remote sensing

MDPI

Article

Feature Decomposition-Optimization-Reorganization Network for Building Change Detection in Remote Sensing Images

Yuanxin Ye [1], Liang Zhou [1], Bai Zhu [1], Chao Yang [1], Miaomiao Sun [2], Jianwei Fan [3] and Zhitao Fu [4,*]

[1] Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 610031, China; yeyuanxin@home.swjtu.edu.cn (Y.Y.); zlup@my.swjtu.edu.cn (L.Z.); kevin_zhub@my.swjtu.edu.cn (B.Z.); yc18483685462@my.swjtu.edu.cn (C.Y.)
[2] SuperMap Software Co., Ltd., Chengdu 610041, China; sunmiaomiao@supermap.com
[3] School of Computer and Information Technology, Xinyang Normal University, Xinyang 464000, China; fanjw@xynu.edu.cn
[4] Faculty of Land Resources Engineering, Kunming University of Science and Technology, Kunming 650093, China
* Correspondence: zhitaofu@kust.edu.cn

Abstract: Building change detection plays an imperative role in urban construction and development. Although the deep neural network has achieved tremendous success in remote sensing image building change detection, it is still fraught with the problem of generating broken detection boundaries and separation of dense buildings, which tends to produce saw-tooth boundaries. In this work, we propose a feature decomposition-optimization-reorganization network for building change detection. The main contribution of the proposed network is that it performs change detection by respectively modeling the main body and edge features of buildings, which is based on the characteristics that the similarity between the main body pixels is strong but weak between the edge pixels. Firstly, we employ a siamese ResNet structure to extract dual-temporal multi-scale difference features on the original remote sensing images. Subsequently, a flow field is built to separate the main body and edge features. Thereafter, a feature optimization module is designed to refine the main body and edge features using the main body and edge ground truth. Finally, we reorganize the optimized main body and edge features to obtain the output results. These constitute a complete end-to-end building change detection framework. The publicly available building dataset LEVIR-CD is employed to evaluate the change detection performance of our network. The experimental results show that the proposed method can accurately identify the boundaries of changed buildings, and obtain better results compared with the current state-of-the-art methods based on the U-Net structure or by combining spatial-temporal attention mechanisms.

Keywords: building change detection; feature decomposition; feature optimization; feature reorganization

Citation: Ye, Y.; Zhou, L.; Zhu, B.; Yang, C.; Sun, M.; Fan, J.; Fu, Z. Feature Decomposition-Optimization-Reorganization Network for Building Change Detection in Remote Sensing Images. *Remote Sens.* **2022**, *14*, 722. https://doi.org/10.3390/rs14030722

Academic Editors: Damian Wierzbicki and Kamil Krasuski

Received: 26 November 2021
Accepted: 29 January 2022
Published: 3 February 2022

1. Introduction

Image change detection denotes the process of recognizing specific differences between multi-temporal images [1,2], which is a key technique for many applications, such as disaster assessment [3,4], land cover change detection [5,6], urban expansion monitoring [7], and so on.

As an important part of the blueprint of cities, the demolition, construction and expansion of buildings are closely related to human existence. It is of great significance to timely and accurately obtain the change information of buildings for human development. With the rapid development of remote sensing imaging technology, massive remote sensing images can be used for building change detection following high-precision co-registration [8]. Building change detection based on remote sensing images has become an area of immense research interest. Research on related change detection methods has also made great progress, from the early pixel-based building change detection methods

to the methods combining object-oriented analysis, as well as the methods employing only the spectral features to the methods combining spectrum, morphological index [9] and other features [10,11]. Although the traditional pixel-level and object-level methods have achieved fruitful research results, there are still many shortcomings in the accuracy and integrity of detection results due to the limited expression capacity of artificial design features and the accuracy of segmentation results.

In recent years, with its unique deep feature expression capacity, deep learning has provided new ideas for remote sensing image processing including semantic segmentation [12,13], object detection [14,15], image matching [16,17], etc. Many remote sensing image change detection methods based on deep learning have been proposed. Among them, the methods [18,19] that introduce deep learning into the traditional change detection process to extract features fail to make full use of the end-to-end structure, resulting in low detection efficiency. Therefore, fully convolutional networks (FCN) [20] are introduced into change detection, owing to their ability to perform pixel-level prediction, and build an end-to-end change detection mode. Although the end-to-end architecture improves the detection efficiency, the down-sampling operation in FCN will degrade the accuracy of the pixels' spatial location, making it difficult to obtain regular building boundaries. A series of enhanced deep learning methods have been proposed. For example, the encoder-decoder structure represented by U-Net [21] can recover spatial location information as much as possible through up-sampling and deconvolution operations to improve the accuracy of detection results. The methods represented by attention models can enhance the identifiability of the network for changed and unchanged pixels through spatial and channel attention, which can optimize the detection effect of building boundaries [22–24].

The building change detection methods based on deep learning have obtained good detection results. However, they improve the internal consistency of the object by modeling the global information of the image, or optimize the edge detection effect by fusing multi-scale features. All these methods do not take into consideration the strong similarity between main body pixels and the great difference between edge pixels. Saw-tooth boundaries are easily generated in detection results, and multiple adjacent buildings are easily identified as a single building.

Herein, based on the successful application of the decoupling idea [25] in semantic segmentation, we introduce a feature decomposition strategy into building change detection, and add the feature optimization structure on this basis. Then, a fast and effective building change detection framework based on the feature decomposition-optimization-reorganization network (FDORNet) is proposed. FDORNet first decomposes the main body and edge features, and trains the network with the multiple supervision strategy. A feature optimization structure is used to optimize the main body and edge features and reduce the irrelevant information in the original image. Finally, a complete building change detection process is formed by feature reorganization and up-sampling operation. The main contributions of this paper are as follows:

(1) We propose a novel framework, namely, feature decomposition-optimization-reorganization network (FDORNet) for building change detection. In our work, we model the main body and edge features of buildings separately based on the characteristics that the similarity between the main body pixels is strong but weak between the edge pixels.
(2) We introduce the decoupling idea into building change detection and employ the feature optimization structure to refine the main body and edge features, which greatly improves the accuracy of the boundary detection of changed buildings.

This paper is structured as follows: Section 2 introduces related work on change detection. Section 3 describes the proposed FDORNet in detail, and Section 4 introduces the experimental details and the building change detection datasets employed. Subsequently, experimental results are presented in Section 5. Finally, we conclude with a recommendation for future work.

2. Related Work

2.1. Traditional Change Detection Methods

The current traditional change detection methods are mainly pixel-based and object-based. Next, we will introduce the traditional change detection algorithms from these two aspects.

2.1.1. Pixel-Based Methods

Pixel-based change detection methods denote a process to determine change information pixel by pixel. The most commonly used algebraic methods include image difference, image ratio, change vector analysis (CVA) [26], etc. The threshold in algebraic methods needs to be specified when obtaining the final change detection results. The detection results are greatly affected by spectral difference since only the image spectral feature is used. Therefore, researchers have propounded many improved methods from the dimensions of threshold selection and feature utilization.

(Semi-)automatic threshold methods can surmount the limitations of the empirical threshold selection methods. Chen et al. [27] apply the double-window flexible pace search (DFPS) to CVA. Bruzzone et al. [28] determine the pixels to be changed and unchanged through the histogram of the difference map, and then use the EM algorithm to determine the optimal threshold. (Semi-)automatic threshold methods can improve the universality of the algorithm, but it is more complex.

To make full use of the image information, combining multiple features has become an important vehicle for improving the performance of change detection. The spectrum, texture, and context information of images are obtained through sliding windows, and the characteristics of ground objects are described more effectively. Li et al. [29] propose a method that combines spectral features and texture difference measurement to optimize change detection results. Mishra et al. [30] introduce local information and optimize detection results by employing fuzzy c-means and Gustafson–Kessel clustering.

2.1.2. Object-Based Methods

Object-based change detection methods consider homogeneous objects as the basic analysis unit, which can comprehensively utilize the shape and edges features, and effectively smoothen the salt-and-pepper phenomenon in pixel-based methods.

Object-based methods can be divided into post-classification comparison methods and direct comparison methods. The post-classification comparison methods first classify the dual-temporal images at object level, and then analyze the results based on the classification to achieve change detection. However, its accuracy is restricted by the classification performance. The direct comparison methods are not influenced by the classification results and are more robust. In general, images are segmented to obtain the homogeneous object according to the scale parameters, and the spectral and texture feature of the object are then extracted. Im et al. [31] improve the performance of change detection by generating neighborhood correlation images (NCI) through object-oriented analysis. Wang et al. [32] perform unsupervised change detection via the cross-sharpening of multitemporal data and image segmentation, which can reduce the detection errors precipitated by relief or spatial displacement.

Although the efficacy of object-based methods in improving the accuracy of change detection has been proven, such methods rely mainly on segmentation algorithms to segment images, and scale parameters usually need to be set during the process of segmentation. If the scale parameters are too small, it will induce obvious false alarms; however, if they are too large, it will cause the problem of missing detection.

2.2. Change Detection Methods Based on Deep Learning

Deep learning models can automatically extract robust abstract image features at multiple levels, and have become an effective way to perform change detection. At the outset, scholars mostly combine deep learning and traditional change detection methods.

For example, Liu et al. [33] propose a symmetric convolutional coupling network, which first inputs the optical and SAR image into a symmetric network structure to obtain the feature pairs, then calculates the Euclidean distance to obtain change information and finally obtains a binary change map through threshold segmentation. Zhan et al. [34] introduce the siamese convolutional neural network into change detection, extract the change information through the network, and use the k-nearest neighbor method to optimize the initial change detection results obtained by threshold segmentation.

Deep learning technology can significantly improve the accuracy of change detection results, but the above methods only replace traditional feature extraction with deep networks, and fail to make full use of the advantages of end-to-end structure. Daudt et al. [35] construct a change detection model based on a convolutional neural network and propose two image input methods; one is the early fusion (EF) mode and the other is the siamese mode. In the EF mode, dual-temporal images are stacked together as input, while in the siamese mode, images are used as input of different branches. Liu et al. [36] apply bipartite differential neural network (BDNN) to change detection of heterogeneous remote sensing images based on the strong feature extraction capacity of deep neural networks. BDNN uses the learnable change disguise maps (CDMs) to weaken the differences between the changed regions and enhance the network's ability to identify the unchanged regions. The test results show that BDNN have the capacity not only to detect changing information in the remote sensing images, but also to resist the impact of registration errors on the results to a certain degree.

Remote sensing image change detection is similar to the binary classification in semantic segmentation, but has its own characteristics, that is, the dual-temporal images used have spatial-temporal correlation. The attention model has been successfully applied in natural language processing and semantic segmentation, thus a series of change detection methods combined with attention models have been proposed. Mou et al. [37] designed an end-to-end recursive convolution neural network. The network learns a spectral-spatial-temporal feature representation to generate features with rich spatial-temporal information, and combines a recurrent neural network with CNN to obtain change information. Diakogiannis et al. [38] propose two methods based on ResNet, i.e., self-attention fusion and relative attention fusion. The former is used to enhance the attention of the region of interest in a single image, while the latter focuses on the correlation between dual-temporal images. Zhang et al. [39] extract deep features based on VGG16, and realize efficient fusion between multi-level features through channel attention and spatial attention module. Deep learning technology has greatly improved the performance of change detection, but its detection effect on changed ground object boundaries needs to be improved further. Some methodologies have already made attempts to extract regularized building boundaries. Marcos et al. [40] integrate priors and constraints including continuous boundaries, smooth edges, and sharp corners into the segmentation process. Zorzi et al. [41] use FCN trained with a combination of adversarial and regularized losses to perform building boundary refinement and regularization. Zorzi et al. [42] extend this algorithm by using a GAN-based model to extract regularized building boundaries after obtaining the building mask using FCN.

3. Methods

3.1. Overview

The FDORNet model is composed of four parts: feature extraction, feature decomposition, feature optimization, and feature reorganization. Specifically, in the feature extraction module, we extract multi-scale difference features of dual-temporal images. Main body features and edge features are separated by constructing a learnable flow field in the feature decomposition module. Subsequently, in the feature optimization module, multiple supervision strategies are adopted to accurately optimize the main body and edge features by using the corresponding main body and edge ground truth. Finally, we reorganize the

optimized features to form a complete building change detection process in the feature reorganization module. The overall design of the model is shown in Figure 1.

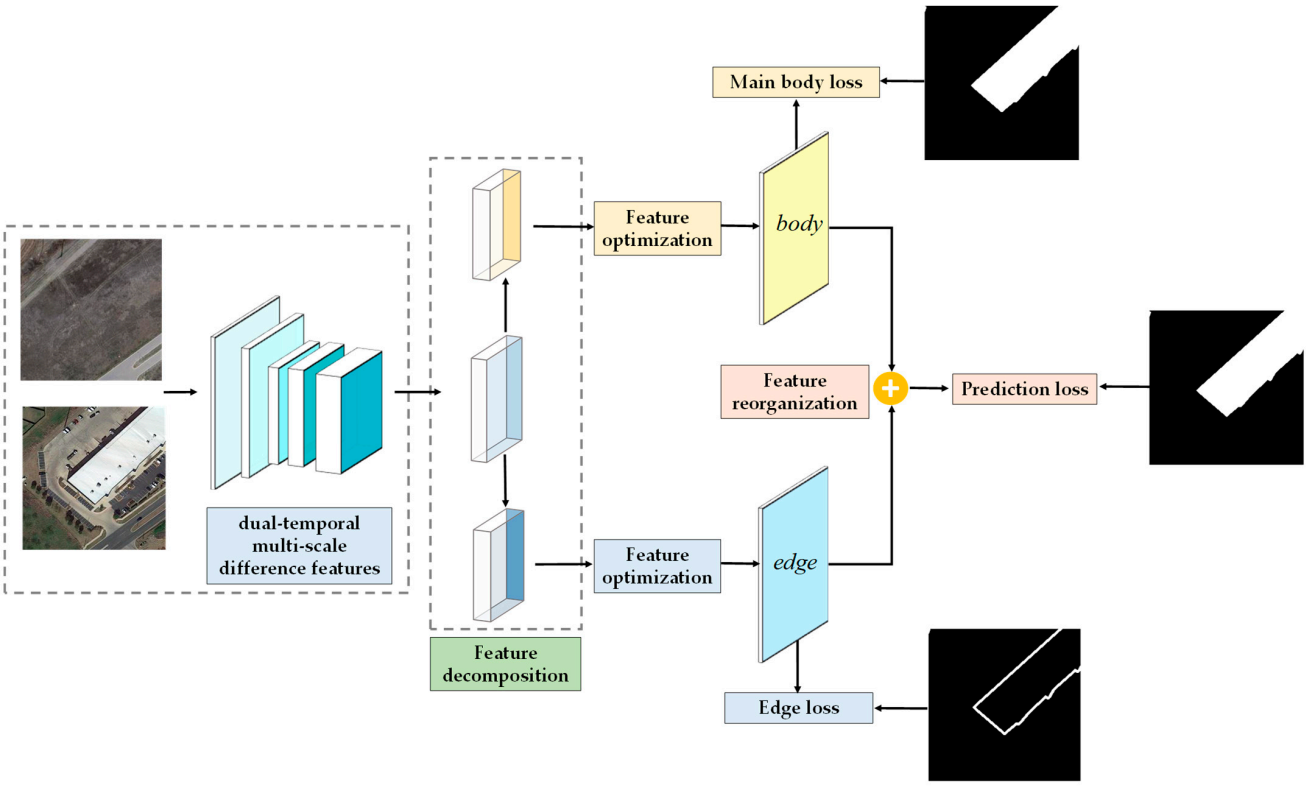


Figure 1. Overview of the proposed framework.

3.2. Feature Extraction

The feature extraction module has great effect on the final detection performance because of the spectral and temporal differences between remote sensing images used for building change detection. Herein, we employ the ResNet [43] framework, which has shown outstanding performance in object detection and image segmentation. In addition, the siamese framework consists of two branch networks with the same architecture that share the same weights and has natural advantages for change detection because it can efficiently extract the features of two input branches. Accordingly, a siamese ResNet framework which adopts ResNet to implement two branches of the siamese network is used in the feature extraction part.

3.2.1. ResNet

The depth of the network is very important in remote sensing image processing, but blindly deepening the network will cause the problem of degradation. As the depth of the network increases, accuracy tends to saturate and then declines rapidly. ResNet can solve the problem of degradation in deep neural networks by means of the residual connection.

ResNet uses residual blocks to implement shortcut connections, which learn the relationship between the target and the input not directly but in the form of residuals. The basic unit of residual block is shown in Equation (1),

$$T(x) = F(x, \{w_i, b\}) + x \tag{1}$$

where, $F(x, \{w_i, b\})$, W_i and b denote the residual function, weight, and bias, respectively.

There are slight differences in the form of residual basic units in the ResNet framework with different depths. Figure 2 shows the two commonly used forms of the residual basic unit. The residual basic unit composed of two convolutional layers on the left is the basic

module of ResNet34, and that containing three convolutional layers on the right is the basic structure of ResNet50, ResNet101, and ResNet152. Considering the feature extraction capacity, the number of model parameters and the model complexity, we build the feature extraction module in this work using ResNet50.

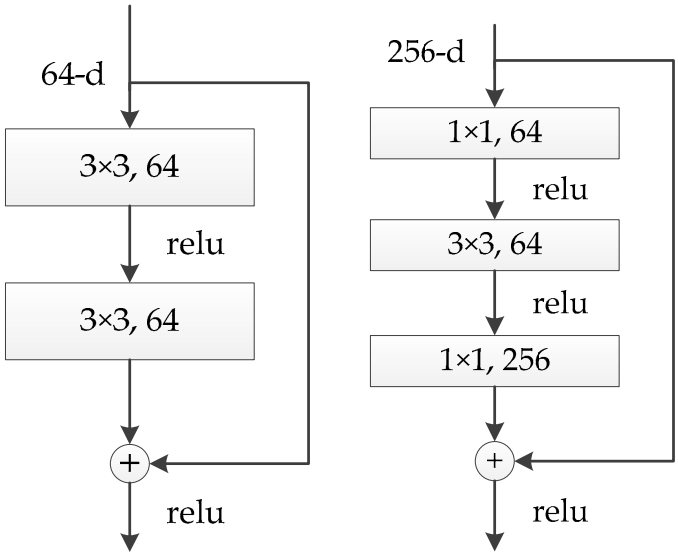


Figure 2. Two commonly used forms of the residual basic unit.

3.2.2. Feature Extraction

The feature extraction part is made up of five stages. The first stage consists of the convolutional layer which does not involve the residual learning unit, namely, conv1 in ResNet50. Stages 2 to 5 correspond to layers conv2_x to conv5_x in ResNet50, respectively, and contain residual learning units. After inputting the dual-temporal images, we can obtain the deep feature maps of each stage $\{F_1^1, F_2^1, F_3^1, F_4^1, F_5^1\}$ and $\{F_1^2, F_2^2, F_3^2, F_4^2, F_5^2\}$, where the superscript represents the temporal of the image, and the subscript represents the corresponding stage. In the original ResNet50 model, through pooling and strided convolution (stride = 2) operations, the output feature maps size of each stage is the result of two times the downsampled, however, the final feature map resolution will be too low, which is not conducive to building change detection. In this work, the strided convolution parameters of the last two stages are changed (setting stride = 1) so that the resolution of the feature map will not decrease further. For input images with a size of 256 × 256, the feature maps size in the five stages is 128 × 128 × 128, 256 × 64 × 64, 512 × 32 × 32, 1024 × 32 × 32, 2048 × 32 × 32, respectively. Compared with the original image, the difference map between images can more intuitively reflect the change information of ground buildings. Therefore, after extracting feature maps through ResNet50, the features of the corresponding stage are subtracted and the absolute value is taken to obtain dual-temporal multi-scale difference features $\{DF_1, DF_2, DF_3, DF_4, DF_5\}$.

3.3. Feature Decomposition

In remote sensing images, the main body usually corresponds to the low-frequency information part of the image, while the edge corresponds to the high-frequency part. Compared to the edge features with large differences, the main body features with stronger internal consistency are easier to extract. Moreover, for multi-scale features with different resolutions, the low-resolution features can better reflect the main information of images, and it is easier to extract the main body features of buildings. During the five stages of the feature extraction module (see Figure 3), dual-temporal multi-scale difference features DF_1 and DF_2 are rich in detailed information, and DF_5 reflect the main information of the image better. Accordingly, the dual-temporal difference features of the deepest level are taken as the input of the feature decomposition module. The learnable flow field provides a vehicle

for the internal pixel features to flow to the center of the changed buildings to extract the main body features. The edge features can then be obtained by subtracting the main body features from the input features DF_5. Figure 4 illustrates the feature decomposition module, where F denotes the input feature maps of the feature decomposition part, F_{Low} and DF are the encoding and decoding feature maps, $Flow$ represents the learnable flow field, and F_{body} and F_{edge} denote the main body features and edge features.

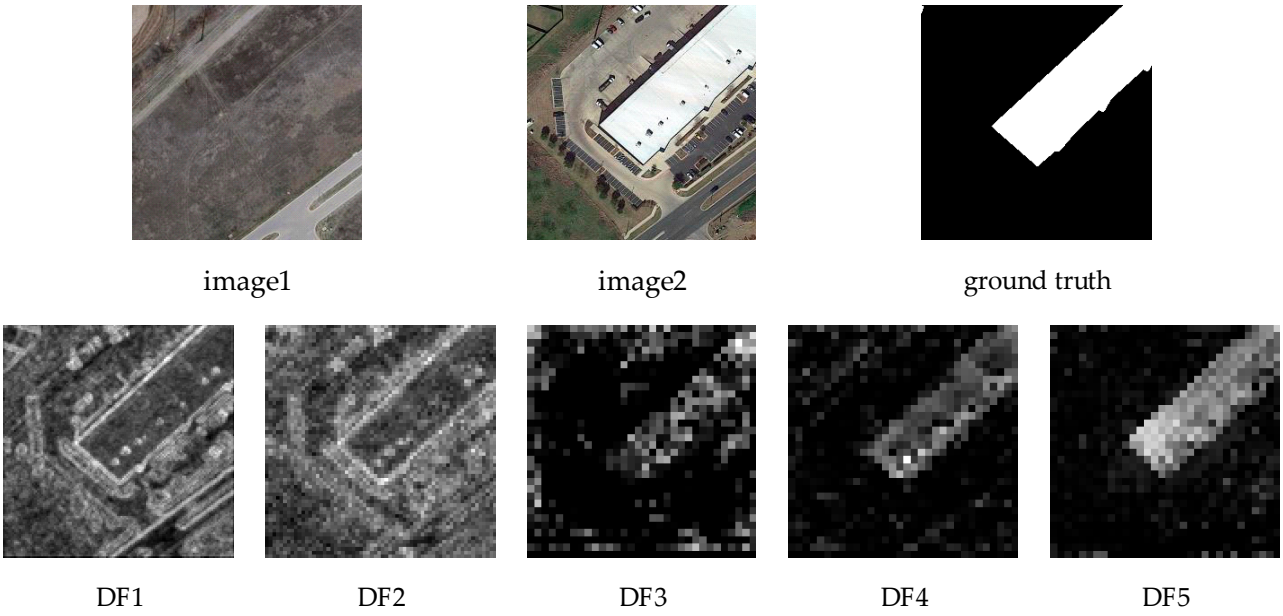


Figure 3. The output feature maps of each stage for feature extraction module.

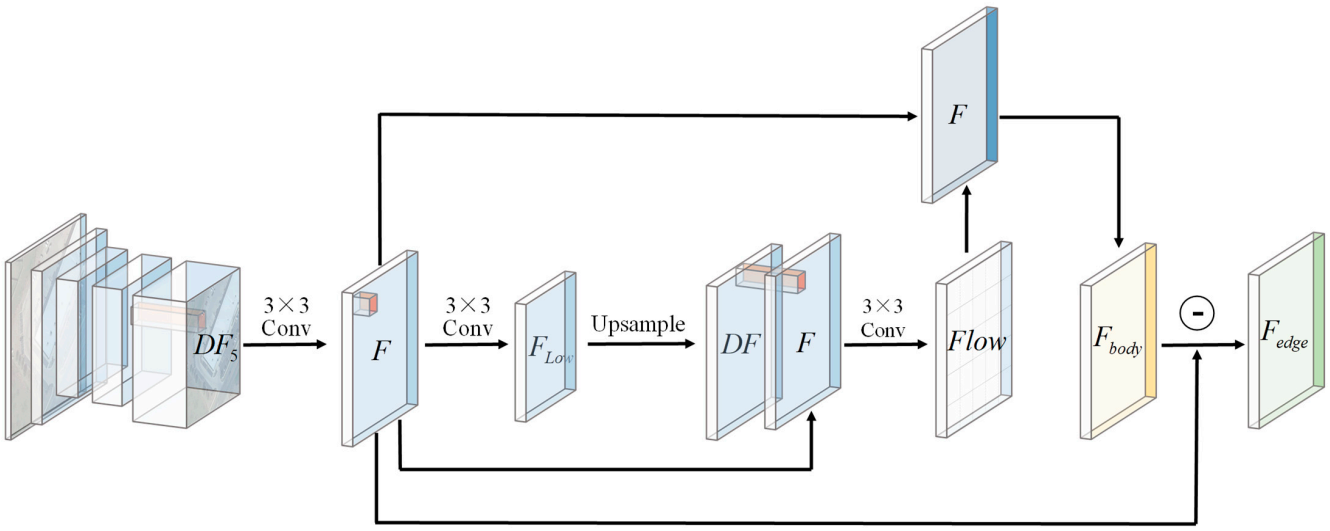


Figure 4. Schematic diagram of feature decomposition module.

3.3.1. Flow Field

In the feature decomposition module, the flow field acquires the main body features through learning of the mapping relationship between F and F_{body}. Its task is similar to that of optical flow, both of which are aimed at learning the movement information between the input and the target. Since the convolutional neural network is very good at learning the relationship between input and output by being given enough training data, this work references the framework of a neural optical flow network [44] to construct the feature decomposition module. We adopt the encoder-decoder structure, and obtain lower frequency encoding feature maps F_{Low} by the strided convolution operation, then generate

low-frequency feature maps DF through up-sampling to the same size as F. Finally, we concatenate F and DF together, and obtain the flow field through a 3×3 convolution operation. The feature map size of F, F_{Low}, DF and $Flow$ is $256 \times 32 \times 32$, $256 \times 16 \times 16$, $256 \times 32 \times 32$, $256 \times 32 \times 32$, respectively.

3.3.2. Main Body Features and Edge Features

The main body features focuses on the representation of the internal pixel information, which guides the flow direction of internal pixel features through $Flow$. In the mapping process, the differentiable bilinear sampling mechanism [45] is used to approximately estimate each point x_l in the low-frequency features. The four neighboring pixel values of x_l are obtained by bilinear interpolation operation.

$$F_{body}(x_l) = \sum_{x \in N(x_l)} \varpi_l F(x) \tag{2}$$

where ϖ_l represents the weight of the bilinear kernel on the separated space grid, which is mainly calculated by $Flow$, and N denotes the pixels in the neighborhood.

The edge feature map F_{edge} is the high frequency information part of the image, and the main body feature map F_{body} is the low frequency information. Thus F_{edge} can be obtained by subtracting the main body feature from the input image feature map. F_{edge} is obtained by subtracting the main body features from the deep features F, which is expressed as

$$F_{edge} = F - F_{body} \tag{3}$$

3.4. Feature Optimization

The main body and edge features decomposed from DF_5 contain more semantic features. which can reflect the main information, but lack detailed features. As a result, the reliability of the boundary information is not adequate. Moreover, up-sampling operation can enlarge the feature map, but cannot increase the total amount of the feature information. These problems will influence the boundary detection of changed buildings. Consequently, a feature optimization module is designed to improve the accuracy of boundary detection by combining multi-scale shallow features. The detail information is added while increasing the size of the feature maps. The feature optimization structure shown in Figure 5 is used to refine both the edge and main body features.

The multi-layer shallow features used in the feature optimization are the dual-temporal multi-scale difference features extracted in the feature extraction module. In the optimization process, the shallow and deep feature maps are stacked layer by layer via skip connections. It can be observed in Figure 3 that DF_1 and DF_2 feature layers have the most copious detailed information. Herein, we first combine the DF_2 feature with $256 \times 64 \times 64$ size to optimize the main body and edge features and generate feature maps optimized once (i.e., F^1_{body} and F^1_{edge} with $256 \times 64 \times 64$ size). Then, the shallowest feature DF_1 is combined by performing a second optimization on the feature map F^1_{body} and F^1_{edge} to increase the detailed information again in a similar manner, which can obtain features optimized twice (i.e., F^2_{body} and F^2_{edge} with $256 \times 128 \times 128$ size). The F^2_{body} and F^2_{edge} are up-sampled to obtain the final feature body and edge maps with $2 \times 256 \times 256$ size. During this process, to reduce the aliasing effect that may exist in the difference feature maps, the shallow feature needs to be dealt with a 3×3 convolution operation before the feature combination, which can enhance the anti-aliasing performance of the network. Moreover, to better combine the feature information of shallow and deep layers, we conduct a 3×3 convolution operation again for the feature maps after the skip connection.

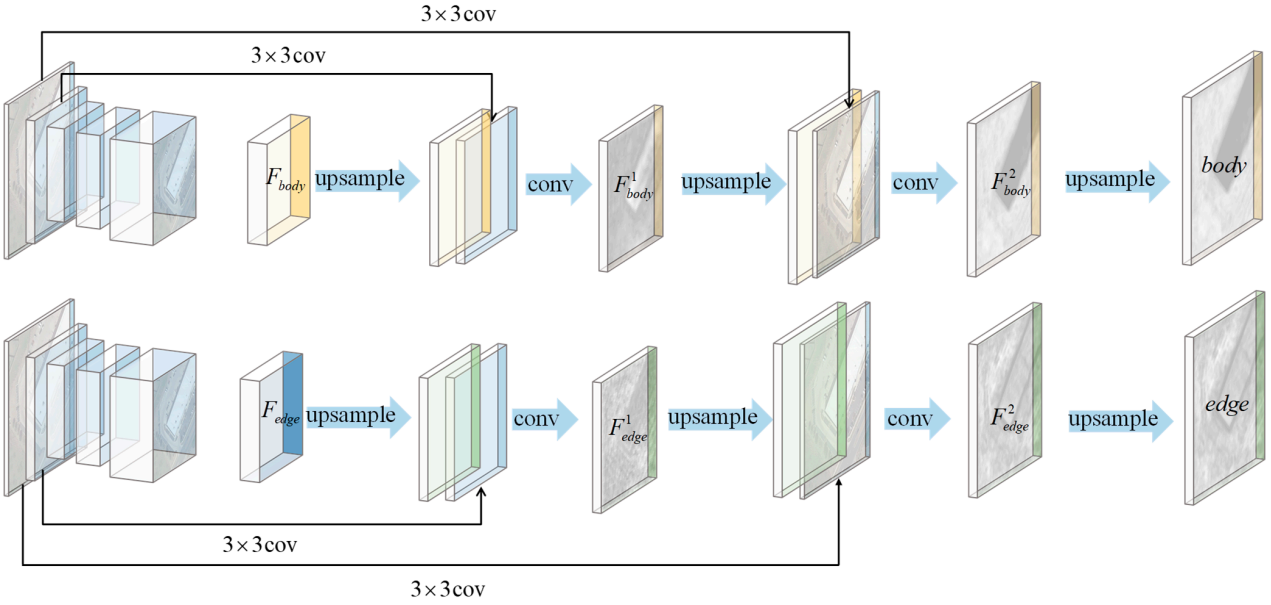


Figure 5. Schematic diagram of feature optimization module.

3.5. Feature Reorganization

Building change detection is aimed at obtaining the final detection result. In the feature optimization part, only the important change information is retained during the precise optimization process, and the unchanged information and pseudo-changed information are filtered out. Feature reorganization is necessary to obtain a single binary map of change detection results. After F^2_{body} and F^2_{edge} are obtained, the reorganized feature map is formed by a direct addition strategy. At this time, the size of the reorganized feature map is $2 \times 128 \times 128$. Subsequently, to ensure that the main body and the boundary of the ground object in the final results are accurately connected during up-sampling of the reorganized feature, the feature maps DF_2 with the 3×3 convolution and up-sampling operation is stacked again to increase the global consistency information. All of the above operations constitute a complete change detection process.

3.6. Loss Function

The existing change detection methods only supervise the final prediction result maps and ignore the interaction between the main body and the edge of the ground buildings. The multiple supervision strategy is adopted to achieve accurate optimization of the main body and edge features in this work. The main body, edge and overall change detection ground truth are used to optimize the main body features, edge features and final prediction results, respectively. The edge label is composed of the outermost pixel of change detection result ground truth. Specifically, each pixel of change detection labels is traversed to determine the category of its 4 neighboring pixels. If there is a difference, the pixel will be identified as the edge, otherwise, it will be identified as the non-edge. The main body label can then be obtained from the change detection label by subtracting the edge label. The loss function is defined as:

$$L = \lambda_1 L_{body}(body, F^{label}_{body}) + \lambda_2 L_{edge}(edge, F^{label}_{edge}) + \lambda_3 L_{out}(out, label) \quad (4)$$

where L_{body} represents the loss of the main body, L_{edge} is the loss of the edge, and L_{out} denotes the loss between the prediction map out and the change detection result ground truth $label$. All three kinds of losses are calculated by using the cross entropy. λ_1, λ_2, and λ_3 respectively denote the weight of the three losses in the total loss L. In this work, the weights of the above three losses are the same, namely, the values of λ_1, λ_2, and λ_3 are all set as 1. F^{label}_{body} and F^{label}_{edge} are the labels of main body and edge.

After the multiple supervised training, the optimization effect of feature optimization module on the main body feature maps and edge feature maps is shown in Figure 6, where the first line is the optimization results of the main body features and the second line is for the edge features. It can be seen in Figure 6a that the main body and edge features formed by feature decomposition alone have fuzzy contours, and it is difficult to give complete ground buildings changed information. As Figure 6b shows, the main body features optimized once (i.e., F^1_{body}) can clearly identify the contours of the changed ground buildings, but the details are lacking. The edge features optimized once (i.e., F^1_{edge}) can find the contour information of some ground buildings, but many boundary lines of the changed buildings are discontinuous, making it difficult to obtain the accurate boundaries. By comparing Figure 6c,d, it can be observed that the main body and edge features optimized twice can obtain more accurate building change information. In other words, the FDORNet model can accurately optimize the main body and edge of buildings through the multiple supervised training.

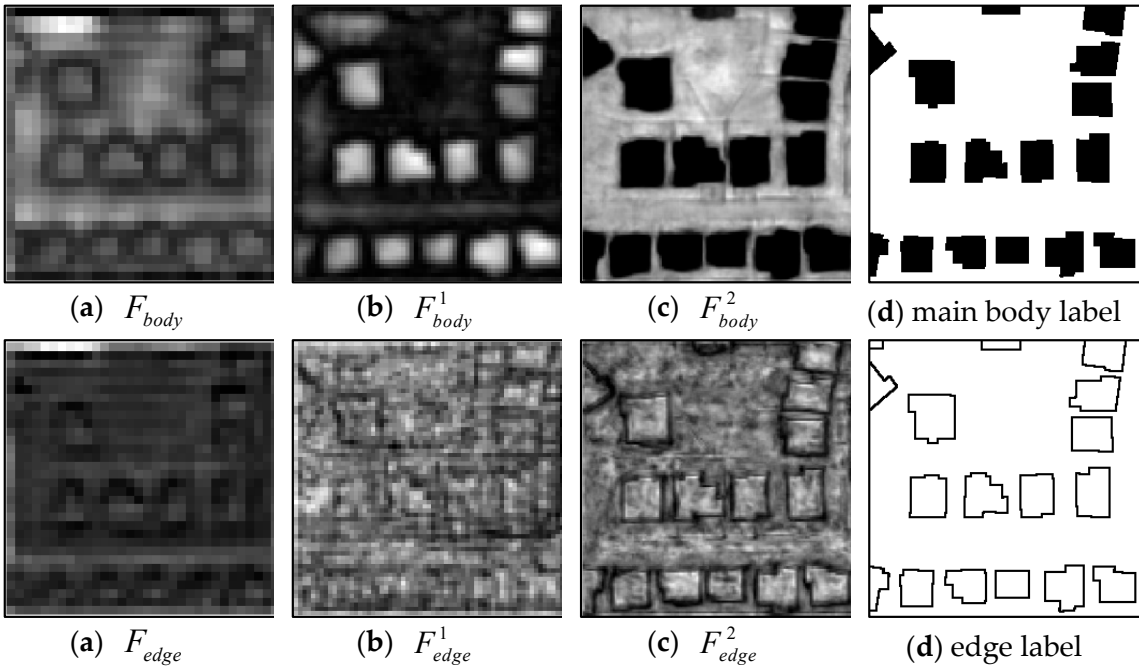


Figure 6. Optimized main body and edge features (**a**–**d**).

4. Experiments

4.1. Dataset

LEVIR-CD, an airborne image building change detection dataset published by Chen et al. [22] in 2020, is adopted in our experiments. The dataset consists of 639 Google Earth images with a size of 1024 × 1024 and a spatial resolution of 0.5 m. The time span between dual-temporal images ranges from 5 to 14 years. The buildings in the dataset include villas, high-rise apartments, small garages and large warehouses. The whole dataset contains a total of 31,333 independent changed buildings, and each image pair contains an average of 50 changed buildings. The changed information is very rich in LEVIR-CD. In this paper, the dataset is divided into the training set, the validation set and the test set according to the data division method of LEVIR-CD. These three sets are composed of 445, 64 and 128 groups of image pairs. A variety of data augmentation methods are adopted, including non-overlapping clipping, random flipping and random rotation. Figure 7 shows some image samples after clipping with a size of 256 × 256 pixels. The sample sizes of the amplified training set, the validation set and the test set are 10,680, 1536 and 2048, respectively.

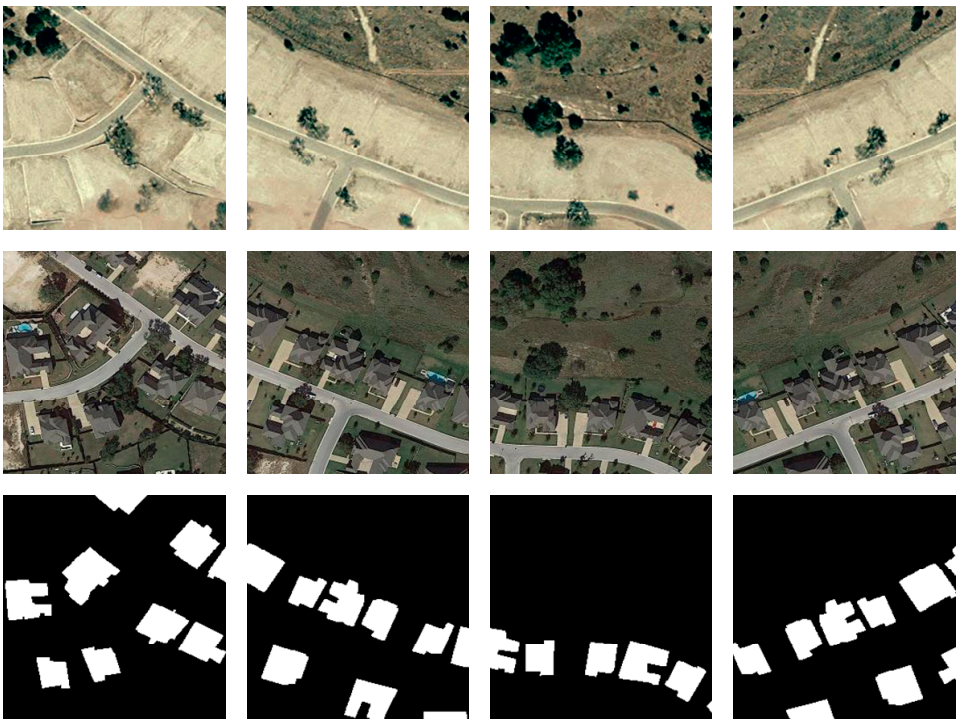


Figure 7. Some image samples after clipping with a size of 256 × 256 pixels. The top two lines are the corresponding images used for building change detection and the last line is the ground truth.

4.2. Experimental Details

In order to highlight the advantages of change detection based on FDORNet and explore the rationality of the network framework, this work designs two groups of experimental comparison schemes. The first group is a comparison of different methods. The FDORNet model is compared with four change detection models based on deep learning, including FC-EF [35], EF-Siam-conc [35], a fully convolutional siamese network combined with a basic spatial-temporal attention model STA-BAM [22] and with pyramid spatial-temporal attention model STA-PAM [22]. The second group is the ablation experiment. The feature optimization module in the FDORNet model is very important for the precise optimization of the main body and edges. To analyze its validity and rationality, the corresponding ablation experiment is performed to test the performance of the feature optimization module. The experimental methods are named FDORNet-base and FDORNet, where FDORNet-base is the FDORNet model without the feature optimization module.

All experiments are based on the Ubuntu 18.04 system. The CPU is Intel(R) Core (TM) i7-10700KF and the GPU is NVIDIA GeForce RTX 3080 with a memory of 10 GB. The deep learning framework used is Pytorch 1.8.0 with a Python version of 3.6. The training epochs are all set as 200.

5. Results and Analysis

5.1. Quantitative Evaluation Cirteria

Quantitative analysis of experimental results is very important and indispensable. We use the confusion matrix to evaluate the performance of the model. Building change detection can be regarded as a binary classification task categorized into changed buildings and background. Its confusion matrix is shown in Table 1, where each column denotes the actual category, and each row represents the predicted category where each pixel belongs. We can obtain overall accuracy, recall, precision, F1-score and mean intersection over union (MIoU) from the confusion matrix.

Table 1. Confusion matrix of building change detection.

		Ground Truth	
Predict	Change buildings	Change buildings	Background
		True Change (TC)	False Background (FC)
	Background	False Change (FB)	True Background (TB)

The overall accuracy represents the ratio of correctly detected pixels to total pixels, usually expressed as a percentage. It is defined as follows:

$$Overall\ accuracy = \frac{TC + TB}{TC + TB + FC + FB} \tag{5}$$

where TC is the number of correctly detected pixels of changed buildings, TB is the correctly detected background pixels, FC represents pixels that belong to background but is misidentified as changed buildings, and FB represents changed building pixels classified in background.

The recall is the metric of accurately predicted changed building pixels from all actual changed building pixels in ground truth:

$$Recall = \frac{TC}{TC + FB} \tag{6}$$

The precision is the metric of the actual changed building pixels predicted to belong to changed buildings:

$$Precision = \frac{TC}{TC + FC} \tag{7}$$

The precision and recall are negatively correlated. To balance the influence of precision and recall and evaluate the model comprehensively, the F1-score is introduced as a comprehensive index:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{8}$$

The MIoU is commonly used to evaluate the efficacy of the model in detecting changing buildings:

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

5.2. Comparison between Different Methods

The results of change detection are analyzed from qualitative and quantitative aspects to evaluate the five deep learning-based change detection methods. In terms of quantitative analysis, the above five indices including overall accuracy, recall, precision, F1-score and MIoU are used to evaluate the performance of change detection methods. For the qualitative evaluation, a detailed analysis is carried out on the detection effect of buildings with general density and higher density in the test set and the detection performance of building boundaries is also compared.

In the quantitative evaluation, Table 2 shows the detection accuracy of the above five deep learning-based change detection methods on the LEVIR-CD dataset. It can be seen that FDORNet obtains the best detection performance and has reached the highest score in all the five evaluation indices. In particular, its precision, F1-score and MIoU are 9.3%, 5.1% and 4.4% higher than those of the method combined with the pyramid spatial-temporal attention model STA-PAM, respectively. For the change detection method of modeling global image information, STA-BAM and STA-PAM obtain better detection performance compared to FC-EF and EF-Siam-conc based on U-Net, but their detection accuracy is still lower than that of FDORNet.

Table 2. Five methods of detection accuracy on the LEVIR-CD dataset. Bold indicates the highest score for each evaluation cirteria.

Methods	Overall Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)	MIoU(%)
FC-EF	96.5459	64.5995	68.7865	66.6273	73.1891
EF-Siam-conc	95.7203	73.4687	57.7948	64.6959	71.6804
STA-BAM	98.0691	87.2022	77.6449	82.1465	83.8408
STA-PAM	98.4768	89.8648	81.9707	85.7364	86.7188
FDORNet	**99.0723**	**90.4158**	**91.2937**	**90.8524**	**91.1335**

In the qualitative evaluation, Figures 8 and 9 show the detection results of building areas with different density (e.g., medium and high density) in the form of color maps and binary change maps, respectively, which can intuitively display the detection effects of the above five change detection methods. Figure 8 illustrates the detection results in the building area with medium density. From left to right are the change detection result labels, the FC-EF detection results, the EF-Sima-conc detection results, the STA-BAM detection results, the STA-PAM detection results, and the FDORNet detection results. The green parts denote the changed areas correctly identified, and the red parts indicate the areas that are incorrectly recognized. As can be seen in the top two lines in Figure 8, FDORNet can better maintain the integrity of the ground object boundary and accurately identify large buildings. The last three lines show that FDORNet can better identify small buildings and describe ground object boundaries in more detail and accuracy.

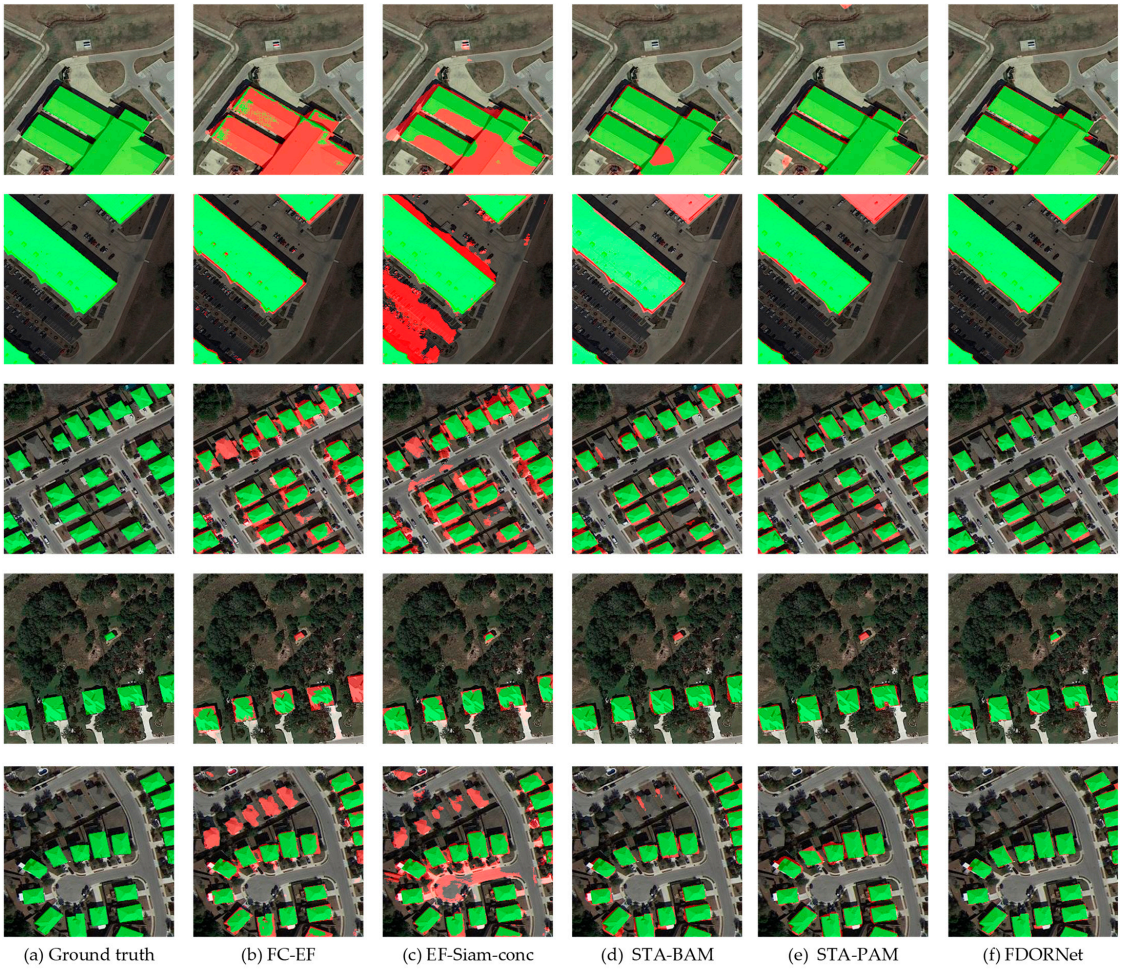
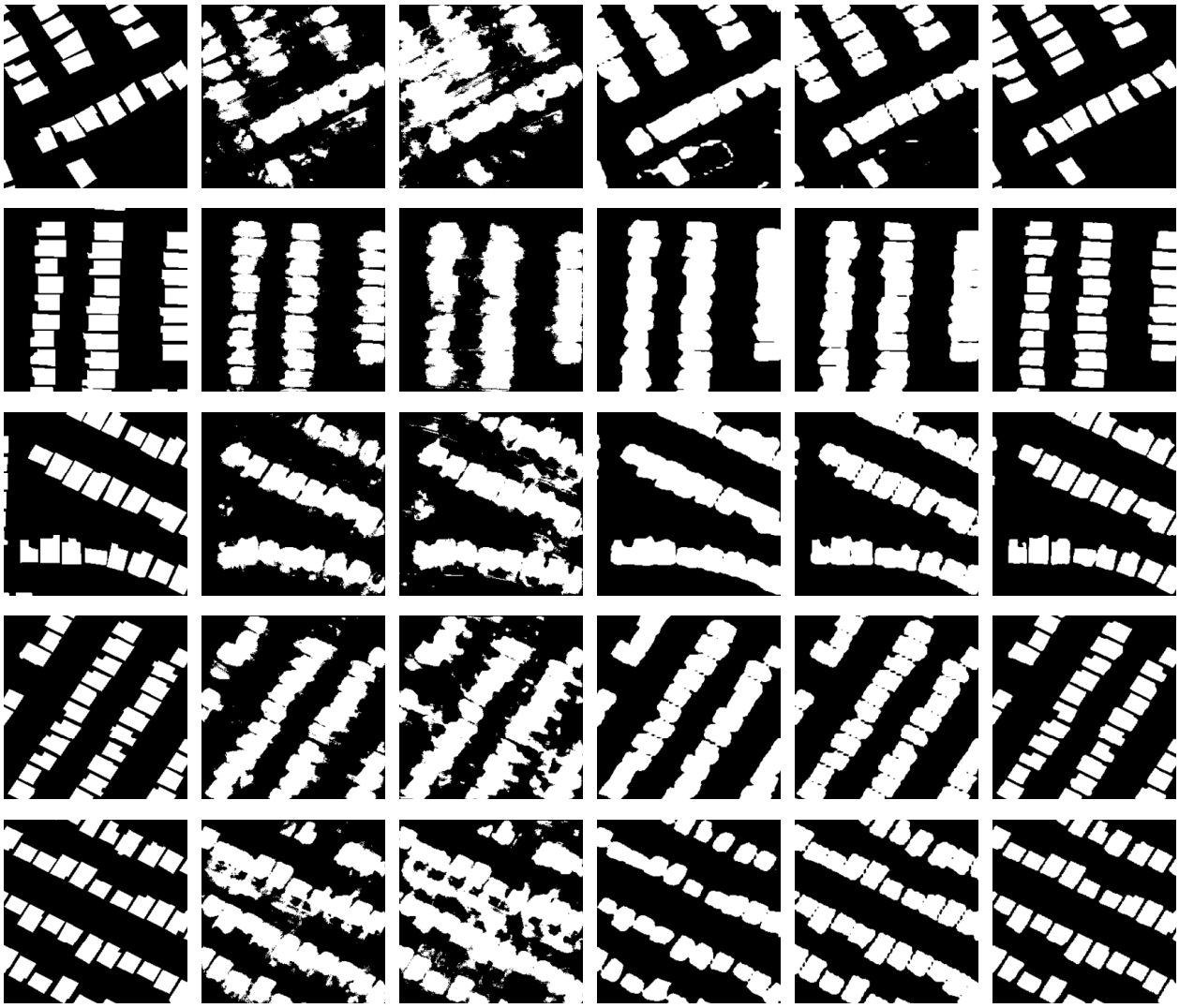


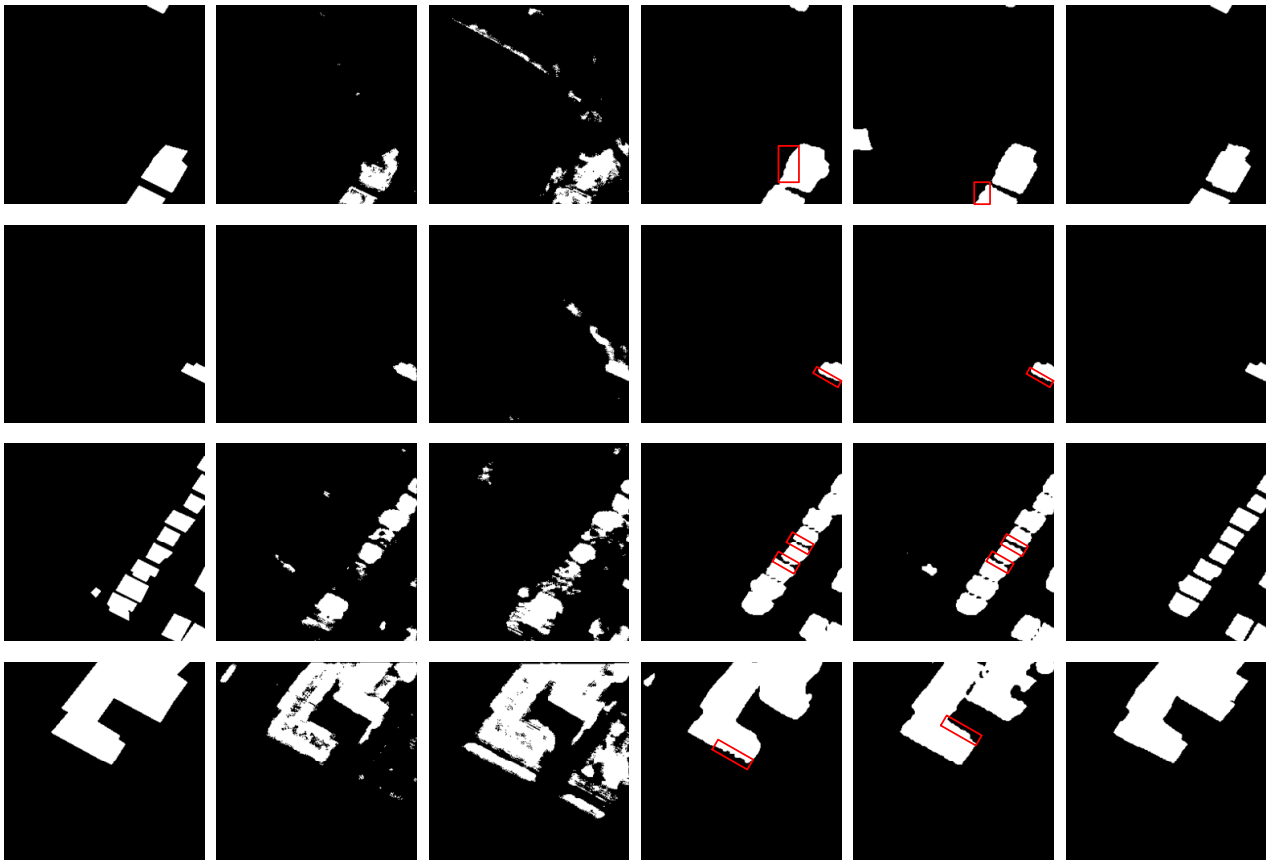
Figure 8. The detection results of the five methods in the medium-density area of buildings, where the green represents the correctly identified areas, and red represents the error-detection areas (**a**–**f**).



(a) Ground truth (b) FC-EF (c) EF-Siam-conc (d) STA-BAM (e) STA-PAM (f) FDORNet

Figure 9. The detection results of the five methods in high-density areas (**a**–**f**).

Figure 9 shows the detection effects of the five methods in the high-density building areas, and Figure 10 compares the accuracy of the five methods in boundary detection of changed buildings. Among them, the detection effects of FC-EF and EF-Siam-conc based on U-Net are observably poor, and it is difficult to identify the gaps between dense buildings. Moreover, the detection results of boundaries are fragmented and have low precision. In contrast, the STA-BAM and STA-PAM methods combined with the spatial-temporal attention model have improved detection performance, but the recognition ability of building gaps is still weak. There are partial connections between buildings and the phenomenon of saw-tooth boundaries (see Figure 10 rectangular box). FDORNet also presents better recognition ability in these areas, and can better maintain the accuracy of ground object contours, and has strong anti-aliasing performance. The above experimental results show that although the method based on the spatial-temporal attention model has certain advantages in improving the performance of change detection, the detection effect in the high-density building areas still needs to be improved further. FDORNet can accurately identify large and small buildings by modeling the main body and edge information of ground objects respectively, and it achieves better performance in maintaining the accuracy of ground object boundaries.



(a) Ground truth (b) FC-EF (c) EF-Siam-conc (d) STA-BAM (e) STA-PAM (f) FDORNet

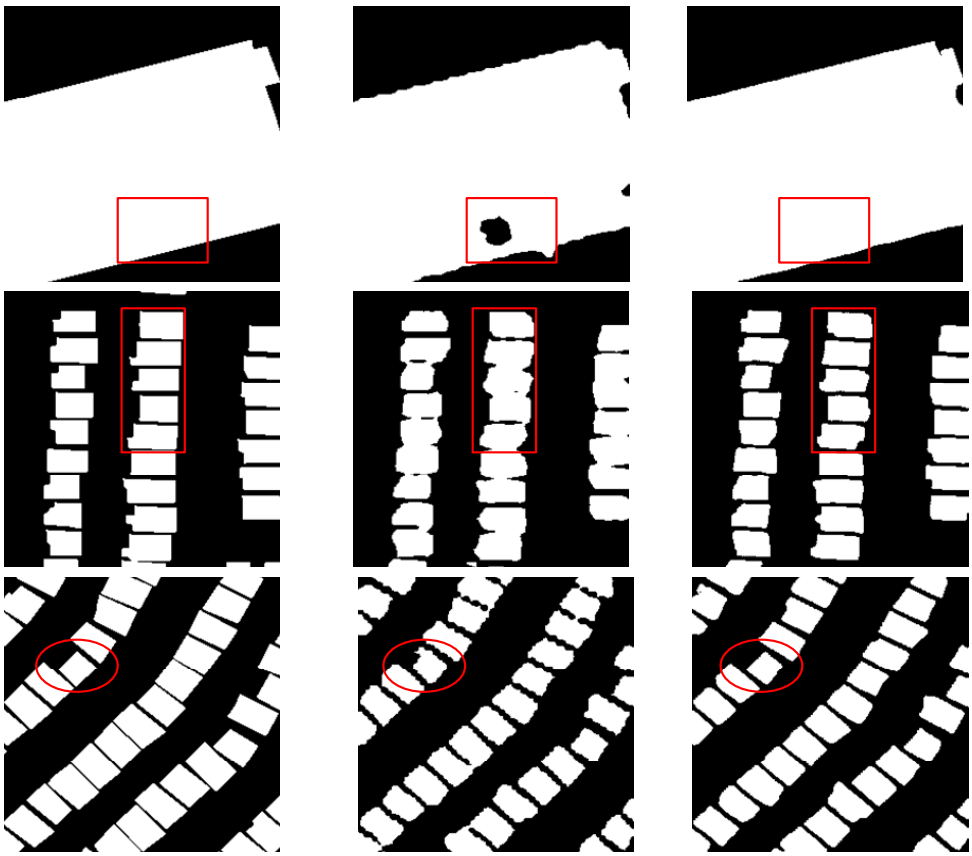
Figure 10. The boundary detection effect of the five change detection methods (**a**–**f**).

5.3. Ablation Experiments

In order to optimize the main body features and edge features accurately, a feature optimization part is designed in the FDORNet model. The ablation experiment is carried out to verify the rationality and efficacy of the design. The experiment compares the FDORNet-base method without the feature optimization module with the FDORNet method using the feature optimization part. The quantitative detection accuracy is shown in Table 3. It is clear that the precision, the F1-score and the MIoU of FDORNet are approximately 2.6%, 2.0% and 1.7% higher than the FDORNet-base, respectively. Specifically, the framework with the feature optimization module can obtain higher detection accuracy. Figure 11 shows the detection results of these two methods, in which the first column illustrates the labels of change detection results, the second column shows the detection results of FDORNet-base, and the last column illustrates the FDORNet detection results. It can be seen that the FDORNet model with the feature optimization module can obtain complete building contours in large building areas, and can better identify small buildings. For both large and small buildings, the saw-tooth phenomenon of boundaries in FDORNet detection results is weak.

Table 3. Detection accuracy of ablation experiments.

Methods	Overall Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)	MIoU(%)
FDORNet -base	98.8611	89.0168	88.6744	88.8452	89.3697
FDORNet	99.0723	90.4158	91.2937	90.8524	91.1335



(a) Ground truth (b) FDORNet-base (c) FDORNet

Figure 11. Results of ablation experiment (**a**–**c**).

6. Conclusions

In order to improve the performance of boundary detection in dense building areas, the idea of feature decoupling is applied to change detection, and a building change detection method based on the feature decomposition-optimization-reorganization network (FDORNet) is proposed in this work. The proposed method includes four modules: feature extraction, feature decomposition, feature optimization and feature reorganization. In the feature extraction module, we extract multi-scale difference features of dual-temporal images. The main body features and edge features are separated by constructing a learnable flow field in the feature decomposition module. Subsequently, in the feature optimization module, multiple supervision strategies are adopted to accurately optimize the main body and edge features by using the corresponding ground truth. Finally, we reorganize optimized features to form a complete building change detection process in the feature reorganization module. The building change detection dataset LEVIR-CD is employed to evaluate the performance of our work. The experimental results show that the FDORNet model can obtain better detection results in the building area compared with the four state-of-the-art methods based on U-Net or combining with spatial-temporal attention models. The diversity of training samples is very important to the change detection performance of the model, but most of the existing datasets are homologous remote sensing images. Future work will additionally make a change detection dataset of remote sensing images from different sensor types to test and further improve the FDORNet model.

Author Contributions: Methodology, Y.Y. and L.Z.; Validation, B.Z., C.Y. and M.S.; Writing—original draft, Y.Y. and L.Z.; Writing—review & editing, Y.Y., J.F. and Z.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 41971281.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://justchenhao.github.io/LEVIR/, accessed on 25 November 2021.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Singh, A. Review article digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* **1989**, *10*, 989–1003. [CrossRef]
2. Radke, R.J.; Andra, S.; Al-Kofahi, O.; Roysam, B. Image change detection algorithms: A systematic survey. *IEEE Trans. Image Process.* **2005**, *14*, 294–307. [CrossRef]
3. Gong, M.; Zhao, J.; Liu, J.; Miao, Q.; Jiao, L. Change detection in synthetic aperture radar images based on deep neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *27*, 125–138. [CrossRef]
4. Mahdavi, S.; Salehi, B.; Huang, W.; Amani, M.; Brisco, B. A PolSAR Change Detection Index Based on Neighborhood Information for Flood Mapping. *Remote Sens.* **2019**, *11*, 1854. [CrossRef]
5. Xian, G.; Homer, C. Updating the 2001 National Land Cover Database impervious surface products to 2006 using Landsat imagery change detection methods. *Remote Sens. Environ.* **2010**, *114*, 1676–1686. [CrossRef]
6. Rokni, K.; Ahmad, A.; Selamat, A.; Hazini, S. Water Feature Extraction and Change Detection Using Multitemporal Landsat Imagery. *Remote Sens.* **2014**, *6*, 4173–4189. [CrossRef]
7. Huang, X.; Zhang, L.; Zhu, T. 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.* **2013**, *7*, 105–115. [CrossRef]
8. Ye, Y.; Bruzzone, L.; Shan, J.; Bovolo, F.; Zhu, Q. Fast and Robust Matching for Multimodal Remote Sensing Image Registration. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9059–9070. [CrossRef]
9. Javed, A.; Jung, S.; Lee, W.H.; Han, Y. Object-Based Building Change Detection by Fusing Pixel-Level Change Detection Results Generated from Morphological Building Index. *Remote Sens.* **2020**, *12*, 2952. [CrossRef]
10. Wang, C.; Wang, X. Building change detection from multi-source remote sensing images based on multi-feature fusion and extreme learning machine. *Int. J. Remote Sens.* **2021**, *42*, 2246–2257. [CrossRef]
11. Cao, S.; Du, M.; Zhao, W.; Hu, Y.; Mo, Y.; Chen, S.; Cai, Y.; Peng, Z.; Zhang, C. Multi-level monitoring of three-dimensional building changes for megacities: Trajectory, morphology, and landscape. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 54–70. [CrossRef]
12. Liu, S.; Ding, W.; Liu, C.; Liu, Y.; Wang, Y.; Li, H. ERN: Edge Loss Reinforced Semantic Segmentation Network for Remote Sensing Images. *Remote Sens.* **2018**, *10*, 1339. [CrossRef]
13. Dechesne, C.; Lassalle, P.; Lefèvre, S. Bayesian U-Net: Estimating Uncertainty in Semantic Segmentation of Earth Observation Images. *Remote Sens.* **2021**, *13*, 3836. [CrossRef]
14. Cheng, G.; Zhou, P.; Han, J. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7405–7415. [CrossRef]
15. Chen, C.; Zhong, J.; Tan, Y. Multiple-oriented and Small Object Detection with Convolutional Neural Networks for Aerial Image. *Remote Sens.* **2019**, *11*, 2176. [CrossRef]
16. He, H.; Chen, M.; Chen, T.; Li, D. Matching of Remote Sensing Images with Complex Background Variations via Siamese Convolutional Neural Network. *Remote Sens.* **2018**, *10*, 355. [CrossRef]
17. Zhou, L.; Ye, Y.; Tang, T.; Nan, K.; Qin, Y. Robust Matching for SAR and Optical Images Using Multiscale Convolutional Gradient Features. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 4017605. [CrossRef]
18. Hou, B.; Wang, Y.; Liu, Q. Change detection based on deep features and low rank. *IEEE Geosci. Remote. Sens. Lett.* **2017**, *14*, 2418–2422. [CrossRef]
19. Saha, S.; Bovolo, F.; Bruzzone, L. Unsupervised deep change vector analysis for multiple-change detection in VHR images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3677–3693. [CrossRef]
20. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
21. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Cham, Switzerland, 2015; pp. 234–241.

22. Chen, H.; Shi, Z. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sens.* **2020**, *12*, 1662. [CrossRef]
23. Chen, J.; Yuan, Z.; Peng, J.; Chen, L.; Huang, H.; Zhu, J.; Liu, Y.; Li, H. DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 1194–1206. [CrossRef]
24. 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.* **2021**, *18*, 811–815. [CrossRef]
25. Li, X.; Li, X.; Zhang, L.; Cheng, G.; Shi, J.; Lin, Z.; Tan, S.; Tong, Y. Improving semantic segmentation via decoupled body and edge supervision. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Springer International Publishing: Cham, Switzerland, 2020; pp. 435–452.
26. Johnson, R.D.; Kasischke, E.S. Change vector analysis: A technique for the multispectral monitoring of land cover and condition. *Int. J. Remote Sens.* **1998**, *19*, 411–426. [CrossRef]
27. Chen, J.; Gong, P.; He, C.; Pu, R.; Shi, P. Land-use/land-cover change detection using improved change-vector analysis. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 369–379. [CrossRef]
28. Bruzzone, L.; Prieto, D.F. Automatic analysis of the difference image for unsupervised change detection. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 1171–1182. [CrossRef]
29. Li, Z.; Shi, W.; Hao, M.; Zhang, H. Unsupervised change detection using spectral features and a texture difference measure for VHR remote-sensing images. *Int. J. Remote Sens.* **2017**, *38*, 7302–7315. [CrossRef]
30. Mishra, N.S.; Ghosh, S.; Ghosh, A. Fuzzy clustering algorithms incorporating local information for change detection in remotely sensed images. *Appl. Soft Comput.* **2012**, *12*, 2683–2692. [CrossRef]
31. Im, J.; Jensen, J.R.; Tullis, J.A. Object-based change detection using correlation image analysis and image segmentation. *Int. J. Remote Sens.* **2008**, *29*, 399–423. [CrossRef]
32. Wang, B.; Choi, S.; Byun, Y.; Lee, S.; Choi, J. Object-based change detection of very high resolution satellite imagery using the cross-sharpening of multitemporal data. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1151–1155. [CrossRef]
33. Liu, J.; Gong, M.; Qin, K.; Zhang, P. A deep convolutional coupling network for change detection based on heterogeneous optical and radar images. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *29*, 545–559. [CrossRef]
34. Zhan, Y.; Fu, K.; Yan, M.; Sun, X.; Wang, H.; Qiu, X. Change detection based on deep siamese convolutional network for optical aerial images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1845–1849. [CrossRef]
35. Daudt, R.C.; Le Saux, B.; Boulch, A. Fully convolutional siamese networks for change detection. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 4063–4067.
36. Liu, J.; Gong, M.; Qin, A.K.; Tan, K.C. Bipartite Differential Neural Network for Unsupervised Image Change Detection. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 876–890. [CrossRef] [PubMed]
37. Mou, L.; Bruzzone, L.; Zhu, X. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 924–935. [CrossRef]
38. Diakogiannis, F.I.; Waldner, F.; Caccetta, P. Looking for Change? Roll the Dice and Demand Attention. *Remote Sens.* **2021**, *13*, 3707. [CrossRef]
39. Zhang, C.; Yue, P.; Tapete, D.; Jiang, L.; Shangguan, B.; Huang, L.; Liu, G. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 183–200. [CrossRef]
40. Marcos, D.; Tuia, D.; Kellenberger, B.; Zhang, L.; Bai, M.; Liao, R.; Urtasun, R. Learning deep structured active contours end-to-end. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8877–8885.
41. Zorzi, S.; Fraundorfer, F. Regularization of Building Boundaries in Satellite Images Using Adversarial and Regularized Losses. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5140–5143.
42. Zorzi, S.; Bittner, K.; Fraundorfer, F. Machine-Learned Regularization and Polygonization of Building Segmentation Masks. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 3098–3105.
43. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, CA, USA, 26 June–1 July 2016; pp. 770–778.
44. Dosovitskiy, A.; Fischer, P.; Ilg, E.; Hausser, P.; Hazirbas, C.; Golkov, V.; van der Smagt, P.; Cremers, D.; Brox, T. Flownet: Learning optical flow with convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 2758–2766.
45. Jaderberg, M.; Simonyan, K.; Zisserman, A.; Kavukcuoglu, K. Spatial Transformer Networks. In Proceedings of the Neural Information Processing Systems 2015, Montreal, QC, Canada, 7–12 December 2015; pp. 2017–2025.