

Article

EIMGDNet: An Edge-Induced and Multi-Dimensional Grouped Difference Network for Remote Sensing Image Change Detection

Le Sun ^{1,2} , Mingxuan Ding ², Qiaolin Ye ³, Yuhui Zheng ⁴ , Zebin Wu ⁵ and Wen Lu ^{4,*}

¹ School of Computer Science and the Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), Nanjing University of Information Science and Technology, Nanjing 210044, China; sunlecncom@nuist.edu.cn

² School of Computer Science, Nanjing University of Information Science and Technology, Nanjing 210044, China; cs_dingmx@nuist.edu.cn

³ School of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China; yqlcom@njfu.edu.cn

⁴ College of Computer, Qinghai Normal University, Xining 810016, China; zheng_yuhui@nuist.edu.cn

⁵ School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; wuzb@njust.edu.cn

* Correspondence: luwen@qhnu.edu.cn

Highlights

What are the main findings?

- We propose EIMDGNet, an end-to-end framework for remote sensing change detection that integrates edge-induced enhancement with multi-dimensional grouping. It incorporates a novel grouped interaction mechanism to capture subtle spatial variations, alongside a lightweight attention module for generating discriminative edge-saliency maps.
- The framework employs a hierarchical guidance and progressive optimization strategy to achieve robust multi-scale feature integration, which effectively enhances feature discrimination in change detection tasks.

What are the implications of the main findings?

- EIMDGNet significantly improves detection accuracy and edge detail preservation in complex remote sensing scenes, offering a more reliable solution for change detection. Its consistently superior performance across multiple datasets demonstrates strong generalization capability, supporting practical applications such as land cover monitoring, urban development analysis, and environmental change assessment.
- The model integrates edge-induced enhancement with multi-dimensional grouping for finer feature representation and improved handling of challenging, high-variation scenarios. Furthermore, its hierarchical and progressive optimization strategy provides a scalable, adaptable approach for multi-scale feature fusion, which can be extended to other vision tasks requiring fine-grained spatial reasoning.



Academic Editor: Zhenwei Shi

Received: 28 January 2026

Revised: 15 February 2026

Accepted: 17 February 2026

Published: 20 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Abstract

Change detection in remote sensing imagery is crucial for monitoring temporal variations in surface characteristics; nevertheless, it presents significant challenges owing to indistinct boundaries, limited semantic differentiation, and inadequate incorporation of multi-scale contextual information. To solve these problems, we propose EIMDGNet (Edge-Induced and Multi-Dimensional Grouped Difference Network), a novel architecture that enhances boundary representation and cross-scale feature interaction for accurate and robust change detection. EIMDGNet adopts a dual-branch ResNet18 backbone to extract multi-scale features from bi-temporal images, capturing both fine spatial detail and high-level semantic

context. To improve boundary awareness and reduce pseudo-change interference, we introduce the Edge-Induced Differential Multi-Dimensional Group Enhancement Module (EID-MDGEM). This module enriches fine-grained spatial features through grouped pooling across spatial and channel dimensions, enabling precise localization of change contours. Within EID-MDGEM, the Edge Feature Enhancement Module (EFEM) integrates a parameter-free attention mechanism to generate edge-saliency maps, highlighting true change regions while suppressing background noise and irrelevant variations. To further enhance semantic consistency across feature scales, we design the Multi-Scale Hierarchical Progressive Fusion Module (MSHPM). This component employs a bottom-up progressive strategy to hierarchically integrate low-level spatial details with high-level semantic abstractions, thus increasing the continuity and completeness of detected change regions. By tightly coupling edge-aware enhancement with multi-scale hierarchical fusion, EIMDGNet effectively addresses major obstacles in change detection, including boundary ambiguity, inconsistent scale information, and feature misalignment. We evaluated EIMDGNet on five remote sensing change detection datasets: LEVIR-CD, DSIFN-CD, S2Looking, CLCD-CD and GVLM-CD. Our method consistently outperformed state-of-the-art approaches, achieving 91.49% F1 and 82.93% IoU on LEVIR-CD, 77.32% F1 and 69.39% IoU on DSIFN-CD, the highest 49.19% IoU and 99.20% OA on S2Looking, 81.65% F1 and 72.91% IoU on CLCD-CD, and 85.49% F1 and 76.08% IoU on GVLM-CD. These results demonstrate the superior accuracy and robustness of EIMDGNet across diverse change detection scenarios.

Keywords: remote sensing image; change detection; edge induction; grouped difference network; multi-scale feature fusion

1. Introduction

Remote sensing image change detection serves as a core technology in remote sensing and earth observation. By acquiring image data at different time points and performing a comparative analysis, this technique enables the identification of surface or object changes over time, including both semantic-level changes and variations at the target or pixel level. Emerging paradigms in foundational vision tasks [1–3] provide a necessary algorithmic groundwork for change detection. Change detection not only supports accurate detection and quantification of land cover transformations but also proves essential in various applications including urban development [4], natural disaster assessment [5], and farmland protection [6].

Traditional remote sensing image change detection techniques typically fall into pixel-level and object-level classifications. Pixel-level methods examine remote sensing images at a per-pixel basis over time to detect changes. Representative techniques include change vector analysis (CVA) [7] and slow feature analysis (SFA) [8]. In contrast, object-level methods analyze both the spatial and spectral information of the images to detect changes in specific objects like buildings, roads, or vegetation. Examples include the use of Kullback–Leibler (KL) divergence [9] and image correlation analysis [10], which emphasize low-level semantic objects and extract changes from bi-temporal or multi-temporal imagery. However, traditional methods struggle to effectively handle complex multi-source and multi-phase data. Furthermore, their limited capability to process multi-scale features and adapt to irregular or complex object shapes constrains their performance in high resolution remote sensing scenarios [11].

Deep learning models [12–14] have gained widespread attention in remote sensing image change detection (CD) due to their ability to capture complex, nonlinear relationships. These models not only enable end-to-end change detection, but also show advantages

over traditional approaches and machine learning (ML)-based methods, particularly in the processing of high-resolution remote sensing imagery. Numerous deep learning architectures have been deployed for CD, including both Convolutional Neural Network (CNN) [15] and Generative Adversarial Networks (GANs) [16], as well as Recurrent Neural Network (RNN) [17]. Among them, the CNN, as a core technology in deep learning, demonstrates a strong ability to automatically extract image features, support end-to-end learning, and accurately capture multi-level spatial information from images [18]. Hedjam et al. [19] introduced an unsupervised Siamese CNN framework. This framework uses contrastive learning to identify change patterns in unlabeled optical remote sensing data. Meanwhile, Peng et al. [20] developed SemiCDNet. This semi-supervised model trains on a few labeled data along with unmarked mass data. As a result, it effectively enhances the efficiency and accuracy of change detection. SemiCDNet is suitable for the application scenarios of ground object change monitoring in remote sensing images. Wang et al. [21] presented a mixed change detection network which integrates cross-axis attention multi-layer perceptrons (MLPs) with CNNs. Specifically, by fusing the global information-processing strength of MLPs with the local feature extraction power of CNNs, the architecture gains enhanced resilience to complex changes; moreover, the embedded cross-axis attention mechanism further refines feature extraction and, consequently, boosts detection accuracy. In addition, Ye et al. [22] presented a 3D-CNN-based approach tailored for remote sensing image change detection. Firstly, features of different levels were extracted respectively, and then features of adjacent levels were fused to improve detection accuracy through cross-fusion features of adjacent levels. Finally, 3D CNN was used for change detection. A two-branch fusion architecture proposed by Song et al. [23] embeds axial cross-attention within CNNs, allowing comprehensive multi-directional and multi-scale features to be harvested. The two-branch design further processes spatial and channel dimensions separately, enhancing information fusion efficiency. As application scenarios diversify and task complexity escalates, the deep learning change detection landscape has evolved beyond conventional CNNs, GANs, RNNs, and Vision Transformers to include recent advances such as the Mamba state-space architecture and diverse hybrid frameworks.

In recent years, Transformer [24–26], originally designed for natural-language tasks, has drawn considerable attention across computer vision. Compared to traditional pure CNN models, Transformer offers significant advantages in modeling long-term dependencies [27,28], better capturing global contextual information and handling remote correlation features. Recently, Transformer-based algorithms have been increasingly explored for remote sensing image change detection. TransUNetCD, a hybrid framework combining Transformer and U-Net architectures, was introduced by Li et al. [29]. The Transformer encoder in this model enhances feature extraction. Meanwhile, the U-Net decoder enables fine-grained, pixel-level change detection. A novel approach employing a channel attention mechanism (CBAM) and multi-scale converter fusion was developed by Wang et al. [30]. This approach strengthens the representation of fine details. It also improves detection accuracy by effectively integrating information across scales. More recently, a multi-scale feature extraction strategy that couples the local-detail sensitivity of CNNs with the long-range context awareness of Transformers was introduced by Li et al. [31]. This strategy effectively captures both fine-grained changes and long-range dependencies in remote sensing images, achieving more comprehensive change detection results. GCFormer was introduced by Yu et al. [32]; this Transformer-based architecture embeds a global context module and a global feature aggregation mechanism to capture long range dependencies in multi-temporal remote sensing imagery. The multi-scale design of this strategy further enhances semantic representation. It also improves the detection of complex and subtle change regions.

While Transformer-based models have significantly advanced the field of change detection, their efficacy in capturing fine-grained spatial details remains limited. The recent emergence of Mamba-based architectures, such as CDMamba [33], offers promising alternatives with efficient long-range modeling capabilities. Despite these advances, several fundamental challenges persist across CNN-, Transformer-, and Mamba-based approaches. A primary limitation is the persistent blurring of change boundaries and inconsistent integration of multi-scale features, which collectively undermine detection performance. This often stems from a common tendency to prioritize high-level semantic representations at the expense of fine-grained spatial cues, resulting in insufficient sensitivity to subtle, small-scale changes. Furthermore, an overemphasis on global contextual modeling frequently impedes the effective extraction and integration of low-level spatial details—elements that are critical for precise boundary localization and nuanced change characterization. As underscored by recent studies, the complementary fusion of high-level semantics with low-level features is essential for achieving substantial improvements in change detection accuracy [34].

To tackle these challenges, we present EIMDGNet, an Edge-Induced and Multi-Dimensional Grouping Collaborative Enhanced Difference Network designed for remote sensing image change detection. Our contributions are designed to directly address the aforementioned gaps: an edge-induced enhancement strategy is proposed to sharpen change boundaries and a progressive multi-scale fusion mechanism is introduced to ensure feature consistency. The architecture employs a dual-branch backbone formed by two pre-trained ResNet18 models that jointly extract multi-scale features from the bi-temporal image pair. To further strengthen edge-related feature representation, we design the Edge-Induced Differential Multi-Dimensional Group Enhancement Module (EID-MDGEM). Boundary information in changed areas is first refined by an Edge Feature Enhancement Module (EFEM), sharpening the model's response to subtle transitions; subsequently, global average and max pooling are applied along spatial dimensions to the enhanced edge maps, further amplifying temporal disparities at matching locations. This dual pooling strategy enhances feature expressiveness from multiple perspectives: global average pooling captures stable overall characteristics, while max pooling accentuates local extremes indicative of significant changes. The complementary fusion of these operations strengthens the feature difference representation and provides valuable information for subsequent fusion. Furthermore, we incorporate a Multi-Scale Hierarchical Progressive Fusion Module (MSHPM) that systematically orchestrates the exchange of information across adjacent stages, guiding the upper layers to absorb fine-grained cues from lower levels while injecting high-order semantics downward; this bidirectional refinement yields a richer, more balanced descriptor in which detailed edge evidence and abstract change semantics are jointly emphasized. To leverage the full spectrum of fused representations, a summary-based prediction head is finally employed, aggregating feature maps of every scale before delivering the ultimate bi-temporal change map. The principal contributions of this study are detailed hereafter.

(1) We develop an Edge-Induced and Multi-Dimensional Grouping Collaborative Enhanced Difference Network (EIMDGNet) based on the Siamese architecture. The encoder focuses on extracting edge features from bi-temporal images to highlight change regions by enhancing feature disparities, while the decoder employs an inter-layer progressive fusion strategy to achieve complementary advantages of multi-scale features.

(2) We innovatively design two core modules: Edge-Induced Differential Multi-Dimensional Group Enhancement Module (EID-MDGEM) and Edge Feature Enhancement Module (EFEM). EID-MDGEM captures subtle spatial variations through grouped channel interactions and spatial-channel pooling, enhancing local detail representation. EFEM (Edge-Focused Enhancement Module) employs a lightweight attention mechanism to ex-

tract edge-salient feature maps, effectively highlighting discriminative edge regions while suppressing pseudo-change noise.

(3) To counteract granularity mismatch across layers, the Multi-Scale Hierarchical Progressive Fusion Module (MSHPM) is introduced; following a bottom-up route, it successively blends features so that fine spatial details ascend the hierarchy step-by-step, thereby drawing fully upon the complementary merits of multi-scale representation.

(4) Comprehensive evaluations conducted on the publicly available LEVIR-CD, DSIFN-CD, and S2Looking benchmarks reveal that the proposed approach secures top-ranking performance across the majority of assessment indicators, thereby substantiating its efficacy in addressing the change detection (CD) task.

The manuscript is organized as follows: A comprehensive survey of deep learning change detection approaches, attention techniques, and multi-scale fusion strategies is provided in Section 2. Section 3 progressively unpacks the architecture of the proposed network. Experimental analyses are detailed in Section 4, followed by a discussion of the method's strengths and limitations in Section 5. Concluding remarks are presented in Section 6.

2. Related Work

This section provides a comprehensive review of deep learning approaches for CD, attention mechanisms used in CD, and strategies for multi-scale feature fusion applied in CD tasks.

2.1. Deep Learning Approaches for CD

Deep learning has fundamentally transformed the field of change detection (CD) in remote sensing by facilitating the comprehensive learning of intricate spatial and temporal patterns through an end-to-end approach. Compared to traditional edge detection methods such as Canny and Sobel operators, which rely on the gradient distribution of local pixel intensities, DL-based approaches can learn more abstract and complex edge features, which play a critical role in capturing contours and texture characteristics in remote sensing images for CD. Early CNN-based approaches extracted edge-related local features through multi-layer convolutions to characterize structural boundaries and semantic contours within images. Daudt et al. [35] presented a Fully Convolutional Siamese network employing weight-shared encoders to enhance edge consistency across temporal images, eliminating the reliance on handcrafted edge features while achieving precise pixel-level localization of change boundaries. Building upon this Siamese architecture, Liu et al. [36] further advanced the paradigm by introducing a dual-task constrained network model. This model integrates building extraction with change detection within a unified framework, leveraging the structural prior from building masks to refine change boundaries and suppress false alarms. Meanwhile, N et al. [37] introduced a robust CNN framework capable of resisting illumination and noise interference for reliable edge feature extraction. Although these CNN-based methods significantly strengthen local edge feature extraction and representation, they continue to encounter difficulties in effectively modeling long-range edge dependencies and holistic contour structures.

To retain finer edge details and spatial contours, encoder–decoder architectures such as U-Net [38] and FCN [39] have been widely adopted, enabling edge-preserving pixel-level CD through skip connections and multi-scale feature fusion. Panda et al. [40] further employed ResNet to mitigate vanishing gradients and support deeper hierarchical edge representations. However, CNNs are constrained by the intrinsic locality of their convolution operations, which inherently limits their ability to capture long-range edge dependencies and global contour coherence in remote sensing imagery. To overcome this limitation,

Transformer architectures were introduced to enhance edge modeling in remote sensing CD. Chen et al. [41] exploited the self-attention mechanism in a Transformer-based CD framework to capture global contextual relationships, while Bandara et al. [42] proposed ChangeFormer, which models bi-temporal dependencies through hierarchical self-attention. Chen et al. [43] employed a hierarchical transformer architecture integrated with an edge-aware mechanism to concurrently enhance feature representation for both intra-region semantic changes and fine-grained edge details, and Ding et al. [44] integrated spatial and temporal cues for enhanced semantic edge understanding in CD. Although Transformers significantly improve global perception, they often blur object boundaries and subtle edge variations. To address this, we propose EIMDGNNet, which introduces a novel Edge-Induced Differential Multi-Dimensional Group Enhancement Module (EID-MDGEM) built upon a pre-trained ResNet backbone. This module systematically enhances edge-aware feature extraction through differential grouping and multi-dimensional aggregation, thereby strengthening structural edge coherence across scales and achieving more accurate and consistent change detection.

2.2. Attention Mechanisms in CD

Attention mechanisms let models concentrate on informative regions and suppress irrelevant background interference, greatly enhancing feature discriminability in CD. Jiang et al. [45] introduced Visual Change Transformer (VCT), which employs self-attention to jointly capture local and global contextual dependencies, improving the detection of multi-scale and complex changes. Yin et al. [46] combined CNNs, Transformers, and CBAM (Channel-Spatial Attention Module) to adaptively re-weight channel and spatial responses, effectively integrating CNN-based local detail extraction with Transformer-based contextual awareness. Lv et al. [47] additionally introduced a spatial-spectral attention mechanism informed by change-magnitude data, which adaptively highlights critical spatial and spectral features to enhance the accuracy of change localization.

Beyond spatial-channel attention, several studies have explored temporal and relational dependencies. Li et al. [48] designed STADE-CDNet, which couples spatio-temporal attention with differential enhancement to highlight subtle evolving regions. Wang et al. [49] introduced a position-aware Graph Attention Network (GAT) that integrates geospatial information to model inter-pixel relationships in multi-category scenes. Ji et al. [50] developed a domain-adaptation framework with differential attention to align source-target features, thereby improving cross-domain generalization. Collectively, these efforts demonstrate that attention mechanisms significantly enhance the robustness and adaptability of CD models under complex conditions. Nevertheless, most attention-based networks prioritize semantic saliency while neglecting fine boundary structures, often leading to discontinuous or blurred contours. High-resolution remote sensing imagery demands accurate delineation of change edges, which remains a persistent challenge. In response, we construct the Edge Feature Enhancement Module (EFEM) to strengthen the boundary-aware representational capacity of the model. EFEM utilizes a lightweight, parameter-free attention mechanism to generate edge-saliency maps that emphasize boundary-relevant features while suppressing noise. These refined edge cues improve structural completeness and yield more precise, continuous representation of change regions.

2.3. Multi-Scale Feature Fusion Strategies in CD

Change patterns in remote sensing imagery exhibit substantial variability in scale—ranging from extensive structural transformations to subtle local alterations—making multi-scale fusion essential for robust CD. Feng et al. [51] proposed mechanisms for intra-scale interaction and inter-scale fusion to enable the exchange of features across varying

resolutions, thereby enhancing both detail expression and overall perception. Luo et al. [52] developed a multi-scale differential-fusion framework for hyperspectral CD, which amplifies contrasts between changed and unchanged regions to highlight subtle variations. These approaches underline the effectiveness of integrating complementary features across multiple scales. Recent advances emphasize hybrid fusion of local and global information. Jiang et al. [53] presented MSFCTNet, which combines CNNs and Transformers to couple local feature extraction with long-range dependency modeling. Huang et al. [54] adopted a hierarchical strategy to merge low- and high-level features while maintaining spatio-temporal consistency. Wang et al. [55] proposed a lightweight super-resolution network that leverages a multi-scale graph attention mechanism to effectively model non-local spatial dependencies and refine structural details for enhanced remote sensing image reconstruction. Although these methods effectively enhance multi-scale complementarity, most rely on straightforward concatenation or upsampling, which may cause semantic inconsistency and feature misalignment. To overcome these limitations, we propose the Multi-Scale Hierarchical Progressive Fusion Module (MSHPM). MSHPM extracts multi-level features through a multi-scale convolutional structure and employs attention to dynamically calibrate key spatial-channel cues during fusion. Through progressive bottom-up integration, it unifies fine spatial details with high-level semantic information, yielding coherent multi-scale representations with continuous boundaries and semantically consistent change maps.

3. Methodology

In this section, we first provide an overview of the network's overall framework and offer a detailed description of its architecture. Next, we provide an in-depth explanation of the key components of the model, including EID-MDGEM and MSHPM. Finally, we introduce the loss function that combines BCE and Dice.

3.1. Architecture Overview

The architectural overview of the proposed EIMDGNet is illustrated in Figure 1. Designed to augment spatio-temporal disparities, the framework incorporates two parallel ResNet18 encoders pre-trained on large-scale vision tasks. These encoders operate in a Siamese fashion, generating multi-scale feature representations across four progressively downsampled stages with resolutions of 64×64 , 32×32 , 16×16 , and 8×8 . These extracted features are subsequently fed into EID-MDGEM, where boundary information of change regions is precisely captured by the embedded EFEM. The EFEM combines the parameter-free attention mechanism SimAM with convolutional operations, effectively enhancing edge feature perception while reducing computational overhead during training. Thereafter, the differential features are partitioned into G groups, and for the edge features of each layer, both global average pooling and global max pooling are performed across the spatial dimensions (height and width). Following this, a bottom-up Multi-Scale Hierarchical Progressive Fusion Module (MSHPM) is applied to achieve hierarchical multi-scale feature integration. Ultimately, all multi-level outputs are resampled to a uniform resolution and aggregated to produce the final change detection map.

3.2. Edge-Induced Differential Multi-Dimensional Group Enhancement Module (EID-MDGEM)

In CD work, various methods have been established to generate difference maps from bi-temporal images. These approaches encompass direct image subtraction, feature concatenation, and a variety of enhancement techniques. However, these traditional methods often neglect spatial discrepancies in edge details or fail to effectively model inter-channel feature correlations. To overcome these limitations, this paper proposes the Edge-Induced Differential Multi-Dimensional Group Enhancement Module (EID-MDGEM), as illustrated

in Figure 2. The proposed module performs multi-dimensional grouping and spatial-channel pooling on edge-induced differential features, thereby simultaneously capturing fine-grained spatial detail variations and reinforcing cross-channel feature dependencies. This design enables more precise representation of subtle change boundaries and more discriminative feature learning for complex change detection scenarios.

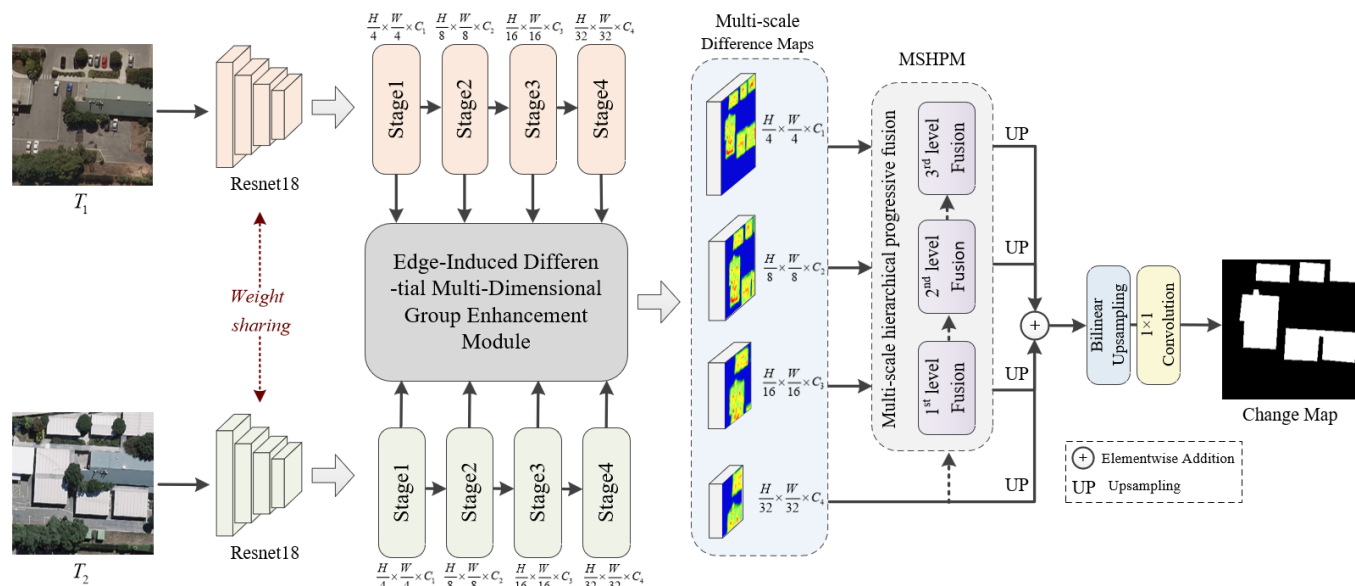


Figure 1. The architecture of our proposed EIMDGNNet.

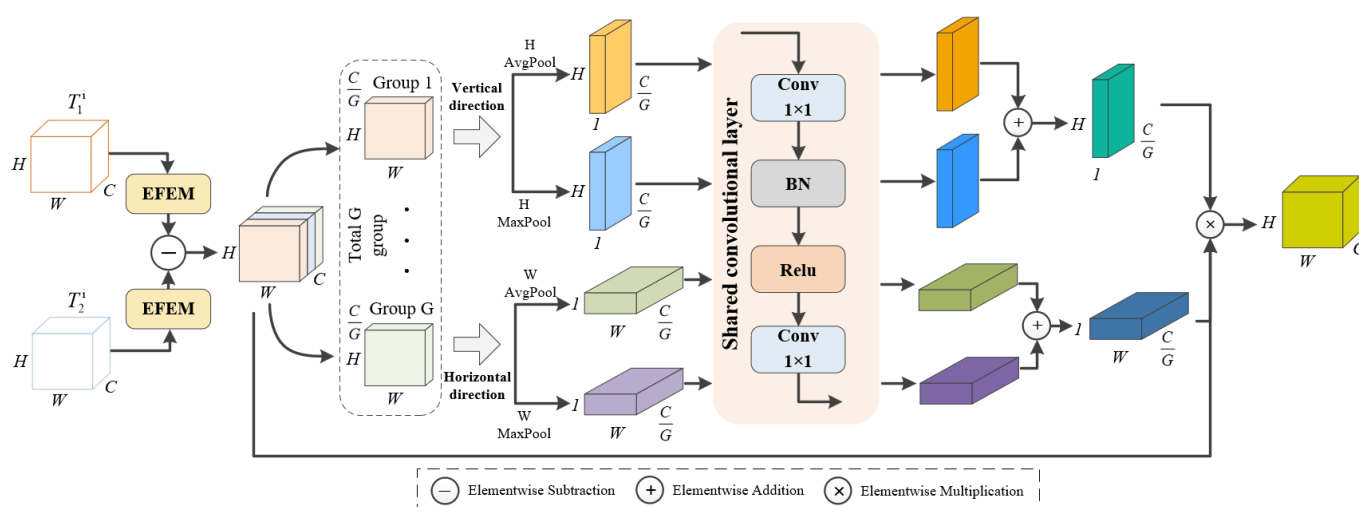


Figure 2. The Edge-Induced Differential Multi-Dimensional Group Enhancement Module (EID-MDGM).

Specifically, we integrate edge information into the feature difference computation to more effectively delineate the boundaries and shapes of change regions. To this end, an Edge Feature Enhancement module (EFEM) based on boundary information is designed to accurately identify areas of change. This module preserves fine-grained boundary details while enhancing the representation of edge-aware features; therefore, enhancing the model's sensitivity to subtle variations is a key objective. Figure 3 illustrates the architectural design of the EFEM.

In our method, features (T_k^i ($k = 1, 2$)) from different stages of the bi-temporal images were processed using the EFEM to extract and enhance edge information from both tem-

poral domains. Initially, the input features of the first layer were preprocessed using the SimAM attention mechanism [56] to identify salient regions.

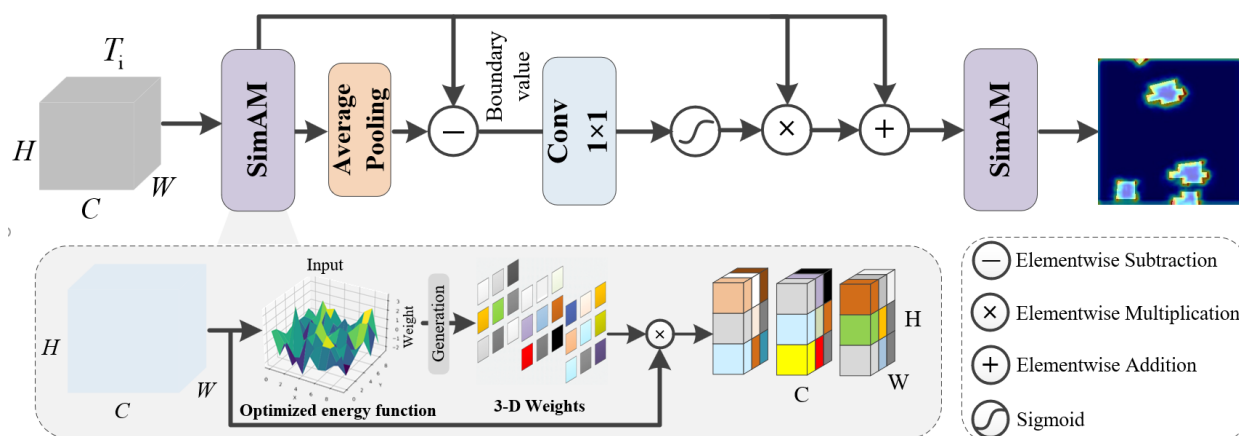


Figure 3. The Edge Feature Enhancement Module (EFEM).

SimAM is a parameter-free three-dimensional attention mechanism inspired by neuroscience theories, which determines the significance of each neuron through the minimization of an energy function. To strengthen attention, the mechanism evaluates the contribution of each neuron to the overall activation, where neurons carrying rich information typically exhibit activation patterns distinct from their neighbors and inhibit them. Consequently, the significance of a neuron can be characterized by its linear separability from surrounding neurons, emphasizing those with spatially inhibitory responses. The minimum energy function associated with the SimAM mechanism is formally defined as follows:

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

In the above expression, $\hat{\mu}$ and σ^2 denote the mean and variance calculated over all neurons except for neuron t . A lower energy value signifies a greater deviation of the neuron from its adjacent neurons, which consequently suggests an elevated degree of significance. Once the minimum energy value is obtained through Equation (1), the importance of the neuron is determined by $1/E$. After evaluating the importance of the neurons, the feature matrix is improved:

$$\hat{X} = \text{sigmoid}\left(\frac{1}{E}\right) \times X \quad (2)$$

where E aggregates all σ_t^2 across the channel and spatial dimensions and X is defined in Equation (6).

Next, a series of steps is performed to enhance the edge information in the image. This process begins with pooling, subtraction, and convolution operations, as edge regions are typically characterized by significant gradient variations. The application of average pooling serves to smooth the feature representations. Concurrently, subtracting the pooled features from the original features accentuates the edge regions by effectively emphasizing local intensity differences. Then, combining 1×1 convolution with the Sigmoid activation function serves to further amplify the contrast and prominence of the edge features. After that, multiplicative and additive operations are used to further reinforce the edge information. Finally, the SimAM attention mechanism is re-applied to adaptively

emphasize the most informative regions, thereby generating the final edge features $\hat{F}_e^t, \hat{f}_e^t$. The corresponding mathematical expressions are defined as follows:

$$\begin{aligned}\hat{F} &= \text{SimAM}(T_i^1) \\ \hat{f} &= \text{SimAM}(T_i^2)\end{aligned}\quad (3)$$

$$\begin{aligned}F_e^t &= \text{Conv}_{1 \times 1}(\hat{F} - \text{AP}(\hat{F})) \times \hat{F} + \hat{F} \\ f_e^t &= \text{Conv}_{1 \times 1}(\hat{f} - \text{AP}(\hat{f})) \times \hat{f} + \hat{f}\end{aligned}\quad (4)$$

$$\begin{aligned}\hat{F}_e^t &= \text{SimAM}(F_e^t) \\ \hat{f}_e^t &= \text{SimAM}(f_e^t)\end{aligned}\quad (5)$$

where SimAM denotes the attention mechanism, AP stands for a 3×3 average pooling layer, and $\text{Conv}_{1 \times 1}$ refers to a convolutional block consisting of a 1×1 convolution with batch normalization coupled with a Sigmoid activation function.

Traditional neural networks often have difficulty capturing comprehensive information spanning the spatial dimensions of an image, which are height and width. This limitation affects their ability to express features effectively. To solve this problem, a channel grouping strategy is employed, whereby the feature map is equally partitioned into G subsets along the channel dimension, with each group containing C/G channels. Here, G is the number of groups, typically chosen as a small integer that evenly divides the total number of channels (in this study, $G = 4$). Consequently, the grouped feature map can be expressed in the following manner:

$$X \in R^{B \times G \times \frac{C}{G} \times H \times W} \quad (6)$$

The channel grouping strategy effectively partitions parameter interactions, thereby reducing computational overhead by limiting the complexity within each independent subset. It also maintains the diversity and richness of the feature representations. Furthermore, in the context of processing high-resolution images and large-scale data, computational complexity and memory overhead become critical considerations. The grouping strategy effectively optimizes these aspects.

Next, global average pooling and global max pooling are applied in parallel to the feature maps of each group, aggregating spatial information separately along the height and width dimensions. Due to the inherent locality and static nature of traditional convolution operations, they are incapable of dynamically calibrating feature importance across varying channels and spatial positions, thereby hindering the effective emphasis on critical information. To address this, we introduce shared convolutional layers combined with an attention mechanism to generate direction-aware attention maps along the vertical and horizontal spatial dimensions. The attention maps are employed to adaptively re-weight the input feature maps, thereby enhancing the representation of salient features while concurrently suppressing redundant or irrelevant information. The shared convolutional layers include two 1×1 convolutional layers, followed by batch normalization and a ReLU activation function, which together refine the channel relationships and adjust dimensionality. Subsequently, the resultant activations from these convolutional layers are aggregated through summation. These intermediate features are normalized to the range $[0, 1]$ via a Sigmoid activation to produce attention weights. The original input feature maps are then

recalibrated via element-wise multiplication with these weights, yielding the final output representation. The result can be expressed as

$$O = X \times A_h \times A_w \quad (7)$$

3.3. Multi-Scale Hierarchical Progressive Fusion Module (MSHPM)

Multi-scale feature fusion within the decoder establishes a hierarchical information flow that offers contextual compensation for the localized representations generated during encoding. Feature maps from various scales are often combined in conventional decoders via simple upsampling procedures. However, simple integration like this can result in feature inconsistency and the loss of fine-grained information due to the vast semantic and structural differences between features from different layers. As depicted in Figure 4, we suggest a Multi-Scale Hierarchical Progressive Fusion Module (MSHPM) to tackle this problem. This module uses channel and spatial attention methods in conjunction with multi-scale feature extraction to increase the fusion process's effectiveness and robustness. In particular, MSHPM uses multi-scale convolutions to capture contextual data at various scales and attention weighting to dynamically tune the significance of each feature, ensuring that during the fusion process, key features are given greater weight and priority.

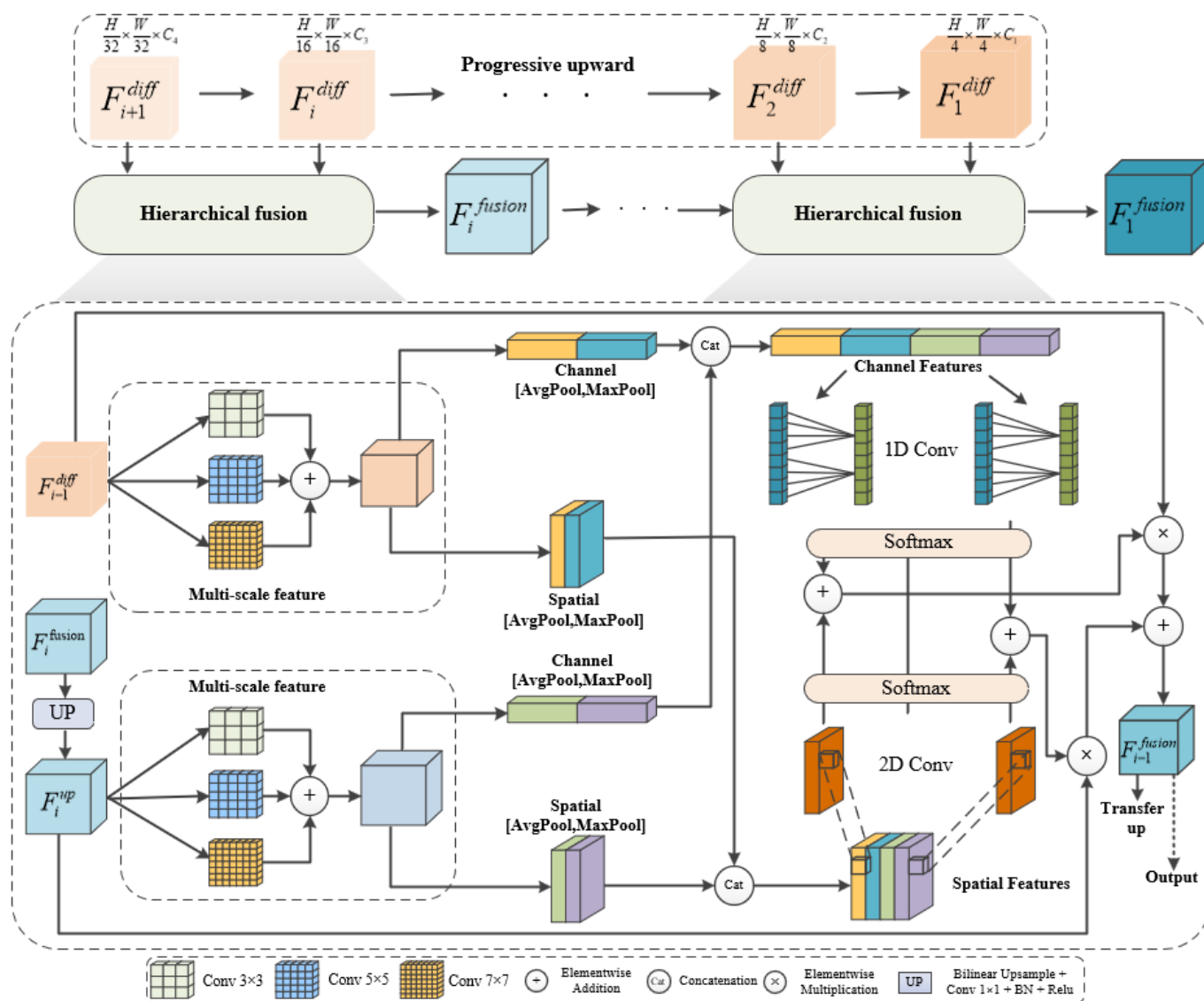


Figure 4. The Multi-Scale Hierarchical Progressive Fusion Module (MSHPM).

First, the high-resolution feature map F_{i+1}^{diff} undergoes bilinear interpolation for spatial alignment with its corresponding low-resolution counterpart F_i^{diff} . Subsequently, a 1×1 convolutional layer is employed, followed by a ReLU activation function, to decrease the channel dimensionality of the upsampled feature map. Therefore, the obtained feature map $F_{i+1}^{diff-up}$ matches the size and dimensions of the low-resolution difference map. For both feature maps F_i^{diff} and $F_{i+1}^{diff-up}$, convolution operations with three kernel sizes (3×3 , 5×5 , and 7×7) are performed to obtain multi-scale representations. Subsequently, the convolution outputs of varying sizes are aggregated to construct a multi-scale feature map. This feature map undergoes pooling operations along both the channel and spatial dimensions. The pooled features obtained from the channel and spatial dimensions are concatenated independently. Following concatenation, the channel features are fed into a 1D convolutional layer to produce temporal-specific channel weights for each of the dual-temporal image representations. These weights are then normalized via the Softmax function to guarantee that their sum equals one. Likewise, the concatenated spatial features are processed by a 2D convolution to obtain spatial weights, also normalized via Softmax. Finally, these learned channel and spatial weights are used to perform weighted summation and multiplication on $T1$ and $T2$, thereby completing the adaptive multi-scale fusion process. The detailed formulation of this procedure is as follows:

$$\begin{aligned} M_i &= \sum_{j=3,5,7} \text{Conv}_{k_j}(F_i^{diff}) \\ M_{i+1} &= \sum_{j=3,5,7} \text{Conv}_{k_j}(F_{i+1}^{diff-up}) \end{aligned} \quad (8)$$

$$\begin{aligned} M_c &= \text{ChannelPool}(\text{cat}(M_i, M_{i+1})) \\ M_s &= \text{SpatialPool}(\text{cat}(M_i, M_{i+1})) \end{aligned} \quad (9)$$

$$\begin{aligned} \omega_c^{1,2} &= \text{Soft max}(\text{Conv}_{1D}(M_c)) \\ \omega_s^{1,2} &= \text{Soft max}(\text{Conv}_{2D}(M_s)) \end{aligned} \quad (10)$$

$$F_i^{fusion} = F_i^{diff} \times (\omega_c^1 + \omega_s^1) + F_{i+1}^{diff-up} \times (\omega_c^2 + \omega_s^2) \quad (11)$$

3.4. Loss Function

In change detection tasks, the majority of regions are typically ‘unchanged’, with only a small portion being ‘changed’ areas. This class imbalance often leads the model to focus on learning the ‘unchanged’ regions, thereby neglecting the critical ‘changed’ areas. To address this issue, a hybrid loss function is constructed by fusing Binary Cross-Entropy (BCE) loss [57] with Dice coefficient (Dice) loss [58]. The BCE loss L_{BCE} and Dice loss L_{Dice} are respectively defined as

$$L_{BCE}(y, \hat{y}) = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) \quad (12)$$

$$L_{Dice}(y, \hat{y}) = 1 - \frac{2y\hat{y} + 1}{y + \hat{y} + 1} \quad (13)$$

where y represents the ground truth label and $\hat{y}$ denotes the predicted change map. The pixel values in y are binary, taking values of either 0 or 1, whereas the values in $\hat{y}$ lie within the interval $[0, 1]$, reflecting the probability that a given pixel is classified as changed or unchanged. Consequently, the overall loss function is formulated as follows:

$$L(y, \hat{y}) = \omega L_{BCE}(y, \hat{y}) + (1 - \omega) L_{Dice}(y, \hat{y}) \quad (14)$$

where the parameter ω strikes a balance between L_{BCE} and L_{Dice} . The composite loss function utilizes L_{BCE} to aid convergence while L_{Dice} addresses class imbalance. In the discussion part, later experimental studies will ascertain the correct weight parameter ω .

4. Experiments

This section presents the experimental datasets, implementation configurations, and evaluation criteria employed in this study. Subsequently, a comprehensive comparative analysis between the proposed method and current state-of-the-art (SOTA) methods is provided.

4.1. Datasets

The representative CD datasets used in the experiments are LEVIR-CD [59], DSIFN-CD [60], S2Looking [61], CLCD-CD [6], and GVLM-CD [62], which are used to validate the effectiveness of the proposed model.

(1) LEVIR-CD contains 637 pairs of high-resolution (0.5 m/pixel) bi-temporal images of size 1024×1024 pixels, focusing on building changes across diverse scenes and conditions. For model training and evaluation, each image pair was cropped into 256×256 -pixel patches, resulting in 7120, 1024, and 2048 patch pairs for the training, validation, and test sets, respectively.

(2) DSIFN-CD is composed of bi-temporal images sourced from Google Earth, featuring a spatial resolution of 2 m. It captures building changes under varying seasonal and illumination conditions, providing a testbed for multi-scenario adaptability. The dataset is partitioned into 14,389 training samples, 13,590 validation samples, and 192 test samples.

(3) S2Looking specializes in detecting building changes from side-looking (off-nadir) viewpoints, with image resolutions between 0.5 and 0.8 meters per pixel. This dataset addresses the challenge of geometric distortion and perspective variation in urban and rural settings. It is divided into 56,000, 8000, and 16,000 images for training, validation, and testing, respectively.

(4) CLCD-CD comprises 600 pairs of high-resolution (0.5–2 m/pixel) bi-temporal images 512×512 pixels in size, which were acquired by Gaofen-2 over Guangdong Province, China. The dataset specifically targets cropland changes from 2017 to 2019, capturing diverse change categories such as construction, road expansion, water bodies, and bare land. For model training and evaluation, each image pair was partitioned into 256×256 -pixel patches, yielding 1440, 480, and 480 patch pairs for the training, validation, and test sets respectively.

(5) GVLM-CD dataset comprises 17 pairs of high-resolution (0.59 m/pixel) bi-temporal Google Earth images, covering 163.77 km² for landslide mapping under varied conditions. Images are cropped into 256×256 patches and split into training (4480 pairs), validation (960 pairs), and test sets (960 pairs) in a 7:1.5:1.5 ratio.

4.2. Implementation Details and Metrics

(1) Implementation details: For all experiments, the proposed model was implemented in PyTorch and trained on a single NVIDIA RTX 4090 D (24 GB). An AdamW optimiser together with a linearly decaying learning-rate schedule was employed; the initial rate was set to 1×10^{-4} , while momentum decay and weight decay were fixed at 0.99 and 0.01, respectively. Training ran for 200 epochs with a batch size of 8, and every epoch leveraged augmentation by flipping, cropping, and blurring to strengthen robustness and ensure fair evaluation.

(2) Evaluation metrics: Performance was gauged with five standard criteria that compared predicted change maps against ground truth: Precision (Pre), Recall (Rec), F1-

score (F1), intersection over union (IoU), and overall accuracy (OA). The formal definitions of these metrics are provided below:

$$\text{Pre} = \frac{TP}{TP + FP'} \quad (15)$$

$$\text{Rec} = \frac{TP}{TP + FN'} \quad (16)$$

$$\text{F1} = \frac{2 \times \text{Rec} \times \text{Pre}}{\text{Rec} + \text{Pre}}, \quad (17)$$

$$\text{IoU} = \frac{TP}{TP + FP + FN'} \quad (18)$$

$$\text{OA} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (19)$$

TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively; all metrics range from 0 to 1, with higher values indicating better results. Network efficiency is also reported through parameter count (Params) and floating-point operations (FLOPs).

4.3. Comparison with Current State-of-the-Art Methods

We selected several state-of-the-art change detection methods for comparison, including: FC-EF [35], DTCDSCN [36], SNUNet [38], BIT [41], ChangeFormer [42], VCT [45], SEIFNet [54], CF-GCN [63], MCECF [64], RHighNet [65], and CDMamba [33]. The details are as follows:

(1) FC-EF [35]: This approach combines Fully Convolutional Networks (FCNs) and Siamese network structures, comparing image pairs through a shared-weight network architecture to automatically learn change features.

(2) DTCDSCN [36]: This method relies on a Deep Siamese Convolutional Network. Dual-task constraints are embedded in the architecture. These constraints guide every training step. They enforce feature sharing between branches. Consequently, knowledge is transferred across tasks. The overall learning process becomes more stable and efficient.

(3) SNUNet [38]: This densely connected Siamese network is designed for VHR remote sensing imagery. It employs a Siamese layout to perform pixel-level comparison of dual-temporal pairs, while DenseNet-style connections are embedded throughout to strengthen feature transmission and encourage reuse across layers.

(4) BIT [41]: The method builds on a Transformer backbone tailored for remote sensing image change detection. It taps the Transformer's ability to capture global dependencies, giving convolutional features a broader context. Semantic labeling is first applied to enrich each frame. A feature-differencing step then highlights changes, and the combined cues yield a stronger contextual representation for accurate detection.

(5) ChangeFormer [42]: The approach introduces a Transformer-based Siamese network. Hierarchical Transformer encoders are combined with multi-layer perceptron decoders. A multi-scale processing strategy is employed throughout the pipeline. In this way, fine-grained details in remote sensing images are captured effectively.

(6) VCT [45]: This method combines the self-attention mechanism inherent in Transformer architectures with a visual feature extraction network, building a multi-layer feature extraction framework that combines multi-scale information to enhance the ability to perceive subtle changes.

(7) SEIFNet [54]: This is a spatial-temporal enhancement and inter-layer fusion network that improves change detection accuracy and robustness by combining spatial and temporal feature enhancement techniques and multi-level feature fusion.

(8) CF-GCN [63]: The approach relies on Graph Convolutional Networks (GCNs). A dedicated graph structure is built to encode both spatial and semantic cues within the images. On this graph, change features are extracted and fused in a unified manner. Consequently, the method handles variations of different types and scales with consistently good performance.

(9) MCECF [64]: This is a multi-scale complementary enhanced context fusion network that improves RSCD performance by integrating multi-scale feature extraction, complementary enhancement strategies, and context information fusion.

(10) RHighNet [65]: This is a relation-aware high-order interaction network that improves CD accuracy by integrating relation-aware feature modeling and high-order interaction mechanisms.

(11) CDMamba [33]: This is a novel Mamba-based architecture that enhances binary change detection accuracy by integrating multi-scale local-contextual modeling into the selective state-space paradigm.

We implement the aforementioned CD networks by utilizing their publicly available codebases, employing the default hyperparameter settings.

4.3.1. Quantitative Evaluation

Table 1 reports the quantitative results on LEVIR-CD, DSIFN-CD, and S2Looking. All methods are compared under the criteria of Pre, Rec, F1, IoU, and OA. Across these benchmarks, EIMDGNet mostly achieves the highest scores. The numbers thus highlight the clear advantage offered by the proposed network.

Table 1. The difference between EIMDGNet and other comparison algorithms on three datasets.

Method	LEVIR-CD (%)					DSIFN-CD (%)					S2looking (%)				
	Pre	Rec	F1	IoU	OA	Pre	Rec	F1	IoU	OA	Pre	Rec	F1	IoU	OA
FC-EF [35]	86.91	80.17	83.41	71.53	98.39	72.61	52.73	61.09	43.98	88.59	54.73	33.51	37.62	29.33	95.13
DTCDSN [36]	88.53	86.83	87.67	78.05	98.77	53.87	77.99	63.72	46.76	84.91	62.49	52.89	57.42	40.27	98.99
SNUNet [38]	89.18	87.17	88.16	78.83	98.82	60.60	72.89	66.18	49.45	87.34	68.37	53.24	55.62	44.91	99.17
BIT [41]	89.24	89.37	89.31	80.68	98.92	68.36	70.18	69.26	52.97	89.41	72.79	53.99	68.81	44.73	99.09
ChangeFormer [42]	92.05	88.37	90.22	82.21	99.01	76.81	69.36	72.90	57.35	91.24	69.30	62.57	65.76	48.99	99.16
VCT [45]	92.57	87.65	90.04	81.89	99.01	83.91	66.47	74.18	58.95	92.14	72.36	53.28	61.37	44.27	99.19
SEIFNet [54]	89.46	90.49	90.95	83.40	99.10	78.32	78.22	67.77	51.11	91.33	68.31	58.27	63.23	48.22	99.16
CF-GCN [63]	92.31	88.67	91.12	82.10	99.07	76.22	72.35	62.19	62.34	92.89	71.77	46.26	56.26	39.14	99.13
MCECF [64]	91.88	90.17	90.81	81.51	99.01	77.86	68.94	73.10	58.96	90.83	70.32	59.02	62.92	47.88	99.07
RHighNet [65]	92.12	89.94	91.31	84.01	99.13	79.52	78.53	69.67	51.64	91.73	71.38	56.94	67.54	48.38	99.14
CDMamba [33]	91.43	90.08	90.75	83.07	99.06	70.62	66.13	67.21	50.72	89.98	72.89	61.94	66.98	48.78	99.15
EIMDGNet (ours)	92.91	90.74	91.49	82.93	99.17	80.52	79.37	77.32	69.39	93.26	74.93	55.37	63.34	49.19	99.20

The best-performing values are highlighted in red, while the second-best are marked in blue.

On LEVIR-CD, EIMDGNet pushes every metric beyond the levels recently reported by SEIFNet and CF-GCN. Specifically, EIMDGNet achieves 92.91%, 90.74%, 91.49%, 82.93%, and 99.17% on the Pre, Rec, F1, IoU, and OA. As shown in Table 1, traditional CNN methods such as FC-EF [35] (F = 83.41%) are limited by local feature extraction, sensitive to illumination changes, and prone to mistakenly detecting shadows as changing areas. Among the Transformer-like methods, BIT [41] (F1 = 89.31%) utilizes global dependency modeling to enhance contextual understanding but lacks edge detail perception and has missed detections in small-scale building change detection. ChangeFormer [42] (F1 = 90.22%) enhances detail capture through multi-scale Transformer but has insufficient suppression ability for complex backgrounds (such as changes in vegetation coverage). The Precision of VCT [45] is ranked as sub-optimal at 92.57%, and it performs well in the accuracy of capturing the changing area. In the multi-scale feature fusion method, the Recall of SEIFNet [54] is

90.49%, and it is better at fully detecting the changed areas. While RHighNet [65] attains the highest IoU (84.01%) among comparative methods, its F1-score (91.31%) remains slightly lower than our approach. EIMDGNet accurately outlines the building profile through the EID-MDGEM. Relative to earlier approaches, the outlined method yields boundaries that are both sharper and more continuous. Consequently, the combined effect markedly lowers omission and errors in the final change map.

On the DSIFN-CD dataset, EIMDGNet also achieves further improvements in all metrics. Specifically, EIMDGNet achieves 80.52%, 79.37%, 77.32%, 69.39%, and 93.26% on the Pre, Rec, F1, IoU, and OA. Table 1 indicates that the proposed F1 and IoU values leave rival results behind: F1 surpasses ChangeFormer [42] at 72.90% by 4.42 percentage points, while IoU beats VCT [45] at 58.95% by an extra 10.44 percentage points. Although the dense connection of SNUNet [38] (F1 = 66.18%) improves feature reuse, the fixed convolution kernel is difficult to adapt to the multi-scale variation pattern. In the attention mechanism method, SEIFNet [54] (F1 = 67.77%) enhances robustness through spatio-temporal augmentation, but the IoU is only 51.11%, reflecting inaccurate regional segmentation. EIMDGNet enhances feature differences through spatial-channel pooling and is more accurate when segmenting the building boundaries captured by different sensors.

On the S2looking dataset, EIMDGNet achieves the highest Pre (74.93%), IoU (49.19%), and OA (99.20%), as shown in Table 1, where Pre is 2.04% higher than the sub-optimal CDMamba [33] (72.89%), OA is 0.01% higher than the sub-optimal VCT [45] (99.19%), and IoU leads the sub-optimal ChangeFormer [42] (48.99%) by 0.2%. The F1-score (63.34%) is 5.50% lower than BIT [41] (68.81%), and Recall (55.37%) is lower than ChangeFormer [42] (62.57%) but higher than most comparative methods. In terms of method characteristics, as a Transformer-based method, BIT [41] performs prominently well in F1 (68.81%) by virtue of global dependency modeling, but geometric distortion in side-view perspectives results in its IoU being only 44.73%. ChangeFormer [42] improves detail capture through multi-scale Transformer, with Recall (62.57%) and F1 (65.76%) ranking second, but it is sensitive to registration errors in side-view images, being prone to generating false change detections. The traditional GCN method CF-GCN [63] is limited by the graph structure's modeling capability for irregular changes, which yields an F1 of only 56.26%. EIMDGNet balances global scene understanding and local detail positioning through edge feature enhancement and cross-dimensional feature collaboration mechanisms in side-view remote sensing scenarios, demonstrating strong robustness against complex geometric deformations and background interference.

In addition, we also compare the datasets of cropland and landslide. On the CLCD-CD dataset, As shown in Table 2. EIMDGNet outperforms all comparison methods across nearly every metric, achieving the highest scores in Precision (80.92%), Recall (79.14%), and F1-score (81.65%), while also securing the top IoU (72.91%) and a competitive OA (95.57%). Compared to runner-up methods, it shows clear improvement over MCECF [64] in Recall and F1, and surpasses CDMamba [33] in overall detection consistency despite the latter's higher OA. Similarly, on the GVLM-CD dataset, the proposed method demonstrates robust performance, attaining the best scores in Precision (86.55%), F1-score (85.49%), IoU (76.08%), and OA (98.67%), with Recall (83.43%) ranking a close second. These results indicate that EIMDGNet maintains strong detection accuracy and completeness across diverse geographical conditions and change types, effectively balancing Precision and Recall while yielding well-defined change boundaries with reduced omission and commission errors in the final change maps.

Table 2. The difference between EIMDGNet and other comparison algorithms on CLCD and GVLM.

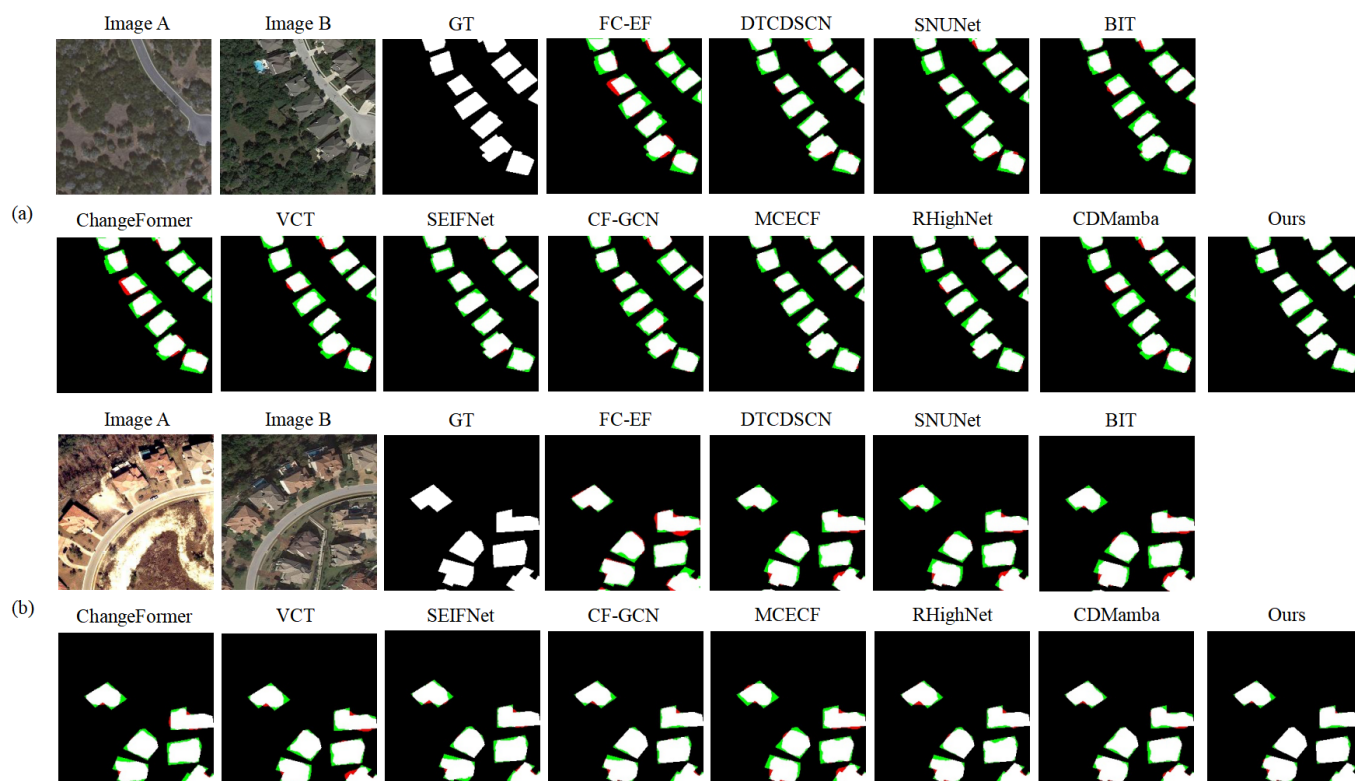
Methods	CLCD-CD (%)					GVLM-CD (%)				
	Pre	Rec	F1	IoU	OA	Pre	Rec	F1	IoU	OA
BIT [41]	76.47	76.93	78.59	68.36	93.84	85.52	81.76	83.60	71.82	98.17
ChangeFormer [42]	76.39	75.12	76.43	65.87	93.54	80.10	78.96	79.48	70.11	97.89
SEIFNet [54]	77.42	78.06	79.72	70.94	94.80	85.31	82.27	83.23	75.22	98.34
MCECF [64]	80.73	78.52	80.46	72.44	95.37	82.36	83.28	84.37	74.27	98.26
CDMamba [33]	74.78	69.05	71.81	56.02	95.96	86.58	83.67	85.10	74.85	98.48
EIMDGNet (ours)	80.92	79.14	81.65	72.91	95.57	86.55	83.43	85.49	76.08	98.67

The best-performing values are highlighted in red, while the second-best are marked in blue.

4.3.2. Qualitative Evaluation

Figures 5–7 visualize the compared algorithms on three datasets, coloring predictions as TP in white, FP in red, TN in black, and FN in green. Across the presented visual comparisons, EIMDGNet demonstrates clear superiority over the competing networks.

Advantage in managing scale variations: Traditional approaches often struggle with incomplete detection of large objects and loss of features for small objects due to limited receptive fields when dealing with scale changes. For instance, large complexes near the background in Figure 6a tend to be overlooked by other methods, while EIMDGNet successfully detects these areas. Moreover, EIMDGNet precisely identifies small buildings, with changes shown in Figure 7a. This highlights EIMDGNet’s strength in capturing and integrating multi-scale differences, as its multi-scale feature progressive fusion mechanism enables complementary information integration and effectively minimizes feature loss.

**Figure 5.** Visual comparisons of the proposed method and SOTA methods on the LEVIR-CD dataset.

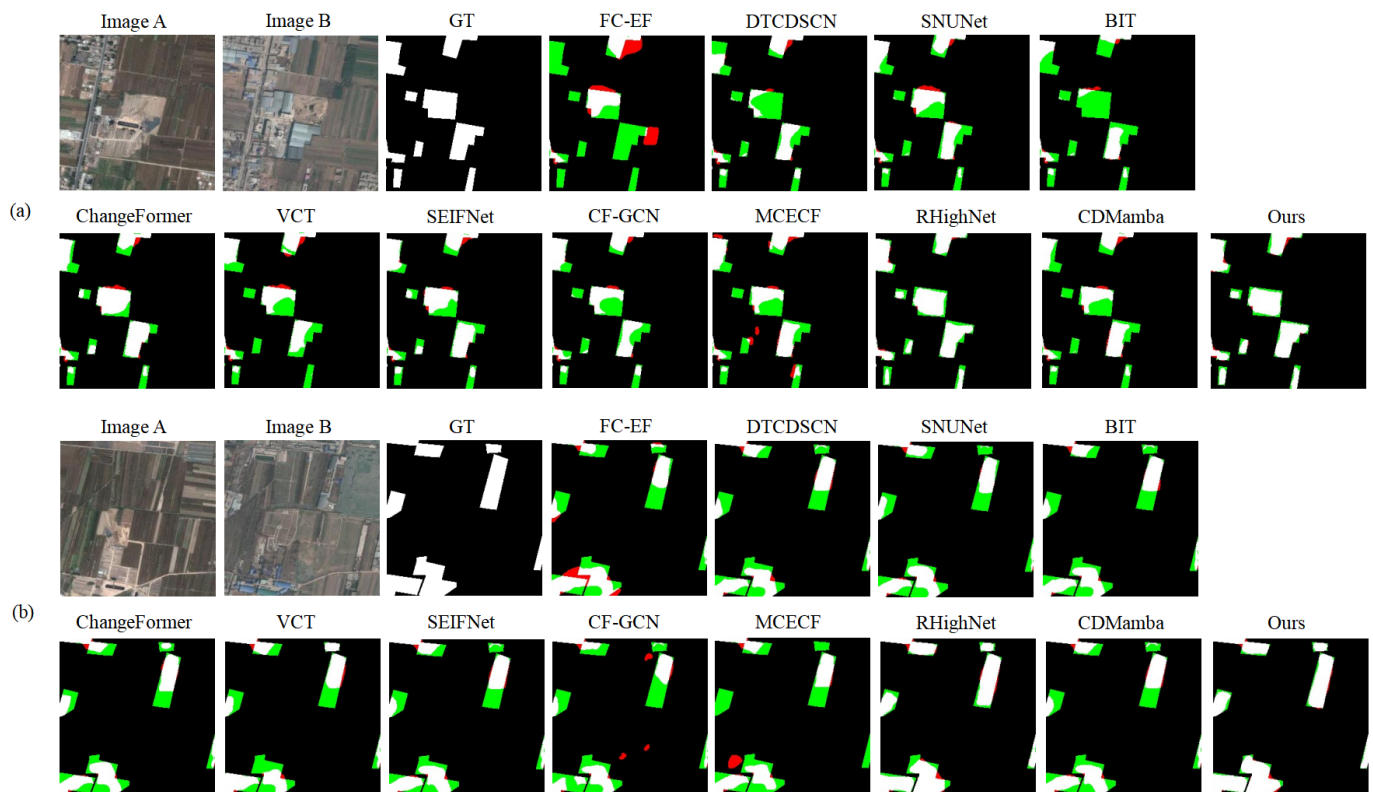


Figure 6. Visual comparisons of the proposed method and SOTA methods on the DSIFN-CD dataset.

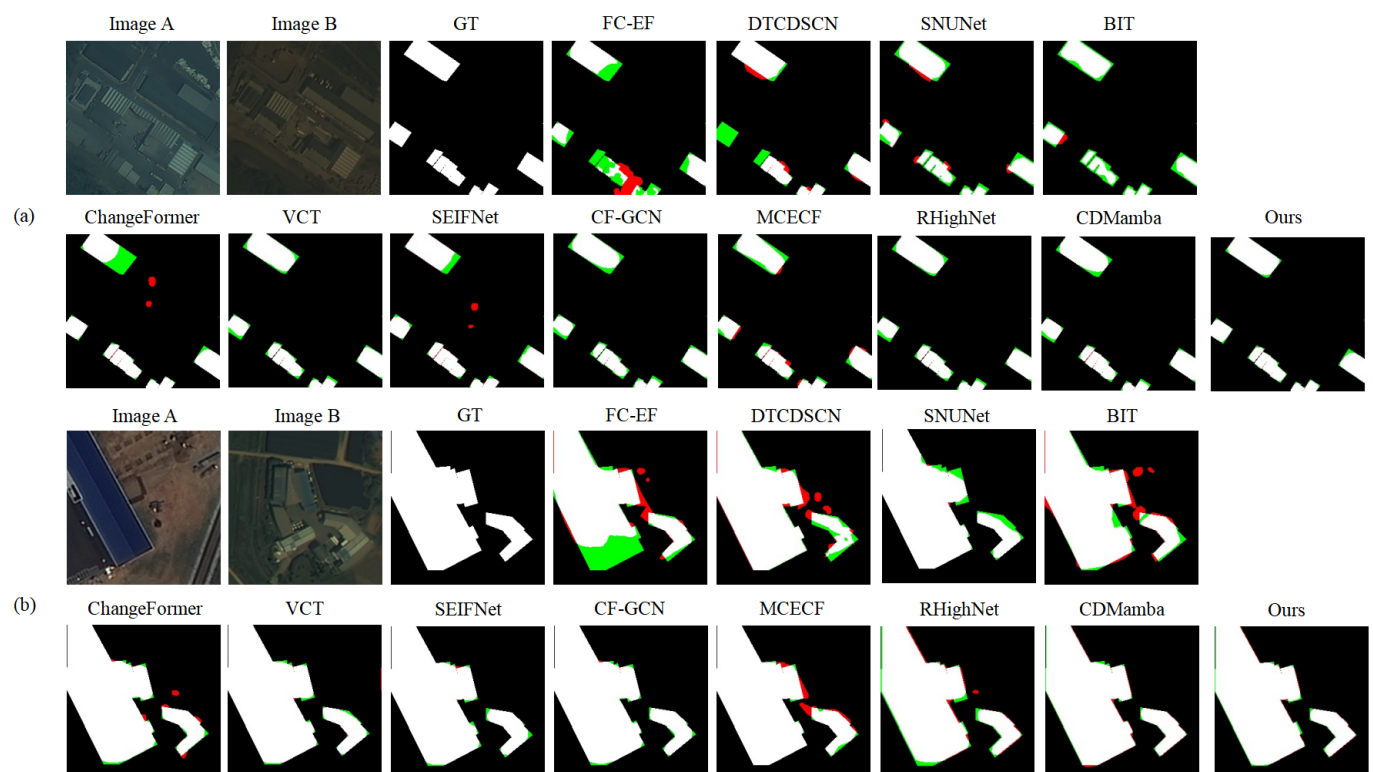


Figure 7. Visual comparisons of the proposed method and SOTA methods on the S2Looking dataset.

Sharpening edges frequently erodes internal detail, yet keeping that detail blurs boundaries. In the visual comparison on the DSIFN-CD dataset, Figures 5b and 6b show that most algorithms still fail to tell neighboring, look-alike buildings apart. While CF-GCN is capable of generating satisfactory change detection (CD) outcomes, it tends to miss

changes occurring along edge areas. In contrast, EIMDGNet excels at precisely identifying both object boundaries and the objects themselves, as shown in Figure 6 (ours). The prediction results exhibit the sparsest and most fragmented distribution of green (missed detection) segments along building contours. This observation indicates that our model possesses superior capability in accurately localizing change boundaries and maintaining their structural continuity. Concurrently, the red (false alarm) regions appearing in complex-texture areas are both the fewest in number and the smallest in extent, demonstrating effective suppression of pseudo-change signals induced by background clutter. The gain stems from the Edge Feature Enhancement Module (EFEM), which employs an attention mechanism to generate edge-saliency maps. By doing so, EFEM amplifies salient edge cues, delivering crisper boundaries without sacrificing inner detail.

To confirm the contribution of each module, exhaustive ablation tests are carried out. Every variant is evaluated on the same three datasets, and the impact of each component on EIMDGNet's detection performance is reported separately.

4.3.3. Ablation of EID-MDGEM

In EIMDGNet, EID-MDGEM (Edge-Induced Differential Multi-Dimensional Group Enhancement Module) plays a critical role in augmenting the discriminative capacity of feature representations. By performing multi-dimensional grouping and spatial-channel pooling on edge-induced differential features, it adeptly captures spatial detail variations while strengthening inter-channel feature correlations. To evaluate its significance, we verify the contribution of EID-MDGEM by incorporating it into the baseline architecture. Focusing on the LEVIR-CD dataset as a representative case, a close examination is conducted by contrasting the numeric outcomes listed in the first and second rows of Table 3.

Table 3. Ablation experiments on three datasets.

Method	EID-MDGEM	EFEM	MSHPM	LEVIR-CD			DSIFN-CD			S2Looking		
				F1	IoU	OA	F1	IoU	OA	F1	IoU	OA
Baseline				89.85	80.36	98.58	74.93	66.79	92.18	60.87	46.16	98.37
EIMDGNet-1	✓			90.73	81.73	98.90	75.43	67.77	93.10	62.11	47.15	99.07
EIMDGNet-2		✓		90.93	81.77	98.95	75.34	67.63	93.08	62.03	47.16	99.07
EIMDGNet-3			✓	90.92	81.54	98.87	75.31	67.59	93.07	61.87	46.96	98.98
EIMDGNet-4	✓	✓		91.17	82.49	99.07	76.77	68.87	93.16	62.14	47.35	99.12
EIMDGNet-5		✓	✓	91.21	82.63	99.12	76.87	68.96	93.17	62.87	48.06	99.13
EIMDGNet-6	✓		✓	91.25	82.64	99.13	76.95	69.13	93.17	63.17	48.93	99.15
EIMDGNet	✓	✓	✓	91.49	82.93	99.17	77.32	69.39	93.26	63.34	49.19	99.20

This direct comparison clearly exposes the incremental gain introduced by the EID-MDGEM component. The F1, IoU, and OA results from Baseline + EID-MDGEM (EIMDGNet-1) are 90.73%, 81.73%, and 98.90%, respectively, which are better than those of the baseline network. A glance at rows six and eight in Table 3 highlights the difference: the F1, IoU, and OA of EIMDGNet without EID-MDGEM (EIMDGNet-5) decrease to 91.21%, 82.63%, and 99.12%, respectively. These metric shifts underline that EID-MDGEM is indispensable for accentuating subtle changes and increasing change detection accuracy, thereby confirming its major contribution to the EIMDGNet framework.

We present the feature maps before and after EID-MDGEM processing to substantiate the efficacy of EID-MDGEM in Figure 8f,g. The feature map (f)–(g) resulting from EID-MDGEM processing exhibits a notable enhancement for the target of interest, accompanied by the accurate extraction of buildings in disparate scenes, thereby rendering them more discernible from the background.

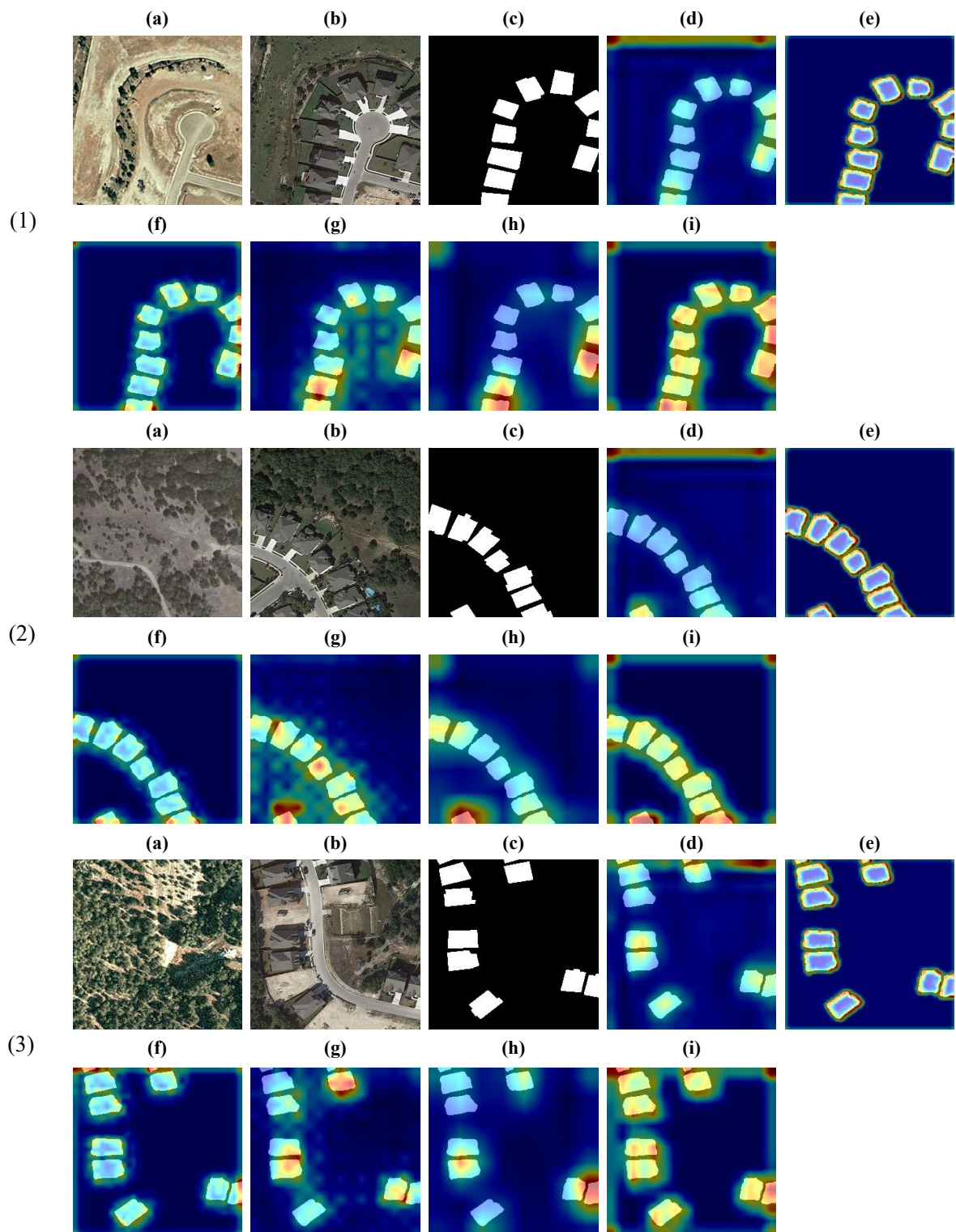


Figure 8. Visualization of feature maps from EFEM, EID-MDGEM, and MSHPM. (1)–(3) Three images of scenes in LEVIR-CD. (a) Image A; (b) Image B, (c) GT, (d) feature maps before EFEM, (e) feature maps after EFEM (warm-colored regions indicate enhanced boundaries), (f) feature maps before EID-MDGEM, (g) feature maps after EID-MDGEM (highlighted areas represent outcomes enhanced through grouping and pooling operations), (h) feature maps before MSHPM, (i) feature maps after MSHPM (the continuity and consistency of regional coloration reflect effective integration of multi-scale semantic information).

4.3.4. Ablation of EFEM

In EIMDGNet, EFEM (Edge Feature Enhancement module) adopts the attention mechanism to generate edge-saliency maps, effectively highlighting prominent edge features and proving pivotal for enhancing edge continuity by refining feature boundaries. To validate its contribution, we verify the contribution of EFEM by incorporating it into the baseline architecture. Focusing on the LEVIR-CD dataset as a representative case, a direct comparison between the first and third rows of Table 3 clearly illustrates the performance gain introduced by the EFEM. The F1, IoU, and OA results from Baseline + EFEM (EIMDGNet-2) are 90.93%, 81.77%, and 98.95%, respectively, which are better than those of the baseline network.

A side-by-side look at rows seven and eight of Table 3 reveals that, once EFEM is removed (EIMDGNet-6), the F1, IoU, and OA values slip to 91.25%, 82.64%, and 99.13%, respectively. To further substantiate the efficacy of EFEM, we present its processed feature maps alongside the unprocessed counterparts in Figure 8d,e, where the observed metric fluctuations unequivocally validate EFEM's indispensable role in maintaining the spatial consistency of change boundaries and accentuate its pivotal contribution to the EIMDGNet framework.

4.3.5. Ablation of MSHPM

In EIMDGNet, MSHPM (Multi-Scale Hierarchical Progressive Fusion Module) is essential for aggregating multi-scale contextual information and enhancing feature representation richness. By adopting a bottom-up fusion path to integrate multi-scale features, it achieves hierarchical progressive fusion through the upward propagation of detailed spatial information, thereby fully leveraging the complementary strengths of multi-scale features. To evaluate its significance, we verify the contribution of MSHPM by incorporating it into the baseline architecture. Selecting the LEVIR-CD dataset for detailed inspection, we contrast the numeric results listed in the first and fourth rows of Table 3; this pairwise comparison quantifies the incremental gain delivered by the MSHPM. The F1, IoU, and OA results from Baseline + MSHPM (EIMDGNet-3) are 90.92%, 81.54%, and 98.87%, respectively, which are notably better than the baseline network's 89.85%, 80.36%, and 98.58%. A direct comparison between rows five and eight of Table 3 highlights the performance difference; the F1, IoU, and OA of EIMDGNet without MSHPM (EIMDGNet-4) decrease to 91.17%, 82.49%, and 99.07%, respectively, in contrast to the full model's 91.49%, 82.93%, and 99.17%. To substantiate the efficacy of MSHPM, we present its processed feature maps alongside the unprocessed counterparts in Figure 8h,i, where the consistent performance gains observed with MSHPM integration and subsequent declines upon its removal unequivocally validate the module's indispensable role in capturing multi-scale change patterns, elevating change detection accuracy and thus underscoring its pivotal contribution to the EIMDGNet framework. The heatmaps are based on the Grad_CAM [66].

5. Discussion

5.1. Parametric Analysis

Table 4 illustrates that the proposed method surpasses all existing techniques across three benchmark datasets (LEVIR-CD, DSIFN-CD, S2Looking). To elucidate the impact of our approach, we report the F1-score of EIMGDNet, the accuracy of Params and FLOPs, as well as the comparison method mentioned in Table 4. Our proposed EIMDGNet strikes an optimal trade-off between computational load and detection quality, delivering top-tier results while requiring only 22.56 M parameters and 6.42 G FLOPs. On LEVIR-CD, it outperforms BIT [41] by 2.18% in F1-score while reducing computational cost by 41.2% compared to ChangeFormer [42]. Particularly noteworthy is the comparison with RHighNet [65],

where EIMDGNet achieves comparable LEVIR-CD performance (91.49% vs. 91.31% F1-score) with merely 23.3% of its parameters and 9.7% of the computational cost. On the challenging DSIFN-CD dataset, EIMDGNet advances the state of the art to 77.32% F1-score and 69.39% IoU, surpassing MCECF [64] by 4.22% in F1-score while utilizing only 54.7% of its parameters and 23.5% of its FLOPs. The model also exhibits excellent generalization on complex scenes. While BIT [41] achieves the highest F1-score (68.81%) on the S2Looking dataset, our method attains the best IoU (49.19%), representing a 4.46% improvement over BIT [41] and demonstrating enhanced capability in precisely localizing change regions. These consistent advantages across multiple datasets validate our architecture's effectiveness in capturing discriminative change features while maintaining high operational efficiency. The observed balance between accuracy and resource usage establishes EIMDGNet as a practical option for real-world remote sensing tasks. Subsequent research will target architectural modifications to boost scalability on large-scale missions and to maintain robust performance under tight computational budgets.

Table 4. Comparative analysis of computational effort and number of parameters on three datasets.

Method	Efficiency		LEVIR-CD		DSIFN-CD		S2Looking	
	Params (M)	Flops (G)	F1	IoU	F1	IoU	F1	IoU
FC-EF [35]	1.35	3.57	83.41	71.53	61.09	43.98	37.62	29.33
DTCDSCN [36]	31.26	13.22	87.67	78.05	63.72	46.76	57.42	40.27
SNUNet [38]	12.03	54.83	88.16	78.83	66.18	49.45	55.62	44.91
BIT [41]	3.55	10.92	89.31	80.68	69.26	52.97	68.81	44.73
ChangeFormer [42]	20.75	21.18	90.22	82.21	72.90	57.35	65.76	48.99
VCT [45]	3.57	10.83	90.04	81.89	74.18	58.95	61.37	44.27
SEIFNet [54]	27.91	8.37	90.95	83.40	67.77	51.11	63.23	48.22
CF-GCN [63]	13.58	43.93	91.12	82.10	62.19	62.34	56.26	39.14
MCECF [64]	41.21	27.35	90.81	81.51	73.10	58.96	62.92	47.88
RHighNet [65]	96.81	66.51	91.31	84.01	69.67	51.64	67.54	48.38
CDMamba [33]	11.90	49.68	90.75	83.07	67.21	50.72	66.98	48.78
EIMDGNet (ours)	22.56	6.42	91.49	82.93	77.32	69.39	63.34	49.19

5.2. Effects of Coefficient of Loss Function

To evaluate how weighting coefficients influence loss balance, we fixed all experimental settings and swept the trade-off parameter ω between 0 and 1 in steps of 0.25 for the BCE and Dice losses. Figure 9 confirms that the proposed method remains stable across this range. Peak scores on LEVIR-CD, DSIFN-CD, and S2looking appear when ω equals 0.75, underlining the importance of the coefficient for balanced supervision. For straightforward tuning without sacrificing accuracy, we recommend setting ω to 0.75 during practical model selection.

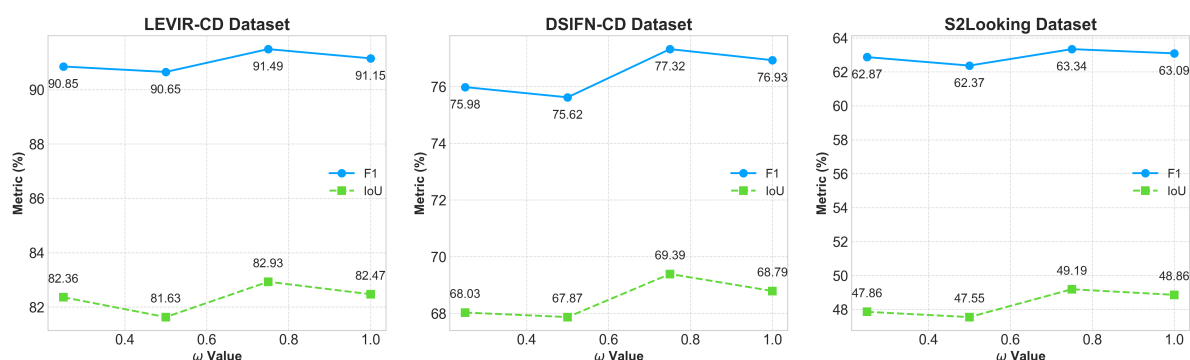


Figure 9. Comparison of different coefficient of loss functions on three datasets.

6. Conclusions

In this paper, we introduce EIMDGNet, a network architecture for CD that addresses two important issues: boundary ambiguities and inadequate multi-scale feature integration across bi-temporal images. Architecturally, the model couples a deep feature extraction module, specifically enhanced for boundary discrimination, with a multi-scale fusion scheme that operates in a hierarchical and progressive manner to optimize feature aggregation. Strong multi-scale representation extraction from dual-temporal inputs is made possible by a dual-stream architecture that is based on pre-trained ResNet18 encoders. We present the Edge-Induced Multi-Dimensional Grouping Enhancement Module (EID-MDGEM), which uses spatial-channel pooling in conjunction with multi-dimensional grouping operations to improve discriminative boundary cues and strengthen cross-channel feature interdependencies in order to improve edge-sensitive feature representation. At the same time, the Edge Feature Enhancement Module (EFEM) highlights structurally salient areas by creating edge-saliency maps using an attention-driven mechanism. At the same time, the Edge Feature Enhancement Module (EFEM) uses an attention-driven mechanism to make edge-saliency maps. This highlights structurally important areas and reduces false change signals through boundary refinement. The Multi-Scale Hierarchical Progressive Fusion Module (MSHPM) also uses a bottom-up fusion pathway to gradually combine multi-scale features by spreading spatial detail across hierarchical levels. This takes advantage of the unique strengths of different feature scales. Comprehensive evaluations are conducted on five benchmark datasets, LEVIR-CD, DSIFN-CD, S2Looking, CLCD-CD, and GVLM-CD, which consistently demonstrate the superior performance and efficiency of EIMDGNet, substantiating the efficacy of its modularized design. While effective, EIMDGNet exhibits two primary limitations: sensitivity to pronounced seasonal color shifts arising from its RGB-based ResNet18 backbone amplifying irrelevant chromatic variance and reliance on manual tuning of the loss weight parameter ω . To enhance its robustness and autonomy, future work will focus on three key directions: (1) developing automated parameter-optimization pipelines such as Bayesian optimization or gradient-based hyperparameter methods to dynamically and efficiently determine the optimal loss-balance weight ω without manual tuning; (2) integrating color-invariant backbone networks or feature representations to mitigate seasonal interference; and (3) leveraging the model's inherent flexibility to explore more advanced backbone architectures and finer-grained cross-hierarchical fusion mechanisms for further performance gains in complex remote sensing scenarios.

Author Contributions: Conceptualization, M.D. and L.S.; methodology, M.D.; software, M.D.; validation, M.D., Y.Z., and L.S.; formal analysis, M.D.; investigation, M.D. and L.S.; resources, L.S. and M.D.; data curation, M.D.; writing—original draft preparation, M.D.; writing—review and editing, M.D., L.S., and Q.Y.; visualization, M.D.; supervision, Z.W. and L.S.; project administration, M.D.; funding acquisition, L.S. and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Research and Development and Transformation Plan Projects of Qinghai Province under grant 2025-QY-215, and also by the Basic Research Program of Jiangsu under grant no. BK20250043.

Data Availability Statement: The LEVIR-CD, DSIFN-CD, S2Looking, CLCD-CD, and GVLM-CD datasets used in this study are publicly available. LEVIR-CD: <https://justchenhao.github.io/LEVIR/> (accessed on 1 December 2025). DSIFN-CD: <https://github.com/ChenHongruixuan/ChangeDetectionRepository/blob/master/Dataset/DSIFN> (accessed on 1 December 2025). S2looking: <https://github.com/S2Looking/Dataset/tree/main/S2Looking/> (accessed on 1 December 2025).

CLCD-CD: <https://github.com/liumency/CropLand-CD> (accessed on 28 January 2026). GVLM-CD: <https://github.com/zxk688/GVLM> (accessed on 28 January 2026).

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

References

1. Hu, X.; Zhang, Z.; Zhai, J.; Zhang, L.; Tang, Y.; Peng, Y.; Zhou, T. EDTST: Efficient Dynamic Token Selection Transformer for Hyperspectral Image Classification. *Remote Sens.* **2025**, *17*, 3180. [\[CrossRef\]](#)
2. Liu, F.; Chen, D.; Guan, Z.; Zhou, X.; Zhu, J.; Ye, Q.; Fu, L.; Zhou, J. RemoteCLIP: A Vision Language Foundation Model for Remote Sensing. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5622216. [\[CrossRef\]](#)
3. Xu, Y.; Xu, Z.; Wang, H.; Wei, Z.; Wu, Z. Instance-Level Orientation Enhancement for Horizontal Box Supervised Oriented Object Detection in Remote Sensing Images. *IEEE Trans. Image Process.* **2025**, *34*, 7613–7626. [\[CrossRef\]](#)
4. Ji, S.; Wei, S.; Lu, M. Fully Convolutional Networks for Multisource Building Extraction from an Open Aerial and Satellite Imagery Data Set. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 574–586. [\[CrossRef\]](#)
5. Hänsch, R.; Arndt, J.; Lunga, D.; Gibb, M.; Pedelose, T.; Boedihardjo, A.; Petrie, D.; Bacastow, T.M. Spacenet 8-the detection of flooded roads and buildings. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 1471–1479.
6. Liu, M.; Chai, Z.; Deng, H.; Liu, R. A CNN-Transformer Network with Multiscale Context Aggregation for Fine-Grained Cropland Change Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4297–4306. [\[CrossRef\]](#)
7. Bovolo, F.; Bruzzone, L. A Theoretical Framework for Unsupervised Change Detection Based on Change Vector Analysis in the Polar Domain. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 218–236. [\[CrossRef\]](#)
8. Du, B.; Ru, L.; Wu, C.; Zhang, L. Unsupervised Deep Slow Feature Analysis for Change Detection in Multi-Temporal Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9976–9992. [\[CrossRef\]](#)
9. Cui, S.; Luo, C. Feature-based non-parametric estimation of Kullback–Leibler divergence for SAR image change detection. *Remote Sens. Lett.* **2016**, *7*, 1102–1111. [\[CrossRef\]](#)
10. Im, J.; Jensen, J.R.; Tullis, J.A. Object-based change detection using correlation image analysis and image segmentation. *Int. J. Remote Sens.* **2008**, *29*, 399–423. [\[CrossRef\]](#)
11. Wang, Y.; Shao, Z.; Zuo, X.; Lu, T.; Wang, J.; Wang, Y.; Wang, S.; Zhang, Z.; Zhao, X. NSBRNet: Non-Local Spatio-Temporal Bidirectional Recurrent Network for Satellite Video Super-Resolution. *IEEE Trans. Circuits Syst. Video Technol.* **2026**, *36*, 1473–1486. [\[CrossRef\]](#)
12. Liu, Y.; Ye, Q.; Sun, L.; Wu, Z. SOD-YOLOv8n: Small Object Detection in Remote Sensing Images Based on YOLOv8n. *IEEE Geosci. Remote Sens. Lett.* **2025**, *22*, 6008405. [\[CrossRef\]](#)
13. Liu, F.; Yao, L.; Zhang, C.; Wu, T.; Zhang, X.; Jiang, X.; Zhou, J. Boost UAV-Based Object Detection via Scale-Invariant Feature Disentanglement and Adversarial Learning. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5622113. [\[CrossRef\]](#)
14. Zhu, J.; Wang, H.; Xu, Y.; Wu, Z.; Wei, Z. Self-Learning Hyperspectral and Multispectral Image Fusion via Adaptive Residual Guided Subspace Diffusion Model. In Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 10–17 June 2025; pp. 17862–17871.
15. Li, X.; He, M.; Li, H.; Shen, H. A Combined Loss-Based Multiscale Fully Convolutional Network for High-Resolution Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *19*, 8017505. [\[CrossRef\]](#)
16. Miao, J.; Li, S.; Bai, X.; Gan, W.; Wu, J.; Li, X. RS-NormGAN: Enhancing change detection of multi-temporal optical remote sensing images through effective radiometric normalization. *ISPRS J. Photogramm. Remote Sens.* **2025**, *221*, 324–346. [\[CrossRef\]](#)
17. Bai, B.; Fu, W.; Lu, T.; Li, S. Edge-Guided Recurrent Convolutional Neural Network for Multitemporal Remote Sensing Image Building Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5610613. [\[CrossRef\]](#)
18. Liu, Y.; Li, S.; He, Z.; Liu, K. Edge and Flow Guided Iterative CNN for Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4406422. [\[CrossRef\]](#)
19. Hedjam, R.; Abdesselam, A.; Melgani, F. Change Detection in Unlabeled Optical Remote Sensing Data Using Siamese CNN. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2020**, *13*, 4178–4187. [\[CrossRef\]](#)
20. Peng, D.; Bruzzone, L.; Zhang, Y.; Guan, H.; Ding, H.; Huang, X. SemiCDNet: A Semisupervised Convolutional Neural Network for Change Detection in High Resolution Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5891–5906. [\[CrossRef\]](#)
21. Wang, L.; Li, H. HMCNet: Hybrid Efficient Remote Sensing Images Change Detection Network Based on Cross-Axis Attention MLP and CNN. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5236514. [\[CrossRef\]](#)
22. Ye, Y.; Wang, M.; Zhou, L.; Lei, G.; Fan, J.; Qin, Y. Adjacent-Level Feature Cross-Fusion With 3-D CNN for Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5618214. [\[CrossRef\]](#)
23. Song, L.; Xia, M.; Weng, L.; Lin, H.; Qian, M.; Chen, B. Axial Cross Attention Meets CNN: Bibranch Fusion Network for Change Detection. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2023**, *16*, 21–32. [\[CrossRef\]](#)

24. Sun, L.; Liu, H.; Wang, X.; Zheng, Y.; Chen, Q.; Wu, Z.; Fu, L. Local–Global Information Perception Network for Salient Object Detection in Optical Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5625518. [[CrossRef](#)]
25. Miao, R.; Zhang, Y.; Dong, Y.; Du, B. Cross-Task Meta-Learning Network with Graph-Enhanced Attention Module for Hyperspectral Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5529415. [[CrossRef](#)]
26. Sun, L.; Zhang, H.; Zheng, Y.; Wu, Z.; Ye, Z.; Zhao, H. MASSFormer: Memory-Augmented Spectral-Spatial Transformer for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5516415. [[CrossRef](#)]
27. Sun, L.; Zhao, G.; Zheng, Y.; Wu, Z. Spectral–Spatial Feature Tokenization Transformer for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5522214. [[CrossRef](#)]
28. Sun, L.; Ma, X.; Wang, X.; Chen, Q.; Wu, Z. HSI Reconstruction: A Spectral Transformer with Tensor Decomposition and Dynamic Convolution. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2025**, *18*, 14988–15000. [[CrossRef](#)]
29. Li, Q.; Zhong, R.; Du, X.; Du, Y. TransUNetCD: A Hybrid Transformer Network for Change Detection in Optical Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5622519. [[CrossRef](#)]
30. Wang, W.; Tan, X.; Zhang, P.; Wang, X. A CBAM Based Multiscale Transformer Fusion Approach for Remote Sensing Image Change Detection. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2022**, *15*, 6817–6825. [[CrossRef](#)]
31. Li, W.; Xue, L.; Wang, X.; Li, G. ConvTransNet: A CNN–Transformer Network for Change Detection with Multiscale Global–Local Representations. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5610315. [[CrossRef](#)]
32. Yu, W.; Zhuo, L.; Li, J. GCFormer: Global Context-Aware Transformer for Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4703212. [[CrossRef](#)]
33. Zhang, H.; Chen, K.; Liu, C.; Chen, H.; Zou, Z.; Shi, Z. CDMamba: Incorporating Local Clues Into Mamba for Remote Sensing Image Binary Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4405016. [[CrossRef](#)]
34. Zhang, X.; Cheng, S.; Wang, L.; Li, H. Asymmetric Cross-Attention Hierarchical Network Based on CNN and Transformer for Bitemporal Remote Sensing Images Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 2000415. [[CrossRef](#)]
35. Caye Daudt, R.; Le Saux, B.; Boulch, A. Fully Convolutional Siamese Networks for Change Detection. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 4063–4067.
36. Liu, Y.; Pang, C.; Zhan, Z.; Zhang, X.; Yang, X. Building Change Detection for Remote Sensing Images Using a Dual-Task Constrained Deep Siamese Convolutional Network Model. *IEEE Trans. Geosci. Remote Sens.* **2021**, *18*, 811–815. [[CrossRef](#)]
37. N, S.; Bhaduka, K.; Lal, S.; Nalini, J.; Reddy, C.S. A Robust CNN Framework for Change Detection Analysis from Bitemporal Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2024**, *17*, 12637–12648. [[CrossRef](#)]
38. Fang, S.; Li, K.; Shao, J.; Li, Z. SNUNet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 8007805. [[CrossRef](#)]
39. Chen, P.; Zhang, B.; Hong, D.; Chen, Z.; Yang, X.; Li, B. FCCDN: Feature constraint network for VHR image change detection. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 101–119. [[CrossRef](#)]
40. Panda, M.K.; Sharma, A.; Bajpai, V.; Subudhi, B.N.; Thangaraj, V.; Jakheti, V. Encoder and decoder network with ResNet-50 and global average feature pooling for local change detection. *Comput. Vis. Image Underst.* **2022**, *222*, 103501. [[CrossRef](#)]
41. Chen, H.; Qi, Z.; Shi, Z. Remote Sensing Image Change Detection with Transformers. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5607514. [[CrossRef](#)]
42. Bandara, W.G.C.; Patel, V.M. A Transformer-Based Siamese Network for Change Detection. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 207–210.
43. Chen, Z.; Zhou, Y.; Wang, B.; Xu, X.; He, N.; Jin, S.; Jin, S. 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.* **2022**, *191*, 203–222. [[CrossRef](#)]
44. Ding, L.; Zhang, J.; Guo, H.; Zhang, K.; Liu, B.; Bruzzone, L. Joint Spatio-Temporal Modeling for Semantic Change Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5610814. [[CrossRef](#)]
45. Jiang, B.; Wang, Z.; Wang, X.; Zhang, Z.; Chen, L.; Wang, X.; Luo, B. VcT: Visual Change Transformer for Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 2005214. [[CrossRef](#)]
46. Yin, M.; Chen, Z.; Zhang, C. A CNN-Transformer Network Combining CBAM for Change Detection in High-Resolution Remote Sensing Images. *Remote Sens.* **2023**, *15*, 2406. [[CrossRef](#)]
47. Lv, Z.; Wang, F.; Cui, G.; Benediktsson, J.A.; Lei, T.; Sun, W. Spatial–Spectral Attention Network Guided With Change Magnitude Image for Land Cover Change Detection Using Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4412712. [[CrossRef](#)]
48. Li, Z.; Cao, S.; Deng, J.; Wu, F.; Wang, R.; Luo, J.; Peng, Z. STADE-CDNet: Spatial–Temporal Attention with Difference Enhancement-Based Network for Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5611617. [[CrossRef](#)]

49. Wang, M.; Li, X.; Tan, K.; Mango, J.; Pan, C.; Zhang, D. Position-Aware Graph-CNN Fusion Network: An Integrated Approach Combining Geospatial Information and Graph Attention Network for Multiclass Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4402016. [[CrossRef](#)]
50. Ji, Y.; Sun, W.; Wang, Y.; Lv, Z.; Yang, G.; Zhan, Y.; Li, C. Domain Adaptive and Interactive Differential Attention Network for Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5616316. [[CrossRef](#)]
51. Feng, Y.; Xu, H.; Jiang, J.; Liu, H.; Zheng, J. ICIF-Net: Intra-Scale Cross-Interaction and Inter-Scale Feature Fusion Network for Bitemporal Remote Sensing Images Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4410213. [[CrossRef](#)]
52. Luo, F.; Zhou, T.; Liu, J.; Guo, T.; Gong, X.; Ren, J. Multiscale Diff-Changed Feature Fusion Network for Hyperspectral Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5502713. [[CrossRef](#)]
53. Jiang, M.; Chen, Y.; Dong, Z.; Liu, X.; Zhang, X.; Zhang, H. Multiscale Fusion CNN-Transformer Network for High-Resolution Remote Sensing Image Change Detection. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2024**, *17*, 5280–5293. [[CrossRef](#)]
54. Huang, Y.; Li, X.; Du, Z.; Shen, H. Spatiotemporal Enhancement and Interlevel Fusion Network for Remote Sensing Images Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5609414. [[CrossRef](#)]
55. Wang, Y.; Shao, Z.; Lu, T.; Huang, X.; Wang, J.; Zhang, Z.; Zuo, X. Lightweight remote sensing super-resolution with multi-scale graph attention network. *Pattern Recognit.* **2025**, *160*, 111178. [[CrossRef](#)]
56. Yang, L.; Zhang, R.; Li, L.; Xie, X. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In Proceedings of the 38th International Conference on Machine Learning, Virtual Event, 18–24 July 2021, Volume 139, pp. 11863–11874.
57. Mannor, S.; Peleg, D.; Rubinstein, R. The cross entropy method for classification. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 7–11 August 2005; pp. 561–568.
58. Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Cardoso, M.J. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In Proceedings of the International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Québec City, QC, Canada, 14 September 2017; pp. 240–248.
59. Chen, H.; Shi, Z. A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection. *Remote Sens.* **2020**, *12*, 1662. [[CrossRef](#)]
60. Zhang, C.; Yue, P.; Tapete, D.; Jiang, L.; Shangguan, B.; Huang, L.; Liu, G. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 183–200. [[CrossRef](#)]
61. Shen, L.; Lu, Y.; Chen, H.; Wei, H.; Xie, D.; Yue, J.; Chen, R.; Lv, S.; Jiang, B. S2Looking: A Satellite Side-Looking Dataset for Building Change Detection. *Remote Sens.* **2021**, *13*, 5094. [[CrossRef](#)]
62. Zhang, X.; Yu, W.; Pun, M.-O.; Shi, W. Cross-Domain Landslide Mapping from Large-Scale Remote Sensing Images Using Prototype-Guided Domain-Aware Progressive Representation Learning. *ISPRS J. Photogramm. Remote Sens.* **2023**, *197*, 1–17. [[CrossRef](#)]
63. Wang, W.; Liu, C.; Liu, G.; Wang, X. CF-GCN: Graph Convolutional Network for Change Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5607013. [[CrossRef](#)]
64. Huang, Z.; Qiu, H.; Hou, M.; Yu, Z.; Wang, S.; Li, X.; Wang, J.; Yan, Y.; Liu, Y. MCECF: A Multiscale Complementary Enhanced Context Fusion Network for Remote Sensing Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4407414. [[CrossRef](#)]
65. Dong, H.; Du, X.; Li, Z.; Ma, Z.; Wang, Y.; Zhu, H.; Ma, W. Relation-Aware High-Order Interaction Network for Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5640116. [[CrossRef](#)]
66. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626.

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