



## Article

# Hierarchical Feature Association and Global Correction Network for Change Detection

Jinquan Lu <sup>1</sup>, Xiangchao Meng <sup>1,\*</sup>, Qiang Liu <sup>1</sup>, Zhiyong Lv <sup>2</sup>, Gang Yang <sup>3</sup> , Weiwei Sun <sup>3</sup> and Wei Jin <sup>1</sup>

<sup>1</sup> Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China; 2111082165@nbu.edu.cn (J.L.); 2001100032@nbu.edu.cn (Q.L.); jinwei@nbu.edu.cn (W.J.)

<sup>2</sup> School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China; zhiyonglyu@xaut.edu.cn

<sup>3</sup> Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo 315211, China; yanggang@nbu.edu.cn (G.Y.); sunweiwei@nbu.edu.cn (W.S.)

\* Correspondence: mengxiangchao@nbu.edu.cn

**Abstract:** Optical satellite image change detection has attracted extensive research due to its comprehensive application in earth observation. Recently, deep learning (DL)-based methods have become dominant in change detection due to their outstanding performance. Remote sensing (RS) images contain different sizes of ground objects, so the information at different scales is crucial for change detection. However, the existing DL-based methods only employ summation or concatenation to aggregate several layers of features, lacking the semantic association of different layers. On the other hand, the UNet-like backbone is favored by deep learning algorithms, but the gradual downscaling and upscaling operation in the mainstream UNet-like backbone has the problem of misalignment, which further affects the accuracy of change detection. In this paper, we innovatively propose a hierarchical feature association and global correction network (HFA-GCN) for change detection. Specifically, a hierarchical feature association module is meticulously designed to model the correlation relationship among different scale features due to the redundant but complementary information among them. Moreover, a global correction module on Transformer is proposed to alleviate the feature misalignment in the UNet-like backbone, which, through feature reuse, extracts global information to reduce false alarms and missed alarms. Experiments were conducted on several publicly available databases, and the experimental results show the proposed method is superior to the existing state-of-the-art change detection models.

**Keywords:** change detection (CD); semantic association; feature reuse; global information; optical satellite images



**Citation:** Lu, J.; Meng, X.; Liu, Q.; Lv, Z.; Yang, G.; Sun, W.; Jin, W. Hierarchical Feature Association and Global Correction Network for Change Detection. *Remote Sens.* **2023**, *15*, 4141. <https://doi.org/10.3390/rs15174141>

Academic Editors: Hongyan Zhang and Xiong Xu

Received: 18 July 2023

Revised: 7 August 2023

Accepted: 16 August 2023

Published: 24 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Change detection, which is dedicated to monitoring the dynamic change of land surface features, plays an increasing role in remote sensing applications [1], such as urban sprawl monitoring [2], forest cover change surveys [3], disaster damage assessment (e.g., landslides, earthquakes) [4,5], and others [6].

### 1.1. Background Studies

To date, there have been proposed various change detection methods, and the existing work can be divided into two major categories: traditional methods and deep learning (DL)-based methods.

- (1) Traditional methods. Traditional change detection methods can be distinguished into arithmetic-operation-based methods [7–11], image-transformation-based methods [12,13], classification-based methods [14], and clustering-based methods [15–17]. The arithmetic-operation-based method obtains the difference map of remote sensing images in different

phases, and then the threshold value is determined to classify the changed and unchanged areas. The typical approach includes the image-difference-based method [7,8], image-ratio-based method [9], and change-vector-based method [10,11], etc. However, the arithmetic-operation-based methods are generally computed at the pixel level and ignore the overall information. The image-transformation-based method increases the image difference by transferring images to feature space through image transformation, such as principal component analysis (PCA) [12] and tasseled cap transformation (KT) [13], and then the final result is obtained through the division of a threshold value. The classification-based method obtains the change detection result on the classification map [14]; however, the accuracy may be affected by the classification results. The clustering-based method clusters the difference map to obtain change detection results. For example, Liu et al. [15] used the typical K-mean clustering, Cui et al. [16] introduced fuzzy *c*-means (FCM) clustering, and Shao et al. [17] proposed a new fuzzy clustering change detection method. Most of the clustering-based methods consider the spectral information of the image but ignore the spatial texture information. In general, traditional methods with simple and fast features are in demand in most applicable cases; however, the accuracy is barely satisfactory.

- (2) Deep-learning-based methods. DL-based change detection methods have attracted attention in recent years. Most existing DL-based methods are established on convolutional neural networks (CNN) [18–20] or Transformer networks [21,22]. The characteristics of CNN networks brought by their convolutional operators enable them to extract rich local detail information, and, due to the characteristics brought by the cascade of convolutional layers, they extract detail-rich information and abstract information with semantic associative properties. Transformer can be good at extracting global information and has a nonlinear fitting capability, no less than that of the CNN. Most deep learning methods achieved competitive results [23–25], and analysis of existing DL-based change detection methods to further improve their performance is necessarily expected.

### 1.2. Analysis of Existing Deep-Learning-Based Methods

In change detection tasks, the fundamental task is to determine whether the semantic state of a feature changes in different images; though there are various implementations of existing methods, it is possible to analyze the methods which are effective for change detection tasks, to learn the advantages and improve the shortcomings.

#### 1.2.1. CNN

There are many CNN-based change detection methods. Here, we summarize some useful structures and methods for change detection from existing CNN-based algorithms.

- (1) Feature extraction on a Siamese network

Feature extraction by convolutional neural networks is the first and critical step of the CNN-based change detection algorithms. In early CNN-based change detection algorithms [18], bitemporal RS images are generally stacked in the channel dimension and fed into the feature extraction network by a single-stream network without a dedicated variation difference analysis module, and the feature extraction network has to undertake the tasks of image feature extraction and difference analysis at the same time, which makes the feature extraction network overburdened and unable to achieve the best feature extraction effect. The emergence of the Siamese network structure can effectively distinguish feature extraction from feature analysis to achieve a better feature extraction effect, that is, the bitemporal RS images are separately extracted by two feature extraction networks with the same structure and the same parameter weights, such as is described by Fang et al. [26] and Liu et al. [27] The extracted features are constrained by using the Siamese structure, so that the bitemporal RS images feature information can be mapped to different feature levels, and thus the subsequent difference analysis module can have higher efficiency and thus

improve the accuracy of the final change detection results. The feature extraction structure of the Siamese network is suitable for the change detection task for feature extraction.

## (2) Multi-scale feature utilization

Because the causes of feature changes are diverse, it is characteristic that the size of the changing area in the image is inconsistent, sometimes with huge differences in scale size, and this phenomenon is inevitable even in the same type of change. Therefore, the integrated analysis of the information at different scales is important in change detection tasks. With the stacking of convolutional layers, the perceptual field of the convolutional layers keeps increasing, while the features of different layers have different perceptual field sizes; thus, different feature layers of the feature extraction network can be regarded as the features of the image at different scales [28,29]. The extraction of features by using convolution operators with different convolution kernel sizes at the same level at the same time also allows us to obtain feature information at different scales, such as is described by Song et al. [30] and Lv et al. [31,32], who designed the multiscale feature extraction module to obtain different multiscale features. Information at different scales is crucial for the analysis of change detection; however, how to fully and effectively utilize the information of features at different scales and fully explore the information belonging to change regions in features at different scales has not received sufficient attention. Therefore, for this paper, a hierarchical feature association module was designed to address this shortcoming.

## (3) Global information utilization

Global information plays an important role in the determination of semantic information in remote sensing images. For example, if the semantic meaning of 'concrete' is determined by the analysis of local information, it will be difficult to analyze the true semantic meaning of 'roof' or 'highway' without the analysis of global information. However, the type of mutual variation between 'roof' and 'highway' exists; thus, it is evident that without the help of global information, deep learning algorithms will produce certain misjudgments and omissions. To complement the global information desired about the features, existing algorithms introduce global information in the channel or spatial dimension by combining attention mechanisms, such as is described by Zhang et al. [33]. Most methods obtain global information by extending the convolutional field or using the attention mechanism; therefore, this study designed a Transformer-based module to obtain global information, which differs from the method of introducing global information through the CNN method and makes the result of change detection more accurate.

### 1.2.2. Transformer

Transformer has been widely used in the field of natural language processing because of its ability to model global dependencies, which corresponds to the demand for modeling ability in the field of natural language processing. Due to the benefits from its modeling capability, which is no less than that of the convolutional neural networks, the computer vision field has focused on using Transformer model networks for better performance, in areas such as image classification, semantic segmentation, target detection, etc. Transformer has achieved excellent results. In the change detection task, the bitemporal image transformer (BIT) model [34], which models the spatial-temporal context based on features extracted by a CNN, is followed by global information learning through two Transformer encoders, and then a change detection map is obtained by a Transformer decoder with a prediction network. The ChangeFormer model [35], which exploits the excellent global feature extraction capability of Transformer as a feature extraction backbone network, can be followed by a simple prediction head to obtain change detection results. The Intra-Scale Cross-Interaction and Inter-Scale Feature Fusion Network (IFIC) model [36], by combining the CNN and Transformer for extracting features, achieved excellent performance through the interaction of detailed and global information at different scales. By using Transformer, these algorithms achieved good results, but the nature of the global information introduced by Transformer was not considered in these algorithms, so our understanding of how to

effectively use Transformer to introduce the global information needed for change detection tasks still needs to be improved.

### 1.3. Challenges

- (1) Challenge 1: How to make full use of the information among different scale features.

Information at different scales is crucial for change detection; however, how to fully explore their complementary information among different scale features has not received sufficient attention. For example, most existing DL-based methods perform change detection analysis directly on the features at different scales and then obtain the results by simply aggregating the features at different scales via summation or concatenation. This simple summation or concatenation is insufficient for describing the complementary information between different levels since the features in different levels represent different specific information. To enhance the information representation so that the information contained in different hierarchical features can be fully utilized, modeling the association relationships among different hierarchical features is necessary.

- (2) Challenge 2: How to alleviate the feature misalignment.

In the change detection task, most existing DL-based methods employ a UNet-like backbone to extract features of different scales via a gradual downscaling operation, and then reconstruct the detection results through a gradual upscaling operation. These gradual downscaling and upscaling operations have the problem of feature misalignment, which further affects the accuracy of change detection at the decision level. Therefore, an alleviation of feature misalignment method is necessary.

### 1.4. Contribution

To address the above challenges, a hierarchical feature association and global correction network (HFA-GCN) is proposed. Specifically, to overcome challenge 1, a hierarchical feature association (HFA) model is proposed to model the relationship between different layer features so that different layer features can be used more effectively. To address challenge 2, a global correction (GC) is designed to extract global information about bitemporal RS images based on the Transformer structure and to correct change detection features by the interaction of global information to mitigate feature misalignment. The contributions of this paper can be summarized as follows.

- (1) A hierarchical feature association and global correction network, namely, HFA-GCN, is proposed for change detection. The HFA is designed to model the association relationships between hierarchical features, so that the different hierarchical features can be fully utilized. The GC is designed to extract global information more efficiently, and alleviate the feature misalignment.
- (2) HFA-GCN, by modeling the correlations of hierarchical features at different levels, can enhance the information at different levels, making it easier for change detection tasks to obtain change information at different levels; the innovative global information extraction and utilization method makes the global information effective for change detection tasks. Therefore, HFA-GCN obtains excellent performance.

The rest of the paper is organized as follows. The details of the proposed model are presented in Section 2. Section 3 shows the experiments. The last section concludes the paper.

## 2. Methodology

In this paper, we proposed a hierarchical feature association and global correction network (HFA-GCN) for change detection. Unlike systems that combine different levels of features by simple aggregating or concatenation, the HFA-GCN models the correlation between different levels to better obtain information about the changes contained in different levels. Bypassing the method of introducing global information by CNNs, the HFA-GCN better extracts the global information about the image based on Transformer

and uses the global information to correct the effect of feature misalignment in a UNet-like network, to achieve better change detection results. The HFA-GCN is shown in Figure 1. For a pair of bitemporal RS images to be detected, the RS images are first passed through a feature extraction network consisting of residual structures to obtain the multi-level feature maps. To better utilize the extracted features at different levels, we propose a hierarchical feature association (HFA) module to model the relationship between feature information at different levels to better utilize the feature maps at different levels. Using the multi-level feature maps after modeling the relationship with neighboring order features, the change detection results to be corrected are generated by a simple decoder. Usually, the change detection framework has completed the whole process, but in this paper, we describe the global correction module used to extract global information about the features using global mapping, correct the change detection features through the interaction of the global information, and then get the final change detection results through the change detection decision maker.

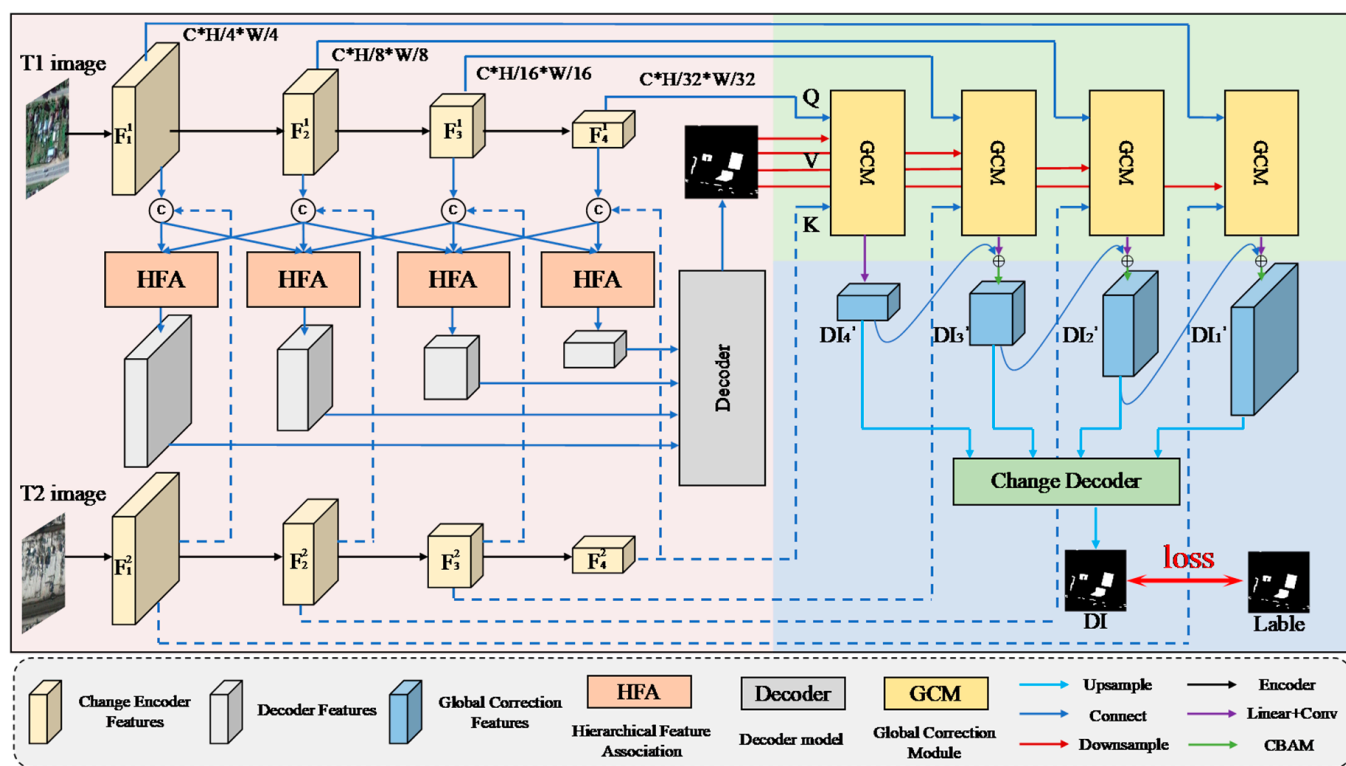


Figure 1. The framework of the hierarchical feature association and global correction network.

### 2.1. Baseline Backbone Network

The baseline backbone of HFA-GCN is a UNet-like network consisting of encoders and decoders.

For the encoder, to efficiently extract feature information from the different layers of the image, a feature extraction module was designed based on the idea of residual connectivity. The module, as a whole, consists of two convolutional modules with one residual connection; the first convolutional module is the feature extraction module, which is used to extract higher-level feature information from the image. The second convolutional module extracts further information from the current level of feature information to obtain more effective feature information. A residual connection is added after the first convolutional module is complete to ensure the validity of subsequent further information extraction and prevent feature degradation. Finally, a weight-sharing feature extraction module is used for bitemporal RS images to constrain the extracted bitemporal features to be projected into a

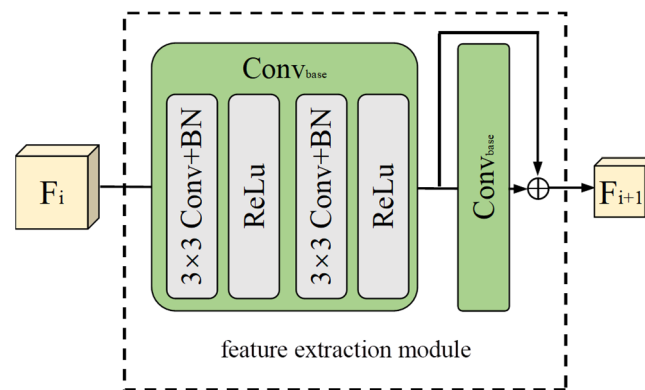
consistent feature space, which is more conducive to the analysis of change detection tasks. For the feature extraction module, as shown in Figure 2, the expressions are as follows.

$$\text{Conv}_{\text{base}}(F_i^t) = \text{ReLU}(\text{BN}(\text{Conv}(\text{ReLU}(\text{BN}(\text{Conv}(F_i^t))))) \quad (1)$$

$$\text{FE}(F_i^t) = \text{Conv}_{\text{base}}(\text{Conv}_{\text{base}}(F_i^t)) + \text{Conv}_{\text{base}}(F_i^t) \quad (2)$$

$$F_{i+1}^t = \text{FE}(F_i^t) \quad (3)$$

where Conv, BN, ReLU represent the two-dimensional convolution operation with a convolution kernel size of  $3 \times 3$ , batch normalization, and ReLU activation, respectively. Equation (1),  $\text{Conv}_{\text{base}}$ , represents a basic convolution block consisting of two basic convolution operations. The first basic convolution is designed to reduce the feature scale to increase the number of feature channels, and the second basic convolution is designed to extract more feature information from the features at this level.  $F_i^t$  is the feature of the input feature extraction module, where  $t = 1, 2$  represents two input RS images at different moments, with  $i = 1, 2, 3, 4$  representing different levels of feature information, when  $i = 1$  is the original input image. With the increasing number of layers, the number of extracted feature channels also increases, respectively, as  $C = 3, 32, 64, 128, 256$ . The scale of the opposite feature decreases, respectively, as  $H = W = 256, 128, 64, 32, 16$ .



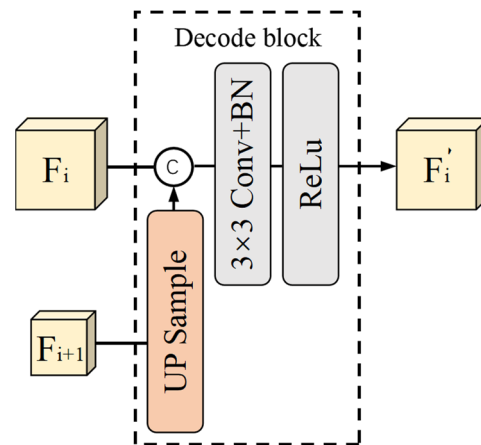
**Figure 2.** Detail of the feature extraction module.

For the decoder, as shown in Figure 3, to efficiently use the front module to generate different scale features, the proposed decoder is progressive for feature decoding. The decoder reconstructs the change detection features by progressively decoding them by simultaneously inputting the feature information of the current level and the previous level. In the decoder, after first upsampling the current level features to the previous level feature size, the change detection features of the previous level are reconstructed by the two levels of features together. The expression of this module is as follows.

$$\text{Decoder}(F_i', F_{i-1}') = \text{ReLU}(\text{BN}(\text{Conv}(\text{Cat}(\text{UP}(F_{i-1}'), F_i')))) \quad (4)$$

$$F_{i-1}' = \text{Decoder}(F_i', F_{i-1}') \quad (5)$$

For the input features of two different levels, using Equation (4), we find that the  $\text{UP}(\cdot)$  operation performs twice the interpolation upsampling for the feature with a small scale, and then the two scales are superimposed on the channel by the  $\text{Cat}(\cdot)$  operation to concatenate the change detection features of the scale size of the previous level by using the two-level features.



**Figure 3.** Detail of Decode block.

## 2.2. Hierarchical Feature Association Module

In remote sensing images, the semantic information about the same feature observed at different scales may be different, but this semantic information is not unrelated. By associating different pieces of semantic information about features at different scales, we can derive the true semantics of the features more accurately. In this study, we designed the hierarchical feature association module to correlate the semantic information about features at different levels by modeling the correlation relationships with neighboring features, to increase the effective representation of feature information at the current level.

As shown in Figure 4, the input of the hierarchical feature association module is the features at the current level and the neighboring levels, and the association weights of information between features are extracted at the spatial level using the structure of spatial attention in the module. For the lower-level features, they are firstly downsampled to make their sizes consistent, and then associated with the current-level features after extracting their association weights by spatial attention; for the higher-level features, they are firstly upsampled by bilinear interpolation to make their sizes consistent, and then associated with the current-level features after extracting their association weights by spatial attention. The current-level features are taken as the main feature information, and the association weight matrices are extracted to the current feature level at four different scales; finally, all the association weight matrices are augmented for the current features to realize the modeling of the neighbor-order feature association relationship, and the final output of the module is obtained. The expressions of the module are as follows.

$$f_{k \times k}(F_i) = \text{ReLU}(\text{BN}(\text{Conv}_{k \times k}(F_i))) \quad (6)$$

$$\text{SA}(F_i) = \text{Sigmoid}(\text{Conv}(\text{Cat}(\text{Maxpool}(F_i), \text{Avepool}(F_i)))) \quad (7)$$

$$\text{SA}_{\text{cur}}(F_i) = \text{SA}(f_{1 \times 1}(F_i)) + \text{SA}(f_{3 \times 3}(F_i)) + \text{SA}(f_{5 \times 5}(F_i)) + \text{SA}(f_{7 \times 7}(F_i)) \quad (8)$$

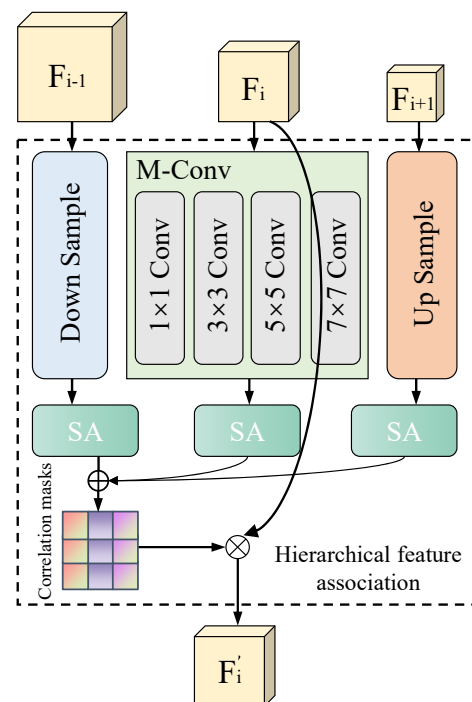
$$\text{SA}_{\text{lat}}(F_{i+1}) = \text{SA}(\text{UP}(F_{i+1})) \quad (9)$$

$$\text{SA}_{\text{pre}}(F_{i-1}) = \text{SA}(\text{Down}(F_{i-1})) \quad (10)$$

$$F'_i = (\text{SA}_{\text{pre}}(F_{i-1}) + \text{SA}_{\text{cur}}(F_i) + \text{SA}_{\text{lat}}(F_{i+1})) \cdot F_i \quad (11)$$

where  $\text{Conv}_{k \times k}$ ,  $\text{BN}$ ,  $\text{ReLU}$  represent the two-dimensional convolution operation with a convolution kernel size of  $k \times k$ , batch normalization, and  $\text{ReLU}$  activation, respectively. In Equation (7),  $\text{Maxpool}(\cdot)$ ,  $\text{Avepool}(\cdot)$  represent the maximum pooling operation on the

channel and the average pooling operation on the channel, respectively. The number of feature channels output after maximum pooling and average pooling is two, while the remaining dimensions remain unchanged. The  $\text{Cat}(\cdot)$  represents a stacking of different features in the dimension of the channel. In Equation (9),  $\text{Up}(\cdot)$  represents an interpolation upsampling of the feature map to make the scale twice the original scale. In Equation (10), the  $\text{Down}(\cdot)$  represents a downsampling of the feature map at twice the scale using the  $2 \times 2$  maximum pooling. In Equations (6)–(11),  $F_i$  is the result of superimposing the  $F_i^1$   $F_i^2$  features of the  $i$  level on the channel, and the matrix addition and matrix dot multiplication in the formula represent the addition and multiplication symbols in Figure 4, respectively, and  $F'_i$  is the feature of  $F_i$  corresponding to the level  $i$  after semantic association by the hierarchical feature association module.



**Figure 4.** Detail of hierarchical feature association module.

### 2.3. Global Correction Module

The change detection task is determined by the semantic state of the features, and essentially compares whether the global semantic information of bitemporal RS images has changed, so emphasizing the global information about the features is powerful for solving the remote sensing image change detection task. There are frequent up-and-down sampling operations of features in the process of change detection, which can also lead to errors due to misalignment between features.

As shown in Figure 5, the global correction module is proposed to solve the above problems. Using the global information about image features with rich change information, the change detection features are corrected to mitigate the errors caused by the misalignment of features, and the absence of global information. The inputs of the module are the features corresponding to the bitemporal RS images and the change detection features' downsampling. The value  $V$  is mapped from the initial change detection feature downsampling, while  $Q$  and  $K$  are mapped from the image features. The interaction of  $Q$  and  $K$  realizes the interaction of the global information of two temporal features, and the change detection features are corrected by the interaction results, and, finally, the corrected change detection features are obtained, which improves the accuracy of the overall network. The expressions of the module are as follows.



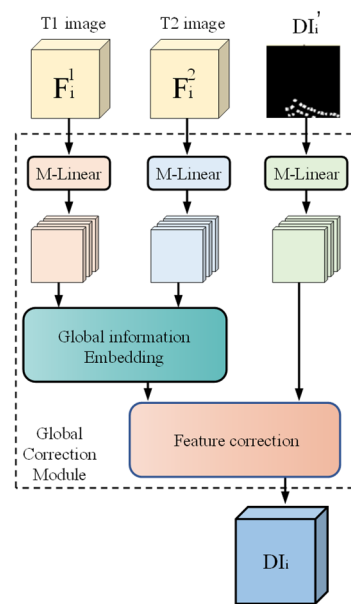
$$K = \text{Linear}(F_i^1), Q = \text{Linear}(F_i^2), V = \text{Linear}(DI_i') \quad (12)$$

$$\text{Gloal Attn map}(Q, K) = \text{Softmax}(Q^T \cdot K) \quad (13)$$

$$\text{Gloal Attn}(Q, K, V) = \text{Gloal Attn map}(Q, K) \cdot V / \sqrt{d} \quad (14)$$

$$DI_i = \text{Gloal Attn}(Q, K, V) \quad (15)$$

where  $F_i^1$  and  $F_i^2$  represent the features corresponding to the bitemporal RS images at the level  $i$  extracted by the feature extraction module, respectively. The initial features are reused to obtain more accurate information. In Equation (12),  $\text{Linear}(\cdot)$  is a linear mapping of the input features to an intermediate state with global information; through the interaction of  $Q$  and  $K$ , the global information about the bitemporal feature map can fully interact to obtain more accurate global information about the feature, and the attention map containing the global information is corrected for the change detection feature  $DI_i'$  features through the operation of  $\text{Gloal Attn}(Q, K, V)$  to generate a more accurate change detection feature map.



**Figure 5.** Detail of global correction module.

#### 2.4. Change Detection Module

For the change detection module, as shown in Figure 6, to alleviate the fusion difficulties caused by the existence of different scales of features at different levels, a simple information fusion is performed for the features at different levels after compressing the channels, and then the change detection results are reconstructed by high-level and low-level features. The expressions of this module are as follows.

$$DI_i' = \text{CBAM}(\text{UP}(DI_i), +\text{Conv}(\text{Liner}(DI_{i-1}')))) \quad (16)$$

$$\text{FFM}(DI_i, DI_{i+1}) = \text{CBMA}(\text{Cat}(DI_i', DI_{i+1}')) \quad (17)$$

$$\text{Change Decoder}(DI_1, DI_2, DI_3, DI_4) = \text{FFM}(\text{FFM}(DI_1, DI_2), \text{FFM}(DI_3, DI_4)) \quad (18)$$

$$DI = \text{Change Decoder}(DI_1, DI_2, DI_3, DI_4) \quad (19)$$

where  $DI'_i$  represents the change detection features present after the global correction module,  $i = 1, 2, 3, 4$  represent the features of different levels, and  $DI_i$  represents the temporary change detection feature, which is generated by combining advanced semantic information for the input features. We first compress its number of channels by linear mapping, then pre-fuse it with the features of different levels by using a CBAM, and then reconstruct the change detection result  $DI$  by using the CBAM to fuse the features twice.

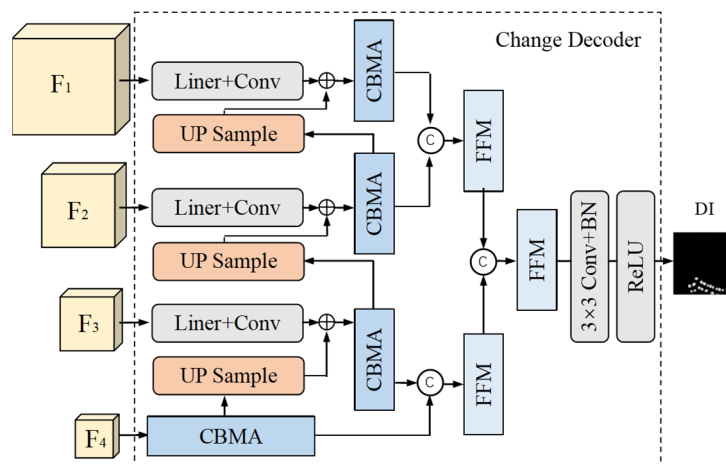


Figure 6. Detail of change detection module.

### 3. Experiment and Analysis

Three benchmark remote sensing (RS) datasets were utilized as opening references to assess the performance of HFA-GCN. First, the experimental datasets were described. Secondly, the details of the experiments were given. After that, we compared and analyzed the classification results of HFA-GCN with different comparison methods. Finally, we conducted ablation experiments to demonstrate the efficacy of various modules.

#### 3.1. Experimental Datasets

##### (1) CDD dataset

The CDD dataset is an early publicly available remote sensing image change detection dataset consisting of 7 pairs of  $4725 \times 2700$  and 4 pairs of  $1900 \times 1000$  images, which are divided into 10,000 (training)/3000 (validation)/3000 (testing) dataset image pairs after being cut into  $256 \times 256$ -size image pairs in a non-overlapping manner. In the experiment, we only used the real remote sensing image pairs, which change with the seasons.

##### (2) LEVIR-CD dataset

LEVIR-CD is a large publicly available building change detection dataset with a time interval of 5–12 years between image pairs, which has change types ranging from large areas such as apartments and villas to small areas such as small garages. The dataset includes a total of 637 pairs of  $1024 \times 1024$  optical remote sensing images (0.5 m). For the original divided training, testing, and validation test sets, we cut the images into  $256 \times 256$  non-overlapping image pairs, as 7120 (training)/1024 (validation)/2048 (testing) images.

##### (3) GZ-CD dataset

GZ-CD is a publicly available remote sensing image change detection dataset for detecting urbanization in the Guangzhou suburbs, and the images contain 19 pairs of images taken in the suburbs of Guangzhou, with image pair sizes ranging from  $1006 \times 1168$  to  $4936 \times 5224$ . Due to the lack of publicly available definite training/testing dataset

divisions, we split the original images into  $256 \times 256$  image pairs without overlap, and randomly divided them into 2194 (training)/936 (test).

### 3.2. Experimental Setup

To confirm the effectiveness of our designed method, we chose several SOTA models for comparison, including three full convolutional models: full convolutional early feature fusion (FC-EF) [2], full convolutional twin difference network (FC-Siam-Di) [2], and full convolutional twin cascade network (FC-Siam-Conv) [2]. For these three methods, the encoder–decoder paradigm was modified into a Siamese architecture, and by using skip connections and different decoders, three results were obtained. The three attention-based methods are the deeply supervised image fusion network (SNUNet) [26], which uses densely connected, deeply supervised, and attention-blocking networks to achieve change detection; the deeply supervised image fusion network IFNet [33], which uses attention modules fused with multi-level deep features with image difference features; and the dual-task constrained deep Siamese convolutional network (DTCDCSN) [27], which introduces a dual attention module (DAM) to exploit the interdependencies between channels and spatial positions, improving feature representation. A CNN has been combined with the Transformer network (BIT) [34], which uses a transformer encoder to model contexts in space–time.

Our proposed model was implemented in the PyTorch framework, and the model was trained and tested on an NVIDIA 3090 GPU (24 G). To iterate the network parameters, we used the AdamW algorithm as an optimizer, with batch size 16, and used the combination of weighted cross-entropy loss and dice loss as the loss function. Cross-entropy loss can measure the gap between the model’s predicted results and the actual results in classification problems, and dice loss is good for solving the problem posed by unbalanced numbers in categorized samples. The expressions of cross-entropy loss and dice loss are expressed in Equations (20) and (21), respectively. The number of training iterations is 400, trained 100 times each at learning rates (0.001, 0.0005, 0.00025, 0.0001).

$$\text{Cross entropy loss} = - \sum_{i=1}^C x_i \log y_i \quad (20)$$

$$\text{Dice loss} = 1 - \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (21)$$

where  $x_i$  denotes the  $i$ th element of the true label and  $y_i$  denotes the probability that the model predicts that  $x$  belongs to the  $i$ th category, and  $TP$ ,  $TN$ ,  $FP$ , and  $FN$  stand for judging change correctly, judging no change correctly, judging a change region incorrectly, and judging no change region incorrectly.

### 3.3. Evaluation Indicators

To evaluate the results obtained by different methods, we used five common evaluation metrics to measure the degree of similarity between the result graphs and the labels, including precision, recall, F1 score, joint intersection, and overall accuracy. Each metric is defined as follows.

$$\text{Precision} = TP / (TP + FP) \quad (22)$$

$$\text{Recall} = TP / (TP + FN) \quad (23)$$

$$F1 = 2 / \left( \text{Recall}^{-1} + \text{Precision}^{-1} \right) \quad (24)$$

$$IOU = TP / (TP + FP + FN) \quad (25)$$

where  $TP$ ,  $TN$ ,  $FP$ , and  $FN$  stand for judging change correctly, judging no change correctly, judging a change region incorrectly, and judging no change region incorrectly, respectively.  $F1$  is a comprehensive index, which consists of both *precision* and *recall*.

### 3.4. Experimental Results

To evaluate the effectiveness of the proposed network model, the results of a large number of experiments were quantitatively summarized in three diachronic remote sensing images, and the performance of different networks in the three datasets CDD, LEVIR-CD, and GZ-CD is reflected in Tables 1–3. To visualize the performance of different network models on different datasets, we visualized the results of all methods by representing  $TN$  as white,  $TP$  as black,  $FN$  as red, and  $FP$  as green, so that we can more intuitively observe the performance advantages and disadvantages of different methods. This allows us to more intuitively reflect the effectiveness of our semantic association, feature reuse, and global information introduction operations.

**Table 1.** Performance of various algorithms and HFA-GCN on CDD dataset.

CDD	Pre (%)	Recall (%)	F1 (%)	IOU (%)
FC-EF(2019)	74.12	55.86	63.70	46.74
FC-Siam-Di(2019)	82.23	54.33	65.43	48.62
FC-Siam-Conv(2019)	77.68	58.63	66.82	50.18
IFN(2020)	96.05	97.01	96.53	93.29
SNUNet-CD/32(2022)	96.14	95.90	96.02	92.34
DTCDSN(2020)	94.98	92.66	93.80	88.33
BIT(2021)	94.72	96.44	96.58	93.38
HFA-GCN	97.40	97.20	97.30	94.75

**Table 2.** Performance of various algorithms and HFA-GCN on LEVIR-CD dataset.

LEVIR-CD	Pre (%)	Recall (%)	F1 (%)	IOU (%)
FC-EF(2019)	78.95	74.82	76.83	62.38
FC-Siam-Di(2019)	87.07	67.10	75.79	61.02
FC-Siam-Conv(2019)	87.14	66.64	75.52	60.67
IFN(2020)	88.45	90.62	89.52	81.03
SNUNet-CD/32(2022)	90.12	89.17	89.64	81.23
DTCDSN(2020)	90.26	87.66	88.94	80.09
BIT(2021)	92.20	87.88	89.99	81.80
HFA-GCN	91.49	89.99	90.73	83.04

**Table 3.** Performance of various algorithms and HFA-GCN on GZ-CD dataset.

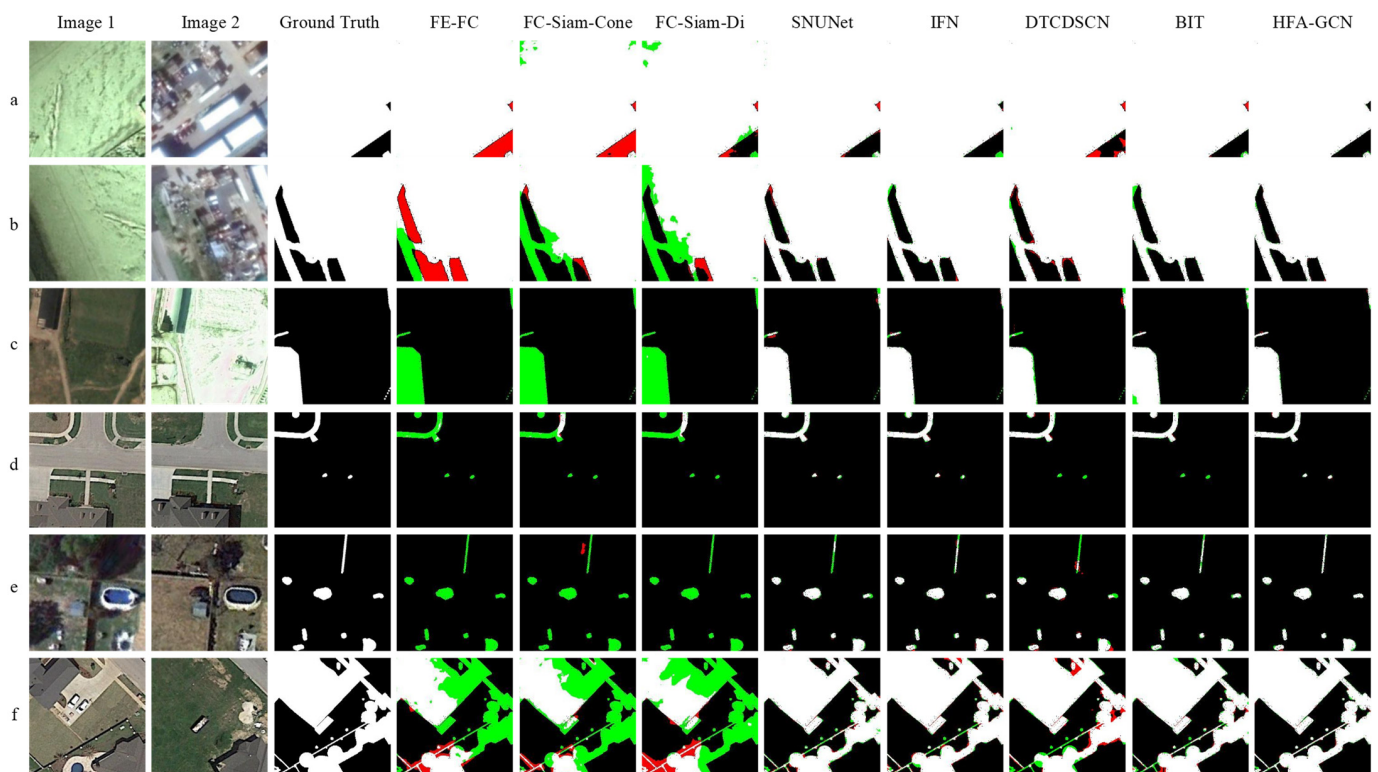
GZ-CD	Pre (%)	Recall (%)	F1 (%)	IOU (%)
FC-EF(2019)	87.52	55.01	67.56	51.01
FC-Siam-Di(2019)	76.99	62.88	69.22	52.93
FC-Siam-Conv(2019)	73.94	66.64	70.10	53.97
IFN(2020)	80.70	84.66	82.64	70.41
SNUNet-CD/32(2022)	86.18	78.73	82.29	69.91
DTCDSN(2020)	88.83	79.40	83.85	72.19
BIT(2021)	92.98	81.29	86.74	76.59
HFA-GCN	91.84	83.40	86.86	76.78

#### 3.4.1. CDD Dataset

In Table 1. The quantitative results reflect that the proposed network in this paper outperforms the other network models in terms of metrics, with Pre/Recall/F1/IOU scores 0.26%/0.76%/0.72%/1.37% higher than the other network, respectively.

For the CDD dataset, we visualize the representative areas in it. Among them, Figure 7a,b show large change regions; remote sensing images are highly affected by

seasons, the edges of buildings in a map are affected by light shadows, and in the b map, the connection between shadows and grasses causes difficulties in detection. Figure 7c,d show small change regions; images in the c map are highly affected by the seasons, and in the d map, the change caused by roadside street lights is very small. Figure 7e,f show a dense change region, affected by the need to detect the details of the texture edge; the e map shows different shapes of small changes in the composition, with vegetation shadows increasing the difficulty of detecting small roads; the f map of residential areas, in addition to the changes in the buildings, also shows a small change that needs to be detected. Comparing the experimental results, as shown in in Figure 7a,b, SNUNet, IFN, DTCDCSCN, BIT, and the proposed method are not affected by the seasonal changes and achieve better results. In Figure 7c,d, SNUNet, IFN, DTCDCSCN, BIT, and the proposed method can detect the general area of change, but the proposed method can detect more small changes. In Figure 7e,f, with numerous small changes, the proposed methods can achieve more accurate predictions with fewer misses and misjudgments. The more accurate prediction plots reflect the powerful change detection ability of the proposed network, which is consistent with the proposed method achieving the best results for the objective index of CDD.



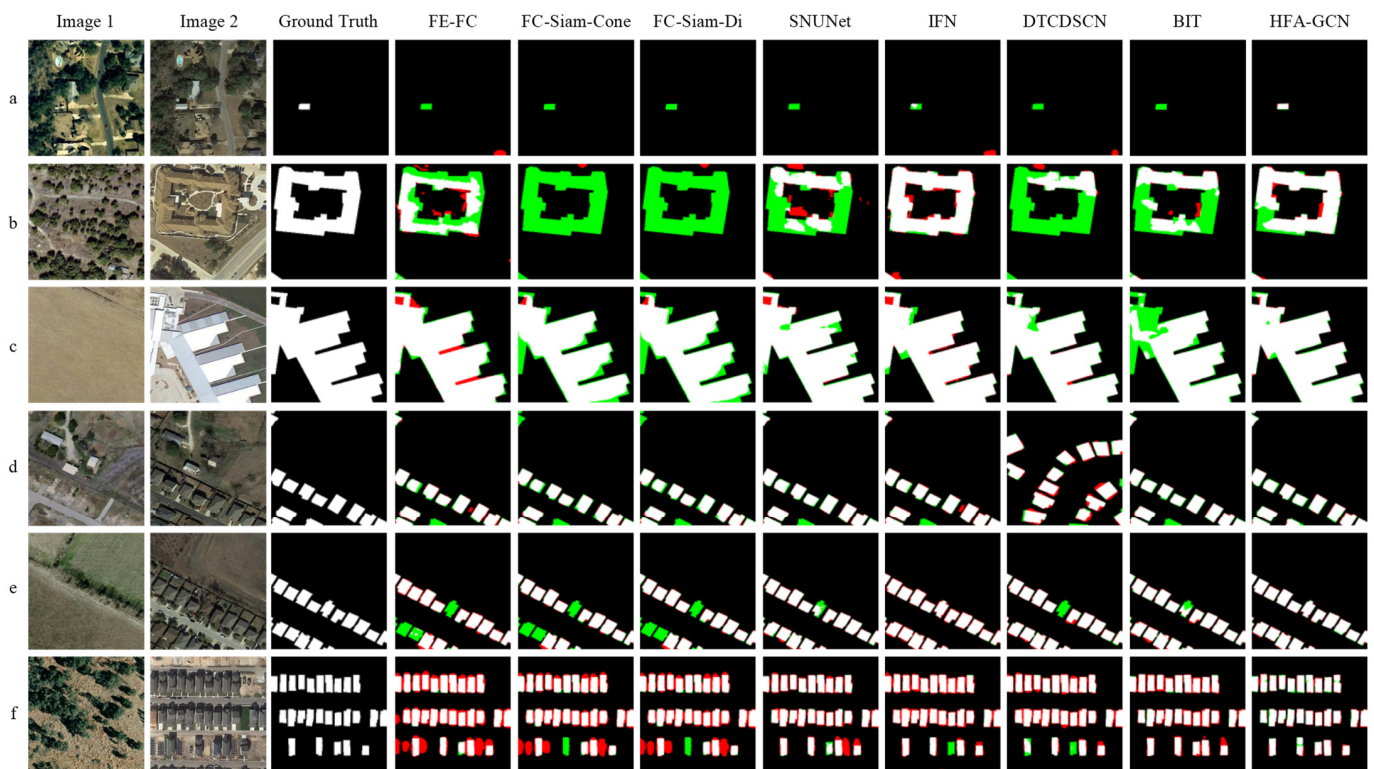
**Figure 7.** Comparison of the result plots of different methods on the CDD dataset; (a–f) are six plots with representative patterns. For better observation, TP is white, TN is black, FP is red, and FN is green.

### 3.4.2. LEVIR-CD Dataset

In Table 2, the quantitative results reflect that the network proposed in this paper outperforms the other network models in terms of metrics, with F1/IOU scores 0.74%/1.24% higher than the other network, respectively.

For the LEVIR-CD dataset, we have also visualized their representative areas. For example, Figure 8a is a small change area; Figure 8b,c are large change areas, which are affected by the roof color being close to the soil color in the b map, and the edges of the buildings are affected by shadows in the c map. Figure 8d–f show intensive change areas with a large number of individual building additions, which need to be divided into a large number of change individuals. In Figure 8a, only the mentioned method, among

all methods, identified the building increase from the vegetation. In Figure 8b, roofs and soils are difficult to distinguish, with SNUNet, BIT, etc. causing a large number of missed detections, and IFN is biased toward increased false detections, but the proposed method achieves a balance between missed and false detections to get a better result. In Figure 8c, all methods can detect the main part of the changing building, and the proposed network can detect the edges of the building better, thus reducing the false detection between the building boundaries, and also reducing the leakage of the building. In Figure 8d,e there are a large number of individual building changes, wherein some buildings are difficult to detect because of the presence of shadows, and SNUNet, IFN, DTCDCSCN, BIT, and other methods all show individual building change detections missed; whereas, using the proposed method, all individual building changes can be detected. In Figure 8f, the proposed method is also more accurate for building change edge detection, while avoiding the comparison method in the case of missing detections of individual building changes. Thus, it can be seen that the proposed method has the best visual performance regarding the prediction map, which is consistent with its best performance of the objective evaluation index in LEVIR-CD.



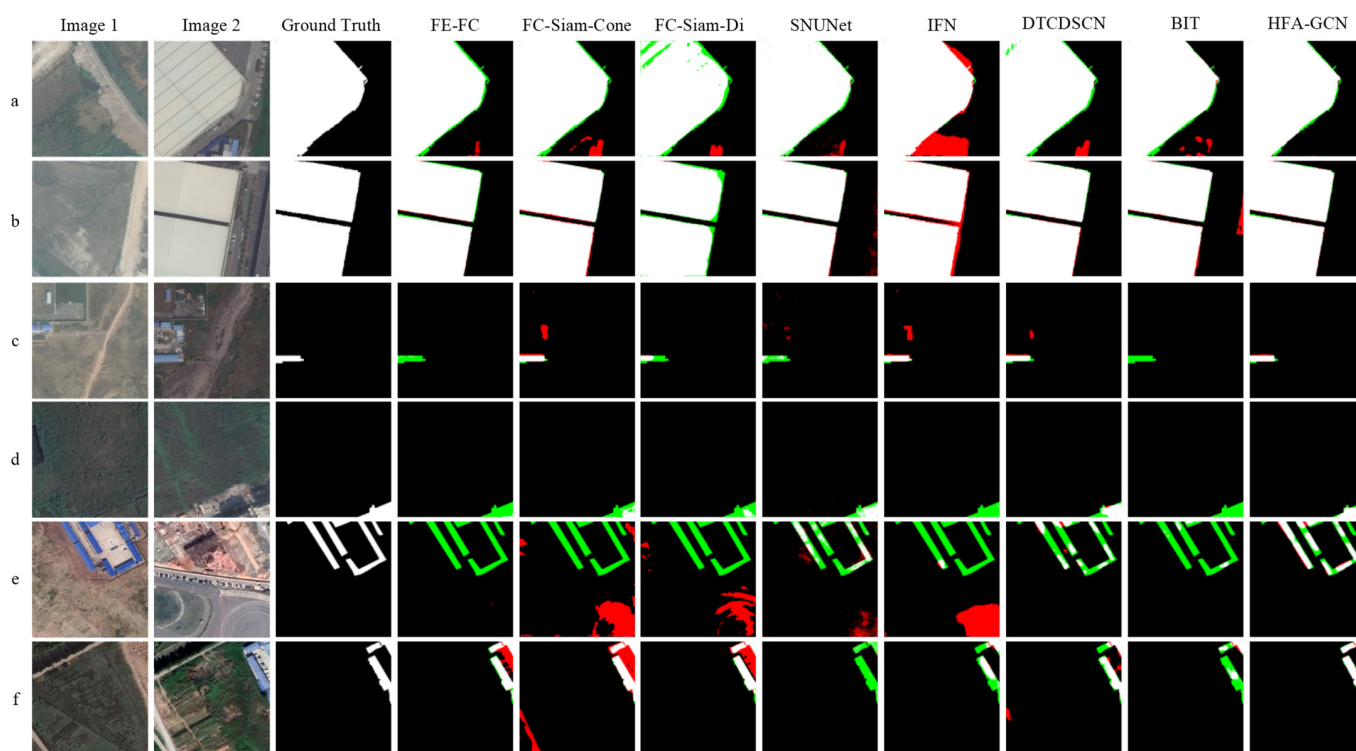
**Figure 8.** Comparison of the result plots of different methods on the LEVIR-CD dataset; (a–f) are six plots with representative patterns. For better observation, TP is white, TN is black, FP is red, and FN is green.

### 3.4.3. GZ-CD Dataset

The quantitative results reflect that the proposed network in this paper outperforms the other network models in terms of metrics, with F1/IOU being 0.12%/0.19% higher than the other network, respectively.

For the GZ-CD dataset, we have visualized its representative areas. For example, Figure 9a,b show the changes in warehouses, which are large change areas. Figure 9c,d show small change areas; the new buildings in c are located at the edges of other buildings whose accurate detection is considered difficult, and the building changes in d are affected by their shadows. Figure 9e,f show the changes in strip dense buildings, for which there is a need to discriminate the detail texture of the change areas. In Figure 9a, it can be seen

that for SNUNet, IFN, BIT, etc., for non-building changes, there are misjudgments, but the proposed method avoids misjudgments, and the overall visual effect achieves better results. In Figure 9b, the edges of the building are affected by shadows; for example, IFN and BIT methods exhibit edge misjudgment, while the FC-Siam-Conv method exhibits missed judgments, but the proposed method has a more accurate visual effect for edge discrimination. In Figure 9c, small building changes are missed in SNUNet, BIT, and other methods, while IFN has misjudgments for unchanged buildings, but the proposed method avoids these misjudgments to accurately discriminate the changing area. In Figure 9d, the change areas are difficult to judge due to shadows, and the proposed method avoids the omissions of the comparison methods to achieve more accurate change detection results generation. In Figure 9e,f, the complete discrimination of strip buildings is challenging; the omission of buildings and the misjudgment of inter-building areas exist in the comparison method, but the proposed method can discriminate the whole of the changing buildings while avoiding misjudgments and also avoiding large-scale omissions. The visualization results on the GZ-CD dataset are better than other methods, and the results are consistent with the objective evaluation indexes; thus, the proposed method achieves the best results.



**Figure 9.** Comparison of the result plots of different methods on the GZ-CD dataset; (a–f) are six plots with representative patterns. For better observation, TP is white, TN is black, FP is red, and FN is green.

### 3.5. Ablation Experiments

After obtaining the overall network results, to further verify the effectiveness of the designed modules in the network model, we conducted ablation experiments for validation, and the results of the ablation experiments are shown in Table 4, confirming that the hierarchical feature association module and the global correction module have a positive effect on the change detection task. Adding either of these modules to the baseline network results in better overall network performance on all three datasets. Finally, the complete network, using these two modules, performs best on all three datasets.

**Table 4.** Experimental results of different ablations on different datasets.

Dataset	Methods	Pre (%)	Recall (%)	F1 (%)	IOU (%)
CDD	Base	98.23	93.62	95.87	92.07
	Res-HFA-GCN	95.10	96.40	95.74	91.84
	w/o HFA	96.41	95.63	96.02	92.34
	w/o GC	98.48	94.12	96.25	92.77
	w/o GC-RIF	97.66	96.60	97.13	94.42
	HFA-GCN	97.40	97.20	97.30	94.75
LEVIR-CD	Base	91.52	87.67	89.55	81.08
	Res-HFA-GCN	92.01	86.79	89.33	80.71
	w/o HFA	92.33	88.91	90.59	82.79
	w/o GC	91.81	87.80	89.76	81.42
	w/o GC-RIF	93.09	88.03	90.49	82.63
	HFA-GCN	91.49	89.99	90.73	83.04
GZ-CD	Base	92.05	78.01	84.45	73.09
	Res-HFA-GCN	92.69	77.53	84.44	73.07
	w/o HFA	88.35	82.37	85.26	74.30
	w/o GC	92.93	78.69	85.22	74.25
	w/o GC-RIF	93.20	80.86	86.59	76.35
	HFA-GCN	91.84	82.40	86.86	76.78

(1) Baseline network (Base): The baseline network consists of an encoder and a decoder. The performance of the baseline network is used to provide a baseline reference for the improvement brought by the innovation module, which can effectively reflect the improvement brought by the proposed module for the change detection task.

(2) Feature extraction module: In HFA-GCN, we designed a feature extraction module to better obtain change information in remote sensing images. To demonstrate the effectiveness of using the feature extraction module, we replaced it with the commonly used feature extraction network resnet-18 (named Res-HFA-GCN). On all three datasets, the performance of Res-HFA-GCN was inferior to that of HFA-GCN, thus proving the effectiveness of the feature extraction module.

(3) Hierarchical feature association module: We analyzed whether the hierarchical feature association module can provide better performance for the overall change detection network. We simply used the feature concat, instead of the proposed hierarchical feature association module, and we can see a significant decrease in the performance of the network, which shows that modeling neighbor-order feature information is beneficial to enhance the representation of effective information and reduces the representation of redundant features.

(4) Global correction module: We analyzed whether the global correction module can provide better performance for the overall change detection network. First, we directly removed the GC module, which means removing the global information introduction and feature reuse, and compared it to the proposed overall network model, which lacks the global information correction, and showed a significant performance degradation. Secondly, in order to verify that the global information extracted by the module came from the bitemporal image features, rather than the change features themselves, we removed the use of bitemporal image features from the GC input and replaced them with the change features themselves. We named it ‘without reuse image features’ (w/o GC-RIF), and the results show that the use of extracted features brings a performance improvement. This shows that the global semantic information embedded in the image features is very important for the change detection task.

#### 4. Conclusions

We investigated the effective utilization of features at different levels and the importance of global information for change detection tasks. We propose HFA-GCN, in which the hierarchical feature association module models the association relationships between



features at different levels to more fully utilize the extracted features, whereas the global correction module achieves more effective extraction and utilization of global features by reusing the features and mining the global information of the image to correct the change detection features. Adequate experiments were conducted on three publicly available datasets, LEVIR-CD, GZ-CD, and CDD. The experimental results show that the extracted networks are highly competitive. Due to the use of the hierarchical feature association module and the global correction module, increased convolution, and the use of global linear mapping, the disadvantages of this model are high computational effort, many parameters, and high resource consumption, as show in Table 5, and future work will investigate this model in a lightweight manner.

**Table 5.** Experimental results of different ablation and HFA-GCN on efficiency comparison.

Methods	Base	Res-HFA-GCN	w/o HFA	w/o GC	w/o GC-RIF	HFA-GCN
FLOPs(G)	153.71	253.47	160.29	306.01	317.32	317.48
Params(M)	3.02	24.73	2.91	15.57	15.60	15.60

**Author Contributions:** Conceptualization, J.L. and X.M.; methodology, J.L. and X.M.; software, J.L.; validation, J.L.; writing—original draft preparation, J.L.; writing—review and editing, J.L., Q.L. and X.M.; supervision, Q.L., X.M., Z.L., G.Y., W.S. and W.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (42171326); National Natural Science Foundation of China (42071323); Zhejiang Provincial Natural Science Foundation of China (LR23D010001, LY22F010014); Zhejiang Province “Pioneering Soldier” and “Leading Goose” R&D Project under Grant 2023C01027; Postdoctoral Research Foundation of China under Grant 2020M672490; Ningbo Natural Science Foundation under Grant 2022J076, and in part by the Ningbo Science and Technology Innovation 2025 Major Special Project under Grants 2021Z107 and 2022Z032.

**Data Availability Statement:** The CDD dataset is available at: <https://pan.baidu.com/s/1Xu0klpThW2koLcyfcJEEfA> (accessed on 15 April 2022) (Password: RSAI). The LEVIR-CD dataset is available at: <https://justchenhao.github.io/LEVIR> (accessed on 18 April 2022). The GZ-CD dataset is available at: <https://github.com/daifeng2016/Change-Detection-Dataset-for-High-Resolution-Satellite-Imagery> (accessed on 21 April 2022).

**Acknowledgments:** The authors wish to acknowledge the team led by Zhe Li of the College of Computer Science and Engineering, Shandong University of Science and Technology, who provided the code of SNUNet.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lv, Z.; Liu, T.; Benediktsson, J.A.; Falco, N. Land cover change detection techniques: Very-high-resolution optical images: A review. *IEEE Geosci. Remote Sens. Mag.* **2021**, *10*, 44–63. [CrossRef]
2. Ji, S.; Wei, S.; Lu, M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 574–586. [CrossRef]
3. Ye, S.; Rogan, J.; Zhu, Z.; Eastman, J.R. A near-real-time approach for monitoring forest disturbance using Landsat time series: Stochastic continuous change detection. *Remote Sens. Environ.* **2021**, *252*, 112167. [CrossRef]
4. Gueguen, L.; Hamid, R. Toward a generalizable image representation for large-scale change detection: Application to generic damage analysis. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3378–3387. [CrossRef]
5. Shi, W.; Zhang, M.; Ke, H.; Fang, X.; Zhan, Z.; Chen, S. Landslide recognition by deep convolutional neural network and change detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 4654–4672. [CrossRef]
6. Asokan, A.; Anitha, J. Change detection techniques for remote sensing applications: A survey. *Earth Sci. Inform.* **2019**, *12*, 143–160. [CrossRef]
7. Quarmby, N.A.; Cushnie, J.L. Monitoring urban land cover changes at the urban fringe from spot hrv imagery in south-east England. *Int. J. Remote Sens.* **1989**, *10*, 953–963. [CrossRef]
8. Zhu, J.; Su, Y.; Guo, Q.; Harmon, T. Unsupervised object-based differencing for land-cover change detection. *Photogramm. Eng. Remote Sens.* **2017**, *83*, 225–236. [CrossRef]

9. Rignot, E.; Vanzyl, J. Change detection techniques for ers-1 sar data. *IEEE Trans. Geosci. Remote Sens.* **1993**, *31*, 896–906. [[CrossRef](#)]
10. Liu, S.; Bruzzone, L.; Bovolo, F.; Zanetti, M.; Du, P. Sequential spectral change vector analysis for iteratively discovering and detecting multiple changes in hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4363–4378. [[CrossRef](#)]
11. Ferraris, V.; Dobigeon, N.; Wei, Q.; Chabert, M. Detecting changes between optical images of different spatial and spectral resolutions: A fusion-based approach. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 1566–1578. [[CrossRef](#)]
12. Richards, J. Thematic mapping from multitemporal image data using the principal components transformation. *Remote Sens. Environ.* **1984**, *16*, 35–46. [[CrossRef](#)]
13. Jin, S.; Sader, A. Comparison of time series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. *Remote Sens. Environ.* **2005**, *94*, 364–372. [[CrossRef](#)]
14. Ahlqvist, O. Extending post-classification change detection using semantic similarity metrics to overcome class heterogeneity: A study of 1992 and 2001 U.S. National Land Cover Database changes. *Remote Sens. Environ.* **2008**, *112*, 1226–1241. [[CrossRef](#)]
15. Liu, J.; Chen, K.; Xu, G.; Sun, X.; Yan, M.; Diao, W.; Han, H. Convolutional neural network-based transfer learning for optical aerial images change detection. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 127–131. [[CrossRef](#)]
16. Cui, B.; Zhang, Y.; Yan, L.; Wei, J.; Wu, H. An unsupervised sar change detection method based on stochastic subspace ensemble learning. *Remote Sens.* **2019**, *11*, 1314. [[CrossRef](#)]
17. Shao, P.; Shi, W.; He, P.; Hao, M.; Zhang, X. Novel Approach to Unsupervised Change Detection Based on a Robust Semi-Supervised FCM Clustering Algorithm. *Remote Sens.* **2016**, *8*, 264. [[CrossRef](#)]
18. Gao, Y.; Gao, F.; Dong, J.; Wang, S. Change detection from synthetic aperture radar images based on channel weighting-based deep cascade network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 4517–4529. [[CrossRef](#)]
19. Zhang, M.; Shi, W. A feature difference convolutional neural network-based change detection method. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7232–7246. [[CrossRef](#)]
20. Mou, L.; Bruzzone, L.; Zhu, X.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](#)]
21. Li, X.; Du, Z.; Huang, Y.; Tan, Z. A deep translation (GAN) based change detection network for optical and SAR remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2021**, *179*, 14–34. [[CrossRef](#)]
22. Wang, D.; Zhao, F.; Yi, H.; Li, Y.; Chen, X. An unsupervised heterogeneous change detection method based on image translation network and post-processing algorithm. *Int. J. Digit. Earth* **2022**, *15*, 1056–1080. [[CrossRef](#)]
23. Yang, B.; Qin, L.; Liu, J.; Liu, X. UTRNet: An Unsupervised Time-Distance-Guided Convolutional Recurrent Network for Change Detection in Irregularly Collected Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [[CrossRef](#)]
24. Lv, Z.; Huang, H.; Li, X.; Zhao, M.; Benediktsson, J.A.; Sun, W.; Falco, N. Land Cover Change Detection With Heterogeneous Remote Sensing Images: Review, Progress, and Perspective. *Proc. IEEE* **2022**, *11*, 1976–1991. [[CrossRef](#)]
25. Lv, Z.; Zhong, P.; Wang, W.; You, Z.; Benediktsson, J.A.; Shi, C. Novel Piecewise Distance Based on Adaptive Region Key-Points Extraction for LCCD With VHR Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–99. [[CrossRef](#)]
26. Fang, S.; Li, K.; Shao, J.; Li, Z. SNUNet-CD: A densely connected Siamese network for change detection of VHR images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [[CrossRef](#)]
27. Liu, Y.; Pang, C.; Zhan, Z. Building change detection for remote sensing images using a dual-task constrained deep siamese convolutional network model. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 811–815. [[CrossRef](#)]
28. Wang, D.; Zhao, F.; Wang, C.; Wang, H.; Zheng, F.; Chen, X. Y-Net: A multiclass change detection network for bi-temporal remote sensing images. *Int. J. Remote Sens.* **2022**, *43*, 565–592. [[CrossRef](#)]
29. Chen, J.; Fan, J.; Zhang, M.; Zhou, Y.; Shen, C. MSF-Net: A Multiscale Supervised Fusion Network for Building Change Detection in High-Resolution Remote Sensing Images. *IEEE Access* **2022**, *10*, 30925–30938. [[CrossRef](#)]
30. Song, L.; Xia, M.; Jin, J.; Qian, M.; Zhang, Y. SUACDNet: Attentional change detection network based on siamese U-shaped structure. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *105*, 102597. [[CrossRef](#)]
31. Lv, Z.; Wang, F.; Cui, G.; Benediktsson, J.A.; Lei, T.; Sun, W. Spatial-Spectral Attention Network Guided with Change Magnitude Image for Land Cover Change Detection Using Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–12. [[CrossRef](#)]
32. Lv, Z.; Huang, H.; Gao, L.; Benediktsson, J.A.; Zhao, M.; Shi, C. Simple Multiscale UNet for Change Detection With Heterogeneous Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 2504905. [[CrossRef](#)]
33. 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](#)]
34. Chen, H.; Qi, Z.; Shi, Z. Remote sensing image change detection with transformers. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [[CrossRef](#)]
35. Bandara, W.; Patel, V. A Transformer-Based Siamese Network for Change Detection. In Proceedings of the IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 207–210.
36. Feng, Y.; Xu, H.; Jiang, J.; Liu, H.; Zheng, J. ICIF-Net: Intra-Scale Cross-Interaction and Inter-Scale Feature Fusion Network for Bitemporal Remote Sensing Images Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [[CrossRef](#)]

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