Semi-Supervised Change Detection With Fourier-Based Frequency Transformation

Ze Zhang, Xue Jiang, *Senior Member, IEEE*, Yue Zhou, and Xingzhao Liu, *Member, IEEE*

***Abstract*—Semisupervised change detection (CD) methods have garnered increasing attention due to their capacity to alleviate the dependency of fully-supervised methods on a large number of pixel-level labels. These methods predominantly leverage generative adversarial network architecture and consistency regularization technology. However, they encounter challenges associated with background noise from cross-temporal images. In this article, we propose a novel multilevel consistency-regularization-based semisupervised CD approach that incorporates Fourier-based frequency transformation and a reliable pseudolabel selection scheme. Specifically, we replace the low-frequency spectrum of one temporal image with a frequency domain transformation derived from the corresponding image in the same bitemporal remote sensing image pair, enhancing the model's capability to discern meaningful changes amidst background noise, thereby contributing to more robust CD. Furthermore, excessively high pseudolabel thresholds in consistency regularization methods may result in the underutilization of valuable unlabeled data. To address this issue, we design a straightforward sigmoid-like function to dynamically adjust the selection threshold for the reliable pseudolabel selection scheme. This strategy takes into consideration the learning status throughout the entire training process, ensuring more effective utilization of unlabeled information. We demonstrate significant performance improvements across three widely-used public datasets, namely, LEVIR-CD, WHU-CD, and CDD. Notably, on the three datasets with only 1% labeled data, our method achieved an IoU^c of 71.29%, 63.90%, and 51.00%, outperforming existing state-of-the-art methods by 2.84%, 1.21%, and 0.98%, respectively. These results robustly substantiate the effectiveness of our approach, showcasing its potential in scenarios where labeled data is limited.**

***Index Terms*—Change detection (CD), dynamic threshold, Fourier-based frequency transformation, remote sensing, semisupervised learning.**

I. Introduction

CHANGE detection (CD) plays a crucial role in remote sensing image (RSI) processing, which is the quantitative analysis of land use changes from RSIs captured at different time periods. Beyond its fundamental role, CD finds application in various real-world scenarios, such as disaster assessment [1], [2], [3], land cover recognition [4], [5], [6], and forest monitoring [7], [8], [9].

Over the past decades, the development of deep learning (DL) has led to significant advancements in the CD community. Many sophisticated CD methods [10], [11], [12], [13] leverage convolutional neural networks (CNNs) [14], [15] and Transformers [16], [17] due to their exceptional capabilities in extracting discriminative features from dual-temporal RSIs, demonstrating outstanding performance. Nevertheless, these methods typically necessitate a substantial collection of manually annotated change labels, which is both time-consuming and expensive. On the contrary, unsupervised CD methods [18], [19], [20] often employ generative models, such as generative adversarial networks (GANs) [21], to artificially generate a training dataset for model training within a supervised pipeline. However, the main drawback of the aforementioned unsupervised CD methods is that they tend to detect more unchanged areas, as no relevant information is provided to the model. Consequently, the practical application in real-world scenarios faces considerable challenges.

To tackle the above issues, researchers have increasingly turned to adopting semisupervised learning technologies [22], [23], especially some semisupervised semantic segmentation methods. Semisupervised methods typically require limited labeled data and exploit the potential of the remaining unlabeled data. The current state-of-the-art semisupervised CD methods predominantly rely on GAN architecture and consistency regularization technology. GAN-based methods [24], [25] employ a segmentation network to modify the generative network, generating initial predictions. However, these methods encounter difficulties in stabilizing the training process and achieving convergence in training loss. In contrast, consistency regularization-based methods [26], [27], [28], [29] harness the power of unlabeled data by subjecting them to diverse perturbations. By doing so, these methods aim to train a model capable of maintaining consistency and robustness in the face of such perturbations.

Despite the significant performance gains achieved by consistency regularization-based methods in semisupervised CD, several inherent challenges persist, limiting their practical effectiveness. Specifically, two main issues are identified.

1) *Cross-temporal background noise:* In the context of the CD task, a pair of bitemporal images serves as input, with the model aiming to predict the difference between them. However, the significant time gap between images

Manuscript received 28 April 2024; revised 31 May 2024; accepted 10 June 2024. Date of publication 14 June 2024; date of current version 1 July 2024. *(Corresponding author: Xue Jiang.)*

Ze Zhang, Xue Jiang, and Xingzhao Liu are with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zzlovexcc@sjtu.edu.cn; xuejiang@sjtu.edu.cn; xzhliu@sjtu.edu.cn).

Yue Zhou is with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China, and also with the S-Lab, Nanyang Technological University, Singapore 639798 (e-mail: yue.zhou@ntu.edu.sg).

Digital Object Identifier 10.1109/JSTARS.2024.3414452

captured in different temporal periods introduces variations in seasons, weather conditions, and time, leading to noticeable changes in color, brightness, and other attributes within the image pair. As a result, the distribution of land cover samples in these two images appears dissimilar and irregular in the chromatic space, regardless of whether changes have occurred or not. This noise poses a challenge as it can mislead the model by interpreting noise differences as changes, resulting in erroneous predictions.

2) *Underutilization of unlabeled data:* Methods utilizing consistency regularization face challenges with unreliable pseudolabels, especially in the initial stages of model training, a phenomenon known as confirmation bias [30]. Current semisupervised learning algorithms [22], [31] often employ a predefined threshold to generate pseudolabels exclusively for high-confidence unlabeled data. However, this approach may result in the underutilization of a significant amount of unlabeled data information, proving less beneficial for effective model training [32], particularly in binary classification prediction tasks [33].

To this end, we introduce a novel multilevel consistency regularization-based semisupervised CD method for RSIs. Our approach involves training a perturbation-invariant model by applying perturbations at various levels [34], [35], including the sample level and feature level. To effectively tackle the identified issues, we propose two innovative strategies: Fourier-based frequency transformation and a reliable pseudolabel selection scheme. Specifically, drawing inspiration from FDA [36], which suggests that low-level amplitude variations can occur without significantly impacting the perception of high-level semantics, we present a straightforward solution: replacing the low-frequency amplitude spectrum of one temporal image with that derived from the corresponding image in the same bitemporal RSI pair. Furthermore, we devise a straightforward sigmoid-like function to dynamically adjust threshold, actively integrating perplexing unlabeled data into model training, thus enabling the selection of more reliable pseudolabels.

In summary, the main contributions of this article are as follows.

1) *Innovative Fourier-based Frequency Transformation:* We introduce a novel approach by integrating the Fourier-based transformation strategy into the semisupervised framework as a form of data augmentation. Specifically, we selectively replace the low-frequency amplitude information of one image in a bitemporal RSI pair with the corresponding information from the other image. The key to this strategy lies in adjusting the background distribution to align the low-frequency information of the bitemporal images, effectively reducing the impact of noise on the CD process.
2) *Dynamic Thresholding for Reliable Pseudolabel Selection:* To optimize the effective utilization of unlabeled data, we propose a sigmoid-like function that dynamically adjusts the threshold for reliable pseudolabel selection. This adaptive mechanism considers the learning status throughout the training process, ensuring a more effective integration of unlabeled data.
3) Experimental results indicate that the proposed method can improve the model performance in comparison to several state-of-the-art methods on three widely used CD datasets. Notably, our method introduces these enhancements without necessitating the addition of extra training parameters.

II. Related Work

In this section, we will provide a concise overview of pertinent research in CD, semisupervised semantic segmentation, and semisupervised CD.

A. Change Detection

Traditional CD methods predominantly employ feature disparities and pixel pair ratios as input, identifying changes through threshold-based strategies. For instance, change vector analysis [37] performed differencing operations on data from different periods of images to obtain the variation of each pixel. Deng et al. [38] employed principal component analysis to enhance the change information derived from stacked multisensor data for the purpose of classification. Tang et al. [39] proposed an object-oriented CD method grounded in the Kolmogorov–Smirnov two-sample test.

In recent years, with the advances of DL, DL-based CD methods have witnessed continuous improvements. Daudt et al. [10] presented three archetypal CD architectures utilizing fully convolutional network [40]. In these architectures, Siamese networks are utilized to extract features from RSI pairs, which are then employed to predict the change maps through feature fusion. Fang et al. [11] introduced a pioneering general CD architecture, named MetaChanger, characterized by a set of alternative interaction layers integrated into the feature extractor. Recent works [41], [42], [43], [44], [45] also emphasized noise suppression and boundary optimization. For example, Wang et al. [44] introduced a two-branch multitask framework of CD and superpixels to alleviate edge blurring. In addition to CNN-based approaches, transformer-based methods have been extensively investigated owing to the potent global attention mechanism inherent in transformers. Chen et al. [12] introduced the bitemporal image transformer to proficiently model contextual information in the spatial-temporal domain. TransY-Net, introduced by Yan et al. [13], aimed to enhance feature extraction globally and integrate multilevel visual features in a pyramid structure. Besides, the evolution of diffusion models [46] has led to the introduction of generative methods into the realm of CD. A pretrained denoising diffusion probabilistic model on million off-the-shelf RSIs was leveraged to employ the multiscale feature representations in DDPM-CD [47].

While fully supervised CD methods have shown outstanding performance, their reliance on a large volume of ground-truth change labels imposes significant time and labor constraints, restricting their practical applicability in real-world settings. As an alternative approach, researchers have turned to semisupervised techniques to alleviate the demand for extensive labeled data. Notably, the application of semisupervised semantic segmentation methods to CD tasks has attracted considerable attention

due to its potential to enhance efficiency and diminish annotation requirements.

B. Semi-Supervised Semantic Segmentation

In contrast to the substantial manual annotation demands of fully-supervised methods, semisupervised semantic segmentation methods maximize the utilization of unlabeled data. Numerous semisupervised semantic segmentation algorithms have been proposed, encompassing GAN-like models, pseudolabeling methods, and consistency regularization methods, all of which have yielded notable results even with limited labeled data.

For example, GAN-like framework was initially utilized in [48] to address semisupervised scenarios. This framework was designed to extract knowledge from a substantial amount of unlabeled data and generate synthetic images for training. Mittal et al. [25] introduced s4GAN, which replaced the conventional generative network in classical GANs with a segmentation network. It also incorporated an additional processing branch wherein a classifier is trained to filter the obtained segmentation maps. Vu et al. [49] employed the entropy-based method known as AdvEnt, incorporating entropy loss and adversarial loss for unsupervised domain adaptation in semantic segmentation. Pseudolabeling methods iteratively generate pseudolabels and apply self-training process. Yang et al. [50] incorporated strong data augmentation techniques into the self-training process to enhance the selection of more accurate pseudolabels. To address the distribution change induced by strong augmentation, distribution-specific batch normalization was proposed in [51]. Built upon the assumption of smoothness [52], consistency regularization methods excavate the potential of unlabeled data via incorporating diverse perturbations to them, and training a perturbation-invariant model. In this context, Chen et al. [53] enforced consistency on two segmentation networks that shared the same architecture but were initialized with different random weights for the same image. Yang et al. [54] proposed a dual-stream unified perturbations method that exploited both sample-level and feature-level perturbations to enhance model training.

C. Semisupervised CD

To alleviate high dependence on manual annotations, semisupervised CD methods pay more attention to the effective utilization of unlabeled data. Existing semisupervised CD methods primarily fall into three categories: metric learning-based, GAN architecture, and consistency regularization-based approaches. For instance, Bovolo et al. [55] introduced a novel semisupervised CD approach for multispectral RSI. This is achieved through the utilization of a binary semisupervised support vector machine (S^3VM) classifier accomplished by a selective Bayesian thresholding. Subsequently, Peng et al. [24] introduced SemiCDNet, a GAN-based method. In this approach, a segmentation network is employed to generate initial predictions, including entropy maps, and two discriminators are utilized to ensure the consistency of feature distribution between segmentation maps and entropy maps for both labeled and unlabeled data.

Recently, numerous researchers have proposed various semisupervised CD methods based on consistency regularization, achieving state-of-the-art results. Bandara and Patel [26] introduced RCR, where they devised an unsupervised consistency loss alongside the supervised cross-entropy loss [56]. This is achieved by imposing constraints on the consistency of the output change probability map for a given unlabeled bitemporal RSI pair under a few random perturbations applied to the feature difference map. Inspired by contrastive learning, Wang et al. [27] introduced the reliable contrastive learning method for semi-supervised CD. In this approach, they designed a novel contrastive loss specifically focused on changed areas. Furthermore, to address confirmation bias [30], they leveraged the uncertainty of unlabeled data to exclude unreliable pseudolabels. Zhang et al. [28] further introduced a progressive semisupervised CD method named feature-prediction alignment (FPA) with physical meanings. In this work, they integrated class-aware feature alignment (FA) and pixel-aware prediction alignment (PA) to diminish the feature difference within the same classes and reduce the prediction uncertainty under different image augmentations. Addressing the challenges of rotation nonequivariance and imbalanced distribution in semisupervised CD tasks, Zhang et al. [29] simultaneously applied consistency learning to rotated unlabeled data for reducing rotation inconsistency and employed uncertainty-based classwise weighting to alleviate imbalanced distribution.

III. Methodology

This section begins with an exploration of the preliminaries and the overarching architecture of the proposed method. Subsequently, we provide detailed explanations of two innovative strategies: Fourier-based frequency transformation and the reliable pseudolabel selection scheme. Finally, we unveil the comprehensive loss function. The overall workflow is illustrated in Fig. 1.

A. Preliminaries

For a comprehensive overview of our methodology, we begin by outlining the standard configurations employed in semisupervised CD. The training data consists of two main subsets: a labeled set and an unlabeled set. The labeled set is denoted as $D^l = \{(x_A^{l,i}, x_B^{l,i}, y^{l,i})\}_{i=1}^{N_l}$, where $x_A^{l,i}$ and $x_B^{l,i}$ represent the ith pair of bitemporal RSIs capturing the same geographical area with label $y^{l,i}$. Conversely, the unlabeled set is denoted as $D^u = \{(x_A^{u,i}, x_B^{u,i})\}_{i=1}^{N_u}$, where $x_A^{u,i}$ and $x_B^{u,i}$ represent the ith unlabeled image pair. Here, x_A and x_B denote the prechange image and the postchange image. N_l and N_u represent the corresponding number of samples.

B. Overall Architecture

As illustrated in Fig. 1, we adopt the widely used multiple-encoder and single-decoder (MESD) structure [10], [57], [58] as the architecture for our CD network. MESD comprises Siamese encoders with shared weights and a single decoder dedicated to CD. The multiple encoders are responsible for extracting

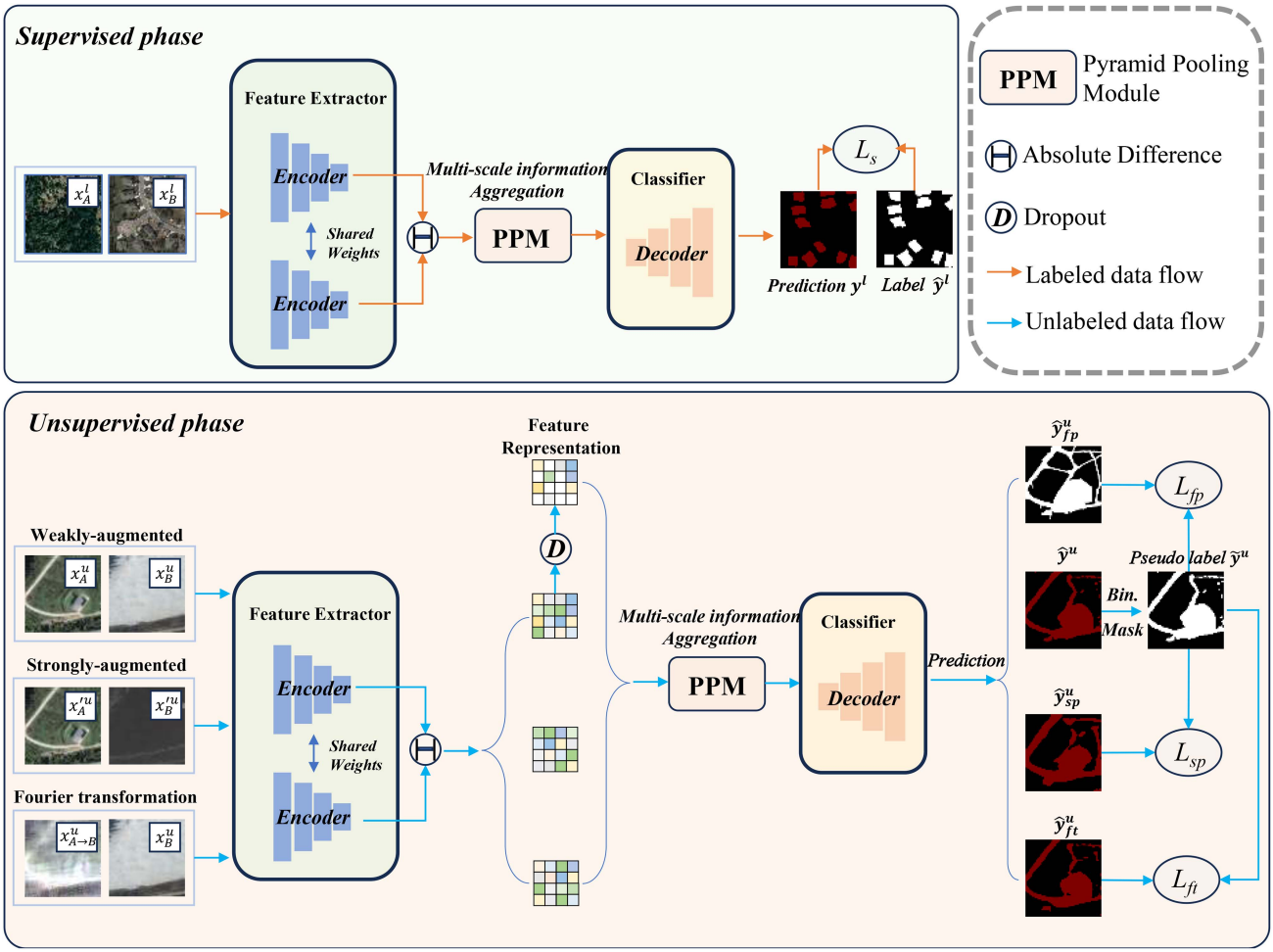


Fig. 1. Overarching architecture of the proposed method. The encoder and decoder are shared during the supervised and unsupervised phases.

bitemporal semantic features, which are subsequently fused at the feature level in the single decoder to identify changed areas.

Specifically, the entire CD procedure is as follows. For the encoder, we utilize a pretrained dilated ResNet50 [14]. Initially, the bitemporal RSI pairs $\{x_A, x_B\}$ are fed into the Siamese encoders, denoted as f_e, to extract bitemporal semantic feature representations. It can be formulated as follows:

$$\begin{aligned} F_A &= f_e(x_A) \\ F_B &= f_e(x_B) \end{aligned} \tag{1}$$

where F_A and F_B denote the feature representations corresponding to the prechange image x_A and the postchange image x_B, respectively.

Following the extraction of hidden features F_A and F_B from the encoder, we calculate their absolute difference, denoted as F_d. This disparity is subsequently processed through a pyramid pooling module (PPM) [59] for an effective exploration of change areas across various scales. Mathematically, the operations can be succinctly articulated as follows:

$$F_d = \text{PPM}(|F_A - F_B|). \tag{2}$$

Finally, we employ a simple CNN in the decoder, denoted as f_d, to estimate the output change probability map $\hat{y}$ from the hidden feature difference F_d. We can formulate the process as follows:

$$\hat{y} = f_d(F_d). \tag{3}$$

Thus, we can express our CD model mathematically as follows:

$$f_{\text{CD}}(x_A, x_B) = f_d\left(\text{PPM}\left(|f_e(x_A) - f_e(x_B)|\right)\right). \tag{4}$$

For supervised learning, we optimize the parameters of our CD network by leveraging the labeled set D^l. The change prediction, represented as $\hat{y}^l$, is directly derived through $f_{\text{CD}}(x^l_A, x^l_B)$. The corresponding supervised loss can be denoted as follows:

$$\mathcal{L}_s = \frac{1}{B}\sum_{i=1}^{B} \mathcal{H}(y^{l,i}, \hat{y}^{l,i}) \tag{5}$$

where B is the batch size of the labeled data and $\mathcal{H}$ refers to the cross-entropy loss.

Throughout the unsupervised training phase, we leverage the unlabeled set D^u to effectively harness the wealth of off-the-shelf unlabeled bitemporal images. Specifically, our method

integrates the multilevel consistency regularization approach with our proposed Fourier-based frequency transformation and the reliable pseudolabel selection scheme to facilitate model training. With regard to multilevel consistency regularization, our approach incorporates two levels of augmentations: sample-level and feature-level. At the sample-level perturbation, this entails generating two distinct perspectives for the original RSI pair: a weakly augmented view and a strongly augmented view. Weak data augmentations mainly involves relatively simple processing, including flipping, rotation, and cropping, whereas strong data augmentation entails more complicated techniques, such as ColorJitter and CutMix [60]. Regarding feature-level perturbation, the utilization of simple channel dropout has proven to be effective and yields satisfactory performance.

To be specific, we integrate unlabeled weakly augmented RSI pairs $\{x_A^u, x_B^u\}$ and strongly augmented views $\{x_A'^u, x_B'^u\}$ into the network. The resulting outputs consist of pixelwise change probability maps, denoted as $\hat{y}^u$ and $\hat{y}_{\text{sp}}^u$ for the weakly and strongly augmented cases, respectively. Then, we obtain the one-hot pseudolabel $\widetilde{y}^u$ corresponding to the weakly augmented view. Therefore, the sample-level perturbation consistency loss is formulated as follows:

$$\mathcal{L}_{\text{sp}} = \frac{1}{\mu B}\sum_{i=1}^{\mu B} \mathbb{1}(\max(y^{u,i}) \geq \tau) \cdot \mathcal{H}(\widetilde{y}^{u,i}, \hat{y}_{\text{sp}}^{u,i}) \tag{6}$$

where the hyperparameter μ is an integer, defining the ratio of unlabeled data to labeled data, and τ represents a predefined threshold used to filter out unreliable pseudolabels associated with low-prediction confidence. The notation $\mathbb{1}$ denotes the indicator function employed for confidence-based thresholding.

As to feature-level perturbation consistency loss, we incorporate feature dropout into the hidden feature differences of the weakly augmented view. Following this, we acquire the pixelwise change probability maps by forwarding the perturbed features through the decoder, denoted as $\hat{y}_{\text{fp}}^u$. It can be denoted as follows:

$$\mathcal{L}_{\text{fp}} = \frac{1}{\mu B}\sum_{i=1}^{\mu B} \mathbb{1}(\max(y^{u,i}) \geq \tau) \cdot \mathcal{H}(\widetilde{y}^{u,i}, \hat{y}_{\text{fp}}^{u,i}). \tag{7}$$

C. Fourier-Based Frequency Transformation

In CD scenarios, images taken at different times may exhibit significant variations in background areas due to factors such as inconsistent imaging and weather conditions. These variations can lead to changes in color, brightness, and other attributes, often manifesting as low-frequency variations in the frequency domain, primarily concentrated in the background region. This challenge is particularly significant as the network struggles to effectively convey information across diverse low-level statistical data [36]. Consequently, addressing known perturbation variability from the outset can eliminate the need for complex training processes.

To address this challenge, we propose a Fourier-based frequency transformation strategy. While Fourier transform has been applied in various CD studies, the focus has often been on different objectives. For instance, Chen et al. [61] utilized Fourier transform to analyze structural relationships within multimodal RSIs, whereas Wen et al. [62] focused on noise suppression through Fourier domain transformations. In contrast, our method targets the mitigation of cross-temporal background noise to enhance CD accuracy, presenting a distinct approach in the realm of semisupervised CD methods.

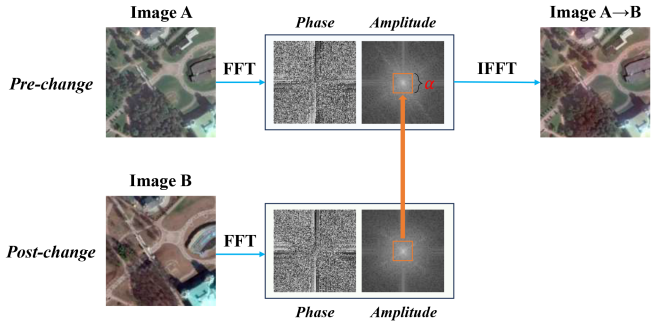


Fig. 2. Illustration of the Fourier-based frequency transformation: Mapping the prechange image to the postchange "style" while preserving its semantic content.

As illustrated in Fig. 2, an unlabeled bitemporal image pair $\{x_A^u, x_B^u\}$ is initially subjected to Fourier transform, denoted as $\mathcal{F}$, transitioning to the frequency domain to derive their respective amplitude spectra and phase spectra. Let $\mathcal{A}, \mathcal{P}$ represent the amplitude and phase components of the Fourier transform of the initial image. The Fourier transform $\mathcal{F}$ can be mathematically formulated as follows:

$$\mathcal{F}(x)(m,n) = \sum_{h=1}^{H}\sum_{w=1}^{W} x(h,w)\mathrm{e}^{-j2\pi\left(\frac{h}{H}m+\frac{w}{W}n\right)}. \tag{8}$$

Therefore, the amplitude spectrum and phase spectrum of the original image can be represented as follows:

$$\mathcal{A}(m,n) = |\mathcal{F}(x)(m,n)| \tag{9}$$

$$\mathcal{P}(m,n) = \arg(\mathcal{F}(x)(m,n)) \tag{10}$$

where the computational efficiency of the presented Fourier transform, as described in [63], is notably enhanced through the utilization of the fast Fourier transform (FFT) algorithm. In mathematical terms, the operator $\mathcal{F}^{-1}$ corresponds to the inverse Fourier transform, facilitating the transformation of spectral signals back into the spatial domain of an image. In addition, we introduce a mask $\mathcal{M}_\alpha$, characterized by zero values across its expanse, with the exception of its central region where $\alpha \in (0,1)$. $\mathcal{M}_\alpha$ is denoted as follows:

$$\mathcal{M}_\alpha(h,w) = \begin{cases} 1, & \text{if } (h,w) \in [-\alpha H : \alpha H, -\alpha W : \alpha W] \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Subsequently, when provided the dual-temporal RSI pairs $\{x_A, x_B\}$ captured at the same location, we have the capability to interchange the amplitude information of the low-frequency region between these images. This exchange allows us to acquire the amplitude information associated with the other temporal instance. The process is finalized by employing the inverse Fourier transform, which restores the modified spectral information back

to the original image domain. Through this method, the primary semantic content of the image remains unaltered, yet its visual appearance mirrors that of another image. The process can be formulated as follows:

$$x_{B\to A} = \mathcal{F}^{-1}([\mathcal{M}_\alpha \cdot \mathcal{A}_A + (1-\mathcal{M}_\alpha)\cdot \mathcal{A}_B]\cdot \mathrm{e}^{j\mathcal{P}_B}) \quad (12)$$

where the amplitude information of the low-frequency region in the postchange image x_B is substituted with that of the prechange image x_A. Due to the potential disruption of local image details and feature positions by altering phase information, it is important to note that the phase spectrum remains unchanged.

Building upon this transformation method, we devised the following training strategy. Over the entire training process, we introduced an additional branch dedicated to bitemporal image transformation. A random parameter, denoted as υ, is stochastically sampled from a uniform distribution in the interval (0,1). If υ exceeds the designated threshold of 0.6, the image pair undergoes transformation from $\{x_A, x_B\}$ to $\{x_{A\to B}, x_B\}$. Conversely, if υ falls below the set threshold of 0.4, the image pair undergoes transformation from $\{x_A, x_B\}$ to $\{x_A, x_{B\to A}\}$. When υ is between 0.4 and 0.6, no transformation is applied to the original image pair. Subsequently, we can derive the predicted probability of change maps for this additional bitemporal image transformation, denoted as $\hat{y}^u_{\text{ft}}$. The loss function is computed by calculating the cross entropy between this prediction and the previously generated one-hot pseudolabels, which is expressed as follows:

$$\mathcal{L}_{\text{ft}} = \frac{1}{\mu B}\sum_{i=1}^{\mu B} \mathbb{1}(\max(y^{u,i}) \geq \tau)\cdot \mathcal{H}(\widetilde{y}^{u,i}, \hat{y}^{u,i}_{\text{ft}}). \quad (13)$$

D. Reliable Pseudolabel Selection Scheme

To mitigate confirmation bias, prevailing semisupervised CD methods [26], [28] employing consistency regularization filter out unreliable pseudolabels by setting a high threshold (e.g., 0.95) in advance. However, employing a predefined threshold may lead to minimal utilization of the entire unlabeled set, particularly at the early stages of the training process when only a few unlabeled pixels have prediction confidences above the threshold. Furthermore, the fixed threshold may result in the selection of more unlabeled examples with wrong pseudolabels, especially at the later stages of the model training process. In other words, the fixed threshold may not be sufficiently effective throughout the training process, potentially degrading the overall performance.

Therefore, we propose a novel and reliable pseudolabel selection strategy, achieved by devising a simple sigmoid-like function to dynamically adjust the threshold. This strategy takes into consideration the learning status of the entire training process, as the threshold dynamically adjusts according to the current training epoch. In theory, we can rewrite the indicator function $\mathbb{1}(\max(y^{u,i}) \geq \tau)$ into an equivalent expression

$$\mathbb{1}(-\log(\max(y^{u,i})) \leq -\log(\tau)) \quad (14)$$

where $-\log(\max(y^{u,i}))$ can be approximated as a cross-entropy loss for the one-hot label of weakly-augmented images.

Intuitively, as the optimization epoch progresses, the loss function would generally decrease, necessitating an increase in $\tau(t)$. Specifically, in the initial stages of model training, due to the lack of prior information, the loss often struggles to decrease and remains at a high value, necessitating $\tau(t)$ to increase slowly and remain at a relatively low level. As the model undergoes gradual training, gaining confidence in its judgment and enhancing accuracy in pseudolabel assignments, the loss will decrease rapidly. At this stage, $\tau(t)$ needs to grow rapidly to ensure the selection of reliable pseudolabels. In the end, as the model training tends to stabilize and the loss tends to reach a plateau, $\tau(t)$ remains at a relatively high value to prevent the introduction of excessive erroneous label information. Hence, we opt for the sigmoid-like function as the threshold adjustment mechanism for selecting reliable pseudolabels. The function is mathematically formulated as follows:

$$\tau(t) = \tau(0) + \frac{\tau(T)-\tau(0)}{1+\mathrm{e}^{-k\left(2\cdot\frac{t}{T}-1\right)}} \quad (15)$$

where $\tau(t)$ represents the evolving threshold value as a function of training epochs, $\tau(0)$ and $\tau(T)$ are the initial threshold and the final threshold, respectively. The parameter k serves as a control for the steepness of the sigmoid-like transition and is experimentally set to be 10.

E. Loss Functions

The comprehensive loss function is a combination of the supervised loss $\mathcal{L}_s$, the feature-level perturbation consistency loss $\mathcal{L}_{\text{fp}}$, the sample-level perturbation consistency loss $\mathcal{L}_{\text{sp}}$, and the Fourier-based frequency transformation loss $\mathcal{L}_{\text{ft}}$, expressed as follows:

$$\mathcal{L} = \mathcal{L}_s + \lambda_{\text{fp}}\mathcal{L}_{\text{fp}} + \lambda_{\text{sp}}\mathcal{L}_{\text{sp}} + \lambda_{\text{ft}}\mathcal{L}_{\text{ft}} \quad (16)$$

where λ_{fp}, λ_{sp}, and λ_{ft} serve as weights to balance the individual components of the loss.

IV. Experimental Results

In this section, we initially provide a detailed illustration of the experimental settings, including dataset description, evaluation metrics, and implementation details. Then, quantitative comparison results against several state-of-the-art methods are performed with some visualization samples. Finally, we conduct several ablation experiments to verify the effectiveness of each component in our proposed method.

A. Experimental Settings

1) Datasets: To verify the effectiveness of our proposed method, we employ three publicly available RSI CD datasets, LEVIR-CD [41], WHU-CD [4], and CDD [64]. Some visualization examples of the three aforementioned datasets are shown in Fig. 3.

LEVIR-CD consists of 637 very high-resolution (0.5 m/pixel) RSI patch pairs, each measuring 1024×1024 pixels. These bitemporal images span a time span of 5–14 years, ranging from 2002 to 2018, illustrating significant land use changes,

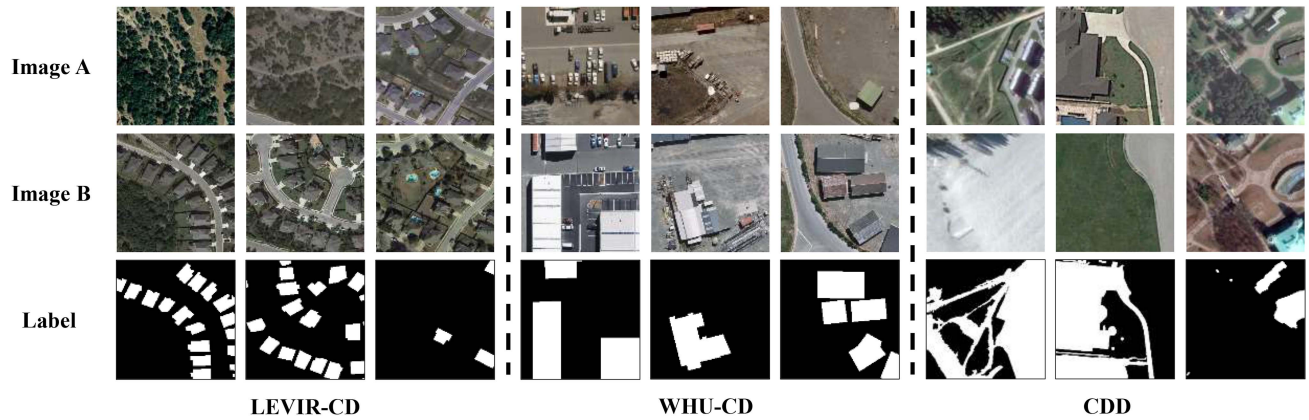


Fig. 3. Some visualization examples of the LEVIR-CD, WHU-CD, and CDD datasets. Image A and B depict the prechange images and the postchange images, respectively.

particularly in the expansion of the construction industry. The dataset covers various types of buildings, including villas, high-rise apartments, small garages, and large warehouses. In line with recent semisupervised CD methods [26], [27], we subdivided the original images into nonoverlapping patches of size 256×256. Following the processing of the original data, we further partitioned these samples into training, validation, and test datasets. Ultimately, we obtained 7120, 1024, and 2048 pairs of patches for training, validation, and testing, respectively.

WHU-CD comprises two-period aerial images captured over the same area of Christchurch, New Zealand, in 2012 and 2016, respectively. The original images have a size of 32507×15354 pixels with a pixel resolution of 0.075 m. The dataset predominantly covers the area reconstructed after an earthquake. Following common processing steps, we cropped these images into nonoverlapping patches of size 256×256 and distributed them into training (80%), validation (10%), and test (10%) set. Consequently, we obtained 5947, 743, and 744 pairs of patches for training, validation, and testing, respectively.

CDD comprises seven pairs of season-varying images, each measuring 4725×2700 pixels, and an additional four pairs of images with minimal changes, each measuring 1900×1000 pixels, all obtained from Google Earth. The dataset is subsequently cropped into 16 000 real RSI patches, each with dimensions of 256×256 pixels and a pixel resolution ranging from 0.03 to 1 m. Specifically, the dataset is divided into 10 000, 2998, and 3000 pairs of patches for training, validation, and testing, respectively.

2) Evaluate Metrics: Various evaluation metrics are employed to assess the model performance, including intersection over union (IoU), F1-Score (F1), Kappa coefficient (Kappa), true positive rate (TPR), and true negative rate (TNR). Given our emphasis on changed areas, we specifically utilize the IoU of the change class, denoted as $\mathrm{IoU}^{\mathrm{c}}$, as the primary evaluation metric in our experiments. For all these metrics, higher values indicate better CD performance. The formulations for the five metrics are as follows:

$$\mathrm{IoU}^{\mathrm{c}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{17}$$

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{18}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{19}$$

$$\mathrm{F1} = \frac{2 \times P \times R}{P + R} \tag{20}$$

$$\mathrm{OA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{21}$$

$$\mathrm{PRE} = \frac{(\mathrm{TP} + \mathrm{FN}) \times (\mathrm{TP} + \mathrm{FP})}{(\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN})^2} + \frac{(\mathrm{FP} + \mathrm{TN}) \times (\mathrm{TN} + \mathrm{FN})}{(\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN})^2} \tag{22}$$

$$\mathrm{Kappa} = \frac{\mathrm{OA} - \mathrm{PRE}}{1 - \mathrm{PRE}} \tag{23}$$

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{24}$$

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{25}$$

where TP and FP denote the count of correctly identified pixels and incorrectly classified pixels, respectively, within the changed areas. Conversely, TN and FN represent the counts for correctly identified and incorrectly classified pixels, respectively, within the unchanged areas.

3) Implementation Details: The experiments in the article are executed on GeForce RTX 3090Ti using the PyTorch framework. To ensure a fair comparison, we maintain all training hyperparameters identical to those utilized in RCR [26]. Stochastic gradient descent with a weight decay of 10^{-4} and a momentum of 0.9 serves as the optimizer, with an initial learning rate set to 0.01. All models undergo training for 80 epochs with a batch size of 8. The backbones of each method are initialized with the pretrained parameters of dilated ResNet50 [14] using ImageNet [65]. For the model training loss function, the weight parameters for each level are set equally. Weak data augmentations, such as flipping, rotation, and cropping, are applied

TABLE I
COMPARISON WITH STATE-OF-THE-ART METHODS ON THE LEVIR-CD DATASET

Method	1%					5%					10%					20%				
	IoU^c	F1	Kappa	TPR	TNR	IoU^c	F1	Kappa	TPR	TNR	IoU^c	F1	Kappa	TPR	TNR	IoU^c	F1	Kappa	TPR	TNR
Only-sup	43.43	60.56	0.5899	51.53	99.09	65.55	79.19	0.7822	75.89	99.23	72.13	83.81	0.8306	80.60	99.43	75.97	86.35	0.8571	82.93	99.55
AdvEnt [49]	43.19	60.33	0.5890	48.21	99.43	71.74	83.54	0.8279	79.42	99.48	75.98	86.35	0.8572	81.97	99.62	78.59	88.01	0.8745	84.53	99.63
s4GAN [25]	26.35	41.71	0.4044	27.06	**99.87**	55.86	71.68	0.7060	59.33	**99.69**	65.20	78.93	0.7794	76.82	99.13	76.16	86.47	0.8586	80.88	**99.70**
CPS [53]	53.04	69.31	0.6805	60.28	99.34	72.53	84.08	0.8334	79.95	99.50	77.34	87.22	0.8663	83.26	**99.63**	79.29	88.45	0.8790	86.12	99.58
RCR [26]	61.35	76.04	0.7507	65.89	99.64	75.75	86.20	0.8554	85.40	99.38	77.97	87.62	0.8703	86.16	99.49	80.00	88.89	0.8836	**87.18**	99.56
Unimatch [54]	68.45	81.27	0.8032	77.10	99.32	74.94	85.67	0.8494	81.65	99.52	76.76	86.85	0.8619	82.05	**99.63**	77.18	87.12	0.8237	87.12	99.64
FPA [28]	57.24	72.81	0.7169	63.38	99.48	75.82	86.25	0.8562	81.85	99.61	78.65	88.05	0.8749	84.71	**99.63**	79.61	88.65	0.8811	87.05	99.55
Ours	**71.29**	**83.24**	**0.8237**	**80.40**	99.31	**78.36**	**87.87**	**0.8723**	**85.96**	99.48	**80.57**	**89.24**	**0.8868**	**87.07**	99.57	**80.55**	**89.23**	**0.8867**	86.67	99.59

Bold values indicate the best results.

TABLE II
COMPARISON WITH STATE-OF-THE-ART METHODS ON THE WHU-CD DATASET

Method	1%					5%					10%					20%				
	IoU^c	F1	Kappa	TPR	TNR	IoU^c	F1	Kappa	TPR	TNR	IoU^c	F1	Kappa	TPR	TNR	IoU^c	F1	Kappa	TPR	TNR
Only-sup	37.14	54.17	0.5288	40.48	99.61	67.81	80.82	0.7998	82.21	99.09	74.11	85.13	0.8449	85.67	99.33	76.95	86.97	0.8641	86.99	99.44
AdvEnt [49]	33.93	50.67	0.4663	35.82	99.17	76.67	86.80	0.8625	84.12	99.58	78.96	88.24	0.8774	87.35	99.54	79.95	88.86	0.8838	88.96	99.52
s4GAN [25]	28.87	44.80	0.4369	29.44	**99.91**	49.75	66.44	0.6542	51.86	**99.82**	72.15	83.82	0.8314	83.05	99.35	67.02	80.25	0.7954	69.73	**99.83**
CPS [53]	39.55	56.68	0.5536	44.25	99.49	71.76	83.56	0.8286	82.85	99.34	77.30	87.20	0.8666	85.63	99.54	82.82	90.61	0.9021	88.96	99.68
RCR [26]	43.25	60.39	0.5911	48.68	99.46	75.26	85.89	0.8525	**89.53**	99.19	79.94	88.85	0.8838	87.14	99.61	81.68	89.92	0.8949	88.38	99.65
Unimatch [54]	62.35	76.81	0.7598	67.92	99.63	78.27	87.81	0.8735	81.82	99.81	81.15	89.60	0.8919	85.07	99.80	82.78	90.58	0.9021	86.26	**99.83**
FPA [28]	62.69	77.07	0.7619	**69.50**	99.53	76.95	86.97	0.8640	89.09	99.32	**82.74**	**90.56**	**0.9015**	**90.22**	99.61	**84.00**	**91.30**	**0.9093**	**90.47**	99.67
Ours	**63.90**	**77.97**	**0.7720**	68.14	99.73	**79.90**	**88.82**	**0.8839**	84.26	99.77	81.70	89.93	0.8954	85.43	**99.81**	82.97	90.69	0.9032	87.06	99.80

Bold values indicate the best results.

TABLE III
COMPARISON WITH STATE-OF-THE-ART METHODS ON THE CDD DATASET

Method	1%					5%					10%					20%				
	IoU^c	F1	Kappa	TPR	TNR	IoU^c	F1	Kappa	TPR	TNR	IoU^c	F1	Kappa	TPR	TNR	IoU^c	F1	Kappa	TPR	TNR
Only-sup	36.05	53.00	0.4820	41.30	97.84	59.28	74.43	0.7105	67.96	97.83	69.10	81.73	0.7922	76.67	98.38	76.84	86.90	0.8507	82.77	98.85
AdvEnt [49]	33.93	50.67	0.4663	35.82	99.17	63.06	77.35	0.7437	70.28	98.30	72.98	84.38	0.8230	77.33	99.11	81.08	89.55	0.8808	85.96	99.11
s4GAN [25]	3.35	6.48	0.0540	3.39	**99.82**	46.10	63.11	0.5905	51.20	98.36	67.44	80.55	0.7802	72.77	98.83	79.56	88.62	0.8702	84.67	99.05
CPS [53]	38.57	55.67	0.5099	44.11	97.87	62.35	76.81	0.7373	70.36	98.09	72.94	84.36	0.8225	77.86	99.00	80.63	89.27	0.8775	86.22	98.97
RCR [26]	34.31	51.09	0.4654	38.11	98.36	66.67	80.00	0.7739	72.37	98.73	74.64	85.48	0.8349	80.21	98.89	80.91	89.45	0.8794	86.47	98.98
Unimatch [54]	50.02	66.68	0.6340	53.78	98.99	69.95	**83.32**	0.8025	74.02	**99.22**	76.24	86.52	0.8486	80.71	99.22	84.34	91.50	0.9041	88.49	99.34
FPA [28]	47.18	64.11	0.5925	**59.45**	96.14	67.71	80.75	0.7824	72.88	98.87	75.63	86.12	0.8423	80.57	99.03	81.21	89.63	0.8818	85.18	99.27
Ours	**51.00**	**67.55**	**0.6405**	57.55	98.28	**71.14**	83.14	**0.8110**	**76.47**	99.00	**77.24**	**87.16**	**0.8558**	**81.56**	**99.25**	**86.13**	**92.55**	**0.9158**	**90.30**	**99.35**

Bold values indicate the best results.

to the original RSI pairs. Specifically, for consistency-based methods, strong data augmentations, including ColorJitter and CutMix [60], are utilized to perturb the weakly augmented unlabeled RSI pairs. The size ratio of low-frequency exchange in the frequency domain is controlled by the parameter α. In the experiment, we set α to 0.01 for the LEVIR-CD and WHU-CD datasets, whereas for the CDD dataset, α is 0.09. Furthermore, we establish the parameters for balancing the loss, setting λ_{fp} to 0.5 and λ_{sp}, λ_{ft} to 0.25 each, respectively.

B. Comparison Results

We select supervised baseline and six state-of-the-art methods for comparison to assess the performance of our proposed method, including AdvEnt [49], s4GAN [25], CPS [53], RCR [26], Unimatch [54], and FPA [28]. Here, the supervised baseline indicates that we exclusively employ restricted proportions of labeled data for model training. Both AdvEnt and s4GAN are GAN-based methods adapted from semisupervised semantic segmentation. The remaining four methods are consistency-based semisupervised approaches. Specifically, RCR and FPA are designed explicitly for CD tasks, whereas CPS and Unimatch are adapted from semisupervised semantic segmentation. In our experimental setup, four distinct proportions of labeled data, namely 1%, 5%, 10%, and 20%, are employed during the supervised phases, with the remaining unlabeled data utilized in the unsupervised phases.

The experimental results for the three public datasets are presented in Tables I–III, respectively. The experimental results indicate that our method achieves the best performance when considering the majority ratios of labeled data across the three datasets.

1) Results of the LEVIR-CD Dataset: The comparison results of the LEVIR-CD dataset, as presented in Table I, highlight

the exceptional performance of our method across all four ratio settings, ranging from 1% to 20%. Notably, the considerable temporal span of 5 to 14 years between bi-temporal RSI pairs introduces a noticeable amount of cross-temporal background noise in the dataset, alongside the actual changes requiring detection. Hence, the Only-sup method, trained directly from labeled data, performs notably poorly, and GAN-based methods frequently result in model collapse and subpar performance due to unstable training. In contrast, the method based on the consistency regularization strategy achieved the best performance relatively. Our proposed method outperforms others, especially in scenarios with extremely limited labeled data (1% and 5%). We achieve the impressive 2.84% and 2.54% improvements in the $\mathrm{IoU}^{\mathrm{c}}$ metric, which is of utmost concern. In the other two settings, our method can also achieve improvements of at least 1.92%, and 0.55%, respectively, compared with the current most advanced semisupervised CD algorithm. On other evaluation metrics, our method consistently demonstrates impressive results.

2) Results of the WHU-CD Dataset: In the WHU-CD dataset, our method and FPA yield comparable results, with our results being superior at 1% and 5% ratios, whereas FPA performs slightly better than our method in the remaining settings, as shown in Table II. The primary challenge in the WHU-CD dataset lies in the intraclass diversity within changing regions, which includes various types of buildings rather than background noise. Consequently, the consistency alignment strategies FA and PA proposed in the FPA method significantly reduce the uncertainty in intraregion prediction. Specifically, under the $\mathrm{IoU}^{\mathrm{c}}$ metric, our method demonstrated the significant 1.21% and 2.95% improvements compared with FPA when trained on a dataset with 1% and 5% labeled data, respectively. However, in the remaining two settings, our method narrowly trailed behind the FPA algorithm with differences of -1.04%, and -1.03% at 10%, and 20% ratios of labeled data, respectively. Nevertheless, compared with other semisupervised CD algorithms, our method can establish a substantial advantage on most of the evaluation metrics.

3) Results of the CDD Dataset: Results of the CDD dataset are shown in Table III. In the CDD dataset, the bitemporal images exhibit a more distinct appearance, attributed to diverse weather conditions during capture. Therefore, unlike the above two datasets, we opt for a larger area ($\alpha = 0.09$) for low-frequency exchange. Our proposed Fourier-based frequency transformation strategy adeptly alleviates the challenges posed by the inherent variability introduced by both known disturbances and changes that necessitate identification. This approach demands minimal effort, particularly in distinguishing between noise and genuine changing regions. From the results in the table, it can be observed that our method significantly outperforms existing state-of-the-art algorithms comprehensively. For example, in the metric of $\mathrm{IoU}^{\mathrm{c}}$, our method can achieve improvements of at least 0.98%, 1.19%, 1.00%, and 1.79%, respectively.

The presented quantitative results affirm the superiority of our proposed method. In addition, for a more comprehensive visual comparison between our approach and alternative methods, we have chosen specific visual prediction results showcased in Fig. 4. The analysis of visual samples demonstrates the superior accuracy of our method in predicting boundary details compared with other methods, accompanied by a significant reduction in the proportion of false positives. This enhancement is credited to our innovative Fourier-based frequency transformation and the reliable pseudolabel selection scheme.

TABLE IV
COMPONENTS ANALYSIS RESULTS ON THE LEVIR-CD DATASET AT THE RATIOS OF 5% AND 10%

$\mathcal{L}_s$	$\mathcal{L}_{\mathrm{sp}} + \mathcal{L}_{\mathrm{fp}}$	$\mathcal{L}_{\mathrm{ft}}$	$\tau(t)$	5%			10%		
				$\mathrm{IoU}^{\mathrm{c}}$	F1	Kappa	$\mathrm{IoU}^{\mathrm{c}}$	F1	Kappa
✓				65.55	79.19	0.7822	72.13	83.81	0.8306
✓	✓			74.45	85.36	0.8461	75.25	85.88	0.8519
✓	✓	✓		76.30	86.55	0.8586	78.38	87.88	0.8725
✓	✓	✓	✓	78.36	87.87	0.8723	80.57	89.24	0.8868

C. Ablation Experiment

To comprehensively explore our proposed method, we conduct extensive ablation experiments on the LEVIR-CD, WHU-CD, and CDD datasets. Initially, we discuss the effectiveness and role of each component within our proposed method. Then, we examine the influence on model performance of adjusting the size of the central region where low-frequency amplitude information is replaced in the Fourier-based frequency transformation strategy. Subsequently, we assess the impact of different threshold adjustment functions for selecting reliable pseudolabels on the overall performance of the model. In addition, we analyze the effect of the initial threshold in our strategies. Finally, we present a detailed analysis of the computational consumption of our method in comparison to other models.

1) Components Analysis: As illustrated in Section III, the proposed method comprises three key components: a multilevel consistency regularization baseline, a Fourier-based frequency domain transformation branch, and a reliable pseudolabel selection scheme. To validate the significance of the aforementioned components within the proposed method, we iteratively train ablation models with different components on the LEVIR-CD dataset, considering labeled ratios of 1%, 5%, 10%, and 20%. Quantitative results are presented in Table IV. Here, $\mathcal{L}_s$ represents the supervised approach, which utilizes only labeled data for model training, $\mathcal{L}_s + \mathcal{L}_{\mathrm{sp}} + \mathcal{L}_{\mathrm{fp}}$ denotes the multilevel consistency regularization baseline, $\mathcal{L}_s + \mathcal{L}_{\mathrm{sp}} + \mathcal{L}_{\mathrm{fp}} + \mathcal{L}_{\mathrm{ft}}$ signifies the method that additionally incorporates the Fourier-based frequency transformation branch, and $\mathcal{L}_s + \mathcal{L}_{\mathrm{sp}} + \mathcal{L}_{\mathrm{fp}} + \mathcal{L}_{\mathrm{ft}} + \tau(t)$ indicates dynamic threshold adjustment function for pseudolabel selection based on the aforementioned methods. Due to the inherent tradeoff between TPR and TNR, achieving higher values for both simultaneously is often unattainable. Consequently, we have selected the other three metrics to precisely characterize the effects of each component.

As observed in Table IV, the multilevel consistency regularization baseline, i.e., $\mathcal{L}_s + \mathcal{L}_{\mathrm{sp}} + \mathcal{L}_{\mathrm{fp}}$, significantly enhances

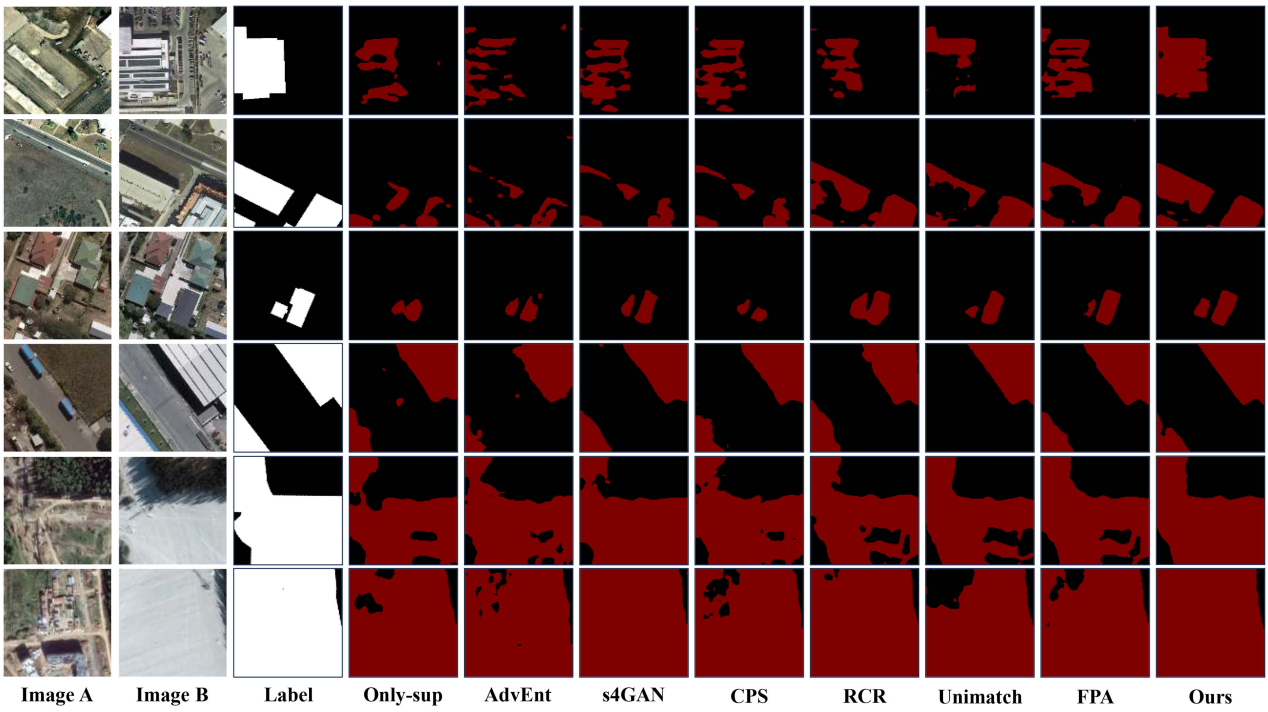


Fig. 4. Some sample visualizations obtained from various comparison methods, including our method. The first and second rows depict the RSI pairs of LEVIR-CD, the third and fourth rows show WHU-CD, and the fifth and sixth rows display CDD, all at the 5% labeled ratio.

model performance, exhibiting a notable improvement compared to the supervised method $\mathcal{L}_s$. For instance, the IoU^c increases from 65.55% to 74.45% by 8.9% at the ratio of 5%. Furthermore, building upon the baseline, the introduction of the Fourier-based frequency transformation branch, namely, $\mathcal{L}_{\mathrm{ft}}$, results in a substantial boost in model performance. Specifically, it improves the IoU^c by 1.85% and 3.13%, respectively, under the two different ratios. This underscores the positive impact of integrating the Fourier-based frequency transformation branch. Finally, based on the aforementioned methods, we effectively select more reliable pseudolabel by dynamically adjusting the threshold, denoted as $\tau(t)$. The results in the table highlight that our threshold adjustment strategy significantly improves the model's detection capability, with notable increases in IoU^c by 2.06% and 2.19% at the ratios of 5% and 10%, respectively.

In summary, the experimental results underscore the pivotal role of each component in our method during model training. Integrating these components synergistically yields optimal results.

2) Choice of α: In (12), the parameter α is a critical factor governing the Fourier-based frequency transformation. Setting α to 0 renders the transformed image $x_{B\to A}$ identical to the original postchange image x_B, whereas α at 1 replaces the amplitude of x_B with that of x_A. Prior research indicates that for $\alpha \geq 0.1$, the transformed image $x_{B\to A}$ tends toward the target image x_A, but noticeable artifacts may be present. Therefore, in this article, we limit α to ≤ 0.1 for the three datasets. We provide visualizations of images under four different α values in Fig. 6. These visualizations illustrate how varying α affects the image: As α increases, the background of the transformed image becomes more similar to that of the other image in the pair,

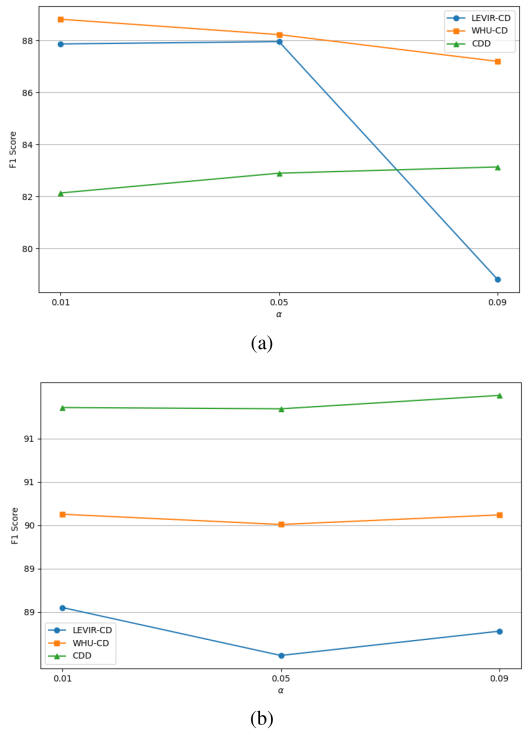


Fig. 5. F1 of (a) 5% and (b) 20% labeled with different α on the three datasets.

but at the cost of introducing artifacts. The red boxes in the figure highlight regions where these artifacts are more pronounced as α increases. Table V shows the effect of various choices of α for the three datasets at the ratios of 5% and 20%. The F1 scores

Fig. 6. Impact of varying the size of α on the Fourier-based frequency transformation operation is noteworthy: As α increases, the cross-temporal gap between bitemporal RSI decreases, yet it also tends to introduce artifacts, as illustrated in the zoomed insets.

TABLE V
CHOICE OF α FOR THE THREE DATASETS AT THE RATIOS OF 5% AND 20%

Dataset	α	5%					20%				
		IoU^c	F1	Kappa	TPR	TNR	IoU^c	F1	Kappa	TPR	TNR
LEVIR-CD	0.01	78.36	87.87	0.8723	85.96	99.48	80.55	89.23	0.8867	86.67	99.59
	0.05	78.51	87.96	0.8733	85.71	99.51	79.34	88.48	0.8788	86.38	99.52
	0.09	65.04	78.82	0.7786	68.85	99.69	79.96	88.86	0.8829	85.49	99.63
WHU-CD	0.01	79.90	88.82	0.8839	84.26	99.77	82.97	90.69	0.9032	87.06	99.80
	0.05	78.94	88.23	0.8778	82.15	99.83	82.69	90.53	0.9015	87.52	99.76
	0.09	77.30	87.20	0.8672	80.27	99.84	82.95	90.68	0.9032	86.68	99.81
CDD	0.01	69.70	82.14	0.8006	73.76	99.22	85.81	92.36	0.9137	90.02	99.34
	0.05	70.79	82.90	0.8087	75.43	99.12	85.77	92.34	0.9135	89.83	99.37
	0.09	71.14	83.14	0.8110	76.47	99.00	86.13	92.55	0.9158	90.30	99.35

of these three datasets with 5% and 20% labeled data under different α values are shown in Fig. 5.

Observing the table, it is evident that selecting $\alpha = 0.09$ yields superior results for the CDD dataset. This preference arises from the dataset's seasonal variable nature, necessitating the replacement of a larger low-frequency region for a comprehensive frequency domain conversion. Therefore, in our experiments, we opted for the value 0.09. In the case of the other two datasets, the distinctions among these three choices are not substantial, except for an unfavorable outcome when choosing $\alpha = 0.09$ for the LEVIR-CD dataset at the ratio of 5%. However, in comparison to other options, opting for $\alpha = 0.01$ proves to be a relatively favorable choice for these two datasets.

3) Influence of Threshold Adjustment Function: As depicted in Fig. 7, we have chosen several threshold adjustment functions to assess their impact on selecting reliable pseudolabels and the overall performance. According to the analysis in the

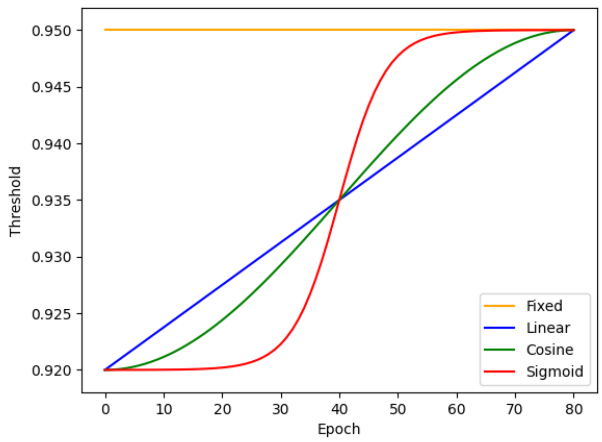


Fig. 7. Comparison of fixed threshold and several threshold adjustment functions. Fixed threshold is in the scale of 0.95.

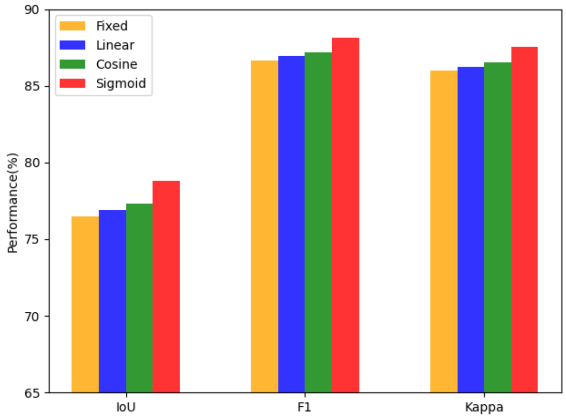


Fig. 8. Comparison results between various threshold adjustment strategies and fixed threshold on the LEVIR-CD dataset at the ratio of 5%.

Section III, we are required to propose a reliable pseudolabel selection strategy, so we have chosen these three threshold adjustment functions, including linear function, cosine function, and sigmoid-like function.

The final results of the ablation experiments on the LEVIR-CD dataset with a 5% ratio of labeled data are illustrated in Fig. 8. Notably, the fixed threshold yields the poorest results across the three commonly used metrics. Conversely, our designed sigmoid-like function demonstrates the best performance. The other two adjustment strategies achieve intermediate performance. This result confirms that our proposed sigmoid-like threshold adjustment function can better adapt to the training process of the model, thereby selecting more reliable pseudolabels.

Furthermore, throughout the entire training process, we compare the number of correct and incorrect pixels in the pseudolabels selected for changed areas between the fixed threshold and our dynamic threshold function. As illustrated in Fig. 9, we verify that the proposed dynamic threshold function selects a greater number of pixels with correct pseudolabels compared to the fixed threshold. While the proposed dynamic threshold function initially preserves more pixels with incorrect pseudolabels, it gradually eliminates more of these pixels after several epochs, in contrast to the fixed threshold.

4) Effect of the Initial Threshold $\tau(0)$: In our proposed dynamic threshold adjustment function, we introduce an initial threshold $\tau(0)$ that gradually increases to the previously established high threshold in the conventional method. This approach facilitates the assimilation of additional information from unlabeled data in the early stages of model training, thereby promoting improved model convergence. We investigate the effect of configuring this initial threshold on model training and conduct relevant experiments on the LEVIR-CD dataset, as detailed in Table VI.

Based on the results presented in the table, it is evident that the selection of the initial threshold should not be excessively low. While an excessively high threshold may diminish the utilization of unlabeled data, a low initial threshold introduces the risk of confirmation bias, potentially leading to unreliable pseudolabels and consistent misalignment of the model's decision boundaries. Striking a balance in this tradeoff, we conducted extensive experiments and determined an optimal initial threshold of 0.92.

TABLE VI
EFFECT OF THE INITIAL THRESHOLD $\tau(0)$ ON THE LEVIR-CD DATASET AT THE RATIOS OF 5% AND 10%

$\tau(0)$	5%			10%		
	IoU^c	F1	Kappa	IoU^c	F1	Kappa
0.86	76.71	86.82	0.8615	79.45	88.55	0.8795
0.88	77.66	87.43	0.8675	76.25	86.52	0.8583
0.90	77.89	87.57	0.8692	78.40	87.89	0.8728
0.92	78.36	87.87	0.8723	80.57	89.24	0.8868
0.93	77.42	87.28	0.8661	79.05	88.30	0.8769
0.94	77.51	87.77	0.8713	78.94	88.23	0.8762

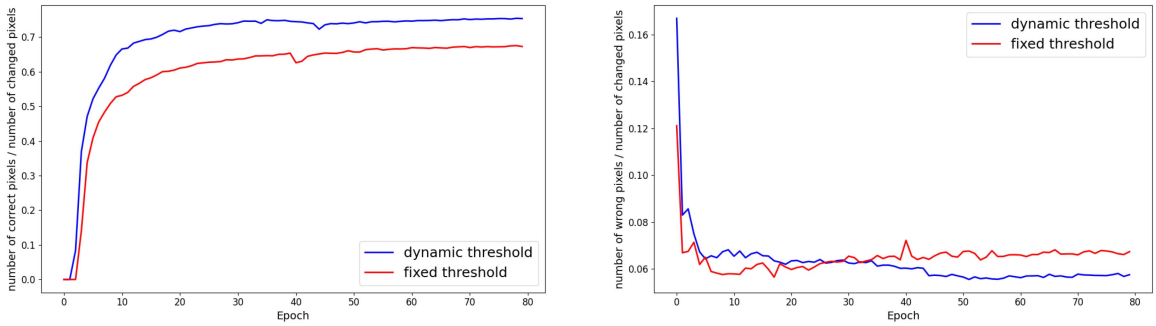
TABLE VII
COMPUTATIONAL ANALYSIS ON THE WHU-CD DATASET AT THE RATIO OF 5%

Method	Params(M)	FLOPs(G)	Training Time (min)	IoU^c
Only-sup	46.85	73.23	55.0	67.81
AdvEnt	46.85	73.23	443.0	76.67
s4GAN	46.85	73.23	646.6	49.75
CPS	47.00	73.41	367.5	71.76
Unimatch	46.71	73.05	929.5	78.27
RCR	46.85	73.23	563.6	75.26
FPA	46.85	73.23	485.7	76.95
Ours	46.71	73.05	925.3	79.90

5) Computational Analysis: In this section, we conduct a comparative analysis between the proposed method and existing semisupervised CD methods, considering computational consumption, training time, and CD performance metrics on the WHU-CD dataset at the ratio of 5%. All comparison frameworks, as well as our method, are trained on a single GeForce RTX 3090Ti, employing identical architectures and training parameters for a fair evaluation. We select the number of parameters, the floating-point operations (FLOPs), and the training time as efficiency metrics, where a smaller value for these metrics indicates higher computational efficiency for the methods.

As illustrated in Table VII, the enhancements we have made do not increase the computational complexity of the model. Our Fourier-based frequency transformation and pseudolabel selection strategies are computationally lightweight and can be seamlessly integrated into the training process of existing models. However, it is important to note that our method requires additional processing of weakly-augmented data, strongly-augmented data, and data with Fourier-based frequency transformation, which results in a longer overall training time compared to other methods.

Despite the longer training time, our method achieves significant improvements in CD performance while maintaining a similar or slightly reduced number of parameters and computational complexity. When compared to methods such as Unimatch, which also processes three batches of data, our method provides



(a) (b)

Fig. 9. Visualization example of the pseudolabels selection quality of changed areas on the LEVIR-CD dataset at the ratio of 5%. (a) Ratio of correctly classified pixels in the changed areas of the selected pseudolabels to the total number of pixels. (b) Ratio of incorrectly classified pixels in the changed areas of the selected pseudolabels to the total number of pixels.

superior performance without requiring additional training time. Overall, our method significantly improves CD performance with only a marginal increase in training duration.

V. Discussion

Our semisupervised CD method presents a significant advancement in addressing the pervasive challenges of cross-temporal background noise and the underutilization of unlabeled data. In this section, we reflect on the strategic choices made during the development of our method, discuss the implications of these strategies, and acknowledge the limitations that require further attention.

First, the Fourier-based frequency transformation aligns the background distribution of bitemporal images. This process mitigates cross-temporal background noise, which could otherwise mislead the model into generating pseudo changes. The parameter α in this transformation is crucial; it governs the extent of amplitude replacement. Through our experiments, we observed that smaller α values effectively reduce background noise without introducing artifacts, whereas larger values (beyond 0.1) may lead to noticeable artifacts, as shown in Fig. 6. This necessitates careful tuning of α depending on the dataset characteristics.

Second, our reliable pseudolabel selection scheme employs a sigmoid-like function to dynamically adjust the threshold for pseudolabel selection. The threshold adjustment is critical to balance the integration of unlabeled data and the accuracy of pseudolabels. An excessively high threshold reduces the utilization of unlabeled data, whereas a low threshold may lead to confirmation bias and unreliable pseudolabels. Our dynamic thresholding adapts throughout the training process, as illustrated in Table VI, ensuring the model's learning status is appropriately considered. Despite its effectiveness, this predefined function might not fully capture the nuanced training status, suggesting room for further optimization.

We recognize that our method is not without limitations. The reliance on a manually selected α parameter, while essential, introduces an element of subjectivity that could be mitigated through more sophisticated hyperparameter optimization techniques. In addition, our dynamic threshold adjustment, though an improvement over static thresholds, may not fully encapsulate the complexity of the model's training dynamics. Future work will explore self-adaptive thresholding strategies that can evolve in tandem with the model's learning progress.

VI. Conclusion

In this article, we introduce a novel semisupervised CD method aimed at tackling two persistent challenges in the field: cross-temporal background noise and underutilization of extensive unsupervised data. Our method incorporates a Fourier-based frequency transformation strategy and a reliable pseudolabel selection scheme to address these challenges. Specifically, the Fourier-based frequency transformation strategy replaces the amplitude information from the low-frequency region of one temporal image with that derived from the corresponding image in the same bitemporal RSI pair. This alignment of background distribution helps mitigate cross-temporal background noise. In addition, we devise a simple sigmoid-like function to dynamically adjust the selection threshold for the reliable pseudolabel selection scheme. This scheme takes into consideration the learning status of the entire training process, ensuring more effective utilization of unlabeled information. It is noteworthy that our proposed method does not incur additional training costs. Extensive experimental results demonstrate that our proposed method exhibits excellent performance on three public datasets, LEVIR-CD, WHU-CD, and CDD, surpassing currently advanced algorithms in the field.

References

[1] L. Gueguen and R. Hamid, "Toward a generalizable image representation for large-scale change detection: Application to generic damage analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3378–3387, Jun. 2016.

[2] Z. Zheng, Y. Zhong, J. Wang, A. Ma, and L. Zhang, "Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters," *Remote Sens. Environ.*, vol. 265, Nov. 2021, Art. no. 112636.

[3] D. Wen et al., "Change detection from very-high-spatial-resolution optical remote sensing images: Methods, applications, and future directions," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 68–101, Dec. 2021.

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

[5] L. Li et al., "Competition between biogeochemical drivers and land-cover changes determines urban greening or browning," *Remote Sens. Environ.*, vol. 287, Mar. 2023, Art. no. 113481.

[6] N. Wang, W. Li, R. Tao, and Q. Du, "Graph-based block-level urban change detection using Sentinel-2 time series," *Remote Sens. Environ.*, vol. 274, Jun. 2022, Art. no. 112993.

[7] B. Desclée, P. Bogaert, and P. Defourny, "Forest change detection by statistical object-based method," *Remote Sens. Environ.*, vol. 102, no. 1–2, pp. 1–11, May 2006.

[8] S. H. Khan, X. He, F. Porikli, and M. Bennamoun, "Forest change detection in incomplete satellite images with deep neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5407–5423, Sep. 2017.

[9] E. L. Bullock et al., "Timeliness in forest change monitoring: A new assessment framework demonstrated using Sentinel-1 and a continuous change detection algorithm," *Remote Sens. Environ.*, vol. 276, Jul. 2022, Art. no. 113043.

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

[11] S. Fang, K. Li, and Z. Li, "Changer: Feature interaction is what you need for change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5610111.

[12] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5607514.

[13] T. Yan, Z. Wan, P. Zhang, G. Cheng, and H. Lu, "TransY-Net: Learning fully transformer networks for change detection of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4410012.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[15] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "RepVGG: Making VGG-style convnets great again," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13733–13742.

[16] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–13.

[17] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 10 012–10 022.

[18] C. Ren, X. Wang, J. Gao, X. Zhou, and H. Chen, "Unsupervised change detection in satellite images with generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10047–10061, Dec. 2021.

[19] H. Noh, J. Ju, M. Seo, J. Park, and D.-G. Choi, "Unsupervised change detection based on image reconstruction loss," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2022, pp. 1352–1361.

[20] W. Zhao, X. Chen, X. Ge, and J. Chen, "Using adversarial network for multiple change detection in bitemporal remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 19, 2020, Art. no. 8003605.

[21] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 2672–2680.

[22] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 5050–5060.

[23] X. Yang, Z. Song, I. King, and Z. Xu, "A survey on deep semi-supervised learning," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 9, pp. 8934–8954, Sep. 2023.

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

[25] S. Mittal, M. Tatarchenko, and T. Brox, "Semi-supervised semantic segmentation with high-and low-level consistency," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 4, pp. 1369–1379, Apr. 2021.

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

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

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

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

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

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

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

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

[34] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 12 674–12 684.

[35] Y. Wang et al., "Semi-supervised semantic segmentation using unreliable pseudo-labels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 4248–4257.

[36] Y. Yang and S. Soatto, "FDA: Fourier domain adaptation for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 4085–4095.

[37] L. Bruzzone and D. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.

[38] J. Deng, K. Wang, Y. Deng, and G. Qi, "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data," *ISPRS J. Photogramm. Remote Sens.*, vol. 29, no. 16, pp. 4823–4838, 2008.

[39] Y. Tang, L. Zhang, and X. Huang, "Object-oriented change detection based on the Kolmogorov–Smirnov test using high-resolution multispectral imagery," *Int. J. Remote Sens.*, vol. 32, no. 20, pp. 5719–5740, 2011.

[40] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

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

[42] K. Tan et al., "Change detection on multi-sensor imagery using mixed interleaved group convolutional network," *Eng. Appl. Artif. Intell.*, vol. 133, 2024, Art. no. 108446.

[43] Y. Qu, J. Li, X. Huang, and D. Wen, "TD-SSCD: A novel network by fusing temporal and differential information for self-supervised remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5407015.

[44] X. Wang et al., "Double U-Net (W-Net): A change detection network with two heads for remote sensing imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 122, 2023, Art. no. 103456.

[45] M. Wang, X. Li, K. Tan, J. Mango, C. Pan, and D. Zhang, "Position-aware graph-CNN fusion network: An integrated approach combining geospatial information and graph attention network for multi-class change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4402016.

[46] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 6840–6851.

[47] W. G. C. Bandara, N. G. Nair, and V. M. Patel, "DDPM-CD: Remote sensing change detection using denoising diffusion probabilistic models," 2022, *arXiv:2206.11892*.

[48] N. Souly, C. Spampinato, and M. Shah, "Semi supervised semantic segmentation using generative adversarial network," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 5688–5696.

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

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

[51] J. Yuan, Y. Liu, C. Shen, Z. Wang, and H. Li, "A simple baseline for semi-supervised semantic segmentation with strong data augmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 8229–8238.

[52] O. Chapelle, B. Schlkopf, and A. Zien, "Semi-supervised learning," *IEEE Trans. Neural Netw.*, vol. 20, no. 3, pp. 542–542, Mar. 2009.

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

[54] L. Yang, L. Qi, L. Feng, W. Zhang, and Y. Shi, "Revisiting weak-to-strong consistency in semi-supervised semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 7236–7246.
[55] F. Bovolo, L. Bruzzone, and M. Marconcini, "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2070–2082, Jul. 2008.
[56] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
[57] C. Zhang et al., "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 183–200, 2020.
[58] S. Zhao, X. Zhang, P. Xiao, and G. He, "Exchanging dual-encoder–decoder: A new strategy for change detection with semantic guidance and spatial localization," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4508016.
[59] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2881–2890.
[60] 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 *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 6023–6032.
[61] H. Chen, N. Yokoya, and M. Chini, "Fourier domain structural relationship analysis for unsupervised multimodal change detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 198, pp. 99–114, 2023.
[62] Y. Wen, X. Ma, X. Zhang, and M.-O. Pun, "GCD-DDPM: A generative change detection model based on difference-feature guided DDPM," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5404416.
[63] M. Frigo and S. G. Johnson, "FFTW: An adaptive software architecture for the FFT," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 1998, pp. 1381–1384.
[64] M. A. Lebedev, Y. V. Vizilter, O. V. Vygolov, V. A. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 565–571, May 2018.
[65] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. F.-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.

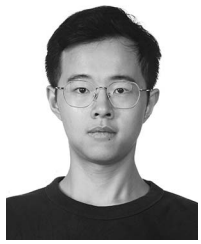
Ze Zhang received the B.S. degree in electronic and information engineering from Harbin Institute of Technology, Harbin, China, in 2022. He is currently working toward the Ph.D. degree in information and communication engineering from Shanghai Jiao Tong University, Shanghai, China.

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

Xue Jiang (Senior Member, IEEE) received the M.S. degree in acoustics and signal processing from the Chinese Academy of Sciences, Beijing, China, in 2007, and the Ph.D. degree in electrical engineering from McMaster University, Hamilton, ON, Canada, in 2014.

From 2007 to 2010, she was a Research Engineer with Sunplus Inc., Beijing. From 2014 to 2015, she was a Postdoctoral Fellow with McMaster University and from 2015 to 2016 with Arizona State University, Tempe, AZ, USA. In 2016, she joined the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China, where she is currently a Professor. Her research interests include array processing, robust beamforming, sparse and low-rank learning, data science, and numerical optimization, with applications to radar, sonar, and remote sensing.

Dr. Jiang was one of the finalists of the Best Student Paper Award of the IEEE 2013 Radar Conference. She was the recipient of the McMaster University International Excellence Award in 2012 and 2013. She is also an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING. She is the Editorial Board Member for *IET Radar, Sonar & Navigation*.



Yue Zhou received the Ph.D. degree in electronics and information from the Shanghai Jiao Tong University, Shanghai, China, in 2024.

He is currently working as a Research Fellow with S-Lab, Nanyang Technological University, Singapore, focusing on remote sensing vision-language model. He is also the leading contributor to the MM-Rotate open-source project for oriented object detection. His research interests include deep learning, computer vision, and remote sensing.

Xingzhao Liu (Member, IEEE) received the B.S. and M.S. degrees from the Harbin Institute of Technology, Harbin, China, in 1984 and 1992, respectively, and the Ph.D. degree from the University of Tokushima, Tokushima, Japan, in 1995, all in electrical engineering.

From 1984 to 1998, he was an Assistant Professor, an Associate Professor, and a Professor with the Harbin Institute of Technology. Since 1998, he has been a Professor with Shanghai Jiao Tong University, Shanghai, China. His research interests include high-frequency radar and synthetic aperture radar signal processing.