B2CNet: A Progressive Change Boundary-to-Center Refinement Network for Multitemporal Remote Sensing Images Change Detection

Zhiqi Zhang, Liyang Bao, Shao Xiang, Guangqi Xie, and Rong Gao

***Abstract*—Change detection is an important method of analyzing information about changes in geographical features. However, existing deep learning feature difference methods often lead to the loss of detailed information. Differences in features can arise from factors like illumination or geometric variations rather than actual change regions, resulting in inaccurate change detection. This leads to poor detection of fine-grained boundaries and internal hole problems. To alleviate this, we propose a novel change detection network guided by change boundary awareness and incorporating the concept of boundary-to-center. Our network introduces a change boundary-aware module to capture boundary information of change regions. This module enhances boundaries, reducing the influence of noise in feature differences and providing rich contextual information to improve the accuracy of change boundaries. Additionally, we propose a bitemporal feature aggregation module (BFAM) based on spatial-temporal features. The BFAM aggregates multiple receptive fields features and complements texture information. Both modules utilize the SimAM attention mechanism to enhance the finegrained nature of the features. In addition, we introduce a deep feature extraction module to extract deep features and minimize information loss during the decoupling process. The proposed change detection network in this article is guided by change boundary perception, progressively integrating semantic and spatial texture information to refine edges and enhance internal integrity. The performance and efficiency of B2CNet have been validated on four publicly available remote sensing image change detection datasets. Through extensive experiments, the effectiveness of the proposed method has been demonstrated. For example, in terms of IOU for LEVIR, WHU, SYSU, and HRCUS datasets, the improvements compared to the baseline are 1.89%, 2.86%, 4.70%, and 3.79%, respectively.**

Manuscript received 16 February 2024; revised 7 April 2024; accepted 24 April 2024. Date of publication 4 June 2024; date of current version 19 June 2024. This work was supported in part by the National Key R&D Program of China under Grant 2022YFB3902800, in part by the National Natural Science Foundation of China under Grant 62301214, in part by the Scientific Research Foundation for Doctoral Program of Hubei University of Technology under Grant XJ2022005901, and in part by the Ministry of Education Chunhui Plan Cooperation Project under Grant HZKY20220350. *(Zhiqi Zhang and Liyang Bao contributed equally to this work.) (Corresponding author: Guangqi Xie.)*

Zhiqi Zhang is with the School of Computer Science, Hubei University of Technology, Wuhan 430068, China, and also with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zzq540@hbut.edu.cn).

Liyang Bao, Guangqi Xie, and Rong Gao are with the School of Computer Science, Hubei University of Technology, Wuhan 430068, China (e-mail: 102211123@hbut.edu.cn; xieguangqi@hbut.edu.cn; gaorong@hbut.edu.cn).

Shao Xiang is with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: xiangshao@whu.edu.cn).

The code of the proposed approach can be found at https://github.com/bao11seven/B2CNet.

Digital Object Identifier 10.1109/JSTARS.2024.3409072

***Index Terms*—Bitemporal feature aggregation, change boundary-aware, change detection, remote sensing image.**

I. Introduction

CHANGE detection (CD) plays a crucial role in monitoring and analyzing temporal changes in specific geographic areas. With the advancements of remote sensing technology, the imaging capability and quality of remote sensing satellites have greatly improved [1]. High-resolution images provide valuable information about the Earth's surface and have a wide range of applications, including environmental monitoring [2], urban planning [3], disaster management [4], and natural resource management [5]. However, change detection for remote sensing images is a challenging task due to various factors, such as angle, illumination, seasonal variations and more.

Early applications can be traced back to pixel-based change detection methods [6]. Pixel-based methods can be specifically classified as algebraic-based methods, transformation-based methods, and classification-based methods. Algebra-based methods analyze the intensity and direction of changes between pixels in multitemporal images using mathematical operations, such as image difference [7], image ratio [8]. Transformation-based methods aim to extract effective difference features by transforming or analyzing the original images, like principal component analysis (PCA) [9], tasseled cap transformation [10], and change vector analysis [11]. Classification-based methods focus on classifying pixels in remote sensing images to detect changes. Machine learning techniques like support vector machines [12], random forests [13], and K-nearest neighbors [14] are commonly used as the classifier. Generally, pixel-based methods are more susceptible to factors like lighting conditions, shadows, and noise, resulting in high false detection and missed detection rates.

The rapid development of deep learning technology in recent years has demonstrated its superiority across various fields [15], [16]. Deep learning's powerful feature learning and generalization capabilities have also made it applicable for diverse remote sensing images change detection [17], [18], such as multisource, multimodal, and multitemporal. Methods based on convolutional neural network (CNN) [19], [20], [21], recursive convolutional neural network (RNN) [22], [23], graph convolutional neural network (GCN) [24], [25], and transformer-based methods [26] are commonly used in current change detection

technology. They play a crucial role in autonomous feature learning and efficient processing for large-scale remote sensing data.

Compared to traditional change detection methods, deep learning methods have higher accuracy and robustness. Deep learning models are able to automatically learn a feature representation of the data, eliminating the need to rely on domain experts to manually extract features. Recent studies, including EGRCNN [23], MFIN [27], and EGCTNet [28], have highlighted the significance of edge information in enhancing the fine-grained nature of the boundaries of change regions. These studies have successfully integrated edge features as prior knowledge in change detection networks. Edges are usually the boundaries or edge contours between objects or scenes in an image. A change region is a region that has changed in two different periods of remote sensing images. Edge information plays an important role in change detection because changes usually result in alteration of object or scene boundaries. By extracting and analyzing edge information, it can help to identify and locate change regions. Edge information can provide clues about the shape, location, and boundary characteristics of the changed region. Therefore, accurate extraction and utilization of edge information can help improve the performance of change detection and reduce false and missed detection. However, most methods often overlook the relationship between edge information and change information. These methods only use the extracted edge information as a supplement to the prediction results, omitting the role of the edge information in characterizing the change region, such as shape, location, and boundary characteristics. Furthermore, these methods always introduce the change boundary labels extracted from the labels as extra input or supervision to help the extraction of boundary features, which may bring about the loss of some internal features of the change area and reduce the change feature extraction ability of the model.

For the CD task, there are two major challenges that need to be faced at present. The first one is the problem of false detection due to pseudochanges caused by differences in illumination conditions and noise in bitemporal images. The second is the problem of missed detection such as insufficient fine-grained boundaries and internal holes caused by the loss of detail information and insufficient semantic information. We utilize the edge information more comprehensively to solve the above problems. This article proposes a change detection network based on change boundary-aware guidance. To enhance the precision of detected edges and mitigate the impact of pseudochanges, we incorporate a change boundary-aware module (CBM) branch with supervision. The proposed CBM enhances differential edge features and improves their expression. By supervising the CBM branch, the enhanced edge features effectively outline the contours of the change region, distinguishing it significantly from the background region and providing guidance for learning in other branches. To improve the completeness of internal features within the change region, we design a spatial-temporal texture feature aggregation branch based on the bi-temporal feature aggregation module (BFAM). This branch considers multiple receptive fields information and utilizes a 3-D attention mechanism to dynamically adjust the coupling of low-level and high-level semantic information. To enhance the feature extraction capability of the network, we introduce the deep feature extraction module (DFEM). The DFEM utilizes multiple residual operations to fuse the features of CBM and BFAM deeply, extracting deeper high-level change features. Moreover, it leverages the high-level features to guide the aggregation of low-level features from BFAM. The main contributions of this article are as follows.

1) We design B2CNet, a change boundary-to-center refinement change detection network. There is no need to use edge extraction algorithms and boundary labels to help extract boundary features. Just simple operations and regular labels can activate boundary features and guide the model to achieve better results.
2) For the problem of insufficient utilization of edge information and insufficient fine-grained detection boundaries, we propose a CBM branch that can roughly extract boundary information from changing regions with only simple operations. The obtained boundary information is refined using accurate change region information from the general labels, ensuring precise extraction of the change region boundary. This module improves boundary granularity and mitigates the effects of pseudochanges, and guides information aggregation in other branches.
3) For the problem of internal holes, we propose BFAM that aggregates low-level texture information in multiple receptive fields with high-level semantic information. This module enhances the internal integrity of the change region and compensates for the loss of detailed information in deeper features.
4) For the problem of insufficient high-level semantic information in complex scenarios, we propose a DFEM that extracts deeper high-level semantic information by enhancing the fusion of feature information between branches. This module enhances the feature extraction capability of the network and strengthens the feature interaction between branches to improve the efficiency of decoupling.

The rest of this article is organized as follows. Section II provides a brief review of related work. Section III describes the specific details of the proposed framework. Section IV presents the evaluation results on public datasets and compares them with current state-of-the-art algorithms. Section V discusses the proposed approach, and finally, Section VI concludes this article.

II. Relate Work

A. Deep-Learning-Based CD Methods

Deep learning networks have a hierarchical architecture. This characteristic has led to the use of Siamese network architectures for change detection tasks. Including skip connections [29] in these architectures enhances prediction accuracy. For example, skip-connected fully convolutional networks [19] (FC-EF [19], FC-Siam-Diff, and FC-Siam-Conc) extract deep information and perform predictions through distinct approaches like spatial-temporal features, feature differencing, and feature splicing respectively. Fang et al. [30] propose a integrates dense skip connections Siamese network (SNUNet), which alleviates the

loss of deep and localized information through the integration of multiscale features. Several studies have proposed different models to enhance change detection performance. For instance, Jiang et al. [31] propose a joint learning framework (SSANet), which incorporates fusion and difference extraction branches to enhance contextual information aggregation. Wen et al. [32] developed the RS-CADM model, which combines denoising diffusion probabilistic models (DDPMs) and adaptive calibration techniques to enhance the extraction of change information. Dual-task constrained convolutional network models [33], [34] have also been proposed to accomplish change detection and semantic segmentation simultaneously. Additionally, researchers used CNNs and transformers individually or jointly. For example, Chen et al. [35] propose a bitemporal image transformer network (BIT), which captures and models spatial-temporal contextual information between two different time-phase images, understanding the correlation between local content and the global scene. Feng et al. [36], Tang et al. [37], and Ji et al. [38] utilize CNN and transformer as feature extraction networks (ICIFNet, WNet, and PASSNet), which facilitate the interaction between locally extracted CNN features and globally extracted transformer features to enhance detail information while preserving their own features.

However, the utilization of transformer-based change detection (CD) methods has led to an increase in time complexity. Moreover, CD methods that integrate edge information often overlook the ability of edge information to characterize change regions and the information loss within change regions due to boundary extraction. This study seeks to tackle these challenges by leveraging edge-enhanced difference feature and incorporating multiple receptive fields spatial-temporal feature.

B. Attention-Mechanism-Assisted CD Methods

Attention mechanisms [39] have gained significant attention in various visual tasks, including image fusion [40], [41], detection [42], [43], and semantic segmentation [44], highlighting their importance in the field. In CD tasks, traditional methods primarily focus on pixel-level comparisons, neglecting semantic relationships and contextual information between pixels, resulting in low detection accuracy. However, the ability of attention mechanisms to capture positional relationships within input sequences, which presents a new opportunity for remote sensing change detection. It can effectively capture global relationships and focus limited energy on important locations from a global perspective, resulting in more useful information. There are three common basic attention mechanisms, namely self-attention [45], spatial attention mechanism (SAM) [46], and channel attention mechanism (CAM) [47]. Currently, many change detection networks are based on the enhanced version of the basic attention mechanisms to further emphasize the change features. For instance, SNUNet [30] proposes ECAM to automatically select and focus on more effective information between different groups. DSIFNet [48] alternates between CAM and SAM to refine fused features from the channel-space dimension. Cross-attention mechanisms have also been widely adopted in change detection, which considers the relationship between different input sequences based on self-attention. Researchers have explored combining self-attention and cross-attention mechanisms with CNNs to extract more accurate features for remote sensing change detection. HANet [49] implements an HAN module, which is capable of capturing long-term dependencies separately from the column and row dimensions. DMINet [50] combines self-attention and cross-attention into a single module to obtain a global attention distribution for information interaction. Inspired by the self-attention mechanism, MFIN [27] designs the feature interaction module to interactively process the feature information of the two images. SARASNet [51] employs self-attention and cross-attention features, effectively detecting changes caused by objects of different scales. HDANet [52] introduces a multiresolution parallel structure and designs an innovative differential attention module to preserve spatial and change information. In addition, SAM and CAM focus on a single dimension which might lack a fine-grained focus on features. Inspired by neuroscience theories, Yang et al. [53] design a parameter-free 3-D attention mechanism (SimAM), which optimizes the energy function to determine the importance of each neuron and achieve more accurate feature focusing. In this article, we introduce the SimAM attention mechanism to enhance feature granularity.

C. Edge-Assisted CD Methods

The task of refined change detection is a hot topic at the moment, and the finegrained boundary of change detection is an ongoing problem. To further mitigate this problem, many domain experts have carried out work on the extraction and retention of boundary features. Common edge extraction algorithms include Canny algorithm, Sobel algorithm, and Laplacian algorithm. Zhang et al. [27] design a detail feature guidance module, which uses the Laplacian operator to extract multiscale edge features from labels and insert them into the backbone network for learning detail features. Another method is to enhance the preservation of edge features of the CD network. Bai et al. [23] propose edge-guided change map estimation, which not only predicts changed buildings and their edges but also integrate edge features directly into discriminative features to further improve the quality of the prediction results. Xia et al. [28] design an edge detection branch and uses the edge change map to constrain the output change mask.

III. Methodology

In this section, we first present an overview of the proposed method. Then, we give the details of the proposed CBM, BFAM, and DFEM. Finally, we provide the hybrid loss function.

A. Overall Structure of B2CNet

As shown in Fig. 1, our proposed B2CNet is based on the encoder–decoder architecture, which consists of four components, a multilevel feature extraction, a change boundary-aware module (CBM), a BFAM, and a DFEM. Following Siamese network approach, the encoder employs a pretrained CNN model (ResNet18 [29]) as the feature extraction network. As for the

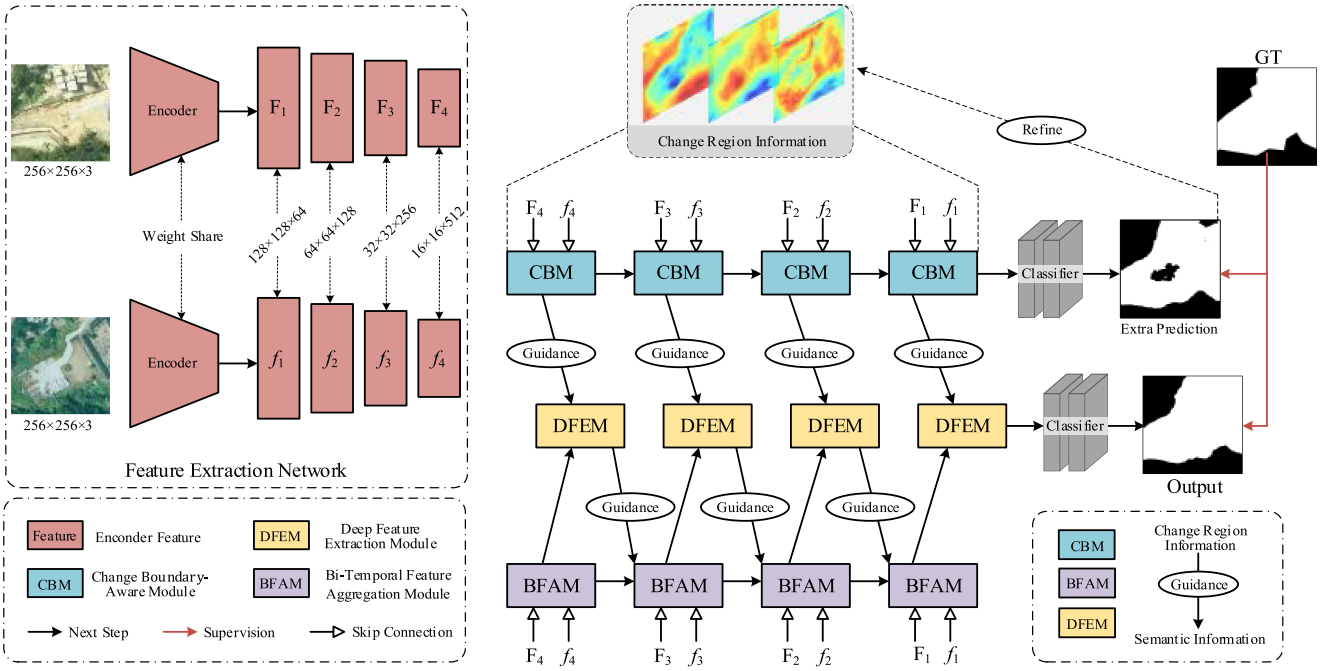

Fig. 1. Architecture of the proposed B2CNet.

decoder, it consists of three branches of CBM, BFAM, and DFEM, as shown the right panel in Fig. 1.

First, by feeding two remote sensing images at different periods into the feature extraction network, we can obtain feature vectors $\{F_n, f_n,\ n \in (1,2,3,4)\}$ at different scales, as shown the left panel in Fig. 1. Let the output feature vectors of the CBM, BFAM, and DFEM be the $\{f^i_{\text{cbm}}, f^i_{\text{bfam}}, f^i_{\text{dfem}}, i \in (1,2,3,4)\}$. We input feature vector F_4, f_4 into the CBM to obtain the edge-enhanced difference feature f^1_{cbm}. Subsequently, the feature $\{F_n, f_n, (n=1,2,3), f^{i-1}_{\text{cbm}}\}$ is used as input for CBM $\{f^i_{\text{cbm}}, (i=2,3,4)\}$. We extracted the edge-enhanced difference feature by CBM, which has a stronger representation of edge information. Thus, it can promote the retention of change edge information and better characterize the change region. Finally, the CBM branch passes the classifier to obtain extra predicted results. The expression of edge features in the CBM branch is stronger, resulting in extra prediction that focus more on edge prediction. In addition, it retains some internal information of the changed area, reducing the loss of valuable information compared to retaining only edge information. Extra predictions are supervised, and the difference between the changing region information and the ground truth is calculated through a loss function. Backpropagation is performed on the network based on supervision, and the gradient of the generated changing region information is updated. Through iterative optimization in the training phase, the generated results are gradually improved, and the ability to locate changed areas and express the characteristics of changed areas is enhanced. It prompts the CBM branch to focus on more accurate change features, guides other branches to better decouple the change region information, and deepens the decoupling of the overall network. Input F_4, f_4 into the BFAM to obtain the texture detail aggregation feature f^1_{bfam} of the multiple receptive fields and the input features for the BFAM $\{f^i_{\text{bfam}}, i \in (2,3,4)\}$ contain features $\{F_n, f_n, (n=1,2,3), f^{i-1}_{\text{dfem}}, f^{i-1}_{\text{bfam}}\}$. The high-level semantic features are used to guide the transition from low-level information to high-level information. For the DFEM branch, input features $\{f^i_{\text{cbm}}, f^i_{\text{bfam}}, i \in (1,2,3,4)\}$ to obtain high-level change semantic features $\{f^i_{\text{dfem}}, i \in (1,2,3,4)\}$ and strengthen the feature extraction capability of the network. Finally, this branch outputs the final result of the model through a classifier, and GT is used to supervise the final prediction result during the training stage.

The complete structure of the above is B2CNet. In addition, we propose B2CNet_S, which is a lightweight version. B2CNet_S removes the features $\{F_4, f_4, f^1_{\text{cbm}}, f^1_{\text{bfam}}, f^1_{\text{dfem}}\}$.

B. Change Boundary-Aware Module

The combination of edge information and feature difference operations can effectively depict the boundaries and contours of change regions. To achieve this, we propose a CBM that acquires change regions using boundary information and facilitates the retention of change region boundary information. The module structure is depicted in Fig. 2.

Edge information serves as valuable prior knowledge to assist change detection by capturing the fine structure and complex shape of objects. However, extracting edge information without differentiation may introduce irrelevant information in non-change regions. Conversely, considering only edge information in change regions may result in the loss of internal features. To alleviate this, we propose extracting and enhancing the boundary information of bi-temporal features before performing feature difference. This approach allows us to capture both the

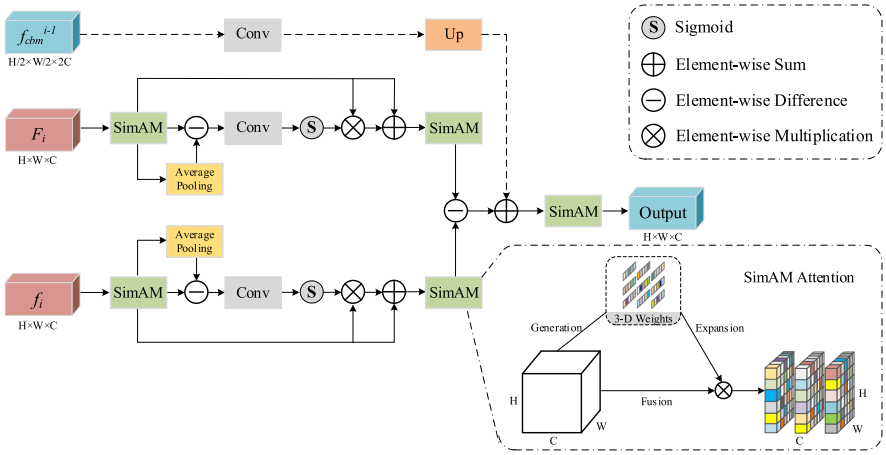

Fig. 2. Details of the CBM.

difference in boundary information and the difference features through feature difference.

In our approach, we input features (f_i) from the features extraction network into the CBM to extract and enhance the edge information of the bitemporal features. Initially, the first-level input features are preattended using the SimAM attention [53] mechanism to identify noteworthy regions. The SimAM attention mechanism is a parameter-free 3-D attention mechanism based on some well-known neuroscience theories and proposes to optimize the energy function to discover the importance of each neuron. To improve attention, the model needs to evaluate the importance of each neuron. Information-rich neurons often exhibit different firing patterns than surrounding neurons and inhibit the latter. The importance can be judged by measuring the linear separability between neurons. Therefore, we need to pay more attention to neurons with spatial inhibitory effects. Its minimum energy function is defined as follows:

$$e_t^* = \frac{4\left(\hat{\sigma}^2 + \lambda\right)}{(t-\hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \tag{1}$$

$$\hat{\mu} = \frac{1}{M}\sum_{i=1}^{M} x_i \tag{2}$$

$$\hat{\sigma}^2 = \frac{1}{M}\sum_{i=1}^{M} (x_i - \hat{\mu})^2 . \tag{3}$$

In the above equation, $\hat{\mu}$ and $\hat{\sigma}^2$ are the mean and variance of all neurons except t. The lower the energy, the greater the difference and importance of neuron t from surrounding neurons. After obtaining the minimum value through (1), the importance of the neuron can be determined by $1/E$. After evaluating the importance of neurons, the feature matrix should be improved according to the following equation:

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{E}\right) \times X \tag{4}$$

where E groups all e_t^* in the channel and spatial dimensions.

Subsequently, edge features are extracted using pooling, subtraction, and convolution. This is because edges typically exhibit high gradient values, and applying average pooling helps to smooth out the features. The subtraction operation, which involves subtracting the average pooled features from the original features, highlights the edge regions. Subsequently, the features undergo 1×1 convolution with a sigmoid activation function to enhance the contrast and saliency of the edge features. The edge features are then re-enhanced through multiplication and addition. Finally, the SimAM attention mechanism is employed to sense the significant regions once again, resulting in the final features $\hat{f}_e^t$. The equations for these operations are as follows:

$$\hat{f} = \text{SimAM}\left(f_i\right) \tag{5}$$

$$f_e^t = \text{Conv}_{1\times1}\left(\hat{f} - \text{AP}\left(\hat{f}\right)\right) \times \hat{f} + \hat{f} \tag{6}$$

$$\hat{f}_e^t = \text{SimAM}\left(f_e^t\right) \tag{7}$$

where SimAM denotes the attention mechanism, AP denotes the 3×3 average pooling layer, and here Conv_{1X1} denotes the convolutional block (1×1 convolution, BN, Sigmoid) using Sigmoid as the activation function.

Feature difference, as a direct operation to derive difference features, is susceptible to noise and information loss. In change detection tasks, it can lead to blurred edges of the change region due to the inherent similarity of bitemporal features. To alleviate these issues, we enhance the edges of bitemporal features before performing feature difference. This edge-enhanced feature difference provides richer contextual information and aids in localizing the change region. We repeat this operation for features at different levels, obtaining edge-enhanced features $(\hat{f}_e^1, \hat{f}_e^2)$ at each level. Subsequently, the edge-enhanced features are processed for feature differences. We recalculate the value information using SimAM. Equations are as follows:

$$f_{\text{cbm}}^i = \begin{cases} \text{SimAM}\left(\text{Abs}\left(\hat{f}_e^1 - \hat{f}_e^2\right)\right) & i = 1 \\ \text{SimAM}\left(\text{Abs}\left(\hat{f}_e^1 - \hat{f}_e^2\right)\right. & \\ \quad \left. + \text{Up}\left(\text{Conv}_{1\times1}\left(f_{\text{cbm}}^{i-1}\right)\right)\right) & i = 2,3,4 \end{cases} \tag{8}$$

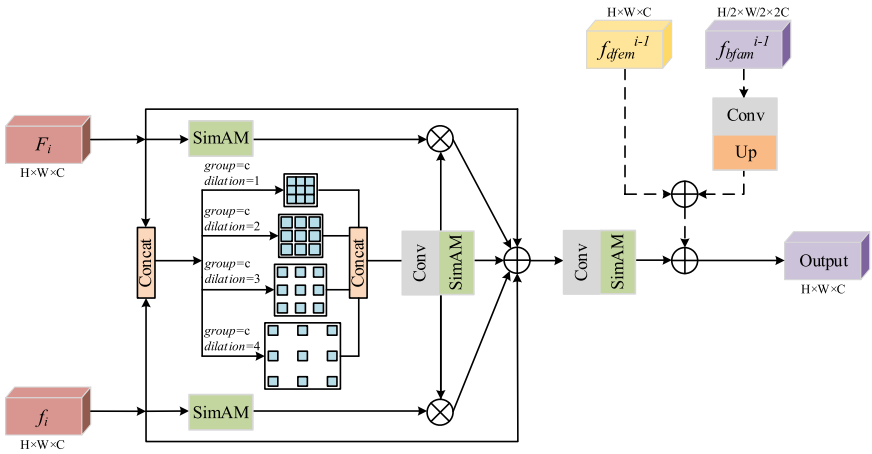

Fig. 3. Details of the BFAM.

$$\text{Prediction} = \text{Conv}_{3\times3}\left(\text{Relu}\left(BN\left(\text{Conv}_{3\times3}\left(f_{\text{cbm}}^{4}\right)\right)\right)\right) \quad (9)$$

where i denotes features at different depths, Abs represents taking absolute values, and Up represents upsampling.

Of course, the change region localized by the above approach is not completely accurate. Therefore, we employ supervision on the decoupling branch, which consists of CBM. This helps correct the localization information, ensuring accurate identification of the change region. Furthermore, we utilize the localized change region as a guide for the decoupling process in other branches.

C. Bitemporal Feature Aggregation Module

Low-level texture information encompasses fundamental image details and local characteristics, such as texture, color, and grayscale. This information captures small changes within an image. On the other hand, bitemporal features exhibit rich details while also preserving spatial relationships due to their inherent nature. To enhance the comprehensiveness of change detection, we propose a BFAM, which progressively combines low-level detail features and multilevel features. The process is illustrated in Fig. 3.

To extract detailed features and maintain spatial relationships, we perform channel splicing on the input bitemporal features (F_i, f_i). We utilize four parallel dilated group convolutions (3×3) with different dilation rates (dilation = 1,2,3,4, group = c) to extract features, where c represents the number of channels. This approach allows for capturing change regions of various sizes through different receptive fields while preserving the spatial integrity of the features using grouped convolutions. Subsequently, the four convolutional outputs are channel concatenated, and a 1×1 convolutional block is employed for channel downsampling. This step further refines the features using the SimAM attention mechanism [53]. Equations are as follows:

$$f_{3\times3}^{d,g} = \text{Conv}_{3\times3}^{d,g}\left(\text{Concat}\left(F_i, f_i\right)\right) \qquad d \in \{1,2,3,4\}, g = c \quad (10)$$

$$f_{cat} = \text{Conv}_{1\times1}\left(\text{Concat}\left(f_{3\times3}^{1,c}, f_{3\times3}^{2,c}, f_{3\times3}^{3,c}, f_{3\times3}^{4,c}\right)\right) \quad (11)$$

$$\hat{f}_{\text{cat}} = \text{SimAM}\left(f_{\text{cat}}\right) \quad (12)$$

where d denotes dilation rate, g denotes group and Concat denotes channel concatenated.

Considering the shared characteristics present in the bitemporal features. It provides more precise low-level texture information. We calculate the significance of pixel-level features from different temporal features using the SimAM attention mechanism individually. Then, we multiply the extracted commonality feature $(\hat{f}_{\text{cat}})$ with the respective temporal features (F_i, f_i) to obtain their similarity. Equations are as follows:

$$\hat{F}_i = \text{SimAM}\left(F_i\right) \times \hat{f}_{\text{cat}} \quad (13)$$

$$\hat{f}_i = \text{SimAM}\left(f_i\right) \times \hat{f}_{\text{cat}}. \quad (14)$$

Next, we sum the low-level detail features $(\hat{F}_i, \hat{f}_i, \hat{f}_{\text{cat}})$ and the advanced change feature (f_{dfem}^{i-1}). The final aggregated feature is obtained through the SimAM attention mechanism. As the CNN model deepens, there is a gradual loss of low-level information. To ensure that we preserve sufficient texture information even in the deeper layers of the network, we incorporate residual connections. Equations are as follows:

$$f_{\text{bfam}}^{i} = \begin{cases} \text{SimAM}\left(\text{Conv}_{3\times3}\left(\hat{F}_i + \hat{f}_i + \hat{f}_{\text{cat}}\right)\right) & i = 1 \\ \text{SimAM}\left(\text{Conv}_{3\times3}\left(\hat{F}_i + \hat{f}_i + \hat{f}_{\text{cat}} + f_{\text{dfem}}^{i-1}\right)\right) & \\ \quad + Up\left(\text{Conv}_{1\times1}\left(f_{\text{bfam}}^{i-1}\right)\right) & i = 2,3,4. \end{cases} \quad (15)$$

Finally, we integrated the multilevel features by utilizing a BFAM. The purpose of this integration was to enhance the spatial texture information within the change region by incorporating multiple receptive fields. Furthermore, we combined these features with high-level semantic information to obtain a more detailed and comprehensive understanding of the change.

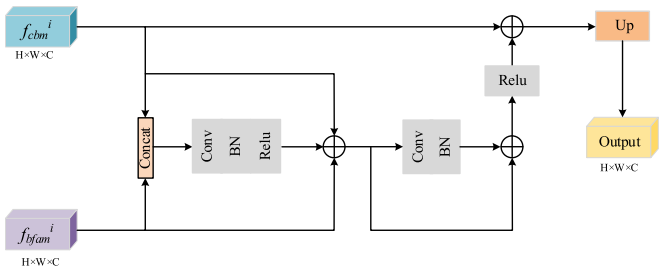

Fig. 4. Details of the DFEM.

D. Deep Feature Extraction Module

High-level semantic features play a crucial role in accurately locating and identifying change areas. On the other hand, detailed texture features provide more precise boundary and texture information. Our proposed DFEM aims to combine the two to learn deeper high-level semantic features, as depicted in Fig. 4.

The DFEM integrates the change region information obtained from the CBM and the spatial detail texture information from the BFAM. We perform concatenation and summation of the input features from the upper-level CBM (f_{cbm}^{i}) and the BFAM (f_{bfam}^{i}), respectively. To reduce computation, we employ 1×1 convolution to reduce the number of channels by half. Additionally, to preserve information completeness, we use residual concatenation to multiply the combined information with it. Next, we extracted the depth features using a 3×3 convolutional block. To minimize information loss, we opted for residual concatenation before applying Relu.

Finally, we sum the output features (f_{cbm}^{i}) of the previous level CBM with the deep features. The complete features of the change boundary and the high-level semantic features are then passed into the BFAM. This ensures that the change boundary and high-level features are guided by the aggregation module. Simultaneously, it minimizes feature information loss through multiple residual operations within the module. Equations are as follows:

$$f_c^i = \text{Conv}_{1\times1}\left(\text{Concat}\left(f_{\text{bfam}}^i, f_{\text{cbm}}^i\right)\right) + \left(f_{\text{bfam}}^i + f_{\text{cbm}}^i\right) \quad i = 1,2,3,4 \tag{16}$$

$$f_{\text{dfem}}^i = \text{Relu}\left(\text{BN}\left(\text{Conv}_{3\times3}\left(f_c^i\right)\right) + f_c^i\right) + f_{cbm}^i \quad i = 1,2,3,4 \tag{17}$$

$$\text{Output} = \text{Conv}_{3\times3}\left(\text{Relu}\left(\text{BN}\left(\text{Conv}_{3\times3}\left(\text{f}_{\text{dfem}}^4\right)\right)\right)\right) \tag{18}$$

where i denotes the features of different depths, Conv, BN, and Relu represent a convolutional layer, batch normalization, and Relu activation function. f_c^i is the intermediate operation of the channel dimension transformation, and f_{dfem}^i is the extracted high-level features.

Therefore, we integrate the extracted high-level semantic features with the low-level detailed texture features obtained from the BFAM. This integration aims to enhance the image characterization and improve the understanding and analysis of complex scenes within the image.

E. Loss Function

In the domain of change detection, there is a significant imbalance between the number of unchanged pixels and the number of pixels that have changed. To attenuate the effect of sample imbalance, we use a hybrid loss function [54], a combination of weighted cross entropy loss and dice loss, defined as follows:

$$L = L_{\text{wce}} + L_{\text{dice}} \tag{19}$$

$$L_{\text{wce}} = \frac{1}{W \times H} \sum_{i=1}^{H\times W} \text{weight}\left[\text{class}\right] \cdot \left(\log\left(\frac{\exp\left(\hat{y}\left[i\right]\left[\text{class}\right]\right)}{\sum_{l=0}^{1}\exp\left(\hat{y}\left[i\right]\left[l\right]\right)}\right)\right) \tag{20}$$

$$L_{\text{dice}} = 1 - \frac{2 \cdot S_o\left(\hat{Y}\right)}{Y + S_o\left(\hat{Y}\right)} \tag{21}$$

where L_{wce} represents weighted cross-entropy loss,L_{dice} represents dice loss, weight represents weight, and class has a value of 1 or 0 (corresponding to changing and unchanging pixels, respectively),$\hat{y}[i]$ represents the i th point in $\hat{Y}$, i and l are indexes, $\hat{Y} = \{\hat{y}[i], i = 1, 2, \ldots, W \times H\}$ represents the change map, and $\hat{Y}$ represents the ground truth.

Our network architecture is guided by CBM for change detection. Therefore, the accuracy of the CBM branch directly impacts the subsequent depth change feature extraction. To alleviate this, we employ the DFEM branch for feature extraction and through the prediction header to generate predictions. Besides, we provide additional supervision for the CBM branch. We utilize complete ground truth for supervision, which includes supervising the boundaries of the change region. Stronger expression of edge information in the CBM branch contributes to accurate edge prediction. Moreover, complete supervision promotes the retention of internal feature information in the change region, minimizing information loss. Thus, our total training loss can be expressed as follows:

$$L_{\text{total}} = \lambda_1 L\left(GT, \text{Output}\right) + \lambda_2 L\left(GT, \text{Prediction}\right) \tag{22}$$

where λ_1 and λ_2 are the tradeoff parameter used to regulate the impact of each loss, which we set $\lambda_1 = 1,\ \lambda_2 = 0.5$.

IV. Experiment and Results

A. Datasets

1) LEVIR-CD [55]: The large-scale building change detection dataset used in our study consists of 637 pairs of high-resolution (0.5 m/pixel) remote sensing images. Each image has a size of 1024×1024. The study area covers 20 different urban areas in the state of Texas, USA, spanning from the year 2002 to 2018, with a specific focus on building changes. The dataset includes various types of buildings and land cover changes, such as villas, high-rise apartments, small garages, and large warehouses. During training, we divided the 1024×1024 images into smaller images of size 256×256. In the end, we

obtained a training set of 7120 image pairs, a validation set of 1024 image pairs, and a test set of 2048 image pairs.

2) WHU-CD [56]: This dataset primarily focuses on the analysis of building damage in the city of Christchurch, New Zealand, following the earthquake in 2012. The dataset consists of a pair of high-resolution (0.2 m/pixel) remote sensing images and the original image size is 32 507 × 15 354. In our experiment, we divide the image into 256 × 256 size image blocks and the division process has no duplicate regions The ratios of training, verification and testing are 8:1:1, respectively. In the end, we obtained a training set of 6096 image pairs, a validation set of 762 image pairs, and a test set of 762 image pairs.

3) SYSU-CD [57]: The SYSU dataset used in our study consists of 20 000 pairs of high-resolution (0.5 m/pixel) remote sensing images. Each image has a size of 256 × 256. The dataset includes ground objects of high-rise buildings, harbor, and suburban variations from Hong Kong area. During training, we utilize the standard data segmentation provided by the researchers. The ratios of training, verification and testing are 6:2:2, respectively. In the end, we obtained a training set of 12 000 image pairs, a validation set of 4000 image pairs, and a test set of 4000 image pairs.

4) HRCUS-CD [58]: The high-resolution complex urban scene BCD (HRCUS-CD) dataset integrates multiple time spans and multiple building change types. The time span is from 2019 to 2022 and 2010 to 2018, respectively. This dataset has a large sample size and includes a variety of complex environmental scenarios such as urban villages, vegetation disturbance, high-rise apartments, industrial parks, cultural tourism facilities, and other large contiguous buildings. The dataset contains cropped 11 388 pairs of high-resolution RSIs at 0.5 m resolution at 256 × 256 pixels, and over 12 000 labeled variation instances.

B. Evaluation Metrics

To validate the performance of the proposed network, we use six indicators to assess the similarity between the predicted results and the true variation, including precision (Pre), recall (Rec), F1 Score (F1), intersection over union (IOU), overall accuracy (OA) and Kappa coefficient (Kappa). The metrics can be individually defined as follows:

$$\text{Pre} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{23}$$

$$\text{Rec} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{24}$$

$$\text{F1} = \frac{2\text{Pre}\cdot\text{Rec}}{\text{Pre}+\text{Rec}} \tag{25}$$

$$\text{IOU} = \frac{\text{TP}}{\text{TP}+\text{FP}+\text{FN}} \tag{26}$$

$$\text{OA} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{27}$$

$$\text{Kappa} = \frac{\text{OA}-P}{1-P} \tag{28}$$

where TP, FP, TN, and FN indicate the number of true positive, false positive, true negative, and false negative, respectively.

P in Kappa denotes the hypothesized probability of chance agreement between the reference and predicted values, which can be expressed as follows:

$$P = \frac{(\text{TP}+\text{FP})(\text{TP}+\text{FN})+(\text{FN}+\text{TN})(\text{TP}+\text{TN})}{(\text{TP}+\text{FP}+\text{TN}+\text{FN})^2}. \tag{29}$$

C. Implementation Details

To ensure fairness, all comparative methods were trained and tested on the same device. Our proposed network model is structured in the pytorch architecture, which is trained and tested on an NVIDIA GeForce GTX3090 GPU with 24 GB memory. AdamW optimizer [59] is used to minimize the loss, weight decay is set to 5e–4. The learning rate was updated using StepLR, with an initial learning rate ($\text{initial}_{\text{lr}}$) of 5e–4, a step size of 8, and a gamma (γ) of 0.5, as shown in (30). Batch size is 16, we train the model 100 epochs until convergence

$$\text{New}_{\text{lr}} = \text{initial}_{\text{lr}} \cdot \gamma^{\text{epoch}//\text{step size}}. \tag{30}$$

D. Comparison With the State-of-the-Art Methods

To comprehensively evaluate the performance of our proposed model, we compare it with three categories of state-of-the-art change detection methods. These categories include convolutional neural network-based methods (FC-EF [19], FC-Siam-Conc [19], FC-Siam-Diff [19], and FresUNet [33]), attention mechanism-based methods (DSIFN [48], SNUNet [30], ICIFNet [36], DMINet [50], and CGNet [60]), and transformer-based method (BIT [35]). The specifical details are as follows.

1) *FC-EF [19]:* A single-stream network model based on the fully convolutional advance fusion of the UNet structure, which is combined through skip connections to obtain semantic information and spatial details at the same time.
2) *FC-Siam-Diff [19]:* The model framework based on Siamese network directly calculates the absolute value of the same-scale feature difference in the encoder, and predicts the changed area based on the concatenation of differences at all levels.
3) *FC-Siam-Conc [19]:* A framework based on Siamese network, which fuses the features of two-phase images through channel connection, and finally predicts the change area based on the fused features.
4) *FresUNet [33]:* A fully convolutional dual-task change detection network not only detects the changed areas, but also further detects the semantics of the changed areas. On the basis of FC-EF, residual blocks are added for the encoder–decoder architecture with skip connections, which promotes the network to obtain deeper feature information.
5) *DSIFN [48]:* A deeply supervised Siamese network. Perfectly combine heterogeneous features from the perspective of channels and space for differential discrimination. Further enhance difference discrimination ability through in-depth supervision.
6) *SNUNet [30]:* A dense skip-connected Siamese network, which incorporates multiscale features to alleviate the

TABLE I
COMPARISON RESULTS ON THE THREE TEST SET

Methods	LEVIR-CD	WHU-CD	SYSU-CD
	Pre / Rec / F1 / IOU / OA / Kappa	Pre / Rec / F1 / IOU / OA / Kappa	Pre / Rec / F1 / IOU / OA / Kappa
FC-EF$_{18}$	78.31 / 73.56 / 75.86 / 61.11 / 97.61 / 74.61	67.37 / 78.76 / 72.62 / 57.01 / 97.72 / 71.44	72.50 / 78.25 / 75.27 / 60.34 / 87.87 / 67.25
FC-Siam-Diff$_{18}$	83.87 / 76.64 / 80.09 / 66.80 / 98.06 / 79.07	74.54 / 76.19 / 75.35 / 60.45 / 98.08 / 74.36	**85.76** / 58.32 / 69.42 / 53.17 / 87.89 / 62.21
FC-Siam-Conc$_{18}$	82.65 / 79.72 / 81.16 / 68.29 / 98.11 / 80.17	73.93 / 76.84 / 75.36 / 60.46 / 98.07 / 74.35	72.24 / 81.42 / 76.55 / 62.01 / 88.24 / 68.74
FresUNet$_{19}$	87.54 / 85.89 / 86.71 / 76.54 / 98.66 / 86.00	83.36 / 78.45 / 80.83 / 67.83 / 98.57 / 80.09	75.84 / **84.07** / 79.74 / 66.31 / 89.93 / 73.06
DSIFNet$_{20}$	89.93 / **90.89** / 90.41 / 82.50 / 99.02 / 89.89	91.45 / 85.29 / 88.26 / 78.99 / 99.13 / 87.81	**82.81** / 73.87 / 78.08 / 64.05 / 90.22 / 71.82
SNUNet$_{22}$	91.34 / 88.63 / 89.96 / 81.76 / 98.99 / 89.43	88.86 / 87.67 / 88.26 / 78.99 / 99.10 / 87.80	77.32 / 80.11 / 78.69 / 64.87 / 89.77 / 71.96
BIT$_{22}$	91.51 / 88.05 / 89.75 / 81.40 / 98.98 / 89.21	92.96 / 88.04 / 90.43 / 82.53 / 99.28 / 90.06	79.86 / 76.99 / 78.40 / 64.47 / 89.99 / 71.89
ICIFNet$_{22}$	91.41 / 89.02 / 90.20 / 82.15 / 99.01 / 89.68	**94.90** / 87.63 / 91.12 / 83.69 / 99.34 / 90.78	78.44 / 78.40 / 78.42 / 64.50 / 89.82 / 71.76
DMINet$_{23}$	**92.12** / 89.29 / 90.68 / 82.95 / 99.07 / 90.19	**95.61** / 88.76 / **92.06** / **85.28** / **99.41** / **91.75**	**82.24** / 81.47 / **81.85** / **69.28** / **91.48** / **76.29**
CGNet$_{23}$	**92.95** / 90.12 / **91.51** / **84.35** / **99.15** / **91.06**	93.65 / **89.16** / 91.35 / 84.08 / 99.35 / 91.02	80.26 / 79.18 / 79.72 / 66.28 / 90.50 / 73.52
B2CNet	91.27 / **91.01** / **91.14** / **83.72** / **99.10** / **90.66**	**93.70** / **90.97** / **92.31** / **85.73** / **99.42** / **92.01**	81.80 / **84.71** / **83.23** / **71.28** / **91.95** / **77.94**
B2CNet_S	**91.55** / **90.25** / **90.90** / **83.31** / **99.08** / **90.41**	93.53 / **90.47** / **91.97** / **85.14** / **99.39** / **91.66**	77.53 / **85.48** / **81.31** / **68.51** / **90.73** / **75.17**

All the scores are described in percentage (%).

loss of deep local information in neural networks. The channel attention mechanism is used to enhance the expression of value features.

7) *BIT [35]:* A bit-time image transformer network, which introduces a transformer to effectively capture and model the spatial-temporal context information between two different phase images, and understand the relationship between the local content of the image and the global scene.
8) *ICIFNet [36]:* CNN and transfomer are used as feature extraction networks respectively, and the local features extracted by CNN interact with the global features extracted by transfomer, which enhances the detailed information while retaining its own characteristics.
9) *DMINet [50]:* A Siamese network that interacts bit-temporal features before obtaining differential features, which combines self-attention and cross-attention to guide the global feature distribution of each input. Facilitate information coupling between intrinsic hierarchical representations.
10) *CGNet [60]:* A method to solve the problem of insufficient expression of change features by the traditional U-Net structure, which uses deep features to generate change maps as prior knowledge to guide multiscale feature fusion.

The quantitative results of our method compared with the state-of-the-art methods are presented in Tables I and II. Our B2CNet achieves considerable results on four datasets, which can be seen from the higher F1, IOU, OA, and Kappa metrics. Furthermore, it is well known that the number of feature channels grows exponentially as the model encoding deepens, which leads to a significant increase in model parameters. To alleviate this issue, we made some modifications to improve the model efficiency. Specifically, we removed features with a feature channel of 512 from the final encoder layer and adjusted the decoupling part accordingly, eliminating the first stage. Consequently, the proposed three modules were collectively defined as one stage, resulting in a lightweight version of our model. Importantly, B2CNet_S also achieves excellent performance on all four datasets. Qualitatively, Figs. 5–7 show the qualitative visualization results with our method and existing state-of-the-art methods. For convenience, several colors are used to facilitate a clearer visualization of results, where TP (white) is the true positive, TN (black) is the true negative, FP (red) represents false positive, and FN (green) denotes the false negative.

TABLE II
COMPARISON RESULTS ON THE HRCUS-CD TEST SET

Methods	HRCUS-CD
	Pre / Rec / F1 / IOU / OA / Kappa
FC-EF$_{18}$	37.92 / 51.66 / 43.74 / 27.99 / 97.51 / 42.50
FC-Siam-Diff$_{18}$	46.97 / 37.88 / 41.94 / 26.53 / 98.04 / 40.95
FC-Siam-Conc$_{18}$	39.16 / 44.98 / 41.87 / 26.48 / 97.66 / 40.68
FresUNet$_{19}$	48.05 / 67.64 / 56.19 / 39.07 / 98.02 / 55.21
DSIFNet$_{20}$	65.12 / 65.20 / 65.16 / 48.32 / 98.69 / 64.49
SNUNet$_{22}$	65.73 / 66.75 / 66.24 / 49.52 / 98.73 / 65.59
BIT$_{22}$	65.08 / 46.70 / 54.38 / 37.35 / 98.53 / 53.66
ICIFNet$_{22}$	**70.06** / **66.98** / 68.49 / 52.08 / **98.85** / 67.90
DMINet$_{23}$	**76.91** / 65.97 / **71.02** / **55.07** / **98.99** / **70.51**
CGNet$_{23}$	65.19 / 59.19 / 62.04 / 44.97 / 98.64 / 61.36
B2CNet	**69.64** / **74.03** / **71.77** / **55.97** / **98.91** / **71.21**
B2CNet_S	66.73 / **72.40** / **69.45** / **53.20** / 98.81 / **68.84**

1) Experimental Results on LEVIR-CD Dataset: Table I displays the results of our proposed B2CNet method and several comparative experiments on the LEVIR-CD dataset. The table shows that B2CNet performs second only to CGNet, achieving considerable results on five evaluation metrics including Rec, F1, IOU, OA, and Kappa. The lightweight model also maintains excellent scores on all metrics. Specifically, B2CNet shows relative improvements in IOU compared to DSIFNet, SUNNet, BIT,

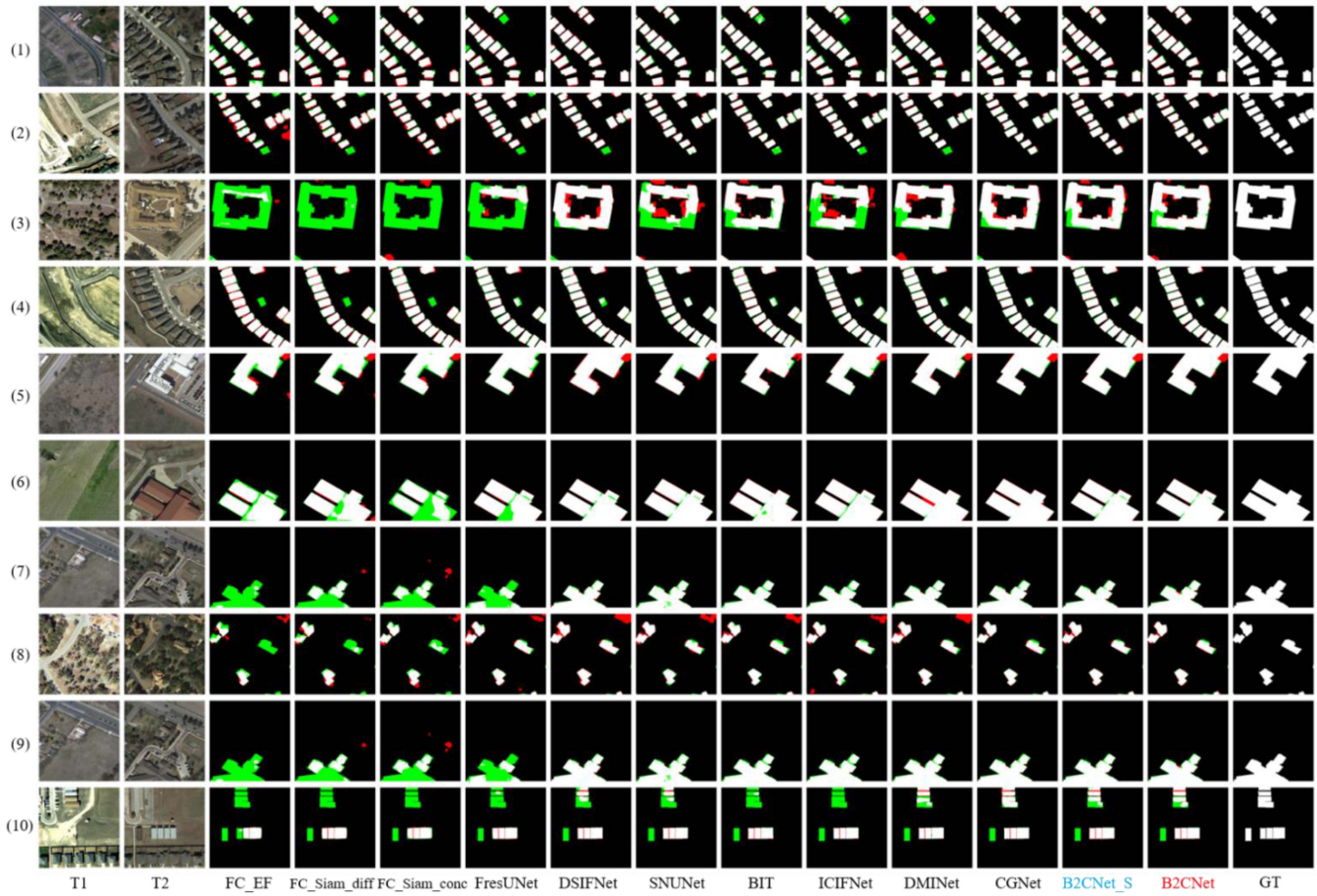

Fig. 5. Qualitative results on LEVIR-CD.TP (white), TN (black), FP (red), and FN (green).

ICIFNet, and DMINet, with improvements of 1.22, 1.96, 2.32, 1.57, and 0.77, respectively. The lightweight version B2CNet_S also demonstrates improvements of 0.81, 1.55, 1.91, 1.16, and 0.36, respectively. These results demonstrate that, compared to other experimental methods, proposed methods exhibit considerable performance in CD.

Fig. 5 shows the results attained by different methods on the LEVIR-CD dataset. Specifically, Fig. 5(1)–(4) show buildings that have undergone significant or intensive changes, and Fig. 5(5)–(7) show large buildings affected by lighting conditions or occlusions. Furthermore, Fig. 5(8)–(10) shows new buildings on original foundations. In Fig. 5(1)–(4), other methods can obtain a rough change area for changes in small buildings, but suffer a lot of loss of edge information. On the contrary, the prediction results of our method can not only effectively detect small buildings, but also retain more detailed information, and the boundaries of the changed areas are more complete. In Fig. 5(5)–(7), it can be seen that for large building changes, the prediction results of most methods produce varying degrees of loss of the shaded portion due to the light angle. However, in Fig. 5(5), the red area in the upper right corner represents a detection error, but actually corresponds to a correctly identified building change. This shows that most networks exhibit some degree of generalization. In contrast, our network has more complete predictions and exhibits some immunity to interference. As can be seen from Fig. 5(8), regarding the impact of the original building on change detection, our method avoids false detections of the original building and more accurately identify the changed area. In summary, the qualitative comparison results are consistent with the quantitative results shown in Table I. These images show the feasibility of B2CNet, producing fewer detection errors and more accurate edge predictions.

2) Experimental Results on WHU-CD Dataset: Table I displays the results of our proposed B2CNet method and several comparative experiments on the WHU-CD dataset. Our method continues to achieve the best performance in metrics. Notably, our proposed method outperforms the highest score achieved by other experimental methods by 2.21, 0.25, 0.45, 0.01, and 0.26 points on Rec, F1 score, IOU, OA, and Kappa respectively. Although the lightweight version experiences a decline in performance compared to the deeper model, it still remains among the top performers. This reduction in performance is attributed to the shallower model depth, which results in the partial loss of semantic information related to large buildings in the dataset. However, our approach strikes a good balance between Precision and Recall, enhancing Recall through change boundary perception and maintaining superior Precision. This allows us to achieve higher F1 scores while ensuring accuracy,

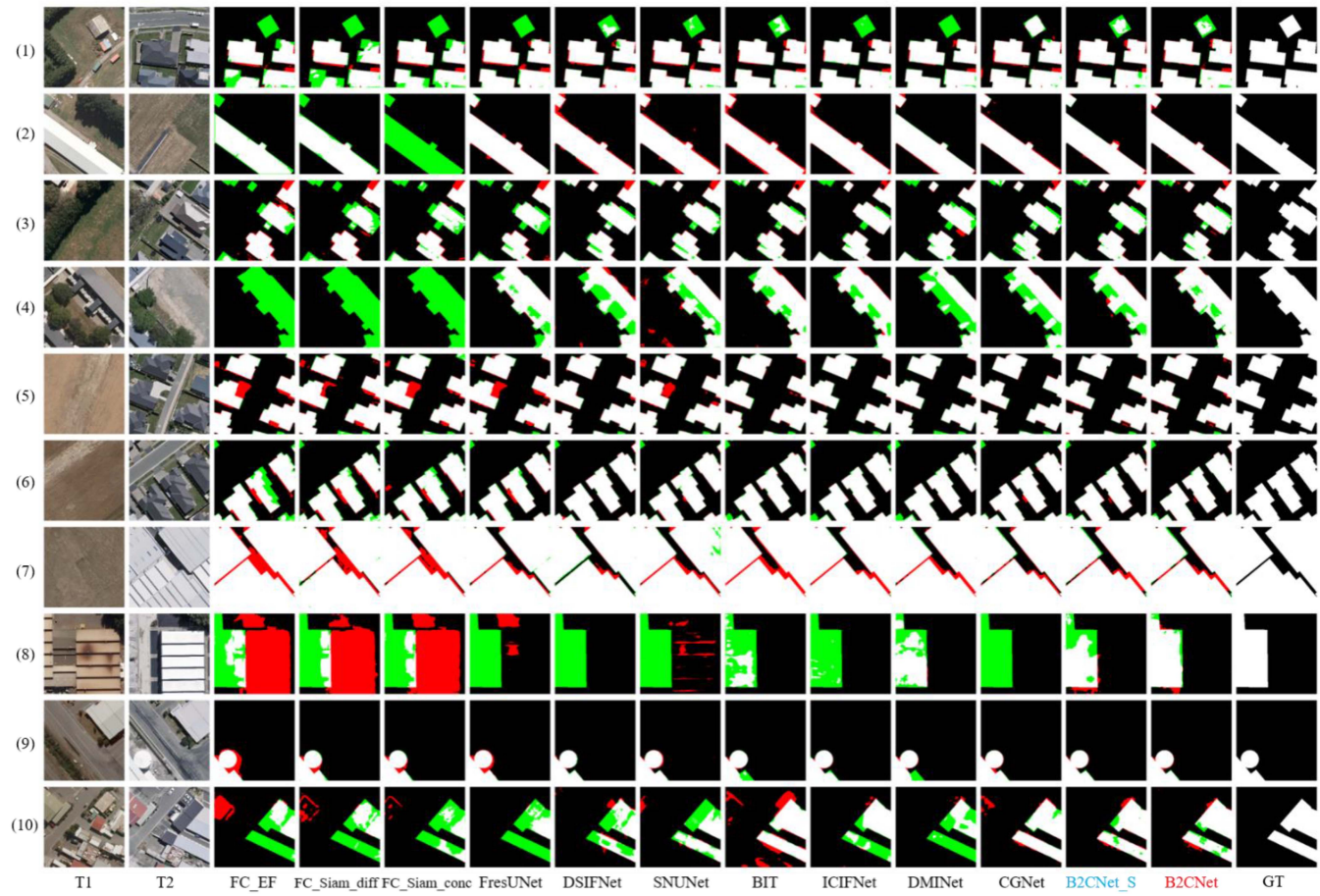

Fig. 6. Qualitative results on WHU-CD.TP (white), TN (black), FP (red), and FN (green).

completeness, and reducing false detection. Fig. 6 shows the results attained by different methods on the WHU-CD dataset.

Fig. 6(1)–(4) shows buildings shaded by trees or vegetation. Fig. 6(5)–(7) shows buildings affected by environmental factors, such as concrete, asphalt floors. Fig. 6(8)–(10) shows buildings affected by the original building with light intensity. Since the WHU-CD dataset has higher resolution (0.2 m/pixel) and rich color texture information, it shows higher false detection rate and missed detection rate. As shown in Fig. 6(1), due to the obvious color difference between the vegetation and the original structure, the vegetation of the adjacent house is mistakenly identified as part of the changed building. Similarly, Fig. 6(5) and (6) show false detection due to similar colors of concrete or asphalt floors and buildings. Compared with other methods, our method can effectively handle the interference of spurious changes. In Fig. 6(8), the rest of the methods almost lose their predictive power for changing regions and exhibit a large number of false detection. In contrast, our method better overcomes the interference between different imaging exposures and the original architecture. Compared with DMINet, B2CNet has a slightly higher false detection rate, but detects fewer holes and has a lower missed detection rate, which is consistent with our quantitative results. Overall, these findings indicate that our method is effective in detecting changes, especially under interference involving spurious changes and illumination changes.

3) Experimental Results on SYSU-CD Dataset: Table I displays the results of our proposed B2CNet method and several comparative experiments on the SYSU-CD dataset. The SYSU-CD dataset presents greater challenges due to its higher diversity, larger data volume, and more complex change scenarios. The table shows that B2CNet outperforms all other experimental methods and achieves the highest scores across four evaluation metrics. Notably, our proposed method obtains scores that are 1.38, 2.00, 0.47, and 1.65 points higher than the highest scores obtained by other experimental methods for F1 score, IOU, OA, and Kappa, respectively. These results demonstrate that, compared to other experimental methods, B2CNet has more efficient feature extraction when dealing with complex and diverse changing scenes.

Fig. 7 shows the results attained by different methods on the SYSU-CD dataset. Fig. 7(1)–(4) show vegetation changes due to seasonal factors and suburban road expansion. Fig. 7(5)–(8) depict changes caused by internal reshaping in complex urban environments. Fig. 7(9) and (10) show the changes in the dock and river due to light intensity and water level. Due to the complexity of the scene in this dataset, the prediction results exhibit a significant number of false detection and missed detection. As can be seen from Fig. 7(2) and (8), other methods have a large number of false detection and missed detection, and the prediction ability of changing region boundaries is limited.

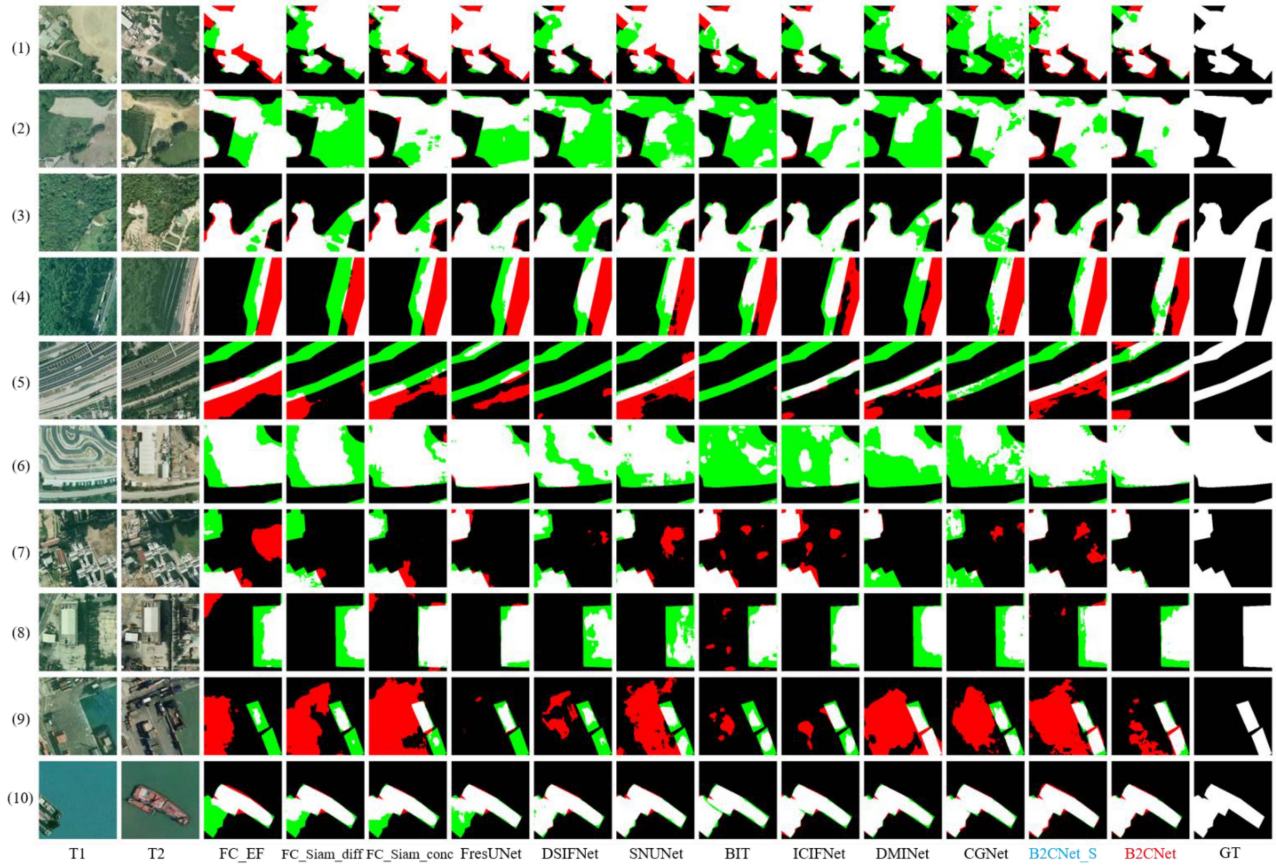

Fig. 7. Qualitative results on SYSU-CD. TP (white), TN (black), FP (red), and FN (green).

Fig. 7(4) and (5) roughly determine the change area, but there are obvious detection errors at the boundaries. Nonetheless, our visual prediction maps show superior performance overall, highlighting the robustness of B2CNet in facing complex scenes. As shown in Fig. 7(7), the same location appears different colors at different times, which can lead to some extraneous changes as opposed to real changes. Many methods fail to exclude these changes, while the proposed method can well handle the pseudochanges resulting from appearance diversity. Our method adopts a boundary-to-center idea, which contributes to the comprehensiveness of changing areas. B2CNet achieves state-of-the-art performance on this dataset.

4) Experimental Results on HRCUS-CD Dataset: In order to further verify the effectiveness of our model in complex scenes, we compared the performance of the model under the high-resolution complex urban scene BCD (HRCUS-CD) dataset. Table II displays the results of our proposed B2CNet method and several comparative experiments on the HRCUS-CD dataset. Compared to the best performing DMINet. Recall, F1, IOU, and Kappa increased by 8.06, 0.75, 0.90, and 0.70, respectively. DSIFNet and SNUNet perform averagely, and the best-performing competitive method is DMINet, followed by ICIFNet.

Fig. 8 shows the results attained by different methods on the HRCUS-CD dataset. This dataset is an extremely challenging task due to the complexity of the scene surrounding the building. Fig. 8(1)–(5) shows the surrounding environment and its complex urban village environment. The visualization results of B2CNet are closer to the labels. Fig. 8(6)–(10) shows scenes with similar colors, including changes in farmland, factories, etc. B2CNet has a stronger ability to resist interference from spurious changes. And it is worth noting that B2CNet has better perception of small buildings and better completeness of perception of large buildings. In contrast, B2CNet alleviates the problem of undetectable changes in small buildings or incomplete change detection in large buildings.

E. Ablation Study

To evaluate the effectiveness of CBM, BFAM, DFEM, and branch supervision (BS) in the architecture of B2CNet. All methods in Table III employ the same backbone. The decoder of baseline includes the feature difference (FD) branch as the base network. In Method_1, Method_2, and Method_3, replace unadded modules with feature concatenation (FC). The aim is to assess the efficacy of these modules within the architecture. The details of the method are as follows.

1) *Baseline:* ResNet18 + FD.
2) *Methods_1:* ResNet18 + CBM + FC.
3) *Methods_2:* ResNet18 + CBM + BFAM + FC.
4) *Methods_3:* ResNet18 + CBM + BFAM + DFEM.
5) *B2CNet:* RseNet18 + CBM + BFAM + DFEM + BS.

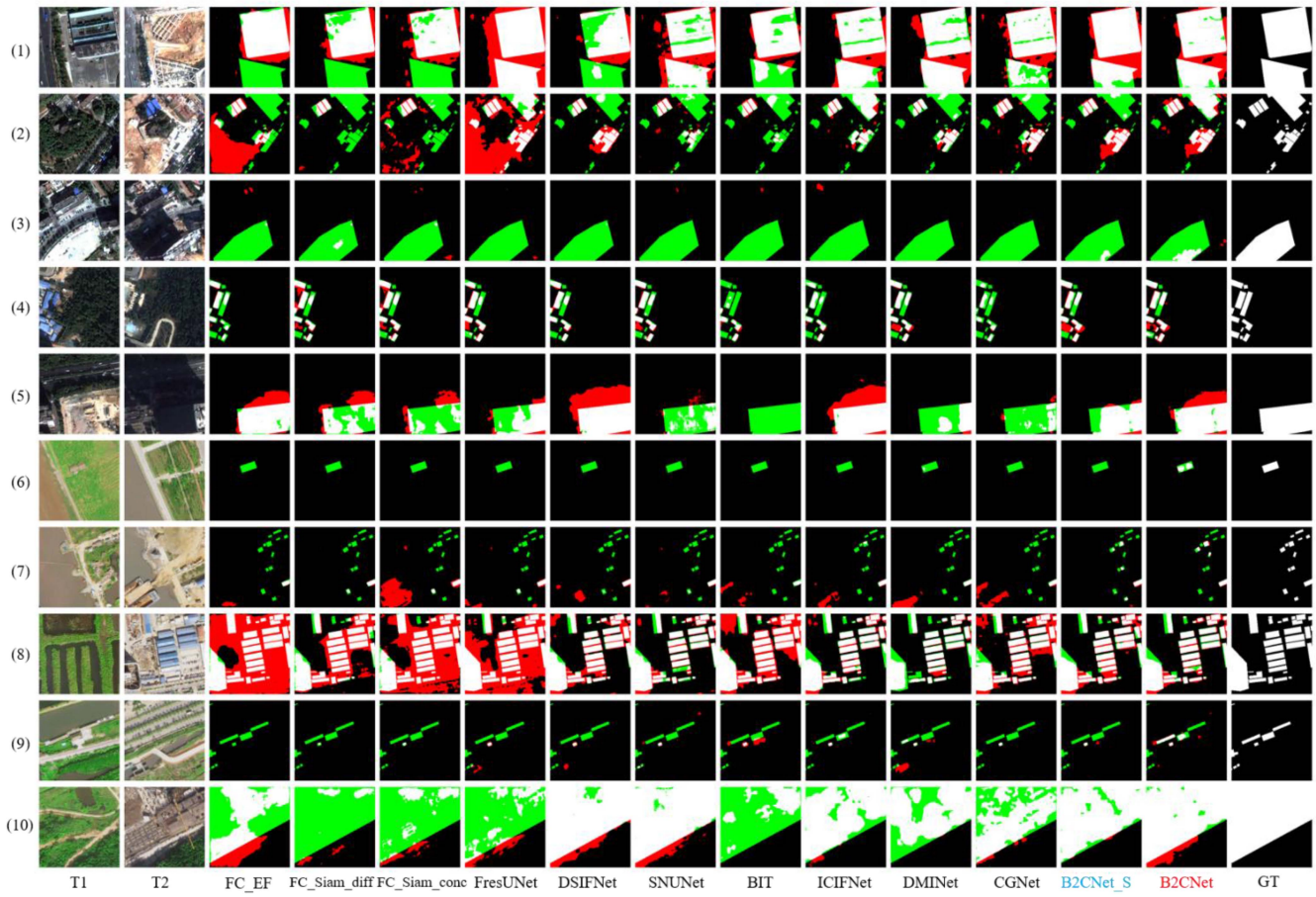

Fig. 8. Qualitative results on HRCUS-CD. TP (white), TN (black), FP (red), and FN (green).

Gradually substituting feature concatenation with our proposed modules reveals a continuous improvement in the model's performance. In Method_1, CBM is used as a branch of the guidance network. This model shows more obvious enhancements on both LEVIR-CD and WHU-CD datasets, and the IOU is increased by 0.96 and 0.62, respectively. This highlights the effectiveness of change boundary-aware guides in change feature detection. However, the improvement of the SYSU-CD and HRCUS-CD datasets is limited, with IOU only improved by 0.35 and 0.33. This may be due to the fact that complex scenes require stronger and more accurate semantic information and richer detailed texture information. Method_2 means that we add BFAM into Methods1 to improve the internal integrity of change region. Compared with Method_1, the improvements achieved by Method_2 in IOU are 0.38 (LEVIR), 0.69 (WHU), 1.20 (SYSU), and 0.68 (HRCUS). Furthermore, In Method_3 which further embeds DFEM. It can be seen that compared with Method_2, the IOU of the four datasets has further improved by 0.27, 0.59, 0.89, and 0.46.

This positively shows that DFEM is better at converting change region information into high-level semantic information. Consequently, significant improvements are observed after implementing supervision on CBM. In B2CNet, the IOU/F1 scores increased again by 0.28/0.71 (LEVIR), 0.96/0.55 (WHU), 2.26/1.56 (SYSU), and 2.32/1.93 (HRCUS), respectively. The application of supervision in the CBM branch is more conducive to improving the accuracy of the information in the changed areas, thereby further helping to identify changes.

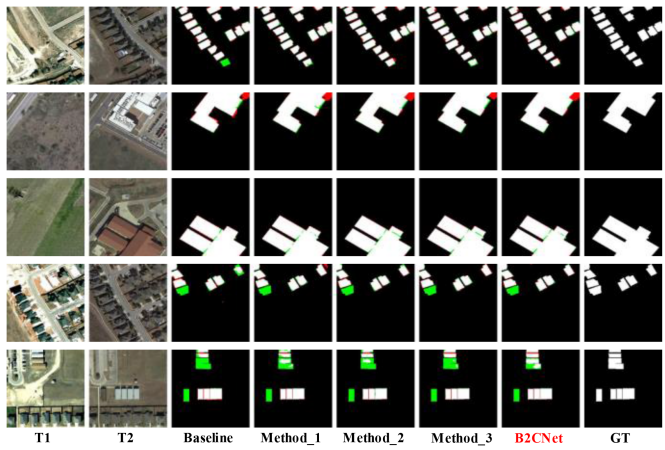

Fig. 9. Qualitative results of ablation study on LEVIR-CD dataset. TP (white), TN (black), FP (red), and FN (green).

Moreover, Fig. 9 represents that refining the network's ability for change feature detection leads to a decrease in the detection missed rate, clearer edge details, and enhanced immunity to

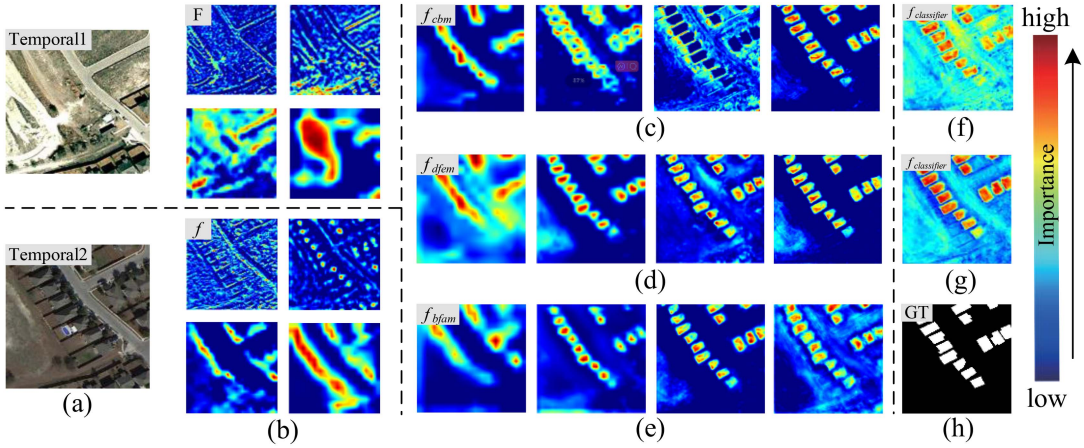

Fig. 10. Network visualization taking images from LEVIR-CD as examples. (a) Input images. (b) Selected multilevel feature maps F and f generated by resnet18. (c), (d), and (e) Selected feature maps f_{cbm}, f_{dfem}, and f_{bfam} coupled through CBM, DFEM and BFAM. (f) And (g) selected feature maps through extra prediction and output classifiers. (h) Ground truth.

TABLE III
QUANTITATIVE PERFORMANCE COMPARISON OF ABLATION EXPERIMENTAL RESULTS ON DIFFERENT DATASETS

Dataset	Methods	CBM	BFAM	DFEM	BS	Pre	Rec	F1	IOU
LEVIR	Baseline					90.34	89.68	90.01	81.83
	Method_1	√				91.25	89.93	90.58	82.79
	Method_2	√	√			90.78	90.80	90.81	83.17
	Method_3	√	√	√		91.25	90.70	90.97	83.44
	B2CNet	√	√	√	√	**91.27**	**91.01**	**91.14**	**83.72**
WHU	Baseline					93.18	88.22	90.63	82.87
	Method_1	√				**94.27**	87.96	91.00	83.49
	Method_2	√	√			94.04	88.92	91.41	84.18
	Method_3	√	√	√		93.00	90.55	91.76	84.77
	B2CNet	√	√	√	√	93.70	**90.97**	**92.31**	**85.73**
SYSU	Baseline					78.93	80.96	79.94	66.58
	Method_1	√				79.19	81.21	80.19	66.93
	Method_2	√	√			79.72	82.42	81.05	68.13
	Method_3	√	√	√		79.23	84.27	81.67	69.02
	B2CNet	√	√	√	√	**81.80**	**84.71**	**83.23**	**71.28**
HRCUS	Baseline					66.34	70.96	68.57	52.18
	Method_1	√				65.50	72.58	68.86	52.51
	Method_2	√	√			63.99	75.92	69.45	53.19
	Method_3	√	√	√		67.55	72.28	69.84	53.65
	B2CNet	√	√	√	√	**69.64**	**74.03**	**71.77**	**55.97**

Note: All the scores are described in percentage (%).
The bold data indicates the best experimental results on the current dataset.

various influences. Meanwhile, in order to more clearly verify the effectiveness of the proposed module, we visualize the feature maps of each stage of B2CNet and provide an example including LEVIR-CD images. Through channel visualization, we selected representative cases to display in Fig. 10. It is obvious that Fig. 10(c) pays more attention to the boundaries of change, Fig. 10(e) pays more attention to the center of the changed individual, and Fig. 10(d) integrates the two and focuses on the whole. Fig. 10(f) is the extra predicted classifier that discriminates Fig. 10(c). It can be seen that its attention to the center of the changing area is still lacking. On the contrary, Fig. 10(g) is the output classifier that classifies Fig. 10(d), and the changed area gets more accurate and complete attention. It is highly coincident with ground truth. These outcomes are attributed to the complementary spatial texture details of the change features provided by BFAM and the augmented feature extraction capability of the model through DFEM. Therefore, these extensive experiments can be observed that the effectiveness of B2CNet in mitigating the proposed problem.

F. Comparison of Efficiency

In addition to conducting qualitative and quantitative analyses of the change detection results of all compared methods, this article also validates the network's efficiency in terms of the number of parameters (Params), floating-point operations per second (FLOPs), and the time required to process a pair image. The number of parameters (Params) represents the model's learning requirements during the training process, corresponding to the spatial complexity of the model. FLOPs denote the number of floating-point operations performed by the model, serving as a measure of the model's time complexity. Time metrics indicate the runtime needed by the model to process a single image. These metrics effectively reflect the model's efficiency. Consistent with all experiments, we conducted 300 rounds of testing on Nvidia RTX 3090 GPUs, with each test involving a size of $256 \times 256 \times 3$ image. The final test results were averaged. As shown in Table IV, our model B2CNet exhibits intermediate levels of Params and FLOPs compared to other networks. However, when compared to DSIFNet, SNUNet, and ICIFNet at the same level, our model demonstrates significantly reduced time while maintaining optimal performance among all networks. Although compared to CGNet, its performance dropped slightly on the LEVIR-CD dataset, but it greatly reduced Params and FLOPs. Furthermore, we propose B2CNet_S, which is a more efficient model. The performance of B2CNet_S is slightly degraded

TABLE IV
EFFICIENCY COMPARISON OF DIFFERENT METHODS ON LEVIR-CD

Methods	Backbone	Params(M)	FLOPs(G)	Time(ms)	F1(%)
FC-EF$_{18}$	UNet	1.35	3.58	2.297	75.86
FC-Siam-Diff$_{18}$	UNet	1.35	4.73	2.998	80.09
FC-Siam-Conc$_{18}$	UNet	1.55	5.33	3.063	81.16
FresUNet$_{19}$	UNet	1.10	2.02	2.414	86.71
DSIFNet$_{20}$	VGG16	35.73	82.27	12.362	90.41
SNUNet$_{22}$	NestedUNet	12.03	54.83	11.141	89.96
BIT$_{22}$	ResNet18	3.50	10.63	5.251	89.75
ICIFNet$_{22}$	ResNet18&PVTv2-B1	23.84	25.41	18.103	90.20
DMINet$_{23}$	ResNet18	6.24	14.55	4.668	90.68
CGNet$_{23}$	VGG16	33.68	82.23	9.194	91.51
B2CNet	ResNet18	16.10	22.36	6.590	91.14
B2CNet_S	ResNet18	4.02	17.11	4.988	90.90

compared to B2CNet. However, it improves the inference speed and significantly reduces the spatial complexity of the model.

Combining the results of subjective and objective analyses, we can conclude that B2CNet achieves the best results on various change detection datasets, while B2CNet_S also achieves excellent results. Thus, our method strikes a favorable balance between model complexity, time cost, and accuracy. Additionally, it demonstrates excellent robustness and generalization.

V. DISCUSSION

Considering the requirement for accurate change detection and the need to alleviate issues such as insufficient finegrained change detection boundaries and poor internal integrity of the change region, we propose a novel change detection network based on change boundary-aware guidance. Our approach aims to retain finer boundary and internal information of the change region, which focuses on obtaining change region information. We introduce three modules to achieve this: the CBM captures change region boundary information, the BFAM emphasizes detail texture information, and the DFEM integrates global information and extracts high-level semantic features. Additionally, our network architecture adopts a boundary-to-center approach, utilizing boundary information for shape and outline description and leveraging detail texture information to supplement boundary and internal information. This approach promotes the completeness and refinement of the detection effect. By keeping the number of model parameters and computational effort at a moderate level, our proposed B2CNet method achieves faster inference with improved detection performance. Furthermore, considering the complexity of real-world application scenarios and hardware limitations [61], [62], we further reduce the number of model parameters and introduce a lightweight version, B2CNet_S. This lightweight version reduces the computational burden and improves inference speed while achieving only a slight decrease in detection accuracy, still maintaining excellent results. The choice of the model depends on the specific scene requirements. For instance, when dealing with a single change category and images with higher resolution, the lightweight version can be selected for higher efficiency with only a minor impact on detection accuracy. On the other hand, in complex and diverse scenes with multiple change categories, B2CNet can be chosen to achieve superior detection results, albeit with some reduction in efficiency.

In summary, our proposed approach ensures better detection results with relatively good efficiency. We have achieved a good balance between efficiency and detection effectiveness, but the model's computational effort did not shrink to a low level due to the large number of multiplication operations involved.

VI. CONCLUSION

We propose B2CNet, a change detection network guided by change boundary-aware. B2CNet consists of three modules: CBM, BFAM, and DFEM. The CBM perceives change region boundaries, enhances boundary information, and guides deeper feature learning for shape description. The BFAM complements spatial-temporal texture information to improve network accuracy. The DFEM combines CBM and BFAM features for feature aggregation and enhanced completeness. These modules collaborate to improve fine-grained change detection boundaries and internal feature integrity, enhancing performance. We conducted comparative experiments on the LEVIR-CD, WHU-CD, SYSU-CD, and HRCUS-CD datasets, demonstrating B2CNet's robustness and generalization. Ablation studies on the four datasets verified the effectiveness of CBM, BFAM, and DFEM. Our approach achieves a favorable balance between efficiency metrics. While B2CNet_S exhibits slightly degraded performance, it excels in efficiency.

Although our approach achieves a good balance between parameters and inference time, there is room for further improvement in computational efficiency. Moreover, complex change scenes are still affected by factors such as large changes, occlusion, shadows, and complex backgrounds. Future research will focus on enhancing anti-jamming ability, improving real-time performance, and exploring DL-based lightweight models for on-board change detection. Additionally, to cope with diverse and real-time task requirements, we will further explore multisource, multitask, and multitemporal phase change detection. We will also consider the possibility of deploying the system in embedded devices or resource-constrained environments.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and members of the editorial team for their comments and suggestions.

REFERENCES

[1] G. Cheng, Y. Huang, X. Li, S. Lyu, Z. Xu, and Q. Zhao, "Change detection methods for remote sensing in the last decade: A comprehensive review, 2023, *arXiv: 2305.05813*.

[2] F. Gao, W. Xiao, Y. Gao, J. Dong, and S. Wang, "Sea ice change detection in SAR images based on convolutional-wavelet neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1240–1244, Aug. 2019, doi: 10.1109/lgrs.2019.2895656.

[3] U. Khusni, H. I. Dewangkoro, and A. M. Arymurthy, "Urban area change detection with combining CNN and RNN from Sentinel-2 multispectral remote sensing data," in *Proc. 3rd Int. Conf. Comput. Inform. Eng.*, 2020, pp. 171–175, doi: 10.1109/ic2ie50715.2020.9274617.

[4] D. Brunner, L. Bruzzone, and G. Lemoine, "Change detection for earthquake damage assessment in built-up areas using very high resolution optical and SAR imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2010, pp. 3210–3213, doi: 10.1109/igarss.2010.5651416.

[5] P. Du, X. Li, W. Cao, Y. Luo, and H. Zhang, "Monitoring urban land cover and vegetation change by multi-temporal remote sensing information," *Mining Sci. Technol.*, vol. 20, no. 6, pp. 922–932, Nov. 2010, doi: 10.1016/S1674-5264(09)60308-2.

[6] P. Gamba, F. Dell'Acqua, and G. Lisini, "Change detection of multitemporal SAR data in urban areas combining feature-based and pixel-based techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2820–2827, Oct. 2006, doi: 10.1109/tgrs.2006.879498.

[7] L. Bruzzone and D. F. 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, doi: 10.1109/36.843009.

[8] Y. Bazi, L. Bruzzone, and F. Melgani, "Automatic identification of the number and values of decision thresholds in the log-ratio image for change detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 3, pp. 349–353, Jul. 2006, doi: 10.1109/lgrs.2006.869973.

[9] L. I. Kuncheva and W. J. Faithfull, "PCA feature extraction for change detection in multidimensional unlabeled data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 1, pp. 69–80, Jan. 2014, doi: 10.1109/TNNLS.2013.2248094.

[10] Q. Liu, G. Liu, C. Huang, S. Liu, and J. Zhao, "A tasseled cap transformation for Landsat 8 OLI TOA reflectance images," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2014, pp. 541–544, doi: 10.1109/IGARSS.2014.6946479.

[11] H. Zhuang, K. Deng, H. Fan, and M. Yu, "Strategies combining spectral angle mapper and change vector analysis to unsupervised change detection in multispectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 681–685, May 2016, doi: 10.1109/lgrs.2016.2536058.

[12] 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, doi: 10.1109/tgrs.2008.916643.

[13] W. Feng, H. Sui, J. Tu, W. Huang, and K. Sun, "A novel change detection approach based on visual saliency and random forest from multi-temporal high-resolution remote-sensing images," *Int. J. Remote Sens.*, vol. 39, no. 22, pp. 7998–8021, Jul. 2018, doi: 10.1080/01431161.2018.1479794.

[14] G. Verdier and A. Ferreira, "Adaptive mahalanobis distance and K-nearest neighbor rule for fault detection in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 1, pp. 59–68, Feb. 2011.

[15] L. Zhang and L. Zhang, "Artificial Intelligence for remote sensing data analysis: A review of challenges and opportunities," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 2, pp. 270–294, Jun. 2022, doi: 10.1109/MGRS.2022.3145854.

[16] L. Zhang, M. Lan, J. Zhang, and D. Tao, "Stagewise unsupervised domain adaptation with adversarial self-training for road segmentation of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5609413, doi: 10.1109/TGRS.2021.3104032.

[17] Z. Lv, H. Huang, W. Sun, T. Lei, J. A. Benediktsson, and J. Li, "Novel enhanced UNet for change detection using multimodal remote sensing image," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 2505405, doi: 10.1109/lgrs.2023.3325439.

[18] Z. Lv, J. Liu, W. Sun, T. Lei, J. A. Benediktsson, and X. Jia, "Hierarchical attention feature fusion-based network for land cover change detection with homogeneous and heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4411115, doi: 10.1109/tgrs.2023.3334521.

[19] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional Siamese networks for change detection," in *Proc. IEEE 25th Int. Conf. Image Process.*, 2018, pp. 4063–4067, doi: 10.1109/icip.2018.8451652.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Interv.: 18th Int. Conf.*, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[21] X. Jiang, S. Xian, M. Wang, and P. Tang, "Dual-pathway change detection network based on the adaptive fusion module," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8018905, doi: 10.1109/lgrs.2021.3103991.

[22] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, Feb. 2019, doi: 10.1109/tgrs.2018.2863224.

[23] B. Bai, W. Fu, T. Lu, and S. Li, "Edge-guided recurrent convolutional neural network for multitemporal remote sensing image building change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5610613, doi: 10.1109/tgrs.2021.3106697.

[24] X. Zheng, D. Guan, B. Li, Z. Chen, and L. Pan, "Global and local graph-based difference image enhancement for change detection," *Remote Sens.*, vol. 15, no. 5, pp. 1194–1194, Feb. 2023, doi: 10.3390/rs15051194.

[25] J. Wang, F. Gao, J. Dong, S. Zhang, and Q. Du, "Change detection from synthetic aperture radar images via graph-based knowledge supplement network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1823–1836, 2022, doi: 10.1109/jstars.2022.3146167.

[26] W. G. C. Bandara and V. M. Patel, "A transformer-based Siamese network for change detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 207–210, doi: 10.1109/IGARSS46834.2022.9883686.

[27] C. Zhang, Y. Zhang, and H. Lin, "Multi-scale feature interaction network for remote sensing change detection," *Remote Sens.*, vol. 15, no. 11, pp. 2880–2880, Jun. 2023, doi: 10.3390/rs15112880.

[28] L. Xia, J. Chen, J. Luo, J. Zhang, D. Yang, and Z. Shen, "Building change detection based on an edge-guided convolutional neural network combined with a transformer," *Remote Sens.*, vol. 14, no. 18, Sep. 2022, Art. no. 4524, doi: 10.3390/rs14184524.

[29] 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, doi: 10.1109/cvpr.2016.90.

[30] S. Fang, K. Li, J. Shao, and Z. Li, "SNUNet-CD: A densely connected siamese network for change detection of VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8007805, doi: 10.1109/LGRS.2021.3056416.

[31] K. Jiang, W. Zhang, J. Liu, F. Liu, and L. Xiao, "Joint variation learning of fusion and difference features for change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4709918, doi: 10.1109/tgrs.2022.3226778.

[32] 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, doi: 10.1109/TGRS.2024.3381752.

[33] R. Caye Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Multitask learning for large-scale semantic change detection," *Comput. Vis. Image Understanding*, vol. 187, Oct. 2019, Art. no. 102783, doi: 10.1016/j.cviu.2019.07.003.

[34] S. Xiang, M. Wang, X. Jiang, G. Xie, Z. Zhang, and P. Tang, "Dual-task semantic change detection for remote sensing images using the generative change field module," *Remote Sens.*, vol. 13, no. 16, Aug. 2021, Art. no. 3336, doi: 10.3390/rs13163336.

[35] 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, doi: 10.1109/tgrs.2021.3095166.

[36] Y. Feng, H. Xu, J. Jiang, H. Liu, and J. Zheng, "ICIF-Net: Intra-scale cross-interaction and inter-scale feature fusion network for bitemporal remote sensing images change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4410213, doi: 10.1109/tgrs.2022.3168331.

[37] X. Tang, T. Zhang, J. Ma, X. Zhang, F. Liu, and L. Jiao, "WNet: W-shaped hierarchical network for remote-sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5615814, doi: 10.1109/tgrs.2023.3296383.

[38] R. Ji, K. Tan, X. Wang, C. Pan, and L. Xin, "PASSNet: A spatial–Spectral feature extraction network with patch attention module for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 5510405, doi: 10.1109/lgrs.2023.3322422.

[39] M.-H. Guo et al., "Attention mechanisms in computer vision: A survey," *Comput. Vis. Media*, vol. 8, no. 3, pp. 331–368, 2022, doi: 10.1007/s41095-022-0271-y.

[40] W. Mi, G. Xie, Z. Zhang, Y. Wang, X. Shao, and Y. Pi, "Smoothing filter-based panchromatic spectral decomposition for multispectral and hyperspectral image pansharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3612–3625, 2022, doi: 10.1109/jstars.2022.3170488.

[41] C. Liu, L. Wei, Z. Zhang, X. Feng, and S. Xiang, "Recursive self-attention modules based network for panchromatic and multispectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 10067–10083, 2023, doi: 10.1109/jstars.2023.3327167.

[42] Z. Zhang, W. Xia, G. Xie, and S. Xiang, "Fast opium poppy detection in unmanned aerial vehicle (UAV) imagery based on deep neural network," *Drones*, vol. 7, no. 9, Sep. 2023, Art. no. 559, doi: 10.3390/drones7090559.

[43] X. Shao, M. Wang, J. Xiao, G. Xie, Z. Zhang, and P. Tang, "Cloud coverage estimation network for remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 6004505, doi: 10.1109/lgrs.2022.3220266.
[44] S. Xiang, Q. Xie, and M. Wang, "Semantic segmentation for remote sensing images based on adaptive feature selection network," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8006705, doi: 10.1109/tgrs.2021.3049125.
[45] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, and A. N. Gomez, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst., Annu. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
[46] K. S. Jaderberg and A. Zisserman, "Spatial transformer networks," in *Proc. Conf. Neural Inf. Process. Syst.*, 2015, pp. 2017–2025.
[47] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141, doi: 10.1109/cvpr.2018.00745.
[48] C. Zhang et al., "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 166, pp. 183–200, Aug. 2020, doi: 10.1016/j.isprsjprs.2020.06.003.
[49] C. Han, C. Wu, H. Guo, M. Hu, and H. Chen, "HANet: A hierarchical attention network for change detection with bi-temporal very-high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3867–3878, 2023, doi: 10.1109/jstars.2023.3264802.
[50] Y. Feng, J. Jiang, H. Xu, and J. Zheng, "Change detection on remote sensing images using dual-branch multilevel intertemporal network," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4401015, doi: 10.1109/tgrs.2023.3241257.
[51] C.-P. Chen, J.-W. Hsieh, P.-Y. Chen, Y.-K. Hsieh, and B.-S. Wang, "SARAS-net: Scale and relation aware siamese network for change detection," in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, pp. 14187–14195, doi: 10.1609/aaai.v37i12.26660.
[52] X. Wang et al., "A high-resolution feature difference attention network for the application of building change detection," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, Aug. 2022, Art. no. 102950, doi: 10.1016/j.jag.2022.102950.
[53] R.-Y. Z. Yang, L. Li, and X. Xie, "SimAM: A simple, parameter free attention module for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 11863–11874.
[54] L. Zhang, S. Zhou, J. Guan, and J. Zhang, "Accurate few-shot object detection with support-query mutual guidance and hybrid loss," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 14424–14432, doi: 10.1109/cvpr46437.2021.01419.
[55] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sens.*, vol. 12, no. 10, May 2020, Art. no. 1662, doi: 10.3390/rs12101662.
[56] 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, doi: 10.1109/tgrs.2018.2858817.
[57] X. Liu, M. Liu, S. Li, X. Liu, F. Wang, and L. Zhang, "A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5604816, doi: 10.1109/tgrs.2021.3085870.
[58] J. Zhang et al., "AERNet: An attention-guided edge refinement network and a dataset for remote sensing building change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5617116, doi: 10.1109/tgrs.2023.3300533.
[59] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. Int. Conf. Learn. Representations*, 2019, pp. 1–18.
[60] C. Han, C. Wu, H. Guo, M. Hu, J. Li, and H. Chen, "Change guiding network: Incorporating change prior to guide change detection in remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 8395–8407, 2023, doi: 10.1109/jstars.2023.3310208.
[61] Z. Zhang, L. Wei, S. Xiang, G. Xie, C. Liu, and M. Xu, "Task-driven on-board real-time panchromatic multispectral fusion processing approach for high-resolution optical remote sensing satellite," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 7636–7661, 2023, doi: 10.1109/jstars.2023.3305231.
[62] Z. Zhang, Z. Qu, S. Liu, D. Li, J. Cao, and G. Xie, "Expandable on-board real-time edge computing architecture for Luojia3 intelligent remote sensing satellite," *Remote Sens.*, vol. 14, no. 15, Jan. 2022, Art. no. 3596, doi: 10.3390/rs14153596.

Zhiqi Zhang received the B.Sc. degree in geographic information system from Huazhong Agricultural University, Wuhan, China, in 2006, the B.Eng. degree in computer science and technology from the Huazhong University of Science and Technology, Wuhan, in 2006, the M.Eng. degree in computer technology, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, in 2015 and 2018, respectively.

He is currently an Associate Professor with the School of Computer Science, Hubei University of Technology, Wuhan. His research interests include system architecture, algorithm optimization, AI, and high-performance processing of remote sensing.

Liyang Bao received the B.Sc. degree in computer science and technology from the Shangqiu University, Shangqiu, China, in 2022. He is currently working toward the M.Eng. degree in computer technology with the School of Computer Science, Hubei University of Technology, Wuhan.

His research interests include intelligent remote sensing image processing, change detection, and deep learning.

Shao Xiang received the B.S. degree in automation engineering from Hefei University, Hefei, China, in 2017, and the M.S. degree in automation from the College of Electrical and Information Engineering, Hunan University, Changsha, China, in 2020. He is currently working toward the Ph.D. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.

His research interests include change detection, image compression, image fusion, object detection, and semantic segmentation of remote sensing.

Guangqi Xie received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2022.

He is currently a Lecturer with the School of Computer Science, Hubei University of Technology, Wuhan, China. His research interest includes image matching and registration, panchromatic sharpening, and image super-resolution.

Rong Gao received the Ph.D. degree in computer science and technology from Wuhan University, Wuhan, China, in 2018.

He is currently an Assistant Professor with the School of Computer Science, Hubei University of Technology, Wuhan. His research interests include artificial intelligence, data mining, and intelligent recommendation.