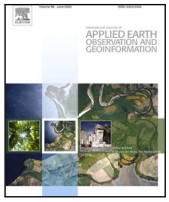
International Journal of Applied Earth Observation and Geoinformation 130 (2024) 103836

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



CDasXORNet: Change detection of buildings from bi-temporal remote sensing images as an XOR problem

Shanxiong Chen [a,b,d], Wenzhong Shi [b,*], Mingting Zhou [c], Min Zhang [b], Yue Yu [b], Yangjie Sun [b], Linjie Guan [a,d], Shuangping Li [a,d]

[a] Changjiang Spatial Information Technology Engineering Co., Ltd, Wuhan 430010, China
[b] Smart Cities Research Institute and Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China
[c] State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[d] Water Resources Information Perception and Big Data Engineering Research Center of Hubei Province, Wuhan 430010, China

ARTICLE INFO

Keywords:
Change detection
Building
Remote sensing
Hierarchical XOR approximation operation
Residual linear attention

ABSTRACT

The up-to-date building information is significant to urban planning and economic assessment. Automatic building change detection (BCD) from bi-temporal remote sensing images is essential for updating building status efficiently. Nevertheless, BCD remains challenging due to the complex building appearance, the diverse imaging conditions, and the building's positional inconsistencies between the bi-temporal images. Recent convolutional neural network-based BCD methods have achieved impressive performance. However, most existing methods employed subtraction or concatenation to identify building changes. Such simple change-deciding operations ignore the spatial–temporal correlation between the bi-temporal features and cannot model the building changes effectively, resulting in overmuch misclassifications. This paper proposes a hierarchical XOR approximating network CDasXORNet to model building changes robustly. An XOR approximation operation is proposed to produce discriminative building differential features from the bi-temporal inputs. We assume that BCD and the logical XOR function have the same nature (i.e., when the two inputs are identical, the output is unchanged/False; otherwise, it is changed/True). This applies to the building change and unaltered pixels simultaneously. Thus, by approximating XOR operation, CDasXORNet can simultaneously exploit the spatial–temporal correlation and the changed and changeless information of buildings. Hierarchical XOR approximation operations are subsequently designed, which process only high-level features to mitigate the influence of substantial irrelevant spectral differences. In addition, the residual linear attention mechanism is introduced to refine the building change features further. Experiments on three publicly challenging datasets demonstrate that our method achieves promising BCD results with fewer commission errors and higher overall performance than the comparative approaches.

1. Introduction

Building change detection (BCD) delivers the newest building status for applications such as digital map updating (Knudsen and Olsen, 2003), urban monitoring (Benedek et al., 2011), and disaster assessment (Chen et al., 2022). High-resolution remote sensing images, with a short revisit period and comprehensive coverage, provide essential data support for detecting building changes. Nevertheless, the manual interpretation of building changes from massive bi-temporal remote sensing images is laborious and time-consuming. Thus, it is of significant application value to investigate automatic BCD from bi-temporal high-resolution remote sensing images to enhance its efficiency (Chen et al., 2022). However, BCD encounters various challenges due to the characteristics of buildings, the complex surrounding environment, and the diverse imaging conditions of bi-temporal images. Buildings with diverse scales, shapes, and construction materials leads to false alarms of building changes and omission errors of small changes. Most buildings are distributed in urban areas, with challenges like shadow occlusion, closely adjacent distribution, or confusion with nearby artificial objects. Moreover, the sensor status, atmospheric conditions, phenological changes in different imaging periods, and the building positional inconsistency caused by the building's height and the satellite side-view imaging mechanism (Shen et al., 2022) further increased the difficulty of BCD. Various BCD algorithms based on bi-temporal remote sensing images have been proposed to address these challenges.

BCD approaches can be sorted into traditional and deep learning (DL)-based methods (Bai et al., 2021). The significant difference between them lies in the feature extraction manner and change-deci-

* Corresponding author.
E-mail address: john.wz.shi@polyu.edu.hk (W. Shi).

https://doi.org/10.1016/j.jag.2024.103836
Received 20 December 2023; Received in revised form 24 March 2024; Accepted 12 April 2024
Available online 6 May 2024

ding strategy. Traditional methods extract spectral, shape, texture (Janalipour and Taleai, 2017), context, and spatial features (Huang et al., 2013) of buildings based on expert knowledge. The extracted features were combined with algebraic, transformation, classification-based, and object-oriented analysis to identify building changes. Traditional pixel-based BCD methods achieved promising results on the test dataset. Nevertheless, these methods suffer from salt-and-pepper noise because they cannot eliminate the spectral and spatial interference of adjacent pixels (Zhang et al., 2021), and they are vulnerable to radiometric correction and geometric registration (Hussain et al., 2013). Object-oriented BCD methods discover changes by comparing spectral attributes, geometric properties, or high-level semantic features retrieved from objects (Xiao et al., 2017; Zhang et al., 2017), which reduces salt and pepper noise and mitigates the sensitivity to registration accuracy. Traditional BCD methods achieved acceptable results under certain conditions. Nevertheless, the handcrafted features have limited representativeness. Moreover, the conventional methods cannot effectively model the change of complex scenes, resulting in overmuch misclassification in the detected building changes. Unlike handcrafted feature-based methods, DL-based BCD approaches automatically learn high-dimensional hierarchical feature representations and have achieved performance far exceeding traditional algorithms.

DL-based methods have been a current mainstream to detect building changes from multi-temporal remote sensing images. The DL-based BCD networks are mostly single-stream and double-stream structures (Shi et al., 2020). The single-stream method fuses the bi-temporal input before inputting them into the model, while the double-stream method processes the input image pair with a Siamese network (Daudt et al., 2018). Owing to the effective feature extraction capability of the Siamese network from image pairs collected at different times, the deep Siamese network is the most popular change detection architecture (Shi et al., 2020). Despite rapid advances in DL-based algorithms to detect building changes, high-quality automatic BCD remains challenging. By introducing attention mechanisms (Chen and Shi, 2020; Liu and Shi, 2021; Ding et al., 2021; Liu et al., 2021; Wang et al., 2022; Eftekhari et al., 2023), preserving high-frequency details (Bai et al., 2021; Zheng et al., 2022; Song et al., 2024), and addressing insufficient or unbalanced sample issues (Chen et al., 2021a; Peng et al., 2020; Shu et al., 2022; Cao and Huang, 2023), various studies have improved the representative capability of extracted features to ameliorate the quality of detected building changes. Recently, transformer-based methods (Vaswani et al., 2017; Dosovitskiy et al., 2020) have been introduced in remote sensing change detection tasks (Chen et al., 2021b) due to their excellent long-range feature-capturing capabilities. Unlike conventional CNNs, transformer-based methods divide the input image into patches to obtain the global receptive field. These methods greatly enhanced the BCD models' capability to extract discriminative features. However, they only apply simple operations such as subtraction or concatenation in change-deciding modules, which lacks mining the spatial–temporal relationship between the bi-temporal features. To obtain spatial–temporal contextual features, hybrid change deciding modules have been proposed by integrating subtraction, concatenation, addition, convolution, attention, transformer, and feature-interaction to model changes of interest (Chang et al., 2023; Song et al., 2023; Chen et al., 2023a). These methods achieved competitive results but with high computational complexity and low efficiency. Moreover, previous BCD methods focus on the change information while ignoring the changeless information, leading to sensitivity to the substantial spectral differences in changeless regions and the misplacement of buildings in bi-temporal images.

A framework integrating XOR approximation operation, hierarchical processing, and residual linear attention (ResLA) mechanism could address the abovementioned issues. In our proposed CDasXORNet, the XOR approximation operation exploits the spatial–temporal relationship between the bi-temporal features and simultaneously focuses on the change and unchanged information for change-deciding. The proposed change-deciding module associates BCD with a logical XOR function. In the BCD process, when the two inputs are the same, the output is unchanged (False); otherwise, it is changed (True). Thus, BCD is equivalent to a "logical XOR" problem. Changeless detection is identical to the "logical AND" issue too. The XOR function can be split into logical AND, OR, and NOT operations. Therefore, by approximating the XOR function, the proposed change-deciding module simultaneously focuses on the change and the unaltered information to identify building changes. The XOR function returns False when two inputs are False or True and returns True when the inputs differ. The input the XOR operation requires is binary False and True, that is, unchanged and changed. However, it is unknown beforehand whether the corresponding pixels in the bi-temporal remote sensing images are identical or different. Therefore, we do not directly apply an XOR operation to achieve BCD. Instead, we model building changes using the XOR approximation operation between bi-temporal hierarchical features extracted by the Siamese convolutional neural network. The hierarchical XOR approximation operation was subsequently designed for robust building change feature capture that overcomes false alarms caused by irrelevant spectral variance and building position misplacement between bi-temporal images. In addition, the ResLA mechanism was introduced to optimize the extracted change features further. These three improvements enable CDasXORNet to fully exploit the relationship between the bi-temporal features and simultaneously focus on the building change and unchanged information to reduce false alarms and achieve high-quality BCD results. The main contributions of our study are as follows.

(1) A novel change-deciding module is created to capture robust change features, enhance the use of spatial–temporal information, and overcome the neglect of changeless information,
(2) A hierarchical XOR approximation operation only on high-level bi-temporal features is designed to reduce false detection caused by irrelevant spectral differences and building position inconsistencies between bi-temporal images, and
(3) A residual linear attention module is introduced to refine the change features further, and competitive results were obtained on two challenging building change detection datasets.

2. Related work

Building change detection (BCD) aims to monitor the dynamic changes in buildings, such as newly built, reconstructed, and demolished buildings, by analyzing multitemporal remote sensing data covering the same region. Various BCD methods have been proposed based on multisource and multitemporal remote sensing data in the past decades. The key to accurately identify building changes from images is to calculate building change features to enhance the discriminability between changed and unchanged areas. According to the building change feature extraction manner, BCD methods can be sorted into handcrafted feature-based traditional approaches and deep learning-based methods. In this section, we focus on reviewing the deep learning BCD network related to this article, as deep learning has become a mainstream method in the field of remote sensing image information extraction since 2017.

Various DL-based methods have been proposed to improve BCD results' quality by enhancing feature extraction capabilities. By creating multiscale feature extraction modules and multilevel feature fusion modules, some studies have improved the feature representation of neural networks to improve the quality of CD results. For example, Wang et al. (2022) proposed to employ a multi-resolution parallel structure and ASPP module to reduce spatial information loss and tackle the diverse scale features of buildings. Bai et al. (2021) suggested an EGRCNN model to extract differential features and integrated edge structure information to identify building changes accurately. To accurately model contextual features, attention mechanisms were

introduced to the BCD model. Chen and Shi (2020) proposed the spatiotemporal attention neural network (STANet). STANet can capture the spatiotemporal dependencies of various scales by segmenting the image into multiscale subregions and introducing self-attention in each subregion. Thus, superior feature representations are produced for adapting to multiscale objects. Attention modules based on spatial (Ding et al., 2021) and channel attention, self-attention mechanisms (Chen and Shi, 2020), and their variants (Jiang et al., 2020) have greatly enhanced the discriminative ability of the extracted features and improved the quality of CD results. Supervised learning-based BCD methods require a large amount of labeled data, which is difficult and expensive to obtain in practical applications. Chen et al. (2021a) utilized generative adversarial training to generate change instances in unchanged regions. It blends synthetic building instances into appropriate locations in one of the bi-temporal images to reduce the model's demand for samples. Peng et al. (2020) proposed a semisupervised change detection network (SemiCDNet) based on FCN and GANs to mine the latent information of unlabeled data to alleviate the insufficient training sample issue. Recently, transformer-based methods were proposed to improve the long-range feature modeling ability, and superior performance was achieved. Liu et al. (2023) have introduced AMTNet, which leverages the benefits of deep convolution, multi-scale modules, attention mechanisms, and transformers to extract distinctive features, thereby achieving superior change detection performance. By creating multiscale feature extraction modules, multilevel feature fusion modules, introducing attention mechanisms, and transformers, various studies have improved the feature extraction capability of neural networks to enhance the quality of CD results. However, change detection involves images from two periods. How to construct the relationship between the images/features of the two periods to discover changes of interest and suppress irrelevant changes still requires further research.

Given the robust feature extraction abilities of DL-based CD methods, the design of the change-deciding strategies is the key to further enhancing the performance of the CD algorithm. Traditional change-deciding strategies include image difference, image ratio, change vector analysis (CVA), and various distance (e.g., cosine distance, Euclidean distance, etc.). In DL-based methods, the most commonly used change-deciding strategy are subtraction and concatenation. The concatenation operation retains bi-temporal features but lacks the change information, while the subtraction operation pays closer attention to the changed regions but ignores the prior knowledge of bi-temporal features (Lu et al., 2024). These simple operations alone cannot model building changes effectively. Hence, various hybrid change-deciding strategies have been proposed by integrating subtraction, concatenation, addition, convolution, attention mechanisms, transformers and feature-interaction to model changes of interest (Chen et al., 2022; Chang et al., 2023; Liu et al., 2023; Zhu et al., 2023; Chen et al., 2023b). Chen et al. (2022) introduced a feature differential enhancement module to improve the global and local information within the differential map, resulting in high-integrity building change regions. Chang et al. (2023) combined CNN and transformer for change deciding and refining the differential feature three times to reduce pseudo-changes. Chen et al. (2023b) adopt feature interaction in bi-temporal images to enhance the model's ability to detect edge and small targets. However, introducing complex structures will increase the CD model's computational complexity. Furthermore, these operations pay closer attention to the changed regions but overlook the constraints of changeless information. The changeless region can assist in eliminating false detections, especially in building change detection. For example, the building positional inconsistency caused by the building's height and the satellite side-view imaging mechanism (Shen et al., 2022) will easily cause false alarms in detected building changes. However, the unchanged part accounts for most of the tested area and the changeless information can constrain obvious false detections and thus improve the quality of detected changes. Thus, it is crucial to enhance the capability of change-deciding modules by utilizing change and changeless information simultaneously.

3. Methodology

A CDasXORNet that considers BCD from bi-temporal remote sensing images as an XOR problem is proposed to model building changes precisely, thus reducing the misclassification caused by substantial irrelevant spectral differences and building rooftop misplacement between the bi-temporal images. The proposed CDasXORNet is shown in Fig. 1. Inputting the co-registered bi-temporal images, CDasXORNet generates a building change probability map that can be binarized to indicate changed and unchanged building areas. CDasXORNet consists of a Siamese encoder, a hierarchical XOR approximation operation, and a decoder. Based on representative features extracted by the Siamese encoder, our framework leverages building change and unaltered information between bi-temporal features by using the hierarchical XOR approximation operation for change deciding. The decoder produces an output the identical size as the initial input image. The Siamese encoder, hierarchical XOR approximation operation, and ResLA decoder are depicted in Sections 3.1 to 3.3.

3.1. Siamese encoder

The Siamese encoder refers to an encoder with two sub-networks with the same architecture and shared weights. The feature extraction module of a modified ResNet34 (He et al., 2016) is selected as the sub-network of the Siamese encoder to extract representative features. To reduce spatial detail loss, the max-pooling layer before the second stage of the original ResNet34 was removed. Thus, the input will only be down-sampled four times. As shown in the area marked with a pink rectangle in Fig. 1, the input image first passed through a convolution block with a kernel size of seven and a stride of two to obtain the first-stage feature F^1. The following four ResNet stages generated F^2, F^3, F^4, and F^5. The five-stage outputs had sizes that were successively $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, and $\frac{1}{16}$ of the input, with $[64, 64, 128, 256, 512]$ channels. The Siamese encoder extracted low- to high-level hierarchical features from input images. The extracted bi-temporal features were delivered to the hierarchical XOR approximation operation for change modeling.

3.2. Hierarchical XOR approximation operation

The hierarchical XOR approximation operation consists of four XOR approximating operations between different stages features. The XOR approximating operation takes the features of the identical Siamese encoder stage to model building changes. Approximating the XOR function can be decomposed into approximating logical AND, OR, and NOT operations since the logical XOR operation can be decomposed into a combination of these three operations, as shown in Eq. (1).

$$A \oplus B = (A \wedge \neg B) \vee (\neg A \wedge B) \tag{1}$$

where $\oplus$ represents logical XOR, $\wedge$ represents logical AND, $\neg$ represents logical NOT, and $\vee$ represents logical OR.

Inspired by Wu et al. (2019), the logical AND operation can be approximated by feature-level multiplication, as shown in Eq. (2). The logical OR operation is non-differentiable and cannot be implemented directly at the feature level. Nevertheless, alternative strategies are used to integrate bi-temporal Siamese features by concatenation operation (Wu et al., 2019), as shown in Eq. (3). Considering that the XOR function is nonlinear, a 1×1 convolution layer combined with the BN and ReLU layers is utilized to enhance the nonlinear expression capability of extracted features.

$$F_1^i \wedge F_2^i = Conv(F_1^i \otimes F_2^i), i = 2, 3, 4, 5 \tag{2}$$

$$F_1^i \vee F_2^i = Conv(Cat(F_1^i, F_2^i)), i = 2, 3, 4, 5 \tag{3}$$

where F_1^i represents the feature of the first period at stage i, F_2^i represents the feature of the second period at stage i, $Conv$ represents

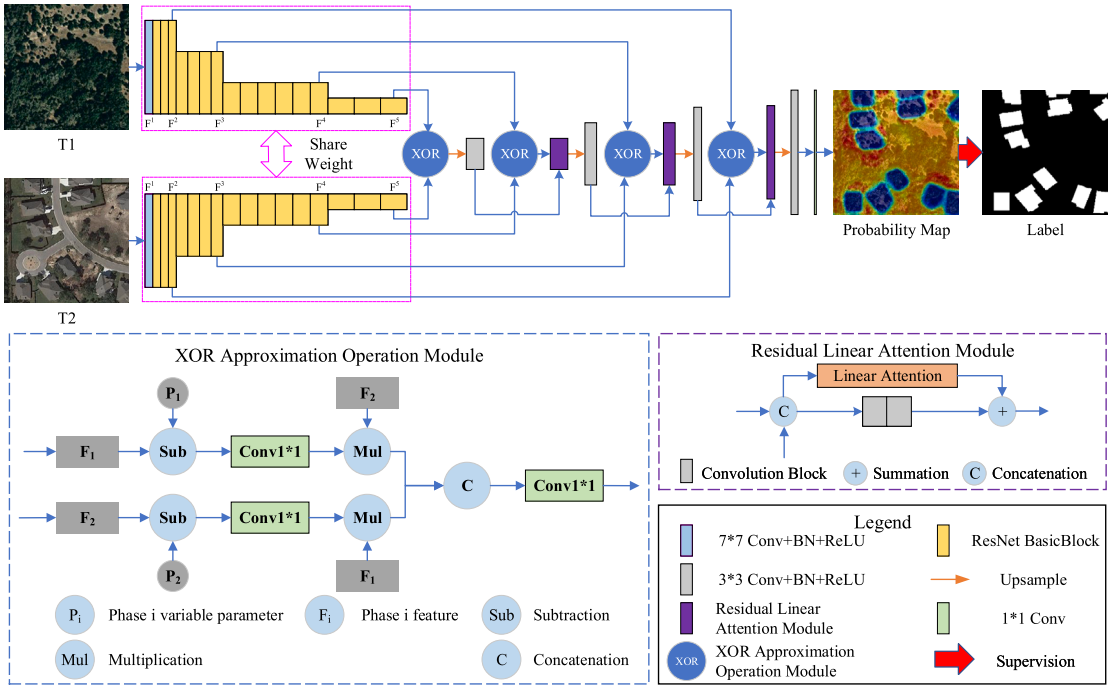


Fig. 1. Overall framework of CDasXORNet. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the 1 × 1 convolution block, ⊗ represents the element-wise product, and Cat represents the concatenation operation.

The next step is approximating the logical NOT operation. To enhance the generalization capability of our framework, we introduce variable parameters and feature subtraction to approximate the logical NOT operation, as shown in Eq. (4). Variable parameters are initialized to 1 and are adaptively adjusted according to the feedback of the loss function during model training.

$$\neg F_1^i = P_1^i - F_1^i, i = 2,3,4,5 \tag{4}$$

where P_1^i represents the variable parameters applied to the i stage of the first phase.

The hierarchical XOR approximation operation was devised based on the XOR approximating operation. As the shallow features extracted in the Siamese encoder's first stage will have undesirable noises, only the bi-temporal features of the last four stages were employed for the hierarchical XOR approximation operation. Combine Eqs. (1) to (4), and the formula for approximating the XOR function is:

$$F_1^i \oplus F_2^i = Conv(Cat(Conv((P_1^i - F_1^i) \otimes F_2^i), Conv((P_2^i - F_2^i) \otimes F_1^i))), i = 2,3,4,5 \tag{5}$$

Eight variable parameters were introduced in Eq. (5) to approximate the logical NOT operation for the bi-temporal features of the second to fifth stages. CDasXORNet performs four XOR approximation operations to model both change and changeless information. The hierarchical approximating XOR operations output four-stage hierarchical change features. The change features were delivered to the ResLA decoder for further processing.

3.3. Residual linear attention decoder

The ResLA decoder further enhances the discriminability of the extracted building change features. As shown in the upper center of Fig. 1, the ResLA decoder includes four up-sampling operations, three ResLA modules, four 3 × 3 convolution blocks, and one 1 × 1 convolution operation. Given the hierarchical change features, the ResLA decoder optimizes their discriminative through the ResLA module. The ResLA module integrated the linear attention (LA) module (Li et al., 2021) with the residual learning technique to optimize the features efficiently and steadily. Li et al. (2021) proposed the LA module to increase the computational efficiency of dot product attention and achieve similar performance. The LA module and its submodules are shown in Fig. 2, in which subfigure (a) is the LA module, (b) is the dot product channel attention, and (c) is the linear spatial attention. As shown in Fig. 2(a), the LA module first applied a convolution block to handle the inputted features. The feature is then processed in parallel using a dot product channel attention and a linear spatial attention module, followed by a dropout layer and a convolution block. The output of the two sub-attention modules is added and handled by a dropout layer and a convolution block again to generate the final output features. The computational complexity of the linear spatial attention module is reduced from $O(N^2)$ to the $O(N)$ level by removing the Softmax operation and applying the associative matrix multiplication law. The residual connection can stabilize the forward and backward information propagation in deep networks via identity mapping (He et al., 2016). Therefore, we introduce the ResLA module by combining the residual connection and the LA module.

As shown in the lower right corner of Fig. 1, the ResLA module initially concatenated the inputted two-stage XOR features. The concatenated features were processed using two branches. The first branch performed feature re-correction on the concatenated features using the LA module, and the other branch applied two consecutive 3 × 3 convolution block operations to the concatenated features. The decoder feature at this stage was created by adding the feature maps produced by the two branches pixel by pixel. The resulting decoder features were up-sampled and processed with a 3 × 3 convolution. The procedure was repeated with the higher-resolution XOR feature until the feature map was restored to the original image size.

3.4. Loss function

To supervise the outputted building change probability map, we use binary cross-entropy (BCE) and add dice loss as loss functions to

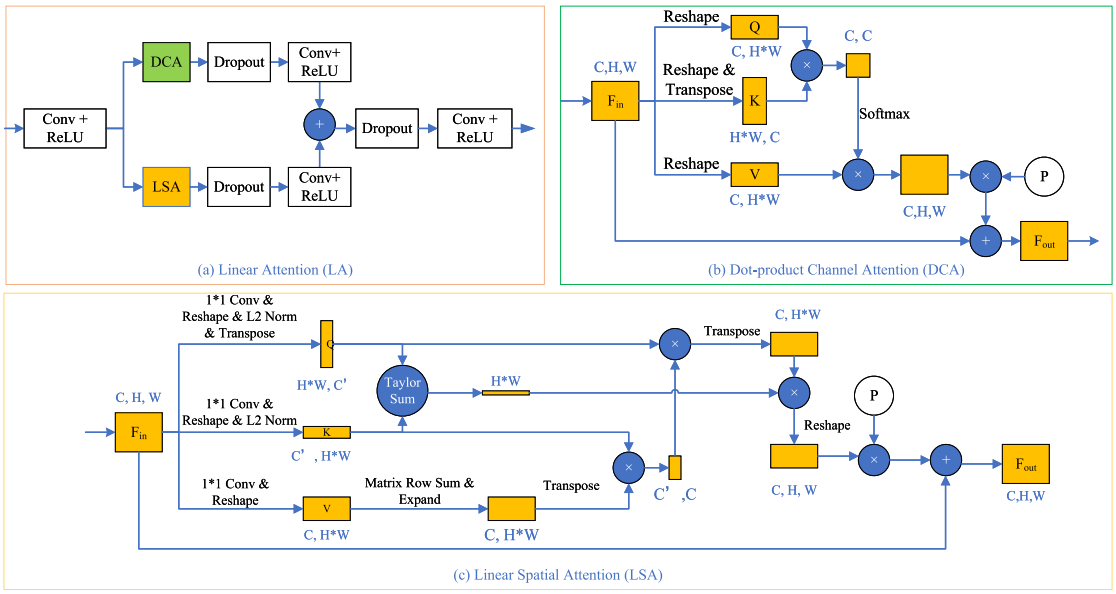


Fig. 2. Schematic diagram of the linear attention module, dot product channel attention, and linear spatial attention.

mitigate the sample imbalance problems (Peng et al., 2019).

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[g_i \log(p_i) + (1-g_i) log(1-p_i)]$$
$$L_{Dice} = \frac{1}{N}(1 - \frac{2\sum_{i=1}^{N} p_i g_i + eps}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i + eps}) \quad (6)$$
$$L_{total} = L_{BCE} + L_{Dice}$$

where N represents the number of samples, g_i represents the sample i's ground-truth value; p_i represents the likelihood that sample i belongs to the positive class, and eps is a constant value 1e-6 to prevent division by zero.

4. Experimental setup

4.1. Dataset

Experiments were conducted on the LEVIR-CD (Chen and Shi, 2020), WHU-BCD (Ji et al., 2018), and NJDS (Shen et al., 2022) datasets to verify the effectiveness of CDasXORNet. LEVIR-CD contains 637 pairs of very high-resolution (0.5 m) images with a 1024 × 1024-pixel size. The image pairs spanned from 5 to 14 years and covered twenty different regions in multiple cities in Texas, USA. The LEVIR-CD dataset introduces the perturbation of irrelevant changes caused by seasonal changes, illumination changes, etc. We follow the original division of the provider (Chen and Shi, 2020) and split each large image into 16 blocks without overlapping. Finally, 7120/1024/2048 image pairs sized 256 × 256 pixels are produced for training/validation/testing.

The WHU-BCD dataset covers an area where a magnitude 6.3 earthquake occurred in February 2011 and was reconstructed in the following years. The bi-temporal images were captured in 2012 and 2016. Fig. 3 shows the bi-temporal images and building change labels covering the entire study area. Several newly built and reconstructed buildings were observed. As shown in the first column, we roughly divide the dataset equally into training and test sets according to the image rows. The training set is further randomly divided into training and validation sets at a ratio of 4:1. By cropping the large image into 256 × 256 size tiles without overlapping, 3016 training samples, 764 validation samples, and 3528 test samples are finally generated. Notably, the training set only contains a few building changes that often occur in the test set (i.e., large-scale building changes and changes from bare ground to bright roofs). Hence, it poses an extraordinary challenge to the generalization ability of the tested algorithm.

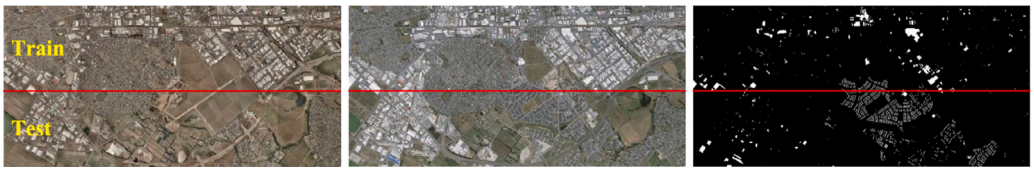
The NJDS dataset comprises very high-resolution (0.3 m) bi-temporal images taken in 2014 and 2018. The dataset incorporates the influence of irrelevant changes caused by building height displacement. We split the large image into image pairs sized at 256 × 256 pixels without overlapping. Finally, 470/233/1717 image pairs are produced for training/validation/testing.

4.2. Implementation details

FC-EF (Daudt et al., 2018), FC-Siam-Conc (Daudt et al., 2018), FC-Siam-Diff (Daudt et al., 2018), UNet++_MSOF (Peng et al., 2019), STANet (Chen and Shi, 2020), DSAMNet (Liu and Shi, 2021), LGP-Net (Liu et al., 2021), EGRCNN (Bai et al., 2021), BIT-CD (Chen et al., 2021b), CTFINet (Feng et al., 2023), and IFTSDNet (Wang et al., 2023) were chosen for comparison. These were general change detection algorithms and BCD methods published in the last six years and have open-source codes. Our code will be released at GitHub (link: https://github.com/MrChen18/CDasXORNet)

CDasXORNet is implemented using the PyTorch framework and trained with a single 3090 GPU. For the fairness of the model parameter numbers, another version of CDasXORNet termed CDasXORNet-ResNet18 is implemented. CDasXORNet-ResNet18 utilizes the modified ResNet18 as the Siamese encoder. The training of CDasXORNet/CDasXORNet-ResNet18 used the AdamW optimizer. We initialized the learning rate to 1e−4, the weight decay is 1e−6, and other parameters default. The batch size for all experiments was set to 32 if memory permitted. The input data are preprocessed by mean–variance normalization. Data augmentation is performed on the training set with random rotations and horizontal and vertical flips to prevent overfitting. The weights of the Siamese encoder are initialized using ResNet34/ResNet18 pre-trained on ImageNet. For the WHU-BCD and NJDS datasets, 150 epochs were trained, and 120 epochs were trained for the LEVIR-CD dataset.

In the experiments, the training hyperparameters of the comparative methods are as consistent as possible with the proposed CDasXORNet. The model parameter settings are as similar as possible to their original paper. Notably, the parameter numbers of FC-EF, FC-Siam-Conc, and FC-Siam-Diff are tripled to match the comparative methods. Due to memory constraints, the batch size is configured to 8 for STANet and



(a) T1 (b) T2 (c) Label

Fig. 3. Overview of the WHU-BCD dataset.

16 for LGPNet. The model is trained with its original loss if a released loss function is observed. Otherwise, it is trained with the proposed loss function. We employ precision, recall, intersection over union (IoU), F1-score, and overall accuracy (OA) as accuracy evaluation indices, as shown in Eqs. (7)–(11).

$$Precision = \frac{TP}{(TP + FP)} \tag{7}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{8}$$

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

$$F1\text{-}score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \tag{10}$$

$$OA = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{11}$$

where TP represents True Positive, FP represents False Positive, FN represents False Negative, and TN represents True Negative.

5. Experimental analysis

5.1. Comparative analysis

5.1.1. Experiments on the LEVIR-CD dataset

Irrelevant spectral variations, multiscale building changes and building positional inconsistencies are typical challenges for BCD from bi-temporal remote sensing images. Hence, we selected three representative image pairs that shows the results of these challenges for visual comparison, as shown in Fig. 4. We stitched 12 test image tiles sized $256 * 256$ pixels to the original image size of the LEVIR-CD dataset to comprehensively compare the tested methods and chose $512 * 512$ size blocks in each scene for more explicit comparison. Pixels colored yellow, green, and red indicate correctly detected, missed-detected, and falsely-detected building changes.

It can be seen from the first, third, and fifth rows of Fig. 4 that CDasXORNet detected the highest quality building change results because its visual accuracy evaluation results mainly consist of yellow pixels, with only a few green and red pixels. As shown in the first and second row of Fig. 4, EGRCNN, CTFINet, LGPNet, and UNet++_MSOF mistakenly recognize the change from bare ground to artificial excavation as the building changes. CDasXORNet accurately identified no building change occurring in this region. This indicates that the suggested approach models building change information and mitigate the impact of unrelated significant spectral differences between bi-temporal images effectively. Examples of densely distributed multi-scale building changes are shown in the third and fourth row. From the fourth row, FC-Siam-Conc and UNet++_MSOF wrongly recognize building changes (i.e., from bare ground to buildings) as changeless due to the large spectral variations between other unchanged buildings. CDasXORNet, CTFINet, and LGPNet achieved better building change detection results, but issues still abound for nearby buildings clinging together. The fifth and last row shows building change detection results with misalignment of buildings caused by differences in imaging angles. Most of the tested comparative methods generated falsely alarmed building changes near building boundaries due to building positional misplacement. The red pixels are few in the qualitative results of EGRCNN and LGPNet. However, LGPNet identified more unchanged buildings as changes than the results of CDasXORNet. The building change area detected by EGRCNN was incomplete. Even considering feature interaction in the change detection model, CTFINet and IFTSDNet still produced some false detections. CDasXORNet achieved the highest quality results, demonstrating that CDasXORNet better overcomes the challenge of buildings caused by differences in imaging angles than the other tested algorithms.

The qualitative analysis of three representative images in the LEVIR-CD dataset illustrates that CDasXORNet effectively copes with challenges posed by irrelevant spectral variations and inconsistent positions between bi-temporal images, which could otherwise result in false detections. The proposed method also performs better in addressing challenges such as multi-scale changes, densely distributed buildings, and confusing backgrounds, which will lead to omission errors. The proposed CDasXORNet has produced better results with fewer false alarms and higher overall performance than the comparative methods. To quantitatively evaluate the accuracy of the detected building changes, the five evaluation indices of 2048 test images were calculated and shown in Table 1. The highest value per column is bolded, and the secondary values are underlined.

CDasXORNet achieved the highest value of the other four indices except for the precision index, and the precision ranked third among the nine tested methods. CDasXORNet-ResNet18 outperformed the comparative approaches in most evaluation metrics, demonstrating that the proposed framework achieves the highest-quality building change results. Among the comparative methods, LGPNet had the highest precision, IoU, F1-score, and OA value. The IFTSDNet achieved the highest recall. The former employs a local–global pyramid network and weights pre-trained on a building extraction dataset for cross-task BCD. The latter captures long-range spatial–temporal relationships among the bi-temporal semantic features to enhance the quality of detected building changes. The IoU of CDasXORNet is 1.56% higher than LGPNet, demonstrating that CDasXORNet produces superior BCD results.

Among the general change detection algorithms, FC-Siam-Diff had an IoU of 80.7% on the LEVIR-CD dataset. FC-Siam-Diff based on Siamese network and simple differential operation modeling changes outperformed other general change detection methods. However, CDasXORNet-ResNet18 outscored FC-Siam-Diff by 3.81% in terms of the IoU, indicating that the proposed XOR approximation operation has a superior capacity to predict changes compared with the simple differential operation. Other classic general change detection methods only achieved IoU values below 80%, demonstrating CDasXORNet's superior suitability for BCD tasks over these approaches.

5.1.2. Experiments on the WHU-BCD dataset

Experiments were conducted on the WHU-BCD dataset to verify further the generalization capability of CDasXORNet. Qualitative evaluation results on 3528 test samples were mosaiced to comprehensively illustrate the detected building changes, as shown in Fig. 5. The newly built buildings were the primary change type in the WHU-BCD test dataset, as observed from the difference between images T1 and T2. The images have a 0.2-meter spatial resolution. Only a few pixels are colored green and red in subfigure (d), demonstrating that CDasXORNet

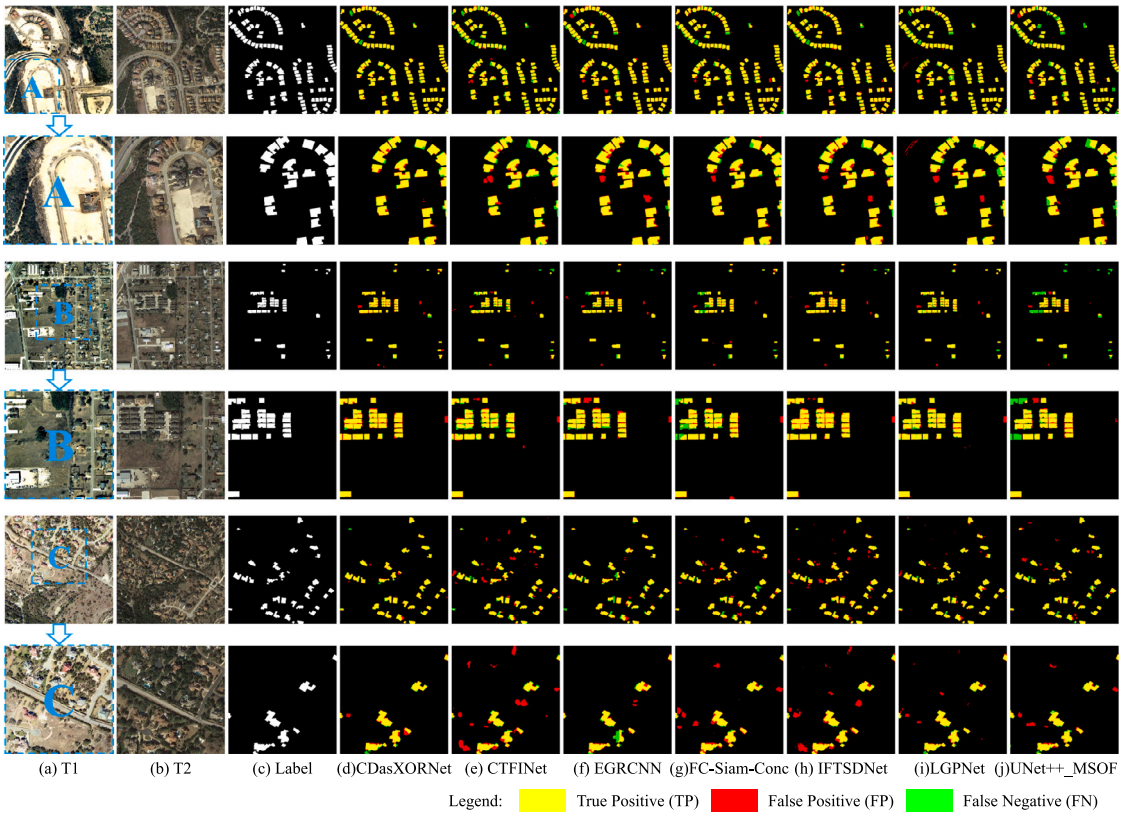


Fig. 4. Qualitative evaluation results on three representative image patches of the LEVIR-CD dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Quantitative evaluation results of comparative experiments on the LEVIR-CD test dataset (%).

Method	Recall	Precision	IoU	F1-score	OA
UNet++_MSOF	86.65	90.98	79.80	88.76	98.88
STANet	89.44	87.23	79.09	88.32	98.80
FC-EF	87.07	88.67	78.36	87.86	98.77
FC-Siam-Conc	89.02	87.89	79.29	88.45	98.82
FC-Siam-Diff	87.23	91.51	80.70	89.32	98.94
DSAMNet	88.97	80.86	73.50	84.73	98.37
LGPNet	88.60	**93.35**	83.34	90.91	99.10
EGRCNN	90.17	87.49	79.87	88.81	98.84
BIT-CD	89.56	90.33	81.72	89.94	98.98
CTFINet	86.30	89.53	78.39	87.89	98.79
IFTSDNet	90.56	88.86	81.33	89.70	98.94
CDasXORNet-ResNet18	91.57	91.64	84.51	91.61	99.15
CDasXORNet	**92.31**	91.36	**84.90**	**91.83**	**99.16**

produced high-quality results with fewer omission and commission errors on the WHU-BCD test dataset. In further investigating the tested approaches, four representative samples with a size of 2048 × 2048 pixels distributed in different test regions were selected, shown as the red square marked areas in subfigure (d). Qualitative evaluations of the tested methods in four chosen regions are shown in Fig. 6.

The examples in the top row of Fig. 6 represent the typical building change types in the training set. The main challenges include multi-scale building changes and confusion between buildings and other artificial objects such as squares. The evaluation results of CTFINet, FC-Siam-Conc, and IFTSDNet revealed apparent false detections, indicating that the concatenation of Siamese features or the feature interaction may not adequately model building changes in such a scenario. Some unchanged buildings were mistakenly labeled as changes in the results of LGPNet. LGPNet incorrectly recognized unaltered buildings with apparent spectral differences as changes. This finding may be attributed to the pre-trained weights from the INRIA dataset when training LGPNet for transfer learning. The generalization ability of LGPNet relies on the similarity between the building dataset used for pre-training and the test scene. The results of LGPNet also have errors at the edges. All methods missed some small building changes, showing that the model's capacity to adjust to scale variation in buildings still needs strengthening. CDasXORNet, and UNet++_MSOF achieved better BCD results. CDasXORNet has almost no false alarms, while other methods concurrently have omission and commission errors.

The second and third rows show change detection of newly built buildings scene, which is the mainstream change type in the WHU-BCD test dataset but is less common in the training dataset. CTFINet and IFTSDNet overlooked many transitions from bare soil to buildings in the two test scenes. Based on the hybrid change-deciding operation, CTFINet did not robustly model building changes between the Siamese features, resulting in its weak generalization capability. IFTSDNet produced promising BCD results on the LEVIR-CD dataset but performed poorly on the WHU-BCD dataset.

The EGRCNN and UNet++_MSOF also ignored some actual building changes. EGRCNN employs an RCNN to model changes. UNet++_MSOF

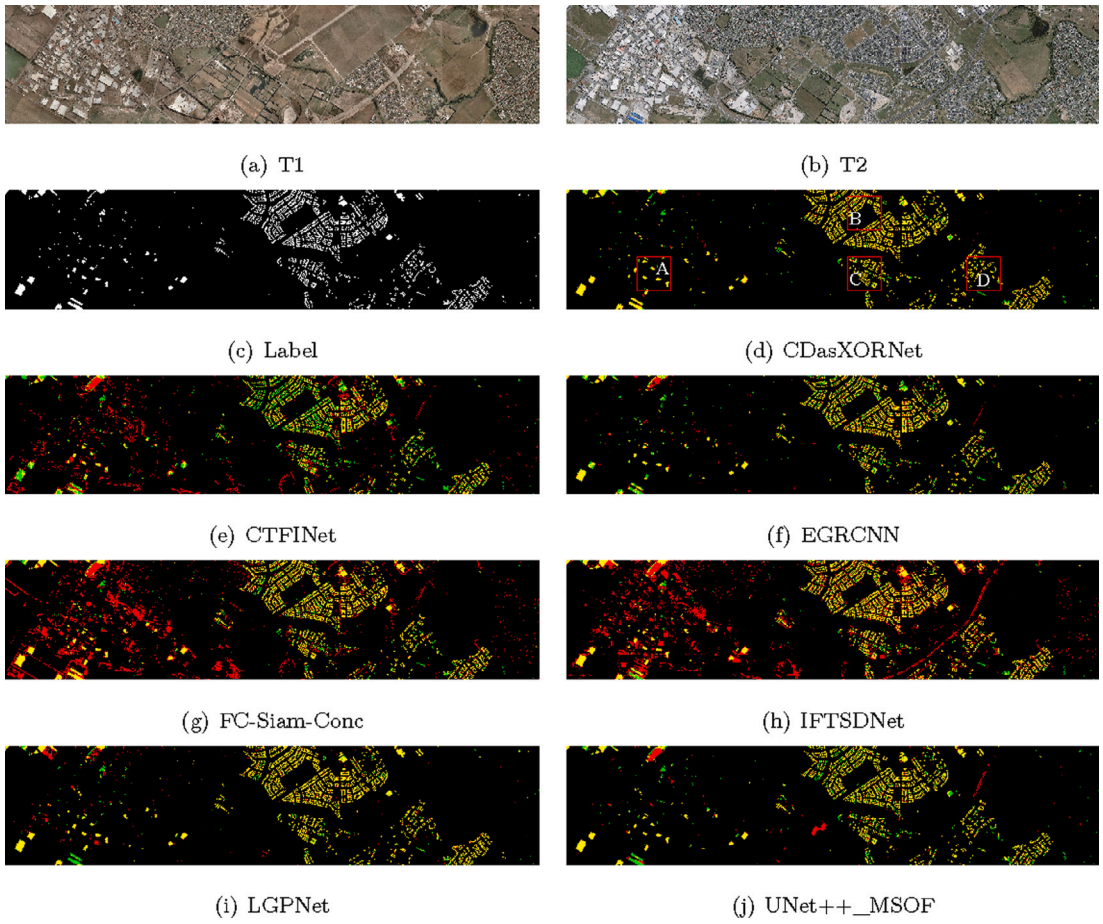


Fig. 5. Overall qualitative evaluation results of the WHU-BCD dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

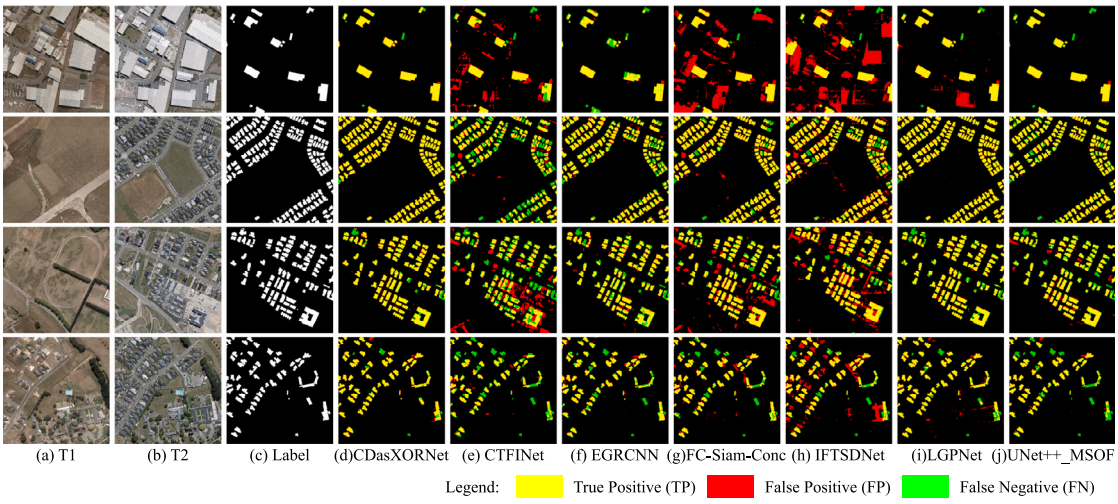


Fig. 6. Detailed qualitative evaluation results of comparative methods on the WHU-BCD dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

is based on multi-scale deep supervision. When the representativeness of training samples is low, neither EGRCNN nor UNet++_MSOF detects building changes efficiently. The proposed CDasXORNet almost has no false alarms, and only a few wrongly recognized building changes were found in regions B and C. CDasXORNet can simultaneously focus on the change and changeless information between the bi-temporal images, thus modeling building changes robustly. Therefore, high-quality BCD results were produced despite the weak representativeness of the training samples.

The final row of Fig. 6 depicts a scene where building changes occurred in the previous image, and the spectral difference between bi-temporal images is noticeable. FC-Siam-Conc incorrectly identified changes from bare soil to hardened surface as the building changes. EGRCNN, FC-Siam-Conc, and UNet++_MSOF overlooked the changes

Table 2
Quantitative evaluation results on the WHU-BCD dataset(%).

Method	Recall	Precision	IoU	F1-score	OA
UNet++_MSOF	76.76	78.50	63.43	77.62	97.75
STANet	**91.38**	80.54	74.86	85.62	98.44
FC-EF	72.37	76.09	58.96	74.18	97.44
FC-Siam-Conc	85.30	44.88	41.66	58.81	93.93
FC-Siam-Diff	29.05	24.81	15.45	26.77	91.93
DSAMNet	89.01	75.94	69.42	81.95	98.01
LGPNet	86.73	87.37	77.06	87.05	98.69
EGRCNN	78.91	81.54	66.95	80.20	98.02
BIT-CD	50.78	61.46	38.52	55.61	95.88
CTFINet	61.24	51.47	38.82	55.93	95.10
IFTSDNet	88.29	43.26	40.91	58.07	93.52
CDasXORNet-ResNet18	91.72	90.46	83.63	91.09	99.09
CDasXORNet	91.12	**92.89**	**85.18**	**91.99**	**99.19**

from hardened ground to buildings due to the stark variance between changes in different buildings in the two periods. CDasXORNet and LGPNet achieved better results than the comparative methods, and CDasXORNet had fewer false detections than LGPNet. The detailed qualitative analysis demonstrated that CDasXORNet could detect building changes more precisely and has enhanced generalization capabilities. To further compare the BCD performance of each algorithm, a quantitative analysis of the entire WHU-BCD test dataset was conducted, with the results in Table 2.

As shown in Table 2, the BCD performance of the proposed CDasXORNet vastly outperformed other comparative methods. The IoU of CDasXORNet is more than 8% higher than that of the second-ranked LGPNet, demonstrating a significant performance gain. The primary types of building change in the training and test datasets differ apparently. The test dataset predominantly consists of transitions from bare dirt to gray roofs. In contrast, the training dataset mainly contains large-scale building changes and changes from bare ground to bright roofs. The significant advancement of CDasXORNet concerning the comparative approaches thoroughly verified that the proposed method has more substantial generalization potential. Among the comparative methods, STANet achieved the highest recall, only 0.26% higher than CDasXORNet. In comparison to STANet, CDasXORNet has a 12.35% higher precision. Hence, the comprehensive evaluation index IoU of CDasXORNet is more than 10% higher than STANet. This trend is consistent with the observation of qualitative evaluation that STANet has many false detections. Most of the comparative algorithms only achieved an IoU lower than 70%, indicating that their building change modeling capability is inferior to the proposed method in this scenario. CDasXORNet exhibits strong generalization ability and robustness.

5.1.3. Experiments on the NJDS dataset

The NJDS dataset introduces the building height displacement problem in change detection (Shen et al., 2022). Fig. 7 shows the qualitative evaluation results of the tested methods on the NJDS dataset. As seen in Fig. 7, our proposed CDasXORNet achieved the highest-quality BCD results. The first and second rows show examples of significant spectral differences between bitemporal images of unchanged and changed parts, respectively. The irrelevant differences confused the comparative methods and resulted in their low-quality BCD results. CTFINet, FC-Siam-Conc, IFTSDNet, and LGPNet missed the building change in the first case. LGPNet, and UNet++_MSOF produced false BCD results when detecting building changes with significant differences. Examples of building misplacement are shown in the third and fourth rows. From the third row, CTFINet and IFTSDNet missed the building change. The two methods applied feature interaction between bi-temporal features to enhance change detection performance. They reduced the false detection of buildings caused by different imaging angles but ignored the actual building changes. The proposed CDasXORNet, however, detected the building changes correctly. To quantitatively evaluate the accuracy of the detected building changes, the five evaluation indices of 1717 test images were calculated and shown in Table 3. The highest value per column is bolded, and the secondary values are underlined.

As shown in Table 3, the proposed CDasXORNet achieved the highest value in all evaluation indicators, indicating that our method achieved the best BCD performance. EGRCNN achieved the second-highest recall value but the lowest IoU value, indicating that edge information cannot enhance model performance when dealing with building misplacement issues. STANet achieved the second-highest IoU value but was 9.43% lower than our method. The superior performance demonstrates that our method significantly reduce misclassification caused by substantial irrelevant spectral differences and building rooftop misplacement between the bi-temporal images.

Comparative experiments on the LEVIR-CD, WHU-BCD, and NJDS datasets demonstrate that CDasXORNet consistently achieved a higher IoU than the comparative methods. The performance of the same comparative approach has different rankings on the two datasets. For instance, FC-Siam-Diff achieved the fifth-highest IoU on the LEVIR-CD dataset but ranked last on the WHU-BCD dataset. CDasXORNet produced reliable results with low false alarm rates in challenging scenarios such as considerable spectral variations, confusion between buildings and backgrounds, and building positional inconsistencies. The proposed CDasXORNet properly models the building change information and has a strong generalization ability even in processing different building change types between the training and test datasets. Therefore, CDasXORNet has strong generalization capability and robustness. The effectiveness of the proposed key modules, including the XOR approximating module, the ResLA module, and the number of XOR approximating operations, were assessed by conducting ablation experiments on the LEVIR-CD and WHU-BCD datasets.

5.2. Ablation analysis

In this section, ablation experiments were conducted to evaluate the effectiveness of the suggested modules. The ablation models were created by gradually adding modules to the baseline model (Baseline). The developed modules of CDasXORNet are completely removed to produce the Baseline. Therefore, the Siamese encoder of Baseline is a ResNet34 backbone with five-stage down-sampling. The Baseline's decoder is the stacking ordinary convolution block corresponding to the encoder, and the change deciding module is the absolute of the subtraction of bitemporal features. Baseline+XOR represents replacing the absolute of the differential with the XOR approximating operation for the change decision. The Baseline+XOR+ResLA stands for further added a ResLA decoder. In CDasXORNet, the five-stage down-sampling and the number of XOR approximating operations are replaced by the four-stage ones. Table 4 shows the quantitative evaluation results on two datasets.

CDasXORNet achieved the highest IoU and F1-score in both datasets, demonstrating the rationality of the proposed overall architecture. The Baseline had the highest recall but the lowest precision on the WHU-BCD dataset. Introducing the XOR approximating operation

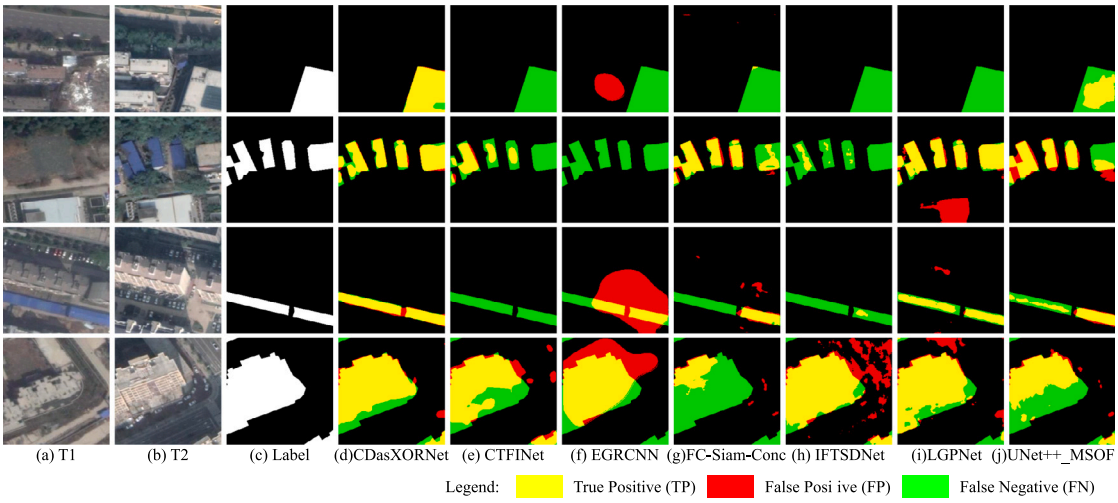


Fig. 7. Detailed qualitative evaluation results of comparative methods on the NJDS dataset.

Table 3
Quantitative evaluation results on the NJDS dataset(%).

Method	Recall	Precision	IoU	F1-score	OA
FC-EF	43.63	53.78	31.73	48.18	97.32
CTFINet	45.90	66.53	37.29	54.32	97.80
LGPNet	48.52	68.00	39.50	56.63	97.88
DSAMNet	48.92	63.58	38.21	55.30	97.74
FC-Siam-Conc	51.15	47.78	32.81	49.41	97.01
IFTSDNet	52.12	45.85	32.26	48.78	96.88
UNet++_MSOF	52.77	61.62	39.72	56.86	97.72
FC-Siam-Diff	52.94	48.03	33.66	50.37	97.02
BIT-CD	54.84	58.95	39.69	56.82	97.62
STANet	55.95	75.62	47.40	64.32	98.23
EGRCNN	62.16	10.08	9.50	17.35	83.12
CDasXORNet_ResNet18	63.14	80.51	54.77	70.77	98.51
CDasXORNet	**65.50**	**81.12**	**56.83**	**72.48**	**98.58**

Table 4
Quantitative evaluation results of ablation experiments on LEVIR-CD and WHU-BCD datasets (%).

Method	LEVIR-CD				WHU-BCD					
	Recall	Precision	IoU	F1-score	OA	Recall	Precision	IoU	F1-score	OA
Baseline	89.61	91.40	82.64	90.49	99.04	**92.82**	84.92	79.69	88.70	98.80
+XOR	89.30	**92.69**	83.43	90.97	99.10	91.64	88.74	82.09	90.16	98.98
+XOR+ResLA	90.03	92.34	83.78	91.17	99.11	92.50	89.05	83.05	90.74	99.04
CDasXORNet	**92.31**	91.36	**84.90**	**91.83**	**99.16**	91.12	**92.89**	**85.18**	**91.99**	**99.19**

to the Baseline achieved a 2.4% IoU improvement on the WHU-BCD dataset, illustrating that the XOR approximating operation can more accurately model the building changes than by absolute subtraction. Additionally, considering the sample difference between the training and test sets of the WHU-BCD, the XOR approximating operation significantly improved the generalization ability of the proposed framework. The four-stage down-sampling and XOR approximating operation is another design that contributed to model performance improvement. Although four-stage down-sampling will increase the number of network parameters, detailed spatial information are preserved. The four-stage XOR approximating operation eliminated the insufficient shallow features susceptible to interference from spectral aliasing. The addition of ResLA further enhances the efficacy of BCD. Baseline+XOR+ResLA ranked second on all accuracy evaluation indices in both datasets, demonstrating the reliability of the XOR approximation operation and ResLA. The three devised modules improved CDasXORNet's BCD performance significantly. The XOR approximating operation mines the correlation between the bi-temporal features and focuses on change and changeless information. The ResLA mechanism improves the feature representativeness of CDasXORNet. The four-stage XOR approximating design further increases the proposed framework's robustness to spurious changes such as spectral confusion and positional inconsistencies, enabling CDasXORNet to generate fewer false alarms and show strong generalization ability.

Aimed at visually comparing the difference in features captured by different ablation models, the Grad-CAM (Selvaraju et al., 2017) technique was utilized to generate heatmaps. The heatmap of three bi-temporal sample images in the LEVIR-CD test dataset is shown in Fig. 8. Comparing the fourth and fifth columns indicates that introducing the XOR approximating operation enhances the model's attention to the changeless region, as the heatmap covering the unchanged area becomes considerably brighter. Compared with the technique based on absolute subtraction, the BCD based on the XOR approximating operation shows superior performance in simultaneously focusing on changed and unchanged information. However, the emphasis on regions with noticeable spectral variations is still excessive in the results of Baseline+XOR because of the use of the five-level XOR approximating operations. The addition of ResLA eliminates misclassifications from Baseline+XOR. Nevertheless, as seen in the second row and sixth column, Baseline+XOR+ResLA may focus too much on the changed region. The proposed CDasXORNet obtains the highest quality BCD

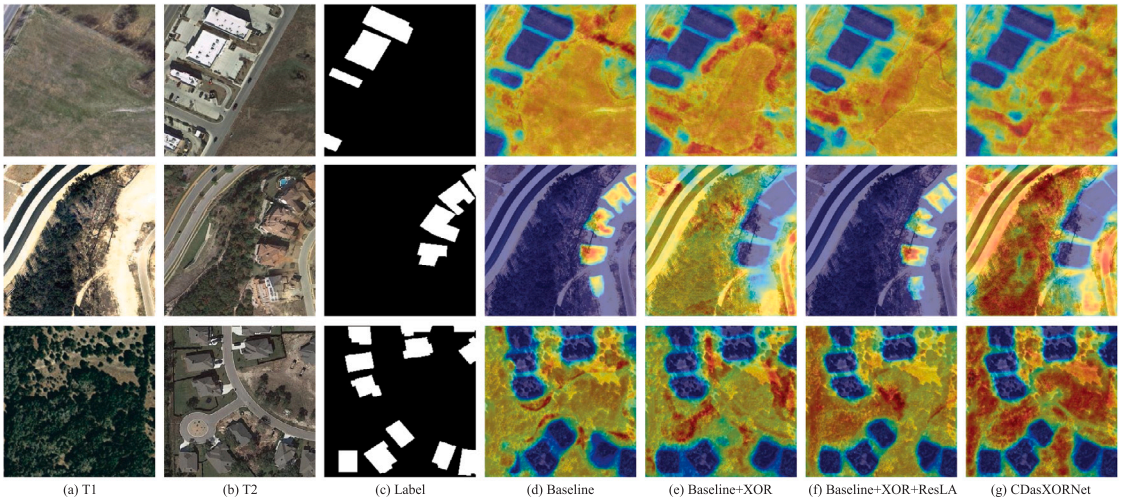


Fig. 8. Heatmaps of CDasXORNet's ablation models.

Table 5
Computational complexity analysis of different comparative and ablation models on the LEVIR-CD test dataset.

Method	Parameters(M)	FPS	IoU(%)
UNet++_MSOF	9.16	92.98	79.80
STANet	–	64.04	79.09
FC-EF	12.12	135.65	78.36
FC-Siam-Conc	13.89	98.02	79.29
FC-Siam-Diff	12.12	103.42	80.70
DSAMNet	16.95	65.35	73.50
LGPNet	–	23.10	83.34
EGRCNN	–	29.59	79.87
BIT-CD	12.40	47.52	81.72
CTFINet	4.96	36.67	78.39
IFTSDNet	4.52	41.11	81.33
CDasXORNet-ResNet18	17.65	50.05	84.51
Baseline	25.27	56.96	82.64
+XOR	26.69	48.97	83.43
+XOR+ResLA	27.86	36.45	83.78
CDasXORNet	27.73	36.88	84.90

results because it simultaneously concentrates on critical changed and changeless regions.

5.3. Complexity analysis

In assessing the computational complexity of the tested approaches, the number of parameters and inference speed of different models were counted. Experiments were conducted on the LEVIR-CD test dataset. The size of the test image is 256 × 256 pixels. Table 5 presents the results of the comparative methods and ablation models. The symbol "–" in the table indicates that the value cannot be counted.

The FPS value of FC-EF is the highest, indicating that FC-EF has the fastest inference speed. However, the IoU of FC-EF is the second smallest. LGPNet achieved the highest IoU among the comparative approaches, but its FPS value was the lowest. The proposed CDasXORNet achieved the highest IoU with moderate inference speed. UNet++_MSOF has the fewest number of parameters. STANet and LGPNet can only be trained with the batch size of 8 and 16 in the same experimental environment as other methods, so their computational complexity is relatively high. The proposed CDasXORNet has relatively more parameters than the other comparative methods. However, the proposed CDasXORNet-ResNet18 has 17.65 M parameters, its FPS exceeds 50, and it outperformed the comparative methods in terms of the IoU. The superior performance of CDasXORNet-ResNet18 illustrates that the suggested framework's design, rather than adding more parameters, is primarily responsible for the IoU improvement. Compared with the Baseline, the Baseline+XOR only increased a few parameters but significantly improved the performance of the proposed framework. The number of parameters is highest for Baseline + XOR + ResLA. By reducing one XOR approximating operation, CDasXORNet reduced the number of parameters and improved the BCD performance. The proposed framework yielded the highest quality BCD results with moderate computational complexity. Furthermore, the devised XOR approximating operation only added a few parameters.

5.4. Extensive analysis

To further validate the robustness of the proposed method across different backbone architectures, we replaced the backbone of the proposed framework with ResNeXt (Xie et al., 2017) and SegFormer (Xie et al., 2021). The ResNeXt architecture replaces the original ResNet's three-layer convolutional blocks with a parallel stack of blocks while maintaining the same topology. This modification enhances model performance while reducing the number of hyperparameters. SegFormer integrates the Transformer with a lightweight MLP, resulting in a straightforward and efficient semantic segmentation framework. Due to the distinct encoding and decoding mechanisms of SegFormer and the convolution-based backbone, we only compared the change-deciding module. Specifically, CD_Diff_ResNeXt and CD_XOR_ResNeXt were implemented by replacing the Siamese encoder from ResNet34 with ResNeXt50_$32 \times 4d$. For SegFormer, we utilized its feature extraction module as the Siamese encoder. Change information was extracted using subtraction (CD_Diff_SegFormer) and approximating XOR operations (CD_XOR_SegFormer). The decoder of CD_Diff_SegFormer and CD_XOR_SegFormer remained identical to the original SegFormer. Experiments were conducted on the LEVIR-CD dataset, and 100 epochs were trained. The batch sizes for ResNeXt50_$32 \times 4d$-based and SegFormer-based change detection methods were set to 8 and 16, respectively. The other experimental settings are the same as in Section 4.2. Table 6 shows the evaluation results of the comparative methods with different backbones.

Table 6 reveals that CD_XOR_ResNeXt outperforms CD_Diff_ResNeXt, and CD_XOR_SegFormer outperforms CD_Diff_SegFormer. It demonstrates that methods utilizing XOR approximation operation yield superior results compared to the subtraction operation-based approaches. Specifically, when compared with methods based on subtraction for change-deciding, the IoU value of XOR-based methods utilizing ResNeXt and SegFormer as backbones increased by 0.26% and 0.76%, respectively. Furthermore, CD_XOR_SegFormer achieved an IoU of 84.66%, indicating that integrating the proposed XOR module with

Table 6
Extensive analysis of the XOR module with different backbones on the LEVIR-CD test dataset.

Method	Recall	Precision	IoU	F1-score	OA
CD_Diff_ResNeXt	90.27	91.08	82.94	90.67	99.05
CD_XOR_ResNeXt	89.40	92.31	83.20	90.83	99.08
CD_Diff_SegFormer	90.40	92.11	83.90	91.25	99.12
CD_XOR_SegFormer	91.05	92.35	84.66	91.70	99.16

transformers can further enhance change detection performance. Extensive experiments demonstrate that the proposed hierarchical XOR approximating operation better models change information and exhibits excellent scalability.

6. Conclusion

This paper proposes a novel building change detection method named CDasXORNet to detect building changes from high-resolution bi-temporal images at high quality. Unlike the conventional building change detection neural network that use concatenation or subtraction to extract building change features, CDasXORNet considers change detection as a XOR problem to explore a new way to extract building change features for improved building change detection effects. In the proposed CDasXORNet, a hierarchical XOR approximating operation was designed to model multi-scale change and changeless information simultaneously. A ResLA mechanism was introduced in CDasXORNet to further improve the discriminability of extracted building change features. Experimental results on three challenging building change detection datasets demonstrated that CDasXORNet generated high-quality results with fewer false alarms when dealing with challenges such as positional inconsistency, notable irrelevant spectral differences, and similarity between buildings and surrounding objects in building change detection tasks. However, there are still limitations in the results of our proposed CDasXORNet as buildings under construction or small-scale building changes were missed. In future research, introducing multi-scale feature extraction modules and augmentation specifically for samples of buildings under construction are possible solutions.

CRediT authorship contribution statement

Shanxiong Chen: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization, Writing – review & editing. **Wenzhong Shi:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **Mingting Zhou:** Conceptualization, Investigation, Funding acquisition, Methodology, Software, Writing – review & editing. **Min Zhang:** Writing – review & editing, Supervision, Formal analysis, Funding acquisition, Methodology. **Yue Yu:** Writing – review & editing, Funding acquisition, Formal analysis. **Yangjie Sun:** Writing – review & editing, Funding acquisition, Formal analysis. **Linjie Guan:** Writing – review & editing, Funding acquisition, Formal analysis. **Shuangping Li:** Writing – review & editing, Funding acquisition, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work was supported by Innovation and Technology Commission, HKSAR Government, China (ITP/041/21LP), Urban Informatics for Smart Cities, The Hong Kong Polytechnic University, China (1-ZVN6, ZVU1), and the Otto Poon Charitable Foundation Smart Cities Research Institute, Hong Kong Polytechnic University, China (Work Program: CD03).

References

Bai, B., Fu, W., Lu, T., Li, S., 2021. Edge-guided recurrent convolutional neural network for multitemporal remote sensing image building change detection. IEEE Trans. Geosci. Remote Sens. 60, 1–13.
Benedek, C., Descombes, X., Zerubia, J., 2011. Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. IEEE Trans. Pattern Anal. Mach. Intell. 34 (1), 33–50.
Cao, Y., Huang, X., 2023. A full-level fused cross-task transfer learning method for building change detection using noise-robust pretrained networks on crowdsourced labels. Remote Sens. Environ. 284, 113371.
Chang, H., Sun, X., Wang, P., Diao, W., Xu, G., 2023. A transformer-based network with differential feature triple refinement for bitemporal remote sensing image change detection.
Chen, C.-P., Hsieh, J.-W., Chen, P.-Y., Hsieh, Y.-K., Wang, B.-S., 2023a. SARAS-net: scale and relation aware siamese network for change detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 37, pp. 14187–14195.
Chen, H., Li, W., Shi, Z., 2021a. Adversarial instance augmentation for building change detection in remote sensing images. IEEE Trans. Geosci. Remote Sens. 60, 1–16.
Chen, H., Qi, Z., Shi, Z., 2021b. Remote sensing image change detection with transformers. IEEE Trans. Geosci. Remote Sens. 60, 1–14.
Chen, H., Shi, Z., 2020. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. Remote Sens. 12 (10), 1662.
Chen, Z., Song, Y., Ma, Y., Li, G., Wang, R., Hu, H., 2023b. Interaction in transformer for change detection in VHR remote sensing images. IEEE Trans. Geosci. Remote Sens. 61, 1–12.
Chen, Z., Zhou, Y., Wang, B., Xu, X., He, N., Jin, S., Jin, S., 2022. 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. 191, 203–222.
Daudt, R.C., Le Saux, B., Boulch, A., 2018. Fully convolutional siamese networks for change detection. In: 2018 25th IEEE International Conference on Image Processing. ICIP, IEEE, Athens, Greece, pp. 4063–4067.
Ding, Q., Shao, Z., Huang, X., Altan, O., 2021. DSA-net: A novel deeply supervised attention-guided network for building change detection in high-resolution remote sensing images. Int. J. Appl. Earth Obs. Geoinf. 105, 102591.
Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.
Eftekhari, A., Samadzadegan, F., Javan, F.D., 2023. Building change detection using the parallel spatial-channel attention block and edge-guided deep network. Int. J. Appl. Earth Obs. Geoinf. 117, 103180.
Feng, Y., Jiang, J., Xu, H., Zheng, J., 2023. Building change detection using cross-temporal feature interaction network. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, pp. 1–5.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR, IEEE, Las Vegas, NV, USA, pp. 770–778.
Huang, X., Zhang, L., Zhu, T., 2013. Building change detection from multitemporal high-resolution remotely sensed images based on a morphological building index. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7 (1), 105–115.
Hussain, M., Chen, D., Cheng, A., Wei, H., Stanley, D., 2013. Change detection from remotely sensed images: From pixel-based to object-based approaches. ISPRS J. Photogramm. Remote Sens. 80, 91–106.
Janalipour, M., Taleai, M., 2017. Building change detection after earthquake using multi-criteria decision analysis based on extracted information from high spatial resolution satellite images. Int. J. Remote Sens. 38 (1), 82–99.
Ji, S., Wei, S., Lu, M., 2018. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. IEEE Trans. Geosci. Remote Sens. 57 (1), 574–586.
Jiang, H., Hu, X., Li, K., Zhang, J., Gong, J., Zhang, M., 2020. PGA-SiamNet: Pyramid feature-based attention-guided siamese network for remote sensing orthoimagery building change detection. Remote Sens. 12 (3), 484.
Knudsen, T., Olsen, B.P., 2003. Automated change detection for updates of digital map databases. Photogramm. Eng. Remote Sens. 69 (11), 1289–1296.
Li, R., Zheng, S., Duan, C., Su, J., Zhang, C., 2021. Multistage attention resu-net for semantic segmentation of fine-resolution remote sensing images. IEEE Geosci. Remote Sens. Lett. 19, 1–5.

Liu, T., Gong, M., Lu, D., Zhang, Q., Zheng, H., Jiang, F., Zhang, M., 2021. Building change detection for VHR remote sensing images via local–global pyramid network and cross-task transfer learning strategy. IEEE Trans. Geosci. Remote Sens. 60, 1–17.

Liu, W., Lin, Y., Liu, W., Yu, Y., Li, J., 2023. An attention-based multiscale transformer network for remote sensing image change detection. ISPRS J. Photogramm. Remote Sens. 202, 599–609.

Liu, M., Shi, Q., 2021. DSAMNet: A deeply supervised attention metric based network for change detection of high-resolution images. In: 2021 IEEE International Geoscience and Remote Sensing Symposium. IGARSS, IEEE, Brussels, Belgium, pp. 6159–6162.

Lu, W., Wei, L., Nguyen, M., 2024. Bi-temporal attention transformer for building change detection and building damage assessment. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 1–20.

Peng, D., Bruzzone, L., Zhang, Y., Guan, H., Ding, H., Huang, X., 2020. SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images. IEEE Trans. Geosci. Remote Sens. 59 (7), 5891–5906.

Peng, D., Zhang, Y., Guan, H., 2019. End-to-end change detection for high resolution satellite images using improved UNet++. Remote Sens. 11 (11), 1382.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision. IEEE, Venice, Italy, pp. 618–626.

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

Shi, W., Zhang, M., Zhang, R., Chen, S., Zhan, Z., 2020. Change detection based on artificial intelligence: State-of-the-art and challenges. Remote Sens. 12 (10), 1688.

Shu, Q., Pan, J., Zhang, Z., Wang, M., 2022. MTCNet: Multitask consistency network with single temporal supervision for semi-supervised building change detection. Int. J. Appl. Earth Obs. Geoinf. 115, 103110.

Song, Z., Wei, X., Kang, X., Li, S., Liu, J., 2023. Towards efficient remote sensing image change detection via cross-temporal context learning. IEEE Trans. Geosci. Remote Sens.

Song, S., Zhang, Y., Yuan, Y., 2024. Iterative edge enhancing framework for building change detection. IEEE Geosci. Remote Sens. Lett. 21, 1–5.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Wang, X., Du, J., Tan, K., Ding, J., Liu, Z., Pan, C., Han, B., 2022. A high-resolution feature difference attention network for the application of building change detection. Int. J. Appl. Earth Obs. Geoinf. 112, 102950.

Wang, L., Zhang, J., Guo, Q., Chen, D., 2023. IFTSDNet: An interact-feature transformer network with spatial detail enhancement module for change detection. IEEE Geosci. Remote Sens. Lett. 20, 1–5.

Wu, Z., Su, L., Huang, Q., 2019. Stacked cross refinement network for edge-aware salient object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, Seoul, Korea, pp. 7264–7273.

Xiao, P., Yuan, M., Zhang, X., Feng, X., Guo, Y., 2017. Cosegmentation for object-based building change detection from high-resolution remotely sensed images. IEEE Trans. Geosci. Remote Sens. 55 (3), 1587–1603.

Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1492–1500.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P., 2021. SegFormer: Simple and efficient design for semantic segmentation with transformers. Adv. Neural Inf. Process. Syst. 34, 12077–12090.

Zhang, L., Hu, X., Zhang, M., Shu, Z., Zhou, H., 2021. Object-level change detection with a dual correlation attention-guided detector. ISPRS J. Photogramm. Remote Sens. 177, 147–160.

Zhang, X., Xiao, P., Feng, X., Yuan, M., 2017. Separate segmentation of multi-temporal high-resolution remote sensing images for object-based change detection in urban area. Remote Sens. Environ. 201, 243–255.

Zheng, H., Gong, M., Liu, T., Jiang, F., Zhan, T., Lu, D., Zhang, M., 2022. HFA-net: High frequency attention siamese network for building change detection in VHR remote sensing images. Pattern Recognit. 129, 108717.

Zhu, P., Xu, H., Luo, X., 2023. MDAFormer: Multi-level difference aggregation transformer for change detection of VHR optical imagery. Int. J. Appl. Earth Obs. Geoinf. 118, 103256.