Progressive Difference Amplification Network With Edge Sensitivity for Remote Sensing Image Change Detection

Yi Liang, Xinghan Xu, Chengkun Zhang, Jianwei Liu, Deyi Wang, and Min Han, *Senior Member, IEEE*

***Abstract*—Capturing finer and discriminative difference features (DFs) is key to obtaining a high-quality change detection (CD) map. However, there is still significant scope for further study on fine-grained detection, especially concerning terms of improving structural integrity and reducing internal holes or sticking in DF. To this end, we propose a progressive difference amplification network (PDANet) with edge sensitivity to detect changed areas in optical remote sensing images (RSIs), where the key point is to amplify DF and reinforce edge detail to improve CD accuracy. The edge sensitivity (ES) encoder is designed to capture the long-distance dependency, which compensates for the limited receptive fields of the convolutional neural network with fixed kernels. Meanwhile, we introduce the prior edge in the network training stage, which collaborates with the ESE to improve the structural integrity of the changed areas. On the other hand, the difference amplification decoder is proposed to enhance the representation of the changed areas, and it is achieved by integrating multiscale DF and reconstructing the original single RSI using DF as full-stage guidance. Finally, the CD map and edge map are predicted based on the reconstructed feature and the maximum scale DF. Extensive experiments on one instance dataset and three CD benchmark datasets demonstrate that PDANet outperforms the state-of-the-art CD competitors both qualitatively and quantitatively.**

***Index Terms*—Change detection (CD), difference features (DFs), edge sensitivity (ES), gate weight modulation, prior, remote sensing images (RSIs).**

I. Introduction

OPTICAL remote sensing images (RSIs) are widely used in real-world applications due to they provide a large range of real-time ground conditions, and change detection (CD) is one of the most interesting research topics in the field of recognition of optical RSI. CD relies on the paired RSIs acquired in the same spatial domain and different temporal domains to provide ground object changes oriented to actual project requirements, such as urban planning [1], fill soil building land [2], land use detection [3], and waterbody monitoring [4].

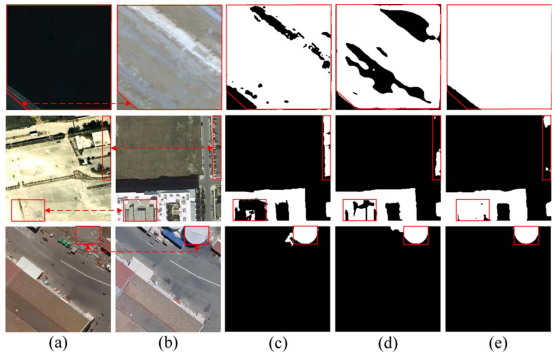


Fig. 1. Visualization of some examples: (a) Time t_1 RSIs. (b) Time t_2 RSIs. (c) P2V. (d) ICIFNet. (e) Proposed PDANet.

As optical sensors become progressively more precise, CD tasks in high-resolution optical RSI present both opportunities and challenges. One of the toughest challenges is the complexity of objects in the scene, such as clear geometric structure, multiple scales, random distributions, and nonuniform spectral features. These factors place higher demands on accurate CD and may result in the CD maps as internal holes, multiple objects sticking, small object loss, etc. As shown in the 1st row of Fig. 1(b), the changed area with the semantic concept of land reclamation (red dotted box) shows significant spectral differences due to the different construction materials and illumination variations. In the 2nd row of Fig. 1, building CD results in the red box are affected by the off-center position of the buildings and similar spectral characteristics to the background roads. In addition, accurate detection is difficult in scenes with mixed a small number of special geometrical configurations, such as the domed building in 3rd row of Fig. 1(b). From these, it can be seen that extracting global structural features and enriching difference features (DFs) to better understand the entire content of the scene is the important requirements for current CD methods.

Nowadays, many studies have been devoted to studying automatic interpretation techniques of computer vision to enrich

Manuscript received 4 December 2023; revised 3 February 2024; accepted 16 March 2024. Date of publication 20 March 2024; date of current version 10 April 2024. This work was supported by the National Natural Science Foundation of China under Grant 62173063. *(Corresponding author: Min Han.)*

Yi Liang is with the Faculty of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China (e-mail: liangyi@mail.dlut.edu.cn).

Xinghan Xu and Jianwei Liu are with the Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian 116024, China (e-mail: xuxh2023@dlut.edu.cn; jwliu@dlut.edu.cn).

Chengkun Zhang is with the Intelligent Computing and Application Laboratory of Qinghai Province, Qinghai University, Xining 810000, China (e-mail: zhangchengkundon@163.com).

Deyi Wang is with the Faculty of Electronics and Communications, Guangdong Mechanical and Electrical Polytechnic, Guangzhou 510550, China (e-mail: 2023010019@gdmec.edu.cn).

Min Han is with the Key Laboratory of Intelligent Control and Optimization for Industrial Equipment of Ministry of Education, Dalian University of Technology, Dalian 116024, China (e-mail: minhan@dlut.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3379565

scene information in CD tasks, i.e., building CD methods based on convolution neural network (CNN) [5] and [6]. Since the nature of CNN is to learn local information hierarchically with a fixed-size window, this causes the network ignore structural context information. To this end, many articles carried out expanding the receptive field to alleviate the above-mentioned problem, such as multiscale feature fusion [7], multiscale convolution [8], atrous spatial pyramid pooling [9], dilated convolution [10], sparse Markov random field [11], and attention mechanism [12]. The recently thriving transformer with long-range dependent properties due to the attention-dependent mechanism is often combined with CNN in the CD task [13], [14], [15], [16], and [17]. The above-mentioned methods show excellent performance in global–local context learning. But it is also critical to obtain long-range context information to further ensure structural details of multiscale objects in high-resolution RSI. Some recent works [18] and [19] adopt the self-attention to model long-range context. However, they do not take into account the effective supervisory role of edge cues, which still has prediction bias.

Existing methods obtain CD maps through distance metrics between DFs [20], [21], [22], and [23] or direct classification [24], [25], and [26]. From this, it is particularly important to separate irrelevant change from DF in RSI with low spectral resolution. Several works proposed effective modules to implement difference recalibration and enhancement. For example, Qu et al. [27] proposed the difference amplification module based on similarity measurement to mask the background superpixels of the RSIs. Lei et al. [28] designed a difference enhancement module based on channel attention to mask irrelevant changes of encoded features at each feature extraction stage. Zhang et al. [29] proposed the feature difference module based on different receptive fields to rich change information. Song et al. [30] inserted a spatial attention-based content difference enhancement module into middleware between the encoder and decoder to refine the encoded features. All of the above-mentioned work achieves difference enhancement based on encoded features. To the best of our knowledge, the encoder focuses more on general feature extraction of the input optical RSI, while the decoder focuses more on feature learning for the target task [31]. The above-mentioned methods do not take into account the fact that there is more redundant information in the encoded features, resulting in the performance of the model facing a bottleneck.

To address the above-mentioned problems, we propose a progressive difference amplification network (PDANet) with edge sensitivity. A pseudosiamese edge sensitivity encoder is proposed to extract global–local context information, where edge sensitivity modules (ESMs) are embedded in the shallow convolutional layers to model long-range context information based on multiscale receptive fields and gate weight modulation. A triple-branches difference amplification decoder is proposed to accurately identify DFs, where the difference mainstream branch (DMB) performs multiscale DF generation based on encoded features, and bilateral auxiliary branches (BABs) perform reconstruction of a single RSI full-stage guided by the DF. In this way, the decoder is forced to reduce irrelevant change information. Finally, CNN-based predictive classifiers (including a CD head and an edge head) generate the CD map and edge map, and a edge prior is introduced to constraint network focus more on the edge of the changed areas in a supervised way. In summary, the main contributions are summarized as follows.

1) We propose a PDANet with edge sensitivity to detect the changes in paired optical RSIs, which consists of an edge sensitivity encoder and a difference amplification decoder.
2) We propose an edge sensitivity scheme in low-level layers of the encoder to improve the structural integrity in a multiscale and gate modulation manner. The mechanism can capture the long-range context through multiscale design, and effectively communicate the multiscale information with the help of information flow selected by the gate modulation.
3) We design a difference amplification mechanism to improve the accurate spatial localization of changed areas with full-stage difference semantic guidance and RSI content reconstruction. We also introduce the different-grained prior edge scheme to constrain the PDANet in a supervised manner to help understand the full content of a scene.

II. Related Works

A. Encoder–Decoder (ED) Backbone in CD Models

ED is one of the widely used structures for CNN-based CD methods. Existing ED structures can be finely categorized into single-stream ED (SED), dual-single stream ED (DSED), and dual-dual stream ED (DDED), as shown in Fig. 2(a)–(c) [32] and [33]. The SED uses an early fusion strategy to learn the DF from the absolute difference between two RSIs, such as FC-EF [24]. The DSED uses a late fusion strategy, where the dual encoders extract the multiscale features of the input RSIs, and one decoder emphasizes the learning of DF, such as SNUNet [25] and P2V [34]. The DDED also employs a late fusion strategy, and unlike DSED, its dual decoder computes the largest scale DF, such as MSCANet [15]. However, SED ignores the private feature learning of the original RSI, leading to the destruction of structural information of ground objects [12] and [35] and lack of resistance to pseudochange interference. DSED is highly dependent on encoded features [36], resulting in poor interpretability. DDED does not fully utilize the semantic information of RSI, resulting in a bottleneck in model performance. In this article, we construct a dual–triple stream encoder–decoder (DTED) structure as shown in Fig. 2(d), which considers the complementary roles of DSED and DDED. In DTED, the multiscale DFs and the private features of the input RSIs are effectively preserved. In this way, more richer representation of the changed object is achieved at the object level, which is conducive to resisting pseudochanges.

B. Edge Detail Learning

A high-quality CD map requires the smooth edge and a complete spatial structure. Edge cues can provide strong structural information of ground objects and are widely used in vision

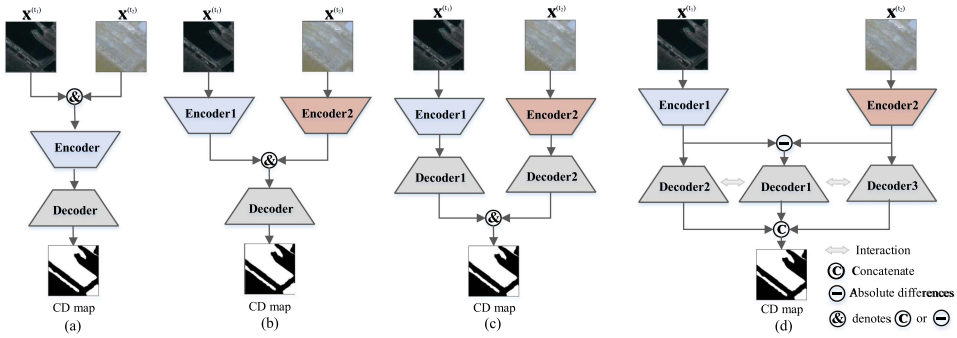


Fig. 2. Summary of the ED structures for CD methods. (a) SED. (b) DSED. (c) DDED. (d) DTED in this article.

tasks, such as salient object detection [37], semantic segmentation [38], image compression [39], and small target augmentation [40]. A sharp-edge detail highly advances global spatial structure, which is helpful in understanding the entire scene. Only a few CD works have focused on edge detail modeling or introducing edge prior for joint learning. In terms of edge detail modeling, Chen et al. [41] proposed an edge-aware module (EAM) based on dilated convolution and gated attention, and embedded EAM in each encoder block to refine the edge details. Yang et al. [42] designed the contour channel attention module (CCAM) to refine the high-resolution decoded features and highlight the edges of changed areas, where CCAM uses superpixel object segmentation as a reference to enhance the internal consistency of the changed area. Xia et al. [14] constructed an edge detection branch to obtain multiscale edge features for masking decoded DFs. In terms of edge prior joint learning, the canny operator acts on CD ground truth to construct edge ground truth. Existing methods usually integrate various convolutional layers at the end of the decoder to generate an edge mask, and optimize the network with an edge loss function. To further improve the model accuracy, some works incorporate the deep supervision strategy to additionally assist the learning of hierarchical edge masks [41] and [43]. Inspired by the above-mentioned ideas, this article designs ESM and the edge detection-based optimization strategy together to learn finer edge details. In addition, the above-mentioned methods use the single-grained edge prior to outline changed areas, ignoring the role of other pixels adjacent to the single-grained edge pixel. Therefore, we actively explore the effect of different grained edge prior on the CD method.

C. Weighting Adjustment Strategy

In order to efficiently utilize CNN features in both spatial and channel dimensions, many weight adjustment strategies are introduced, such as the deep supervision strategy, attention mechanism, and gate mechanism. Deep supervision updates some of the middle layer weights by adding auxiliary classifiers in the training stage, which greatly improves the learning efficiency of change categories in the middle layer [22], [32], and [36]. The attention mechanism computes nonlocal important weights by inner product self-mapping or pooling operation to focus preference information. It effectively compensates for the limitation of the receptive field of the fixed-kernel convolutional that prevents adequate fitting of contextual information [44], [45], [46], and [47]. Gate mechanism adaptively controls information flow by directly affecting the competing or collaborative behavior of neurons via the gating function (i.e., the weight function) [48]. Based on the above-mentioned theory, we embed a attention-based gate integration (GI) unit in the ESM for modulating and fusing the multiple receptive field features.

III. Proposed Model

A. Architecture Overview

Fig. 3 illustrates the PDANet architecture that includes an edge sensitivity encoder, a difference amplification decoder, and the prediction heads (a CD head and an edge head). The details are as follows.

In the edge sensitivity encoder, a pseudosiamese ResNet18 [49] (all output channels are set to 64) is built to extract multiscale features, which are denoted as $\{\mathbf{F}^{*}_{e,k}\}_{k=1}^{5}$ ($* \in \{t_1, t_2\}$), where k represents the index of the encoder level. Two ESMs are embedded in the first two ResBlocks in a skip connection manner. Specifically:

1) input RSI ($\mathbf{X}^{(t_1)}$ or $\mathbf{X}^{(t_2)} \in \mathbb{R}^{3\times H\times W}$) is fed to a 7×7 convolutional layer to extract half-size shallow features $\mathbf{F}_{e,1} \in \mathbb{R}^{64\times H/2\times W/2}$;
2) $\mathbf{F}_{e,1}$ is, respectively, fed to the **1st** ResBlock and **1st** ESM, whose outputs are fused into $\mathbf{F}_{e,2} \in \mathbb{R}^{64\times H/4\times W/4}$ through concatenation and a 1×1 convolution layer;
3) $\mathbf{F}_{e,3} \in \mathbb{R}^{64\times H/8\times W/8}$ is also obtained in the same construction as 2);
4) $\mathbf{F}_{e,4} \in \mathbb{R}^{64\times H/16\times W/16}$ and $\mathbf{F}_{e,5} \in \mathbb{R}^{64\times H/32\times W/32}$ are obtained by another two ResBlocks, respectively.

In the difference amplification decoder, a DMB and two DF-guided BABs (denoted as BAB-1 and BAB-2) jointly learn the details of the changed areas. Specifically:

1) DMB consists of five difference decoding modules (DDMs), and outputs DFs $\{\mathbf{F}_{d,k}\}_{k=1}^{5}$ corresponding to $\{\mathbf{F}_{e,k}\}_{k=1}^{5}$ on the scale;
2) BAB consists of four auxiliary decoding modules (ADMs) and reconstructs private features ($\{\mathbf{F}^{*}_{Ad,k}\}_{k=1}^{4}$ $(* \in \{t_1, t_2\})$ of single input RSI with a greater focus on changed areas;
3) the amplified DF ($\mathbf{F}_{Ad}$) is obtained by absolute difference operation based on the largest scale reconstructed features

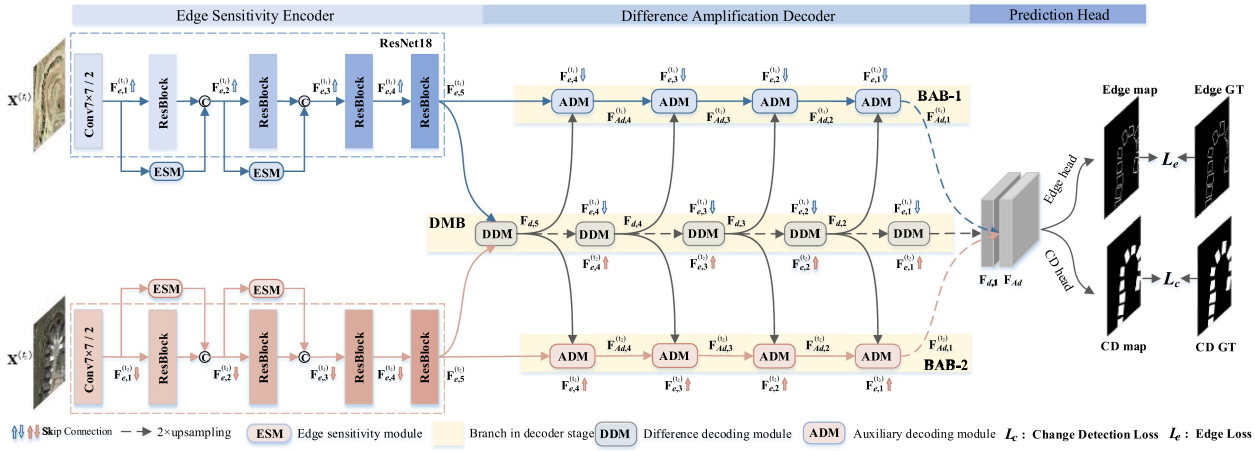


Fig. 3. Architecture of the proposed PDANet.

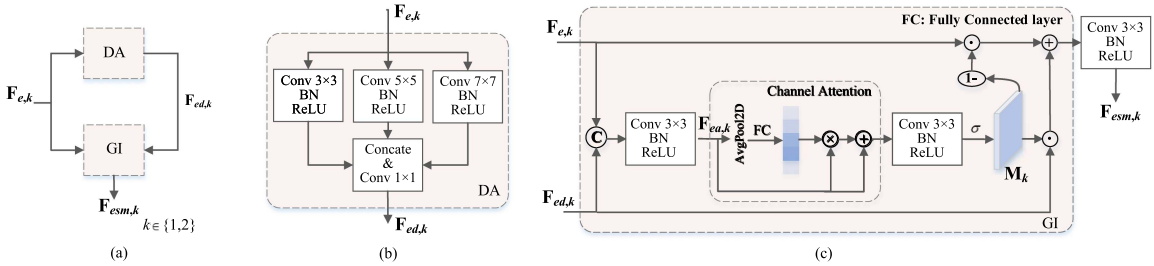


Fig. 4. Architecture of ESM. (a) Edge sensitivity module. (b) Detail-aware unit. (c) Gate integration unit.

$\mathbf{F}_{Ad,1}^{(t_1)}$ and $\mathbf{F}_{Ad,1}^{(t_2)}$, and a 3×3 convolutional layer with batch normalization (BN) and ReLU activation;

4) $\mathbf{F}_{d,1}$ and $\mathbf{F}_{Ad}$ are concatenated and then sent to the CD head and edge head to output the CD probability map and edge probability map, respectively.

B. Edge Sensitivity Encoder

The low-level features have higher spatial resolution and more complete structure information of the ground objects. However, global structural is diluted during the encoding process from shallow to deep, and the upsampling decoding process does not accurately recover it. To this end, the ESM that acts on the shallow encoded feature is proposed and used to reinforce structural information with richer receptive fields. In addition, a GI unit is introduced in ESM to incorporate multiple receptive field features. Fig. 4 shows the structure of ESM, including the detail-aware (DA) unit and GI unit, and they are described as follows.

1) DA Unit: DA can achieve compact spatial information based on multiple receptive fields, and its architecture is shown in Fig. 4(b). DA includes three parallel convolutional blocks with BN and ReLU, which have 3×3, 5×5, and 7×7 convolutional kernels, respectively. Next, the outputs of three convolutional blocks are concatenated and fed into a 1×1 convolutional layer to obtain fined features $\mathbf{F}_{ed,k}$, which can be expressed as follows:

$$\mathbf{F}_{ed,k} = \mathrm{Conv}_{1\times1}[\mathrm{BRConv}_{s\times s}(\mathbf{F}_{e,k})] \tag{1}$$

where, $\mathbf{F}_{e,k}$ ($k \in \{1,2\}$) and $\mathbf{F}_{ed,k}$ are the input and output features of DA, respectively. $\mathrm{BRConv}_{s\times s}$ ($s \in \{3,5,7\}$) represents the convolutional block with $s \times s$ kernel, BN, and ReLU. $\mathrm{Conv}_{1\times1}$ represents the 1×1 convolutional layer. $[\cdot,\cdot]$ denotes the concatenation in channel dimension.

2) GI Unit: DA obtains more structural information by expanding the receptive fields, but the nature of its convolution still loses some information. Therefore, it is necessary to retain the current encoded feature $\mathbf{F}_{e,k}$ before acquiring the next stage encoded features. However, a practical problem, how to effectively select the most valuable information from the multiscale collaborative features (i.e., $\mathbf{F}_{e,k}$ and $\mathbf{F}_{ed,k}$), to be solved because the direct addition and concatenation of different receptive field features may be uncontrollable and miscellaneous. To this end, we proposed a GI unit to learn weight mask $\mathbf{M}_k$ for adaptively controlling the effect of multiscale collaborative features.

The architecture of GI is shown in Fig. 4(c), and it hinges on the learnable weight mask $\mathbf{M}_k \in \mathbb{R}^{C\times H_k\times W_k}$ ($H_k = H/(2^k)$, $W_k = W/(2^k)$, and $k \in \{1,2\}$), which controls how much information in $\mathbf{F}_{ed,k}$ is available to flow into the next encoding stage. Correspondingly, $1 - \mathbf{M}_k$ (1 denotes an all-one tensor of the same size as $\mathbf{M}_k$) measures how much information in $\mathbf{F}_{e,k}$ flow into the next encoding stage. Specifically, we first concatenate the encoded features $\mathbf{F}_{e,k}$ and the DA feature $\mathbf{F}_{ed,k}$ in the channel dimension, and sent it to a 3×3 convolutional block with BN and ReLU to obtain the integrated features $\mathbf{F}_{ea,k}$. Then, $\mathbf{F}_{ea,k}$ is refined by channel attention and an additional 3×3 convolutional block with BN and ReLU. Next, sigmoid

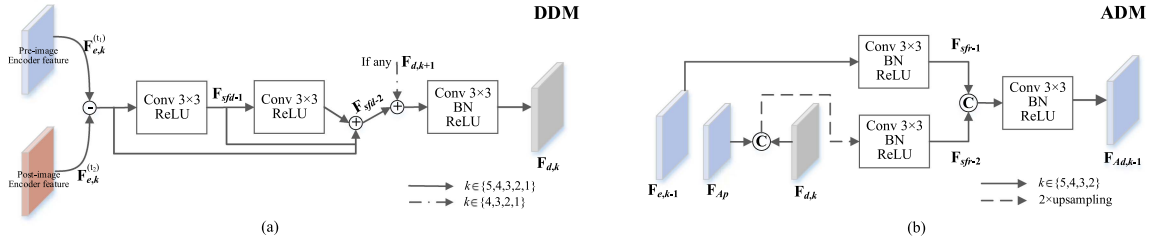


Fig. 5. Structure of DDM and ADM.

activation ($\sigma(\cdot)$) is used to output $\mathbf{M}_k$. The above-mentioned process can be expressed as follows:

$$\mathbf{M}_k = \sigma(\text{BRConv}_{3\times3}(\text{CA}(\text{BRConv}_{3\times3}([\mathbf{F}_{e,k}, \mathbf{F}_{ed,k}])))) \quad (2)$$

where CA() is channel attention. $\text{BRConv}_{3\times3}$ represents the 3×3 convolutional block with BN and ReLU. Finally, $\mathbf{M}_k$ adjusts the weights of $\mathbf{F}_{e,k}$ and $\mathbf{F}_{ed,k}$ based on spatial pixels

$$\mathbf{F}_{\text{esm},k} = \text{BRConv}_{3\times3}(((1-\mathbf{M}_k)\odot\mathbf{F}_{e,k}) \oplus (\mathbf{M}_k \odot \mathbf{F}_{ed,k})) \quad (3)$$

where $\mathbf{F}_{\text{esm},k}$ is output feature of current GI. $\odot$ and $\oplus$ are the elementwise multiplication and summation, respectively.

C. Difference Amplification Decoder

The decoder aims to generate the change-related features. As described in Section II, existing studies adopt SED, DSED, or DDED to deal with encoded features, which results in redundant feature interference and limited model performance. Considering the reconstructed features of a single RSI can retain more ground object information independently, two reconstructed processes corresponding to the two RSIs are introduced in our decoder, i.e., the BABs (BAB-1 and BAB-2) in Fig. 3. In addition, the reconstruction process necessarily amplifies the details of the unchanged areas. In order to reduce redundant interference, we use DFs from the DMB to full-stage guide the reconstruction and to emphasize change-related features. The details of DMB and BABs are as follows.

1) Difference Mainstream Branch: The workflow of the DMB is represented as the middle yellow rectangular box in Fig. 3. Given the multiscale encoded features $\{\mathbf{F}_{e,k}^{(t_1)}\}_{k=1}^{5}$ and $\{\mathbf{F}_{e,k}^{(t_2)}\}_{k=1}^{5}$, DMB progressively generates the DFs $\{\mathbf{F}_{d,k}^{*}\}_{k=5}^{1}$ from deep to shallow through a series of simple structured difference decode modules. Fig. 5(a) shows the structure of DDM. First, $|\,\mathbf{F}_{e,k}^{(t_1)} - \mathbf{F}_{e,k}^{(t_2)}\,|$ is passed to two cascaded 3×3 convolutional blocks with ReLU, and then the semifinished DFs are summed to obtain $\mathbf{F}_{\text{sfd}-2}$ (5). Second, the upsampled prescale DF $\mathbf{F}_{d,k+1}$ and $\mathbf{F}_{\text{sfd}-2}$ are fused by summation and a 3×3 convolutional blocks with BN and ReLU to achieve integration of different scales DFs (4). Specifically, the DDM corresponding to $k = 5$ has no prescale DF as input. In summary, the learning process of DDM can be expressed as follows:

$$\mathbf{F}_{d,k} = \text{BRConv}_{3\times3}(\mathbf{F}_{\text{sfd}-2} + \mathbf{F}_{d,k+1}) \quad (4)$$

$$\mathbf{F}_{\text{sfd}-2} = \text{RConv}_{3\times3}(\mathbf{F}_{\text{sfd}-1}) + \mathbf{F}_{\text{sfd}-1} + \mathbf{F}_{\text{abs},k} \quad (5)$$

$$\mathbf{F}_{\text{sfd}-1} = \text{RConv}_{3\times3}(\mathbf{F}_{\text{abs},k}) \quad (6)$$

$$\mathbf{F}_{\text{abs},k} = |\,\mathbf{F}_{e,k}^{(t_1)} - \mathbf{F}_{e,k}^{(t_2)}\,| \quad (7)$$

where $\mathbf{F}_{d,k}$ ($k \in \{5,4,3,2,1\}$) is the output DF of current DDM. $\mathbf{F}_{d,k+1}$ ($k \in \{4,3,2,1\}$) is the upsampled DF from the previous DDM. $\text{RConv}_{3\times3}$ different from $\text{BRConv}_{3\times3}$ in that there is no BN layer. Both $\mathbf{F}_{\text{sfd}-1}$ and $\mathbf{F}_{\text{sfd}-2}$ represent the semifinished DFs of the DDM middle layers.

2) Bilateral Auxiliary Branches: The workflows of the BABs are represented as the upper and lower yellow rectangular boxes in Fig. 3. Since BAB-1 and BAB-2 are siamese decoding branches with unshared weights, we present the details based on the BAB-1 branch example. Given the multiscale encoded features $\{\mathbf{F}_{e,k}^{(t_1)}\}_{k=1}^{5}$, BAB simultaneously fuses $\mathbf{F}_{e,k}^{(t_1)}$ and $\mathbf{F}_{d,k}$ through a series of ADM and reconstructs the private features $\mathbf{F}_{Ad,k-1}^{(t_1)}$ for the single RSI. Fig. 5(b) shows the structure of ADM. In order to the reconstructed features $\{\mathbf{F}_{Ad,k-1}^{(t_1)}\}_{k=2}^{5}$ focus more on the details of the changed area, we use a 3×3 convolutional block with BN and ReLU to fuse the private feature $\mathbf{F}_{AP}$ and $\mathbf{F}_{d,k}$, where $\mathbf{F}_{AP}$ refers to $\mathbf{F}_{e,5}^{(t_1)}$ for the first ADM and $\mathbf{F}_{Ad,k-1}^{(t_1)}$ for other ADMs (10). Then, the prescale encoded feature $\mathbf{F}_{e,k-1}$ is fused to current semifinished feature $\mathbf{F}_{\text{sfr}-2}$ by concatenation and two 3×3 convolutional blocks with BN and ReLU (8). In summary, the learning process of ADM can be expressed as follows:

$$\mathbf{F}_{Ad,k-1} = \text{BRConv}_{3\times3}([\mathbf{F}_{\text{sfr}-1}, \mathbf{F}_{\text{sfr}-2}]) \quad (8)$$

$$\mathbf{F}_{\text{sfr}-1} = \text{BRConv}_{3\times3}(\mathbf{F}_{e,k-1}) \quad (9)$$

$$\mathbf{F}_{\text{sfr}-2} = \text{BRConv}_{3\times3}([\mathbf{F}_{AP}, \mathbf{F}_{d,k}]) \quad (10)$$

where $\mathbf{F}_{Ad,k-1}$ is the reconstructed feature of current ADM. Both $\mathbf{F}_{\text{sfr}-1}$ and $\mathbf{F}_{\text{sfr}-2}$ represent the semifinished reconstructed features of the ADM middle layers.

D. Prediction Heads and Loss Function

1) Prediction Heads: We use CNN-based classifiers to construct the prediction heads for generating the CD map ($\hat{Y} \in \mathbb{R}^{2\times H\times W}$) and edge map ($\hat{Z} \in \mathbb{R}^{2\times H\times W}$). Specifically, both the largest scale DFs $\mathbf{F}_d$ from DMB and amplification DFs $\mathbf{F}_{Ad} = \text{BRConv}_{3\times3}(|\,\mathbf{F}_{Ad,1}^{(t_1)} - \mathbf{F}_{Ad,1}^{(t_2)}\,|)$ from BABs are concatenated and sent to the CD head and the edge head to generate $\hat{Y}$ and $\hat{Z}$. Both CD head and the edge head contain two convolutional blocks, where the first convolutional block has input and output channels of 128, 64, and the second convolutional block has input and output channels of 64, 2. They are expressed as

follows:

$$\hat{Y} = \sigma(\text{Conv}_{1\times1}(\text{BRConv}_{3\times3}[\mathbf{F}_d, \mathbf{F}_{Ad}])) \tag{11}$$

$$\hat{Z} = \sigma(\text{BRConv}_{1\times1}(\text{BRConv}_{1\times1}[\mathbf{F}_d, \mathbf{F}_{Ad}])) \tag{12}$$

where $\sigma()$ is softmax function.

2) Loss Function: We design a loss function to constrain the whole network to focus not only on the center of changed areas but also on the edge details. The total function L consists of the binary CD loss L_{cd} and edge loss L_{edge}. L_{cd} represents the loss between the CD map $\hat{Y}$ and the reference ground truth Y. L_{edge} represents the loss between the edge map $\hat{Z}$ and the reference edge ground truth Z. L can be expressed as follows:

$$L = L_{\text{cd}} + \alpha \times L_{\text{edge}} = f\left(\hat{Y}, Y\right) + \alpha \times f\left(\hat{Z}, Z\right) \tag{13}$$

where α is the hyperparameter used to balance L_{cd} and L_{edge}, and we discuss it in Section V. The function $f(\cdot,\cdot)$ is constructed from the binary cross-entropy loss and dice loss, described as follows:

$$f(\hat{x}, x) = f_{\text{bce}}(\hat{x}, x) + f_{\text{dice}}(\hat{x}, x) \tag{14}$$

$$f_{\text{bce}}(\hat{x}, x) = \sum_{i=1,j=1}^{H\times W} \frac{x_{ij}\log\hat{x}_{ij} + (1-x_{ij})\log(1-\hat{x}_{ij})}{H\times W} \tag{15}$$

$$f_{\text{dice}}(\hat{x}, x) = 1 - \sum_{i=1,j=1}^{H\times W} \frac{2\hat{x}_{ij}x_{ij} + \varepsilon}{\hat{x}_{ij} + x_{ij} + \varepsilon} \tag{16}$$

where $\hat{x}$ and x are the predicted map and ground truth, respectively. H and W are the height and width of the input RSI. i and j denote the position indexes of the pixel. ε is a smoothing constant that prevents the denominator from being 0, and is set to 0.0001. $x_{ij} = 1$ ($x_{ij} = 0$) indicates a changed (unchanged) pixel in CD ground truth, or an edge (nonedge) pixel in edge ground truth.

In particular, considering that the single-pixel prior edge does not easily constrain the PDANet to anchor the edge of the changed area in high-resolution RSI, we follow the work of Gu et al. [50] and attempt to use coarse-grained prior edge to guide the network training. Unlike existing methods that use single-pixel edge ground truth to guide model [14], [41], [42], and [43], our approach additionally considers the role of other pixels near the single edge pixel. Specifically, we use the canny operator to generate single-pixel edge ground truth Z_{sp} (value is 0 or 1) based on CD ground truth. Then, the coarse-grained edge ground truth is determined by setting the Euclidean distance $s \in \mathbb{N}^+$ between single edge pixel and its surrounding pixels (i.e., assigning 1 value to the pixels whose Euclidean distance from the single edge pixel is at $[0, s]$), and calculated as follows:

$$Z_{\text{cg}}(i,j) = \begin{cases} 1, & \sqrt{(i_e-i)^2 + (j_e-j)^2} \le s \\ 0, & \text{else} \end{cases}$$
$$\text{s.t. } i \ne i_e \in \{1,\ldots,H\}, j \ne j_e \in \{1,\ldots,W\} \tag{17}$$

where $Z_{\text{cg}}(i,j)$ represents the coarse-grained prior edge value at location i and j. $(i_e,\ j_e)$ denotes a position of $Z_{sp} = 1$, $(i,\ j)$ denote other positions different from $(i_e,\ j_e)$. $\sqrt{(i_e-i)^2 + (j_e-j)^2}$ calculates the Euclidean distance between (i_e, j_e) and (i, j). We discuss the effect of different edge scale s in Section V.

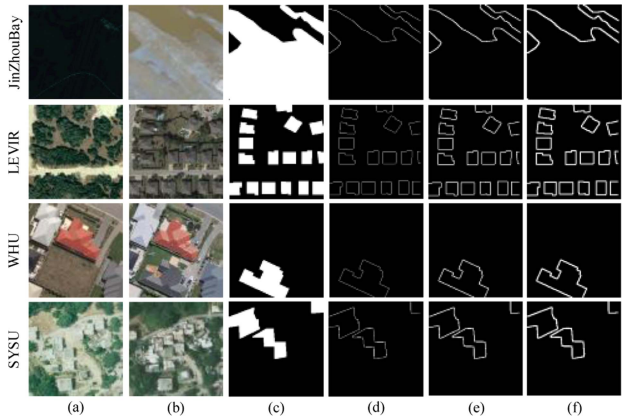


Fig. 6. Display of cropped examples. (a) Time t_1 RSIs. (b) Time t_2 RSIs. (c) CD ground truth. (d), (e), and (f) are the edge ground truths with $s = 1, 2, 3$.

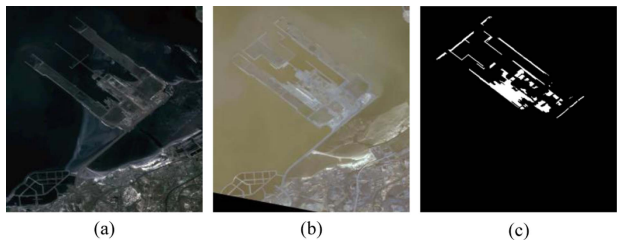


Fig. 7. JinzhouBay dataset. (a) Time t_1 RSI of JinzhouBay collected in 2015. (b) Time t_2 RSI of JinzhouBay collected in 2017. (c) Ground truth.

IV. Experiments

A. Datasets and Implementation Details

We conduct comprehensive experiments on an instance dataset and three public datasets, including:

1) JinzhouBay-CD;
2) LEVIR-CD [23];
3) WHU-CD [51];
4) SYSU-CD [22].

Fig. 6 shows cropped examples of the four datasets, and shows the edge ground truths with different scales (i.e., $s = 1, 2, 3$), which are described in detail as follows.

JinzhouBay-CD: This small dataset contains a pair of nonhomologous RSIs recording the reclamation in Jinzhou-Bay, China. RSIs were collected from the Gaofen-1 satellite in February 2015 and the Gaofen-2 satellite in January 2017, which have a resolution of 2 and 1 m/pixel, respectively. We downsample the high-resolution posttemporal RSI to 2 m/pixel to match the pretemporal RSI, as shown in Fig. 7. The matched RSIs with the size of 5667 × 5423 are cropped to 256 × 256 patches with an overlap on the right and bottom and randomly divided these patches to 311 pairs for training, 44 pairs for validation, and 90 pairs for testing.

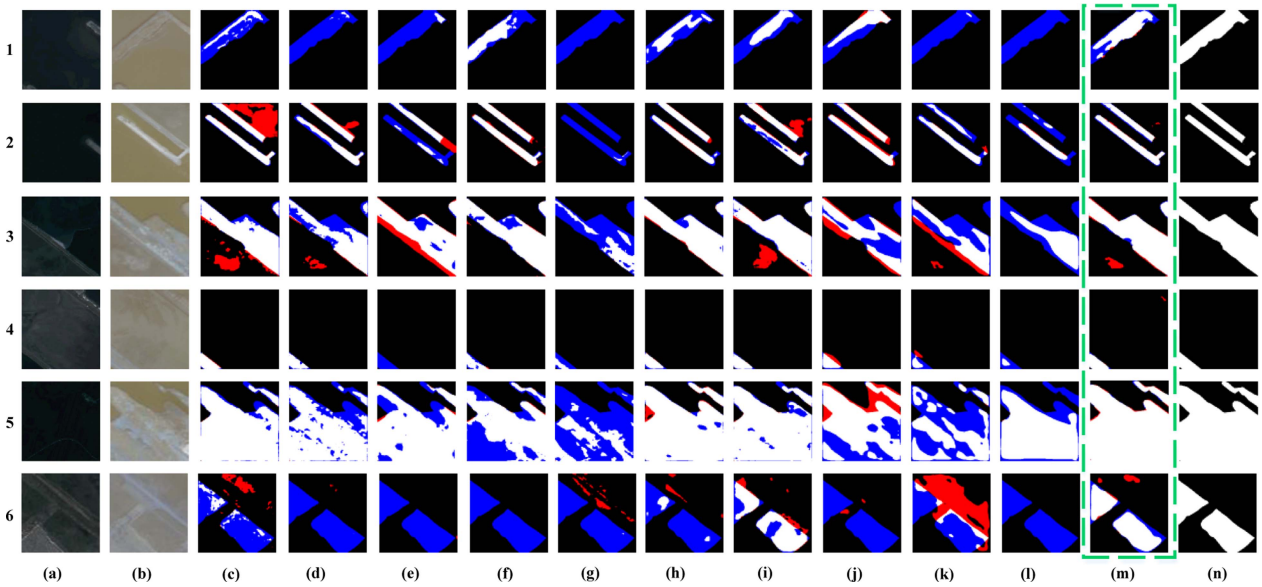


Fig. 8. Visual Comparison of PDANet and SOTA models on JinzhouBay test set. (a) Time t_1 RSIs. (b) Time t_2 RSIs. (c) FC-EF. (d) FC-SiamConc. (e) DSAMNet. (f) STANet. (g) SNUNet. (h) BiT. (i) P2V. (j) MSCANet. (k) ICIFNet. (l) EGCTNet. (m) PDANet (ours). (n) Ground truth. Where white represents true positive, black represents true negative, red represents false positive, and blue represents false negative.

LEVIR-CD: This medium-size dataset contains 637 pairs of RSIs covering various buildings in the city, with a size of 1024 $\times$ 1024 and a resolution of 0.5 m/pixel. We use the training set, validation set, and test set provided by the original authors in a 7 : 1 : 2 ratio and crop the RSIs with the size of 256 $\times$ 256.

WHU-CD: This medium-size dataset contains one pair of RSIs recording urban building changes before and after the earthquake disaster, with a size of 32 508 $\times$ 15 354 and a resolution of 0.3 m/pixel. we crop the original paired image into 256 $\times$ 256 patches with an overlap on the right and bottom and randomly divide these patches into 5460 pairs for training, 779 pairs for validation, and 1561 pairs for testing.

SYSU-CD: This large dataset contains 20 000 pairs of RSIs capturing the various land and marine use type changes in Hong Kong, China, with a size of 256 $\times$ 256 and a resolution of 0.5 m/pixel. These change types include newly built urban buildings, suburban dilation, groundwork before construction, change of vegetation, road expansion, and sea construction. We use the dataset divided by the original authors, where 8000 pairs for training, 4000 pairs for validation, and 4000 pairs for testing.

We perform all comprehensive experiments by using PyTorch on a PC with an Intel Core i9-13900KF 3.00-GHz CPU, 24-GB RAM, and an NVIDIA RTX 3090Ti GPU. During the training process, parameters of PDANet are randomly initialized by Kaiming initialization [52]. Adam algorithm is used to optimize our network with a batch size of 16, and the initial learning rate is 0.001 in the first 50 epochs and decays linearly to e^{-7} in the following 50 epochs. The hyperparameters α is set to 10 and we discuss it in Section V. The edge scale s is set to 1 on JinzhouBay and is set to 2 on other datasets.

In addition, we use the following metrics, including overall accuracy (OA), precision (Pre), recall (Rec), F1-measure (F1), Kappa coefficient (KC), and mean intersection-over-union (mIoU) to quantitatively evaluate the performance of the model. Validation is transacted after training in each epoch, and then the best model on the validation set evaluates the performance of the model on the test set.

B. Comparison With State-of-the-art (SOTA) Methods

We compare the proposed PDANet with ten SOTA CD methods, including FC-EF [24], FC-SiamConc [24], DSAMNet [22], STANet [23], SNUNet [25], BIT [13], P2V [34], MSCANet [15], ICIFNet [26], and EGCTNet [14]. To be fair, we reproduce these comparison methods using their default hyperparameters and the same experimental setup as our study. The qualitative and quantitative comparative results are analyzed as follows.

Qualitative comparison: Figs. 8–11 show the visualization comparison results of PDANet with SOTA CD models on JinzhouBay, LEVIR, WHU, and SYSU test sets, respectively. Where white, black, red, and blue represent true positive, true negative, false positive, and false negative, respectively. From these figures, it can be seen that PDANet more accurately highlights changed areas with clear seams and fewer holes. The detailed analysis of the qualitative comparison results is as follows.

1) Visualization Comparison on JinzhouBay dataset: JinzhouBay dataset has an underexposed t_1 RSI and an overexposed t_2 RSI, which results in blurred details. This makes it more difficult to identify changed areas with long-range dependencies. For example, samples from **1st**, **2nd**, and **3rd** rows in Fig. 8 show illegible object boundaries. Comparison models suffer from more false positives and false negatives. In addition, it is a common phenomenon that the same changed area shows inhomogeneous pixels due to differences in building materials of land reclamation. For example, RSIs from **3rd** and **5th** rows show nonuniform spectrum and luminance distribution. In contrast, PDANet generates better changed areas without defects

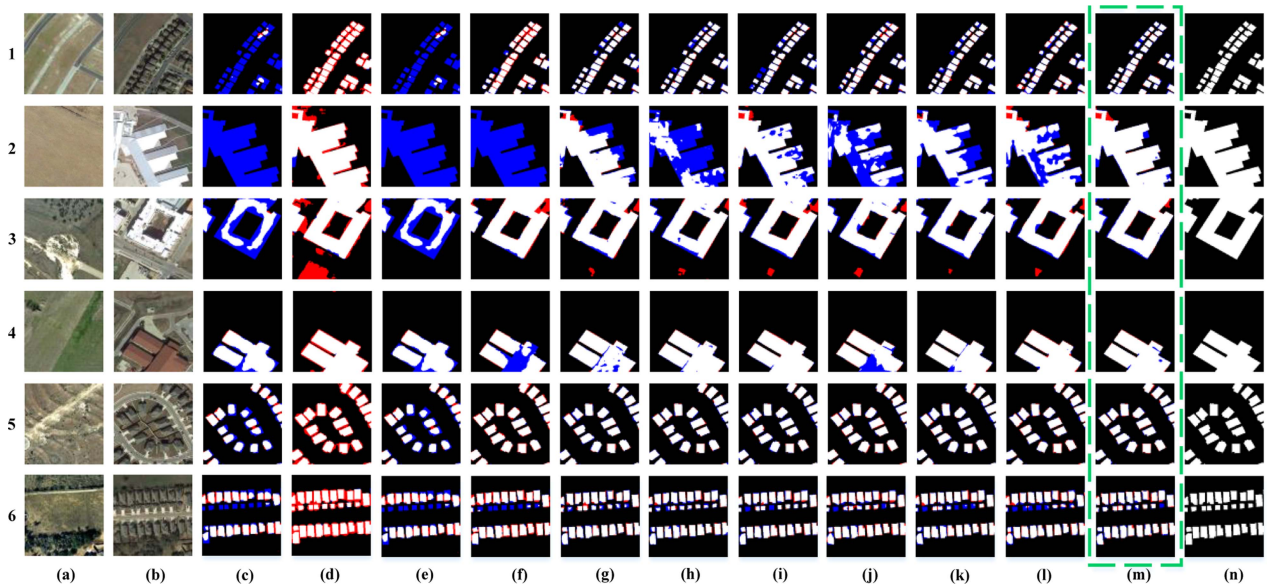


Fig. 9. Visual Comparison of PDANet and SOTA models on LEVIR test set. (a) Time t_1 RSIs. (b) Time t_2 RSIs. (c) FC-EF. (d) FC-SiamConc. (e) DSAMNet. (f) STANet. (g) SNUNet. (h) BiT. (i) P2V. (j) MSCANet. (k) ICIFNet. (l) EGCTNet. (m) PDANet (ours). (n) Ground truth.

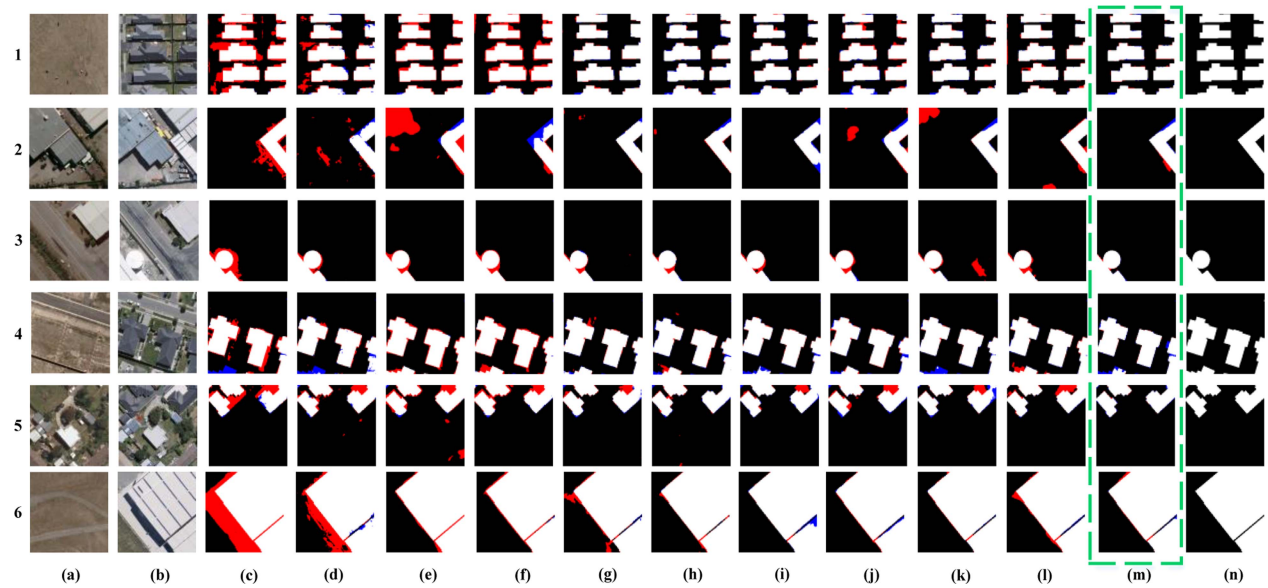


Fig. 10. Visual Comparison of PDANet and SOTA models on WHU test set. (a) Time t_1 RSIs. (b) Time t_2 RSIs. (c) FC-EF. (d) FC-SiamConc. (e) DSAMNet. (f) STANet. (g) SNUNet. (h) BiT. (i) P2V. (j) MSCANet. (k) ICIFNet. (l) EGCTNet. (m) PDANet (ours). (n) Ground truth.

or holes. Similarly, the actual reclamation change between two RSIs in **6th** row is not visually noticeable. Only P2V and the PDANet recognized more changed pixels.

2) Visualization Comparison on LEVIR Dataset: LEVIR dataset is oriented toward changed buildings with dense and multiscales. From **1st**, **5th**, and **6th** rows in Fig. 9, PDANet accurately locates small-scale buildings without adhesions. On the other hand, nonplanar building roof panels are subject to chromatic aberration due to illumination, resulting in low-level feature differences between pixels in the same building. In this scenes, PDANet can overcome this significant underlying difference of the same semantic concept, and generate CD maps that are more closely with ground truth, such as **2nd**, **3rd**, and **4th** rows. This advantage may be attributed to PDANet's deep semantic feature-guided DF progressive decoding and DF-guided independent feature reconstruction.

3) Visualization Comparison on WHU Dataset: WHU dataset focuses on building change but has more pseudochanges, such as the car change in **1st** and **2nd** rows of Fig. 10 and land change in **5th** rows. FC models, DSAMNet, and STANet have weak power to recognize pseudochanges, and other models also failed to obtain accurate edge details due to shadow interference.

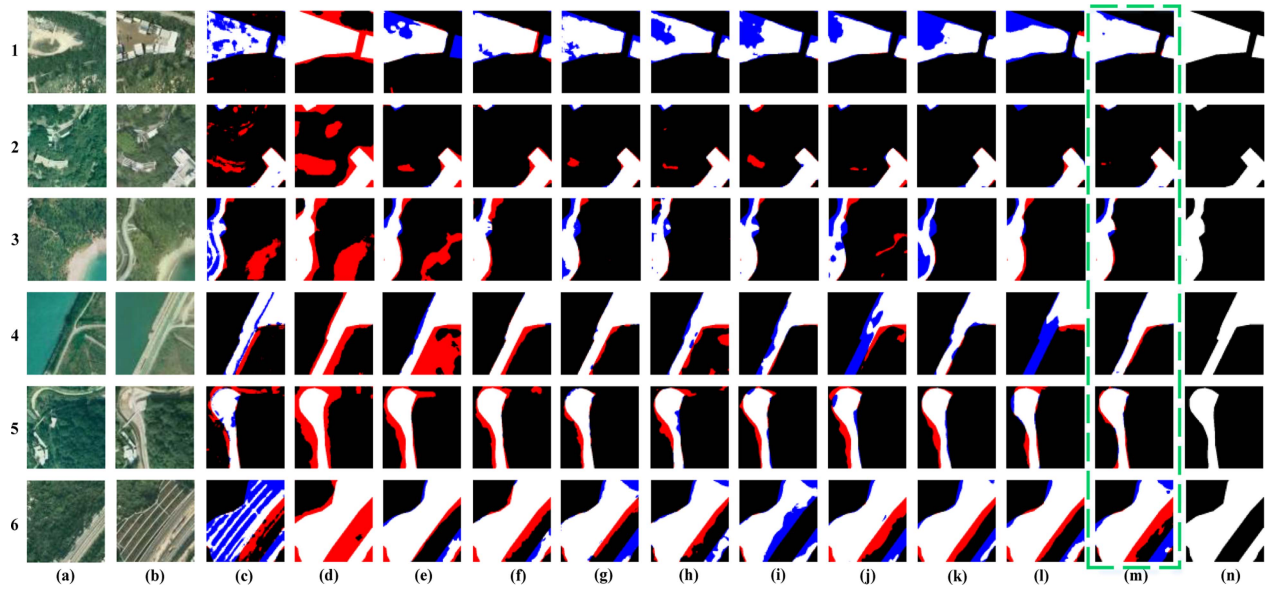


Fig. 11. Visual Comparison of PDANet and SOTA models on SYSU test set. (a) Time t_1 RSIs. (b) Time t_2 RSIs. (c) FC-EF. (d) FC-SiamConc. (e) DSAMNet. (f) STANet. (g) SNUNet. (h) BiT. (i) P2V. (j) MSCANet. (k) ICIFNet. (l) EGCTNet. (m) PDANet (ours). (n) Ground truth.

In particular, it is difficult to accurately identify the changed objects with special geometric configurations, such as the domed building in **3rd** row and the narrow passage between two buildings in **6th** row. Comparison models can barely outline the edges of the changed areas accurately in the above-mentioned scenes, while PDANet benefits from effectively leveraging and transmitting structural information to obtain near-labeled CD maps.

4) Visualization Comparison on SYSU Dataset: SYSU dataset provides multiple types of changed ground objects, which requires the CD model to learn the features with rich semantic attributes. In addition, the dataset has relatively rough labels of the changed areas due to the large sample size. Therefore, all detection results (see Fig. 11) appear general compared to the other three datasets. Fortunately, the proposed PDANet obtains relatively accurate changed areas in most scenes, although the edges are not flat. This is because the skip connections of the encoding–decoding process retain more global information. Similarly, SNUNet with dense connections also performs better. It is worth noting that PDANet is not as good at detecting the scene shown in the 6th row of Fig. 11 as the other methods. Specifically, 1) PDANet misclassifies the repaired roads as changes (red pixels), whereas SYSU is concerned with the road expansion change; 2) PDANet without finding real road expansion changes due to vegetation cover (blue pixels). This indicates that PDANet does not have a good understanding of the semantics of road expansion and road repairs, and has limited performance in scenes where multiple change types are intertwined. Therefore, establishing deep semantic relations and improving the generalization performance of the model is also a key focus of future research.

Quantitative comparison: Table I reports the quantitative comparison results of all models on the four test sets. PDANet has the best performance in comprehensive metrics of OA, F1, KC, and mIoU, and also has competitive performance on Pre and Rec. Specifically:

1) the gaps between PDANet and the SOTA models are more obvious on the small-size JinzhouBay dataset, where the F1/KC/mIOU is 2.66%, 2.74%, and 2.15% higher than that of the second-ranked P2V;
2) PDANet has a more moderate advantage compared to the other methods on the medium-size LEVIR and WHU datasets, where the F1/KC/mIOU is 0.35% (1.04%), 0.36%(1.08%), and 0.30%(0.94%) higher than that of the second-ranked ICIFNet (P2V) on LEVIR (WHU) dataset;
3) PDANet also presents a significant advantage on the large-size SYSU dataset, where the F1/KC/mIOU is 0.97%, 0.99%, and 0.69% higher than the second-ranked BIT.

On the other hand, we summarize the ED type of each model in Table I. Compared with the SED and DSED (FC-SiamConc and SNUNet), our DTED structure has slightly less one-sided performance in Pre or Rec. However, DTED can better balance between Pre and Rec, indicating that our model is more capable of discriminating pixel change types where the ratio is severely imbalanced. Since DTED concept is based on DSED and DDED, it shows superior performance in comparison. In addition, PDANet uses ESM and a coarse-grained prior edge to jointly learning structural information of the changed areas, which achieves excellent performance over EGCTNet which uses an edge map to mask the CD map.

C. Ablation Study

In this part, we validate the role and computational costs of different components in PDANet. All ablation experiments were performed using the same training configuration as PDANet. The details of the simplified models based on PDANet are described as follows.

TABLE I
QUANTITATIVE COMPARISON STUDIES WITH DIFFERENT CD MODELS

Models	ED types	JinzhouBay						LEVIR					
		OA	Pre	Rec	F1	KC	mIOU	OA	Pre	Rec	F1	KC	mIOU
FC-EF	SED	98.07	76.81	71.38	73.99	72.99	78.37	96.32	61.09	76.68	68.00	66.08	73.85
FC-SiamConc	DSED	97.96	83.30	58.81	68.95	67.92	75.26	96.59	60.19	**97.35**	74.39	72.67	77.81
DSAMNet	DDED	97.03	61.70	60.49	61.09	59.55	70.47	98.18	85.62	77.12	81.15	80.19	83.19
STANet	DDED	98.40	**95.95**	60.90	74.51	73.73	78.87	98.51	84.71	86.41	85.55	84.77	86.60
SNUNet	DSED	97.15	88.38	30.00	44.79	43.69	62.99	98.95	91.31	87.71	89.48	88.92	89.93
BIT	DDED	98.80	95.21	72.49	82.31	81.70	84.35	98.93	92.12	86.30	89.11	88.55	89.62
P2V	DSED	98.95	91.86	79.66	85.33	84.78	86.66	98.97	**92.57**	86.64	89.51	88.96	89.96
MSCANet	DDED	97.06	63.09	56.72	59.73	58.21	69.79	98.81	88.58	88.01	88.30	87.67	88.90
ICIFNet	DDED	96.56	56.91	43.44	49.27	47.52	64.59	98.98	91.89	87.72	89.76	89.22	90.18
EGCTNet	DDED	97.90	89.49	51.49	65.36	64.36	73.20	98.78	89.75	88.23	88.09	87.45	88.72
PDANet	**DTED**	**99.09**	89.66	**86.39**	**87.99**	**87.52**	**88.81**	**99.00**	91.13	89.12	**90.11**	**89.58**	**90.48**

Models	ED types	WHU						SYSU					
		OA	Pre	Rec	F1	KC	mIOU	OA	Pre	Rec	F1	KC	mIOU
FC-EF	SED	96.91	52.41	93.40	67.14	65.65	73.67	82.54	62.36	65.53	63.91	52.40	63.16
FC-SiamConc	DSED	97.93	64.06	87.96	74.13	73.08	78.38	77.52	51.28	**93.76**	66.30	51.51	60.36
DSAMNet	DDED	98.33	68.05	**95.42**	79.45	78.60	82.09	88.57	74.45	78.44	76.39	68.85	73.89
STANet	DDED	98.94	79.53	92.49	85.52	84.97	86.80	89.50	80.08	73.84	76.83	70.05	74.83
SNUNet	DSED	98.97	80.98	90.99	85.69	85.16	86.96	90.04	85.31	69.77	76.76	70.50	75.18
BIT	DDED	99.33	88.77	91.60	90.16	89.81	90.70	90.11	84.52	71.06	77.21	70.95	75.50
P2V	DSED	99.47	94.83	89.19	91.93	91.65	92.26	90.07	**85.78**	69.42	76.73	70.51	75.19
MSCANet	DDED	98.74	76.78	90.03	82.88	82.23	84.73	89.59	82.95	70.29	76.10	69.50	74.47
ICIFNet	DDED	98.97	81.18	90.52	85.59	85.06	86.88	88.47	79.96	68.19	73.61	66.29	72.25
EGCTNet	DDED	98.65	73.88	92.75	82.24	81.55	84.22	88.52	77.91	71.60	74.62	67.22	72.85
PDANet	**DTED**	**99.54**	**95.39**	90.68	**92.97**	**92.73**	**93.20**	**90.26**	82.90	73.98	**78.18**	**71.94**	**76.19**

The bold entities in table represents the optimal result under the corresponding metrics.

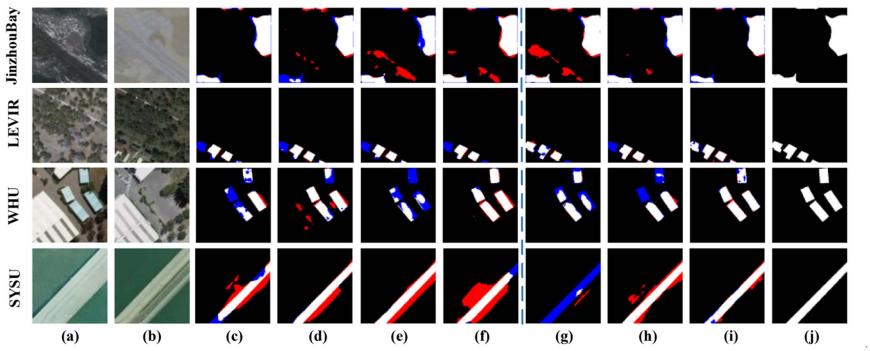


Fig. 12. Visual comparison of the components of PDANet. (a) Time t_1 RSIs. (b) Time t_2 RSIs. (c) W/o 1st ESM. (d) W/o 2nd ESM. (e) W/o ESMs. (f) W/o GIs. (g) W/o DMB. (h) W/o BABs. (i) Full model. (j) Ground truth.

1) Removing the ESM acting on the encoded features $\mathbf{F}^*_{e,1}$ ($* \in \{t_1, t_2\}$), denoted as $\mathbf{w/o\ 1st\ ESM}$.
2) Removing the ESM acting on the encoded features $\mathbf{F}^*_{e,2}$ ($* \in \{t_1, t_2\}$), denoted as $\mathbf{w/o\ 2nd\ ESM}$.
3) Removing two ESMs from the encoder at the same time, denoted as $\mathbf{w/o\ ESMs}$.
4) Removing the GI unit from two ESMs, denoted as $\mathbf{w/o\ GIs}$.
5) Deleting the DMB and do not use the first concatenate operation for ADM in the BABs, denoted as $\mathbf{w/o\ DMB}$.
6) Deleting two BABs, which denoted as $\mathbf{w/o\ BABs}$.

Ablation validation of different components:

In Fig. 12, the simplified models roughly locate the changed areas but lack complete structure and clear edge, and much background interference is not suppressed. Compared with the simplified models without ESM [see Fig. 12(c)–(e)], PDANet reduces many false positives (red pixels) and obtains more consistent structural. Compared with the simplified models without DMB or BABs [see Fig. 12(g) and (h)], PDANet reduces many false negatives (blue pixels) and obtains more finer-grained difference information.

Table II shows the qualitative results. Compared with the simplified models without ESM (IDs 1∼4), the introduction of the ESM brings the relative gains of F1, KC, and mIOU are about 0.21∼6.96%, 0.15∼7.26%, and 0.15∼5.52% on four datasets. The 1st ESM contributes more to the model than the 2nd ESM, which is understandable due to the 1st ESM learns richer contextual information. In addition, the introduction of GI also improved the model's F1, KC, and mIOU about 0.36∼4.89%, 0.37∼5.08%, and 0.33∼3.92%, respectively. It also shows that GI adaptively controlling the effect of multiscale collaborative features helps to improve model performance. On the other hand, the simplified model without DMB (ID∼5) has a dramatic performance sink compared with the PDANet. DMB brings the

TABLE II
ABLATION STUDIES ABOUT THE COMPONENTS OF PDANET

ID	Models	Params(M)	JinzhouBay						LEVIR					
			OA	Pre	Rec	F1	KC	mIOU	OA	Pre	Rec	F1	KC	mIOU
1	w/o 1st ESM	3.81	98.76	86.88	79.78	83.18	82.54	84.96	98.97	91.53	87.85	89.65	89.10	90.08
2	w/o 2nd ESM	3.81	98.87	**90.40**	78.98	84.30	83.72	85.85	98.99	91.60	88.18	89.86	89.32	90.26
3	w/o ESMs	2.78	98.53	80.68	81.38	81.03	80.26	83.29	98.96	91.42	87.77	89.55	89.01	90.00
4	w/o GIs	4.24	98.73	85.22	81.09	83.10	82.44	84.89	98.97	91.77	87.70	89.69	89.15	90.12
5	w/o DMB	3.90	98.71	84.94	80.86	82.85	82.18	84.69	98.93	**91.79**	86.76	89.21	88.64	89.70
6	w/o BABs	4.06	98.91	85.14	**86.91**	86.02	85.45	87.17	98.96	91.32	87.88	89.57	89.02	90.01
7	Full model	4.85	**99.09**	89.66	86.39	**87.99**	**87.52**	**88.81**	**99.00**	91.13	**89.12**	**90.11**	**89.58**	**90.48**

ID	FLOPs(G)	Time(ms)	WHU						SYSU					
			OA	Pre	Rec	F1	KC	mIOU	OA	Pre	Rec	F1	KC	mIOU
1	18.89	14.23	99.46	94.83	88.83	91.73	91.45	92.09	89.99	83.05	72.31	77.31	70.92	75.47
2	31.28	14.54	99.51	95.19	90.17	92.61	92.36	92.87	90.30	83.91	72.81	77.97	71.79	76.09
3	14.76	11.33	99.46	93.78	90.07	91.89	91.61	92.22	90.25	**85.23**	70.98	77.45	71.30	75.75
4	29.34	14.32	99.52	**95.63**	89.78	92.61	92.36	92.87	90.04	83.48	72.01	77.32	70.99	75.52
5	26.74	14.58	99.42	95.11	87.19	90.98	90.68	91.42	87.72	76.84	68.60	72.49	64.62	71.10
6	26.99	13.33	99.51	94.46	**90.73**	92.56	92.30	92.82	90.17	83.32	72.92	77.77	71.50	75.88
7	35.42	15.81	**99.54**	95.39	90.68	**92.97**	**92.73**	**93.20**	**90.26**	82.90	**73.98**	**78.18**	**71.94**	**76.19**

The bold entities in table represents the optimal result under the corresponding metrics.

relative gains of F1, KC, and mIOU are about 0.90∼5.69%, 0.94∼7.32%, and 0.78∼5.09% on four datasets. It indicates that the model with a DDED structure is insufficient to fully learn the DFs due to ground truth only positively supervising the convolutional layer that generates the DF. The introduction of BABs (ID∼6) brings the relative gains of F1, KC, and mIOU are about 0.41∼1.97%, 0.43∼2.07%, and 0.31∼1.64% on four datasets. It proves that BABs suppress change-irrelevant by reinforcing a more useful DF has a positive effect on the model. In conclusion, the ablation studies demonstrate the effectiveness of the different components.

Calculation cost analysis

We use the model parameters (Params), floating point operations per second (FLOPs), and inference time (Time) to measure the calculation costs of the simplified models. Params and FLOPs are calculated based on input RSI with the size of $3 \times 256 \times 256$. Time calculates the time consumed by forward propagation of the paired images with batchsize=1. From Table II, the differences in Params and Time of the simplified models are not significant. This suggests that the DTED structure of our model and the designed ESM did not impose a heavier burden. FLOPs have significant variations but are generally acceptable. The variation is mainly reflected in the fact that multiscale convolutional kernels are used in the DA unit.

V. DISCUSSION

A. Sensitivity Analysis of Hyperparameter

The hyperparameter α is used to adjust the contribution of edge loss to the total loss. Considering that α influences the PDANet's learning on the edge pixels of changed areas, we conduct a series of experiments for different α to analyze how it affects the model. After many attempts, the range of α is set to $\{0, 0.01, 0.1, 1, 10, 100, 1000\}$. Table III shows the performance of PDANet trained under different α. From it, the PDANet's emphasis on edge pixels of the changed areas does not increase with α. It is very easy to understand that a higher α causes the model to overemphasize the edge pixels and penalize the other pixels. PDANet performs optimally on the other three datasets except for SYSU when $\alpha = 10$. For SYSU, the model's F1/KC/mIOU at $\alpha = 10$ is only 0.07%/0.31%/0.23% lower than that at $\alpha = 1$, which is acceptable. Therefore, all PDANet-related experiments were set with an α of 10. In particular, $\alpha = 0$ represents only the binary CD loss function L_{cd} is used to train PDANet, resulting in F1/KC/mIOU that is 0.21%∼3.43%, 0.21%∼3.55%, and 0.18%∼2.77% lower than that at $\alpha = 10$. Fig. 13 also shows that a reasonable introduction of edge loss L_{edge} can enhance the internal structural consistency detection.

TABLE III
COMPARISON STUDIES OF PDANET TRAINING USING DIFFERENT HYPERPARAMETRES

α	JinzhouBay			LEVIR		
	F1	KC	mIOU	F1	KC	mIOU
0	84.56	83.97	86.04	89.90	89.37	90.30
0.01	86.65	86.13	87.71	89.99	89.47	90.38
0.1	87.23	86.74	88.19	90.09	89.57	90.47
1	85.81	85.28	87.04	90.00	89.47	90.39
10	**87.99**	**87.52**	**88.81**	**90.11**	**89.58**	**90.48**
100	78.31	77.50	81.37	89.96	89.43	90.35
1000	73.23	72.28	77.93	89.65	89.11	90.08

α	WHU			SYSU		
	F1	KC	mIOU	F1	KC	mIOU
0	92.63	92.38	92.89	77.74	71.47	75.86
0.01	92.10	91.83	92.41	77.69	71.49	75.88
0.1	91.88	91.61	92.22	77.54	71.17	75.64
1	92.36	92.10	92.64	**78.35**	**72.25**	**76.42**
10	**92.97**	**92.73**	**93.20**	78.18	71.94	76.19
100	90.65	90.33	91.13	77.29	70.69	75.29
1000	83.05	82.50	84.96	76.33	69.74	74.63

The bold entities in table represents the optimal result under the corresponding metrics.

B. Effect of Different-Grained Prior Edges

As described in Section III, s represents the grained scale of edge ground truth. It is worth noting that a more coarse-grained

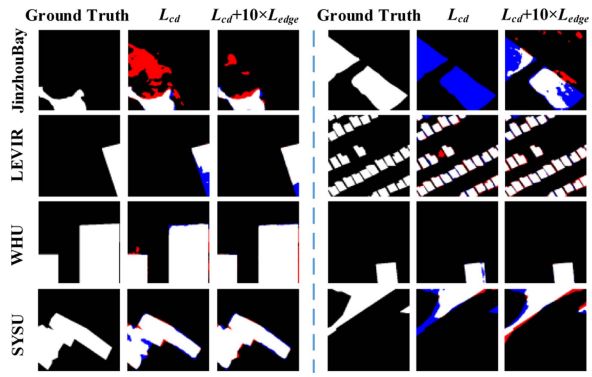


Fig. 13. Visual comparison of PDANet trained with different loss functions on the test sets.

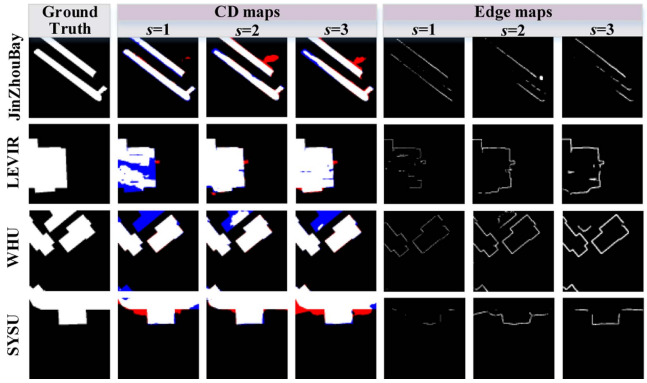


Fig. 14. Visual comparison of PDANet trained with different-grained prior edges $s = \{1, 2, 3\}$ on the test sets.

TABLE IV
COMPARISON STUDIES OF PDANET TRAINING USING DIFFERENT-GRAINED PRIOR EDGES

s	JinzhouBay			LEVIR		
	F1	KC	mIOU	F1	KC	mIOU
1	**87.99**	**87.52**	**88.81**	90.06	89.54	90.44
2	86.94	86.41	87.92	**90.11**	**89.58**	**90.48**
3	84.65	84.05	86.09	89.93	89.41	90.33
s	WHU			SYSU		
	F1	KC	mIOU	F1	KC	mIOU
1	92.28	92.02	92.57	77.57	71.25	75.71
2	**92.97**	**92.73**	**93.20**	**78.18**	**71.94**	**76.19**
3	92.78	92.53	93.02	76.85	70.47	75.15

The bold entities in table represents the optimal result under the corresponding metrics.

prior edge significantly deviates from the target changed areas, which leads to misclassification of the CD result. Therefore, we only explore the effects on PDANet when $s = \{1, 2, 3\}$. Fig. 14 shows the qualitative results of PDANet that is trained using different-grained prior edges. CD maps at $s = 2$ are closest to CD ground truths in the other three datasets except JinzhouBay, and the corresponding edge maps have a relatively compact closed topology. This is supported by the quantitative results shown in Table IV. For JinzhouBay, the CD map obtains the best performance when $s = 1$, but the edge map shows an incomplete

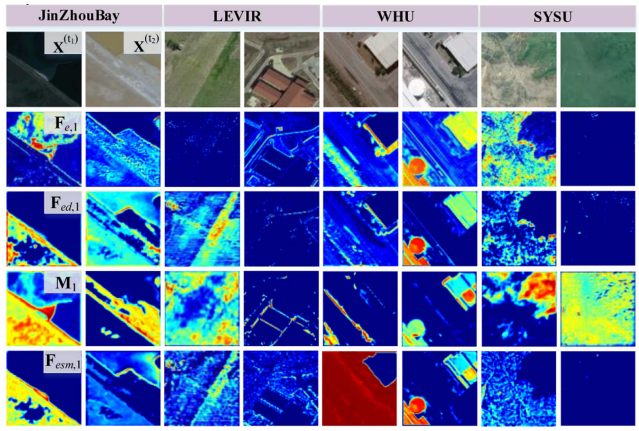


Fig. 15. Visualization feature maps of the ESM.

contour of the changed area. The phenomenon may have arisen because: 1) insufficient edge pixel fitting due to a smaller sample of the JinzhouBay; 2) rough edge details in RSIs due to illumination differences and relatively large resolution (2 m/pixel). Since the performance of PDANet on the JinzhouBay shows large differences at different s, this article uses single-pixel prior edge (i.e., $s = 1$) for experiments on the JinzhouBay, and uses coarse-grained prior edge with $s = 2$ on other three datasets.

C. ESM Feature Visualization

In order to better observe the effect of the ESM on the structural consistency of ground objects, we use the Grad-Cam method to visualize the input feature of ESM ($\mathbf{F}_{e,1}$), DA feature ($\mathbf{F}_{ed,1}$), learnable weight mask ($\mathbf{M}_1$), and integration feature ($\mathbf{F}_{\text{esm},1}$). Since the lowest level encoded features are more intuitive to show the structure of the ground object, only the feature maps associated with 1st ESM are shown in Fig. 15, and all feature maps are reshaped to 256 × 256. In Fig. 15, red indicates high spatial attention values, blue indicates low spatial attention values and darker color represents the stronger the feature. Moreover, based on the gate mechanism (3), high attention values in $\mathbf{M}_1$ indicate that $\mathbf{F}_{ed,1}$ has greater weight in feature integration, and high attention values indicate that $\mathbf{F}_{e,1}$ has greater weight. From Fig. 15, ESM is able to highlight salient objects and separate cluttered background. For example:

1) changed area of $\mathbf{F}_{e,1}^{(t_2)}$ in JinzhouBay gradually gains higher attention with the DA and the GI, and weakens the influence of the background in $\mathbf{F}_{e,1}$ by $\mathbf{M}_1$;
2) $\mathbf{F}_{e,1}^{(t_2)}$ in LEVIR presences both "building" and "road" as significant objects, while the LEVIR aims to focus on changes in "buildings." It can be seen that "road" is absorbed into the background in the $\mathbf{F}_{\text{esm},1}$ due to the background positions in $\mathbf{M}_1$ have similar values;
3) the changed area in the bottom left corner of $\mathbf{F}_{e,1}^{(t_2)}$ in WHU is redder, indicating increasing attention;
4) the blurred vegetation range in the upper right corner of $\mathbf{F}_{e,1}^{(t_1)}$ in SYSU is also becoming clearer. In summary, the ESM contributes to PDANet to improve the structural consistency of changed objects, in which GI can highlight changed objects more.

TABLE V
COMPARISON OF COMPUTATIONAL COST OF DIFFERENT MODELS

Models	Params (M)	FLOPs (G)	Time (ms)
FC-EF	1.35	3.57	4.63
FC-SiamConc	1.55	5.32	5.75
DSAMNet	16.95	75.35	11.18
STANet	12.21	12.55	10.18
SNUNet	10.20	44.13	17.75
BIT	3.01	8.45	5.90
P2V	5.42	32.86	7.81
MSCANet	16.59	14.70	12.43
ICIFNet	25.83	25.28	36.33
EGCTNet	106.16	38.41	9.32
PDANet	4.85	35.42	15.81

D. Model Efficiency Analysis

Table V provides a comparison of computational efficiency between PDANet and SOTA CD models in terms of Params, FLOPs, and Time, whose computational laws are the same as the ablation study in Section IV. It can be seen that PDANet has relatively smaller Params but higher FLOPs and Time. Since the multithreaded parallelized implementation, PDANet enjoys refined CD results with the sacrifice the computational costs. Therefore, PDANet has slightly higher parameters than FC-EF, FC-SiamConc, BIT, and P2V. Since ESMs act on low-level encoded features, PDANet requires more computational space for addition and multiplication operations. Therefore, PDANet has higher FLOPs than comparison models except for DSAMNet, SNUNet, and EGCTNet, and it also has higher inference time except for the SNUNet and ICIFNet. Although PDANet outperforms the recent comparison models, there is still a requirement in the future to explore the possibility of reducing the computational complexity, such as reasonably introducing group convolution or depthwise convolution in spite of a small amount of information is lost.

VI. CONCLUSION

In this article, we propose a PDANet with edge sensitivity for performing CD task in optical RSIs. Considering internal holes in CD results caused by ignoring global context information, we propose an edge sensitivity encoder to capture the long-range dependency and improve the structural integrity of the changed area. On the other hand, considering the importance of separating irrelevant change from DFs in RSIs with low spectral resolution, we propose a difference amplification decoder to enhance the representation of the changed object by reconstructing individual RSI with full-stage DF guidance. In addition, we also introduce a different-grained prior edge scheme in the training stage to guide PDANet to learn more accurate edge details. In this way, PDANet captures high-quality CD maps in four types of datasets that can accurately highlight changed areas. The results of extensive comparative and ablation experiments fully confirm the effectiveness and superiority of PDANet. Since the multithreaded parallelized implementation and fine-gained learning of underlying encoded features, PDANet encounters difficulties in reducing the computational complexity and will work on these limitations. In addition, establishing deep semantic relations to further improve the generalization of the model is also listed as a key research priority for the future.

REFERENCES

[1] G. Liu, Y. Gousseau, and F. Tupin, "A contrario comparison of local descriptors for change detection in very high spatial resolution satellite images of urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3904–3918, Jun. 2019.

[2] J. Fan, K. Lin, and M. Han, "A novel joint change detection approach based on weight-clustering sparse autoencoders," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 2, pp. 685–699, Feb. 2019.

[3] Y. Hao, Z. Chen, Q. Huang, F. Li, B. Wang, and L. Ma, "Bidirectional segmented detection of land use change based on object-level multivariate time series," *Remote Sens.*, vol. 12, no. 3, 2020, Art. no. 478.

[4] M. Song, Y. Zhong, and A. Ma, "Change detection based on multi-feature clustering using differential evolution for landsat imagery," *Remote Sens.*, vol. 10, no. 10, 2018, Art. no. 1664.

[5] Q. Li, R. Zhong, X. Du, and Y. Du, "TransUNetCD: A. hybrid transformer network for change detection in optical remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5622519.

[6] 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.

[7] Q. Shen, J. Huang, M. Wang, S. Tao, R. Yang, and X. Zhang, "Semantic feature-constrained multitask Siamese network for building change detection in high-spatial-resolution remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 189, pp. 78–94, 2022.

[8] J. Li, M. Hu, and C. Wu, "Multiscale change detection network based on channel attention and fully convolutional BiLSTM for medium-resolution remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 9735–9748, 2023.

[9] Deepanshi, R. Barkur, D. Suresh, S. Lal, C. S. Reddy, and P. G. Diwakar, "RSCDNet: A robust deep learning architecture for change detection from bi-temporal high resolution remote sensing images," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 7, no. 2, pp. 537–551, Apr. 2023.

[10] L. Wang, J. Zhang, Q. Guo, and D. Chen, "IFTSDNet: An interact-feature transformer network with spatial detail enhancement module for change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 2505105.

[11] Z. Wang, X. Wang, W. Wu, and G. Li, "Continuous change detection of flood extents with multisource heterogeneous satellite image time series," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4205418.

[12] Y. Zhao, P. Chen, Z. Chen, Y. Bai, Z. Zhao, and X. Yang, "A triple-stream network with cross-stage feature fusion for high-resolution image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5600417.

[13] 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.

[14] 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, 2022, Art. no. 4524.

[15] M. Liu, Z. Chai, H. Deng, and R. Liu, "A CNN-Transformer network with multiscale context aggregation for fine-grained cropland change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4297–4306, 2022.

[16] W. Liu, Y. Lin, W. Liu, Y. Yu, and J. Li, "An attention-based multiscale transformer network for remote sensing image change detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 202, pp. 599–609, 2023.

[17] K. Zhang, X. Zhao, F. Zhang, L. Ding, J. Sun, and L. Bruzzone, "Relation changes matter: Cross-temporal difference transformer for change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5611615.

[18] S. Liang, Z. Hua, and J. Li, "Enhanced self-attention network for remote sensing building change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 4900–4915, 2023.

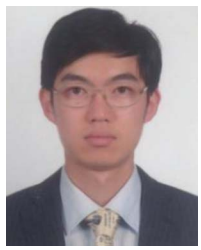
[19] Z. Song, X. Wei, X. Kang, S. Li, and J. Liu, "Toward efficient remote sensing image change detection via cross-temporal context learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5404210.

[20] Y. Zhang, W. Li, Y. Wang, Z. Wang, and H. Li, "Beyond classifiers: Remote sensing change detection with metric learning," *Remote Sens.*, vol. 14, no. 18, 2022, Art. no. 4478.

[21] W. Zhao, L. Mou, J. Chen, Y. Bo, and W. J. Emery, "Incorporating metric learning and adversarial network for seasonal invariant change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2720–2731, Apr. 2020.
[22] Q. Shi, M. Liu, S. Li, X. Liu, F. Wang, and L. Zhang, "A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5604816.
[23] 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.
[24] R. Caye Daudt, B. Le Saux, and A. Boulch, "Fully convolutional siamese networks for change detection," in *Proc. IEEE Int. Conf. Image Process.*, 2018, pp. 4063–4067.
[25] 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.
[26] 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.
[27] J. Qu, Y. Xu, W. Dong, Y. Li, and Q. Du, "Dual-branch difference amplification graph convolutional network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5519912.
[28] T. Lei et al., "Difference enhancement and spatial CSpectral nonlocal network for change detection in VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4507013.
[29] X. Zhang et al., "SMD-Net: Siamese multi-scale difference-enhancement network for change detection in remote sensing," *Remote Sens.*, vol. 14, no. 7, 2022, Art. no. 1580.
[30] D. Song, Y. Dong, and X. Li, "Context and difference enhancement network for change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9457–9467, 2022.
[31] R. Cong et al., "CIR-Net: Cross-modality interaction and refinement for RGB-D salient object detection," *IEEE Trans. Image Process.*, vol. 31, pp. 6800–6815, 2022.
[32] Y. Liang, C. Zhang, and M. Han, "RaSRNet: An end-to-end relation-aware semantic reasoning network for change detection in optical remote sensing images," *IEEE Trans. Instrum. Meas.*, vol. 73, 2024, Art. no. 5006711.
[33] S. Zhao, X. Zhang, P. Xiao, and G. He, "Exchanging dual-encoder CDecoder: A new strategy for change detection with semantic guidance and spatial localization," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4508016.
[34] M. Lin, G. Yang, and H. Zhang, "Transition is a process: Pair-to-video change detection networks for very high resolution remote sensing images," *IEEE Trans. Image Process.*, vol. 32, pp. 57–71, 2023.
[35] X. Hou, Y. Bai, Y. Li, C. Shang, and Q. Shen, "High-resolution triplet network with dynamic multiscale feature for change detection on satellite images," *ISPRS J. Photogramm. Remote Sens.*, vol. 177, pp. 103–115, 2021.
[36] 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.
[37] X. Zhou et al., "Edge-guided recurrent positioning network for salient object detection in optical remote sensing images," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 539–552, Jan. 2023.
[38] B. Sui, Y. Cao, X. Bai, S. Zhang, and R. Wu, "BIBED-Seg: Block-in-block edge detection network for guiding semantic segmentation task of high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1531–1549, 2023.
[39] Y. Guo, Y. Tao, Y. Chong, S. Pan, and M. Liu, "Edge-guided hyperspectral image compression with interactive dual attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5500817.
[40] S. Ren and Q. Liu, "Small target augmentation for urban remote sensing image real-time segmentation," *IEEE Trans. Intell. Transp Syst.*, vol. 25, no. 2, pp. 2076–2088, Feb. 2024.
[41] Z. Chen et al., "EGDE-Net: A building change detection method for high-resolution remote sensing imagery based on edge guidance and differential enhancement," *ISPRS J. Photogramm. Remote Sens.*, vol. 191, pp. 203–222, 2022.
[42] B. Yang, Y. Huang, X. Su, and H. Guo, "MAEANet: Multiscale attention and edge-aware siamese network for building change detection in high-resolution remote sensing images," *Remote Sens.*, vol. 14, no. 19, 2022, Art. no. 4895.
[43] 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.
[44] T. Lei et al., "Ultralightweight spatial–spectral feature cooperation network for change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4402114.
[45] F. Luo, T. Zhou, J. Liu, T. Guo, X. Gong, and J. Ren, "Multiscale diff-changed feature fusion network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5502713.
[46] Y. Jiang, L. Hu, Y. Zhang, and X. Yang, "WRICNet: A weighted rich-scale inception coder network for remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4705313.
[47] N. Shi, K. Chen, and G. Zhou, "A divided spatial and temporal context network for remote sensing change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4897–4908, 2022.
[48] Z. Yang, L. Zhu, Y. Wu, and Y. Yang, "Gated channel transformation for visual recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11791–11800.
[49] 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.
[50] Y. Gu, H. Xu, Y. Quan, W. Chen, and J. Zheng, "ORSI salient object detection via bidimensional attention and full-stage semantic guidance," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5603213.
[51] 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.
[52] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1026–1034.

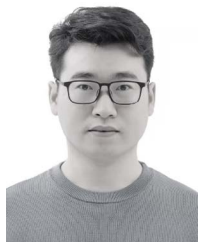
Yi Liang received the B.S. degrees in measurement and control technology and instrument from Liaoning Technical University, Fuxin, China, in 2016. She is currently working toward the Ph.D. degree in electronics and information with the Dalian University of Technology, Dalian, China.

Her research interests include remote sensing image change detection and deep learning.



Xinghan Xu received the B.S. degree in atmosphere science from Nanjing University, Nanjing, China, in 2010, and the M.S. degree in environment engineering from Kyoto University, Kyoto, Japan, in 2015.

He is currently with the Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian, China. His research focuses on the combination of numerical model and learning model in analysis and forecast comprehensive system.



Chengkun Zhang received the B.S. degrees in automation from the Ocean University of China, Qingdao, China, in 2013, and the Ph.D. degree in control theory and control engineering from the Dalian University of Technology, Dalian, China, in 2021.

He is Lecturer with the Intelligent Computing and Application Laboratory of Qinghai Province, Qinghai University, Xining, China. His research interests include feature extraction and classification of hyperspectral images.

Jianwei Liu received the Ph.D. degree in hydrology and water resources from the Dalian University of Technology, Dalian, China, in 2007.

He is currently an Associate Professor with the School of Water Resources Engineering, Faculty of Infrastructure Engineering, Dalian University of Technology. He has presided more than two projects funded by the National Natural Science Foundation and participated in six others. He has led two projects under the National Science and Technology Support Program. He is responsible for and serves as the main Researcher on more than 20 consulting and service projects for enterprises. He has authored or coauthored more than 50 academic papers, with more than 20 of them indexed by SCI and EI. His research interests include the utilization of watershed rainstorm resources and wetland ecological restoration based on remote sensing information.

Deyi Wang received the Ph.D. degree in control theory and control engineering from the Dalian University of Technology, Dalian, China, in 2023.

He is currently a Lecturer with the Faculty of Electronics and Communications, Guangdong Mechanical and Electrical Polytechnic, Guangzhou, China. His research focuses on image processing based on artificial intelligence and remote sensing image processing.

Min Han (Senior Member, IEEE) received the B.S. and the M.S. degrees in industrial enterprise automation from Dalian University of Technology, Dalian, China, in 1982 and 1993, respectively, and the M.S. and Ph.D. degrees in electrical engineering, and electrical and electronic systems, respectively, from Kyushu University, Fukuoka, Japan, in 1996 and 1999, respectively.

She is currently a Professor with the Key Laboratory of Intelligent Control and Optimization for Industrial Equipment of Ministry of Education, Dalian University of Technology. Her research interests include remote sensing image interpretation, complex system modeling and forecasting method, time series analysis and forecasting, and neural network.

Dr. Han is a Deputy Director of the Chinese Society of Instrumentation Youth Work Committee, a Committee Member of the Chinese Society of Artificial Intelligence, and an Organizing Chair of ISNN2013, ICICIP 2014, ICIST2016.