Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.060966

ARTICLE

CE-CDNet: A Transformer-Based Channel Optimization Approach for Change Detection in Remote Sensing

Jia Liu[1], Hang Gu[1], Fangmei Liu[1], Hao Chen[1], Zuhe Li[1], Gang Xu[2], Qidong Liu[2] and Wei Wang[2,*]

[1]School of Computer Science and Technology, Zhengzhou University of Light Industry, Zhengzhou, 450002, China
[2]Department of Computing, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China

*Corresponding Author: Wei Wang. Email: wei.wang03@xjtlu.edu.cn

Received: 13 November 2024; Accepted: 06 January 2025; Published: 26 March 2025

ABSTRACT: In recent years, convolutional neural networks (CNN) and Transformer architectures have made significant progress in the field of remote sensing (RS) change detection (CD). Most of the existing methods directly stack multiple layers of Transformer blocks, which achieves considerable improvement in capturing variations, but at a rather high computational cost. We propose a channel-Efficient Change Detection Network (CE-CDNet) to address the problems of high computational cost and imbalanced detection accuracy in remote sensing building change detection. The adaptive multi-scale feature fusion module (CAMSF) and lightweight Transformer decoder (LTD) are introduced to improve the change detection effect. The CAMSF module can adaptively fuse multi-scale features to improve the model's ability to detect building changes in complex scenes. In addition, the LTD module reduces computational costs and maintains high detection accuracy through an optimized self-attention mechanism and dimensionality reduction operation. Experimental test results on three commonly used remote sensing building change detection data sets show that CE-CDNet can reduce a certain amount of computational overhead while maintaining detection accuracy comparable to existing mainstream models, showing good performance advantages.

KEYWORDS: Remote sensing; change detection; attention mechanism; channel optimization; multi-scale feature fusion

1 Introduction

In the field of remote sensing, change detection (CD) is the task of analyzing land surface variations using remote sensing images from different periods [1,2]. It is not only extensively applied in geological exploration and ecological water conservation, but also can be employed to evaluate the impact of human survival activities [3,4]. Building change detection (BCD) plays an indispensable role in understanding urban scope changes, land use and other fields [5,6]. Advancements in remote sensing satellite and drone technologies have made it possible to quickly and easily acquire high-resolution space imagery [7]. These data contain a large amount of rich detail, such as shape, texture and color features. However, due to differences in shooting equipment, time, angle, etc., change detection does not have a unified standard for detailed features, and it is difficult to define a unique boundary between buildings and non-buildings [8]. This explains why the accuracy of change detection does not necessarily improve with the increasing precision of the device. As remote sensing images from different phases are taken at various times and angles, many pseudo-changes such as seasonal tree changes, rivers changes and image color changes will bring a lot of difficulties to remote

sensing building change detection [9]. Detection methods based on convolutional neural networks (CNN) have shown good results in different tasks using convolution operations and attention mechanisms,but there are still some difficulties in capturing subtle changes between different image regions.

Recently, deep learning has gained popularity in building change detection (BCD) for its strong ability in feature extraction and pattern recognition [10]. Since the introduction of Fully Convolutional Networks (FCNs) into change detection, CNN-based methods have started to dominate this field [11]. CNNs employ multi-layer convolutions and pooling operations to effectively extract local features from images, enabling the detection of fine-grained building changes [12,13]. However, CNNs have an inherent limitation—their limited receptive fields prevent them from capturing long-range pixel dependencies [14]. As urban environments become more complex, relying solely on local feature extraction is no longer sufficient to address the full range of challenges in building change detection [15]. Differences in building scale, image acquisition angles and times, along with pseudo-changes in the surrounding environment (e.g., trees, rivers), can cause confusion and misclassifications during detection [16].

Unlike CNNs, the Transformer architecture, which has risen in popularity in recent years, demonstrates significant potential in capturing long-range dependencies and global features [17], also, applied to remote sensing CD tasks [18]. Compared with traditional CNN, the stacked self-attention module of Transformers is good at managing long-distance relationships in images, overcoming the limitations of CNN in processing global information. In current Transformer-based CD studies, an increasing number of researchers utilize Transformers as encoders to extract robust and representative features, which are then processed by the Transformer decoder [19]. However, simple symmetric encoder-decoder networks often struggle to efficiently process and transfer these features. As a result, significant challenges remain in effectively capturing building changes over time from multi-temporal images. Therefore, the effective integration of multi-scale features, along with advanced spatio-temporal modeling techniques, has emerged as a critical issue in the field of BCD.

However, as model complexity increases, particularly with the adoption of Transformer architectures, the multi-layer self-attention mechanisms demand substantial computational resources, resulting in slower inference speed and higher energy consumption. Although state-of-the-art models deliver excellent performance on perform well metrics (e.g., F1-score), their computational overhead and resource demands are frequently disproportionate to the gains in accuracy. This imbalance is particularly problematic in resource-constrained environments. Consequently, the deployment of such models on edge devices, mobile devices, and other low-resource platforms remains restricted. Due to the limited computational power and memory of these environments, high-resource-consuming models are often impractical for real-world applications.

In this context, reducing the computational cost and resource demands of models while maintaining reasonable accuracy has become a key challenge in remote sensing image processing. To this end, this study introduces an innovative network architecture designed to optimize parameter efficiency and reduce the number of channels, significantly lowering computational overhead while preserving detection performance.

In order to reduce the excessive demand for computing resources in BCD task, this paper proposes a channel-optimized network—Channel-Efficient Change Detection Network (CE-CDNet). By reducing the number of channels for data propagation at each level in the network, the computational cost is reduced. At the same time, the accuracy of CD can be preserved. The main contributions are as follows:

1. We optimized the network structure and channel design, so that the channel-efficient change detection network (CE-CDNet) effectively reduces the model complexity without sacrificing detection accuracy. CE-CDNet controls the number of channels in each layer to ensure the consistency of accuracy in different detection scenarios.

2. We introduced Channel-Aware Multi-Scale Fusion (CAMSF) to generate attention maps using global average pooling and global maximum pooling to enhance key features and suppress irrelevant interference. Accurately capture subtle and complex changes between images at different times.
3. We introduced the Lightweight Transformer Decoder (LTD), which is effectively captured by the self-attention mechanism. The LTD module reduces the dimensionality of the feature map and reduces the computational burden of high-dimensional data. In addition, the multi-head attention mechanism more effectively analyzes the spatial relationship between image regions.

The rest is arranged as follows: Section 2 reviews related works in change detection, focusing on traditional methods and deep learning-based approaches, along with recent developments in Transformer architectures for remote sensing. Section 3 describes the proposed CE-CDNet architecture in detail, including the feature extraction module, channel-aware multi-scale fusion (CAMSF) module, lightweight Transformer decoder (LTD) module, and the channel optimization strategy. Section 4 outlines the experimental setup, datasets, evaluation metrics, and presents a comparative analysis of CE-CDNet against existing models. And presents the results of ablation studies, highlighting the contribution of each module to the overall performance of CE-CDNet. Finally, Section 5 summarizes the contributions of this paper.

2 Related Works

2.1 Classical Change Detection

Previously, change detection in remote sensing images predominantly relied on traditional image processing techniques, includes image differencing, image overlay, change vector analysis (CVA), and principal component analysis (PCA). In addition, the classical method is simpler and more effective. For example, multivariate change detection (MAD) [20] uses the contrast in different images for detection. Slow feature analysis (SFA) [21] often performs well in long-term monitoring tasks. In addition, Fourier domain methods [22] convert data into frequency domain for monitoring and have relatively good results. Du et al. [23] proposed a difference-based image fusion method that combines the outputs of various simple change detectors using feature-level and decision-level fusion techniques. Qin et al. [24] proposed an object-based 3D building detection framework that combines orthophotos and digital surface models (DSMs) to identify building changes by analyzing height, spectral, and shape features using supervised classification with decision trees and support vector machines (SVMs). Tan et al. [25] proposed an object-based change detection method (OB-MMUA) that combines multiple classifiers and multi-scale uncertainty analysis. This method extracts spectral, texture, shapegray-level co-occurrence matrix (GLCM), and Gabor filter features. Random forests are used for feature selection, followed by Dempster-Shafer (D-S) evidence theory to merge SVM, k-nearest neighbor (KNN), and extra-tree classifiers. Yousif et al. [26] proposed a change detection method for high-resolution synthetic aperture radar (SAR) images, representing pixel-level change signals and employing Fourier and wavelet transforms to quantify the intensity of changes in objects, thereby capturing dominant change behaviors while suppressing interference.

Traditional methods detect changes in simple scenarios by directly comparing pixel values, spectral features, and geometric characteristics across multi-temporal images. However, their effectiveness is limited in complex environments.

2.2 Change Detection Using Deep Learning

CNNs have been able to effectively capture the detailed changes in buildings by extracting local image features through convolution operations. The strengths of CNNs in automatic feature extraction and pattern recognition have led to outstanding performance in various image-related tasks. Similarly, the

introduction of CNN into change detection has made it possible to effectively capture changes in building details, bringing considerable progress. Many methods have also emerged that combine multi-scale feature fusion with spatiotemporal modeling to improve the robustness and accuracy of detection. Li et al. [27] proposed a multi-scale fully convolutional network (MFCN) that uses multi-scale convolution kernels to extract detailed features of the target. Zhang et al. [28] proposed to construct a dual-channel network through transfer learning to generate multi-scale and multi-depth feature difference maps. Incorporating the attention mechanism also greatly enhanced change detection performance. Shen et al. [29] developed the global channel attention (GCA) module and multi-scale feature fusion (MSFF) module to enhance detection accuracy of low-level features and integrate multi-scale information. Chen et al. [30] used dual attention mechanisms to capture long-range dependencies, enhancing the discriminative power of feature representation. By using a weighted dual-margin contrastive loss function, they addressed sample imbalance and reduced the impact of pseudo-changes. Shi et al. [31] proposed combining deep metric learning with convolutional blocks and attention blocks to learn variation maps. Zhang et al. [32] aimed to reconstruct the change map by fusing the multi-layer deep features of the original image with the difference features using an attention module. Additionally, other deep learning methods, such as GNNs, have been explored. Liang et al. [33] proposed a multi-scale fusion network model based on Graph Neural Networks (GNNs) called GNN-based multi-scale fusion network model (GCNCD), which learns richer features by message pooling from neighboring vertices in the graph. Holail et al. [34] proposed an attention-based feature difference enhancement (AFDE) network that combines spatial channel attention and deep supervision to effectively address the issue of spatial information loss.

Due to the typically small receptive field of CNNs, the models face difficulties in capturing long-range dependencies between pixels, which is particularly problematic in building change detection, especially when dealing with buildings that exhibit significant changes in scale or viewpoint. Moreover, CNNs primarily focus on extracting local features, which may be insufficient to handle the diverse changes in buildings within complex urban environments. For example, when changes occur in the surroundings of buildings, such as the growth or disappearance of trees, or changes in water bodies, CNNs may mistakenly identify these changes as building changes, affecting the accuracy of the detection results.

2.3 Vision Transformers for Change Detection

Recently, the Transformer architecture has seen successful application across various fields, has gradually been introduced into the field of computer vision. Through its stacked self-attention mechanism, the Transformer efficiently captures global features and long-range relationships in images, addressing the shortcomings of CNNs. Recently, Mamba-based methods [35] have emerged and have shown considerable advantages in the field of change detection. It shows great potential to replace existing solutions. It introduces linear attention and long-range dependencies to reduce complexity and computational overhead, and performs extremely well.

Trying to use Transformer as an encoder to extract the change feature image and using the decoder for change detection is also a new direction [17]. ChangeFormer [36] combined the Transformer's global feature-capturing capability with the demands of change detection, showcasing the potential of the Transformer in this field. TransUNet [37], by combining the Transformer and UNet, demonstrated the strong encoding capabilities of the Transformer in image segmentation tasks, providing a useful reference for the field of remote sensing change detection. As a result, many researchers have adopted hybrid approaches combining Transformers and CNNs. Liang et al. [38] proposed a Transformer-CNN hybrid network for remote sensing building change detection, which combines CNN and Transformer and fuses the features of both to improve accuracy. To address the interaction issues between CNNs and Transformers. Zhang et al. [39] introduced

a hierarchical network with asymmetric cross-attention. Song et al. [40] proposed a method combining Progressive Sampling (PS) and Transformers, iteratively optimizing the feature information of remote sensing images. Inspired by visual Transformers. Shi et al. [41] developed a token-based approach for passing global context information without using additional Transformer decoders or skip connections.

3 Method

This chapter provides a detailed explanation of the overall architecture of the proposed Channel-Efficient Change Detection Network (CE-CDNet) and the design and implementation of its various modules. The goal of CE-CDNet is to improve the model's efficiency by reducing its computational complexity and resource consumption, without significantly compromising detection accuracy.

3.1 Channel-Efficient Change Detection Network

In remote sensing change detection, although the spatial resolution of acquired images is constantly improving and the details are becoming clearer. However, the accuracy has not been further improved with the improvement of technology. On the contrary, the computing resources and accuracy are constantly increasing. How to balance computer resources and accuracy is an issue that cannot be ignored. Especially after the transformer-based models have shown strong performance in image processing, the models now tend to stack transformer blocks and gradually increase the number of parameters in the hope of improving accuracy. As a result, these models are difficult to deploy on edge devices or resource-constrained environments.

We propose a Channel-Efficient Change Detection Network (CE-CDNet) that achieves a balance between accuracy and computing resources by optimizing the number of channels propagated between networks. CE-CDNet combines the advantages of Transformers and CNNs, significantly reducing the computational overhead while maintaining high detection accuracy. The design of this network consists of four main components: a feature extraction module, a multi-scale feature fusion module, a lightweight Transformer decoder module, and a channel optimization strategy. This combination allows CE-CDNet to maintain efficiency and accuracy while processing images. The specific process is shown in Algorithm 1.

Algorithm 1: CE-CDNet algorithm table

Input: Remote sensing images T_0 and T_1.
Output: Change Detection Map (CDM), the final binary result L.
//**Feature Extraction**
1.1: Read the input images T_0 and T_1.
1.2: Pass the images through the Siamese feature extractor to obtain feature matrices X_{T0}, X_{T1}.
//**Feature Update Using CAMSF Module**
2.1: Compute the spatial attention map A using convolutional operations and the attention mechanism.
2.2: Calculate the adaptive multi-scale feature fusion: $Z = f(X_{T0}, X_{T1}, A)$, where Z represents the updated feature matrix incorporating attention.
//**Global Context Extraction Using Lightweight Transformer Decoder (LTD)**
3.1: Encode the multi-scale feature tokens using **Trans Encoder**.
3.2: Apply self-attention mechanism to capture spatial dependencies.
3.3: Decode tokens using **Trans Decoder** to generate refined feature maps.
//**Change Detection Map (CDM) Generation**

(Continued)

Algorithm 1 (continued)

4.1: Compute the feature distance map $D_{i,j}$ based on the difference between feature representations at each pixel.
4.2: Compute the threshold θ using global context information extracted by the transformer.
4.3: For each pixel (i, j)):
- If $D_{i,j} > \theta$, then $P_{i,j} = 1$ (Change detected).
- Else, $P_{i,j} = 0$ (No change).
//Output Binary Map
5.1: Generate the final binary change detection map L.

1. Feature extraction module: This module employs a lightweight CNN architecture to extract multi-scale features from input multi-temporal remote sensing images. This module minimizes computational complexity by reducing the number of channels in each layer, while effectively extracting key features related to building changes, thereby ensuring accuracy.
2. Multi-scale feature fusion module: Adaptively fuses features of different scales to accurately capture building changes in complex scenes.
3. Lightweight Transformer decoder module: The LTD module optimizes the self-attention mechanism and dimensionality reduction, effectively captures complex spatial relationships during feature fusion and upsampling, greatly reduces the computational burden, and achieves efficient change detection.
4. Channel optimization strategy: Channel optimization runs through the entire network design. By streamlining the number of channels, the computational cost of the model is reduced, and the adaptability of the model in resource-constrained environments is enhanced.

CE-CDNet receives two images and uses the twin feature extraction method with shared weights in a dual-branch convolutional neural network to extract features (as shown in Fig. 1). This process generates feature maps at two scales, with dimensions of 64 × 64 and 256 × 256, and all feature maps in the early stages of the network have 32 channels. These feature maps are then fused through the Channel-Aware Multi-Scale Fusion (CAMSF) module, ensuring that building changes are accurately captured at different scales. The fused feature maps are fed into the Transformer encoder and decoder modules, where spatial attention and positional encoding are applied, transforming the feature maps into tokens. Further processing occurs in the Lightweight Transformer Decoder (LTD) module, which optimizes the self-attention mechanism and performs dimensionality reduction, effectively reconstructing complex spatial relationships in the images. The final output is a binary change detection map highlighting building changes.

3.2 Channel-Aware Multi-Scale Fusion (CAMSF)

The Channel-Aware Multi-Scale Fusion module is one of the key components of CE-CDNet. The goal is to minimize the computational overhead while maintaining detection accuracy by fusing features of different scales. The CAMSF module receives two sets of feature maps from the Siamese feature extraction module, with dimensions of 64 × 64 and 32 channels, which correspond to the feature representations of the input images T0 and T1 (as shown in Fig. 2).

First, the feature maps are concatenated to generate a feature map $\mathbf{F}_{\text{cat}}$ with 64 channels, where:

$$\mathbf{F}_{\text{cat}} = \text{Concat}\left(\mathbf{F}_{64,1}, \mathbf{F}_{64,2}\right) \in \mathbb{R}^{64\times64\times64} \tag{1}$$

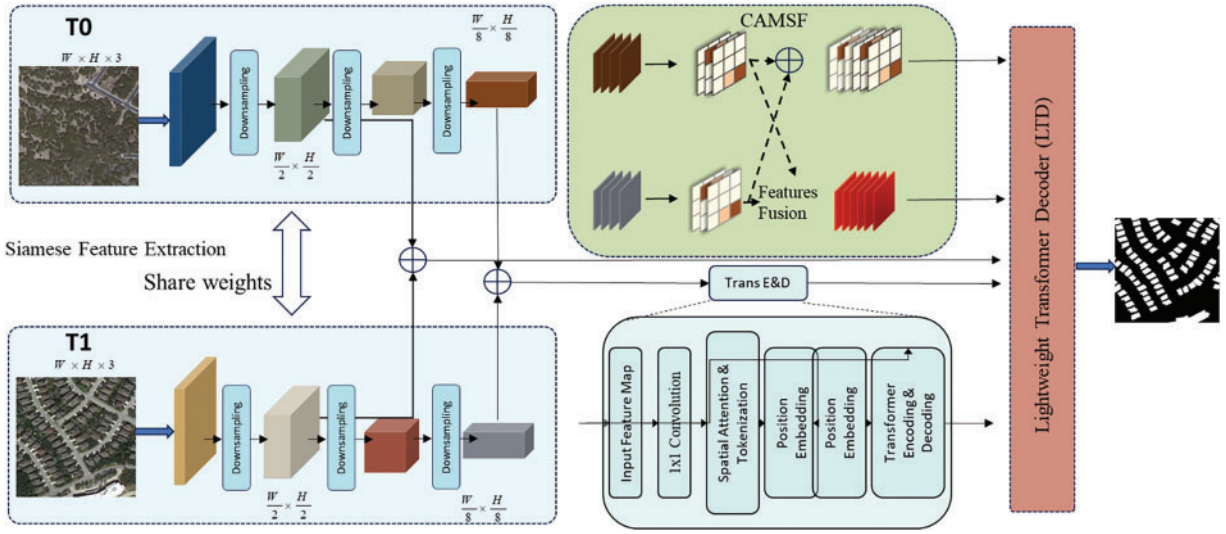

Figure 1: Architecture of Channel-Efficient Change Detection Network (CE-CDNet)

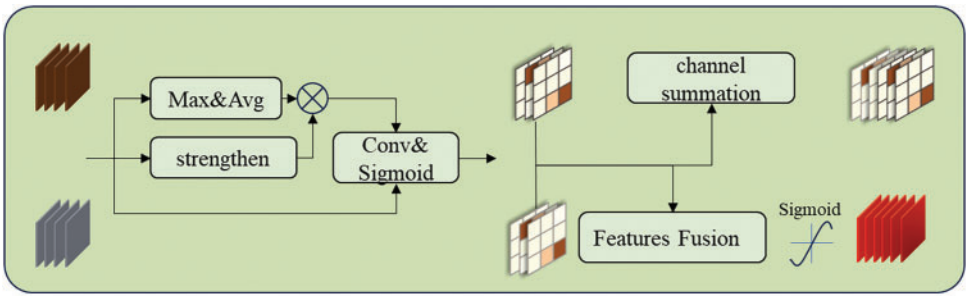

Figure 2: Channel summation and feature fusion process in CAMSF module

For enhanced effect, the CAMSF module first applies adaptive weight allocation to the concatenated feature map $\mathbf{F}_{\text{cat}}$. Through Global Average Pooling (GAP) and Global Max Pooling (GMP), channel-level attention maps are generated as follows:

$$\mathbf{A}_{\text{avg}} = \text{ReLU}\left(\mathbf{W}_1 \cdot \left(\mathbf{F}_{\text{cat}}\right)\right), \mathbf{A}_{\text{max}} = \text{ReLU}\left(\mathbf{W}_1 \cdot \text{GMP}\left(\mathbf{F}_{\text{cat}}\right)\right) \tag{2}$$

Next, these attention maps $\mathbf{A}_{\mathbf{avg}}$ and $\mathbf{A}_{\mathbf{max}}$ are combined with the original feature map, and pixel-wise weighting is applied to generate a feature map with channel attention $\mathbf{F}_{\text{att}}$:

$$\mathbf{F}_{\mathbf{att}} = \sigma\left(\mathbf{A}_{\mathbf{avg}} + \mathbf{A}_{\mathbf{max}}\right) \cdot \mathbf{F}_{\mathbf{cat}} \tag{3}$$

Here, $\sigma(\cdot)$ is the Sigmoid activation function, ensuring that important channel information is enhanced, thus better preserving the key information related to building changes.

After obtaining the feature map with adaptive weights $\mathbf{F}_{\text{att}}$, the CAMSF module performs multi-scale fusion. Through weighted fusion, it combines the original feature maps $\mathbf{F}_{64,1}$ and $\mathbf{F}_{64,2}$ to form the final fused feature map:

$$\mathbf{F}_{\text{fusion}} = \alpha \cdot \mathbf{F}_{64,1} + \beta \cdot \mathbf{F}_{64,2} \tag{4}$$

α and β are the weights learned adaptively through network training, balancing the contributions of different scales during the fusion process to ensure that the final fused feature map fully expresses multi-scale building change information.

The fused feature map $\mathbf{F}_{\text{fusion}}$, processed by the CAMSF module, contains multi-scale information and effectively balances the contributions of this information across the channel dimension. This feature map is then passed to the subsequent Lightweight Transformer Decoder (LTD) module for further building change detection. By applying adaptive weight allocation and multi-scale feature fusion, the CAMSF module reduces computational complexity and enhances the feature representation capabilities.

3.3 Channel-Aware Multi-Scale Fusion (CAMSF)

In the CE-CDNet network, the Transformer mechanism is utilized through the Trans Encoder and Trans Decoder modules to process the global contextual information of the input images, further improving the model's performance in complex building change detection tasks. The Trans Encoder captures spatial dependencies by mapping image features into a set of tokens and enhances global information extraction by incorporating positional embeddings.

Specifically (as shown in Fig. 3), the Trans Encoder module first performs convolution operations on the input multi-scale features, generating a spatial attention matrix to weight the features of different regions and further extract global features through the Transformer mechanism. These token representations incorporate positional information through positional encoding, enabling the network to understand dependencies between distant pixels in the image. The Trans Decoder integrates the encoded global information with the original features and converts the tokens back into spatial features using cross-feature attention mechanisms. It remaps the tokens into feature maps.

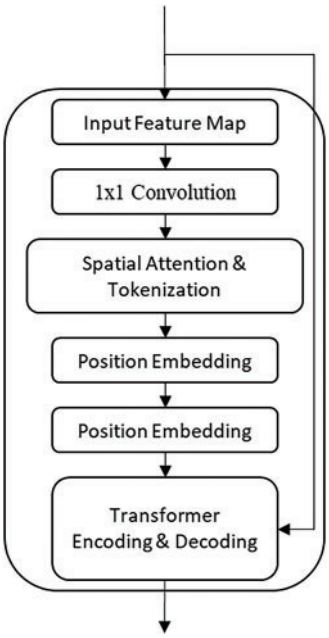

Figure 3: Transformer mechanism for global context extraction in CE-CDNet

The upsampling part of the Decoder module receives input from several key modules, and after fusion and processing, generates the final result of building change detection (as shown in Fig. 4). First, the input

to the Decoder comes from the Channel-Aware Multi-Scale Fusion (CAMSF) module, which adaptively fuses features from multi-temporal images to generate feature maps containing multi-scale building change information, ensuring that the network can accurately capture changes in buildings across different times and scales. Second, feature maps from the Visual Geometry Group (VGG) feature extraction module are also important inputs to the Decoder. These feature maps are extracted through multiple layers of convolution operations, containing both low-level and high-level information, providing rich details for the final change detection. Additionally, the token encoding and decoding modules capture global contextual information of the image through the Transformer structure, ensuring that the network can understand long-range dependencies in the image. By integrating these inputs from the CAMSF, VGG feature extraction, and token encoding and decoding modules, the Decoder effectively fuses multi-scale, multi-level feature information to generate high-precision building change detection results.

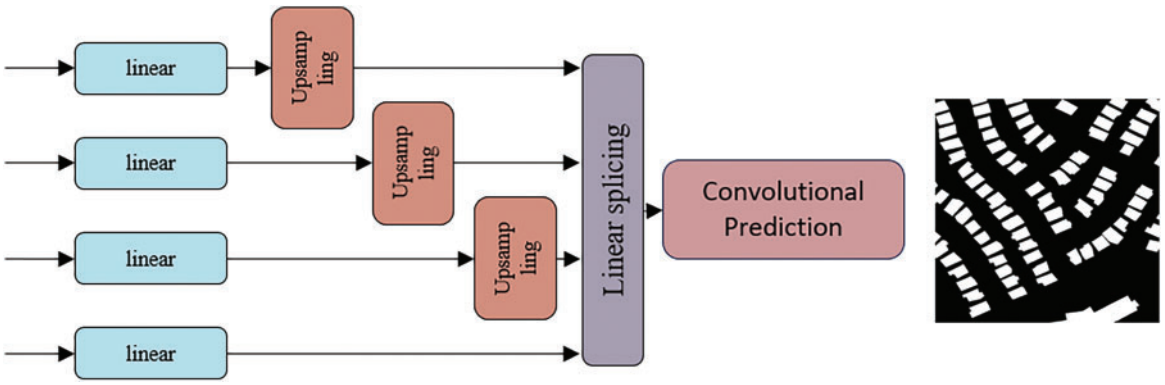

Figure 4: Decoder module and lightweight transformer decoder (LTD) in CE-CDNet

The LTD module receives the fused feature map $\mathbf{F}_{\text{fusion}}$ from the CAMSF module, with dimensions $64 \times 64 \times 64$. First, a convolution operation is applied to $\mathbf{F}_{\text{fusion}}$ to reduce the dimensionality, generating the encoded feature representation:

$$\mathbf{F}_{\text{enc}} = \text{Conv2d}\left(\mathbf{F}_{\text{fusion}}\right) \tag{5}$$

After dimensionality reduction, the LTD module applies a lightweight Transformer self-attention mechanism to capture the global contextual information in the image. The self-attention mechanism linearly maps the input feature map $\mathbf{F}_{\text{enc}}$ to query, key, and value vectors, and calculates the correlations between features using the following formula:

$$\text{Attention}\left(Q, K, V\right) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

This process enables the LTD module to efficiently extract global information, enhancing the model's ability to capture building changes.

After that, the original resolution of the input image is upsampled.

To optimize computational complexity, the LTD module reduces the number of attention heads and feature dimensions through a lightweight Transformer structure. This approach maintains the ability to extract global information while significantly improving computational efficiency. As a result, the LTD can provide high-performance change detection with low computational overhead, even in complex scenarios.

Finally, the feature map output by the LTD module is mapped into a change detection result map, marking the building change areas in the two input multi-temporal remote sensing images.

4 Experiments

4.1 Datasets and Evaluation Metrics

We assessed the performance of the CE-CDNet network for building change detection by conducting experiments on the LEVIR-CD, WHU Building, and SYSU datasets.

1. LEVIR-CD Dataset: A high-resolution remote sensing image change detection dataset specifically designed for change detection. It covers different areas of many cities in China and contains rich information on building changes. Its images cover complex urban environments and have high spatial resolution.

2.WHU Building Dataset: Contains a large number of urban building images taken at different times. The image resolution is up to 0.075 m and contains rich building shape and texture details. Each pair of multi-temporal images is accompanied by accurately annotated building change areas, and the annotation accuracy of the entire dataset is very high. In terms of capturing fine details, this dataset is very suitable for evaluating model performance.

3. SYSU-CD Dataset: The SYSU-CD dataset is a publicly available remote sensing change detection dataset, captured in Hong Kong between 2007 and 2014. The data is split into a training set (12,000 pairs), a validation set (4000 pairs), and a test set (4000 pairs), as provided by the researchers.

By testing and validating the CE-CDNet network using these three datasets, we can effectively assess the model's robustness and generalization capability in BCD tasks.

We used the following evaluation indicators to measure the performance of the model, including:

1. Precision (P): Precision indicates the proportion of correctly predicted changed areas among all predicted positive samples.

$$P = \frac{TP}{TP + FP} \tag{7}$$

2. Recall (R): Recall measures the model's ability to identify true positives, indicating the ratio of correctly detected positive samples to all actual positives.

$$R = \frac{TP}{TP + FN} \tag{8}$$

3. Intersection over Union (IoU): Measures the overlap between predicted and actual regions by dividing their intersection by their union.

$$IoU = \frac{TP}{TP + FP + FN} \tag{9}$$

4. F1-score: The F1-score is the weighted harmonic mean of precision and recall, used to balance the precision and recall of the model. It is particularly suitable for scenarios with imbalanced sample classes.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{10}$$

4.2 Experiment Analysis

In this section, we present and analyze the quantitative results of CE-CDNet on the WHU Building Change, LEVIR-CD, and SYSU-CD Datasets, along with the corresponding metrics. We first evaluate the

model's performance on the WHU Building Change Dataset and the LEVIR-CD Dataset, and conduct a comparative analysis of its segmentation capabilities with other models.

The specific comparative data results on the WHU dataset are shown in Table 1. By comparing the performance of multiple models in the task of building change detection in remote sensing images, CE-CDNet performs well in all four key evaluation indicators, especially in the two indicators of Precision and Recall, surpassing most other models, highlighting its comprehensive detection capabilities. Compared with high-precision models such as IFNet and SNUNet, CE-CDNet significantly improves Recall while maintaining a high accuracy rate, effectively reducing missed detections and ensuring full coverage of the changed area. In addition, CE-CDNet performs stably in complex scenes, and its false alarm rate is significantly lower than that of models such as FC-Siam-Conc and DTCDSCN, further improving its reliability in practical applications. In contrast, although models such as STANet and BIT perform well in some indicators, they are lacking in detail change area detection and IoU, and are not effective in dealing with complex building change scenes.

Table 1: Quantitative comparison of change detection models on the WHU dataset

Model	Precision (%)	Recall (%)	IoU (%)	F1-score (%)
FC-EF [42]	71.63	67.25	53.11	69.37
FC-Siam-Conc [42]	60.88	73.58	49.95	66.63
FC-Siam-Diff [42]	47.33	77.66	41.66	58.51
SNUNet [43]	88.04	87.36	78.09	87.7
BIT [44]	86.64	81.48	72.39	83.98
ChangeFormer [36]	91.83	88.02	81.63	89.88
DTCDSCN [45]	63.92	82.30	56.19	71.95
STANet [46]	79.37	85.50	69.95	82.32
IFNet [32]	96.91	73.19	71.52	83.40
HANet [1]	88.30	88.01	78.82	88.16
CGNet [16]	94.47	90.79	86.21	92.59
CE-CDNet	90.59	94.06	85.69	92.30

In addition, from the performance of the LEVIR-CD dataset in Table 2, the accuracy of CE-CDNet on both datasets (WHU dataset and LEVIR-CD dataset) is quite high, 90.59% and 89.27%, respectively. Although IFNet has a higher accuracy (96.91%) on the WHU dataset, its recall rate is lower (73.19%), which may cause the model to miss important change areas. On the WHU dataset, the recall rate of CE-CDNet is 94.06%, ranking first among all models, much higher than other models (for example, STANet is 85.50% and BIT is 81.48%). On the LEVIR-CD dataset, the recall rate of CE-CDNet is 92.07%, which is also ahead of most models (for example, STANet is 91.00% and IFNet is 82.93%). This shows that CE-CDNet can detect building changes more comprehensively and reduce missed detections.

Table 2: Quantitative comparison of change detection models on the LeViR-CD dataset

Model	Precision (%)	Recall (%)	IoU (%)	F1-score (%)
FC-EF [42]	86.91	80.17	71.53	83.40
FC-Siam-Di [42]	89.53	83.31	75.92	86.31

(Continued)

Table 2 (continued)

Model	Precision (%)	Recall (%)	IoU (%)	F1-score (%)
FC-Siam-Conc [42]	91.99	76.77	71.96	83.69
DTCDSCN [45]	88.35	86.83	78.05	87.67
BIT [44]	89.24	87.37	80.68	89.31
STANet [46]	93.81	91.00	77.40	87.26
IFNet [32]	91.78	82.93	78.77	88.13
SNUNet [43]	89.18	87.17	78.83	88.16
HANet [1]	91.21	89.36	82.27	90.28
ChangeFormer [36]	92.05	88.80	82.48	90.40
CE-CDNet	89.27	92.07	82.90	90.65

The visualization results of comparing WHU-CD and LEVIR-CD are shown in Figs. 5 and 6, our model outperforms other models in both missed detections (green areas) and false detections (red areas). Compared with the DTCDSCN model and the BIT model, our model consistently reduces the green missed detection areas in all test scenarios, indicating that it has a stronger ability to detect building changes in complex urban environments. Specifically, our model is good at accurately identifying changes in areas with rich building details and complex backgrounds (such as trees and roads), effectively reducing missed detections. In addition, when examining the false detection areas, the DTCDSCN and BIT models show higher misclassification rates in some cases, especially in non-building areas such as shadows and roads, causing the red areas to stand out more. In contrast, our model significantly reduces these false detections, maintaining a low false detection rate even in scenes with complex backgrounds or subtle changes, demonstrating its excellent cross-scene generalization ability and detail processing performance. In addition, when analyzing the white correctly detected areas, it can be found that our model shows higher consistency in accurately detecting building changes, more comprehensive coverage, and a higher degree of match with the ground truth labels. Our model shows higher stability and reliability, especially at the edges of buildings and in complex areas. Overall, this shows that our model not only achieves higher accuracy in the building change detection task, but also shows strong robustness and generalization ability in complex backgrounds and multi-scale changing scenes.

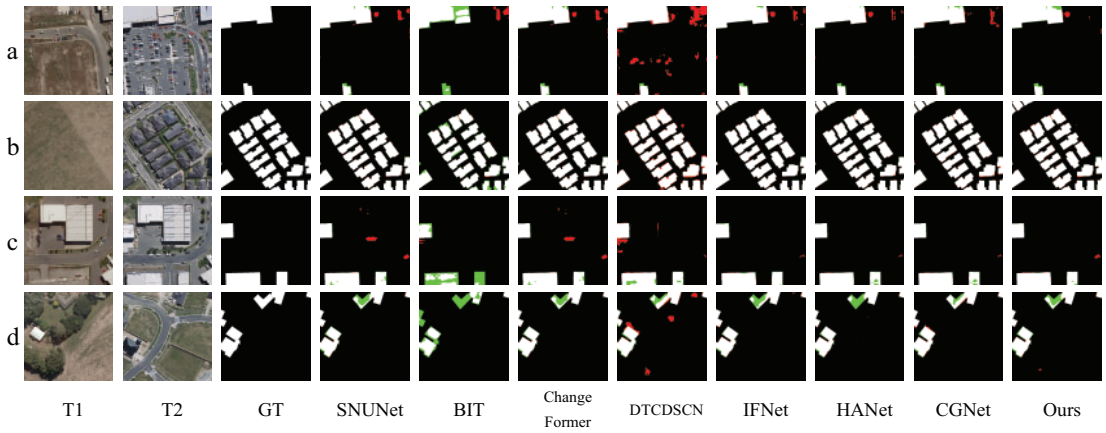

Figure 5: Qualitative experimental results on WHU-CD. TP (white), TN (black), FP (red), and FN (green)

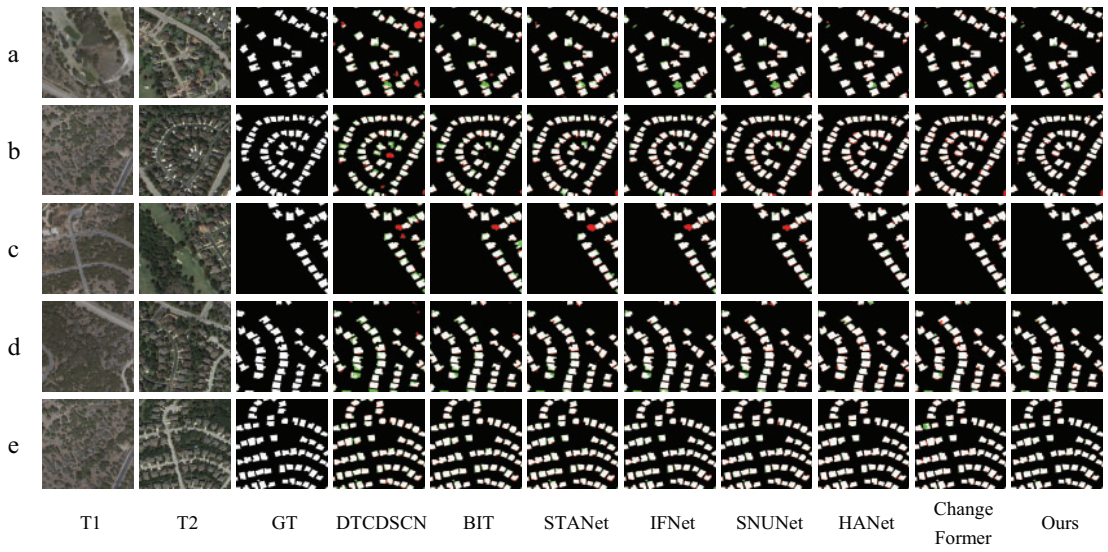

Figure 6: Qualitative experimental results on LEVIR-CD. TP (white), TN (black), FP (red), and FN (green)

For example, in the test image c of Fig. 6, the central areas of T1 and T2 have changed a lot, especially the highlighted part of the suspected building. Our model showed excellent cross-scene detection ability during detection and avoided false detection. Similarly, in Fig. 6b, due to the close or consistent colors of the road and the building, there is a problem of inconsistent building edge division in multi-building scenes. Other models basically have the problem of too large or too small edge division to a large extent, resulting in insufficient accuracy. Compared with other results, our model has certain optimization in this regard.

In order to further assess the performance of CE-CDNet in diverse environments, we conducted experiments on the SYSU dataset.

Table 3 shows the quantitative comparison results of different models on the SYSU dataset. CE-CDNet performs well in various indicators, especially in Recall (76.17%) and IoU (60.77%), which is better than other models. For example, IoU is significantly higher than ChangeFormer (47.73%) and IFNet (46.60%).

Table 3: Quantitative comparison of change detection models on the SYSU dataset

Model	Precision (%)	Recall (%)	IoU (%)	F1-score (%)
Change Former [36]	59.40	70.84	47.73	64.62
FC-conc [42]	81.57	66.69	57.96	73.38
FC-diff [42]	91.27	55.61	52.80	69.11
IFNet [32]	50.56	85.61	46.60	63.57
BIT [44]	75.15	71.58	57.88	73.32
CE-CDNet	75.04	76.17	60.77	75.60

By comparing the results in Fig. 7, it is clear that CE-CDNet outperforms other models in terms of both missed detections and false detections. Specifically, the green missed detection areas in CE-CDNet are minimal, indicating its superior ability to accurately detect building changes even in complex environments. Additionally, the red false detection areas are also relatively small, demonstrating that the model effectively minimizes misclassification in non-change areas. In contrast, other models exhibit higher rates of missed and

false detections in certain scenarios, particularly when handling multi-scale building changes and complex background interference, highlighting some limitations.

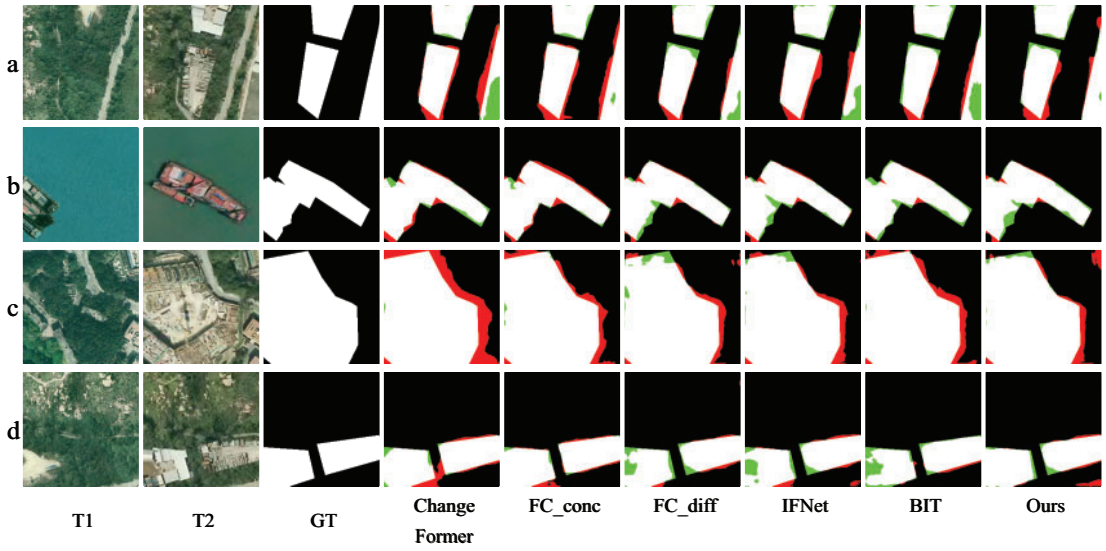

Figure 7: Qualitative experimental results on SYSU. TP (white), TN (black), FP (red), and FN (green)

Overall, CE-CDNet achieves efficient change detection without sacrificing detection accuracy, effectively segmenting building change areas in complex urban environments. This demonstrates its potential as an excellent solution for change detection tasks.

Fig. 8 shows the number of network parameters of several building change detection networks, where the first five networks are CNN-based networks and the others are transformer-based networks. The number of parameters of CE-CDNet is 17.76 M, slightly higher than the CNN-based methods and lower than the parameters of transformer-based methods. From the performance on the two datasets listed in Table 4, CE-CDNet achieves a better balance between parameter overhead and F1.

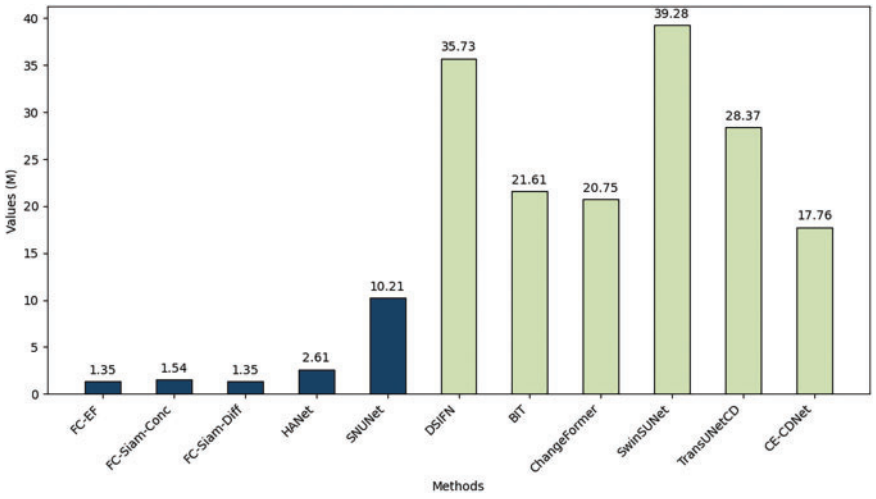

Figure 8: Comparison of methods by parameters count

Table 4: Comparison of parameter count and F1 results on WHU and LEVIR-CD datasets (*Giga Floating-Point Operations Per Second)

Method	Params (M)	GFLOPs*	F1	
			WHU-CD	LEVIR-CD
ChangeFormerV4 [36]	33.61	852.53	87.39	75.87
ChangeFormerV5 [36]	55.27	841.08	87.45	78.23
ChangeFormerV6 [36]	41.03	811.15	83.66	72.71
BIT-101 [44]	43.27	380.62	90.04	82.53
CE-CDNet	17.76	759.23	92.30	90.65

The final comparison results show that CE-CDNet exhibits excellent overall performance in the task. Through the effective integration of channel optimization, adaptive multi-scale feature fusion (CAMSF), and the lightweight Transformer decoder (LTD) modules, CE-CDNet significantly improves detection accuracy and robustness.

4.3 Ablation Studies

To evaluate the impact of each module in the CE-CDNet network on building change detection performance, we designed an ablation study, analyzing the contribution of each module through a comparison of different model versions. The experiments were conducted using the LEVIR-CD building change detection dataset.

Initially, the baseline model demonstrated moderate performance, with an IoU of 0.7869, a Precision of 86.87%, a Recall of 89.32%, and an F1-score of 88.08% (as shown in Table 5). These results indicate that while the model was able to capture local features, it still faced challenges in accurately detecting change regions.

Table 5: Performance comparison of different modules of CE-CDNet on LEVIR-CD and WHU-CD dataset (All values are in %)

	LEVIR-CD				WHU-CD			
Model	Pre.	Rec.	IoU	F1	Pre.	Rec.	IoU	F1
Baseline	86.87	89.32	78.69	88.08	84.44	94.06	80.16	88.99
+CAMSF	87.34	91.98	81.16	89.60	89.09	91.86	82.58	90.46
+LTD	87.76	92.47	81.91	90.05	89.98	91.14	82.74	90.55
CE-CDNet	89.27	92.07	82.90	90.65	90.59	94.06	85.69	92.30

Upon adding the CAMSF module, which adaptively integrates multi-scale features, the model's performance significantly improved. The Recall increased to 91.98%, IoU rose to 0.8116, and the F1-score reached 89.60%. The CAMSF module enabled the model to better handle multi-scale changes in complex scenarios, particularly enhancing its ability to detect change regions. After introducing the lightweight Transformer decoder, the experimental results showed further improvement, with the F1-score rising to 90.05% and IoU increasing to 81.91%, demonstrating that the Transformer effectively enhanced the capture of global features, making the model more robust in complex change detection scenarios.

The full CE-CDNet network, which combines the channel optimization strategy, CAMSF module, and lightweight Transformer decoder, achieved optimal performance, with a Precision of 89.27%, a Recall of 92.07%, an IoU of 82.90%, and an F1-score of 90.65%. The channel optimization strategy maintained high precision while reducing computational costs, allowing the model to perform better in resource-constrained environments.

Overall, the results of the ablation study demonstrate that the CAMSF module improved the model's ability to capture fine details through multi-scale feature fusion, while the lightweight Transformer decoder enhanced the understanding of global information. The combination of these two modules significantly boosted the overall performance of the model.

In Fig. 9c, we can clearly see that our baseline has many false positives and missed positives when predicting the edge of the house change. This directly reflects that the baseline model does not accurately divide the original building edge. After adding CAMSF, the green missed positives in the figure are significantly reduced, indicating that the module has further improved the detection level. After adding LTD alone, the number of false positives (red) pixels in the figure is reduced to a certain extent compared to CAMSF. The complete CE-CDNet further integrates the advantages of these two modules, and the results of the previous tests are improved.

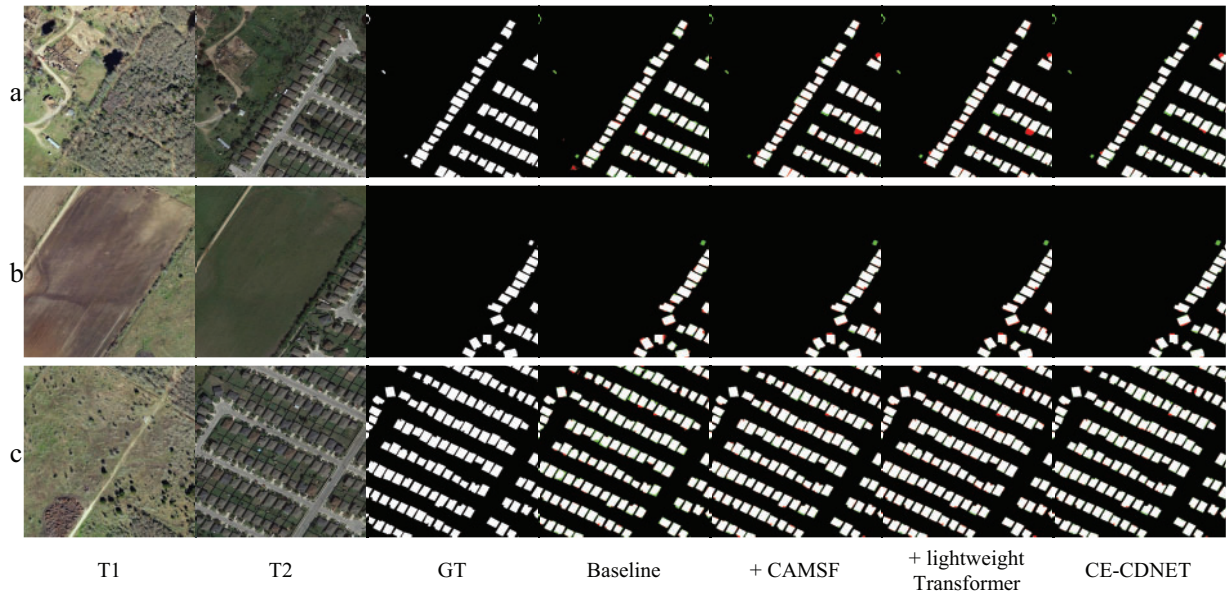

Figure 9: The horizontal displays a, b, and c respectively show the original change graph, label, and corresponding model prediction graph of the dataset. Ablation experiments of different modules of CE-CDNet on the LEVIR-CD dataset. TP (white), TN (black), FP (red), and FN (green)

We selected a case of a small building change. In Fig. 10, there is no building in the area in a, while there is a building with an unconventional roof in the corresponding area in b. In the previous building detection network, this area was not paid any attention. In our network, Figs. 10c,d respectively shows the building detection capabilities of Baseline+CAMSF and Baseline+LTD. After adding these two modules, they both pay attention to this area, especially in d, which basically draws the outline of the building that is easy to ignore. In the complete network in e, it can be seen that our network pays more attention to this area. The building is clearly displayed.

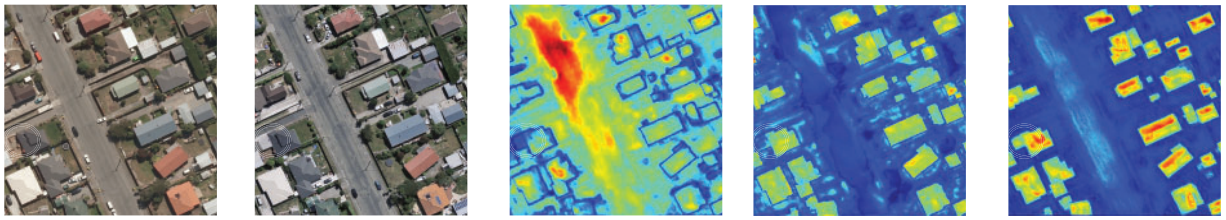

Figure 10: Heatmap of each module in the ablation experiment. a: Original image T1, b: Original image T2, c: Heatmap of Baseline+CAMSF, d: Heatmap of Baseline+LTD, e: Heatmap of CE-CDNet

In conclusion, the comparison shown in the figure clearly illustrates the contribution of each module to the model's performance. The complete CE-CDNet model exhibits significant advantages in handling multi-scale features and complex building change scenarios. In contrast, removing specific modules leads to increased false positives and false negatives, resulting in noticeable performance degradation. This validates the crucial roles of the CAMSF module and the Lightweight Transformer Decoder (LTD) in achieving the model's overall effectiveness in change detection.

5 Conclusions

This paper proposes a channel-optimized change detection network (CE-CDNet) to address the issues of computational cost and detection accuracy in BCD in high-resolution remote sensing images. By incorporating the adaptive multi-scale feature fusion module (CAMSF) and the lightweight Transformer decoder (LTD), CE-CDNet significantly reduces computational costs and resource consumption while maintaining high detection accuracy. Extensive experimental results demonstrate that CE-CDNet performs exceptionally well on multiple public datasets, with notable improvements in reducing both false negative and false positive rates compared to existing mainstream methods. Unlike traditional convolutional neural networks (CNNs), CE-CDNet takes advantage of its Transformer architecture to effectively model long-range dependencies. As a result, it exhibits superior capabilities in change detection and shows greater robustness in complex scenarios.

In summary, CE-CDNet not only excels in detection accuracy and computational efficiency, but also demonstrates excellent generalization ability and application potential. Future research could further optimize the model's lightweight design and explore its application and expansion in other remote sensing tasks.

Acknowledgement: This paper was supported by Henan Province Key R&D Project, Jiangsu Science and Technology Programme, Henan Provincial Science and Technology Research Project and Science and Technology Innovation Project of Zhengzhou University of Light Industry.

Funding Statement: This paper was supported by Henan Province Key R&D Project (241111210400), Henan Provincial Science and Technology Research Project (242102211007 and 242102211020), Jiangsu Science and Technology Programme-General Programme (BK20221260) and Science and Technology Innovation Project of Zhengzhou University of Light Industry (23XNKJTD0205).

Author Contributions: Data Curation: Jia Liu, Hang Gu; Formal Analysis: Hang Gu, Hao Chen, Qidong Liu; Funding Acquisition: Jia Liu, Zuhe Li; Methodology: Jia Liu, Hang Gu, Hao Chen; Project Administration: Jia Liu, Zuhe Li;

Software: Zuhe Li; Supervision: Gang Xu, Wei Wang, Fangmei Liu; Visualization: Zuhe Li, Fangmei Liu; Writing—Original Draft: Hang Gu, Hao Chen; Writing—Review & Editing: Wei Wang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data supporting the findings of this study are publicly available in the WHU Building Dataset at https://gpcv.whu.edu.cn/data/building_dataset.html (accessed on 01 January 2025) and the LEVIR-CD Dataset at https://lesvir.buaa.edu.cn/datasets/index.html (accessed on 01 January 2025) and the SYSU-CD Dataset at https://github.com/liumency/SYSU-CD (accessed on 01 January 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Han C, Wu C, Guo H, Hu M, Chen H. HANet: a hierarchical attention network for change detection with bitemporal very-high-resolution remote sensing images. IEEE J Sel Top Appl Earth Obs Remote Sens. 2023;16:3867–78. doi:10.1109/jstars.2023.3310208.
2. Woodcock CE, Loveland TR, Herold M, Bauer ME. Transitioning from change detection to monitoring with remote sensing: a paradigm shift. Remote Sens Environ. 2020;238:111558. doi:10.1016/j.rse.2019.111558.
3. Zhu Q, Guo X, Li Z, Li D. A review of multi-class change detection for satellite remote sensing imagery. Geo-Spatial Inf Sci. 2024;27:1–15. doi:10.1080/10095020.2022.2128902.
4. Sofina N, Ehlers M. Building change detection using high resolution remotely sensed data and GIS. IEEE J Sel Top Appl Earth Obs Remote Sens. 2016;9:3430–8. doi:10.1109/jstars.2016.2542074.
5. Lv Z, Zhong P, Wang W, You Z, Falco N. Multiscale attention network guided with change gradient image for land cover change detection using remote sensing images. IEEE Geosci Remote Sens Lett. 2023;20:1–5. doi:10.1109/lgrs.2023.3267879.
6. 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. doi:10.1109/tgrs.2024.3360516.
7. Leichtle T, Geiß C, Wurm M, Lakes T. Unsupervised change detection in VHR remote sensing imagery-an object-based clustering approach in a dynamic urban environment. Int J Appl Earth Obs Geoinf. 2017;54:15–27. doi:10.1016/j.jag.2016.08.010.
8. Zhang J, Shao Z, Ding Q, Huang X, Wang Y, Zhou X, et al. AERNet: an attention-guided edge refinement network and a dataset for remote sensing building change detection. IEEE Trans Geosci Remote Sens. 2023;61:5617116. doi:10.1109/tgrs.2023.3300533.
9. Bai T, Wang L, Yin D, Sun K, Chen Y, Li W, et al. Deep learning for change detection in remote sensing: a review. Geo-Spatial Inf Sci. 2023;26:262–88. doi:10.1080/10095020.2022.2085633.
10. Chen H, Li W, Shi Z. Adversarial instance augmentation for building change detection in remote sensing images. IEEE Trans Geosci Remote Sens. 2021;60:1–16. doi:10.1109/TGRS.2021.3066802.
11. Chen T, Lu Z, Yang Y, Zhang Y, Du B, Plaza A. A Siamese network based U-Net for change detection in high resolution remote sensing images. IEEE J Sel Top Appl Earth Obs Remote Sens. 2022;15(1):2357–69. doi:10.1109/JSTARS.2022.3157648.
12. Li Z, Yan C, Sun Y, Xin Q. A densely attentive refinement network for change detection based on very-high-resolution bitemporal remote sensing images. IEEE Trans Geosci Remote Sens. 2022;60:1–18. doi:10.1109/TGRS.2022.3159544.
13. Ji Y, Sun W, Wang Y, Lv Z, Yang G, Zhan Y, et al. Domain adaptive and interactive differential attention network for remote sensing image change detection. IEEE Trans Geosci Remote Sens. 2024;62:5616316. doi:10.1109/TGRS.2024.3382116.
14. Dai X, Xia M, Weng L, Hu K, Lin H, Qian M. Multiscale location attention network for building and water segmentation of remote sensing image. IEEE Trans Geosci Remote Sens. 2023;61:1–19. doi:10.1109/TGRS.2023.3276703.

15. Peng X, Zhong R, Li Z, Li Q. Optical remote sensing image change detection based on attention mechanism and image difference. IEEE Trans Geosci Remote Sens. 2020;59(9):7296–307. doi:10.1109/TGRS.2020.3033009.
16. Han C, Wu C, Guo H, Hu M, Li J, Chen H. Change guiding network: incorporating change prior to guide change detection in remote sensing imagery. IEEE J Sel Top Appl Earth Obs Remote Sens. 2023;16:8395–407. doi:10.1109/jstars.2023.3310208.
17. Zhang C, Wang L, Cheng S, Li Y. SwinSUNet: pure transformer network for remote sensing image change detection. IEEE Trans Geosci Remote Sens. 2022;60:1–13. doi:10.1109/tgrs.2022.3160007.
18. 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;60:1–19. doi:10.1109/tgrs.2022.3169479.
19. Song X, Hua Z, Li J. Remote sensing image change detection transformer network based on dual-feature mixed attention. IEEE Trans Geosci Remote Sens. 2022;60:1–16. doi:10.1109/tgrs.2022.3209972.
20. Nielsen AA. The regularized iteratively reweighted MAD method for change detection in multi-and hyperspectral data. IEEE Trans Image Process. 2007;16:463–78. doi:10.1109/tip.2006.888195.
21. Shang C, Huang B, Yang F, Huang D. Slow feature analysis for monitoring and diagnosis of control performance. J Process Control. 2016;39(2):21–34. doi:10.1016/j.jprocont.2015.12.004.
22. Nagarajaiah S, Varadarajan N. Short time Fourier transform algorithm for wind response control of buildings with variable stiffness TMD. Eng Struct. 2005;27(3):431–41. doi:10.1016/j.engstruct.2004.10.015.
23. Du P, Liu S, Gamba P, Tan K, Xia J. Fusion of difference images for change detection over urban areas. IEEE J Sel Top Appl Earth Obs Remote Sens. 2012;5(4):1076–86. doi:10.1109/JSTARS.2012.2200879.
24. Qin R, Huang X, Gruen A, Schmitt G. Object-based 3-D building change detection on multitemporal stereo images. IEEE J Sel Top Appl Earth Obs Remote Sens. 2015;8(5):2125–37. doi:10.1109/JSTARS.2015.2424275.
25. Tan K, Zhang Y, Wang X, Chen Y. Object-based change detection using multiple classifiers and multi-scale uncertainty analysis. Remote Sens. 2019;11(3):359. doi:10.3390/rs11030359.
26. Yousif O, Ban Y. A novel approach for object-based change image generation using multitemporal high-resolution SAR images. Int J Remote Sens. 2017;38:1765–87. doi:10.1080/01431161.2016.1217442.
27. 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 Geosci Remote Sens Lett. 2021;19:1–5. doi:10.1109/LGRS.2021.3098774.
28. Zhang M, Shi W. A feature difference convolutional neural network-based change detection method. IEEE Trans Geosci Remote Sens. 2020;58:7232–46. doi:10.1109/tgrs.2020.2981051.
29. Shen Q, Huang J, Wang M, Tao S, Yang R, Zhang X. Semantic feature-constrained multitask siamese network for building change detection in high-spatial-resolution remote sensing imagery. ISPRS J Photogramm Remote Sens. 2022;189:78–94. doi:10.1016/j.isprsjprs.2022.05.001.
30. Chen J, Yuan Z, Peng J, Chen L, Huang H, Zhu J, et al. DASNet: dual attentive fully convolutional Siamese networks for change detection in high-resolution satellite images. IEEE J Sel Top Appl Earth Obs Remote Sens. 2020;14:1194–206. doi:10.1109/JSTARS.2020.3037893.
31. Shi Q, Liu M, Li S, Liu X, Wang F, Zhang L. A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection. IEEE Trans Geosci Remote Sens. 2021;60:1–16. doi:10.1109/TGRS.2021.3085870.
32. Zhang C, Yue P, Tapete D, Jiang L, Shangguan B, Huang L, et al. 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. doi:10.1016/j.isprsjprs.2020.06.003.
33. Liang S, Hua Z, Li J. GCN-based multi-scale dual fusion for remote sensing building change detection. Int J Remote Sens. 2023;44:953–80. doi:10.1080/01431161.2023.2173031.
34. Holail S, Saleh T, Xiao X, Li D. Afde-net: building change detection using attention-based feature differential enhancement for satellite imagery. IEEE Geosci Remote Sens Lett. 2023;20:1–5. doi:10.1109/lgrs.2023.3283505.
35. Chen H, Song J, Han C, Xia J, Yokoya N. ChangeMMamba: remote sensing change detection with spatio-temporal state space model. arXiv:2404.03425. 2024. doi:10.1109/tgrs.2024.3417253.

36. Bandara WGC, Patel VM. A transformer-based siamese network for change detection. In: Proceedings of the IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium; 2022; Kuala Lumpur, Malaysia. p. 207–10. doi:10.1109/igarss46834.2022.9883686.
37. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. Transunet: transformers make strong encoders for medical image segmentation. 2021. doi:10.1109/isbi48211.2021.9433886.
38. Liang S, Hua Z, Li J. Hybrid transformer-CNN networks using superpixel segmentation for remote sensing building change detection. Int J Remote Sens. 2023;44:2754–80. doi:10.1080/01431161.2023.2208711.
39. 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;61:1–15. doi:10.1109/tgrs.2023.3245674.
40. Song X, Hua Z, Li J. PSTNet: progressive sampling transformer network for remote sensing image change detection. IEEE J Sel Top Appl Earth Obs Remote Sens. 2022;15:8442–55. doi:10.1109/jstars.2022.3204191.
41. Shi N, Chen K, Zhou G. A divided spatial and temporal context network for remote sensing change detection. IEEE J Sel Top Appl Earth Obs Remote Sens. 2022;15:4897–908. doi:10.1109/jstars.2022.3176858.
42. Daudt RC, 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); 2018; Athens, Greece. p. 4063–7. doi:10.1109/icip.2018.8451652.
43. 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. 2021;19:1–5. doi:10.1109/LGRS.2021.3056416.
44. Chen H, Qi Z, Shi Z. Remote sensing image change detection with transformers. IEEE Trans Geosci Remote Sens. 2021;60:1–14. doi:10.1109/TGRS.2021.3095166.
45. 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 Geosci Remote Sens Lett. 2020;18(5):811–5. doi:10.1109/LGRS.2020.2988032.
46. Chen H, Shi Z. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. Remote Sens. 2020;12(10):1662. doi:10.3390/rs12101662.