Article

Multi-Feature Enhanced Building Change Detection Based on Semantic Information Guidance

Junkang Xue [1], Hao Xu [2], Hui Yang [1,3,4], Biao Wang [1,3], Penghai Wu [1,3], Jaewan Choi [5], Lixiao Cai [6] and Yanlan Wu [1,3,*]

1 School of Resources and Environmental Engineering, Anhui University, Hefei 230601, China; x19301089@stu.ahu.edu.cn (J.X.); yanghui@ahu.edu.cn (H.Y.); wangbiao-rs@ahu.edu.cn (B.W.); wuph@ahu.edu.cn (P.W.)
2 Institute of Spacecraft System Engineering, Beijing 100094, China; haibeihms@163.com
3 Information Materials and Intelligent Sensing Laboratory of Anhui Province Hefei, Hefei 230601, China
4 Institutes of Physical Science and Information Technology, Anhui University, Hefei 230601, China
5 School of Civil Engineering, Chungbuk National University, Chungju 28644, Korea; jaewanchoi@chungbuk.ac.kr
6 School of Design Group, Shandong Jianzhu University, Jinan 250101, China; 13681@sdjzu.edu.cn
* Correspondence: wuyanlan@ahu.edu.cn

Abstract: Building change detection has always been an important research focus in production and urbanization. In recent years, deep learning methods have demonstrated a powerful ability in the field of detecting remote sensing changes. However, due to the heterogeneity of remote sensing and the characteristics of buildings, the current methods do not present an effective means to perceive building changes or the ability to fuse multi-temporal remote sensing features, which leads to fragmented and incomplete results. In this article, we propose a multi-branched network structure to fuse the semantic information of the building changes at different levels. In this model, two accessory branches were used to guide the buildings' semantic information under different time sequences, and the main branches can merge the change information. In addition, we also designed a feature enhancement layer to further strengthen the integration of the main and accessory branch information. For ablation experiments, we designed experiments on the above optimization process. For MDEFNET, we designed experiments which compare with typical deep learning model and recent deep learning change detection methods. Experimentation with the WHU Building Change Detection Dataset showed that the method in this paper obtained accuracies of 0.8526, 0.9418, and 0.9204 in Intersection over Union (IoU), Recall, and F1 Score, respectively, which could assess building change areas with complete boundaries and accurate results.

Keywords: remote sensing; building change detection; deep learning; multi-branch; semantic information

1. Introduction

Changes in building detection are one of the most important tasks in the field of remote sensing, which has been widely used in urban development planning [1], post-earthquake [2] or flood assessment [3], and so on. However, due to the diversity of buildings and the heterogeneity of remote sensing data at different times, there are still fundamental difficulties and challenges involved in building change detections of high-resolution remote sensing images.

The remote sensing detection of building changes has developed two main research branches: traditional methods and deep learning methods [4]. According to different analysis units, the traditional methods are divided into pixel-based change detection (PBCD) and object-based change detection (OBCD) [5]. The PBCD methods take pixels as the main analysis units. By detecting and analyzing the spectral characteristics of a single pixel, such methods can assess the features of change [6]. Typical PBCD methods include change vector analysis (CVA) [7], principal component analysis (PCA) [8], and feature

index differencing [9]. However, these methods only consider the spectral information of a single pixel and assume that the spatial characteristics of adjacent pixels are independent. When high-resolution images are utilized in practical applications, buildings are more refined and there are feature dependencies between adjacent pixels. Such methods will cause fragmentation, blurred boundaries, or the "salt and pepper" effect [10,11]. The OBCD method uses a homogenous object extracted from the image as the processing unit. It combines both spectral and spatial information of the object to detect the change, so as to reduce the detection error caused by the diversity of the features [12]. Representative object-oriented methods used in eCognition include chessboard segmentation [13], quadtree-based segmentation [14], multiresolution segmentation [15], etc. However, OBCD methods regard multi-temporal buildings as homogeneous objects for comparison, which will distort the distribution of buildings as a result, and the detailed information inside the target object cannot be effectively used. Meanwhile, the spectral difference between multi-temporal high-resolution images makes the OBCD methods difficult to apply, which affects the change detection results [16].

Deep learning technology has also been continuously developed in the field of change detection [17], due to the large amount of data and rich features in the image processing process. CNN can optimize the amount of parameters and retain the feature in images through the convolution process. Simultaneously, it has an end-to-end simple form, which can automatically learn the non-linear characteristics between image pairs, making it the mainstream image analysis research method for researchers [18]. Scholars innovated based on the CNN framework and proposed many networks that exhibit different advantages, such as VGGNet [19], CaffeNet [20], SegNet [21], and U-Net [22,23]. Based on the framework of these models, deep learning methods can be categorized into single-branch network methods [24] and multi-branch network methods [25]. Single-branch network methods use fusion multi-temporal images as inputs to construct non-linear fitting for the changed feature [26]. For example, Gong et al. [27] used hierarchical Fuzzy C-Means (FCM) clustering to obtain highly accurate samples for PCANet, which could reduce false alarms and the speckle noise of SAR images. Wang et al. [28] proposed an end-to-end 2-D convolutional neural network (GETNET), in order to derive better hyperspectral image pair information. Peng et al. [29] employed a fusion strategy of multiple side outputs, which could avoid error accumulation problems in processing final change maps. Multi-branch network methods process the image information of a certain time phase separately. Compared with single-branch network processing, such methods can be used to assess the image pairs information more effectively [30]. In [31,32], dual attentive fully convolutional Siamese networks (DASNet) and deep Siamese convolutional network (DSCN) were proposed to effectively learn about the complex information between multi-temporal images by calculating the pairwise Euclidean distance. Chen et al. [33] combined double-branch networks and LSTM for detecting changes in homogeneous VHR images and heterogeneous very-high-resolution (VHR) images. The authors of [34] also discussed the influence of fusing different features on change detection results, while detecting building changes in a specific semantic environment; semantic information is important for enhancing change information. How to use semantic information to guide detections of building changes in a specific scenario has become a way to improve the accuracy of building change detection.

In this paper, a multi-branch feature-enhanced network is proposed for building changes under high-resolution image. The network structure is mainly composed of a main branch and two accessory branches. The main branch is for learning semantic change information, and the accessory branch is for guiding semantic auxiliary building information. After each branch is down-sampled, the feature fusion layer is used to make the network learn the features of the building changes more effectively. In the up-sampled part, we use skip connections to reduce the loss of detailed information. This multi-branch network reduces the noise of change detection and detects building changes more efficiently.

The rest of this article is organized as follows. In Section 2, we introduce the building change detection framework, and in Section 3, some details of the dataset and the experimental setup are illustrated. In Section 4, we report on relevant experiments to prove the effectiveness of this method. We conclude this article in Section 5.

2. Methods

2.1. The Network Structure

In recent years, Siamese networks have been used in the field of change detection [35]. Subsequently, proposed pseudo-Siamese and pseudo-double-branch networks have also achieved good results. Those methods provide more ideas to improve the feature extraction of image pairs. Inspired by these concepts, this paper proposes Multi-Dense-Enhanced-Feature-Net (MDEFNET), a network of a main and accessory branch structure. In order to attain accurate and integrated results in the field of building change detection, we use dense blocks to concatenate the entire network and fuse multiple information points through the proposed feature-enhanced layer. The main branch learns the semantic information of the change area, and the accessory branch guides the semantic information of the building change. The MDEFNET model can be used to fully learn the effective feature information of building changes in multi-temporal images. It can reduce the loss of building change features, improve the accuracy of building extraction from remote sensing images, and better detect the integrity of building boundaries.

The network structure is shown in Figure 1. Firstly, in the input part of the network, different branches perform convolution operations on the dual-phase image. The images from two periods after one down-sampling are fused as the input of the main branch. Secondly, the other two accessory branches simultaneously perform down-sampling operations on a certain phase image. During the down-sampling of the network, dense block and the transition layer are connected. This operation can release the calculation burden, which is very important in change detection. In the process of dense layer processing information, because layers are deeper, the results of non-linear fitting are better for more complex information. Therefore, we designed a mechanism in which the number of dense layers grows with the down-sampling process. The number of dense layers in the main branch grows with two, three, four, and five layers, and the two accessory branches are down-sampled with two, two, three, and two layers. In this way, through the different growth rates of the number of layers, higher-level features can be obtained from the main branch, whereas low-level features of the accessory branches can be used as auxiliary feature learning. Such network structures and layer designs are to better identify the accuracy of areas of building changes. After the second, third, or fourth down-sampling, the accessory branch features are added to the main branch through the feature-enhanced module. The proposed feature-enhanced module integrates the three branch features to form an extraction focusing on buildings change features, which can guide the semantic building information. After the fifth down-sampling, the main branch and the accessory branch are merged, entering the up-sampling process after six dense layers. In the up-sampling part, skip connection is used in this process. We perform a jump connection between the main branch information after feature fusion and the information from the up-sampling process. Additionally, when the main branch is down-sampling for the first time, it is also skip-connected with the corresponding information of the up-sampling. In this way, the detailed feature information can be retained during the learning process, and the loss of low-level information can be also reduced.
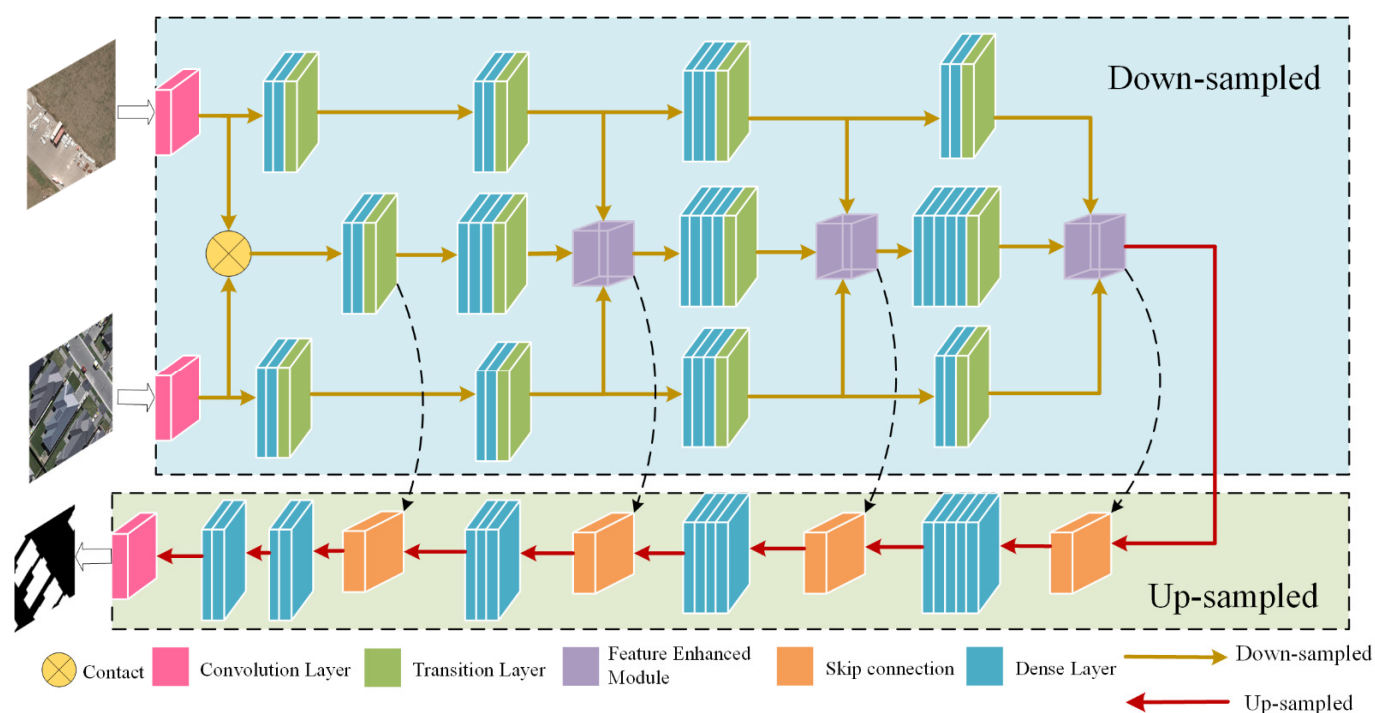
Figure 1. MDEFNET network structure.

2.2. Feature Enhanced Module

High-resolution image information handling is complex. For the detection of building changes in a special semantic environment, it has caused difficulties in the accuracy and completeness of assessing the area of building changes. In order to better combine the features of the accessory branches with the main branch, and enhance the building feature information, in the down-sampling part, we add the feature-enhanced module to fully learn the effective information in the multi-temporal image. As Figure 2 shows, in this module, first, we merge the single-phase information of the accessory branches. The merged feature information is subsequently put into a convolution layer with a convolution kernel of 3×3 and a step size of 1. Then, it combines the extracted features with the main branch information through the global information enhancement module. This operation allows the network to obtain more auxiliary information during the down-sampling process, and thus, it can be used to guide this semantic information to detect changes, which highlights the building changes of the main branch.
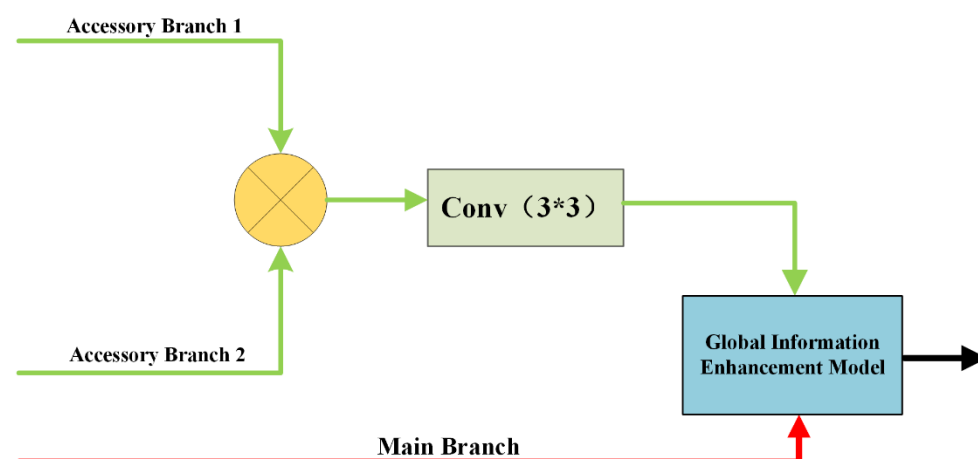

Figure 2. Feature-enhanced module.

Most contemporary deep learning models adopt an end-to-end, encoding–decoding structure. In the down-sampling part, if long time sequence information is input into the model, it may cause part of the initial information to be lost. In the field of semantic segmentation, as the depth of the network increases, low-level semantic information may not be learned effectively. Global pooling has been proven to enhance context information [36]. Therefore, we designed the global information enhancement module. As illustrated in Figure 3, the structure diagram is as follows: this module mainly passes the input information through a series of operations of global pooling, convolution, relu functions, and sigmoid activation functions. At the same time, it includes related calculations such as channel weight multiplication and addition, which changes the input feature weight, enhances the input of effective information, and can improve the segmentation accuracy.
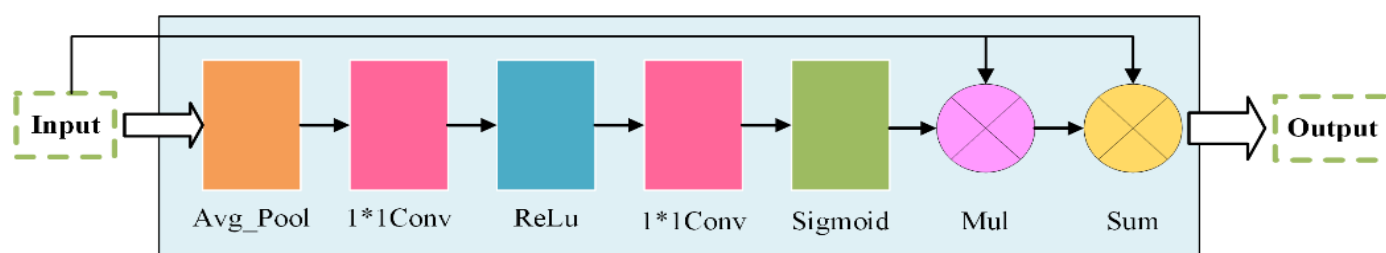

Figure 3. Global information enhancement module.

3. Dataset and Experiments Settings

3.1. Dataset

In the research of building change detection, the lack of public datasets has made the research of this method slower. The Building Change Detection Dataset in the WHU Building Dataset released by Wuhan University in 2019 [37] is a high-quality dataset and meets the training requirements of deep learning samples. Therefore, this dataset was selected to validate the proposed model in this study. The time span of this dataset was April 2011 to 2016, depicting aerial data images of two scenes showing the areas of change in the destruction and reconstruction of buildings caused by an earthquake. It has 12,796 buildings in 2012 and 16,077 buildings in 2016. In addition, the public dataset has 1.6-pixel rectified accuracy, which meets the needs of deep learning methods in change detection. The dataset consisted of a pair of aerial images measuring $32,507 \times 15,354$ pixels at a resolution of 0.2 m/pixel, as well as labeled maps corresponding to changes in the buildings. As shown in Figure 4, this dataset contains many different scenarios for building change detection, such as, demolition buildings, new construction or reconstruction buildings, changes of residential buildings or factories.

Figure 4. The representative building changes of every type in WHU Building Change Detection Dataset. (**a**) Before; (**b**) After; (**c**) Ground Truth.

In order to make the model more suitable for building change feature extraction and learning, this experiment divided the aerial image pairs and the corresponding building change label maps, which were the sample training selection area (blue box in Figure 5) and the test area (red box in Figure 5). In the training sample selection area, 256 × 256 pixels and non-overlapping patches were used to cut the two-phase image and the corresponding label maps of building changes. Due to the seriously imbalanced problem of building change samples, this study saved the image pairs with positive samples and added the background data randomly. Finally, the training set of 1662 images and the verification set of 350 images were formed.
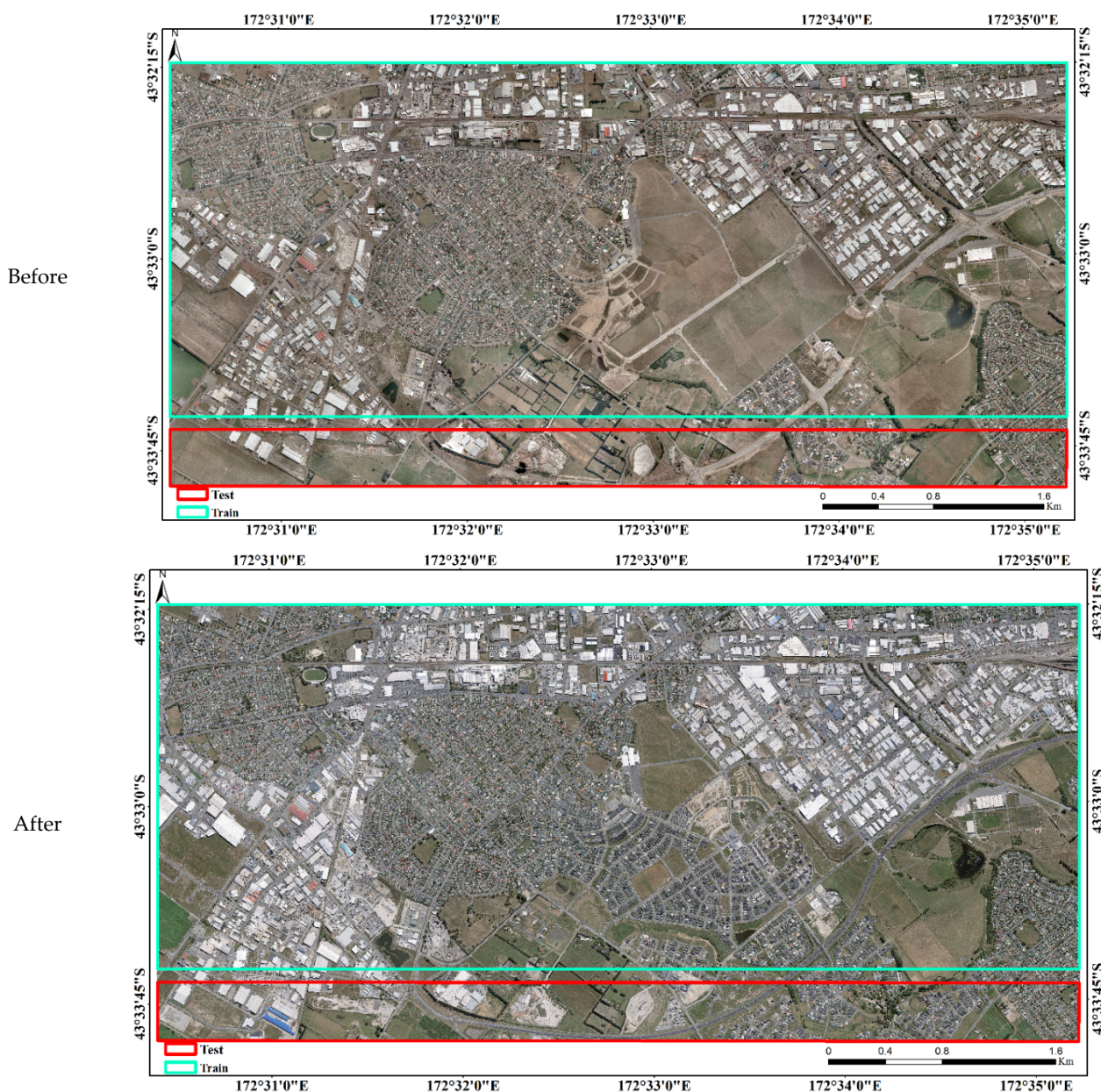

Figure 5. *Cont.*
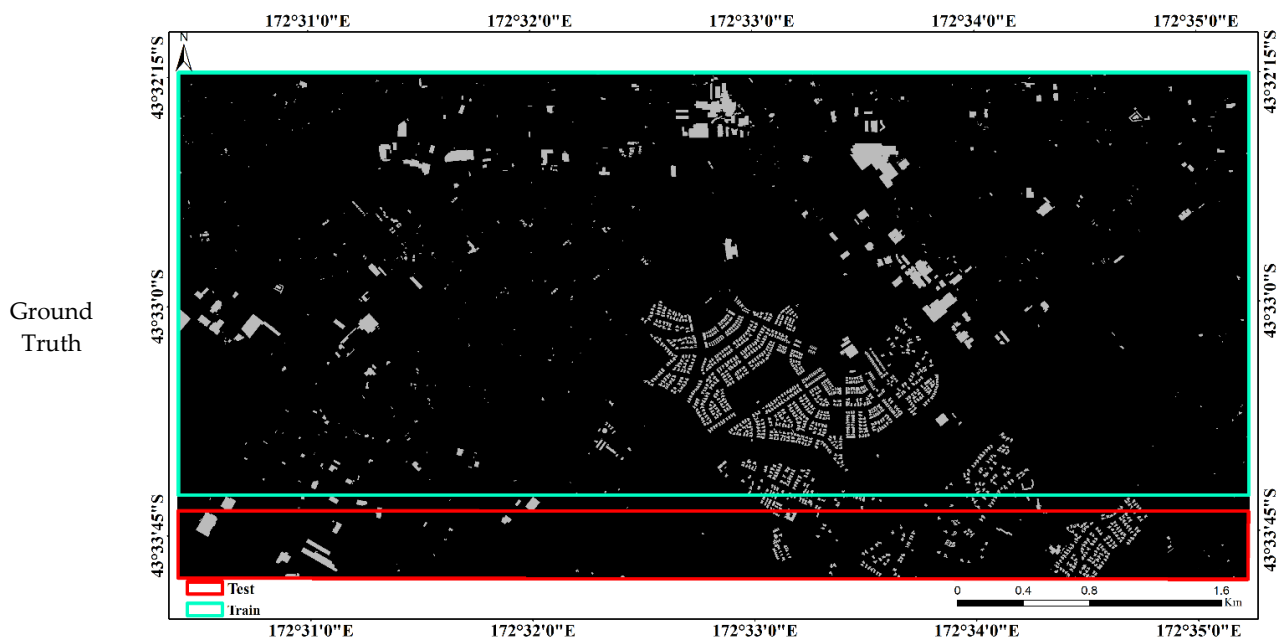
Figure 5. Overview of the dataset.

3.2. Experimental Settings

All experimental settings are shown in Table 1.

Table 1. Experimental Settings.

Experimental Settings	Value
Framework	TensorFlow
Language	Python
Central Processing Unit	Intel Core i9 processor 7980XE@2.60 GHz
Graphics Processing Unit	NVIDIA GEFORCE GTX 1080 Ti (11 GB)
Optimization Algorithm	Adam
Batch Size	12
Base Learning rate	0.001
Train crop size	256×256

3.3. Evaluation Metrics

There were few areas of building change compared to the whole picture; therefore, this studied focused more on the sample area with positive change. Therefore, we chose to use the following accuracy indicators to evaluate the proposed building change methods: Recall focuses on the prediction results of positive samples, which predict the ratio of the number of correct positive pixels to the total number of positive pixels. Precision represents the correct pixels over the prediction result. F1 Score is the weight average index of precision and Recall, and indicators to evaluate the corresponding accuracy of the forecast results is the Intersection over Union (IoU). The above indicators were calculated based on the statistical indicators in the confusion matrix, for example, TP (True Positive), FP (False Positive), TN (True Negative), and FN (False Negative). The formula of each evaluation index is as follows:

$$\text{precision} = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{1}$$

$$IoU = \frac{N_{TP}}{N_{TP} + N_{FP} + N_{FN}} \tag{2}$$

$$Recall = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{3}$$

$$F1 = \frac{2N_{TP}}{2N_{TP} + N_{FN} + N_{FP}} \tag{4}$$

The above indicators were calculated based on the statistical indicators in the confusion matrix. TP (True Positive): The building change pixels are correct. FP (False Positive): The no building change pixels are wrongly detected as building change area. TN (True Negative): The no building change pixels are correct. FN (False Negative): The building change pixels are wrongly detected as no building change area.

At the same time, for ground features such as buildings, we also focused on patch changes in change detection applications. Therefore, in deep learning method comparison experiments, in addition to the abovementioned pixel-based accuracy evaluation index, we added a change index, missing patch (MP), for the missing building change patch. MP was based on the number of building polygons in the results.

4. Results and Analysis

In this part, we designed ablation experiments and comparison experiments with other deep learning methods. In the ablation I and ablation II experiment, we optimized our method from two aspects: network structure and depth of semantic information. In the comparison experiment, we chose typical models in the field of deep learning for analysis. The accuracy evaluation of all experiments is based on all test areas, and the figures are shown as examples.

4.1. Ablation I Experiments

In this section, we set comparison experiments with single-branch structure and without feature enhanced module. DenseNet [38] has been shown to be excellent in semantic segmentation in recent years. Therefore, we chose the dense block to build a network as the baseline of our ablation experiments. We combined it with Early Fusion theory to propose the EF-Dense. Meanwhile, we propose Multi-Dense-Net that combines the advantages of multiple networks in processing multi-temporal complex data. Based on the above work, in order to reduce the loss of detailed information and effectively enhanced the semantic information, we propose the feature enhancement module to form the network of this paper, Multi-Dense-Enhanced-Feature-Net.

As in Table 2, EF-Dense shows great result in evaluation metrics. This shows that dense block is effective for the extraction of change detection tasks. The recall of MDNet is 93.37%, but the net is poor in IoU. This may be because the semantic information has been enhanced but is not well utilized. Our network performed well in every accuracy evaluation index. For the EF-Dens network that performed well in IoU and F1 Score, our method improved the accuracy by 11% and 7%, respectively. Meanwhile, our method also improved the recall of MDNet by about 11%. This is due to the feature enhancement module we proposed, which can reasonably use semantic information.

Table 2. The evaluation metrics results of ablation I experiments.

	IoU	Recall	F1 Score
EF-Dens	0.7423	0.7870	0.8521
MDNet	0.4572	0.9337	0.6275
Ours	**0.8526**	**0.9418**	**0.9204**

In Figure 6, the single-branch network structure, EF-Dense, had lower accuracy in identifying changed areas. This may be due to the early fusion operation, which increased the number of image channels and complicated the change information. Meanwhile, this

operation also made it impossible to effectively extract features, which caused EF-Dense's incorrect extraction of the change results and incomplete change detection. In the first two pairs of images, the multi-branch network had a more complete result for building change detection.
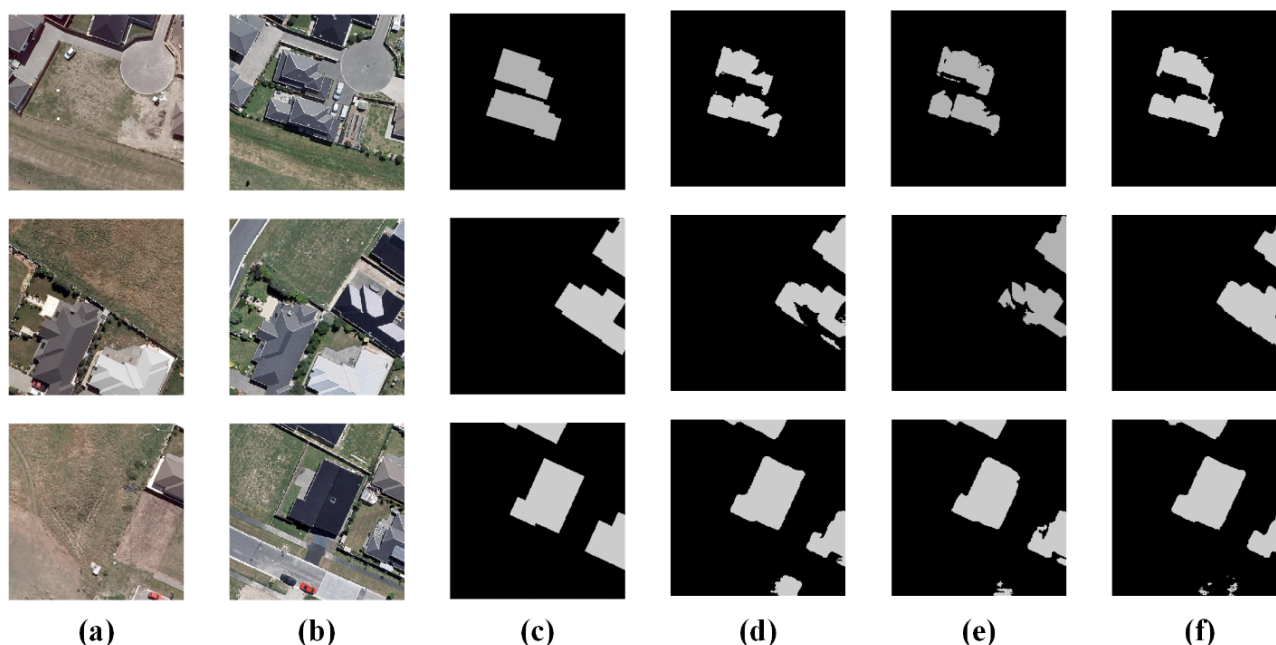

Figure 6. (**a**) Before image; (**b**) After image; (**c**) Label; (**d**) EF-Dens; (**e**) MDNET; (**f**) Ours.

4.2. Ablation II Experiments

On the basis of the MDEFNET proposed in this article, we were concerned with the depth of the feature information fusion for change detection results. Thus, two networks are proposed, MDEFNET-LL and MDEFNET-HL. Of these two multi-branch networks, MDEFNET-LL discusses the effect of low-level feature information fusion; MDEFNET-HL shows the effect of the network on change detection when the high-level information is fused. MDEFNET-LL is obtained by feature fusion of the main and accessory branches after the second down-sampling. Additionally, after the fourth down-sampling, we fused feature information to obtain the MDEFNET-HL.

In the training set and the test set, this study randomly selected several image pairs of the changing areas with a size of 256×256 pixels. Additionally, we output the feature maps of these image pairs. In the test part, the final output results of each network are presented at the same time in order to maintain consistent feature information and the depth of the image information in the down-sampling. This section selected EF-Dense to output the feature image after four down-sampling cycles. Additionally, we present the results of MDEFNET-LL, MDEFNET-HL, and our network after the third down-sampling process and feature fusion of the main branch.

Through the display of the feature heat map in Figure 7, it can be seen that the multi-branch network was roughly correct for the medium-to-high sensitivity change area. Through the network of multi-branch structure feature fusion, the building change area information can basically be learned. However, due to feature fusion at different depths of the network, it will cause the misidentification of features and loss of information, as shown in the feature map of MDEFNET-LL, if fusion is performed when the network obtains low-level feature information. Although the building change area can be learned with high weight, the other subtle changes on the image pair, such as grass becoming other types of features, are also highly sensitive. Similarly, to the MDEFNET-HL, after the deep-level main and accessory branch information is fused, they are extremely sensitive to building

changes, but this is limited to areas with obvious changes. Under the influence of the same spectrum of foreign objects, it cannot be better identified.
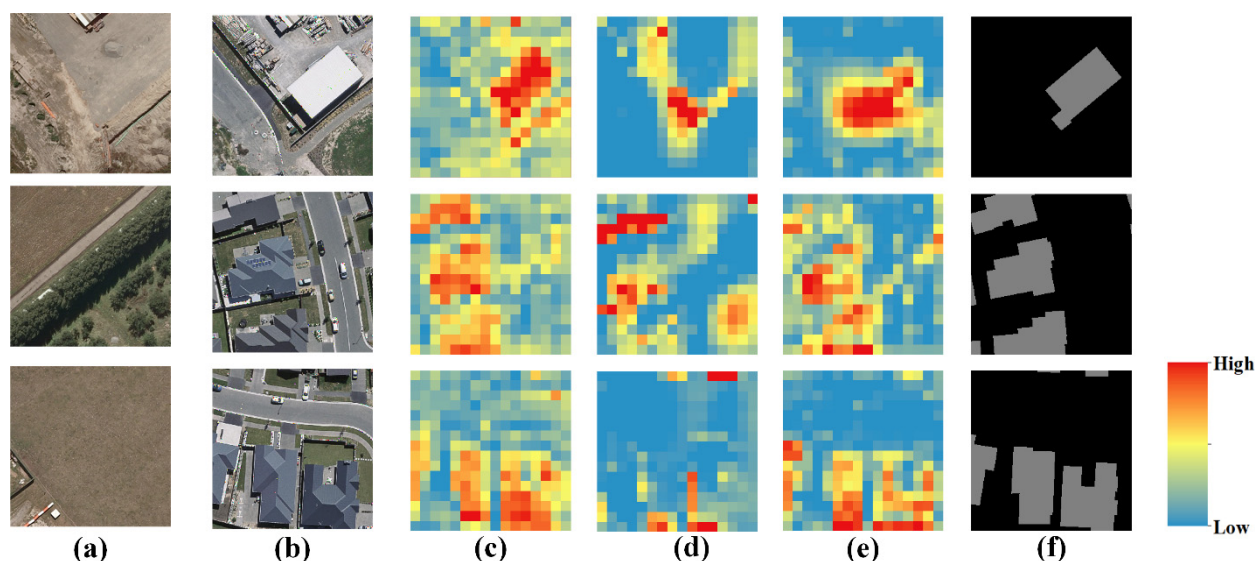

Figure 7. (**a**) Before image; (**b**) After image; (**c**) MDEFNET-LL; (**d**) MDEFNET-HL; (**e**) MDEFNET; (**f**) Label.

In the test image results in Figure 8, the above two methods reflect different degrees of information loss and inaccurate boundary ranges. The method in this paper uses the feature heat map display and the fusion of the feature information in the middle part of the network down-sampling, which can accurately learn the features of the building changes. Although there is a small degree of inaccurate extraction in the detection results, the accuracy and completeness of the building change detection are greatly improved compared with the above methods. This is mainly because we chose the right time for information fusion, so that the accessory branch information could better guide the main branch to learn about building changes.
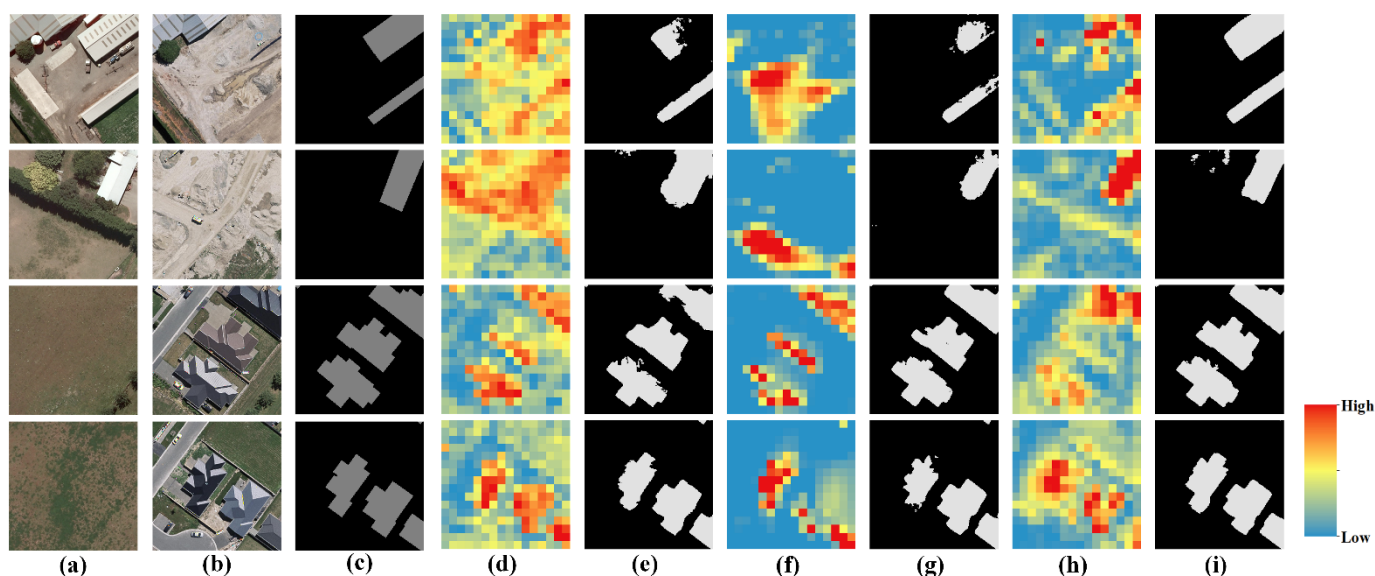

Figure 8. (**a**) Before image; (**b**) After image; (**c**) Label; (**d**) the feature heat map of MDEFNET-LL; (**e**) the result of MDEFNET-LL; (**f**) the feature heat map of MDEFNET-HL; (**g**) the result of MDEFNET-HL; (**h**) the feature heat map of Ours; (**i**) the result of Ours.

After training, we applied the above models to the entire test area and performed quantitative analysis on each model. In Table 3, we have listed the accuracy evaluation results of the above methods for the entire test area. Among them, the method presented in this paper achieved high accuracies of 0.8526, 0.9418, and 0.9204 on the three evaluation indexes of IoU, Recall and F1 Score, respectively. For MDEFNET-shallow fusion, which performed well in Recall, this method improved the accuracy by nearly 15%. Although the above methods are proposed in the process of ablation experiments in this article, the method proposed in this article could develop a good correlation with the range of the building changes in the test area. It can accurately identify the area of building changes, with fewer omissions in image extraction, which verifies the effectiveness of this method in detecting building changes.

Table 3. The evaluation metrics results of ablation II experiments.

	IoU	**Recall**	**F1 Score**
MDEFNET-LL	0.7182	0.7976	0.8360
MDEFNET-HL	0.7344	0.7786	0.8469
Ours	**0.8526**	**0.9418**	**0.9204**

4.3. Comparison Experiments

In order to verify that our method is effective in detecting building changes in this dataset, we selected several deep learning models that have been effective in the field of image recognition for comparison.

SegNet [39]: This network uses VGG16 to extract features of images and connects the output of the encoder to do a non-linear up-sampling.

FC-EF [30]: Fully Convolutional Early Fusion. It uses early fusion theory to stack image pairs. The network is single-branch and uses skip connections to connect the up-sampling and down-sampling.

FC-Siam-conc and FC-Siam-diff [30]: Fully Convolutional Siamese-Concatenation and Fully Convolutional Siamese-Difference. Based on FC-EF, the two networks use Siamese structure. The FC-Siam-conc connects the two down-sampling streams and the other calculates the difference between the feature results.

DASNet [31]: Dual Attentive Fully Convolutional Siamese Network. It includes channel attention mechanism and spatial attention mechanism and uses WDMC loss to balance.

STANet [40]: Spatial–Temporal Attention Neural Network. This method uses spatial-temporal attention module to extract feature of changes.

Table 4, Figures 9 and 10 show the results of each model and the method in the test area. From Table 4, compared with the more classic SegNet network model in the field of semantic segmentation, the model in this paper presented nearly 6%, 8% and 4% improvements in IoU, Recall, and F1 Score metrics, respectively. Compared to FC-EF, FC-Siam-conc and FC-Siam-diff, which have exhibited good performance in change detection in recent years, the accuracy of the model in this paper can be improved by 4%, 8% and 3%. For an index that focuses on changes in patches, MP was used. In terms of building omissions, the method in this article is the same as FC-Siam-conc, better than other methods. The method presented in this paper performed better than FC-Siam-conc in other indexes. Meanwhile, we have compared with two recent deep learning change detection methods, DASNet and STANet. Our method can also achieve better results. Therefore, in general, the proposed method is superior to other deep learning methods. From a network point of view, a network suitable for single-phase target extraction usually cannot achieve good learning features for complex information after the fusion of two image bands. Additionally, a two-branch network suitable for change detection does not need to perform early fusion operations. Therefore, it is possible to process the semantic information of certain times separately. This paper studied the detection of building changes and target segmentation

in a specific semantic scene. Without the guidance of semantic information in detecting building changes, the two-phase change information could not be distinguished well.

Table 4. Comparisons between MDEFNET and several typical deep learning models.

	IoU	Recall	F1 Score	MP
FC-EF	0.7954	0.8696	0.8861	16
FC-Siam-conc	0.8163	0.8651	0.8988	8
FC-Siam-diff	0.7955	0.8689	0.8861	12
SegNet	0.7263	0.8415	0.8414	11
DASNet	0.7941	0.8582	0.8843	9
STANet	0.7056	0.7333	0.8274	18
Ours	**0.8526**	**0.9418**	**0.9204**	**8**

Figure 9. (**a**) Before image. (**b**) After image. (**c**) Label. (**d**) FC-EF. (**e**) FC-Siam-conc. (**f**) FC-Siam-diff. (**g**) SegNet. (**h**) DASNet. (**i**) STANet. (**j**) Ours.

Figure 10. (**a**) Before image. (**b**) After image. (**c**) Label. (**d**) FC-EF. (**e**) FC-Siam-conc. (**f**) FC-Siam-diff. (**g**) SegNet. (**h**) DASNet. (**i**) STANet. (**j**) Ours.

This section also interprets some of the extraction results of each method in the study area, as shown in Figures 9 and 10. Among them, the part of extraction results of DASNet and STANet were similar to the model in this paper and could achieve better results. However, in other parts, the model in this paper still gets better extraction effects. In Figure 9, it can be seen that in the factory change area, because fewer training samples of the public dataset were used, the other methods could not learn the change features with fewer training samples. In the main and accessory branch structure adopted in this article, the main branch could learn the change feature. The accessory branch superimposed the buildings in the changed area to the main branch through the feature-enhanced module. Thereby, the accessory branches can guide the building semantic information to the main branch. The extraction effect of factory changes in the study area is better. For example, for the small patch of building changes in the fourth pair of images, the method in this paper is also superior to other methods for extraction, mainly due to the network structure and Global Information Enhancement module proposed in this paper. The network structure

makes the detailed features after feature fusion more abundant, which is better for the extraction of such small patches. In Figure 10, regarding the boundary accuracies of the building change areas, whether for residential houses with complex boundaries or factories with simple boundaries, the method in this paper performed better than other deep learning methods.

The model in this paper can also identify some small area building changes through training. As Figure 11 shows, in this public dataset, the researchers omitted a small amount of building change detection area in the label. Compared with other models these building changes are not detected or extract well, our model proposed in this paper can still identify them well.
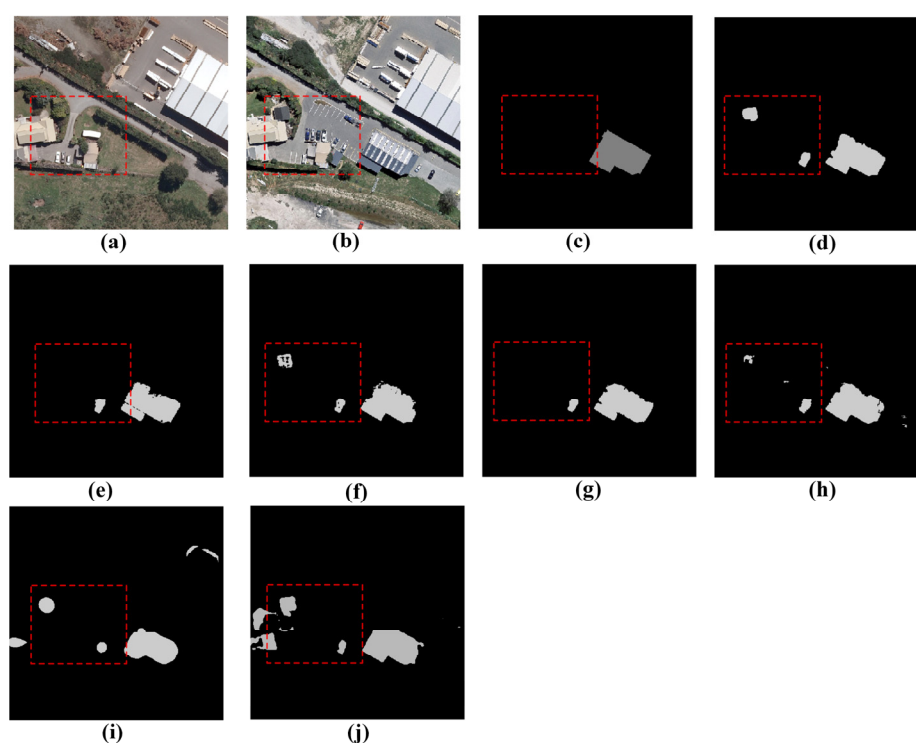

Figure 11. (**a**) Before image. (**b**) After image. (**c**) Label. (**d**) Ours. (**e**) FC-EF. (**f**) FC-Siam-conc. (**g**) FC-Siam-diff. (**h**) SegNet. (**i**) DASNet. (**j**) STANet.

5. Discussion

Scholars improve the effect of deep learning in building change detection usually in two ways: (i) pre-processing of images pairs data; and (ii) the way of model training. Among the first way, constructing a difference map is an important step in change detection. Researchers get the effective information by optimizing the features to make it more distinguishable from the interference information. However, some effective information will be lost, and error propagation will occur, which greatly affects the accuracy and reliability of change detection result no matter how well the model works. Therefore, to retain the information of original image pairs and improve the way of model training, the end-to-end structure of CNN should be used in the way of improving. The method in this paper is inspired by the Siamese network. We chose the multi-branch structure to construct the model for processing the effective features in image pairs separately. The dense block that has already showed good fitting ability was used to connect the whole network.

In the ablation experiment and the comparison experiment, we made a comparison with single-branch networks. In Table 2, the results show that our multi-branch structure is obviously more suitable for change detection tasks. For image pairs learning, a multi-branch network is more reasonable. As in Figures 7 and 8, the feature heat image shows that different depth of semantic information influences the effect of the feature enhanced. For factory changes which have few training samples, we proposed the feature-enhanced

module, and obtained more complete and accurate results than with other methods. The main reason is that not only the network can be performed well for non-linear fitting of the change area, but the semantic information provided by the auxiliary branch enhanced the feature. Meanwhile, a major problem in the research process of change detection methods was also presented. The model feature learning caused by insufficient sample size is not sufficient. In the current several deep learning change detection methods, different spatial attention modules are also used to enhance feature information. In the future deep learning methods in building change detection, we will use different methods which can enhance the feature in multi-temporal images or consider generative adversarial networks (GAN) and other small-sample deep learning techniques to solve the problem of small sample size and complex information.

6. Conclusions

In this paper, based on the WHU Building Change Detection Dataset, MDEFNET has been proposed to automatically extract the building change area on the dataset. In order to make the network more focused on the extraction of buildings, we designed the network structures of the main branch and accessory branch. The main branch mainly learned the change feature, and accessory branches guided the semantic information of building changes, focusing the network on areas of building changes. At the same time, the method proposed in this paper used dense block to connect the entire network, which can reduce the amount of network calculation and has a certain degree of anti-fitting. We also designed a feature-enhanced module, which adopted a global information enhancement module; as the depth of the network increased, the initial detailed information was not lost. MDEFNET has been proven capable of effectively extracting the changing areas of buildings and reducing the interference of background information. The experimental results show that the method presented in this paper obtained accuracies of 0.8526, 0.9418, and 0.9204 in IoU, Recall, and F1 Score metrics, respectively, which could lead to accurately and efficiently identifying areas of change in buildings from aerial images. Meanwhile, in the experimental process of this article, the omission of annotations in the public dataset by the proposed network was checked to a certain degree. Due to the sample imbalance of building change detection, we will focus on improving the robustness of small-sample supervised learning and transfer learning for these tasks in the future.

Author Contributions: J.X. and B.W. designed the experiments; Y.W., H.X., L.C. and P.W. contributed analysis tools; J.C. performed the experiments; J.X. and H.Y. wrote the paper. All authors have read and agreed to the published version of the manuscript.

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

References

1. Zhong, Q.; Ma, J.; Zhao, B.; Wang, X.; Zong, J.; Xiao, X. Assessing spatial-temporal dynamics of urban expansion, vegetation greenness and photosynthesis in megacity Shanghai, China during 2000–2016. *Remote Sens. Environ.* **2019**, *233*, 111374. [CrossRef]
2. Li, Q.; Wang, W.; Wang, J.; Zhang, J.; Geng, D. Exploring the relationship between InSAR coseismic deformation and earthquake-damaged buildings. *Remote Sens. Environ.* **2021**, *262*, 112508. [CrossRef]
3. De Moel, H.; Jongman, B.; Kreibich, H.; Merz, B.; Penning-Rowsell, E.; Ward, P.J. Flood risk assessments at different spatial scales. *Mitig. Adapt. Strateg. Glob. Chang.* **2015**, *20*, 865–890. [CrossRef] [PubMed]
4. Afaq, Y.; Manocha, A. Analysis on change detection techniques for remote sensing applications: A review. *Ecol. Inform.* **2021**, *63*, 101310. [CrossRef]
5. Chen, J.; Liu, H.; Hou, J.; Yang, M.; Deng, M. Improving Building Change Detection in VHR Remote Sensing Imagery by Combining Coarse Location and Co-Segmentation. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 213. [CrossRef]
6. Zhang, Z.; Li, Z.; Tian, X. Vegetation change detection research of Dunhuang city based on GF-1 data. *Int. J. Coal Sci. Technol.* **2018**, *5*, 105–111. [CrossRef]

7. 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.* **2018**, *56*, 1566–1578. [CrossRef]

8. Deng, J.S.; Wang, K.; Deng, Y.H.; Qi, G.J. PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. *Int. J. Remote Sens.* **2008**, *29*, 4823–4838. [CrossRef]

9. Eid, A.N.M.; Olatubara, C.O.; Ewemoje, T.A.; El-Hennawy, M.T.; Farouk, H. Inland wetland time-series digital change detection based on SAVI and NDWI indecies: Wadi El-Rayan lakes, Egypt. *Remote Sens. Appl. Soc. Environ.* **2020**, *19*, 100347. [CrossRef]

10. Niemeyer, I.; Marpu, P.R.; Nussbaum, S. Change detection using object features. In *Object-Based Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 154–196.

11. Hussain, M.; Chen, D.; Cheng, A.; Wei, H.; Stanley, D. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS J. Photogramm. Remote Sens.* **2013**, *80*, 91–106. [CrossRef]

12. Wang, B.; Choi, J.; Choi, S.; Lee, S.; Wu, P.; Gao, Y. Image Fusion-Based Land Cover Change Detection Using Multi-Temporal High-Resolution Satellite Images. *Remote Sens.* **2017**, *9*, 804. [CrossRef]

13. Haiquan, F.; Yunzhong, J.; Yuntao, Y.; Yin, C. River Extraction from High-Resolution Satellite Images Combining Deep Learning and Multiple Chessboard Segmentation. *Acta Sci. Nat. Univ. Pekin.* **2019**, *55*, 692–698.

14. Gong, M.; Yang, Y.-H. Quadtree-based genetic algorithm and its applications to compzter vision. *Pattern Recognit.* **2004**, *37*, 1723–1733. [CrossRef]

15. Baraldi, A.; Boschetti, L. Operational Automatic Remote Sensing Image Understanding Systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA). Part 1: Introduction. *Remote Sens.* **2012**, *4*, 2694–2735. [CrossRef]

16. Hou, X.; Bai, Y.; Li, Y.; Shang, C.; Shen, Q. High-resolution triplet network with dynamic multiscale feature for change detection on satellite images. *ISPRS J. Photogramm. Remote Sens.* **2021**, *177*, 103–115. [CrossRef]

17. Khelifi, L.; Mignotte, M. Deep Learning for Change Detection in Remote Sensing Images: Comprehensive Review and Meta-Analysis. *IEEE Access* **2020**, *8*, 126385–126400. [CrossRef]

18. Wang, Q.; Zhang, X.D.; Chen, G.Z.; Dai, F.; Gong, Y.F.; Zhu, K. Change detection based on Faster R-CNN for high-resolution remote sensing images. *Remote Sens Lett* **2018**, *9*, 923–932. [CrossRef]

19. Wang, Z.; Zheng, X.; Li, D.Y.; Zhang, H.L.; Yang, Y.; Pan, H.G. A VGGNet-like approach for classifying and segmenting coal dust particles with overlapping regions. *Comput Ind* **2021**, *132*, 103506. [CrossRef]

20. Xiao, Y.Q.; Pan, D.F. Robust Visual Tracking via Multilayer CaffeNet Features and Improved Correlation Filtering. *IEEE Access* **2019**, *7*, 174495–174506. [CrossRef]

21. Afify, H.M.; Mohammed, K.K.; Hassanien, A.E. An improved framework for polyp image segmentation based on SegNet architecture. *Int. J. Imag. Syst Technol.* **2021**, *31*, 1741–1751. [CrossRef]

22. Moustafa, M.S.; Mohamed, S.A.; Ahmed, S.; Nasr, A.H. Hyperspectral change detection based on modification of UNet neural networks. *J. Appl. Remote Sens* **2021**, *15*, 028505. [CrossRef]

23. Papadomanolaki, M.; Vakalopoulou, M.; Karantzalos, K. A Deep Multitask Learning Framework Coupling Semantic Segmentation and Fully Convolutional LSTM Networks for Urban Change Detection. *IEEE T Geosci Remote.* **2021**, *99*, 1–18.

24. Gong, M.; Yang, H.; Zhang, P. Feature learning and change feature classification based on deep learning for ternary change detection in SAR images. *ISPRS J. Photogramm. Remote Sens.* **2017**, *129*, 212–225. [CrossRef]

25. Qian, J.; Xia, M.; Zhang, Y.; Liu, J.; Xu, Y. TCDNet: Trilateral Change Detection Network for Google Earth Image. *Remote Sens.* **2020**, *12*, 2669. [CrossRef]

26. Samadi, F.; Akbarizadeh, G.; Kaabi, H. Change detection in SAR images using deep belief network: A new training approach based on morphological images. *Iet. Image Process* **2019**, *13*, 2255–2264. [CrossRef]

27. Gao, F.; Dong, J.; Li, B.; Xu, Q. Automatic Change Detection in Synthetic Aperture Radar Images Based on PCANet. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1792–1796. [CrossRef]

28. Wang, Q.; Yuan, Z.; Du, Q.; Li, X. GETNET: A General End-to-End 2-D CNN Framework for Hyperspectral Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3–13. [CrossRef]

29. Peng, D.; Zhang, Y.; Guan, H. End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++. *Remote Sens.* **2019**, *11*, 1382. [CrossRef]

30. Daudt, R.C.; Saux, B.L.; Boulch, A. Fully Convolutional Siamese Networks for Change Detection. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 4063–4067.

31. 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]

32. 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]

33. Chen, H.; Wu, C.; Du, B.; Zhang, L.; Wang, L. Change Detection in Multisource VHR Images via Deep Siamese Convolutional Multiple-Layers Recurrent Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 2848–2864. [CrossRef]

34. Huang, L.; An, R.; Zhao, S.; Jiang, T.; Hu, H. A Deep Learning-Based Robust Change Detection Approach for Very High Resolution Remotely Sensed Images with Multiple Features. *Remote Sens.* **2020**, *12*, 1441. [CrossRef]

35. Fang, B.; Pan, L.; Kou, R. Dual Learning-Based Siamese Framework for Change Detection Using BiTemporal VHR Optical Remote Sensing Images. *Remote Sens.* **2019**, *11*, 1292. [CrossRef]

36.   Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. Learning a Discriminative Feature Network for Semantic Segmentation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.

37.   Ji, S.; Wei, S.; Lu, M. Fully Convolutional Networks for Multi-Source Building Extraction from An Open Aerial and Satellite Imagery Data Set. *Ieee T. Geosci. Remote* **2019**, *57*, 574–586. [CrossRef]

38.   Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.

39.   Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]

40.   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]