdvEnt [62]	46.85	585.85	405	63	66.1
s4GAN [35]	46.85	585.85	585	58	64.0
SemiCDNet [22]	46.85	585.85	408	75	67.6
RCR [21]	46.85	585.85	742	59	72.5
FPA [23]	46.85	585.85	727	68	73.7
AdaSemiCD	46.85	585.85	915	67	**77.7**
Oracle	46.85	585.85	293	55	77.9

parameter selection will cause large performance attenuation. This is because our method conducts supervised training on labeled samples and unsupervised training on unlabeled samples at the same time. If the relationship between the two cannot be properly balanced, overfitting on labeled samples or excessive noise interference from unlabeled samples will be caused. All of our remaining experiments were performed at this hyperparameter setting, and the hyperparameters of the compared methods were consistent with the best choices in their original paper.

V. CONCLUSION

In this study, we present AdaSemiCD, a flexible semi-supervised framework for change detection. This framework assesses the quality of pseudo-labels on unlabeled training samples and implements adaptive modifications based on the assessment outcomes, which include sample fusion (AdaFusion), and parameter updates (AdaEMA). Despite the complexity of the scenes, our model successfully identifies the areas of interest with minimal interference during training. Empirical evidence from ten publicly available CD datasets attests to the efficacy of our methodology. Looking ahead, this adaptive processing technique holds promise for potential application in other semi-supervised tasks.

REFERENCES

[1] X. Zhang, Y. Yang, L. Ran, L. Chen, K. Wang, L. Yu, P. Wang, and Y. Zhang, "Remote sensing image semantic change detection boosted by semi-supervised contrastive learning of semantic segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–13, 2024.

[2] M. Bouziani, K. Goïta, and D.-C. He, "Automatic change detection of buildings in urban environment from very high spatial resolution images using existing geodatabase and prior knowledge," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 1, pp. 143–153, 2010.

[3] Y. Xing, Q. Zhang, L. Ran, X. Zhang, H. Yin, and Y. Zhang, "Progressive modality-alignment for unsupervised heterogeneous change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023.

[4] ——, "Improving reliability of heterogeneous change detection by sample synthesis and knowledge transfer," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–11, 2024.

[5] R. Anniballe, F. Noto, T. Scalia, C. Bignami, S. Stramondo, M. Chini, and N. Pierdicca, "Earthquake damage mapping: An overall assessment of ground surveys and vhr image change detection after l'aquila 2009 earthquake," *Remote sensing of environment*, vol. 210, pp. 166–178, 2018.

[6] Z. Y. Lv, W. Shi, X. Zhang, and J. A. Benediktsson, "Landslide inventory mapping from bitemporal high-resolution remote sensing images using change detection and multiscale segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 5, pp. 1520–1532, 2018.

[7] H. Zhang, H. Chen, C. Zhou, K. Chen, C. Liu, Z. Zou, and Z. Shi, "BiFA: Remote sensing image change detection with bitemporal feature alignment," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–17, 2024.

[8] H. Chen, H. Zhang, K. Chen, C. Zhou, S. Chen, Z. Zou, and Z. Shi, "Continuous cross-resolution remote sensing image change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–20, 2023.

[9] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[10] W. G. C. Bandara and V. M. Patel, "A transformer-based siamese network for change detection," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 207–210.

[11] C. Wu, B. Du, and L. Zhang, "Fully convolutional change detection framework with generative adversarial network for unsupervised, weakly supervised and regional supervised change detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 9774–9788, 2023.

[12] Z. Li, C. Tang, X. Liu, C. Li, X. Li, and W. Zhang, "MS-Former: Memory-supported transformer for weakly supervised change detection with patch-level annotations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–13, 2024.

[13] L. Yan, J. Yang, and J. Wang, "Domain knowledge-guided self-supervised change detection for remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 4167–4179, 2023.

[14] X. Tang, H. Zhang, L. Mou, F. Liu, X. Zhang, X. X. Zhu, and L. Jiao, "An unsupervised remote sensing change detection method based on multiscale graph convolutional network and metric learning," *IEEE*

Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–15, 2022.

[15] H. Chen, Y. Zao, L. Liu, S. Chen, and Z. Shi, "Semantic decoupled representation learning for remote sensing image change detection," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 1051–1054.

[16] H. Chen, W. Li, S. Chen, and Z. Shi, "Semantic-aware dense representation learning for remote sensing image change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2022.

[17] H. Chen, W. Li, and Z. Shi, "Adversarial instance augmentation for building change detection in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[18] C. Ren, X. Wang, J. Gao, X. Zhou, and H. Chen, "Unsupervised change detection in satellite images with generative adversarial network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 12, pp. 10 047–10 061, 2021.

[19] Z. Wang, D. Liu, Z. Wang, X. Liao, and Q. Zhang, "A new remote sensing change detection data augmentation method based on mosaic simulation and haze image simulation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 4579–4590, 2023.

[20] W. G. C. Bandara, N. G. Nair, and V. M. Patel, "DDPM-CD: Denoising diffusion probabilistic models as feature extractors for change detection," in *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2025.

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

[22] D. Peng, L. Bruzzone, Y. Zhang, H. Guan, H. Ding, and X. Huang, "SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5891–5906, Jul. 2021.

[23] X. Zhang, X. Huang, and J. Li, "Semisupervised change detection with feature-prediction alignment," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023.

[24] L. Ran, W. Zhan, Y. Li, X. Zhang, and S. Zhang, "DTFSeg: A dynamic threshold filtering method for semi-supervised semantic segmentation," in *2023 China Automation Congress (CAC)*. Chongqing, China: IEEE, Nov. 2023, pp. 7571–7576.

[25] L. Ran, L. Wang, T. Zhuo, Y. Xing, and Y. Zhang, "DDF: A novel dual-domain image fusion strategy for remote sensing image semantic segmentation with unsupervised domain adaptation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–13, 2024.

[26] L. Ran, Y. Li, G. Liang, and Y. Zhang, "Pseudo labeling methods for semi-supervised semantic segmentation: A review and future perspectives," *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[27] 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 Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.

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

[29] W. Feng, X. Gao, S. Boukir, Z. Xie, Y. Quan, W. Huang, and M. Xing, "Hypothesis margin-based ensemble method for the classification of noisy remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–21, 2023.

[30] Z. Zhao, L. Yang, S. Long, J. Pi, L. Zhou, and J. Wang, "Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 11 350–11 359.

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

[32] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," in *5th International Conference on Learning Representations, ICLR*, 2017.

[33] P. Mi, J. Lin, Y. Zhou, Y. Shen, G. Luo, X. Sun, L. Cao, R. Fu, Q. Xu, and R. Ji, "Active teacher for semi-supervised object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 14 482–14 491.

[34] L. Yang, L. Qi, L. Feng, W. Zhang, and Y. Shi, "Revisiting weak-to-strong consistency in semi-supervised semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7236–7246.

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

[36] Z. Zhao, S. Long, J. Pi, J. Wang, and L. Zhou, "Instance-specific and model-adaptive supervision for semi-supervised semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 23 705–23 714.

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

[38] X. Li, L. Yu, H. Chen, C.-W. Fu, L. Xing, and P.-A. Heng, "Transformation-consistent self-ensembling model for semisupervised medical image segmentation," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 2, pp. 523–534, 2020.

[39] Y. Zhang, Y. Cheng, and Y. Qi, "SemiSAM: Exploring sam for enhancing semi-supervised medical image segmentation with extremely limited annotations," in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2024.

[40] Q. Xie, Z. Dai, E. Hovy, T. Luong, and Q. Le, "Unsupervised data augmentation for consistency training," *Advances in neural information processing systems*, vol. 33, pp. 6256–6268, 2020.

[41] W. Feng, Y. Quan, G. Dauphin, Q. Li, L. Gao, W. Huang, J. Xia, W. Zhu, and M. Xing, "Semi-supervised rotation forest based on ensemble margin theory for the classification of hyperspectral image with limited training data," *Information Sciences*, vol. 575, pp. 611–638, 2021.

[42] X. Lai, Z. Tian, L. Jiang, S. Liu, H. Zhao, L. Wang, and J. Jia, "Semi-supervised semantic segmentation with directional context-aware consistency," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1205–1214.

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

[44] Y. Wang, H. Wang, Y. Shen, J. Fei, W. Li, G. Jin, L. Wu, R. Zhao, and X. Le, "Semi-supervised semantic segmentation using unreliable pseudo-labels," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 4248–4257.

[45] E. Bousias Alexakis and C. Armenakis, "Evaluation of semi-supervised learning for cnn-based change detection," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp. 829–836, 2021.

[46] Z. Mao, X. Tong, and Z. Luo, "Semi-supervised remote sensing image change detection using mean teacher model for constructing pseudo-labels," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

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

[48] S. Hafner, Y. Ban, and A. Nascetti, "Urban change detection using a dual-task siamese network and semi-supervised learning," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 1071–1074.

[49] J. Liu, K. Chen, G. Xu, H. Li, M. Yan, W. Diao, and X. Sun, "Semi-supervised change detection based on graphs with generative adversarial networks," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 74–77.

[50] S. Yang, S. Hou, Y. Zhang, H. Wang, and X. Ma, "Change detection of high-resolution remote sensing image based on semi-supervised segmentation and adversarial learning," in *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2022, pp. 1055–1058.

[51] J. G. Vinholi, P. R. B. d. Silva, D. I. Alves, and R. Machado, "Enhancing change detection in ultra-wideband vhf sar imagery: An entropy-based approach with median ground scene masking," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–15, 2024.

[52] H. Zhang, M. Cissé, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *6th International Conference on Learning Representations*, 2018.

[53] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sensing*, vol. 12, no. 10, p. 1662, 2020.
[54] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Transactions on geoscience and remote sensing*, vol. 57, no. 1, pp. 574–586, 2018.
[55] S. Holail, T. Saleh, X. Xiao, and D. Li, "Afde-net: Building change detection using attention-based feature differential enhancement for satellite imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.
[56] J. Zhang, Z. Shao, Q. Ding, X. Huang, Y. Wang, X. Zhou, and D. Li, "AERNet: An attention-guided edge refinement network and a dataset for remote sensing building change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023.
[57] M. Lebedev, Y. V. Vizilter, O. Vygolov, V. A. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 565–571, 2018.
[58] C. Zhang, P. Yue, D. Tapete, L. Jiang, B. Shangguan, L. Huang, and G. Liu, "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 183–200, 2020.
[59] 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 Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.
[60] M. Liu, Z. Chai, H. Deng, and R. Liu, "A cnn-transformer network with multiscale context aggregation for fine-grained cropland change detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 4297–4306, 2022.
[61] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "Randaugment: Practical automated data augmentation with a reduced search space," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 702–703.
[62] T.-H. Vu, H. Jain, M. Bucher, M. Cord, and P. Pérez, "Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2517–2526.