



## Article

# Semi-Supervised Change Detection with Data Augmentation and Adaptive Thresholding for High-Resolution Remote Sensing Images

Wuxia Zhang <sup>1,\*</sup> , Xinlong Shu <sup>1</sup>, Siyuan Wu <sup>2</sup>  and Songtao Ding <sup>1</sup>

<sup>1</sup> School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China; xinlong@stu.xupt.edu.cn (X.S.); dst@xupt.edu.cn (S.D.)

<sup>2</sup> School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China; wusiyuan@opt.ac.cn

\* Correspondence: zhangwuxia@xupt.edu.cn

**Abstract:** Change detection (CD) is an important research direction in the field of remote sensing, which aims to analyze the changes in the same area over different periods and is widely used in urban planning and environmental protection. While supervised learning methods in change detection have demonstrated substantial efficacy, they are often hindered by the rising costs associated with data annotation. Semi-supervised methods have attracted increasing interest, offering promising results with limited data labeling. These approaches typically employ strategies such as consistency regularization, pseudo-labeling, and generative adversarial networks. However, they usually face the problems of insufficient data augmentation and unbalanced quality and quantity of pseudo-labeling. To address the above problems, we propose a semi-supervised change detection method with data augmentation and adaptive threshold updating (DA-AT) for high-resolution remote sensing images. Firstly, a channel-level data augmentation (CLDA) technique is designed to enhance the strong augmentation effect and improve consistency regularization so as to address the problem of insufficient feature representation. Secondly, an adaptive threshold (AT) is proposed to dynamically adjust the threshold during the training process to balance the quality and quantity of pseudo-labeling so as to optimize the self-training process. Finally, an adaptive class weight (ACW) mechanism is proposed to alleviate the impact of the imbalance between the changed classes and the unchanged classes, which effectively enhances the learning ability of the model for the changed classes. We verify the effectiveness and robustness of the proposed method on two high-resolution remote sensing image datasets, WHU-CD and LEVIR-CD. We compare our method to five state-of-the-art change detection methods and show that it achieves better or comparable results.

**Keywords:** semi-supervised change detection; pseudo-labeling; adaptive threshold; unbalanced; consistency regularization



Academic Editor: Ganchao Liu

Received: 22 November 2024

Revised: 27 December 2024

Accepted: 6 January 2025

Published: 7 January 2025

**Citation:** Zhang, W.; Shu, X.; Wu, S.; Ding, S. Semi-Supervised Change Detection with Data Augmentation and Adaptive Thresholding for High-Resolution Remote Sensing Images. *Remote Sens.* **2025**, *17*, 178. <https://doi.org/10.3390/rs17020178>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Change detection (CD) in remote sensing typically involves analyzing either bi-temporal or long-time series remote sensing images of the same area captured at different times to identify and quantify changes. Change detection encompasses not only the binary classification task of changed or unchanged, but also methods that identify multiple classes before and after the change. This study focuses on exploring change detection based on bi-temporal remote sensing images, with particular emphasis on the task of binary classification into changed or unchanged. With the continuous improvement of satellite sensors,

the resolution of remote sensing images is also increasing. High-resolution remote sensing images usually have higher spatial resolution and can provide more information about surface features and spatial distribution, which helps in detecting more subtle changes, such as alterations to small buildings. This makes them particularly useful in areas where highly detailed information is required, such as urban planning [1], land use [2], disaster assessment [3], forest monitoring [4], and military [5].

At present, a large number of studies focus on how to effectively extract the change region of dual-time high-resolution remote sensing images [6,7]. With the continuous improvement of image segmentation techniques [8,9], the effectiveness and accuracy of fully supervised change detection models are also increasing. Nowadays, mainstream change detection methods [10–13] still rely on a large amount of reliable labeled data to train models. However, there is a challenging problem with this kind of method; that is, labeling pixel-level change detection labels requires a lot of manpower and material resources. Especially in the military field of target change detection, it is very difficult to obtain a large amount of accurate pixel-level label data. In order to solve the problem of difficulty in obtaining pixel-level label data at a high cost, researchers pay more and more attention to semi-supervised change detection (SSCD) methods using a small amount of label data, weakly supervised change detection methods using images and labels, and unsupervised change detection methods without using label data [14]. Unsupervised change detection methods are mainly divided into two categories. One is to use the pre-trained model trained by the remote sensing images of scene classification as the feature extractor for change detection [15,16]. The other is the combination of classical unsupervised algorithms and supervised learning change detection frameworks [17,18]. This method relies heavily on prior knowledge, which limits its application. Weakly supervised change detection uses coarse-grained image-level labels for training [19,20], which effectively mitigates the problem of the high cost of manual annotation. However, the quality of weakly supervised labels still seriously affects the accuracy of the model. In contrast, SSCD methods use a small amount of labeled data, which not only reduces the cost of labeled data but also achieves good change detection results [21].

Currently, the core idea of SSCD is to train the model using a large amount of unlabeled data. We can broadly divide it into the following three methods: (1) adversarial learning-based methods [22–24], which mainly use generative adversarial networks (GANs) to obtain the data distribution using the generator and use the discriminator to distinguish between different images; (2) consistency regularization-based methods [25–28], which constrain the model's prediction results on different enhanced versions of the input data by applying different forms of perturbations to the data; (3) pseudo-labeling methods [29,30], which use the model to predict the unlabeled data and then select the pseudo-labels with confidence. The pseudo-labels are then mixed with the true labels to train the model.

Although current SSCD methods effectively reduce the cost of data labeling, they still face several challenges. Firstly, adversarial learning-based methods suffer from training instability, making it difficult for the model to achieve optimal results [22]. Secondly, consistency regularization-based methods are often influenced by the degree and type of perturbations applied, which can limit the robustness of the model [25]. Additionally, the approach used to select pseudo-labeling has a significant impact on model performance. Typically, a fixed threshold is set to determine the prediction value for each pixel, but this approach can introduce issues [30]. If the threshold is set too high, the pseudo-labeling will only capture part of the information from the original remote sensing image, compromising the completeness of the change features [29]. On the other hand, if the threshold is set too low, a large number of incorrect labels may be introduced, further reducing model performance [26].

In response to the problems mentioned above, we propose SSCD with joint data augmentation and adaptive threshold updating (DA-AT) for high-resolution remote sensing images. From the perspective of semi-supervised strategy, inspired by FixMatch [31] and SoftMatch [32], the proposed method uses a weak–strong data augmentation method to perturb the initial unlabeled dataset and then uses an adaptive threshold (AT) to filter out high-quality pseudo-labeling from the results generated by the weakly augmented data through the model. Pseudo-labels are then used to supervise the results generated by the strong augmentation passing through the model. At the same time, we introduce channel-level data augmentation (CLDA) methods on the basis of image-level data perturbation to further improve the diversity of data. Considering the varying learning difficulties associated with changed and unchanged classes, we propose adaptive class weight (ACW) to encourage the model to focus more on learning from minority class information. From the perspective of the training flow, our network is based on the encoder–decoder structure [26] and is divided into two parts: supervised and unsupervised. The supervised and unsupervised parts are trained simultaneously and share weight information. The proposed DA-AT method is applied in the unsupervised training part, where the two parts jointly guide the optimization of the model. Overall, the DA-AT method achieves a balance between the number and quality of pseudo-labeling while introducing CLDA to better explore the intrinsic consistency of image features.

The main contributions of the proposed DA-AT method are as follows:

- The channel-level data augmentation is designed to improve the strong enhancement effect by finer-grained enhancement operations. By enhancing the role of consistency regularization, the model is promoted to learn the intrinsic consistency of image features, so as to improve the robustness and accuracy of the model.
- The dynamic adaptive threshold is proposed to achieve the best balance between the quality and quantity of pseudo-labels by adjusting the threshold in real time during the training process. This mechanism can filter out the optimal pseudo-labels, so as to improve the self-training performance of the model and further optimize the training effect of the model.
- The adaptive class weight mechanism is presented to deal with the problem of data imbalance between the changed class and the unchanged class. This mechanism reduces the negative impact of data imbalance on model learning by adjusting the class weight, so as to improve the learning ability and detection accuracy of the model for changing classes.

## 2. Related Work

### 2.1. Semi-Supervised Learning

Semi-supervised learning is a strategy that drives model training by leveraging a small amount of labeled data and a large amount of unlabeled data. Combining these two types of data, semi-supervised learning can improve the model's performance and generalization ability while reducing labeling costs. In the past decade, methods based on semi-supervised learning have been widely studied and applied, especially in the fields of image classification and semantic segmentation [33]. At present, semi-supervised learning methods mainly include generative adversarial networks (GANs), consistency regularization methods, and pseudo-labeling methods.

GAN-based methods usually contain several generators and discriminators, which indirectly provide supervision information to the model through adversarial learning between the generator and discriminator. Wang et al. [34] proposed a CCS-GAN model for classification, which contains a classifier in addition to the generator and discriminator. This method uses enhanced feature matching to constrain the generator to learn the sample

distribution, which further helps the classifier learn a more accurate decision boundary. Hung et al. [35] designed a fully convolutional discriminator and used the discriminator to identify credible regions from the prediction results of unlabeled data as an additional supervision signal in semi-supervised learning.

Consistency regularization-based methods improve performance by optimizing the consistency of model output under different perturbations. It is very important to choose the appropriate perturbation mode. For example, Yun et al. [36] proposed data-level perturbations to enhance the diversity of the input dataset and thus improve the performance of the model. Li et al. [37] proposed feature-level perturbation to add perturbation to the intermediate features of the encoder. Chen et al. [38] proposed a method called Cross Pseudo Supervision (CPS), which imposes different disturbances on two semantic segmentation networks and uses the prediction results of one network to supervise the other network.

Pseudo-labeling methods generate pseudo-labels by using a threshold filter on unlabeled data, thus making the training process appear to be supervised self-training. Feng et al. [39] leveraged the discrepancies in pseudo-label predictions between different models to dynamically mutually train two distinct models, achieving significant results in the fields of semantic segmentation and image classification. Sohn et al. [31] proposed Fix-Match by using a high threshold to filter pseudo-labels from weakly augmented branches and using these pseudo-labels to supervise the prediction results of strongly augmented branches. However, this method may limit the opportunities for the model to learn from the rich unlabeled data at the early stage of training, which in turn affects the training efficiency. Based on this, Yang et al. [40] proposed UniMatch, which uses the pseudo-labeling of the weakly enhanced branch filtered by a fixed threshold to supervise the prediction results of the two strongly enhanced branches. Wang et al. [32] proposed FreeMatch on this basis. By improving the shortcomings of fixed threshold, they proposed an adaptive threshold adjustment strategy to solve the problem of low data utilization caused by fixed or temporary adjustment of threshold, and achieved remarkable results in the field of image classification.

## 2.2. Semi-Supervised CD

SSCD is a method to detect changes in images under the condition of limited labeled data. Supervised change detection methods are limited in practice due to the expensive and time-consuming acquisition of labeled data. By combining limited labeled data with a large amount of unlabeled data, SSCD methods reduce the demand and cost of labeled data while improving detection performance.

To address the annotation cost problem, several SSCD methods have been developed in recent years. Gong et al. [17] proposed a Generative Discriminative Classification Network (GDCN), in which the discriminator classifies the data into changed, unchanged, extra, and fake classes, while the generator is used to augment the data. However, the stability of the generator and the choice of noise have a large impact on the model. Wang et al. [41] achieved semi-supervised semantic segmentation of remote sensing images through pseudo-label average updating and consistency regularization. Peng et al. [23] proposed a semi-supervised convolutional network for CD (SemiCDNet), which uses a generator to generate prediction maps and entropy maps and a discriminator to enforce the feature distribution consistency of segmentation maps and entropy maps between labeled and unlabeled data.

To solve the problem of insufficient training data, Chen et al. [42] proposed instance-level change augmented CDNet (IAug\_CDNet), which synthesized data through GAN to increase sample diversity and focused on building change detection. Nie et al. [43] proposed an SSCD framework based on Siamese GAN networks, which improves the change

detection performance by generating adversarial training that is difficult to distinguish between pseudo-labels and real labels. Bandara and Patel et al. [26] proposed a slight perturbation consistency regularization principle for feature difference maps, which effectively improved the utilization efficiency of the model on unlabeled data. Wang et al. [27] proposed Reliable Contrastive Learning (RCL), which introduces contrastive loss and selects pseudo-labeling from the best pre-trained model to improve change detection accuracy, but the process of separate training may lead to inconsistent model optimization.

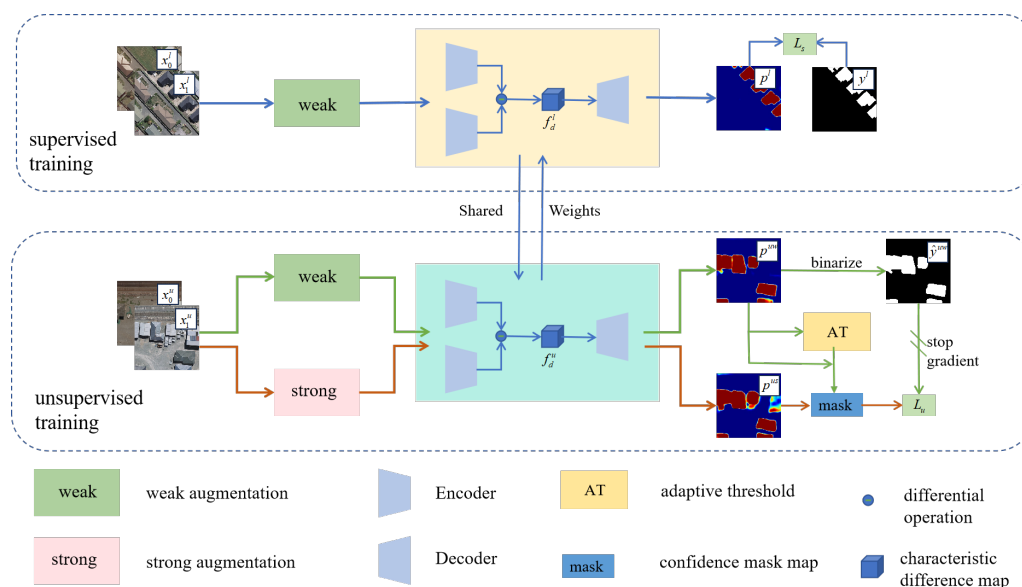
Sun et al. [29] proposed a Siamese nested UNet with graph attention mechanism (SemiSANet), which introduces a variety of data enhancement methods and graph attention mechanisms, but its pseudo-labeling screening uses a fixed threshold, which reduces the diversity of data learned by the model. Mao et al. [44] used the weak enhancement of the teacher network and the strong enhancement of the student network to enforce consistency constraints, making full use of unlabeled data to participate in training. On the basis of consistency regularization of predecessors, Zhang et al. [30] further proposed feature consistency alignment, namely class-based feature alignment (FA) and pixel-level predictive alignment (PA), to further constrain the model on changed and unchanged features so as to better improve the change detection ability of the model.

### 3. Method

This section mainly introduces the basic network framework structure of DA-AT. Then, the channel-level data augmentation (CLDA), adaptive threshold (AT) and adaptive class weight (ACW) are introduced.

#### 3.1. DA-AT Network Architecture

As shown in Figure 1, the training consists of a supervised branch using a labeled dataset and an unsupervised branch using an unlabeled dataset. We first outline the SSCD settings, followed by an introduction to the encoder–decoder structure, and then delineate the flow of both the supervised and unsupervised branches.



**Figure 1.** Architecture diagram of DA-AT. It includes the supervised branch, the unsupervised branch, and sub-modules of the semi-supervised change detection model.

##### 3.1.1. SSCD Settings

In the semi-supervised change detection setting, the following tags and descriptions are primarily used. The training dataset part contains two parts: labeled dataset  $D^l$ , and un-

labeled dataset  $D^l$ . Specifically, the labeled dataset is denoted as  $D^l = \{x_{0,i}^l, x_{1,i}^l, y_i^l\}_{i=1}^n$ , where  $\{x_{0,i}^l, x_{1,i}^l\}$  represents the  $i$ th pair of bi-temporal remote sensing images,  $x_{0,i}^l$  represents the image at time T0,  $x_{1,i}^l$  represents the image at time T1,  $y_i^l$  represents the corresponding binary pixel-level ground truth, and  $n$  represents the size of the dataset. The unlabeled dataset is specifically denoted as  $D^u = \{x_{0,i}^u, x_{1,i}^u\}_{i=1}^N$ , where the  $i$ th pair of bi-temporal remote sensing images in generation  $\{x_{0,i}^u, x_{1,i}^u\}$  table  $N$  represents the size of the dataset. Different from the labeled dataset, it does not have the corresponding label of the  $i$ th pair of bi-temporal remote sensing images. The labeled dataset  $n$  has far fewer image pairs than the unlabeled dataset  $N$ .

### 3.1.2. Shared Encoder–Decoder Model

In the field of semi-supervised remote sensing change detection based on consistency regularization, the core is to explore effective strategies, but the network structure is not the focus of research. Therefore, we adopt the basic encoder–decoder structure [26,45], and the model mainly includes two identical encoders  $E$ , a decoder  $D$ , and a pyramid pooling module (PPM) [46]. Two encoders are responsible for extracting features from two remote sensing images, respectively, and the decoder processes the difference features to obtain the change probability map.

Specifically, a pair of remote sensing images  $x_0, x_1$  are fed into two encoders  $E$  after the same data augmentation, where  $x_0$  and  $x_1$  have the same dimensions  $H \times W \times 3$ . This results in a pair of feature maps  $f_0 \in \mathbf{R}^{H/s \times W/s \times C}$  and  $f_1 \in \mathbf{R}^{H/s \times W/s \times C}$ , which are both of the same size and have high dimensionality. Here,  $H$  and  $W$  represent the height and width of the input images, respectively, while  $C$  and  $s$  denote the number of feature spatial channels and the spatial scale ratio. Notably, we employ a pre-trained ResNet50 model [47], where  $C$  and  $s$  are set to 2048 and 8, respectively.

$$f_0 = E(x_0) \quad (1)$$

$$f_1 = E(x_1) \quad (2)$$

Next, to compute the feature difference of the two images after applying the same type of data augmentation (either weak or strong augmentation),  $f_0$  and  $f_1$  are used in the difference operation. However, a simple difference operation may result in the loss of important details. To address this, we utilize the PPM to further process the feature difference map. This allows us to capture feature information  $f_d$  at different scales, which is crucial for understanding the context of the image.

$$f_d = \text{PPM}(|f_0 - f_1|) \quad (3)$$

Finally, the decoder  $D$  is employed to predict the feature difference map. The decoder  $D$  is composed of a series of convolutional upsampling modules [48], which aim to restore the spatial resolution of the input  $f_d$ , and output a probability map  $p \in \mathbf{R}^{H \times W \times 2}$ . The number 2 in  $p$  represents the two classes: changed and unchanged. The map  $p$  is then scaled by applying a softmax function along the category dimension, resulting in pixel-wise predictions in a range of  $[0, 1]$ . The number of input channels and output classes in our model can be flexibly adjusted according to the specific requirements of the dataset and task at hand.

$$p = D(f_d) \quad (4)$$

$$p(i, j) = \text{softmax}(p(i, j)) \quad (5)$$

Here,  $(i, j)$  denotes the spatial pixel location in the probabilistic prediction map  $p$ .

### 3.1.3. DA-AT Framework

In Figure 1, our training partition is divided into two parts according to whether labels are used or not, the supervised branch and the unsupervised branch are trained at the same time, and the weight information is shared. The labeled dataset  $D^l$  drives the supervised branch. Initially, the input bi-temporal remote sensing image  $\{x_0^l, x_1^l\}$  undergoes weak augmentation to obtain  $\{x_0^{lw}, x_1^{lw}\}$ , and then, it is input into the encoder–decoder. This process yields pixel-level change probability maps  $p^l$ . For the supervised part, cross-entropy (CE) loss [49] is used to minimize the loss between the probability map  $p^l$  and the label  $y^l$ . It is expressed as follows:

$$L_s = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W CE(p^l(i, j), y^l(i, j)) \quad (6)$$

The unlabeled dataset  $D^u$  is utilized in the unsupervised branch. Initially, the input bi-temporal remote sensing image pair  $\{x_0^u, x_1^u\}$  undergoes both weak and strong augmentations to obtain  $\{x_0^{uw}, x_1^{uw}\}$  and  $\{x_0^{us}, x_1^{us}\}$ . These two sets of enhanced image pairs are then fed separately into the encoder–decoder network, resulting in pixel-level change probability maps  $p^{uw}$  and  $p^{us}$ . The map  $p^{uw}$  is subsequently used to update the adaptive threshold filter, and a corresponding pseudo-label  $\hat{y}^{uw}$  is generated through binarization. Here, “stop gradient” refers to the fact that the results of the weak augmentation branch only provide pseudo-labels for the self-training of the strong augmentation branch. The consistency loss function is then applied to minimize the loss between the probability map  $p^{us}$  and the pseudo-label  $\hat{y}^{uw}$ . This process can be summarized as follows:

$$L_u = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \mathbb{1}(\max(p^{uw}(i, j)) \geq \tau) \cdot H(p^{us}(i, j), \hat{y}^{uw}(i, j)) \quad (7)$$

In this context,  $\tau$  represents the predefined confidence threshold used to filter out noisy labels. The condition  $\mathbb{1}(\max(p^{uw}(i, j)) \geq \tau)$  indicates that if the predicted probability exceeds  $\tau$ , it is considered a high-quality pseudo-label and is assigned a value of 1. Otherwise, it is set to 0. The function  $H$  typically refers to the CE loss. The details of weak and strong augmentations here are described in the Channel-Level Data Augmentation section.

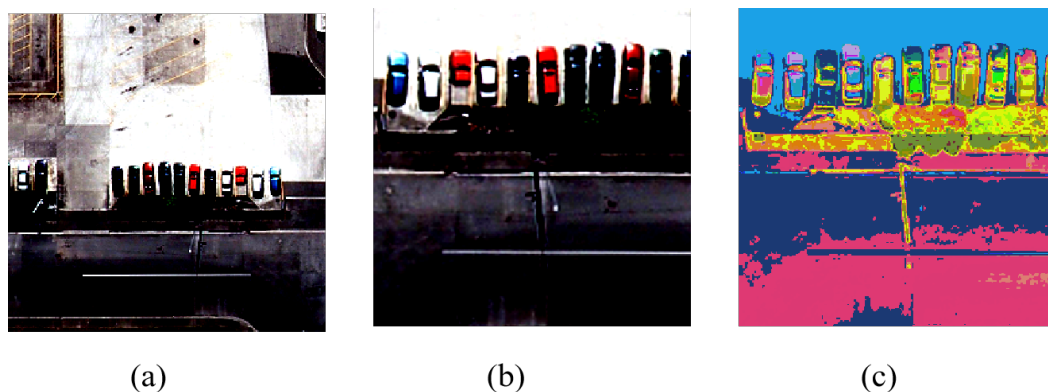
Specifically, in the experimental stage, the proposed semi-supervised framework integrates CLDA, AT, and ACW to achieve robust and effective change detection. First, weak augmentation is applied to the input bi-temporal remote sensing image pair  $\{x_0^u, x_1^u\}$ , generating  $\{x_0^{uw}, x_1^{uw}\}$ , which is processed by the network to produce pixel-level change probability maps  $p^{uw}$ . Simultaneously, CLDA introduces strong augmentation to the same input pair, yielding  $\{x_0^{us}, x_1^{us}\}$  and corresponding probability maps  $p^{us}$ . By enforcing consistency between  $p^{uw}$  and  $p^{us}$ , CLDA improves the model’s robustness to input variations and strengthens its feature representations. Based on  $p^{uw}$ , the AT module dynamically adjusts thresholds for pseudo-label selection, balancing their quality and quantity to enhance training stability. Recognizing the inherent imbalance in change detection tasks, the ACW module applies targeted optimization constraints, assigning higher weights to change categories to mitigate class imbalance. Together, these components work synergistically; CLDA

ensures robust and consistent learning across augmentations, AT refines pseudo-labels for effective semi-supervised learning, and ACW emphasizes minority class learning, forming a unified framework that addresses the challenges of semi-supervised change detection.

### 3.2. Channel-Level Data Augmentation

In consistency regularization-based training, utilizing both strong and weak augmentation methods is essential. General weak augmentation techniques, such as random flipping, cropping, and resizing, can increase the diversity of the training data, helping the model to better learn various features. On the other hand, strong augmentations involve more substantial perturbations, such as adjustments to brightness, color, and image masking [50], which introduce a greater degree of change. However, traditional augmentation methods primarily focus on the superficial aspects of the image and often fail to fully exploit the channel information. To address this limitation, we improve the strong enhancement based on the randomized quantization (RQ) data enhancement method proposed by Wu et al. [51].

In Figure 2, the weak augmentation result is achieved by sequentially applying resize, crop, and flip operations to the original image. Specifically, the resize operation adjusts the image size within a range of [0.5, 2.0]. The crop operation randomly generates a crop size and extracts a corresponding region of the image. Finally, the flip operation applies a horizontal flip with a 50% probability. Building on this, RQ is then applied to the RGB channels of the image to achieve strong augmentation. Specifically, the data in each channel are divided into a certain number of intervals, defined by  $RN$ , and the original value  $x$  within each interval is mapped to a randomly sampled value  $y$  from the same interval. This approach introduces a unique quantization value for each interval, thereby enhancing the diversity of the data. Based on our experiments and the referenced paper, we set the  $RN$  to 8, as a smaller interval size results in a more pronounced augmentation effect.



**Figure 2.** Data augmentation is used for remote sensing images. (a) is the original image, (b) is the weak augmentation, and (c) is the result of the strong augmentation.

### 3.3. Adaptive Threshold

To fully utilize unlabeled datasets and enhance the model's generalization ability, we generate a pseudo-label  $\hat{y}^{uw} \in \mathbf{R}^{H \times W}$  corresponding to the predicted probability map  $p^{uw}$ , which is obtained after weak augmentation and then binarized. This pseudo-label  $\hat{y}^{uw}$  is used to supervise the predicted probability map  $p^{us}$ , which is obtained after strong augmentation. The formula for generating the pseudo-label is expressed as follows:

$$\hat{y}^{uw}(i, j) = \arg \max_{c \in \{0,1\}} p^{uw}(i, j, c) \quad (8)$$

Specifically,  $\hat{y}^{uw}$  represents the class with the maximum predicted probability in  $p^{uw}$  for the image pair at spatial position  $(i, j)$ . Here, 1 in  $c$  denotes the changed class, while 0 denotes the unchanged class.

As mentioned earlier, a fixed threshold  $\tau$  (e.g., 0.5, 0.95, 0.99) is commonly used to filter out high-quality pixel-wise pseudo-labels to constrain the results of strong augmentation. For clarity, we denote the confidence mask corresponding to  $\mathbb{1}(\max(p^{uw}(i, j)) \geq \tau)$  as  $\mathbf{m} \in \mathbf{R}^{H \times W}$ . The corresponding calculation formula is expressed as follows:

$$\mathbf{m}(i, j) = \begin{cases} 1, & \text{if } p^{uw}(i, j) > \tau \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

Generally, setting a threshold of 0.5 allows the use of all pseudo-labels, but it introduces excessive noise, which can reduce training accuracy. Alternatively, setting a higher threshold, such as 0.99, yields higher-quality pseudo-labels, but this approach may miss the opportunity to learn diverse predictions during the early stages of training, leading to low data utilization and hindering the model's ability to fully learn. To achieve a balance between quantity and quality, we implement an adaptive threshold strategy. This approach gradually increases the confidence threshold  $\tau_t$  based on the model's predictions at round  $t$ .

First, the model is used to predict the weakly augmented unlabeled data, and the maximum prediction probability for each pixel is calculated. Then, the probabilities of each class  $c$  are averaged to obtain a local prediction confidence  $\tau_c$ .

$$\tau_c = \frac{1}{BHW} \sum_{k=1}^B \sum_{i=1}^H \sum_{j=1}^W \max(p_k^{uw}(i, j, c)) \quad (10)$$

Here,  $B$  represents the batch size, and  $p_k^{uw}(i, j, c)$  denotes the maximum probability corresponding to class  $c$  at the spatial location  $(i, j)$  in the  $K$ -th image pair.

Considering the imbalance between categories, we resample the data. Specifically, we calculate the proportion  $Con_c$  of pixels corresponding to the maximum probability of class  $c$ , and then determine the inverse weight  $S_c$  of class  $c$  based on this proportion. The expression is as follows:

$$Con_c = \frac{1}{BHW} \sum_{k=1}^B \sum_{i=1}^H \sum_{j=1}^W \text{sum}(\arg \max p_k^{uw}(i, j, c)) \quad (11)$$

$$S_c = 1 - Con_c \quad (12)$$

Here,  $\text{sum}(\bullet)$  refers to the number of pixels belonging to class  $c$ .

Then, to facilitate the update, we estimate the adaptive threshold  $\tau_t$  as the exponential moving average (EMA) of the confidence for each round. It is initialized to  $\frac{1}{C}$ , where  $C$  is the number of classes.

$$\tau_t = \begin{cases} \frac{1}{C}, & \text{if } t = 0 \\ \lambda \tau_{t-1} + (1 - \lambda) S_c \tau_c, & \text{otherwise} \end{cases} \quad (13)$$

Here,  $\lambda$  represents the momentum decay of the EMA, which falls within the range  $(0, 1)$ . For binary change detection,  $C$  is set to 2, meaning the initial threshold  $\tau_0$  is set to 0.5.

After adopting the adaptive threshold, the confidence mask map corresponding to the high-quality and quantity of pseudo-labels obtained through our screening process can be redefined as  $\mathbf{m}_t \in \mathbf{R}^{H \times W}$ .

$$\mathbf{m}_t(i, j) = \begin{cases} 1, & \text{if } p^{uw}(i, j) > \tau_t \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

### 3.4. Adaptive Class Weight

However, adaptive thresholds alone are insufficient because they overlook the varying learning difficulties of different classes. Intuitively, in the semi-supervised change detection task, predicting the changed class is generally more challenging than predicting the unchanged class. To address this, we propose adaptive class weights, which encourage the model to focus more on training the minority class rather than predominantly on the majority class. Similarly, we resample the number of pixels in each category to obtain the corresponding resampling rate  $\lambda_c$  for each category. Unlike the adaptive threshold, where we adjust by multiplying weights inversely, here, we utilize all the minority classes without inversely scaling the weights. This ensures that the model pays greater attention to the minority class information in the image.

$$\lambda_c = \frac{1 - \text{Con}_c}{\max_c \text{Con}_c}, c \in 0, 1 \quad (15)$$

Here,  $\max(\bullet)$  refers to the largest value across all classes.

Therefore, we can adjust the weight of the loss function based on the resampling rate  $\lambda_c$  of each pixel. The weight map of pixels can be represented as  $\mathbf{w}_t \in \mathbf{R}^{H \times W}$ .

$$\mathbf{w}_t(i, j) = \begin{cases} \lambda_0, & \text{if } \arg \max p^{uw}(i, j, c) = 0 \\ \lambda_1, & \text{otherwise.} \end{cases} \quad (16)$$

Here,  $\arg \max p^{uw}(i, j, c)$  refers to the class with the highest probability predicted by the model at the  $(i, j)$  location.

Thus, the loss used in the unsupervised part can be reformulated as follows:

$$L_u = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \text{CE}(p^{us}(i, j), \hat{y}^{uw}(i, j)) \cdot \mathbf{m}_t(i, j) \cdot \mathbf{w}_t(i, j) \quad (17)$$

### 3.5. Overall Loss and Training Procedure of DA-AT

The total loss can be represented by both the supervised loss  $L_s$  and the unsupervised loss  $L_u$ .

$$L = L_s + L_u \quad (18)$$

To better describe our proposed method for semi-supervised change detection in high-resolution remote sensing images using joint data augmentation and the adaptive threshold updating method, we summarize the training process of DA-AT as Algorithm 1.

**Algorithm 1** DA-AT**Input:**

Labeled dataset  $D^l = \{x_{0,i}^l, x_{1,i}^l, y_i^l\}_{i=1}^n$ , unlabeled dataset  $D^u = \{x_{0,i}^u, x_{1,i}^u\}_{i=1}^N$ , backbone network  $E-D$ , batch size  $B$ , number of iterations  $N_i$ , and number of epochs  $N_e$ . The initial value of the adaptive threshold is 0.5.

**Steps:**

- 1: **Data processing:** For the labeled data, we obtain the sample  $\{x_0^{lw}, x_1^{lw}\}$  through weak augmentation. For the unlabeled dataset, the samples  $\{x_0^{uw}, x_1^{uw}\}$  and  $\{x_0^{us}, x_1^{us}\}$  are obtained after weak and strong augmentation, respectively.
- 2: **Model prediction and pseudo-label generation:** For the inputs  $\{x_0^{lw}, x_1^{lw}\}$ ,  $\{x_0^{uw}, x_1^{uw}\}$  and  $\{x_0^{us}, x_1^{us}\}$ , generate the model prediction results  $p^l(i, j)$ ,  $p^{uw}(i, j)$  and  $p^{us}(i, j)$ , respectively, using Equations (1)–(5). Then, generate the pseudo-label  $\hat{y}^{uw}(i, j)$  for  $p^{uw}(i, j)$  using Equation (8).
- 3: **Compute the supervised loss:** Calculate the supervised loss  $L_s$  using the label and model prediction  $p^l(i, j)$ , according to Equation (6).
- 4: **Adaptive threshold update:** First, calculate the prediction confidence  $\tau_c$  for the locality of the changed and unchanged classes in  $p_k^{uw}(i, j, c)$  using Equation (10). Then, resample the pixels to obtain  $S_c$  via Equations (11) and (12). Finally, update the global threshold  $\tau_t$  using Equation (13).
- 5: **Computing the confidence mask map:** Use the adaptive threshold  $\tau_t$  to calculate the confidence mask map  $\mathbf{m}_t(i, j)$  from the prediction result  $p^{uw}(i, j)$ , as described by Equation (14).
- 6: **Adaptive Class Weight:** Calculate the class weight  $\lambda_c$  adaptively based on the prediction results  $p^{uw}(i, j)$ , according to Equation (11) and (15).
- 7: **Compute the weight map:** Calculate the weight map  $\mathbf{w}_t(i, j)$  using  $p^{uw}(i, j)$  and  $\lambda_c$  by Equation (16).
- 8: **Compute unsupervised loss:** Calculate the unsupervised loss  $L_u$  based on  $\hat{y}^{uw}(i, j)$ ,  $p^{us}(i, j)$ ,  $\mathbf{m}_t(i, j)$ , and  $\mathbf{w}_t(i, j)$ , using Equation (17).
- 9: **Update model parameters:** Optimize the backbone network  $E-D$  using both the supervised loss  $L_s$  and the unsupervised loss  $L_u$ .

**Output:**

Optimized backbone network  $E-D$ .

## 4. Experimental Section

This section begins by describing the dataset, evaluation method, and experimental specifics. Next, ablation experiments are conducted on CLDA, AT, and ACW parameters. Finally, a comparison is made with semi-supervised change detection methods from recent years.

### 4.1. Datasets

In this study, we selected two high-resolution remote sensing image datasets, WHU-CD and LEVIR-CD, to design and validate the model's performance. Both datasets are widely used in remote sensing change detection tasks and provide rich and challenging image data. These datasets not only feature diverse geographic scenes but also include images from different time points, making them particularly suitable for remote sensing change detection tasks. By using these two datasets, we were able to test the model's performance across different scenarios and change detection tasks, further improving the model's generalization ability.

#### 4.1.1. WHU-CD: [1]

The WHU-CD dataset, proposed by Ji et al., consists of two sets of high-resolution aerial images captured in 2012 and 2016 in Christchurch, New Zealand. The primary focus of these image pairs is the detection of building changes. The total size of the image pairs is  $32,507 \times 15,354$  pixels, with a pixel resolution of 0.075 m. To facilitate training, the images

were cropped into  $256 \times 256$  blocks with no overlap between blocks, resulting in a total of 7434 image pairs. The dataset was then split into training, validation, and test sets in an 8:1:1 ratio.

#### 4.1.2. LEVIR-CD: [52]

The LEVIR-CD dataset, proposed by Chen et al., consists of 637 pairs of high-resolution remote sensing images taken between 2002 and 2018 in Texas, USA, with a minimum time interval of 5 years and a maximum of 14 years. These image pairs contain significant information about building changes. Each image is  $1024 \times 1024$  pixels in size, with a resolution of 0.5 m. Similar to the WHU-CD dataset, the images were cropped into  $256 \times 256$  blocks with no overlap between blocks, resulting in a total of 10,192 image pairs. The dataset was then split into training, validation, and test sets in a 7:1:2 ratio.

Further, in model training, we divided the training set into labeled and unlabeled datasets with 1% (99%), 5% (95%), 10% (90%), and 20% (80%) proportions. The labeled training set selection process was random, and the remaining part was the unlabeled dataset. See Table 1 for detailed partitioning information.

**Table 1.** Labeled and unlabeled data distribution for WHU-CD and LEVER-CD.

Train Label Ratio	WHU-CD		LEVER-CD	
	Label	Unlabel	Label	Unlabel
1%	60	5889	72	7050
5%	298	5651	357	6765
10%	595	5354	713	6409
20%	1190	4759	1425	5697
<b>Val/Test Image Pairs</b>	744/745		1025/2049	

#### 4.2. Evaluation Metrics

Semi-supervised change detection can be viewed as a binary classification problem. Therefore, in this paper, in order to comprehensively evaluate the performance of the model, the five indicators of IoU, F1, Kappa, *TPR*, and *TNR* are used. Some of the formulas are explained in detail as follows:

$$P_e = \frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{(TP + TN + FP + FN)^2} \quad (19)$$

$$kappa = \frac{OA - P_e}{1 - P_e} \quad (20)$$

$$TPR = \frac{TP}{TP + FN} \quad (21)$$

$$TNR = \frac{TN}{FP + TN} \quad (22)$$

In the above formula, the basic computational units *TP*, *TN*, *FP*, and *FN* represent the changed pixels correctly predicted, unchanged pixels correctly predicted, unchanged pixels incorrectly predicted as changed pixels, and changed pixels incorrectly predicted as unchanged pixels, respectively. IoU can reflect the degree of overlap between the predicted region and the real region. Since the prediction difficulty of changed pixels is much higher than that of unchanged pixels in change detection, we further used the  $IoU^c$  of changed pixels to replace the global IoU. F1 evaluates the overall performance of our model. Kappa assesses the agreement between the model and actual observations. *TPR* evaluates the model's ability to predict changing pixels, known as sensitivity. *TNR* evaluates the ability

of the model to predict unchanged pixels, that is, specificity. The higher the above five average metrics, the better the performance of the model.

#### 4.3. Implementation Details

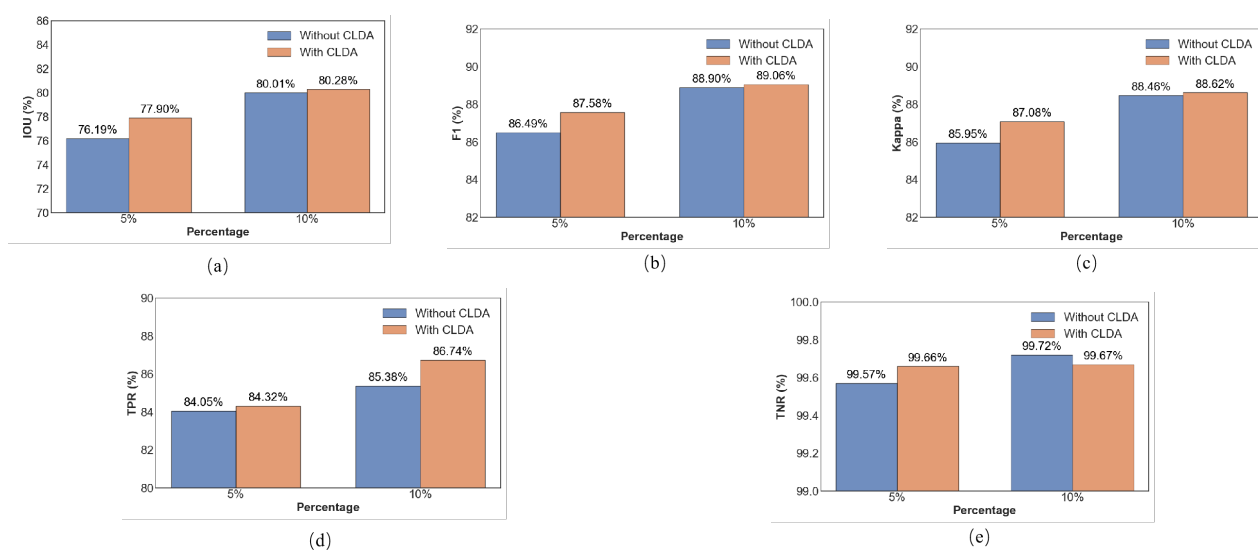
For a fair comparison, all models utilized ResNet50 as the pre-trained model, with each model trained for 80 epochs and a batch size of 4. The optimizer employed was stochastic gradient descent (SGD), consistent with the FPA [30] method, featuring an initial learning rate of 0.01, weight decay of 0.0001, and momentum of 0.9. To ensure an unbiased evaluation, the hyperparameter settings for the other comparison methods were retained as specified in their original publications. For data augmentation, weak augmentation included random flipping, random rescaling between 0.5 and 2.0, and random cropping, while the strongly enhanced region number was set to 8. The adaptive threshold started with an initial value of  $C$  set to 2 (yielding an initial threshold  $\tau_t$  of 0.5), and the momentum attenuation  $\lambda$  of the exponential moving average (EMA) was set to 0.9999. All experiments were carried out using the PyTorch framework with a GeForce RTX 3080Ti (NVIDIA, Santa Clara, CA, USA).

#### 4.4. Ablation Experiments

##### 4.4.1. CLDA Method Ablation Experiment

In semi-supervised change detection based on the consistency principle, weak augmentation preserves the basic structure of the image and slightly perturbs the model, which helps reduce the risk of overfitting. In contrast, strong augmentation alters the original features of the image to a greater extent and more aggressively disrupts the model, thereby enhancing its robustness. To validate the effectiveness of the CLDA method employed in semi-supervised change detection, we conducted ablation experiments using 5% and 10% data proportions of the WHU-CD dataset.

Our base model is the version with AT and ACW removed. We compared the performance of this base model with the model enhanced by CLDA. As shown in Figure 3, our model exhibits significant improvements across all metrics compared to the base model. Specifically, under the 5% data proportion, the IoU value increased by 1.71%, and the F1 score improved by 1.09%. With a 10% data proportion, the Kappa value rose from 88.46% to 88.62%, and the TPR value increased from 85.38% to 86.74%, demonstrating a substantial enhancement in our model's ability to detect changes. Additionally, for detecting unchanged pixels, our TNR metric consistently remained above 99.5%.



**Figure 3.** Effectiveness of channel-level data augmentation in our proposed method. (a) IoU, (change) (b) F1, (c) Kappa, (d) TPR, (e) TNR.

#### 4.4.2. Loss Function Ablation Experiment

During the training process with a large amount of unlabeled data, we introduce AT to provide the model with a substantial number of reliable pseudo-labels. This allows the model to be exposed to richer feature information, enhancing supervised training. Additionally, since the learning difficulty of detecting changes is significantly higher than that of unchanged areas, the introduction of ACW encourages the model to focus more on the feature information of the changed areas, thereby improving overall detection performance. To verify the effectiveness of these components, we conducted ablation experiments on the corresponding loss functions for AT and ACW. In these experiments, w/o  $L_u$  indicates the removal of unsupervised loss, w/o  $\mathbf{m}_t \mathbf{w}_t$  refers to the removal of both AT and ACW, leaving only the basic unsupervised loss, w/o  $\mathbf{w}_t$  stands for the removal of ACW while retaining AT, and w/o  $\mathbf{m}_t$  stands for the removal of AT while retaining ACW.

As shown in Table 2, removing the unsupervised loss significantly impacted the model, especially at lower data proportions. The use of AT (i.e., the “w/o  $\mathbf{w}_t$ ” method) alone improved performance notably, with the IoU increasing from 77.90% to 78.79% at the 5% data scale. Furthermore, the introduction of ACW resulted in an additional IoU improvement of 0.17% under the 5% data proportion. When using ACW (i.e., the “w/o  $\mathbf{m}_t$ ” method) alone, IOU increased by 0.84% at 5% data scale and decreased by 0.77% at 10% data scale. Although some metrics slightly decreased at the 10% data ratio, overall, our proposed method demonstrated strong detection results at both the 5% and 10% data ratios.

**Table 2.** Ablation experiment results of DA-AT on 5% and 10% labeled WHU-CD datasets. The best scores are marked in bold font, and the second scores are underlined.

Method	5%					10%				
	IoU <sup>c</sup>	F1	Kappa	TPR	TNR	IoU <sup>c</sup>	F1	Kappa	TPR	TNR
w/o $L_u$	66.29	79.73	79.00	71.31	99.69	70.67	82.82	82.12	81.95	99.34
w/o $\mathbf{m}_t \mathbf{w}_t$	77.90	87.58	87.08	<b>84.32</b>	99.66	80.28	89.06	88.62	<b>86.74</b>	99.67
w/o $\mathbf{w}_t$	<u>78.79</u>	<u>88.14</u>	<u>87.67</u>	84.19	<u>99.72</u>	<b>80.83</b>	<b>89.40</b>	<b>88.98</b>	85.75	<b>99.75</b>
w/o $\mathbf{m}_t$	78.74	88.10	87.64	<u>84.21</u>	99.71	79.51	88.59	88.13	85.70	99.68
Proposed	<b>78.96</b>	<b>88.24</b>	<b>87.79</b>	83.11	<b>99.78</b>	<u>80.72</u>	<u>89.33</u>	<u>88.91</u>	<u>85.94</u>	<u>99.73</u>

As shown in Table 3, removing the unsupervised loss ( $L_u$ ) significantly negatively impacted the model’s performance, especially at the lower data proportion (5%). When only the AT was used while removing the ACW (i.e., the “w/o  $\mathbf{w}_t$ ” method), the IoU reached 74.33%, and the F1 score was 85.52% at the 5% data proportion, both of which were the best results. At the 10% data proportion, the IoU was 76.35%, and the F1 score was 86.59%, also showing excellent performance. This demonstrates that the AT significantly enhances the model’s performance on the LEVIR-CD dataset, especially under lower data proportions. When only the ACW was introduced while removing the AT (i.e., the “w/o  $\mathbf{m}_t$ ” method), the IoU was 73.60%, and the F1 score was 85.14% at the 5% data proportion, slightly lower than the method using only AT (74.33% and 85.52%). At the 10% data proportion, the IoU was 76.59%, and the F1 score was 86.74%, both of which were the second-best results. This indicates that while ACW contributes to performance enhancement on the LEVIR-CD dataset, its effect is somewhat inferior to AT, particularly at higher data proportions. The complete model (Proposed), which retained both AT and ACW, achieved an IoU of 74.16% and an F1 score of 85.16% at the 5% data proportion, both of which were the second-best results. At the 10% data proportion, the IoU was 76.73%, and the F1 score was 86.83%, both of which were the best results. Additionally, other metrics such as Kappa, TPR, and TNR also showed that the complete model performs excellently in most cases, especially at the 10% data proportion, where Kappa and TPR reach their highest values.

**Table 3.** Ablation experiment results of DA-AT on 5% and 10% labeled LEVIR-CD datasets. The best scores are marked in bold font, and the second scores are underlined.

Method	5%					10%				
	IoU <sup>c</sup>	F1	Kappa	TPR	TNR	IoU <sup>c</sup>	F1	Kappa	TPR	TNR
w/o $L_u$	63.84	77.93	76.91	68.97	<u>99.57</u>	71.40	83.32	82.49	77.53	99.54
w/o $\mathbf{m}_t$ $\mathbf{w}_t$	73.41	84.53	83.89	78.52	99.50	75.32	85.87	85.39	81.16	<u>99.62</u>
w/o $\mathbf{w}_t$	<b>74.33</b>	<b>85.52</b>	<b>84.79</b>	<u>78.89</u>	99.56	76.35	86.59	85.91	81.59	<b>99.63</b>
w/o $\mathbf{m}_t$	73.60	85.14	84.13	<u>78.62</u>	99.54	<u>76.59</u>	<u>86.74</u>	<u>86.07</u>	<u>82.15</u>	99.61
Proposed	<u>74.16</u>	<u>85.16</u>	<u>84.43</u>	<b>79.64</b>	<b>99.60</b>	<b>76.73</b>	<b>86.83</b>	<b>86.17</b>	<b>82.65</b>	99.59

#### 4.4.3. Effect of Threshold EMA Decay $\lambda$

Table 4 demonstrates the robustness of our model under varying thresholding EMA decay rates ( $\lambda$ ) on the 5% labeled WHU-CD dataset. The IoU remains stable across different  $\lambda$  values, with a maximum variation of only 0.14. Similarly, the F1, Kappa, and TPR metrics exhibit minimal fluctuations, indicating that the model's performance is insensitive to changes in  $\lambda$ . These results validate the reliability of our threshold updating mechanism, demonstrating its consistent strong performance without the need for fine-tuning the decay rate.

**Table 4.** Error metrics under different EMA decay rates  $\lambda$  on 5% labeled WHU-CD. The differences in IoU<sup>c</sup> from the baseline (0.9999, IoU = 78.96) are shown in parentheses.

$\lambda$	IoU <sup>c</sup>	F1	Kappa	TPR	TNR
0.99	78.98 (+0.02)	88.25	87.79	84.59	99.71
0.999	79.03 (+0.07)	88.29	87.82	85.41	99.67
0.9999	78.96	88.24	87.79	83.11	99.78
0.99999	78.89 (−0.07)	88.20	87.73	85.81	99.64

#### 4.5. Comparison Experiments

To verify the effectiveness of our method in enhancing change detection performance for semi-supervised high-resolution remote sensing images, we compared it with several recent state-of-the-art methods, including S4GAN [53], SemiCDNet [23], SemiCD [26], RCL [27], and FPA [30]. Additionally, we included comparisons with the baseline models: “Only-sup”, which uses only the corresponding proportion of labeled data, and “Fully-sup”, which uses 100% of the labeled data.

S4GAN is a GAN-based semi-supervised semantic segmentation method that employs a two-branch structure to mitigate low-level and high-level artifacts commonly encountered when training with limited labeled data. SemiCDNet is based on generative adversarial networks. SemiCD, RCL, and FPA are based on consistency regularization and pseudo-label screening techniques. These comparisons aim to demonstrate the robustness and superior performance of our proposed method for handling semi-supervised change detection tasks.

##### 4.5.1. WHU-CD

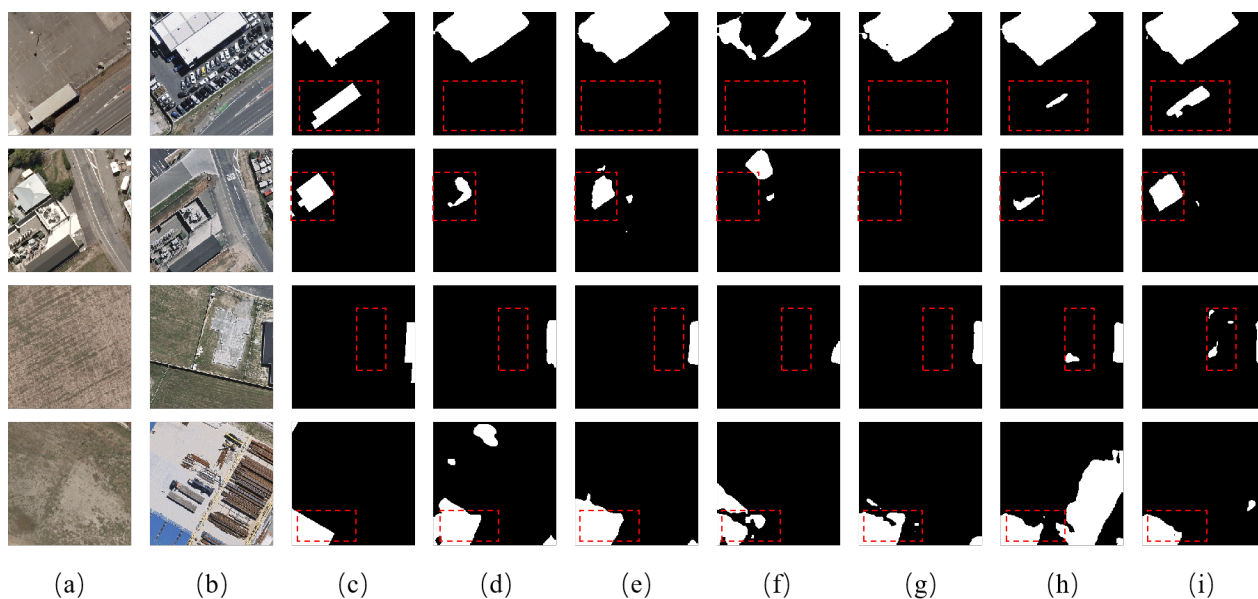
The WHU-CD dataset provides extensive information on changes in building objects. As shown in Table 5, our proposed method consistently achieves either the best or second-best performance across all evaluation metrics on the four data scales. Notably, our method's IoU is 9.53% higher than that of the next best model, FPA, when using just 5% of the labeled data. We attribute this improvement to the fact that the fixed threshold in the FPA method, under small data ratios, results in the loss of learning opportunities for a large number of samples early in training, a problem our AT approach effectively mitigates. Moreover, when using 20% of the labeled data, our method's performance nearly

matches that of using 100% labeled data, significantly reducing the cost of data annotation. Although our model's TPR index is slightly lower than that of the FPA method at the 10% data ratio, our F1 score is 0.38% higher, demonstrating superior overall performance.

**Table 5.** Comparison experiment results on the WHU dataset. The best scores are marked in bold font, and the second scores are underlined.

Method	1%					5%				
	IoU <sup>c</sup>	F1	Kappa	TPR	TNR	IoU <sup>c</sup>	F1	Kappa	TPR	TNR
Only-sup	40.17	57.31	55.64	54.78	98.50	66.29	79.73	79.00	71.31	99.69
S4GAN	53.81	69.97	68.81	65.67	99.09	70.16	82.47	81.81	75.21	99.70
SemiCDNet	41.65	58.81	57.46	48.90	<u>99.28</u>	71.97	83.70	83.06	79.25	99.58
SemiCD	55.60	71.46	70.32	69.31	98.98	73.68	84.97	84.39	79.18	99.70
RCL	49.33	66.07	65.01	53.39	<b>99.66</b>	73.28	84.58	83.99	78.84	99.69
FPA	<u>59.34</u>	<u>74.48</u>	<u>73.44</u>	<u>73.62</u>	99.01	<u>77.28</u>	<u>87.18</u>	<u>86.68</u>	<u>82.79</u>	<u>99.71</u>
<b>Ours</b>	<b>68.57</b>	<b>81.36</b>	<b>80.59</b>	<b>81.23</b>	99.24	<b>78.96</b>	<b>88.24</b>	<b>87.79</b>	<b>83.11</b>	<b>99.78</b>
Method	10%					20%				
	IoU <sup>c</sup>	F1	Kappa	TPR	TNR	IoU <sup>c</sup>	F1	Kappa	TPR	TNR
Super.only	70.67	82.82	82.12	81.95	99.34	74.45	85.36	84.76	83.62	99.49
S4GAN	74.48	85.37	84.78	83.99	99.47	78.63	88.03	87.56	85.13	99.66
SemiCDNet	74.92	85.66	85.08	84.43	99.48	78.88	88.19	87.72	86.17	99.62
SemiCD	78.97	88.25	87.78	85.30	99.67	79.44	88.54	88.07	87.21	99.60
RCL	78.35	87.86	87.38	83.97	<u>99.70</u>	78.13	87.72	87.22	86.03	99.58
FPA	<u>80.10</u>	<u>88.95</u>	<u>88.50</u>	<b>86.87</b>	99.65	<u>81.71</u>	<u>89.93</u>	<u>89.53</u>	<b>87.38</b>	<u>99.71</u>
<b>Ours</b>	<b>80.72</b>	<b>89.33</b>	<b>88.91</b>	<u>85.94</u>	<b>99.73</b>	<b>82.04</b>	<b>90.13</b>	<b>89.74</b>	<u>87.31</u>	<b>99.73</b>
Fully-sup	IoU <sup>c</sup> = 85.16, F1 = 91.98, Kappa = 91.66, TPR = 91.12, TNR = 99.71									

Figure 4 illustrates the visualization results when the proportion of labeled samples in the WHU-CD dataset is 5%. As shown in the first two rows of images in Figure 4, our model outperforms other methods in detecting small change regions, providing more complete detection of these areas. In contrast, other methods struggle to accurately capture real changes in complex scenes. Notably, the SemiCD method suffers from significant omissions in the detected change regions across the two image groups. This issue arises because the edges of the changed regions typically exhibit rapid variations in pixel intensity. During training, a fixed high threshold often results in the loss of crucial edge information, leading to incomplete detection of these regions during inference. Our AT method addresses this by incrementally adjusting the threshold based on the average threshold of the changed and unchanged parts of the image. This approach allows the model to retain lower threshold information at the edges, making it easier to learn detailed boundary information and, consequently, produce more complete change regions during inference. From the third row of Figure 4, we can see that although our model performs reasonably well in maintaining completeness, there remain a few false positives. Given that the adaptive threshold may be relatively low in the early stages, there is still room for improvement in the pseudo-label filtering process. Therefore, going forward, we will explore the following strategy: if a pixel (or a small group of pixels) is significantly distant from other similarly labeled change pixels, we will consider it “isolated” and remove it as early as possible, thereby mitigating the cumulative impact of false positives on the model.



**Figure 4.** Detection results of different algorithms on the WHU-CD dataset: (a) im0 (b) im1 (c) ground truth (d) S4GAN (e) SemiCDNet (f) SemiCD (g) RCL (h) FPA (i) Ours. The red dashed area highlights the significant differences between methods.

The last row of Figure 4 highlights that other methods tend to produce higher false detection rates, particularly due to the influence of irrelevant changes. In contrast, our model excels in accurately detecting unchanged parts. This improvement is especially significant in building datasets, where areas like roads and parking lots, which share structural and color similarities with buildings, often lead to high false detection rates. The introduction of CLDA into our model plays a crucial role in this enhancement. By applying data perturbation at the shallow layer of the image and random quantization at the channel layer, CLDA significantly increases image diversity at a deeper level. This strong augmentation allows the model to learn the correct change regions more effectively, thereby improving the overall performance of change detection.

#### 4.5.2. LEVIR-CD

The LEVIR-CD dataset, while rich in building turnover information similar to the WHU-CD dataset, has a resolution that is an order of magnitude lower. As shown in Table 6, our method achieves the second-best performance at the 10% data proportion, with an IoU just 0.26% lower than the top-performing FPA. It ranks third at the 5% and 20% data proportions, trailing the optimal SemiCD by 0.67% and 1.06%, respectively. We attribute this to the relatively low resolution of the LEVIR-CD dataset, which may hinder the model's ability to learn detailed feature information, thereby affecting detection accuracy. However, when the proportion of labeled samples is reduced to 1%, our method outperforms others, achieving the highest IoU, F1, and Kappa scores, with improvements of 0.86%, 0.65%, and 0.71% over the SemiCD method, respectively. Moreover, across the 1%, 5%, 10%, and 20% data proportions, our method consistently surpasses the RCL method in IoU by 5.35%, 0.51%, 0.92%, and 0.85%, respectively.

Figure 5 provides a visual comparison of the detection results on the LEVIR-CD dataset when the proportion of labeled samples is 1%. In the first two rows of images, our method demonstrates a lower missed detection rate than S4GAN and SemiCDNet, and it more effectively delineates the boundary between changed and unchanged regions compared to FPA. In the fourth row, despite the challenges posed by changes in building color, our method achieves a lower false detection rate than other methods. Although the third row depicts a scenario where the change region is irregular, and none of the methods perform

ideally, our approach, along with SemiCD and RCL, manages to separate the four change parts effectively. Overall, the visual results indicate that our model provides superior detection performance.

**Table 6.** Comparison experiment results on the LEVIR dataset. The best scores are marked in bold font, and the second scores are underlined.

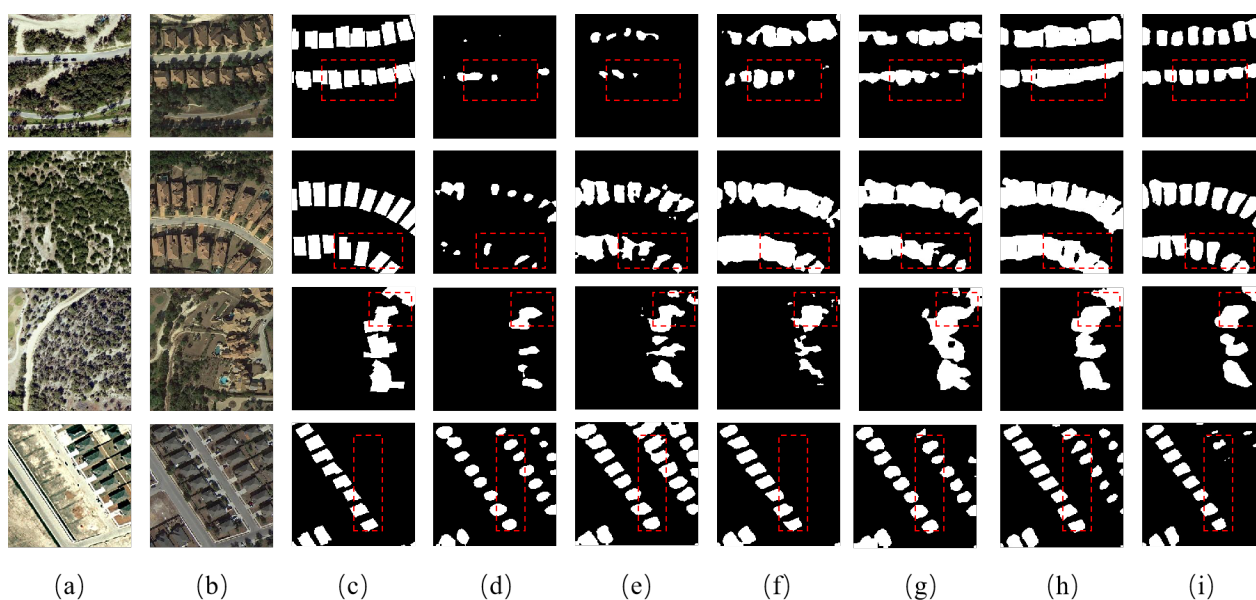
Method	1%					5%				
	IoU <sup>c</sup>	F1	Kappa	TPR	TNR	IoU <sup>c</sup>	F1	Kappa	TPR	TNR
Only-sup	44.23	61.34	59.30	60.17	98.07	63.84	77.93	76.91	68.97	99.57
S4GAN	39.06	56.18	54.36	46.15	99.03	68.87	81.57	80.67	74.89	99.53
SemiCDNet	48.35	65.18	63.28	66.23	98.01	69.80	82.21	81.32	77.00	99.45
SemiCD	<u>61.61</u>	<u>76.25</u>	<u>75.08</u>	<u>70.65</u>	<u>99.21</u>	<b>74.83</b>	<b>85.61</b>	<b>84.87</b>	<u>81.59</u>	99.52
RCL	57.12	72.71	71.31	69.84	98.80	73.65	84.83	84.03	<b>82.72</b>	99.34
FPA	61.59	76.23	74.99	<b>74.42</b>	98.88	<u>74.55</u>	<u>85.42</u>	<u>84.70</u>	79.05	<b>99.68</b>
<b>Ours</b>	<b>62.47</b>	<b>76.90</b>	<b>75.79</b>	69.79	<b>99.37</b>	74.16	85.16	84.43	79.64	<u>99.60</u>

Method	10%					20%				
	IoU <sup>c</sup>	F1	Kappa	TPR	TNR	IoU <sup>c</sup>	F1	Kappa	TPR	TNR
Super.only	71.40	83.32	82.49	77.53	99.54	75.35	85.94	85.22	82.12	99.52
S4GAN	73.90	84.99	84.26	78.12	<b>99.69</b>	75.36	85.95	85.25	80.67	99.62
SemiCDNet	74.64	85.48	84.76	79.80	99.63	76.39	86.62	85.95	81.44	<b>99.65</b>
SemiCD	76.60	86.75	86.08	<u>82.79</u>	99.57	<b>78.60</b>	<b>88.02</b>	<b>87.41</b>	83.89	<u>99.64</u>
RCL	75.81	86.24	85.52	<b>83.64</b>	99.45	76.69	86.81	86.11	<b>85.69</b>	99.37
FPA	<b>76.99</b>	<b>87.00</b>	<b>86.35</b>	81.77	<u>99.67</u>	<u>78.39</u>	<u>87.88</u>	<u>87.25</u>	<u>85.61</u>	99.51
<b>Ours</b>	<u>76.73</u>	<u>86.83</u>	<u>86.17</u>	82.65	99.59	77.54	87.35	86.71	83.01	99.62

Fully-sup	IoU <sup>c</sup> = 79.53, F1 = 88.60, Kappa = 88.01, TPR = 85.17, TNR = 99.62									
-----------	---	--	--	--	--	--	--	--	--	--



**Figure 5.** Detection results of different algorithms on the LEVIR-CD dataset: (a) im0 (b) im1 (c) ground truth (d) S4GAN (e) SemiCDNet (f) SemiCD (g) RCL (h) FPA (i) Ours. The red dashed area highlights the significant differences between methods.

#### 4.5.3. Semi-Supervised Versus Only-Sup

To provide a more intuitive comparison of our method's effectiveness on the two datasets, we evaluated the detection performance across different label proportions—1%, 5%, 10%, and 20%—using the only-sup method, our proposed method, and the fully-sup

method. The results clearly show that semi-supervised methods, including ours, consistently outperform the only-sup method across both the WHU-CD and LEVIR-CD datasets.

On the WHU-CD dataset, although the 1% label ratio results in a higher missed detection rate, as the label ratio increases, our method detects more complete change regions, and its performance increasingly approaches that of the fully-sup method. On the LEVIR-CD dataset, while the false detection rate of irrelevant changes is higher, particularly noticeable at the 1% label ratio, our method shows significant improvement at the 20% label ratio. At this ratio, the false detection rate we marked is nearly as low as that of the fully-sup method. Overall, our method demonstrates low false detection and missed detection rates on both datasets, affirming its robustness and effectiveness in various scenarios.

#### 4.5.4. Semi-Supervised Versus Fully Supervised Methods: Exploring Performance Differences

To further validate the performance of our proposed semi-supervised method, we conducted experiments on the WHU-CD and LEVIR-CD datasets, comparing its performance with only 30% labeled data to that of fully supervised methods, including BIFA [54] and the diffusion model DDPM-CD [55]. BIFA is a lightweight model that aligns bi-temporal features across channels, spatial dimensions, and scales, effectively mitigating the impact of illumination changes and perspective differences on change detection. DDPM-CD leverages a pre-trained diffusion model as a feature extractor, combined with a lightweight change classifier, effectively utilizing unlabeled data.

As shown in Table 7, the semi-supervised method demonstrates excellent performance on both the WHU-CD and LEVIR-CD datasets, approaching or even rivaling fully supervised methods in certain scenarios. On the WHU-CD dataset, the IoU and F1 scores of the semi-supervised method are 84.94% and 91.86%, with only minor differences of 0.22% and 0.12% compared to the Fully-sup, indicating nearly identical performance. The IoU and F1 scores of the proposed semi-supervised method are 3.85% and 2.21% lower than those of BIFA, and 0.90% and 0.52% lower than those of DDPM-CD, respectively. These results highlight the efficiency of our semi-supervised approach even with reduced labeled data. On the LEVIR-CD dataset, while the semi-supervised method performs slightly below BIFA and DDPM-CD, the differences remain relatively small, with IoU and F1 gaps of 4.73% and 2.91% compared to BIFA, and 4.50% and 2.74% compared to DDPM-CD, showcasing its strong adaptability in low-labeled data scenarios.

**Table 7.** Performance Comparison Between Semi-Supervised and Fully Supervised Methods on WHU-CD and LEVIR-CD Datasets

Dataset	Method	IoU <sup>c</sup>	F1	Kappa	TPR	TNR
WHU-CD	30%	84.94	91.86	91.52	91.83	99.67
	BIFA	88.79 (+3.85)	94.07 (+2.21)	94.14 (+2.62)	93.60 (+1.77)	99.74 (+0.07)
	DDPM-CD	85.84 (+0.90)	92.38 (+0.52)	92.19 (+0.67)	92.05 (+0.22)	99.73 (+0.06)
	Fully-sup	85.16 (+0.22)	91.98 (+0.12)	91.66 (+0.14)	91.12 (−0.71)	99.71 (+0.04)
LEVIR-CD	30%	78.23	87.78	87.16	84.14	99.59
	BIFA	82.96 (+4.73)	90.69 (+2.91)	90.19 (+3.03)	89.86 (+5.72)	99.53 (−0.06)
	DDPM-CD	82.73 (+4.50)	90.52 (+2.74)	89.81 (+2.65)	90.39 (+6.25)	99.55 (−0.04)
	Fully-sup	79.53 (+1.30)	88.60 (+0.82)	88.01 (+0.85)	85.17 (+1.03)	99.62 (+0.03)

Overall, the semi-supervised method excels on the WHU-CD dataset and maintains stable performance on the LEVIR-CD dataset, demonstrating its ability to achieve results comparable to fully supervised methods while significantly reducing labeling costs. This highlights its significant research value and practical application potential.

#### 4.6. Training Complexity Analysis

To comprehensively compare the advantages and disadvantages of our method with other change detection methods, we evaluated three aspects: the number of parameters, training time, and IoU (Intersection over Union). The comparison was performed on the 5% labeled WHU-CD dataset, as shown in Table 8. In terms of the number of parameters, our method, like Only-sup, S4GAN, RCL, FPA, and Fully-sup, maintains the smallest parameter count of 46.85M. Regarding the training time per epoch, although our method lags behind Only-sup and Fully-sup in terms of training time, the performance of Only-sup is suboptimal. While Fully-sup achieves the best performance, its high labeling cost makes it impractical for real-world applications. In contrast, our method requires less training time than SemiCD, RCL, and FPA. Although it has a slightly higher training time than S4GAN and SemiCDNet, it shows a 7–8% improvement in IoU over these two methods. Overall, our method achieves a good balance between performance, parameter efficiency, and training time, demonstrating an excellent trade-off between accuracy and efficiency.

**Table 8.** Comparison of parameters and training time of different CD methods on 5% labeled WHU-CD.

Method	Params (M)	Training Time/Epoch (min)	IoU <sup>c</sup>
Only-sup	46.85	0.75	66.29
S4GAN	46.85	6.9	70.16
SemiCDNet	52.37	6.5	71.97
SemiCD	50.69	9.1	73.68
RCL	46.85	8.5	73.28
FPA	46.85	7.8	77.28
Fully-sup	46.85	3.7	85.16
Our	46.85	7.5	78.96

## 5. Conclusions

In this paper, a DA-AT change detection method is proposed. By introducing channel-level data augmentation into strong augmentation, the problem of the insufficient feature expression ability of current data augmentation only at the image level is addressed. The adaptive threshold method is proposed to balance the quantity and quality of pseudo-labels and optimize the diversity of data learned by the model. The adaptive class weights are designed to improve the learning ability of the model for changing classes. Comprehensive experimental results show that the proposed algorithm is effective and robust on the WHU-CD and LEVIR-CD high-resolution remote sensing image datasets. The method has not yet been tested on other remote sensing data types, such as multispectral, hyperspectral, or Synthetic Aperture Radar (SAR) imagery, which present challenges due to differences in spectral information, resolution, and noise. We plan to combine diffusion models with semi-supervised learning to improve performance.

**Author Contributions:** Conceptualization, W.Z. and X.S.; methodology, X.S. and S.W.; software, S.D.; validation, W.Z. and X.S.; data curation, S.W.; writing—original draft preparation, X.S.; writing—review and editing, W.Z. and X.S.; supervision, S.D.; project administration, S.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by National Natural Science Foundation of China General Program 62471389, in part by the Shaanxi Provincial Key Research and Develop Programme General Project under Grant 2024SF-YBXM-572, in part by the National Science Fund for Distinguished Young

Scholars under Grant 61925112, in part by the Shaanxi Provincial Key Research and Development Program 2023-YBSF-455.

**Data Availability Statement:** No new data were created in this manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Ji, S.; Wei, S.; Lu, M. Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 574–586. [\[CrossRef\]](#)
2. Hussain, M.; Chen, D.; Cheng, A.; Wei, H.; Stanley, D. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS J. Photogramm. Remote Sens.* **2013**, *80*, 91–106. [\[CrossRef\]](#)
3. Lei, T.; Zhang, Y.; Lv, Z.; Li, S.; Liu, S.; Nandi, A.K. Landslide Inventory Mapping From Bitemporal Images Using Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 982–986. [\[CrossRef\]](#)
4. Desclée, B.; Bogaert, P.; Defourny, P. Forest change detection by statistical object-based method. *Remote Sens. Environ.* **2006**, *102*, 1–11. [\[CrossRef\]](#)
5. Jianya, G.; Haigang, S.; Guorui, M.; Qiming, Z. A review of multi-temporal remote sensing data change detection algorithms. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2008**, *37*, 757–762.
6. Jiang, H.; Peng, M.; Zhong, Y.; Xie, H.; Hao, Z.; Lin, J.; Ma, X.; Hu, X. A survey on deep learning-based change detection from high-resolution remote sensing images. *Remote Sens.* **2022**, *14*, 1552. [\[CrossRef\]](#)
7. Cheng, G.; Huang, Y.; Li, X.; Lyu, S.; Xu, Z.; Zhao, H.; Zhao, Q.; Xiang, S. Change detection methods for remote sensing in the last decade: A comprehensive review. *Remote Sens.* **2024**, *16*, 2355. [\[CrossRef\]](#)
8. Li, X.; He, H.; Li, X.; Li, D.; Cheng, G.; Shi, J.; Weng, L.; Tong, Y.; Lin, Z. Pointflow: Flowing semantics through points for aerial image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 4217–4226.
9. Zheng, Z.; Zhong, Y.; Wang, J.; Ma, A. Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 4096–4105.
10. Caye Daudt, R.; Le Saux, B.; Boulch, A. Fully Convolutional Siamese Networks for Change Detection. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 4063–4067. [\[CrossRef\]](#)
11. Chen, H.; Qi, Z.; Shi, Z. Remote Sensing Image Change Detection with Transformers. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5607514. [\[CrossRef\]](#)
12. Fang, S.; Li, K.; Shao, J.; Li, Z. SNUNet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 8007805. [\[CrossRef\]](#)
13. Bandara, W.G.C.; Patel, V.M. A transformer-based siamese network for change detection. In Proceedings of the IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 207–210.
14. Wu, C.; Du, B.; Zhang, L. Fully convolutional change detection framework with generative adversarial network for unsupervised, weakly supervised and regional supervised change detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 9774–9788. [\[CrossRef\]](#)
15. Hou, B.; Wang, Y.; Liu, Q. Change detection based on deep features and low rank. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2418–2422. [\[CrossRef\]](#)
16. Saha, S.; Bovolo, F.; Bruzzone, L. Unsupervised deep change vector analysis for multiple-change detection in VHR images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3677–3693. [\[CrossRef\]](#)
17. Gong, M.; Yang, Y.; Zhan, T.; Niu, X.; Li, S. A generative discriminatory classified network for change detection in multispectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 321–333. [\[CrossRef\]](#)
18. Wu, C.; Chen, H.; Du, B.; Zhang, L. Unsupervised change detection in multitemporal VHR images based on deep kernel PCA convolutional mapping network. *IEEE Trans. Cybern.* **2021**, *52*, 12084–12098. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Huang, R.; Wang, R.; Guo, Q.; Wei, J.; Zhang, Y.; Fan, W.; Liu, Y. Background-mixed augmentation for weakly supervised change detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 7919–7927.
20. Dai, Y.; Zhao, K.; Shen, L.; Liu, S.; Yan, X.; Li, Z. A Siamese Network Combining Multiscale Joint Supervision and Improved Consistency Regularization for Weakly Supervised Building Change Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 4963–4982. [\[CrossRef\]](#)
21. Shafique, A.; Cao, G.; Khan, Z.; Asad, M.; Aslam, M. Deep learning-based change detection in remote sensing images: A review. *Remote Sens.* **2022**, *14*, 871. [\[CrossRef\]](#)

22. Liu, J.; Chen, K.; Xu, G.; Li, H.; Yan, M.; Diao, W.; Sun, X. Semi-supervised change detection based on graphs with generative adversarial networks. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 74–77.
23. Peng, D.; Bruzzone, L.; Zhang, Y.; Guan, H.; Ding, H.; Huang, X. SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5891–5906. [\[CrossRef\]](#)
24. Zheng, Z.; Ermon, S.; Kim, D.; Zhang, L.; Zhong, Y. Changen2: Multi-temporal remote sensing generative change foundation model. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**. [\[CrossRef\]](#)
25. Wang, J.; Wang, Y.; Chen, B.; Liu, H. LCS-EnsemNet: A Semisupervised Deep Neural Network for SAR Image Change Detection With Dual Feature Extraction and Label-Consistent Self-Ensemble. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 11903–11925. [\[CrossRef\]](#)
26. Bandara, W.G.C.; Patel, V.M. Revisiting consistency regularization for semi-supervised change detection in remote sensing images. *arXiv* **2022**, arXiv:2204.08454.
27. Wang, J.X.; Li, T.; Chen, S.B.; Tang, J.; Luo, B.; Wilson, R.C. Reliable Contrastive Learning for Semi-Supervised Change Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4416413. [\[CrossRef\]](#)
28. Zhang, X.; Huang, X.; Li, J. Joint Self-Training and Rebalanced Consistency Learning for Semi-Supervised Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5406613. [\[CrossRef\]](#)
29. Sun, C.; Wu, J.; Chen, H.; Du, C. SemiSANet: A semi-supervised high-resolution remote sensing image change detection model using Siamese networks with graph attention. *Remote Sens.* **2022**, *14*, 2801. [\[CrossRef\]](#)
30. Zhang, X.; Huang, X.; Li, J. Semisupervised change detection with feature-prediction alignment. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5401016. [\[CrossRef\]](#)
31. Sohn, K.; Berthelot, D.; Carlini, N.; Zhang, Z.; Zhang, H.; Raffel, C.A.; Cubuk, E.D.; Kurakin, A.; Li, C.L. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. In *Proceedings of the Advances in Neural Information Processing Systems, Online, 6–12 December 2020*; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 596–608.
32. Wang, Y.; Chen, H.; Heng, Q.; Hou, W.; Fan, Y.; Wu, Z.; Wang, J.; Savvides, M.; Shinozaki, T.; Raj, B.; et al. Freematch: Self-adaptive thresholding for semi-supervised learning. *arXiv* **2022**, arXiv:2205.07246.
33. Yang, X.; Song, Z.; King, I.; Xu, Z. A survey on deep semi-supervised learning. *IEEE Trans. Knowl. Data Eng.* **2022**, *35*, 8934–8954. [\[CrossRef\]](#)
34. Wang, L.; Sun, Y.; Wang, Z. CCS-GAN: A semi-supervised generative adversarial network for image classification. *Vis. Comput.* **2022**, *38*, 2009–2021. [\[CrossRef\]](#)
35. Hung, W.C.; Tsai, Y.H.; Liou, Y.T.; Lin, Y.Y.; Yang, M.H. Adversarial learning for semi-supervised semantic segmentation. *arXiv* **2018**, arXiv:1802.07934.
36. Yun, S.; Han, D.; Oh, S.J.; Chun, S.; Choe, J.; Yoo, Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6023–6032.
37. Li, Q.; Shi, Y.; Zhu, X.X. Semi-supervised building footprint generation with feature and output consistency training. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5623217.
38. Chen, X.; Yuan, Y.; Zeng, G.; Wang, J. Semi-supervised semantic segmentation with cross pseudo supervision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 2613–2622.
39. Feng, Z.; Zhou, Q.; Gu, Q.; Tan, X.; Cheng, G.; Lu, X.; Shi, J.; Ma, L. Dmt: Dynamic mutual training for semi-supervised learning. *Pattern Recognit.* **2022**, *130*, 108777. [\[CrossRef\]](#)
40. Yang, L.; Qi, L.; Feng, L.; Zhang, W.; Shi, Y. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 7236–7246.
41. Wang, J.; HQ Ding, C.; Chen, S.; He, C.; Luo, B. Semi-supervised remote sensing image semantic segmentation via consistency regularization and average update of pseudo-label. *Remote Sens.* **2020**, *12*, 3603. [\[CrossRef\]](#)
42. Chen, H.; Li, W.; Shi, Z. Adversarial instance augmentation for building change detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–16. [\[CrossRef\]](#)
43. Nie, W.; Gou, P.; Liu, Y.; Shrestha, B.; Zhou, T.; Xu, N.; Wang, P.; Du, Q. Semi supervised change detection method of remote sensing image. In Proceedings of the 2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Beijing, China, 3–5 October 2022; pp. 1013–1019.
44. Mao, Z.; Tong, X.; Luo, Z. Semi-supervised remote sensing image change detection using mean teacher model for constructing pseudo-labels. In Proceedings of the ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.

45. Bromley, J.; Guyon, I.; LeCun, Y.; Säckinger, E.; Shah, R. Signature verification using a “siamese” time delay neural network. In *Advances in Neural Information Processing Systems*; Morgan Kaufmann: Cambridge, MA, USA, 1993; Volume 6.
46. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
47. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
48. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.
49. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
50. Cubuk, E.D.; Zoph, B.; Shlens, J.; Le, Q.V. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Seattle, WA, USA, 14–19 June 2020; pp. 702–703.
51. Wu, H.; Lei, C.; Sun, X.; Wang, P.S.; Chen, Q.; Cheng, K.T.; Lin, S.; Wu, Z. Randomized Quantization: A Generic Augmentation for Data Agnostic Self-supervised Learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Paris, France 2–6 October 2023; pp. 16305–16316.
52. Chen, H.; Shi, Z. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sens.* **2020**, *12*, 1662. [[CrossRef](#)]
53. Mittal, S.; Tatarchenko, M.; Brox, T. Semi-supervised semantic segmentation with high-and low-level consistency. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 1369–1379. [[CrossRef](#)] [[PubMed](#)]
54. Zhang, H.; Chen, H.; Zhou, C.; Chen, K.; Liu, C.; Zou, Z.; Shi, Z. BiFA: Remote Sensing Image Change Detection with Bitemporal Feature Alignment. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5614317. [[CrossRef](#)]
55. Bandara, W.G.C.; Nair, N.G.; Patel, V.M. Ddpm-cd: Remote sensing change detection using denoising diffusion probabilistic models. *arXiv* **2022**, arXiv:2206.11892.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.