



Article Built-Up Area Change Detection Using Multi-Task Network with Object-Level Refinement

Song Gao ^{1,†}, Wangbin Li ^{1,†}, Kaimin Sun ^{1,*}, Jinjiang Wei ¹, Yepei Chen ² and Xuan Wang ³

- State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China; gaosong@whu.edu.cn (S.G.); lwb1996@whu.edu.cn (W.L.); weijinjiang@whu.edu.cn (J.W.)
- ² School of Computer Science, Hubei University of Technology, Wuhan 430079, China; chenyp@whu.edu.cn
- ³ School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China; xuan.wang@whu.edu.cn
- * Correspondence: sunkm@whu.edu.cn
- + These authors contributed equally to this work.

Abstract: The detection and monitoring of changes in urban buildings, as a major place for human activities, have been considered profoundly in the field of remote sensing. In recent years, comparing with other traditional methods, the deep learning-based methods have become the mainstream methods for urban building change detection due to their strong learning ability and robustness. Unfortunately, often, it is difficult and costly to obtain sufficient samples for the change detection method development. As a result, the application of the deep learning-based building change detection methods is limited in practice. In our work, we proposed a novel multi-task network based on the idea of transfer learning, which is less dependent on change detection samples by appropriately selecting high-dimensional features for sharing and a unique decoding module. Different from other multi-task change detection networks, with the help of a high-accuracy building mask, our network can fully utilize the prior information from building detection branches and further improve the change detection result through the proposed object-level refinement algorithm. To evaluate the proposed method, experiments on the publicly available WHU Building Change Dataset were conducted. The experimental results show that the proposed method achieves F1 values of 0.8939, 0.9037, and 0.9212, respectively, when 10%, 25%, and 50% of change detection training samples are used for network training under the same conditions, thus, outperforming other methods.

Keywords: change detection; multi-task change detection network; remote sensing; object-level refinement algorithm

1. Introduction

Remote sensing-based change detection includes difference identification in remote sensing images of the same area acquired at different times. It plays an important role in the application of remote sensing images and has been widely used in many areas, including land usage and cover change, forest and vegetation change, natural disaster monitoring, and urban development [1–7]. Recently, due to the development of new concepts such as smart city, automated change detection methods of urban buildings have received much attention from the scientific community.

Currently, building change detection methods based on remote sensing images can be roughly divided into traditional methods, including the principal component analysis (PCA) [8–10], change vector analysis (CVA) [11–13], and texture-based transforms [14,15], and deep learning-based methods, including the fully convolutional neural network (FCN) [16–18], deep belief network (DBN) [19–21], and principal component analysis network (PCANet) [22–24]. Regardless of many limitations, the traditional methods still occupy a certain position in the field of specific change detection because they require fewer training samples. However, the traditional methods have certain disadvantages in accuracy,



Citation: Gao, S.; Li, W.; Sun, K.; Wei, J.; Chen, Y.; Wang, X. Built-Up Area Change Detection Using Multi-Task Network with Object-Level Refinement. *Remote Sens.* 2022, *14*, 957. https://doi.org/10.3390/ rs14040957

Academic Editor: Jan Platoš

Received: 6 January 2022 Accepted: 11 February 2022 Published: 16 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/). reliability, and practicability compared to the deep learning-based methods. Due to the limitations of traditional methods, an increasing number of studies on change detection have been based on deep learning-based methods.

Deep learning-based change detection methods can be classified into two types: postclassification comparison methods and direct comparison methods. The post-classification comparison methods [25,26] first extract the ground objects of two images and then compare the extraction results to obtain the change area. Although this approach can reduce the demand for change detection samples, the imaging conditions of two compared images cannot be the same, that is, even with the area unchanged, the result of the ground object extraction will be different, and these errors will be introduced into the subsequent comparison results. The direct comparison methods [27–29] directly compare the high-dimensional features of two images to obtain the change region by constructing a network. Although this approach can overcome the above-mentioned shortcomings of the post-classification comparison methods, the number of required training samples is often larger. It should be noted that unannotated change detection samples are difficult to obtain, let alone an annotated change detection dataset. Therefore, reducing incorrect detections when using fewer samples can help to improve change detection methods.

This paper proposes an end-to-end change detection method called the multi-task change detection network, which is based on prior knowledge of building detection. The proposed method can achieve high accuracy with only a few building change detection samples.

- (1) Aiming to solve the problem of the low accuracy of the change detection methods based on deep learning in the case of a small number of samples, a new multi-task change detection method is proposed. The proposed method uses the building extraction dataset to pre-train the building extraction branch so that the feature extraction module of the network can better extract the building features. Therefore, only a small number of building transformation detection samples are required for the follow-up network training while achieving an excellent detection effect.
- (2) To enhance the network's ability to extract the features of a building, a building detection branch is added to the network, and a multi-task learning strategy is used in network training so that the network has a stronger feature extraction capability.
- (3) For making full use of the results of the building extraction branch to improve the change detection accuracy, a object-level refinement algorithm is proposed. Combined with the results of the change detection branch and building extraction branch, the proposed method selects the building mask with change area greater than a predefined threshold as a final change detection result, which improves the accuracy and visual effect of change detection results.

The effectiveness of the proposed method is verified by the experiments on a public dataset. The experimental results show that under the same conditions, the proposed method is as effective as the existing advanced methods, although it uses a smaller number of change detection samples. When all compared methods use the same number of change detection samples, the proposed method is superior to the advanced methods.

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 introduces the proposed method in detail. Section 4 presents the detailed experimental results of the proposed method. Section 5 discusses open issues, and finally, Section 6 concludes the paper.

2. Previous Work

There have been many studies on change detection methods based on deep learning. This section briefly reviews several advanced change detection methods.

At present, most of the deep learning-based change detection methods have focused on how to improve the accuracy and reduce the need for annotated samples. As a kind of method that has low requirements for annotated samples, post-classification comparison change detection methods [25,26] have certain significance in application. Methods of this kind usually adopt segmentation network (e.g., mask region-based convolutional neural network, U-Net and DeepLabv3+) to extract object masks firstly, and then compared the mask of two image to obtain change detection results. Although this approach can reduce the requirements for annotated samples and does not require using real change detection samples, the change detection accuracy is highly dependent on the building detection accuracy. Considering the similarities between the building detection and change detection tasks, multi-task learning-based methods [28,29] were proposed. By employing the multi-task learning strategy, these models can use the connection between the two tasks to improve the accuracy, but it requires a large number of training samples, and these methods proposed do not fully utilize the results generated by the object detection branches. To reduce the number of samples required for change detection, methods [30–34] based small samples were proposed. Among them, the approaches of Chen et al. [31] and Zheng et al. [32] stand out. Chen et al. proposed a synthesis framework, namely, Instancelevel change Augmentation (IAug), to generate change instances in unchanged regions by leveraging generative adversarial training. In this way, the samples can be augmented and the risk of class imbalance can also be reduced. However, their method also has the problem that the generated architectural image may not be faithful with the image in the target domain, which will affect the training effect of the network. Zheng et al. proposed a single-temporal supervised learning (STAR) method, which leverages different regions in a single period of image to construct pseudo bitemporal image pairs. This method can train change detection network by only using unpaired labeled images. There are also some unsupervised deep change detection methods [35-38], but their accuracy still has a lot of room to improve.

Currently, there is no change detection method that can handle all application scenarios. Due to the limitation of the performance of the object detection network, the accuracy of post-classification comparison change detection methods is difficult to be greatly improved. The large demand for annotated samples of the direct comparison method limits its application. Although some data augmentation methods are mentioned above, they provide limited usefulness. The problem of relying on a large number of labeled samples is that, although unsupervised change detection methods overcome many obstacles, such methods still require a lot of work.

Inspired by the above-mentioned methods, this paper proposes a new change detection method based on a convolutional neural network trained with a small dataset. This method makes full use of the correlation between the building extraction branch and the change detection branch and improves the accuracy of change detection results by making full use of the building extraction branch.

3. Proposed Methods

In this section, the proposed change detection method is described in detail. As shown in Figure 1, the proposed method includes four main parts: a building detection branch, a change detection branch, a learning strategy (pre-training and multi-task learning), and an object-level refinement algorithm.

3.1. Building Detection Branch

At present, there are many building extraction methods based on convolutional neural networks (CNNs) [39–41]. Among CNNs, the U-Net has been widely used in building extraction applications due to its high segmentation accuracy and fine edges. Among many improved building extraction algorithms based on the U-Net, the improved U-Net algorithm proposed by Wei et al. [42] has the advantages of high accuracy on multiple open datasets and stronger multi-scale feature fusion ability; so, this network is selected as a building extraction module in the proposed method. The structure of the improved U-Net is shown in Figure 2, where it can be seen that it consists of two parts. The left part denotes the feature extraction structure of the network. Through four pooling layers, the feature extraction network generates a total of five feature maps with different scales. Correspondingly, the decoding structure on the right uses four upsampling layers to ensure that the segmentation results can be restored to the original size. Moreover, the decoding

structure of the network adopts the skip connection structure on the same stage, which makes the decoding structure integrate more low-level features when recovering the feature map so that the edge information of the network segmentation result can be more refined. Compared to the traditional U-Net, the feature extraction structure in Figure 2 has three additional convolution layers, and a certain number of batch normalization (BN) layers are introduced to realize feature normalization and to accelerate network convergence.

In the network structure shown in Figure 2, features P2, P3, P4 and P5 contain richer semantic and structural information in the same stage [42], which can also be used in the change detection branch. Therefore, the difference of features P2, P3, P4 and P5 are selected as the input of the change detection branch. See section for specific technical details.



Figure 1. Overview of the proposed method. The building extraction modules have been pre-trained. Modules of the same color have the same weight.



Figure 2. The improved U-Net Structure.

3.2. Change Detection Branch

The feature extraction structure in the building detection branch can provide not only building feature information but also necessary information on building change features for change detection. As shown in Figure 1, the input of the change detection branch is Pd2–Pd5, which can be obtained as follows:

$$Pdi = Pi - P'i, \tag{1}$$

where Pi represents the feature i extracted from image A, and P'i represents the feature *i* extracted from image B.

The detailed structure of the change detection branch is shown in Figure 3, where it can be seen that it includes two parts, the neck part and the head section. The neck part is used to fuse multi-scale difference feature maps Pd2–Pd5. The multi-scale feature fusion modules, such as path aggregation network (PANet), neural architecture search feature pyramid network (NAS-FPN), and bi-directional feature pyramid network (BiFPN) [43–45], have been widely studied. Considering that the proposed change detection method is based on a small number of change detection samples, to avoid the overfitting problem, a feature pyramid network (FPN) with fewer parameters and a simple structure is chosen as a feature fusion module of the change detection branch. As a classic feature fusion method [46], the FPN has been used in many studies [47–49]. The FPN represents a top-down network structure with lateral connections, which can integrate and make full use of the features of multiple scales in the network, making the segmentation result more refined.





The head section of the change detection branch refers to the semantic segmentation branch introduced by Kirillov et al. [50]. In this part, four high-dimensional feature maps with different scales are restored to the original size by the repeated structure consisting of two upsampling, convolution, BN, and rectified unit (ReLU) layers first and then summed up element by element. Finally, a pixel-based segmentation map with the same resolution as the original one is generated by a 3×3 convolution kernel and sigmoid function.

3.3. Pre-Training and Multi-Task Learning

Due to their powerful feature extraction capabilities, CNNs have become a common method for remote sensing image processing. However, their powerful feature extraction ability requires using a large number of learning samples. In the change detection task, it is very difficult to obtain the pixel-level labeled change detection samples; therefore, reducing the dependency on the number of samples is the main challenge to be addressed. As a common training strategy in CNN, pre-training based on transfer learning can solve the problem of insufficient samples. In the proposed method, the building extraction samples are used to pre-train the building detection branch with shared weights to extract building features better. The high-dimensional feature map of building change required by the change detection branch can be obtained by (1). In this way, the highly accurate extraction effect can be achieved while using a small number of change detection samples for training. In the pre-training stage, the cross-entropy function is used as a loss function, and it can be defined as follows:

$$loss = -\frac{1}{N} \sum_{n=1}^{N} [y_n \log \hat{y}_n + (1 - y_n) \log (1 - \hat{y}_n)],$$
(2)

where *N* represents the number of samples; y_n is the true value of the samples, which can be 0 or 1, and \hat{y} is the prediction of the network, ranging from 0 to 1.

As a way for a machine to imitate human learning activities, multi-task learning can transfer similar knowledge from one task to another [51]. In the proposed method, due to the similarity between the change detection task and the building extraction task, the multi-task learning strategy is adopted to make full use of the useful information in multiple learning tasks to help each branch learn data patterns better. During the proposed network training, supervised training of two branches is performed at the same time, and each branch adopts cross-entropy as a loss function. The training strategy is given by:

$$\mathcal{L}oss = \sum_{i=1}^{Q} loss_i,\tag{3}$$

where Q denotes the number of branches, $loss_i$ is the loss of a branch i.

Our training strategy is as follows: firstly, we only train the building detection module (as shown in Figure 2) in the pre-training stage, and then we adopt the multi-task learning strategy to train the building extraction and change detection modules simultaneously.

3.4. Object-Level Refinement Algorithm

In the proposed change detection method, the segmentation results of the building detection branch have high accuracy and fine edges, and the segmentation results of the two branches (building mask and change building mask) are similar. Therefore, the changing mask can be used to pick out the two-phase building mask and to retain the change building mask. For instance, in the area where no buildings are detected in the two-phase images, the change detection results will be considered as a pseudo-change if changes were detected in this area. In this way, the pseudo-change can be eliminated, thereby improving detection accuracy.

Based on this idea, an object-level refinement algorithm is developed, and the specific process is shown in Figure 4.

The object-level refinement algorithm includes three steps. The first step is to obtain the union m1 of two phases of the building mask. The second step is to overlap m1 with the change detection mask to obtain m2. The third step is to process m2. The processing rules are as follows:

- (1) If a change mask is detected in the area where a building mask has been detected, and the number of the pixels in the intersection of the two masks divided by the number of the pixels in the building mask is greater than the predefined threshold *α*, then the building mask is set as a result of the processed change mask.
- (2) If a change mask is detected in the area where a building mask has been detected, and the number of the pixels in the intersection of the two masks divided by the number of the pixels in the building mask is less than the predefined threshold *α*, then all masks in the area are set as non-change as a processed result.

- (3) If there is only a change mask and no building mask in an area, all masks in the area are set as non-change as a processed result.
- (4) If there is only a building mask and no change mask in an area, all masks in the area are set as non-change as a processed result.



Figure 4. Object-level refinement algorithm. (1) The first step is to find the union of the masks of the two phases of buildings. (2) The second step is to overlap the union of the building masks with the change mask. (3) The third step is to pick out the building mask that meets the threshold requirements as a final change detection result.

In the subsequent experiments, it is shown that the processed change detection results not only have a significant improvement in the evaluation index but the mask edges of the changes are also more refined.

4. Experiments and Analysis

To evaluate the proposed method, a series of ablation experiments were conducted, and the proposed method was compared with other change detection methods on the WHU Building Change Dataset [52].

4.1. Dataset and Implementation Details

The WHU Building Dataset [52] was used for model pre-training, and the WHU Building Change Dataset was used to evaluate the proposed method. The study area was located in Christchurch, New Zealand, having an area of 450 km² at a 0.2-m spatial resolution, as shown in Figure 5. In Figure 5, the red box is a part of the WHU Building Dataset, and the images of this dataset were obtained in 2016; the yellow box denotes the WHU Building Change Dataset, whose images were obtained in 2011 and 2016. The images in the black (A), blue (B), and green (C) boxes are used as training data, and they accounted for 10%, 25%, and 50% of the WHU Building Change Dataset, respectively. In Figure 5, the images in the purple (D) box are used as test data, which accounted for 50% of the WHU Building Change Dataset; 10% of the test data were randomly selected for verification. The images were recorded into 512×512 patches without overlapping and then resized to 256 \times 256. Finally, 1702 tiles were included in the red box, and the black box included 192×2 patches; 480×2 patches were included in the blue box, and 960×2 patches were included in the green box. The purple box included 960 × 2 patches. In order to prove that our method also has advantages in general cases, we also cropped the dataset into training, validation and test sets in a ratio of 8:1:1.



Figure 5. Study area in Christchurch, New Zealand, at a 0.2-m spatial resolution. The red and yellow boxes indicate the pre-training data and change detection data, respectively.

The proposed algorithm was implemented using the Pytorch framework on a PC with Windows 10 system. In model pre-training, the Adam optimizer with a batch size of 16 and an epoch number of 150 was used. The initial learning rate was 1×10^{-3} , and it which decreased to 5×10^{-4} and 2.5×10^{-4} when the epoch number was 80 and 120, respectively. Furthermore, random resized crop, rotation, and horizontal flip and color jitter were used as data augmentation methods. In training, the Adam optimizer with the same batch size and epoch number as in pre-training was used. The initial learning rate was set to 1×10^{-4} , and it which decreased to 5×10^{-5} and 2.5×10^{-5} when the epoch number was 80 and 120, respectively. The data augmentation methods were the same as those used in pre-training. All experiments were performed on a device with an Inter(R) Xeon(R) Silver 4114 CPU, 128-GB RAM and an NVIDIA TITAN 24-GB GPU. The training took 7 h in total.

In the algorithm evaluation, we use precision, recall and F1 value three as our evaluation indicators:

$$precision = \frac{TP}{TP + FP'},\tag{4}$$

$$recall = \frac{TP}{TP + FN'}$$
(5)

$$F1 = \frac{2 * precision * recall}{precision + recall},$$
(6)

9 of 17

where *TP* is the number of true positives, *FP* is the number of false positives, *TN* is the number of true negatives, and *FN* is the number of false negatives.

4.2. Results of Change Detection

The comparison results of the proposed method and several advanced methods on the WHU Building Change Dataset with the threshold α of the object-level refinement algorithm as 0.5 for different sample sizes are presented in Table 1. As shown in Table 1, the proposed method achieved the highest F1 value among all the methods when 50% of the total change detection samples were used as training samples. When only 10% of the total change detection samples were used as training samples, the proposed method achieved a better performance (F1 value of 0.8939) than most methods, which were trained with a significantly larger number of change detection samples than the proposed method. This was because the proposed method adopts transfer learning to strengthen the network's ability to extract building features. Additionally, high-accuracy building detection branch supplements the missing part of change detection results and also provides necessary information for noise elimination through the object-level refinement algorithm. Therefore, the proposed method can achieve good detection performance under a small training dataset size. As shown in Table 2, we also compared with several current advanced methods in the case of large samples, from which we can see that our method still has advantages.

Table 1. Results on WHU Building Change Dataset (Small Samples).

Method	Year	Precision	Recall	F1 Value
Ji et al., 2019a [25] (25% samples used for training)	2019	0.9520	0.8040	0.8745
Ji et al., 2019a [25] (50% samples used for training)	2019	0.9310	0.8920	0.9111
IFN [53] (50% samples used for training)	2020	0.9026	0.8054	0.8512
STANet [54] (50% samples used for training)	2020	0.7429	0.8935	0.8121
Our method (10% samples used for training)	2022	0.9220	0.8675	0.8939
Our method (25% samples used for training)	2022	0.9173	0.8906	0.9037
Our method (50% samples used for training)	2022	0.9512	0.8930	0.9212

Table 2. Results on WHU Building Change Dataset (Large Samples).

Method	Year	Precision	Recall	F1 Value
DASNet [55] (80% samples used for training)	2020	0.8920	0.9050	0.8980
DTCDSCN [28] (80% samples used for training)	2020	0.9015	0.8935	0.8975
IFN (80% samples used for training)	2020	0.9159	0.9374	0.9265
STANet (80% samples used for training)	2020	0.9007	0.8040	0.8496
SRCDNet [56] (80% samples used for training)	2021	0.8484	0.9013	0.8740
Our method (80% samples used for training)	2022	0.9457	0.9293	0.9374

Figure 6 shows five examples of change detection results of the proposed method and other methods on 256 × 256 patches for different training sample sizes. Based on the comparison results, the proposed method was more accurate and had finer edge details than the other methods. As shown in the second and third rows in Figure 6, Figure 6d,e include more missed detection results than Figure 6f, but the overall difference is not significant. This is because the final change detection mask in the proposed method is obtained by the object-level refinement algorithm, which is actually a screened building mask. Therefore, even if there is a certain detection error or a missed detection in the change detection branch results, it will have little impact on the final refinement results. The effectiveness of the object-level refinement algorithm will be discussed in Section 4.3.



Figure 6. Change detection results of different methods on the WHU Building Change Dataset. (a) 2011 images; (b) 2016 images; (c) ground truth; (d) the proposed method (10% of samples were used for training); (e) the proposed method (25% of samples were used for training); (f) the proposed method (50% of samples were used for training); (g) Ji et al. [25] (50% of samples were used for training); (h) IFN [53] method; (i) STANet [54] method.

Figure 7 shows the final test results of the proposed method in the entire test area for different training sample sizes, where white pixels represent true positives, black pixels represent true negatives, red pixels represent false positives, and blue pixels represent false negatives. From the perspective of the whole test area, the proposed algorithm achieved a high accuracy under different training sample sizes, and as the training sample size increased, the error detection gradually decreased.





Figure 7. Results of the proposed method on the whole test area. (**a**) 2011 image; (**b**) 2016 image; (**c**) 10% training data; (**d**) 25% training data; (**e**) 50% training data.

4.3. Efficiency of Object-Level Refinement Algorithm

Figure 8d,f,h show the results of the proposed method without the object-level refinement algorithm for different training sample sizes. Comparing these results with the results in Figure 8e,g,i, there were more incomplete masks in the results of the change detection branch under 10% and 25% training data. However, after the introduction of the object-level refinement algorithm, the results have been improved to a large extent.



Figure 8. The results of the proposed method before and after the introduction of the object-level refinement algorithm for different sample sizes. (a) 2016 images; (b) 2011 images; (c) ground truth data; (d) unprocessed data (10% of samples were used for training); (e) processed data (10% of samples were used for training); (f) unprocessed data (25% of samples were used for training); (g) processed data (25% of samples were used for training); (h) unprocessed data (50% of samples were used for training); (i) processed data (50% of samples were used for training).

Table 3 shows the comparison of the evaluation metrics before and after introducing the object-level refinement algorithm under different training set sizes. As shown in Table 3, there were large differences in precision, recall, and F1 value under 10% and 25% training data before and after the object-level refinement algorithm processing. This result is consistent with the results in Figure 6, which confirms the effectiveness of the object-level refinement algorithm in the proposed method.

Training Data Size	Processed			Unprocessed		
	Precision	Recall	F1 Value	Precision	Recall	F1 Value
10%	0.9220	0.8675	0.8939	0.8770	0.8247	0.8501
25%	0.9173	0.8906	0.9037	0.8927	0.8879	0.8903
50%	0.9512	0.8930	0.9212	0.9420	0.8966	0.9187

Table 3. Results Before and After being Processed.

In the previous experiments, the threshold α in the object-level refinement algorithm was set to 0.5, which is the threshold selected when neither precision nor recall is the primary concern. In fact, the selection of threshold α has a significant impact on the object-level refinement algorithm results. Figure 9 shows the relationship between threshold α and evaluation metrics, precision, recall, and F1 value when the object-level refinement algorithm was used for different training set sizes.



Figure 9. The relationship between the threshold α and three evaluation indicators, precision, recall, and F1 value.

As shown in Figure 9, with the increase in threshold α , precision showed an upward trend, recall showed a downward trend, and F1 value first increased and then decreased. This was because, with the object-level refinement algorithm, buildings in the area where no change is detected are removed. In the area where the change is detected, with the reasonable increase of threshold α , the object-level refinement algorithm can make up for the missing part of the change detection results and eliminate noise. Although some correct results will inevitably be deleted when eliminating noise, its impact is small. Therefore, the indicator F1, which represents the comprehensive performance, will be improved. As threshold α continues to increase, only results of change detection branch with higher accuracy will be retained. Therefore, precision will increase, while recall and f1 will decrease. Theoretically, the decline of the three indicators of the change detection branch trained by more samples is smaller, and the experimental results shown in Figure 9 also prove this point of view.

Therefore, when selecting the value of threshold α , if it is necessary to ensure that, after being processed by the object-level refinement algorithm, the result has a high recall, a smaller value of α should be selected. In that case, the precision will not decrease much compared with the one where the object-level refinement mechanism is not used . If it is necessary to ensure that the processed result has a higher precision, a larger α should be chosen, but to ensure that the recall does not drop too much, the value of α should not be too large. If the processed results need to be improved regarding the precision, recall, and F1 value at the same time, the value of α needs to be moderate. For instance, on the WHU Building Change Dataset, the moderate range of α is approximately 0.4–0.5.

4.4. Efficiency of Pre-Training and Multi-Task Learning

The experiment presented in Section 4.3 demonstrates the effectiveness of the objectlevel refinement algorithm. As presented above, the effect of the object-level refinement algorithm depends on the performance of the change detection branch. Therefore, to obtain a higher change detection accuracy, it is necessary to improve the accuracy of the change detection branch. When strategies such as pre-training and multi-task learning are not used, the learning effect of a network is relatively poor under a small dataset. Table 4 shows the evaluation indicators of the proposed method when the multi-task learning, pre-training, and object-level refinement algorithm are used for different sample sizes at $\alpha = 0.5$.

Sample Size	Strategy	Precision	Recall	F1 Value
10%	none	0.7627	0.5451	0.6358
	multi-task learning	0.8718	0.7202	0.7888
	multi-task learning and pre-train	0.8770	0.8247	0.8501
	multi-task learning, pre-train and object-level refinement algorithm	0.9220	0.8675	0.8939
25%	none	0.7350	0.8546	0.7903
	multi-task learning	0.8424	0.8959	0.8588
	multi-task learning and pre-train	0.8927	0.8879	0.8903
	multi-task learning, pre-train and object-level refinement algorithm	0.9173	0.8906	0.9037
50%	none	0.8908	0.8824	0.8866
	multi-task learning	0.9271	0.9019	0.9143
	multi-task learning and pre-train	0.9420	0.8966	0.9187
	multi-task learning, pre-train and object-level refinement algorithm	0.9512	0.8930	0.9212

Table 4. Results of Adopting Different Strategies.

As shown in Table 4, when 50% of samples were used for training, strategies such as pre-training and multi-task learning did not significantly improve the precision, recall, and F1 value. This was because, in the case of a sufficient number of samples, the network learned the building features well, and additional learning strategies had a slight impact on the network results. However, when 10% or 25% of samples were used for training, these strategies improved the precision, recall, and F1 value significantly. This was because when the training samples were insufficient, additional learning strategies could significantly improve the feature extraction capability of the network. Due to the similarity between the building detection task and the change detection task, when using multi-task learning, the two tasks could promote each other's learning effect during learning. In this case, building segmentation samples could be used to compensate for the shortage of change detection samples. On this basis, by introducing the pre-training, the network can obtain building feature extraction capabilities in advance through building segmentation samples, thereby further improving the detection accuracy of the network under a small sample number. Since the pre-training and multi-task learning strategies used in this work both use building segmentation samples to improve the network's ability to extract building features, the effect of the building detection branch on the network performance will also be improved. This type of improvement can affect the final change detection result obtained by the object-level refinement algorithm. By applying the above-mentioned strategies, the proposed method makes full use of the building detection branch in both the training phase and the testing phase and compensates for the limitations of other multi-branch change detection methods.

The results of gradually adopting various strategies under 10% training data are shown in Figure 10, where it can be seen that the improvement from column d, where no strategy is used, to column e, where multi-task learning is used, is the most obvious, followed by the improvement from column e with multi-task learning to column f with multi-task learning and pre-training, and finally the improvement from column f with multi-task learning and pre-training to column g with multi-task learning, pre-training, and object-level refinement algorithm. This result was because in the case of insufficient samples, multi-task learning could better enhance the building feature extraction ability of the network, and useful information could be used to enhance the learning effect of each branch. Since the feature extraction capability of the network was enhanced to a certain extent by using multi-task learning, the improvement effect was reduced when the pre-training strategy was adopted, which was the most obvious in the case of 50% training data, as shown in Table 4. In general, each of the adopted strategies had a large effect on the detection result for a small data size.



Figure 10. Samples of Adopting Different Strategies under 10% Training Samples. (**a**) 2016 images. (**b**) 2011 images. (**c**) ground truth. (**d**) none. (**e**) multi-task learning. (**f**) multi-task learning and pre-training. (**g**) multi-task learning, pre-training and object-level refinement.

5. Discussion

In this section, the advantages and future improvements of the proposed method are discussed. The comparison between the proposed method and several advanced deep learning-based change detection methods on the WHU Building Change Dataset has shown that the proposed method can achieve better performance than the other methods under small change detection sample sizes. This improvement depends to a certain extent on building extraction samples for pre-training. However, this cannot be regarded as a defect of the proposed algorithm because, in addition to the WHU Building Change Dataset used in the experiment, there are currently many other public building detection datasets, such as Inria and AIRS datasets, but few change detection datasets. In addition, the mapping and urban planning departments of central and local governments also have global architectural GIS maps [25], which can be used in projects. Currently, the main challenge is to overcome the lack of change detection samples, and the proposed method can address this challenge well.

According to the ablation experiment, the object-level refinement algorithm has a certain effect on the detection result under a reasonable threshold α , and this effect is more significant when the training sample size is small. The relationship curve between the threshold α and evaluation indicators shows that in most cases, threshold α can have a

positive effect on the network performance. When the threshold α is small, the proposed algorithm can obtain a high recall while ensuring good precision within a reasonable range. In most practical applications, it is required to reduce the detection error without introducing missing detections as much as possible.

The introduced object-level refinement algorithm also has certain defects, that is, when only one part of a building changes, the object-level refinement algorithm will conclude that the whole building has changed. Therefore, the object-level refinement algorithm can have a negative effect on the detection result when this situation occurs. In addition, the object-level refinement algorithm requires adjusting the threshold α manually according to different engineering requirements, which can be time-consuming in the case of large data volumes. In the future, we will consider how to improve the object-level refinement algorithm to make it learnable so as to solve the mentioned shortcomings.

6. Conclusions

This paper proposed a novel end-to-end building change detection network under the idea of multi-task learning. Through unique decode module and appropriately selecting the high-dimensional features for sharing, the network's dependence on change detection samples is reduced greatly. The results of experiments in Tables 1 and 2 show that our method trained only with 10% samples exceeds most of the methods we compared, with the F1 value of 0.8993. For taking full advantage of the building detection branch, the objectlevel refinement algorithm is proposed, which improved final change detection results. Through ablation experiments, the effectiveness of object-level refinement algorithm was proved in the case of small samples. Experiments show that after using proposed method, F1 value improves by 4.38% when training with only 10% of samples, and improves by 1.34% when training with 25% of samples. In addition, it is shown that in the case of fewer change detection training samples, the proposed method can achieve high-accuracy change detection results and can adjust the precision, recall, and F1 values of the results by adjusting the threshold value α of the object-level refinement algorithm, which is of great engineering significance because the existing public change detection datasets are limited and engineering requirements for test indicators are not the same for all applications. Our method can be regarded as an effective way when the change detection samples are insufficient. Furthermore, it provides a new idea for small sample change detection method.

Author Contributions: S.G., W.L. and K.S. designed the experiments; S.G. and W.L. conducted the experiments; Y.C. and J.W. prepared the data; S.G., W.L., X.W. and K.S. discussed the results. All authors contributed to the writing and revising of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (NO. 92038301, NO. 42192583, NO. 41471354).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data are available from the Grop of Photogrammetry and Computer Vision http://gpcv.whu.edu.cn/data/building_dataset.html (accessed on 5 January 2022).

Acknowledgments: We would like to thank the Grop of Photogrammetry and Computer Vision for providing WHU Building Dataset and WHU Building Change Dataset. Moreover, we would like to thank the authors of these methods we used and compared, including MA-FCN, FPN, Panoptic FPN, DASNet, Si-HRNet, DTCDSCN, IFN, STANet.

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

References

- Singh, A. Review article digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* 1989, 10, 989–1003. [CrossRef]
- 2. Chen, G.; Hay, G.; Carvalho, L.; Wulder, M. Object-based change detection. Int. J. Remote Sens. 2012, 33, 4434–4457. [CrossRef]

- 3. Zanetti, M.; Bruzzone, L. A theoretical framework for change detection based on a compound multiclass statistical model of the difference image. *IEEE Trans. Geosci. Remote Sens.* 2018, *56*, 1129–1143. [CrossRef]
- Bovolo, F.; Bruzzone, L. An adaptive multiscale random field technique for unsupervised change detection in vhr multitemporal images. *Geosci. Remote Sens. Symp.* 2009, *4*, 777–780.
- Torres-Vera, M.A.; Prol-Ledesma, R.M.; García-López, D. Three decades of land use variations in mexico city. *Int. Remote Sens.* 2009, 30, 117–138. [CrossRef]
- Jensen, J.R.; Toll, D.L. Detecting residential land-use development at the urban fringe. *Photogramm. Eng. Remote Sens.* 1982, 48, 629–643.
- Deng, J.; Wang, K.; Deng, Y.; Qi, G. Pca-based land-use change detection and analysis using multitemporal and multisensor satellite data. *Int. J. Remote Sens.* 2008, 29, 4823–4838. [CrossRef]
- Celik, T. Unsupervised change detection in satellite images using principal component analysis and k-means clustering. *IEEE Geosci. Remote. Lett.* 2009, 6, 772–776. [CrossRef]
- Ortizrivera, V.; Vélezreyes, M.; Roysam, B. Change detection in hyperspectral imagery using temporal principal components. In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII*; SPIE: Bellingham, WA, USA, 2006; Volume 6233, pp. 368–377.
- 10. Munyati, C. Use of principal component analysis (pca) of remote sensing images in wetland change detection on the kafue flats, zambia. *Geocarto Int.* 2004, 19, 11–22. [CrossRef]
- 11. Bovolo, F.; Bruzzone, L. A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain. *IEEE Trans. Geosci. Remote Sens.* 2006, 45, 218–236. [CrossRef]
- 12. Johnson, R.D.; Kasischke, E. Change vector analysis: A technique for the multispectral monitoring of land cover and condition. *Int. Remote Sens.* **1998**, *19*, 411–426. [CrossRef]
- 13. Malila, W.A. Change vector analysis: An approach for detecting forest changes with landsat. LARS Symp. 1980, 385, 326–335.
- Erener, A.; Düzgün, H.S. A methodology for land use change detection of high resolution pan images based on texture analysis. *Ital. J. Remote Sens.* 2009, 41, 47–59. [CrossRef]
- Tomowski, D.; Ehlers, M.; Klonus, S. Colour and texture based change detection for urban disaster analysis. In Proceedings of the 2011 Joint Urban Remote Sensing Event, Munich, Germany, 11–13 April 2011; pp. 329–332.
- Lei, T.; Zhang, Q.; Xue, D.; Chen, T.; Meng, H.; Nandi, A.K. End-to-end change detection using a symmetric fully convolutional network for landslide mapping. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 3027–3031.
- Zhang, H.; Tang, X.; Han, X.; Ma, J.; Zhang, X.; Jiao, L. High-Resolution Remote Sensing Images Change Detection with Siamese Holistically-Guided FCN. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 4340–4343. [CrossRef]
- 18. 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]
- 19. Huang, F.; Xu, C.; Zhu, Z. Change detection of hyperspectral remote sensing images based on deep belief network. *Int. J. Earth Sci. Eng.* **2016**, *9*, 2096–2105.
- 20. 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]
- 21. Argyridis, A.; Argialas, D.P. Building change detection through multi-scale geobia approach by integrating deep belief networks with fuzzy ontologies. *Int. J. Image Data Fusion* **2016**, *7*, 148–171. [CrossRef]
- Gao, F.; Dong, J.; Li, B.; Xu, Q. Automatic change detection in synthetic aperture radar images based on pcanet. *IEEE Geosci. Remote Lett.* 2016, 13, 1792–1796. [CrossRef]
- Wan, Y.; Liu, Y.; Peng, Q.; Jie, F.; Ming, D. Remote sensing image change detection algorithm based on bm3d and pcanet. In *Chinese Conference on Image and Graphics Technologies*; Springer: Singapore, 2019; Volume 1043, pp. 524–531.
- Li, M.; Li, M.; Zhang, P.; Wu, Y.; Song, W.; An, L. Sar image change detection using pcanet guided by saliency detection. *IEEE Geosci. Remote Sens. Lett.* 2018, 16, 402–406. [CrossRef]
- 25. Ji, S.; Shen, Y.; Lu, M.; Zhang, Y. Building instance change detection from large-scale aerial images using convolutional neural networks and simulated samples. *Remote Sens.* **2019**, *11*, 1343. [CrossRef]
- 26. Maiya, S.R.; Babu, S.C. Slum segmentation and change detection: A deep learning approach. arXiv 2018, arXiv:1811.07896.
- Cao, Z.; Wu, M.; Yan, R.; Zhang, F.; Wan, X. Detection of small changed regions in remote sensing imagery using convolutional neural network. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2020; Volume 502, pp. 12–17.
- 28. Liu, Y.; Pang, C.; Zhan, Z.; Zhang, X.; Yang, X. Building change detection for remote sensing images using a dual-task constrained deep siamese convolutional network model. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 811–815. [CrossRef]
- 29. Sun, Y.; Zhang, X.; Huang, J.; Wang, H.; Xin, Q. Fine-Grained Building Change Detection From Very High-Spatial-Resolution Remote Sensing Images Based on Deep Multitask Learning. *IEEE Geosci. Remote Sens. Lett.* **2020**, *58*, 8000605. [CrossRef]
- Zhang, M.; Shi, W. A feature difference convolutional neural network-based change detection method. *IEEE Trans. Geosci. Remote* 2020, 58, 7232–7246. [CrossRef]
- Chen, H.; Li, W.; Shi, Z. Adversarial instance augmentation for building change detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* 2021, 60, 5603216. [CrossRef]

- 32. Zheng, Z.; Ma, A.; Zhang, L.; Zhong, Y. Change is everywhere: Single-temporal supervised object change detection in remote sensing imagery. *CVF Int. Conf. Comput. Vis.* **2021**, *15*, 15193–15202.
- Yang, M.; Jiao, L.; Liu, F.; Hou, B.; Yang, S. Transferred Deep Learning-Based Change Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* 2019, 57, 6960–6973. [CrossRef]
- Liu, J.; Chen, K.; Xu, G.; Sun, X.; Han, H. Convolutional Neural Network-Based Transfer Learning for Optical Aerial Images Change Detection. *IEEE Geosci. Remote Sens. Lett.* 2020, 17, 127–131. [CrossRef]
- 35. Saha, S.; Bovolo, F.; Bruzzone, L. Unsupervised deep change vector analysis for multiple-change detection in vhr images. *IEEE Trans. Geosci. Remote Sens.* 2019, *57*, 3677–3693. [CrossRef]
- 36. Wu, C.; Do, C.H.B.; Zhang, L. Unsupervised change detection in multi-temporal vhr images based on deep kernel pca convolutional mapping network. *arXiv* **2019**, arXiv:1912.08628.
- Saha, S.; Mou, L.; Qiu, C.; Zhu, X.X.; Bruzzone, L. Unsupervised deep joint segmentation of multitemporal high-resolution images. *IEEE Trans. Geosci. Remote Sens.* 2020, 58, 8780–8792. [CrossRef]
- D'Addabbo, A.; Pasquariello, G.; Amodio, A. Urban change detection from vhr images via deep-features exploitation. In Proceedings of Sixth International Congress on Information and Communication Technology, London, UK, 25–26 February 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 487–500.
- Chen, Q.; Wang, L.; Waslander, S.L.; Liu, X. An end-to-end shape modeling framework for vectorized building outline generation from aerial images. *ISPRS J. Photogramm. Remote Sens.* 2020, 170, 114–126. [CrossRef]
- Chen, Z.; Li, D.; Fan, W.; Guan, H.; Wang, C.; Li, J. Self-attention in reconstruction bias u-net for semantic segmentation of building rooftops in optical remote sensing images. *Remote Sens.* 2021, 13, 2524. [CrossRef]
- 41. Ziaee, A.; Dehbozorgi, R.; Döller, M. A novel adaptive deep network for building footprint segmentation. *arXiv* 2021, arXiv:2103.00286.
- 42. Wei, S.; Ji, S.; Lu, M. Toward automatic building footprint delineation from aerial images using cnn and regularization. *IEEE Trans. Geosci. Remote Sens.* 2019, 58, 2178–2189. [CrossRef]
- Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
- Ghiasi, G.; Lin, T.Y.; Le, Q.V. Nas-fpn: Learning scalable feature pyramid architecture for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp.7036–7045.
- 45. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. *CVF Conf. Comput. Pattern Recognit.* 2020, 10, 781–790.
- Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
- 47. Bao, T.; Fu, C.; Fang, T.; Huo, H. Ppcnet: A combined patch-level and pixel-level end-to-end deep network for high-resolution remote sensing image change detection. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1797–1801. [CrossRef]
- Zhang, L.; Hu, X.; Zhang, M.; Shu, Z.; Zhou, H. Object-level change detection with a dual correlation attention-guided detector. ISPRS J. Photogramm. Remote Sens. 2021, 177, 147–160. [CrossRef]
- 49. Lian, X.; Yuan, W.; Guo, Z.; Cai, Z.; Song, X.; Shibasaki, R. End-to-end building change detection model in aerial imagery and digital surface model based on neural networks. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 2020, 43, 1239–1246. [CrossRef]
- Kirillov, A.; Girshick, R.; He, K.; Dollár, P. Panoptic feature pyramid networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 6399–6408.
- 51. Zhang, Y.; Yang, Q. An overview of multi-task learning. Natl. Rev. 2018, 5, 30–43. [CrossRef]
- 52. 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.* 2018, 57, 574–586. [CrossRef]
- 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]
- 54. 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]
- Chen, J.; Yuan, Z.; Peng, J.; Chen, L.; 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.* 2020, 14, 1194–1206. [CrossRef]
- 56. Lin, M.; Shi, Q.; Marinoni, A.; He, D.; Zhang, L. Super-resolution-based Change Detection Network with Stacked Attention Module for Images with Different Resolutions. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 4403718.