En-Decoded Index Guided Edge Refinement Network for Change Detection of Remote Sensing Image

Chunyan Yu, *Senior Member, IEEE*, Chi Yu, Feihong Zhou, Yulei Wang, *Member, IEEE*, and Qiang Zhang, *Member, IEEE*

***Abstract*—Change detection (CD) aims to analyze pairs of remote sensing images (RSIs) that are captured at different times to identify valuable information regarding changes in land features, which plays a crucial role in the fields of urban planning, environmental monitoring, and disaster assessment. Due to the impact of edge blur and occlusion, existing CD methods vulnerably generate the phenomenon of imprecise edge detection. In this article, we propose the en-decoded index guided edge refinement network (EIGER-Net) for CD of RSI by establishing a novel indexed edge representation mechanism, which effectively improves edge depiction with the combination of high-level semantic features and multilevel edge index information. Specifically, we construct the dual-time feature exchange module to reduce the inter-domain variance of the multilevel features and achieve refined feature extraction through the small target feature enhancement module. Subsequently, the presented en-decoded-index module is responsible for edge reconstruction with the index information involved in the multilevel fusion features during the decoding phase. With the indication of encoded and decoded index information, the proposed model generates the precise edge prediction for the CD task. Experimental results show that the EIGER-Net outperforms other compared CD models, achieving the highest IoU values of 93.54%, 83.98%, and 69.48% on the CDD, LEVIR-CD, and SYSU-CD datasets, respectively. Besides, the proposed method obtains the highest F1 score of 93.54% on the WHU-CD dataset. Edge detection experiments further demonstrate the effectiveness of EIGER-Net in identifying detailed and blurred edges in RSI.**

***Index Terms*—Change detection (CD), edge refinement, en-decoded index, remote sensing image (RSI), small target enhancement.**

I. Introduction

CHANGE detection (CD) of remote sensing image (RSI) refers to the process of comparing and analyzing dual-time RSIs to detect land surface changes [1]. With the rapid development of remote sensing technology, CD of RSI has provided the scientific basis for decision-making in various fields such as environmental monitoring [2], land use change survey [3], disaster analysis and mapping [4], urbanization construction research [5], and disaster exploration [6].

Early CD methods primarily focused on processing low-resolution RSI and were classified into pixel-based change detection (PBCD) [7] method and object-based change detection (OBCD) [8] method. The PBCD method treats each pixel of RSI as the basic analysis unit and compares the pixel variations of the temporal images [9], [10], [11]. Image differencing method [12] and principal component analysis [13], [14] are typical PBCD techniques. In contrast, the OBCD method considers the image object as the basic analysis unit from image pairs for comparison, which avoids the interference caused by light variations and occlusion to a certain extent. Consequently, the OBCD method is enabled to explore the spatial background and shape information of the object in RSI [15], [16], [17]. Although the above-mentioned methods have the advantages of fewer samples for training and strong interpretability, the detection accuracy of PBCD and OBCD methods is relatively insufficient for feature extraction due to the inherent complexity of dual-time RSIs.

In the last decade, deep learning (DL) technology has significantly accelerated the development of CD in high-resolution RSI [18]. Numerous CD methods have been proposed regarding the model and feature optimization. At present, the mainstream CD networks are divided into single-branch [19] and Siamese [20] networks. The single-branch networks perform feature extraction and fusion, and judge the fused features to obtain detection results by classifier. In contrast, the Siamese network extracts features from dual-time images through different branches with shared weights for CD [21]. Due to the complexity and diversity of the detected scenes, popular state-of-the-art (SOTA) CD models employ Siamese networks for the feature extraction framework [22], [23]. Typically, in [22], the authors proposed a series of CD models including FC-EF, FC-Siam-conc, and FC-Siam-diff. Specifically, FC-EF was a UNet-based model that concatenated temporal images as input channels for CD. FC-Siam-conc performed a jump connection between the multilevel feature maps from the two encoder branches and the corresponding layer of the decoder, while FC-Siam-diff extracted the feature difference calculation before the connection. Recently, a bunch of methods have made considerable progress in edge recognition for CD tasks. Typical implementations include edge loss constraint [24], edge-guidance modules [25], [26], superpixel segmentation edge constraints [27], and feature residual connectivity [28]. Although the aforementioned

Received 17 February 2025; revised 29 May 2025; accepted 28 June 2025. Date of publication 2 July 2025; date of current version 7 August 2025. The work was supported in part by the National Nature Science Foundation of China under Grant 62471079 and Grant 62401095, and in part by the Fundamental Research Funds for the Central Universities under Grant 3132017124. *(Corresponding author: Yulei Wang.)*

The authors are with the Center for Hyperspectral Imaging in Remote Sensing (CHIRS) at Information and Technology College, Dalian Maritime University, Dalian 116026, China (e-mail: yucy@dlmu.edu.cn; 17861271082@163.com; fhz@dlmu.edu.cn; wangyulei@dlmu.edu.cn; qzhang95@dlmu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2025.3585308

methods provide solutions, the variability in scale of RSI still leads to the inadequate edge representation and tiny object detection issue, making edge ambiguity an important problem.

In this article, we establish a novel indexed edge representation mechanism to construct the en-decoded-index edge refinement network (EIGER-Net) for CD of RSI. EIGER-Net effectively achieves the detection of change feature edges by combining the edge index information of multilevel features with semantic features. Specifically, we design a small target feature enhancement (STFE) module with dilated convolution that aims to enhance these representations of minor objects in RSIs. To further improve the accuracy of CD, especially in edge detection of RSI, we proposed the en-decoded-index (EDIdx) module to extract the index features that are closely associated with edge information. Notably, these index features are derived from the combined low-level features. Subsequently, the EDIdx module incorporates index features into the upsampling process and enables the model to achieve accurate edge detection of RSI.

The main contributions of this article are listed as follows:

1) We propose a novel EIGER-Net that is the first attempt to employ multilevel index information to guide the edge prediction for CD of RSI. Innovatively, the presented EDIdx module captures encoded and decoded indexes sourced from the fused low-level features, which contain rich, detailed information for accurate edge reconstruction. Besides, the collaboration of the generated indexes and semantic features is beneficial for continuous change detection.
2) We present the STFE module to refine the dual-temporal exchanged features, the offered module captures a range of contextual information for identifying small targets. In addition, we incorporate the attention mechanism to STFE to reassign channel weights and further enhance the representation of minor and detailed regions.
3) We propose a simple yet effective fusion method to obtain the information difference of multilevel features. The fused module is integrated into the sampling procedure to strengthen the edge refinement, which benefits the sequential IndexNet in extracting the encoded index information.
4) We have conducted sufficient experiments on four popular CD datasets including CDD [29], LEVIR-CD [30], SYSU-CD [31], and WHU-CD [32]. The extensive experimental results and analysis demonstrate that EIGER-Net significantly outperforms the detection metrics of the current SOTA models.

The rest of this article is organized as follows. Section II provides a brief review of existing methods. The details of our proposed model are described in Section III. Section IV introduces the experimental results and comparison analyses. Section V summarizes the conclusion of the article.

II. Related Work

In recent years, advancements in remote sensing technologies have significantly enhanced the resolution of RSI [33], sparking considerable interest and extensive methods in the field of high-resolution CD [34], [35]. In this section, we provide a review of CD methods by focusing primarily on traditional and DL approaches.

A. Traditional CD Methods

Over the past few years, with the development of machine learning (ML) technology, the innovation of CD methods has been greatly promoted. Briefly, ML-based methods require the training sample set consisting of labeled and image pairs to detect change targets in RSIs by extracting and selecting effective features. Popular ML algorithms are comprised of support vector machine [36], [37], artificial neural network [38], random forest [39], and decision tree [40]. Through the advantages of RSI processing and the advancement of detection technology, the ML-based CD approach generates better detection performance than PBCD and OBCD methods. Nevertheless, the limitation of feature extraction capability hinders the full exploitation of the rich information in high-resolution RSI. In addition, the ML algorithm struggles with feature fusion and hampers the ability to effectively integrate regions of varying sizes, which causes unsatisfactory edge and detail detection.

B. DL-Based Siamese CD Networks

In the traditional Siamese network structure, the image features extracted by the backbone network undergo change recognition through a decoder. Typically, the FC-Siam-conc network proposed by Daudt et al. [22] adopted U-Net [41] in the Siamese structure and combined the jump connections of two encoder branches. To suppress noise interference in detection results, Chen et al. [42] built a Siamese-AUNt network, which embedded a nonlocal attention mechanism in U-Net to reduce noise impact. Chen et al. [43] proposed a SARAS-Net network to solve the noise problem caused by scene changes and objects of different scales with the relation-aware module, the scale-aware module, and the cross transformer. Recently, the Transformer has gained widespread interest in CD tasks due to the perception of global context information. Liu et al. [44] integrated the attention mechanism and Transformer into the multiscale CNN network and presented an AMTNet to solve CD tasks with long- and short-distance dependencies simultaneously. Bandara and Patel [45] introduced a ChangeFormer network that reduces the complexity and the number of parameters by employing a hierarchical Transformer encoder and an MLP decoder in the network structure. To address the deficiency of Mamba in lacking frequency information, Xing et al. [46] proposed the FEM-Net model to employ DCT for frequency information extraction, thereby addressing the problem of poor detection of small and texture-variant features. Aiming at the deficiency in radiometric normalization of multitemporal images, Miao et al. [47] proposed the RS-NormGAN model, which was designed based on pseudo-invariant features to process different features, and a global-local attention mechanism was employed to address the spatial distortion. Zhu et al. [48] proposed the ChangeViT framework, which employed a simple ViT backbone to enhance large-scale change detection and achieved multiscale feature fusion through a detail-capturing module and a feature injector. Although the above-mentioned methods achieve relatively

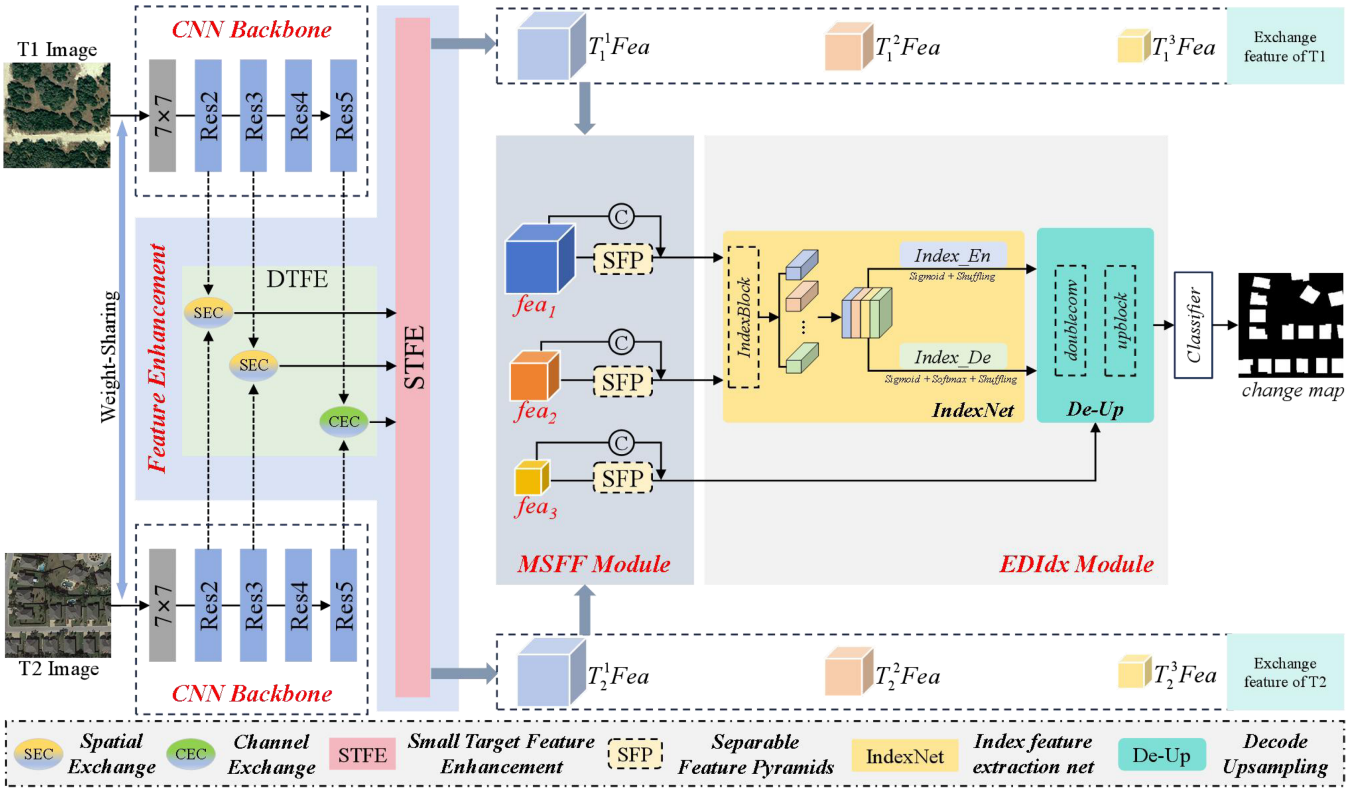

Fig. 1. Flowchart of the proposed EIGER-Net. Overall, the presented architecture is a Siamese CNN backbone network, which consists of the feature enhancement module, the multiscale feature fusion module, and the en-decoded index structure, where blue, orange, and yellow cubes represent the fused multilevel features.

satisfactory detection results in specific scenarios, employing the Transformer structure for CD increases the model complexity. Simultaneously, the traditional CD methods neglect the effective extraction of fine-grained object edge features in RSI, resulting in a certain degree of edge blurring or inaccuracy.

C. Edge-Guided Siamese CD Networks

Edges provide crucial clues for the definition of change areas and directly affect the accuracy and reliability of CD. Typicaliy, combining feature residual connectivity and attention mechanisms leads to effective edge representation. Fang et al. [49] proposed the SNUNet for CD, which has dense skip connections between the encoder and decoder to alleviate the uncertainty of pixels. Cheng et al. [50] proposed the ISNet network, which achieved clarification of semantic edges and enhancement of positional change responses by employing edge maximization and attention mechanisms. Extracting edge features through an edge detection module is also an effective approach to tackling the problem of edge blurring. Bai et al. [24] proposed EGRCNN to integrate the discriminant with edge structure information and achieved the accurate detection of buildings in RSI by designing a difference analysis module. Zhu et al. [25] presented a parallel encoding framework EGPNet to fully extract detailed and change information to enhance edge representation. Moreover, the edge-loss constraint exerts a significant influence on enhancing the expression of edge features. Although the above-mentioned model achieves the detection of edge changes to a certain extent, there are deficiencies in the extraction of features for the edges, resulting in the edge detection effect is still unsatisfactory. Unlike the existing approaches, we propose the novel EIGER-Net model for change detection, which achieves precise edge detection by effectively extracting the encoder-decoder index containing edge information from multilevel hierarchical features. The specific structure of the model proposed in this article will be described in detail in Section III.

III. Proposed Approach

A. Overall Structure of EIGER-Net

The proposed EIGER-Net adopts a CNN-based Siamese structure as shown in Fig. 1, which consists of a backbone network, feature enhancement module, multilevel feature fusion module, and an encoded-decoded-index module. Specifically, the backbone network aims to obtain multilevel features from dual-time RSIs. The feature enhancement module refines the feature with the presented DTFE and STFE modules. Among them, DTFE improves the representation of cross-temporal features in change regions through the operations of channel and spatial dimension exchange. The STFE module enhances the representation of small target features by capturing changes in contextual information employing dilated convolutional groups with different expansion rates to process the exchange features. Importantly, the multilevel EDIdx module serves to extract the encoded and decoded indexes containing edge information.

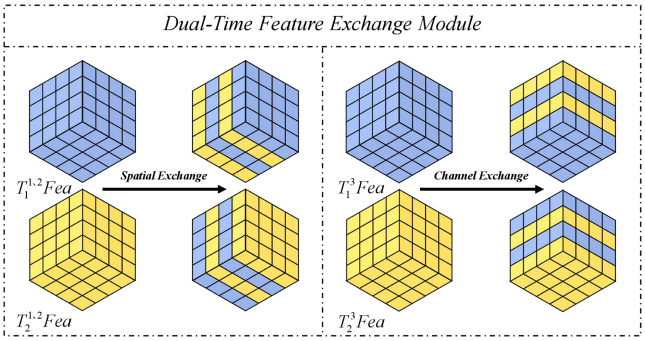

Fig. 2. Structural diagram of the DTFE model.

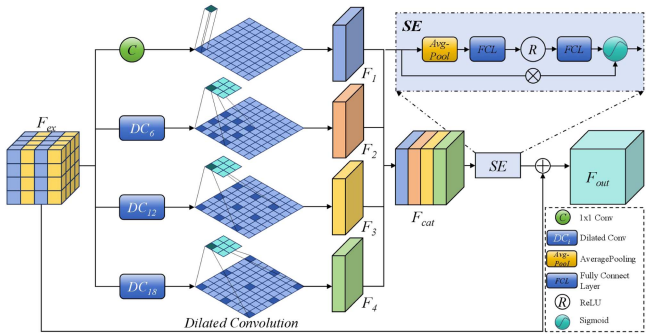

Fig. 3. Structural diagram of the STFE model.

Specifically, the encoded index improves the extraction of edge information in the feature encoding stage, and the decoded index integrates the edge information into high-level semantic features by upsampling, strengthening the representation of edge features in the detection process and enabling accurate edge prediction.

B. Siamese CNN-Based Backbone

In the EIGER-Net model, we employ ResNet as the backbone to extract low-level features containing edge information and high-level semantic features from dual-temporal images. We adopt different residual layer outputs to preserve rich details and global structure. Compared with the original structure of ResNet [51], we remove the initial fully connected layer and merely retain the 7×7 convolutional layer and residual structure, which reduces the computational complexity of the network. First, the initial features were extracted from the bitemporal images denoted as T1 and T2 through the max-pooling operation (MP). Next, through various levels of residual layers, we obtain the multilevel feature which is denoted as $\text{Res}_i (i \in \{2, 3, 5\})$. Each residual block contains a convolutional layer, a batch normalization layer [52], and a ReLU [53]. Finally, to reduce the computational complexity of the subsequent modules, we adjusted the feature channel to 32 by average pooling convolution with Sigmoid activation processing to obtain the multilevel feature $T_i^n \text{Fea} (i \in \{1, 2\}; n \in \{1, 2, 3\}) \in \mathbb{R}^{32 \times H \times W}$.

C. Multilevel Feature Enhancement

1) DTFE Module: In the model, to address the interdomain differences of dual-time images caused by light and seasonal changes, we perform parameter-free exchange [54] of features in the channel and spatial dimensions. As shown in Fig. 2, $T_1^{(n)}\text{Fea}$ and $T_2^{(n)}\text{Fea}$ represent the features produced by the same-level residual block of the double-branch network. Among them, the low-level features from the Res2 and Res3 residual blocks are responsible for spatial exchange, and the high-level features from the Res5 residual block aim to channel exchange. By exchanging elements at corresponding positions, the features incorporate dual-phase information, and the feature distributions between the dual branches become closer, which effectively reduces the gap between different domains and facilitates the model to learn a more general feature representation.

The formula for the feature exchange operation $T_i^{(n)} \text{Fea} (i \in \{1, 2\}; n \in \{1, 2, 3\})$ is defined as follows:

$$T_i^{(n)}\text{Fea}\,(\text{bs}, c, h, w) = \begin{cases} T_i^{(n)}\text{Fea}(\text{bs}, c, h, w) M(\text{bs}, c, h, w) = 0 \\ T_{3-i}^{(n)}\text{Fea}(\text{bs}, c, h, w) M(\text{bs}, c, h, w) = 1 \end{cases} \tag{1}$$

where bs, c, h, and w represent the batch size, channels, height, and width of the feature, respectively. M represents the mask, 0 denotes no exchange of feature elements, and 1 signifies the exchange in that dimension.

2) STFE Module: To enhance the ability to identify small targets and feature representation in RSI, the EIGER-Net model employs the advantages of dilated convolution in enlarging the receptive field and maintaining resolution to construct the STFE module to effectively extract the features of small targets. The specific processing procedure is shown in Fig. 3. First, we employ 1×1 convolution and a set of dilated convolution groups with dilation rates of 6, 12, and 18 to handle the exchange features to obtain the feature outputs F_1, F_2, F_3, and F_4. Subsequently, the acquired features are concatenated along the channel dimension to effectively integrate the features. Notably, by effectively leveraging information from various receptive fields, the model enhances the perception of small targets. In this article, dilated convolution is denoted as $\text{DilConv}_{3\times3}^{n}$, where n represents the dilation rate. The concatenated feature F_{cat} is calculated as follows:

$$\begin{cases} F_1 = \text{Conv}_{1\times1}(F_{\text{input}}) \\ F_2 = \text{DilConv}_{3\times3}^{6}(F_{\text{input}}) \\ F_3 = \text{DilConv}_{3\times3}^{12}(F_{\text{input}}) \\ F_4 = \text{DilConv}_{3\times3}^{18}(F_{\text{input}}) \end{cases} \tag{2}$$

$$F_{\text{cat}} = [F_1, F_2, F_3, F_4]\,. \tag{3}$$

Furthermore, we adopt the SE [55] attention mechanism in the STFE module to reassign channel weights to the spliced feature F_{cat} for feature optimization. The SE employs avg-pooling to extract spatial information in channels. It converts pooled features into one-dimensional vectors through a fully connected layer and processes with the sigmoid function to get normalized weights for each channel. Finally, element-wise addition generates the output feature F_{out} as defined

$$F_{\text{out}} = F_{\text{ex}} + F_{\text{cat}} \cdot \sigma(\text{FCL}(\text{ReLU}(\text{FCL}(\text{Avg}(F_{\text{cat}}))))) \tag{4}$$

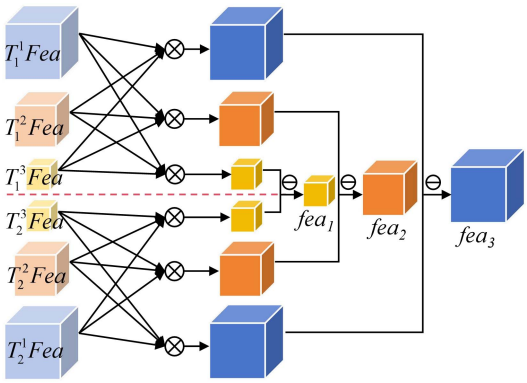

Fig. 4. Illustration of the MSFF module process.

where σ represents the sigmoid activation function, FCL means the fully connected layer, and Avg is the average pooling. F_{out} and F_{ex} denote the output feature and input dual-time exchange features, respectively.

D. MSFF Module

To fully leverage multilevel features and eliminate information discrepancies, we propose the simple and effective feature fusion module MSFF to generate discriminative features. The module takes the feature-enhanced multilevel features as inputs, which are processed by the MSFF module to obtain feature outputs, denoted as fea1, fea2, and fea3, respectively. The specific processing method as shown in Fig. 4, the feature fusion is achieved by sampling features of different scales in uniform dimensions, and by element-by-element multiplication and difference computation. The features are fused at multilevel by sampling as shown in (4), (5), and (6). Equation (7) is the difference calculation of the features

$$\text{fea}_{(x)}^1 = T_{(x)}^1\text{Fea} \otimes \text{US}(T_{(x)}^2\text{Fea}) \otimes \text{US}(T_{(x)}^3\text{Fea}) \quad (5)$$

$$\text{fea}_{(x)}^2 = \text{DS}(T_{(x)}^1\text{Fea}) \otimes T_{(x)}^2\text{Fea} \otimes \text{US}(T_{(x)}^3\text{Fea}) \quad (6)$$

$$\text{fea}_{(x)}^3 = \text{DS}(T_{(x)}^1\text{Fea}) \otimes \text{DS}(T_{(x)}^2\text{Fea}) \otimes T_{(x)}^3\text{Fea} \quad (7)$$

$$\text{fea}_i = \text{Sub}\left[\text{fea}_1^{(i)}, \text{fea}_2^{(i)}\right] \; i \in \{1,2,3\} \quad (8)$$

where $x \in \{1,2\}$,"$\otimes$"represents element-wise multiplication, US denotes the up-sampling operation, DS is the down-sampling operation, and Sub denotes the subtraction calculation.

Further, to reduce the computational cost of the module, the features fea_i are incorporated into the separable feature pyramid structure, which employs deep separable convolution [56] instead of traditional convolution and decreases the number of model parameters. After the above fusion, the fea_i is provided for the subsequent EDIdx module for edge refinement.

E. EDIdx Module

Compared to the traditional bilinear interpolation up-sampling operation, the unpooling operation employs max-pooling to guide the up-sampling process [57], which achieves the effective extraction of edge information from multilevel features. Motivated by this opinion, we propose the EDIdx module guided by max-pooling.

The EDIdx module is implemented by IndexNet for en-decoded index feature extraction and decode upsampling operations. The core idea of IndexNet is to capture salient edge information in feature maps through MPs and encode it into index features, enabling precise reconstruction of edge information during the decoding process. Fig. 5 shows the detailed structure of IndexNet. First, the feature fea_i is processed by the indexblock module, which consists of four sets of parallel convolution operations and MPs. Each set of operations employs convolution kernels of different sizes (n) and strides (m) to capture multiscale edge information. The MP is applied to extract local maxima from the feature maps, which typically correspond to the locations of edges or salient features. In this way, multilevel features generate a set of feature matrices X1, X2, X3, and X4, containing edge index information at different scales

$$X_n = \text{MP}(\text{Conv}_{1\times1}^1(\text{ReLU}(\text{BN}(\text{Conv}_{4\times4}^2(\text{fea}_i))))) \quad (9)$$

where $n \in \{1,2,3,4\}$, MP means the max-pooling operation, $\text{fea}_i(i \in \{1,2\})$represents a low-level fusion feature, $\text{Conv}_{n\times n}^m$ denotes the convolution operation, n is the size of the convolution kernel, and m is the step number.

Next, X_n is spliced along the channel dimension to generate the index matrix X_{cat}, which is processed by the following Sigmoid and Shuffling operations. Specifically, the preservation of edge information by Sigmoid nonlinear processing and pixel shuffling and rearranging of channels by shuffling operations, which facilitates the transition from low to high resolution. Consequently, the decoded index Index_De is obtained after the first shuffling operation. In contrast, the encoded index Index_En is generated at the end of the EDIdx block through an additional softmax processing layer, which further enhances the discriminability of the index features. The specific computation of encoded and decoded indices is shown as follows:

$$X_{\text{cat}} = [X_1, X_2, X_3, X_4] \quad (10)$$

$$\begin{cases} \text{Index\_De} = \text{shuffling}(\sigma(X_{\text{cat}})) \\ \text{Index\_En} = \text{shuffling}(\text{softmax}(\sigma(X_{\text{cat}}))) \end{cases}. \quad (11)$$

Furthermore, the Index_En is integrated into the features through matrix multiplication, enriching the feature representation with abundant edge information. Similarly, the Index_De is integrated into the decode upsampling process through matrix multiplication to achieve precise reconstruction of edge information. The specific execution process is shown in Fig. 6. Specifically, the black arrows represent the extraction process of multilevel encoded and decoded index features by IndexNet, and the red arrows indicate the De-Up process of incorporating decoded index into feature up-sampling operations to achieve edge refinement detection. Where ASPP_Decoder enables decoding of input high-level features to better incorporate low-level features, doubleconv is capable of extracting effective features from image data and upsampling high-level features. Feature fusion is achieved through upblock to obtain the detection result.

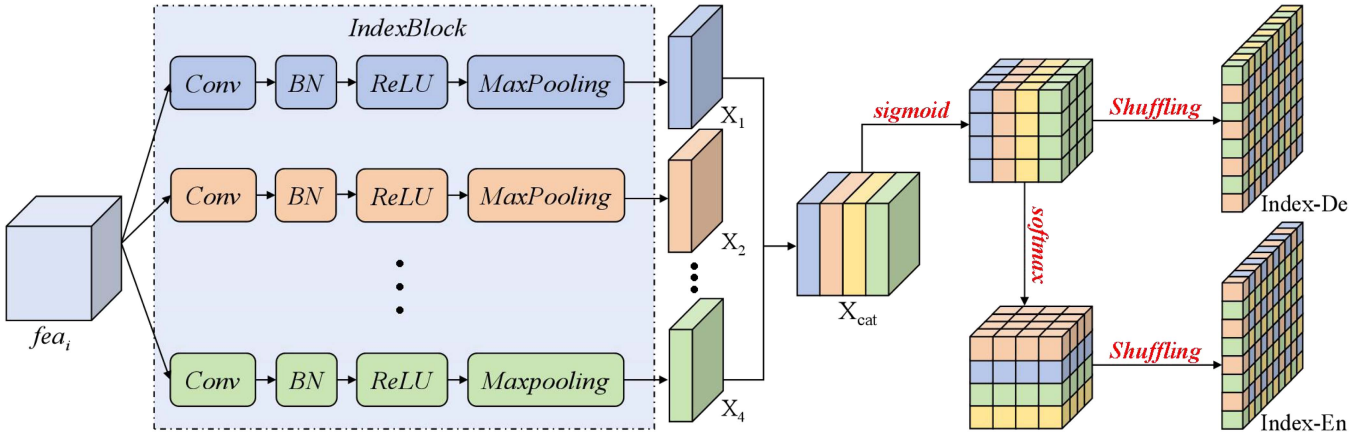

Fig. 5. Schematic diagram of the IndexNet structure.

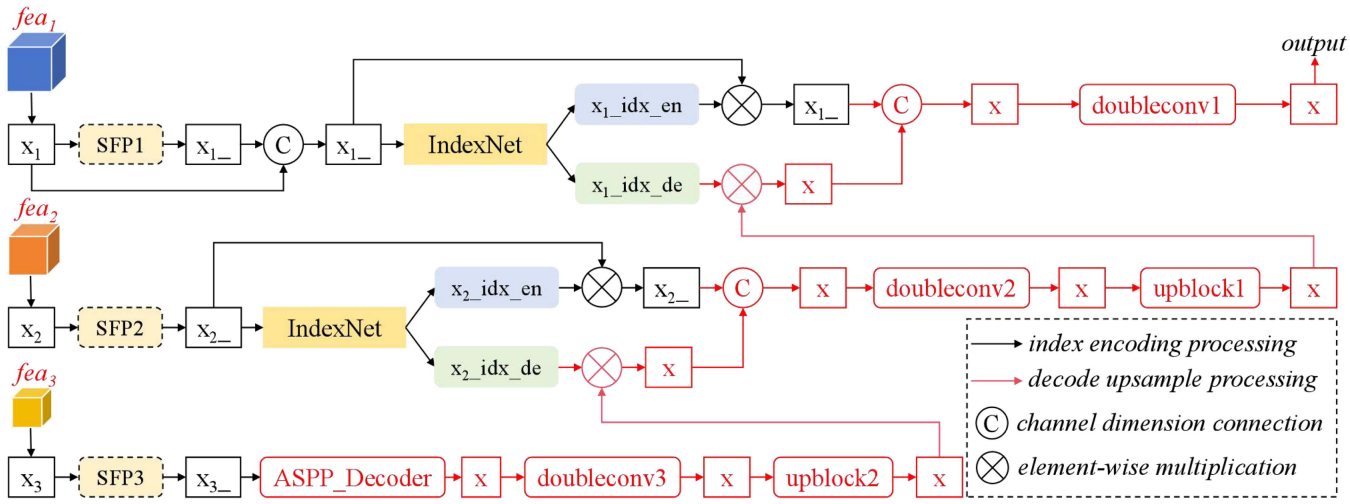

Fig. 6. Specific execution process of edge refinement of the EDIdx module.

F. Loss Function

To decrease the impact of class imbalance, we adopt Focal loss [58] as the loss function during the model training, which prompts the network to focus more on hard-to-identify samples. In the following experiments, we set the value of γ to 2, and the FocalLoss function is defined as follows:

$$L_{\text{focal}} = -(1-p_t)^{\gamma}\log(p_t) \tag{12}$$

$$p_t = \begin{cases} \hat{p} & \text{if} y = 1 \\ 1-\hat{p} & \text{otherwise} \end{cases} \tag{13}$$

where p_t represents the probability predicted by the model for the positive class, γ is the balance coefficient.

IV. Experiments and Analysis

In this section, we introduce the datasets required for experimental validation, the evaluation metric, a brief depiction of the compared approaches, and the final detection results. The public datasets include CDD [29], LEVIR-CD [30], SYSU-CD [31] and WHU-CD [32], and the specific parameters of each dataset are listed in Table I.

A. Datasets and Evaluation Metrics

CDD Dataset: The CDD dataset contains image pairs collected from Google Earth (GE). The dataset comprises 16 000 pairs of RSIs captured from the same area under different seasons and lighting conditions. Among them, 10 000 pairs are employed for training, 3000 pairs for validation, and 3000 pairs for testing. Each image has 256 × 256 pixels, with a spatial resolution ranging from 0.03 to 1 m/pixel. The image pairs cover various change objects of different sizes, such as large buildings, cars, trees, and roads.

LEVIR-CD Dataset: LEVIR-CD is obtained from GE and released by Beihang University. The dataset comprises 637 pairs of images captured at different times, with each image having 1024 × 1024 pixels and a spatial resolution of 0.5 m/pixel. The image pairs cover various change objects, including large buildings, residential houses, high-rise apartments, and small garages. Due to the large size of the image pairs, they are cropped

TABLE I
SPECIFIC PARAMETER INFORMATION OF THE DATASETS USED FOR COMPARATIVE EXPERIMENTS

Datasets	Image Source	Bands	Image Pairs	Resolution/m	Image Size	Train Pairs	Val Pairs	Test Pairs
CDD	Google Earth	3	16000	0.03~1	256×256	10000	3000	3000
LEVIR-CD	Google Earth	3	637	0.5	1024×1024	7120	1024	2048
SYSU-CD	Aerial Image	3	20000	0.5	256×256	12000	4000	4000
WHU-CD	Aerial Image	3	1	0.2	32507×15354	5204	742	1488

into nonoverlapping small image blocks of size 256 × 256 pixels for CD processing. Among these small blocks, 7120 pairs are selected for the training set, 1024 pairs for the validation set, and 2048 pairs for the test set.

SYSU-CD Dataset: SYSU-CD is released by Sun Yat-sen University with images collected from Aerial Images. The dataset contains various types of changes, with the main types including new urban buildings, suburban expansion, preconstruction groundwork, vegetation changes, road expansion, and offshore expansion. Due to the diversity of change types, this dataset presents certain detection challenges. It comprises 20 000 pairs of aerial images taken in Hong Kong from 2007 to 2014. The sample has a size of 256 × 256 pixels and a resolution of 0.5 m/pixel. Among all the samples, 12 000 pairs are utilized for the training set, 4000 pairs for the validation set, and 4000 pairs for the test set.

WHU-CD Dataset: The WHU-CD dataset was released by Wuhan University and consists of two pairs of aerial images collected in 2012 and 2016, covering large-scale changed buildings due to earthquakes and subsequent reconstruction over the years. The image pairs in the dataset have 32 507 × 15 354 pixels and a resolution of 0.2 m/pixel. We crop the image pairs into nonoverlapping blocks with a size of 256 × 256, which are randomly divided into training, validation, and test sets with a ratio of 7:1:2.

To intuitively demonstrate the performance of the EIGER-Net model compared to other detection algorithms in the CD field, five common metrics including precision (Pre), recall, F1-score (F1), intersection over union (IoU), and overall accuracy (OA) were adopted in our paper experiment. The formulas for calculating metrics are deined as follows:

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{14}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{15}$$

$$F1 = \frac{2 \times \text{Pre} \times \text{Recall}}{\text{Pre} + \text{Recall}} \tag{16}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \tag{17}$$

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{18}$$

where TP represents true positive samples predicted as positive by the model, TN indicates true negative samples predicted as negative by the model, FP means false positive samples predicted as positive by the model, and FN denotes false negative samples predicted as negative by the model.

B. Comparison Algorithm Description

In this section, we evaluate other SOTA methods including FC-EF [22], FC-Siam-conc [22], FC-Siam-diff [22], SNUNet [49], ChangeFormer [45], AMTNet [44], ISNet [50], EGPNet [25], EGRCNN [24], ChangeViT [48], and FEM-Net [46]. The source codes of the comparison models are obtained from the URLs provided by the authors in the papers. For articles that are without URLs, we utilize the official code from *paperswithcode.com* for experimental comparison. All comparative experimental results are based on the assay results provided in the original papers.

C. Experimental Implementation Details

In this experiment, all comparative experiments were conducted on a single NVIDIA GeForce RTX 3050ti GPU with 4GB of VRAM. The code was implemented with PyTorch. EIGER-Net employs ResNet-50 pretrained on ImageNet as the backbone for feature extraction in the overall network model. To accelerate model training and improve convergence efficiency, we adopted the Adam optimizer with a weight decay of 0.0001. During training, the initial learning rate was set to $lr = 0.0001$.

For the CDD dataset, LEVIR-CD dataset, SYSU-CD dataset, and WHU-CD dataset, the batch size was uniformly configured to batch_size = 4, and the number of epochs was arranged to num_epoch = 100. Throughout the training process, we employed Early Stopping to prevent overfitting, where the training process would be terminated prematurely if the validation loss did not decrease for ten consecutive epochs. All experiments utilized the Focal Loss function as the optimization objective, and the performance metrics on the validation set were computed at the end of each epoch to monitor the training progress of the model.

D. Compare Results and Discussion

1) Experimental Results of the CDD: Table II illustrates the quantitative comparison results on the CDD dataset. As observed, our proposed EIGER-Net is significantly better than other models in terms of Recall, F1, and IoU parameter scores. Specifically, the IoU and F1 increased by 0.23% and 0.85% respectively compared with the suboptimal method (FEM-Net), and the Recall improved by 0.42% compared with the third-best method (AMTNet).

Fig. 7 shows the visualization results of each comparison method and the edge detection result of the EIGER-Net method. According to the results, for the regions with overall regular changes [see Fig. 7(1) and (5)], the EIGER-Net method achieves effective detection. For the microregions and edge changes in

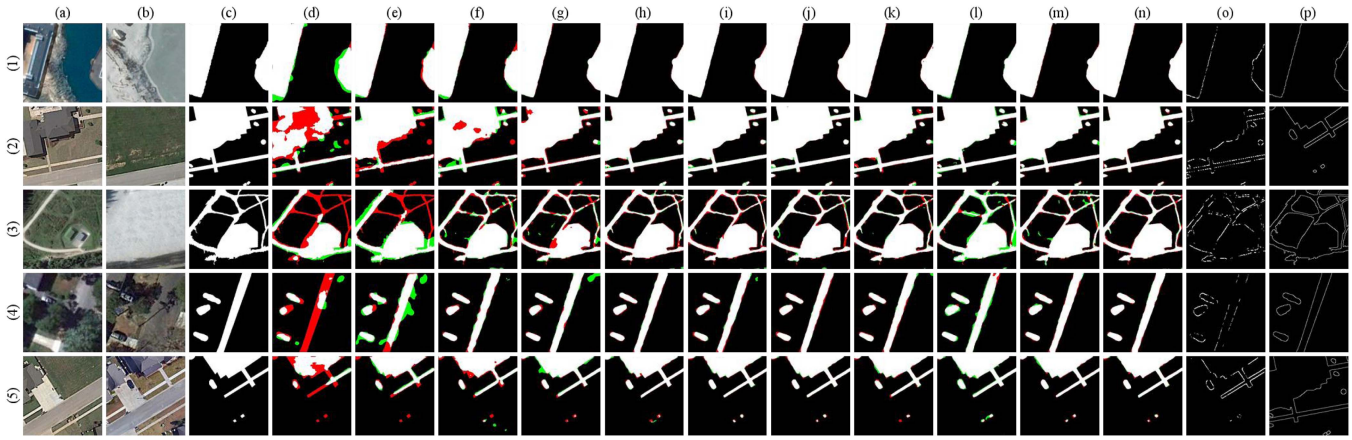

Fig. 7. Compares the detection results of eight detection methods on the CDD dataset. (a) Time1 image. (b) Time2 image. (c) Ground Truth (GT). (d) FC-Siam-conc. (e) FC-Siam-diff. (f) EGRCNN. (g) ISNet. (h) ChangeFormer. (i) SNUNet. (j) EGPNet. (k) AMTNet. (l) ChangeViT. (m) FEM-Net. (n) EIGER-Net (ours). (o) Edges of EIGER-Net. (p) Edges of GT. The colors represent different meanings: white indicates true positives, black indicates true negatives, red indicates false positives, and green indicates false negatives.

TABLE II
DETECTION RESULTS OF EACH NETWORK IN THE CDD DATASET (%)

Method	Pre	Recall	F1	IoU	OA
FC-EF	90.41	69.60	78.65	64.81	96.07
FC-Siam-conc	85.36	81.39	83.33	71.43	96.16
FC-Siam-diff	88.97	81.52	85.08	74.04	96.63
SNUNet	95.65	95.43	95.55	91.47	98.95
ChangeFormer	95.79	95.06	95.42	91.25	98.92
ISNet	95.29	94.41	94.85	90.21	98.79
EGPNet	**97.21**	95.10	**95.86**	**93.12**	**99.16**
EGRCNN	93.82	92.15	92.98	86.87	98.36
AMTNet	**96.17**	**96.20**	95.73	92.77	99.08
ChangeViT	94.30	93.26	93.79	92.87	98.40
FEM-Net	96.08	**97.61**	**95.89**	**93.31**	**99.48**
EIGER-Net(ours)	**96.38**	**96.62**	**96.74**	**93.54**	**99.13**

Note: "*" indicates that the value of the network model detection result is taken from the change detection result given in the original paper. Color-coded: **best**, **2nd-best** and **3rd-best**.

TABLE III
DETECTION RESULTS OF EACH NETWORK IN THE LEVIR-CD DATASET (%)

Method	Pre	Recall	F1	IoU	OA
FC-EF	86.91	80.17	83.40	71.53	98.15
FC-Siam-conc	84.61	83.54	84.07	72.52	98.39
FC-Siam-diff	90.29	82.76	86.36	75.99	98.67
SNUNet	**92.88**	85.93	89.27	80.62	98.95
ChangeFormer*	92.05	88.80	90.40	82.48	99.04
ISNet*	92.46	88.27	90.32	82.35	99.04
EGPNet*	92.03	**89.93**	**90.96**	**83.43**	**99.09**
EGRCNN	92.53	87.44	89.91	81.66	99.00
AMTNet	91.96	**89.57**	90.75	83.07	99.07
ChangeViT	92.04	89.56	90.78	83.12	99.07
FEM-Net	**93.27**	88.94	**91.21**	**83.69**	**99.11**
EIGER-Net(ours)	**93.79**	**89.87**	**91.14**	**83.98**	**99.10**

Note: "*" indicates that the value of the network model detection result is taken from the change detection result given in the original paper. Color-coded: **best**, **2nd-best** and **3rd-best**.

complex scenarios [Fig. 7(2) and (3)], the encoded and decoded index guides the preservation of edge features and restores the real change boundaries. From the edge visualization detection results, we conclude that EIGER-Net restores the edges of regions and small targets to a large extent, which is beneficial for edge prediction and accuracy enhancement. Particularly, the areas outlined by the red frame in the detection result image indicate the unique advantage of the presented STFE module in detecting small targets.

2) Experimental Results of the LEVIR-CD: The quantitative evaluation index results of the LEVIR-CD are shown in Table III. Evidently, the F1 and IoU based on the fully convolutional FC-EF model are the lowest at 83.40% and 71.53% respectively. Other models based on convolution and attention mechanisms achieve relatively higher performance. Among all the comparison methods, the EIGER-Net model achieves excellent data in terms of Pre, F1, IoU, and OA, which are 93.79%, 91.14%, 83.98%, and 99.10%, respectively. Compared with the second-best method (FEM-Net), IoU and Pre increase by 0.29% and 0.52% individually, but F1 and OA are lower compared to FEM-Net.

Fig. 8 shows the visualization detection results of each method in different scenarios (houses and woods) on the LEVIR-CD dataset. For large-scale regional changes [as shown in Fig. 8(2)], our method yields the best recognition effects. In the detection of the edges of the changing region [especially in (5) in Fig. 8], our model still achieves accurate recognition in the detection results. In addition, in terms of the accurate detection of the area, the false (red areas) and missed detection (green areas) rates are also lower than those of other methods.

3) Experimental Results of the SYSU-CD: The comparison results of each detection model on the SYSU-CD dataset are demonstrated in Table IV. According to the data in the table, the EIGER-Net method is superior to other SOTA models in terms of Recall, F1 and IoU, reaching the highest values of 79.29%, 81.12%, and 69.48% respectively. For this more challenging detection dataset, our IoU has increased by 0.53% compared to the suboptimal method (FEM-Net), indicating that EIGER-Net accurately captures changes when facing complex area detection. However, the Pre value of this model method on SYSU-CD is relatively low of 83.08% .

Fig. 9 shows the visualization results of all the compared networks on the SYSU-CD dataset. Besides, we provide the

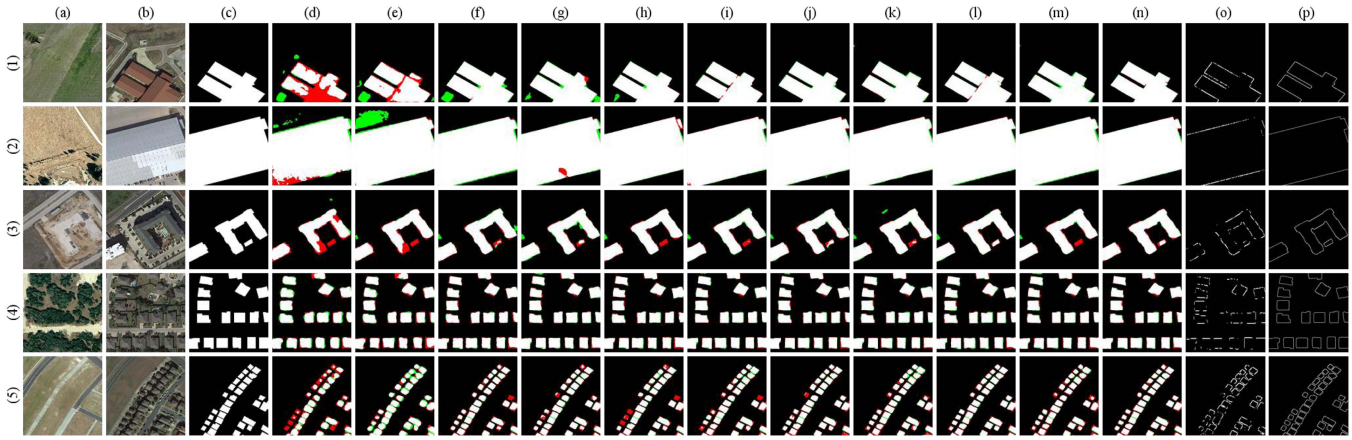

Fig. 8. Compares the detection results of eight detection methods on the LEVIR-CD dataset. (a) Time1 image. (b) Time2 image. (c) Ground Truth (GT). (d) FC-Siam-conc. (e) FC-Siam-diff. (f) SNUNet. (g) EGRCNN. (h) ISNet. (i) ChangeFormer. (j) EGPNet. (k) AMTNet. (l) ChangeViT. (m) FEM-Net. (n) EIGER-Net (ours). (o) Edges of EIGER-Net. (p) Edges of GT. The colors represent: white for true positives, black for true negatives, red for false positives, and green for false negatives.

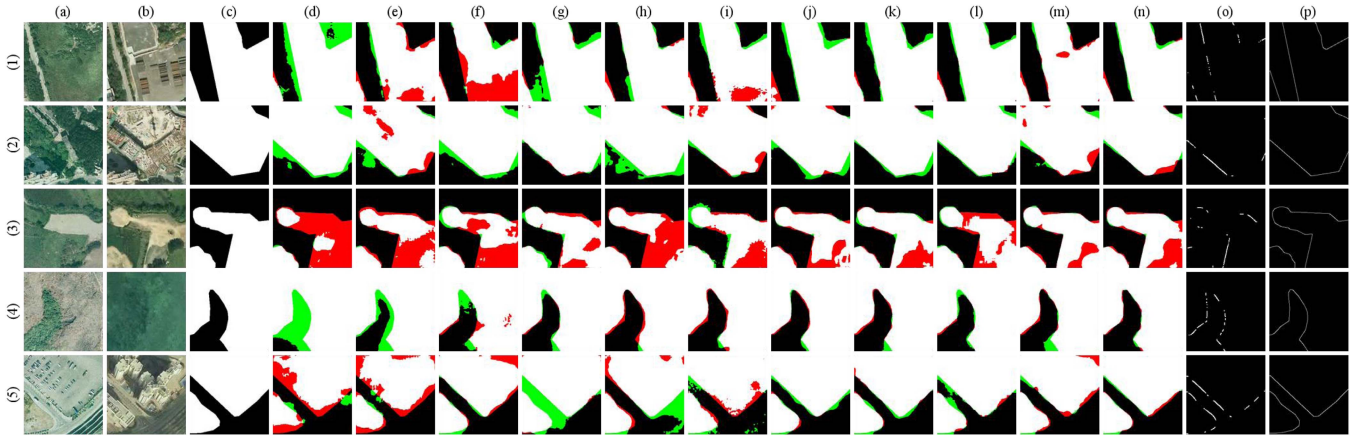

Fig. 9. Compares the detection results of eight detection methods on the SYSU-CD dataset. (a) Time1 image. (b) Time2 image. (c) Ground Truth (GT). (d) FC-Siam-conc. (e) FC-Siam-diff. (f) ChangeFormer. (g) SNUNet. (h) ISNet. (i) EGRCNN. (j) EGPNet. (k) AMTNet. (l) ChangeViT. (m) FEM-Net. (n) EIGER-Net (ours). (o) Edges of EIGER-Net. (p) Edges of GT. The color representation is as follows: white indicates true positives, black indicates true negatives, red indicates false positives, and green indicates false negatives.

TABLE IV
DETECTION RESULTS OF EACH NETWORK IN THE SYSU-CD DATASET (%)

Method	Pre	Recall	F1	IoU	OA
FC-EF	77.07	72.86	74.91	59.88	88.49
FC-Siam-conc	80.61	71.52	75.79	61.02	89.23
FC-Siam-diff	75.15	73.79	76.41	60.36	88.91
SNUNet	83.21	72.29	77.37	63.09	90.02
ChangeFormer	83.25	69.99	76.04	61.35	89.60
ISNet*	80.27	76.41	78.29	64.44	90.01
EGPNet	83.06	78.61	**80.98**	**68.13**	**91.33**
EGRCNN	**83.94**	75.29	79.38	65.81	90.78
AMTNet	81.79	**79.13**	80.44	67.27	90.92
ChangeViT	**85.37**	75.92	80.37	67.18	91.25
FEM-Net	**85.50**	**78.93**	**80.80**	**68.95**	**92.88**
EIGER-Net(ours)	83.08	**79.29**	**81.12**	**69.48**	**91.92**

Note: "*" indicates that the value of the network model detection result is taken from the change detection result given in the original paper. Color-coded: **best**, **2nd-best** and **3rd-best**.

edge detection result of the EIGER-Net and GT edge in the figure. Although the dataset is challenging, EIGER-Net performs well compared to other methods, with relatively accurate edge detection [as shown in Fig. 9(2) and (5)] and a relatively small area of missed and false detection regions. In addition, when facing small change discrimination situations [as shown in Fig. 9(4), from grassland to land], the EIGER-Net also captures satisfactory performance and achieves a higher detection accuracy than other methods.

4) Experimental Results of the WHU-CD: Table V presents the detection comparison experiment results of each detection model on the WHU-CD dataset. As known from the table, the EIGER-Net model achieves excellent performance in Recall, F1 score, IoU, and OA, reaching 93.25%, 92.86%, 86.39%, and 99.36%, respectively. Compared with the suboptimal model (FEM-Net), the value of Recall is increased by 0.95%, and the F1 score has improved by 0.93% . However, in terms of the IoU coefficient, the detection result of 86.39% is slightly inferior to that of the FEM-Net model, with a decrease of 0.02% . Fig. 10 shows the visualization detection results of each model on the WHU-CD dataset. As the images in this dataset are mostly with buildings affected by the light and shadows, WHU-CD is more challenging on edge detection than other datasets. The proposed

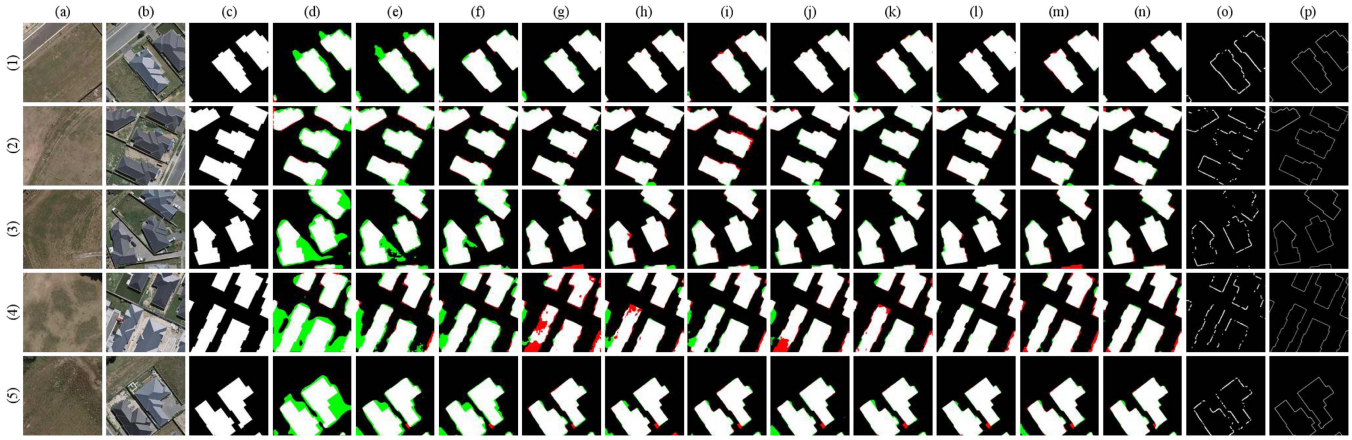

Fig. 10. Compares the detection results of eight detection methods on the WHU-CD dataset. (a) Time1 image. (b) Time2 image. (c) GT. (d) FC-Siam-conc. (e) FC-Siam-diff. (f) SNUNet. (g) ChangeFormer. (h) ISNet. (i) EGRCNN. (j) EGPNet. (k) AMTNet. (l) ChangeViT. (m) FEM-Net. (n) EIGER-Net (ours). (o) Edges of EIGER-Net. (p) Edges of GT. The color representation is as follows: white indicates true positives, black indicates true negatives, red indicates false positives, and green indicates false negatives.

TABLE V
DETECTION RESULTS OF EACH NETWORK IN THE WHU-CD DATASET (%)

Method	Pre	Recall	F1	IoU	OA
FC-EF	65.15	89.04	75.24	60.31	97.44
FC-Siam-conc	63.46	88.37	74.77	59.71	97.31
FC-Siam-diff	70.51	90.94	79.44	65.89	97.94
SNUNet	84.71	91.27	87.86	78.36	98.89
ChangeFormer	85.38	**93.12**	89.08	80.32	99.01
ISNet	89.04	92.52	90.75	83.06	99.17
EGPNet	90.29	**93.22**	91.74	84.74	99.27
EGRCNN	88.47	91.97	90.18	82.11	99.12
AMTNet	**92.68**	91.92	**92.29**	**85.31**	99.33
ChangeViT	**95.63**	86.66	90.93	85.07	**99.64**
FEM-Net	91.95	92.30	**91.93**	**86.41**	**99.78**
EIGER-Net(ours)	**92.15**	**93.25**	**92.86**	**86.39**	**99.36**

Note: "*" indicates that the value of the network model detection result is taken from the change detection result given in the original paper. Color-coded: **best**, **2nd-best** and **3rd-best**.

EIGER-Net model preserves detailed and abundant information for edge extraction and achieves relatively ideal results in edge detection.

5) Ablation Study: To further validate the effectiveness of the key modules in the EIGER-Net, we performed ablation experiments in this section. Ablation studies of critical modules are achieved by removing the corresponding modules from the complete network structure. Table VI shows the results of our ablation experiments on the four datasets with F1 score and IoU as evaluation metrics.

a) Effect of the DTFE module: To explore the effectiveness of exchanging dual-temporal features, we performed an ablation study on the DTFE module. Specifically, on the CDD and LEVIR-CD datasets, as shown in the first row of Table VI, the F1 is increased by 0.96% and 0.80%, and the IoU value is increased by 1.62% and 1.03%, respectively. It is analyzed that feature exchange promotes the effective fusion of different features between dual-time images and strengthens the feature representation ability. On the challenging SYSU-CD and WHU-CD datasets, our method achieved remarkable improvements with 1.55% and 1.31% enhancements in F1-score, accompanied by 2.41% and 1.98% boosts in IoU, respectively. These results unequivocally validate the superiority of DTFE in complex scenarios, which facilitates precise target identification and localization while significantly mitigating false positives and false negatives.

b) Effect of the STFE module: In the EIGER-Net, the STFE module innovatively enhances the recognition and characterization ability of small target features. As observed in the second row of Table VI, for the CDD and LEVIR-CD datasets with a large number of small targets, the IoU index has a relatively obvious improvement of 2.29% and 1.82% respectively. The improved IoU indicates that the STFE module enhances the identification and detection of small targets.

Moreover, the increased accuracy of the SYSU-CD and WHU-CD datasets that contain many changes in regions and buildings demonstrates the effectiveness of the STFE in complex scenarios.

Furthermore, different settings of STFE were selected to conduct comparative experiments. The relative results are shown in Table VII. Evidently, larger expansion rate combinations of 12, 24, and 36 paired with smaller rates of 3, 6, and 9 led to lower F1 and IoU performance. In contrast, the medium expansion rate obtains higher index scores than the small expansion rates, which limit the receptive field and restrict the expression of broader contextual information. While the expansion rates of 6, 12, and 18 achieved the best detection performance and were selected as the optimal configuration in the above-mentioned comparison experiments.

c) Effect of the MSFF module: To verify the effectiveness of the MSFF structure, we modified the backbone by removing the feature output of the initial two residual blocks in ResNet and only retaining the final output of the multilevel residuals. The specific ablation results are shown in the third row of Table VI.

When the low-level residual blocks are removed, for the CDD dataset where most of the changes are in detail (with complex edge curves of the changed objects) and the LEVIR-CD dataset where small targets account for a large proportion, the IoU of

TABLE VI
ABLATION EXPERIMENTAL RESULTS ON FOUR DATASETS (%)

Module	DTFE	STFE	MSFF	EDIdx	CDD		LEVIR-CD		SYSU-CD		WHU-CD	
					F1	IoU	F1	IoU	F1	IoU	F1	IoU
EIGER-Net	×	✓	✓	✓	95.78	91.92	90.34	82.95	79.57	67.07	91.55	84.41
	✓	×	✓	✓	95.57	91.25	90.01	82.16	80.11	68.26	91.83	84.90
	✓	✓	×	✓	88.67	80.39	86.17	76.32	76.83	60.79	87.92	78.45
	✓	✓	✓	×	95.11	90.67	90.73	83.03	78.44	64.53	91.06	83.59
	✓	✓	✓	✓	**96.74**	**93.54**	**91.14**	**83.98**	**81.12**	**69.48**	**92.86**	**86.39**

Note: "DTFE" stands for Dual-Time Feature Exchange module, "STFE" stands for Small Target Feature Enhancement module, "MSFF" stands for Multi-Scale Feature Fusion module, and "EDIdx" stands for Encoding-Decoding Index Edge Refinement module. The symbol "×" represents the removal of the corresponding module from the network. Black bold font indicates the highest value for each metric(**best**).

TABLE VII
COMPARISON RESULTS OF DILATED CONVOLUTION WITH DIFFERENT EXPANSION RATES OF STFE MODULE (%)

Module	DilConv Setting	CDD		LEVIR-CD		SYSU-CD		WHU-CD	
	dilation_rate	F1	IoU	F1	IoU	F1	IoU	F1	IoU
EIGER-Net	dr=3, 6, 9	95.97	92.46	90.31	82.83	80.29	68.78	92.03	85.86
	dr=5, 10, 15	96.32	93.11	91.02	83.81	80.91	69.19	92.57	86.15
	dr=12, 24, 36	96.03	92.94	90.44	83.18	80.50	68.97	92.24	86.03
	dr=6, 12, 18	**96.74**	**93.54**	**91.14**	**83.98**	**81.12**	**69.48**	**92.86**	**86.39**

CDD decreased by 13.15%, and the F1 decreased by 8.07% . The IoU and F1 of LEVIR-CD decreased by 7.66% and 4.97%, respectively.

The results indicate that the MSFF module has a significant advantage in handling small changes. Similarly, the performance in the SYSU-CD and WHU-CD is satisfactory. Compared to the model without MSFF, the IoU increased by 8.69% and 7.94%, and the F1 increased by 4.29% and 4.94% . In conclusion, MSFF shows a significant performance improvement on all datasets by capturing changes of different types and scales.

d) Effect of the EDIdx module: To validate the effectiveness of the EDIdx module, we conducted an ablation study with specific experimental results shown in the fourth row of Table VI. As observed, the implementation without EDIdx module decreases in F1 scores and IoU indices on the datasets. Specifically, there are decreases of 2.87%, 0.95%, 4.95%, and 2.80% in IoU indices of CDD, LEVIR-CD, SYSU-CD, and WHU-CD, respectively. In addition, we supplied visualization maps and edge detection comparison between the network model with and without EDIdx, which is shown in and Fig. 11. As can be observed, it is evident that the approach without the EDIdx module has significantly inferior edge detection results and generates more numerous false positives (green annotations) and missed detections (red annotations). Overall, the experimental results demonstrate that the EDIdx module is a crucial component of the network model and is capable of achieving fine-grained edge detection.

6) Model Efficiency Analysis: To measure the computational complexity and space complexity of the models, we further analyze the CD models regarding the criteria of floating point operations (FLOPs) and parameters (Params). We have reported the comparison of FLOPs and Params for both the proposed method and the comparative methods in Table VIII. As can be seen, FC-EF, FC-Siam-conc, and FC-Sima-diff have the lowest FLOPs and Params. The three pure CNN-based models have very small computational and spatial complexities due to the

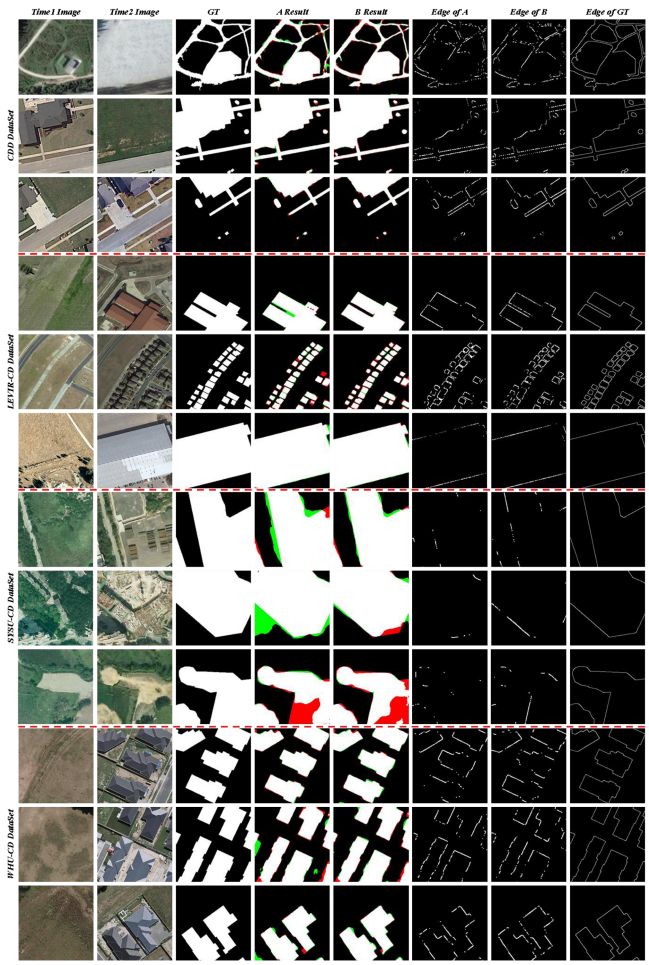

Fig. 11. Visualization and edge detection results of the ablation study of the EDIdx module on the CDD dataset. "A Result" represents the visualization result without the EDIdx module, while "B Result" represents the visualization result of EIGER-Net. "Edge of A" depicts the edge detection result without the EDIdx module, and "Edge of B" represents the EIGER-Net edge detection result.

TABLE VIII
EFFICIENCY COMPARISON WITH DIFFERENT MODELS

Method	Params(M)	FLOPs(G)
FC-EF	1.35	3.58
FC-Siam-conc	1.55	5.33
FC-Siam-diff	1.35	4.73
SNUNet	12.03	54.83
ChangeFormer*	41.03	202.79
ISNet*	34.55	21.61
EGPNet*	44.32	77.33
EGRCNN	9.63	17.64
AMTNet	24.67	21.56
ChangeViT	32.13	38.80
FEM-Net	51.33	136.90
EIGER-Net (ours)	28.31	37.65

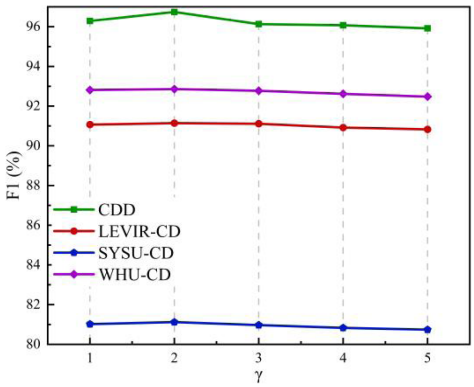

Fig. 12. F1 scores of the dataset with different γ.

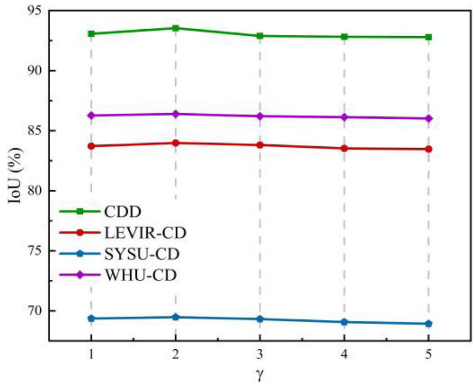

Fig. 13. IoU index of the dataset with different γ.

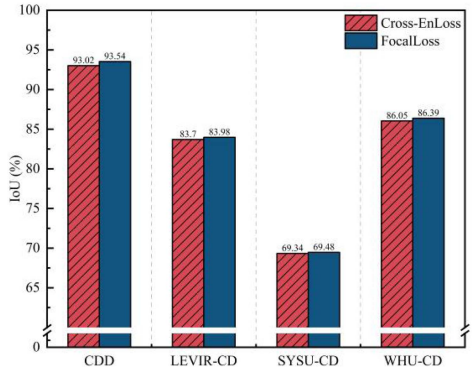

Fig. 14. Comparison of FocalLoss and cross-entropy loss on IoU index.

simple structures, while the performances on the four datasets are unsatisfactory. ChangeFormer has the highest numbers of FLOPs and Params among the transformer-based networks due to the hierarchical transformer encoder. In the comparison of the number of parameters of all CD models, our proposed EIGER-Net has achieved relatively competitive results, with 37.65G FLOPs and 28.31M Params. Compared with the AMT model, the improvement in detection accuracy is achieved by increasing the computational parameters of the model, but the overall result remains within the expected range. In summary, our method not only excels in accuracy but also maintains a commendable computational efficiency, demonstrating a well-balanced tradeoff between the two critical factors.

7) Parametric Analysis:

a) Analysis of hyperparameter of γ: In this section, we conducted an experiment to verify the impact of different γ values in the loss function on the four datasets. In specific, the values of γ are from the set of {1, 2, 3, 4, 5}, F1, and IoU, the results were recorded in Figs. 12 and 13 for comparison. According to the results, for all the datasets, the values of F1 and IoU are relatively similar with different values of γ. Specifically, the detection accuracy is the highest when $\gamma = 2$ for all the datasets, while the accuracy decreases when $\gamma = 1$, 3, 4, or 5. The main reason is that γ controls the model focusing on easy and hard samples. With the setting of $\gamma = 1$, the model primarily focuses on hard samples and slightly reduces detection accuracy. Conversely, higher values of γ, such as 3, 4, or 5, lead the model to neglect easy samples and slightly degrade performance. The setting with γ to 2 avoids bias toward either easy or hard samples and generates the best performance.

TABLE IX
PERFORMANCE COMPARISON WITH DIFFERENT LOSSES ON LEVIR-CD (%)

Coefficient setting		Performance	
FocalLoss	Cross-entropy Loss	F1	IOU
√	/	91.14	83.98
/	√	90.53	82.71
0.5	0.5	90.80	83.12
0.3	0.7	90.26	82.25

b) Loss function analysis: To explore the effectiveness of different losses, we conducted comparative experiments with the results shown in Fig. 14. As can be observed, the approach with FocalLoss on the four datasets generates better performance than the implementation with cross-entropy loss, which proves the effectiveness of FocalLoss in CD tasks. To further validate the effect of the different functions for our model, we conducted comparative experiments by assigning different weights to the focal loss and cross-entropy loss functions. The experiment was merely conducted on the LEVIR-CD dataset, and the setting of the coefficients for each loss function are shown in Table IX. As observed, the adoption of focal loss achieves more favorable detection results compared to the cross-entropy loss. In addition, the mixed loss function also delivers satisfactory outcomes. Overall, the results indicate that while various loss functions

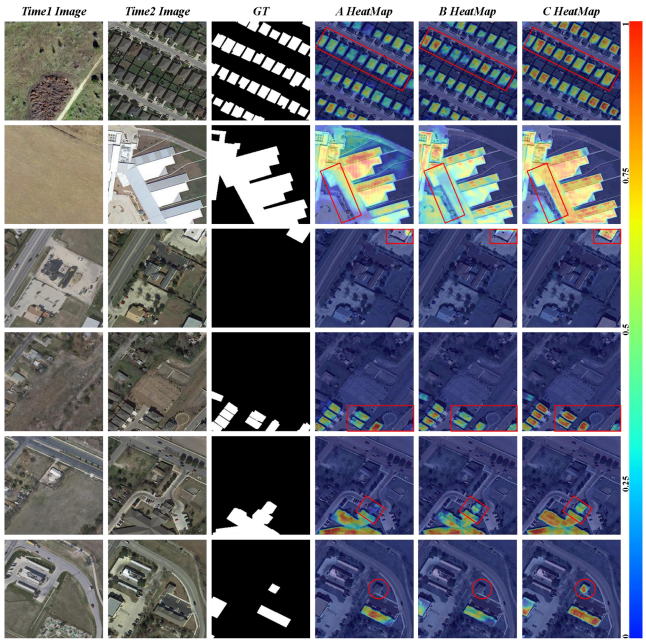

Fig. 15. Visualize heatmap inspection results. "A HeatMap" represents the heatmap detection results of the network model without using the STFE module; "B HeatMap" represents the heatmap detection results of the network model without using the EDIdx module; and "C HeatMap" represents the overall heatmap detection results of the EIGER-Net network.

produce acceptable performance, focal loss emerges as the most effective choice in the proposed model.

8) Grad-Cam Heatmap Detection Analysis: Further, to validate the effectiveness of our model in detecting changing small targets and edges, we provided the visual heatmaps of embedding features on the LEVIR-CD dataset, where highlighted areas indicate specific locations of attention, with brighter regions reflecting higher levels of attention in the heatmaps. According to the results in Fig. 15, the EIGER-Net with the EDIdx and STFE modules exhibits significantly higher attention to small targets and edges compared to the model without these modules. The highlighted regions in the heatmaps demonstrate that the STFE module focuses more on changing parts of small targets, while the network with the EDIdx module is beneficial for object edges and the overall changing regions.

V. Conclusion

In this article, we propose a feature en-decoded-index network denoted as EIGER-Net, which is the first attempt to combine the edge index information of dual-temporal images with multilevel features for the CD of RSI. Innovatively, the model integrates the edge index information extracted from the low-level features into the high-level semantic features, enhancing the representation of edge information during feature processing. In addition, to fully exploit the advanced semantic information contained in features of different scales, the model achieves feature representation and edge characterization through a multilevel index structure based on Siamese networks. Besides, the model incorporates a STFE module to enhance the expression ability of small targets and edge details through dual temporal exchange and dilated convolution operations. In comparison with the SOTA models, our proposed method demonstrates a significant enhancement of detection performance on the CDD, LEVIR-CD, SYSU-CD, and WHU-CD datasets. Despite the impressive detection capabilities demonstrated by EIGER-Net, the adaptability to multisource data and complex environments need further enhanced. In the future, we plan to develop CD models based on multisource data fusion and spatio-temporal joint analysis. By integrating multiple types of data, we aim to enhance the processing capacity for large-scale remote sensing data and enable efficient and timely detection of changes, enabling rapid responses to dynamic environmental conditions.

References

[1] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989.
[2] D. Apda and E. So, "Enhanced change detection index for disaster response, recovery assessment and monitoring of accessibility and open spaces (camp sites)," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 57, pp. 49–60, 2017.
[3] M. Chang, X. Meng, W. Sun, G. Yang, and J. Peng, "Collaborative coupled hyperspectral unmixing based subpixel change detection for analyzing coastal wetlands," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8208–8224, 2021.
[4] F. Huang, L. Chen, K. Yin, J. Huang, and L. Gui, "Object-oriented change detection and damage assessment using high-resolution remotesensing images, Tangjiao landslide, three gorges reservoir, China," *Environ. Earth Sci.*, vol. 77, no. 183, pp. 1–19, 2018.
[5] S. Tian, Y. Zhong, A. Ma, and L. Zhang, "Three-dimensional change detection in urban areas based on complementary evidence fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5608913.
[6] T. A. Abera, J. Heiskanen, E. E. Maeda, B. T. Hailu, and P. K. Pellikka, "Improved detection of abrupt change in vegetation reveals dominant fractional woody cover decline in Eastern Africa," *Remote Sens. Environ.*, vol. 271, pp. 1–13, 2022.
[7] M. İlsever et al., "Pixel-based change detection methods," in *Two-Dimensional Change Detection Methods: Remote Sensing Applications*. Berlin, Germany: Springer, 2012, pp. 7–21.
[8] A. Javed, S. Jung, W. H. Lee, and Y. Han, "Object-based building change detection by fusing pixel-level change detection results generated from morphological building index," *Remote Sens.*, vol. 12, no. 18, pp. 2952–2971, 2020.
[9] N. Quarmby and J. Cushnie, "Monitoring urban land cover changes at the urban fringe from spot HRV imagery in South-East England," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 953–963, 1989.
[10] E. H. Wilson and S. A. Sader, "Detection of forest harvest type using multiple dates of landsat TM imagery," *Remote Sens. Environ.*, vol. 80, no. 3, pp. 385–396, 2002.
[11] F. Yuan, K. E. Sawaya, B. C. Loeffelholz, and M. E. Bauer, "Land cover classification and change analysis of the twin cities (Minnesota) metropolitan area by multitemporal landsat remote sensing," *Remote Sens. Environ.*, vol. 98, no. 2, pp. 317–328, 2005.
[12] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Review articledigital change detection methods in ecosystem monitoring: A review," *Int. J. Remote Sens.*, vol. 25, pp. 1565–1596, 2004.
[13] J. S. Deng, K. Wang, Y. H. Deng, and G. J. Qi, "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data," *Int. J. Remote Sens.*, vol. 29, no. 16, pp. 4823–4838, 2008.
[14] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and K-means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, Oct. 2009.
[15] O. Miller, A. Pikaz, and A. Averbuch, "Objects based change detection in a pair of gray-level images," *Pattern Recognit.*, vol. 38, no. 11, pp. 1976–1992, 2005.
[16] A. Lefebvre, T. Corpetti, and L. Hubert-Moy, "Object-oriented approach and texture analysis for change detection in very high resolution images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2008, pp. IV–663–IV–666.

[17] O. Hall and G. J. Hay, "A multiscale object-specific approach to digital change detection," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 4, no. 4, pp. 311–327, 2003.
[18] 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.
[19] 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, pp. 3336–3350, 2021.
[20] H. Zhai, H. Zhang, P. Li, and L. Zhang, "Hyperspectral image clustering: Current achievements and future lines," *IEEE Geosci. Remote Sens.*, vol. 9, no. 4, pp. 35–67, Dec. 2021.
[21] M. Zhang, Z. Liu, J. Feng, L. Liu, and L. Jiao, "Remote sensing image change detection based on deep multi-scale multi-attention siamese transformer network," *Remote Sens.*, vol. 15, pp. 842–866, 2023.
[22] R. C. Daudt, B. L. Saux, and A. Boulch, "Fully convolutional siamese networks for change detection," in *Proc. IEEE Int. Conf. Image Process.*, 2018, pp. 4063–4067.
[23] F. Rahman, B. Vasu, J. V. Cor, J. Kerekes, and A. Savakis, "Siamese network with multi-level features for patch-based change detection in satellite imagery," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2018, pp. 958–962.
[24] 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.
[25] Y. Zhu, K. Lv, Y. Yu, and W. Xu, "Edge-guided parallel network for VHR remote sensing image change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 7791–7803, 2023.
[26] T. Lei, Y. Zhai, Y. Xu, Y. Wang, and M. Gong, "Edge guided and dynamically deformable transformer network for remote sensing images change detection," *Acta Electronica Sinica*, vol. 52, no. 1, pp. 108–117, 2024.
[27] Y. Cao, H. Xie, L. Yang, W. Zhang, and P. Gong, "Building change detection based on multi-feature fusion and edge constraints," *Surveying Mapping Spatial Geographic Inf.*, vol. 45, no. 3, pp. 8–15, 2022.
[28] Z. Song, X. Li, R. Zhu, Z. Wang, Y. Yang, and X. Zhang, "ERMF: Edge refinement multi-feature for change detection in bitemporal remote sensing images," *Signal Process., Image Commun.*, vol. 116, 2023, Art. no. 116964.
[29] M. Lebedev, Y. V. Vizilter, O. Vygolov, V. A. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *Int. Arch. Photogrammetry, Remote Sens., Spatial Inf. Sci.*, vol. 42, pp. 565–571, 2018.
[30] 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, pp. 1662–1684, 2020.
[31] Q. Shi, M. Liu, S. Li, X. Liu, F. Wang, and L. Zhang, "A deeply supervised attentionmetric-based network and an open aerial image dataset for remote sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5604816.
[32] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery dataset," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 574–586, Jan. 2019.
[33] R. Zhang, H. Zhang, X. Ning, X. Huang, J. Wang, and W. Cui, "Global-aware Siamese network for change detection on remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 199, pp. 61–72, 2023.
[34] M. Ehlers et al., "Cest analysis: Automated change detection from very-high-resolution remote sensing images," *Int. Arch. Photogrammetry Remote Sens.*, vol. XXXIX-B7, pp. 317–322, 2012.
[35] M. Carlo, B. Francesca, and B. Lorenzo, "Building change detection in multitemporal very high resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2664–2682, May 2015.
[36] 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.
[37] J. J. Gapper, H. El-Askary, E. Linstead, and T. Piechota, "Coral reef change detection in remote pacific islands using support vector machine classifiers," *Remote Sens.*, vol. 11, no. 13, pp. 1525–1552, 2019.
[38] X. Liu and R. G. Lathrop Jr., "Urban change detection based on an artificial neural network," *Int. J. Remote Sens.*, vol. 23, no. 12, pp. 2513–2518, 2002.
[39] D. Quispe and J. Sulla-Torres, "Automatic building change detection on aerial images using convolutional neural networks and handcrafted features," *Int. J. Adv. Comput. Sci.*, vol. 11, no. 6, pp. 679–684, 2020.
[40] X. Huang, Y. Xie, and J. Wei, "Automatic recognition of desertification information based on the pattern of change detection-CART decision tree," *J. Catastrophol.*, vol. 32, pp. 36–42, 2017.
[41] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.
[42] T. Chen, Z. Lu, Y. Yang, Y. Zhang, B. Du, and A. Plaza, "A Siamese network based U-net for change detection in high resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2357–2369, 2022.
[43] C. Chen, J. W. Hsieh, P. Chen, Y. K. Hsieh, and B. Wang, "SARAS-Net: Scale and relation aware siamese network for change detection," in *Proc. 37th AAAI Conf. Artif. Intell. 35th Conf. Innov. Appl. Artif. Intell. 13th Symp. Educ. Adv. Artif. Intell.*, 2022, pp. 14187–14195.
[44] 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. Photogrammetry Remote Sens.*, vol. 202, pp. 599–609, 2023.
[45] 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.
[46] Y. Xing, Y. Jia, S. Gao, J. Hu, and R. Huang, "Frequency-enhanced mamba for remote sensing change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 22, 2025, Art. no. 2501605.
[47] J. Miao, S. Li, X. Bai, W. Gan, J. Wu, and X. Li, "RS-NormGAN: Enhancing change detection of multi-temporal optical remote sensing images through effective radiometric normalization," *ISPRS J. Photogrammetry Remote Sens.*, vol. 221, pp. 324–346, 2025.
[48] D. Zhu, X. Huang, H. Huang, Z. Shao, and Q. Cheng, "ChangeViT: Unleashing plain vision transformers for change detection," 2024, *arXiv:2406.12847v1*.
[49] 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.
[50] G. Cheng, G. Wang, and J. Han, "ISNet: Towards improving separability for remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5623811.
[51] 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.
[52] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *Mach. Learn.*, vol. 37, pp. 448–456, 2015.
[53] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
[54] S. Fang, K. Li, and Z. Li, "Changer: Feature interaction is what you need for change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5610111.
[55] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
[56] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1800–1807.
[57] H. Lu, Y. Dai, C. Shen, and S. Xu, "Indices matter: Learning to index for deep image matting," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 3265–3274.
[58] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2999–3007.

Chunyan Yu (Senior Member, IEEE) received the Ph.D. degree in environmental engineering from Dalian Maritime University, Dalian, China, in 2012.

She is currently an Associate Professor with the Information Science and Technology College, Dalian Maritime University. Her research interests include image segmentation, hyperspectral image classification, and pattern recognition.

Chi Yu received the bachelor's degree from Weifang University, Weifang, China, in 2023. He is currently working toward the master's degree in artificial intelligence with Dalian Maritime University, Dalian, China.

His research interests include hyperspectral image processing and deep learning.

Feihong Zhou received the bachelor's degree in software engineering from Anhui University of Science and Technology, Huainan, China, in 2023. He is currently working toward the master's degree in computer science and technology with Dalian Maritime University, Dalian, China.

His research interests include hyperspectral image processing and deep learning.

Yulei Wang (Member, IEEE) received the B.S. and Ph.D. degrees in signal and information processing from Harbin Engineering University, Harbin, China, in 2009 and 2015, respectively.

She is awarded by the China Scholarship Council in 2011 as a joint Ph.D. Student to study in Remote Sensing Signal and Image Processing Laboratory, University of Maryland, Baltimore, MD, USA, for two years. She is an Associate Professor with the Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian, China. Her research interests include hyperspectral image processing and vital signs signal processing.

Qiang Zhang (Member, IEEE) received the B.E. degree in surveying and mapping engineering and the M.E. and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2017, 2019, and 2022, respectively.

He is currently an Xinghai Associate Professor with the Center of Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian, China. He has authored more than twenty journal articles in IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, *Earth System Science Data*, and *ISPRS Journal of Photogrammetry*, and *Remote Sensing*. His research interests include remote sensing information processing, computer vision, and machine learning.