

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/333638570>

Building Instance Change Detection from Large-Scale Aerial Images using Convolutional Neural Networks and Simulated Samples

Article in Remote Sensing · June 2019

DOI: 10.3390/rs11111343

CITATIONS

0

READS

103

4 authors, including:



Shunping Ji

Wuhan University

23 PUBLICATIONS 141 CITATIONS

[SEE PROFILE](#)



Meng Lu

Utrecht University

21 PUBLICATIONS 55 CITATIONS

[SEE PROFILE](#)



Yongjun Zhang

Wuhan University

134 PUBLICATIONS 993 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:




deforestation monitoring using multispectral time series [View project](#)



Satellite and UAV image processing [View project](#)

Article

Building Instance Change Detection from Large-Scale Aerial Images using Convolutional Neural Networks and Simulated Samples

Shunping Ji ^{1,*} , Yanyun Shen ¹, Meng Lu ² and Yongjun Zhang ¹

¹ School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China; shenyanyun@whu.edu.cn (Y.S.); zhangyj@whu.edu.cn (Y.Z.)

² Department of Physical Geography, Faculty of Geoscience, Utrecht University, Princetonlaan 8, 3584 CB Utrecht, The Netherlands; m.lu@uu.nl

* Correspondence: jishunping@whu.edu.cn

Received: 3 May 2019; Accepted: 31 May 2019; Published: 4 June 2019



Abstract: We present a novel convolutional neural network (CNN)-based change detection framework for locating changed building instances as well as changed building pixels from very high resolution (VHR) aerial images. The distinctive advantage of the framework is the self-training ability, which is highly important in deep-learning-based change detection in practice, as high-quality samples of changes are always lacking for training a successful deep learning model. The framework consists two parts: a building extraction network to produce a binary building map and a building change detection network to produce a building change map. The building extraction network is implemented with two widely used structures: a Mask R-CNN for object-based instance segmentation, and a multi-scale full convolutional network for pixel-based semantic segmentation. The building change detection network takes bi-temporal building maps produced from the building extraction network as input and outputs a building change map at the object and pixel levels. By simulating arbitrary building changes and various building parallaxes in the binary building map, the building change detection network is well trained without real-life samples. This greatly lowers the requirements of labeled changed buildings, and guarantees the algorithm's robustness to registration errors caused by parallaxes. To evaluate the proposed method, we chose a wide range of urban areas from an open-source dataset as training and testing areas, and both pixel-based and object-based model evaluation measures were used. Experiments demonstrated our approach was vastly superior: without using any real change samples, it reached 63% average precision (AP) at the object (building instance) level. In contrast, with adequate training samples, other methods—including the most recent CNN-based and generative adversarial network (GAN)-based ones—have only reached 25% AP in their best cases.

Keywords: building change detection; deep learning; CNN; aerial images

1. Introduction

Change detection is the process of identifying differences in the state of an object, a scene or a phenomenon by comparing them at different times [1]. Remote sensing data has become a major data source in change detection due to its high time-frequency, wide spectrum of spatial and spectral resolutions and the broad bird's eye view [2–4]. It has been widely applied to detecting land cover and land-use changes [5–8], urban development [9–11], natural disaster evaluation [12] and forestry [13–15]. A typical challenge of detecting changes from remote sensing data is that the spectral behavior of remote sensing imagery (e.g., reflectance values, local textures) may lead to false alarms due to anthropogenic behavior, atmospheric conditions, illumination, viewing angles and soil moisture [1,8,16,17], which has accordingly led to the development of a variety of change detection methodologies.

The change detection process consists of three major steps: pre-processing, change detection technique selection and accuracy assessment [18,19]. The pre-processing stage primarily deals with issues related to atmospheric and radiometric correction, topographic correction and image registration. Since image displacement will cause many false changes, accurate geometric registration between multi-temporal images is in many cases the key to a successful change detection. Radiometric correction rectifies errors from changes in atmospheric conditions, illumination angles, sensor characteristics and viewing angles to ensure radiance consistency [20,21].

A variety of change detection methods on remote sensing imagery have been developed in the past few decades. These methods can be grouped into pixel-based and object-based change detection [18]. The sizes of the study areas and the spatial resolution of remote sensing data are important factors for choosing a certain change detection technique. Generally, low resolution images (e.g., MODIS) are mainly used in large-scale change detection. At the regional scale, the use of higher spatial resolution remote sensing data such as Landsat Thematic Mapper (TM) may be adequate. Pixel-based techniques have been widely applied to these remote sensing data. Using high resolution (HR) or very high resolution (VHR) images (e.g., QuickBird, IKONOS, aerial images) could provide much more details in land cover changes. However, with the pixel-based change detection approach, a large number of small false changes may occur [2] due to unpredictable high-frequency components in HR images, geometric registration errors and imperfect radiation correction. The object-based approaches have been shown to achieve improved results on HR or VHR images against these problems at the local scale [22,23], and are becoming increasingly popular.

We firstly provide a brief review of the development of classic pixel-based and object-based change detection methods. Image differencing [24], image ratioing [25] and regression analysis [26] are intuitive and straightforward pixel-based technologies that assume one image is a (generalized) linear function of other images, and locate the mutant pixels as changes. Besides the pixel intensity, more sophisticated feature maps are generated from pixel values, such as vegetation indices, which are widely used in imagery time series analysis [27]. In addition to the linear and rational polynomial transformations, principal component analysis (PCA) [28], Tasseled Cap transformation [29], change vector analysis (CVA) [30] and texture-based transforms [31,32] are also widely applied. Classification-based change detection, often known as the post-classification comparison method, classifies multi-temporal images separately and then compares the classified pixels [33,34]. The classification could be realized with supervised [35,36] or unsupervised [37,38] learning. Many machine learning algorithms have been introduced as a part (a feature extractor or a classifier) of the pixel-based change detection, such as artificial neural networks [39,40], support vector machines [41,42] and decision trees [33].

Object-based change detection treats objects instead of pixels as the unit to detect change by comparing spectral information [43], geometric properties [44] or high semantic features [44] extracted from the objects, which reduces the problems caused by boundary effects [45] and misalignment. In most cases, objects should be extracted and located firstly from bi-temporal images separately, then the classified objects can be compared to obtain change information [46]. There are also methods that directly find changed objects from stacked multi-temporal images [47].

As buildings are the main place for human activities and the representative of manmade structure, building change detection has been an important topic in remote sensing data change detection. A variety of methods for building change detection have been proposed, especially for HR and VHR optical sensors that could provide detailed land cover information. Huang et al. [22] proposed a morphological building index (MBI), to build a relationship between the spectral-spatial characteristics of buildings and morphological operators [22]. Du et al. [48] detected building changes in urban areas using aerial images and LiDAR data. Liu et al. [49] used a line-constrained shape feature to capture the shape characteristics of a building. Xiao et al. [23] developed a co-segmentation method for building change detection from multitemporal HR images.

Besides these classic change detection methods, the deep learning methods have been applied to remote sensing data (e.g., multispectral [50–53], hyperspectral [54], synthetic aperture radar (SAR) [55])

to classify land cover types such as forests [56], rivers and farmland [54] and landslides [57], and to detect land cover changes. These have obtained better performances than classic methods. Among fundamental network models for deep learning, such as convolutional neural networks (CNNs) [58], deep belief networks (DBNs) [59], sparse autoencoders (AEs) [60], recurrent neural networks (RNNs) and generative adversarial networks (GANs) [61], the CNN, which consists of a series of convolutional layers, is the most widely used structure in image classification and change detection.

CNNs have been applied to building change detection. Daudt et al. [62] utilized three fully convolutional neural networks (FCNs) to detect changes in registered images. In this study, satellite images of only 10–60 m resolutions were used, which inevitably results in low change detection accuracy. Nemoto et al. [63] used a CNN to extract buildings from a new image, and then used the building classification map and the two images as inputs for another CNN to detect building changes. This study was tested on large aerial images, but the change detection results were unsatisfactory. In [64], a Siamese CNN network was applied to detect changes of buildings and trees between a laser point cloud and an aerial image in a very small area. Amin et al. [65] used a super-pixel segmentation on registered bi-temporal images and then applied a CNN to detect pixel-based building change. This work used only two small images for testing, and cannot recognize changes in building instances.

Generally, these recent CNN-based building change detection studies have contributed to automatic building change detection, but a variety of challenges still exists. For example, the previous studies have either utilized very small images, or did not perform well on large datasets; they need enough samples of changed buildings, which are commonly sparse, to train the CNN. Furthermore, most of those studies only detected changes at the pixel level, whereas the statistics on building instances are often more important in practice.

In this paper, we present a novel framework for building change detection from VHR aerial images, which incorporates a building extraction network and a building change detection network, the performances of which are thoroughly evaluated on a large open building dataset and compared to some of the most recent methods. The framework is fully automatic, end-to-end, and could easily compute pixel- and object (building instance)-based change maps. The main idea and contribution of this paper can be summarized in three aspects:

(1) A new and end-to-end framework is proposed to not only detect changed buildings in pixels, but also in building instances. The latter case realizes a true “object change detection” instead of treating a group of arbitrary pixels as an object, as most object-based studies have done. The front-end building detection network is implemented with a pixel-based semantic segmentation network and an object-based instance segmentation network.

(2) The back-end change detection network we proposed can not only detect accurate building changes, but also greatly mitigates one of the most prevalent problems of the deep-learning-based change detection: the requirement of large, high-quality training samples, which is rarely met since changes (positive samples) are usually less common. The back-end network can be well pretrained with automatically generated positive samples in building classification maps to achieve high accuracy without one real change sample. This simulation manner also improves the robustness of our method in situations where accurate image registration cannot be achieved due to sensor angles and building parallaxes, which affect the performance of most of change detection methods.

(3) Experiments demonstrated that our method is promising: without any real change sample, it greatly outperformed other methods trained with adequate samples by at least 38% AP (average precision) at the object level. Our algorithm was evaluated on a larger urban area compared to most of the relevant studies, which guaranteed more rigorous statistical significance and better practical reference values. Specifically, the experiments were executed on an open-source dataset, namely, the WHU building dataset [66]. The test area covers about 120,000 buildings (including 2007 changed buildings) with diverse architectural styles and usages.

2. Methodology

The framework of our CNN-based building change detection is shown in Figure 1. The input bi-temporal images are classified into a binary map of buildings and background using a building extraction network. The noises in the maps caused by imperfect building extraction were filtered out with a designed filter, then the maps of different times were concatenated and input into the change detection network to detect building changes and produce a building change map.

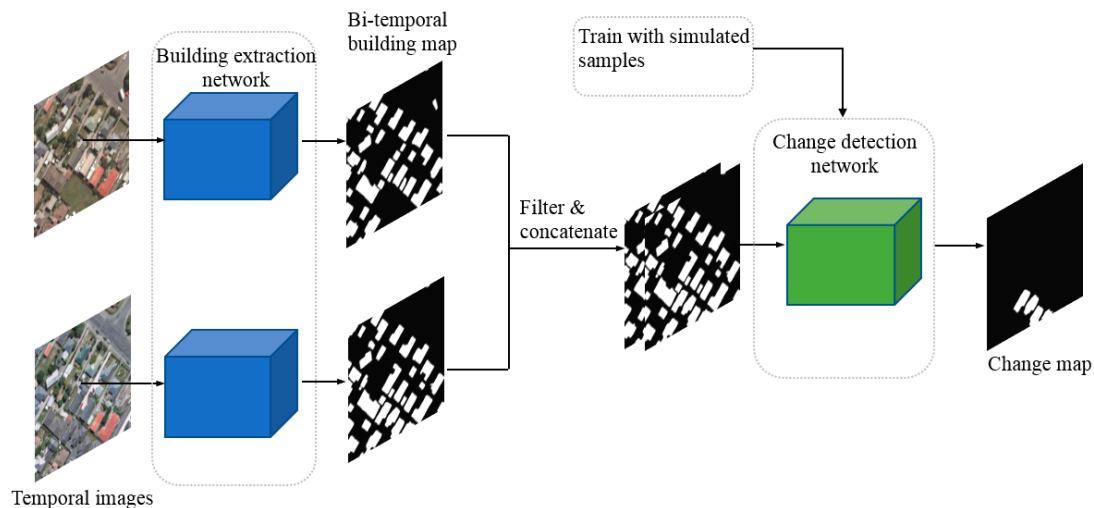


Figure 1. Overview of our classification-based change detection framework.

2.1. Building Extraction Network

In order to evaluate the impact of the building extraction network at the object and pixel levels on the subsequent change detection, this paper uses two network structures to extract buildings, namely, the Mask R-CNN [67] and the MS-FCN (multi-scale fully convolutional network), which are based on the U-Net [68] structure.

Although the Mask R-CNN was proposed in 2017, it is the most powerful instance segmentation method up to now. The structure of the Mask R-CNN is shown in Figure 2. It consists of a backbone CNN framework, a regional proposal network (RPN) [69], a region of interest (RoI) alignment process (RoIAlign) and three output branches: classification, box regression and mask prediction. The Mask R-CNN adopts a two-stage strategy. In the first stage, the feature map is searched through the RPN for regions that may contain foreground. Rectangles with different sizes and proportions are used to cover such regions, and the suggested rectangles are used as the bounding box for the candidates. The second step uses the bounding box to obtain the RoI from the feature maps of CNN layers, then performs classification and bounding box regression. We kept all the parameters and super parameters the same as the original version, except for changing from the multi-class object detection to a single object detection (building).

We designed a U-Net structure combining with multi-scale aggregation for pixel-based segmentation named the MS-FCN. We find this light structure efficient and especially effective in remote sensing data classification. The MS-FCN structure is shown in Figure 3. The encoder consists of a series of 3×3 convolution and 2×2 max-pooling to extract a higher level of semantic features; the decoder of the original U-Net [68] gradually enlarges the feature maps by a series of 3×3 convolutions and 2×2 up-samplings up to a feature map with the same size of the original input image. However, due to the different sizes of buildings, only applying this feature map for building extraction will cause an incomplete inspection of some large-sized buildings, as well as some small-sized non-building objects, such as cars and containers, which could be mistakenly classified as buildings. In order to improve the robustness of multi-scale building extraction, we add a convolution layer with one channel in each

scale (red) and up-sampled them to the original scale and concatenate them to form a final four-channel feature map.

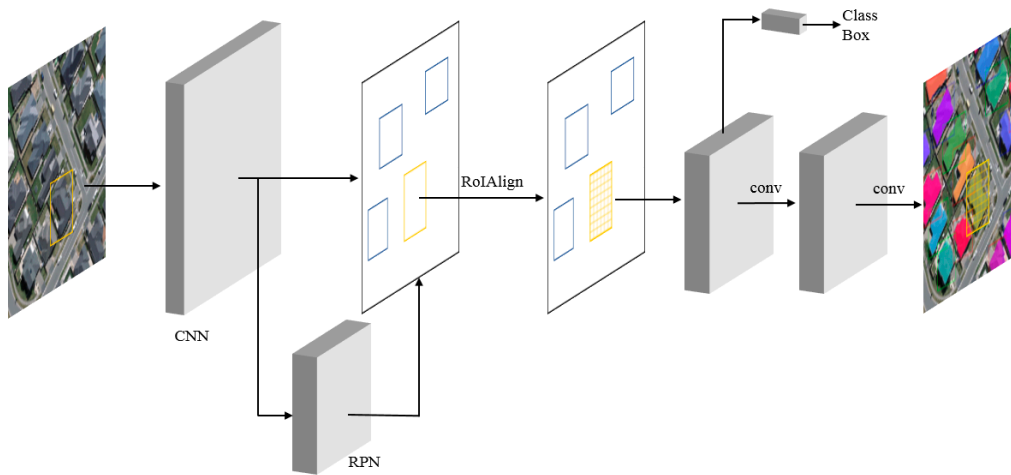


Figure 2. The Mask R-CNN framework for instance segmentation.

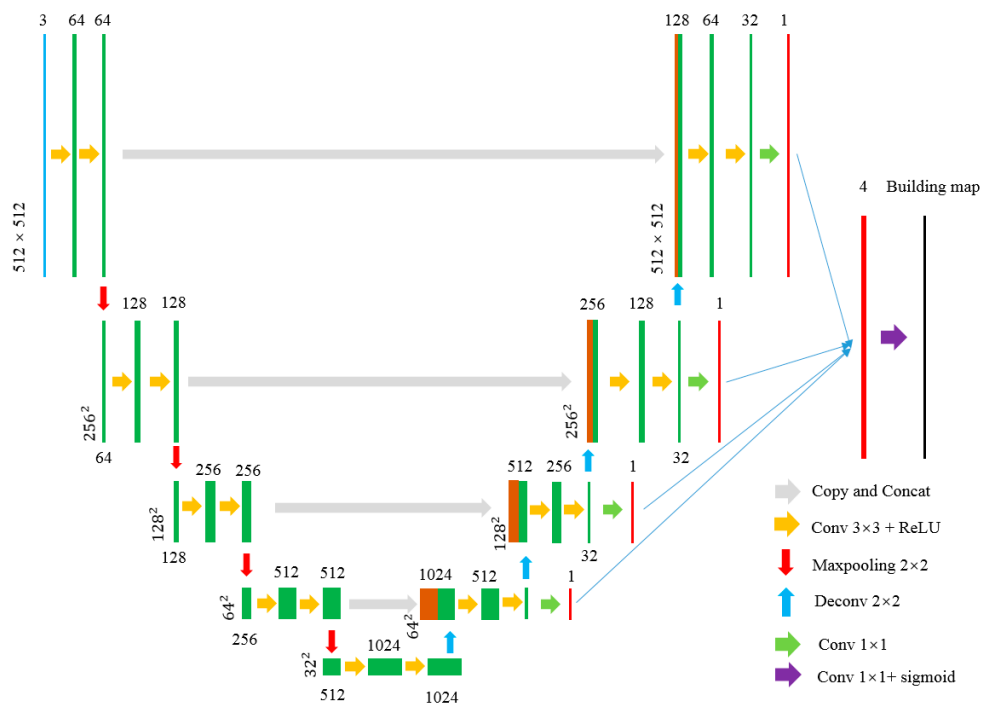


Figure 3. The MS-FCN framework for semantic segmentation.

The binary building map produced by the Mask R-CNN and the MS-FCN contains some errors. We filtered out pixel segments that are smaller than a given threshold to produce a more accurate classification map.

2.2. Self-Trained Building Change Detection Network

Different from directly using image pairs as the input of the change detection network [70], we use the binary maps produced by the building extraction network as the input. This modification has tremendous advantages. First, we simulate arbitrary building changes in the binary maps, which is almost impossible to simulate directly in the original images. With the simulated samples, the rigid demand for a large number of manual samples of a supervised deep learning method is greatly reduced.

Note that images with a high proportion of changes are rare in practice, as typically only a small fraction of buildings are changed even in a large area.

Second, we simulate registration errors of buildings in the binary maps by randomly shifting a building's mask within a given threshold (e.g., 10 pixels) to train the network that is resistant to this change (i.e., treating it as an unchanged building). Note that the parallax of buildings from different view angles of VHR images captured from pin-hole cameras always leads to a geometric registration error. This simple self-learning strategy is beneficial in comparison with empirical, data-specific and unstable post-processing methods, which typically involve many parameters trying to filter out these false changes. Figure 4 shows some examples of simulated change samples and buildings' parallaxes.

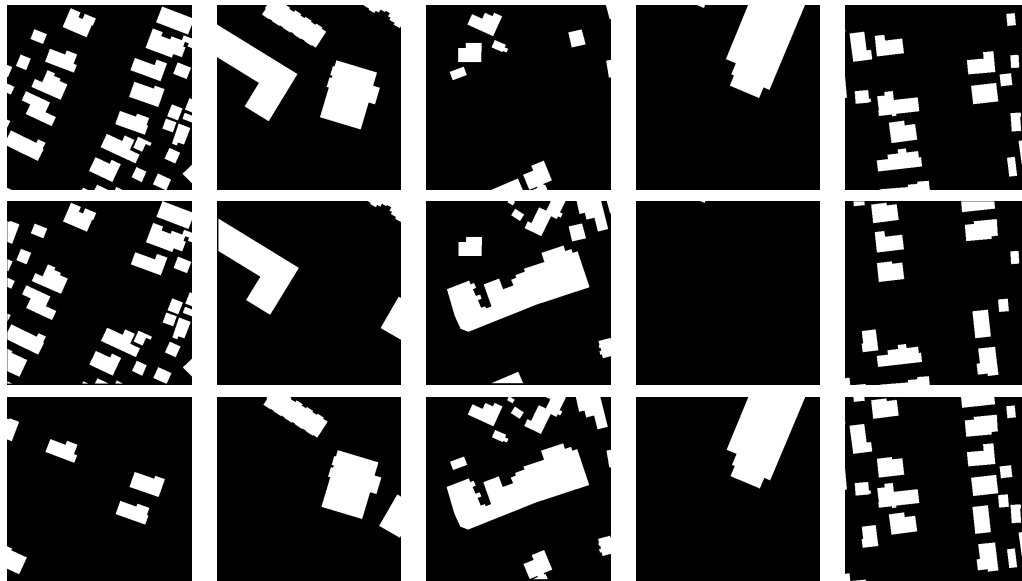


Figure 4. Examples of the simulated change detection dataset. The first row is the original building mask maps, the second row shows simulated buildings added or original buildings eliminated upon the mask maps, where the original buildings are slightly shifted according to a random parallax. The last row is simulated change labels.

As we only need to learn changes from binary maps, a simple CNN is suitable. Our change detection network structure is a simplified U-Net with less feature map channels, as shown in Figure 5. This structure has been empirically demonstrated to be better than the original version of U-Net, our MS-FCN and more recent structures such as the DeepLab v3+ [71].

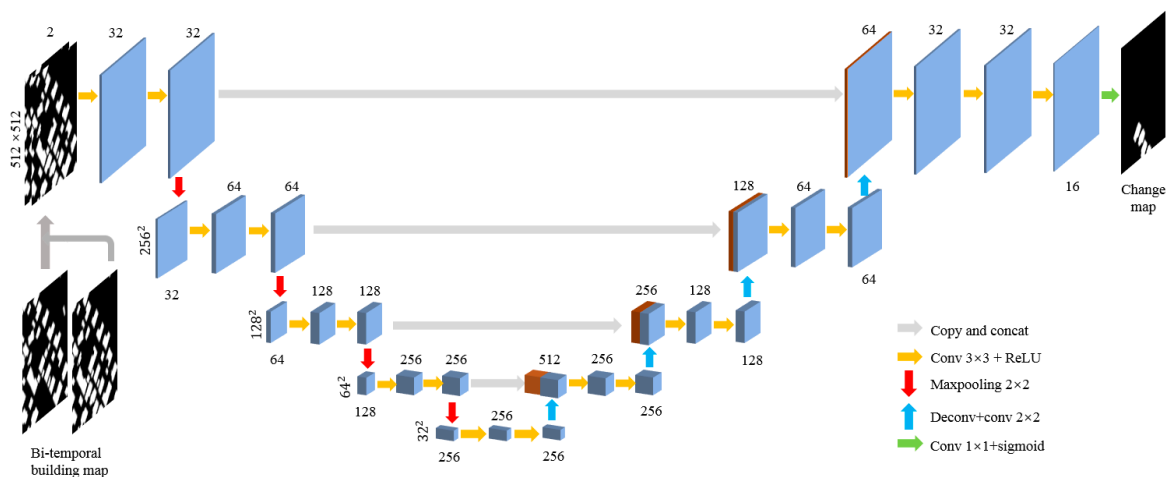


Figure 5. The building change detection network.

3. Experiments and Analysis

3.1. Data Set and Evaluation Measures

The dataset used in this paper comes from the WHU building change detection dataset [66]. The study area is in Christchurch, New Zealand, and covers about 120,000 buildings with various architectural styles and usages. According to different usages, we divided the dataset into five sub-datasets enclosed in colored boxes, as shown in Figure 6. In the training and prediction of a CNN model, all data was divided into small blocks of 512×512 pixels with a ground resolution of 0.2 m to adapt to an NVIDIA GTX 1080Ti 11G GPU, which was used in all the experiments. The details are listed in Table 1.

The experimental strategy is: in the first step, we used the training dataset of 2016 (SC-2016, yellow) to train the building extraction network, and predicted on the target area (TA-2016, red) to produce the building classification map of 2016. In the second step, the building extraction network pretrained on SC-2016 is trained on the training dataset of 2011 (SC-2011, green) and then applied to predict buildings on the target area (TA-2011, red) to produce the building map of 2011.

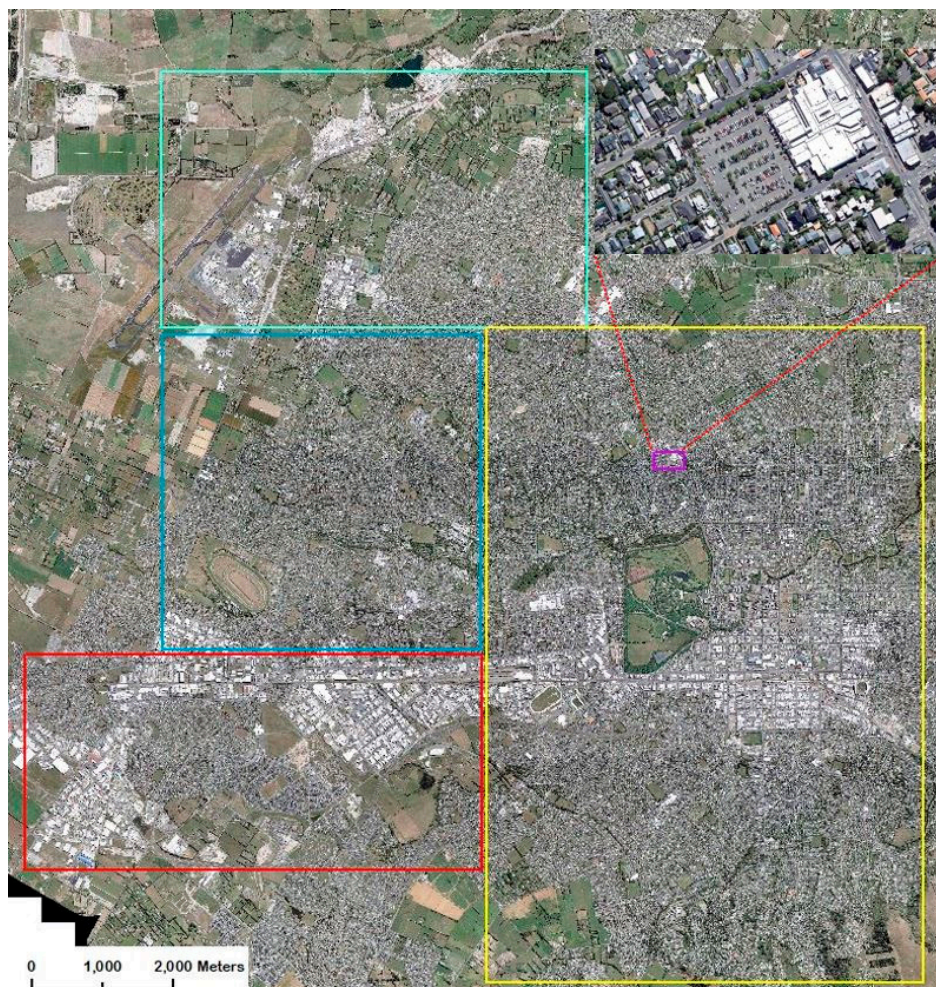


Figure 6. The five sub-datasets in the WHU building dataset with 0.2 m GSD (ground sampling distance). The red box is the main study area with images of 2011 (TA-2011) and 2016 (TA-2016) for evaluating change detection. The yellow box contains images of 2016 (SC-2016) to train the building extraction model for 2016. The green box contains images of 2011 to train the building extraction model for 2011. In the blue box, we simulate changed building samples automatically upon existing building masks of 2016 (SI-2016) to train the building change detection model.

Table 1. Details of the sub-datasets in the WHU building change detection dataset.

Datasets	GSD (m)	Area (km ²)	Tiles	Pixels	Building Number	Box Color (Figure 6)
SC-2016	0.2	57.744	5400	512 × 512	67,190	Yellow
SC-2011	0.2	22.035	2065	512 × 512	11,495	Green
TA-2016	0.2	19.964	1827	512 × 512	11,595	Red
TA-2011	0.2	19.964	1827	512 × 512	9588	Red
SI-2016	0.2	20.294	1892	512 × 512	21,876	Blue

In the third step, the simulated building change detection data set (SI-2016, blue) was created and used to train the change detection network. First, building masks were randomly shifted with 0–5 pixels in an arbitrary direction to simulate misplacement of bi-temporal images. Then, the buildings were randomly reduced and added to a change map to simulate building changes. Specifically, for each tile, we randomly dropped 0–3 buildings, or added to the background 0–3 buildings that were randomly selected from building masks of the simulation area as change labels. Finally, the change detection network was trained on the simulated change dataset and applied to predict those change buildings in the target area (red box). The model can also be fine-tuned on available real building change maps to produce a better change detection map of the target region.

Figure 7 lists examples of 512 × 512 tiles of different sub-datasets. The diversity of building styles and usages makes the study area with more than 120,000 buildings and 2007 changed buildings in the red box an ideal place to study building extraction and change detection.

Figure 8 shows examples of change labels. New buildings were built on the bare ground of images of 2011. Aside from changes to the buildings, a lot of land cover changes such as roads, parking lots and gardens are visible, which could result in many false alarms for any change detection methods except those based on post-classification comparison methods.

As we investigated both object (building instance) accuracy and pixel accuracy of the change detection algorithm, we applied two types of evaluation measures. The first one uses the intersection of union (IoU) as the main index for pixel-based evaluation, which is defined as

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (1)$$

In building extraction, true positive (TP) indicates the number of pixels correctly classified as buildings, false positive (FP) indicates the number of pixels misclassified as buildings and false negative (FN) indicates the number of pixels misclassified as background. In building change detection, TP indicates the number of pixels correctly classified as changed buildings, FP indicates the number of pixels misclassified as changed buildings and FN indicates the number of pixels misclassified as background.

The second evaluation measure uses average precision (AP) as the main index for object-based evaluation, which is defined as

$$\text{AP} = \int_0^1 p(r) dr \quad (2)$$

where p denotes precision and r denotes recall. AP is the area under the precision-recall curve with precision as the vertical axis and recall as the horizontal axis.

In building extraction, true positive (TP) indicates the number of buildings that are correctly detected (IoU > 50%), false positive (FP) indicates the number of buildings that are falsely detected and false negative (FN) indicates the number of buildings that are falsely detected as background. In building change detection, TP indicates the number of change buildings that are correctly detected, FP indicates the number of change buildings that are falsely detected and FN indicates the number of change buildings that are falsely detected as background.

All experiments were executed on a single NVIDIA GeForce 1080 TI GPU with 11 GB RAM.

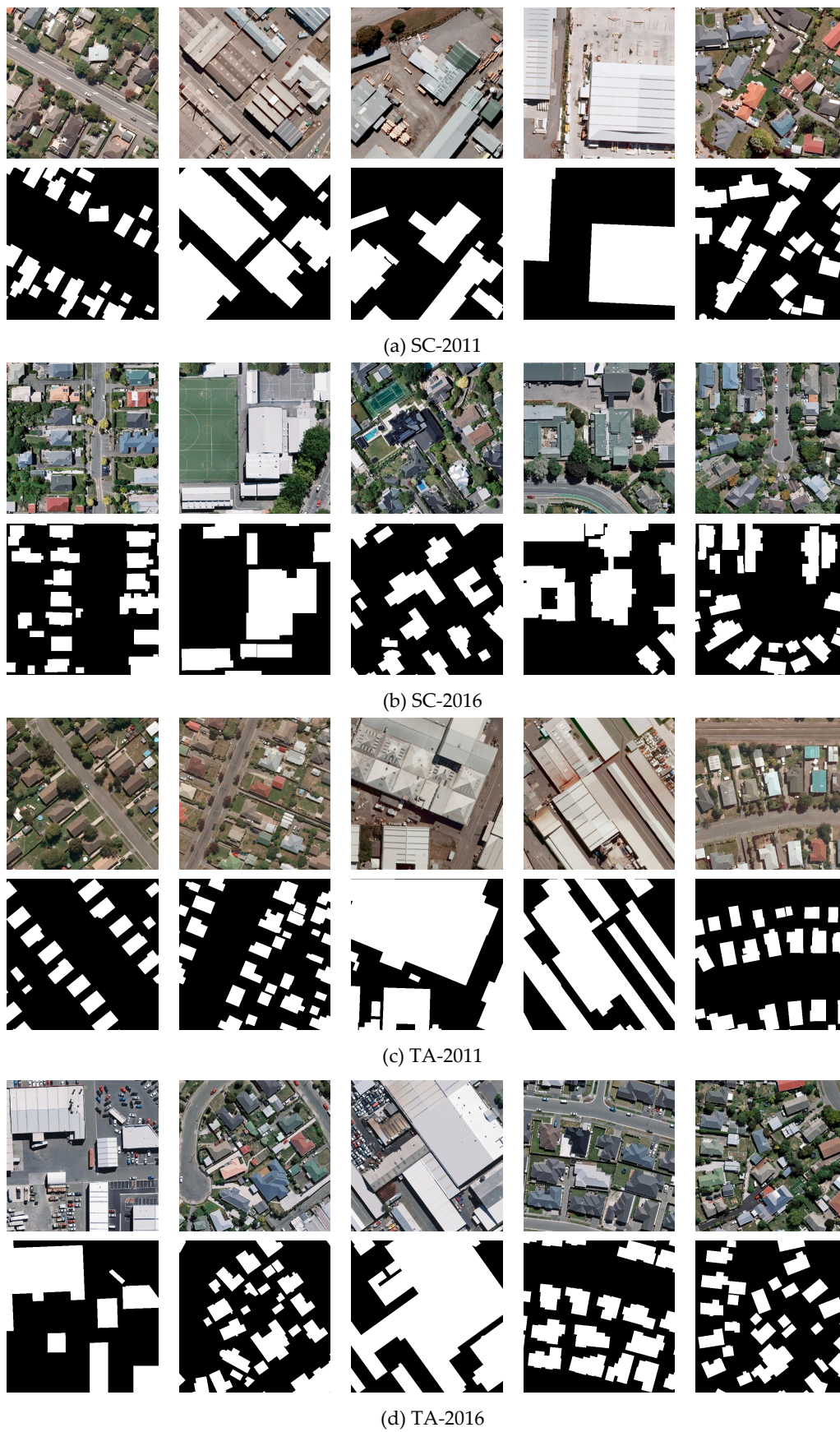


Figure 7. Examples of the SC-2011, SC-2016, TA-2011 and TA-2016 sub-datasets.



Figure 8. Examples of changed buildings. The top row is images of 2011, middle row is images of 2016 and bottom row is the change labels.

3.2. Building Extraction Results

We used the Mask R-CNN and the MS-FCN for building extraction. First, 5200 tiles in SC-2016 were used as the training set, 200 tiles were used as the validation set and TA-2016 was used as the test set. For the Mask R-CNN, the model pretrained on the open-source COCO dataset [72] and was converged after 40 epochs. The process took about 60 h. For the MS-FCN, the model had initial random weights and was converged after 30 epochs. The process took about 6 h.

Second, 1900 and 165 tiles in SC-2011 were used as the training set and verification set, respectively, while TA-2011 was used as the test set. For the Mask R-CNN, the model pretrained on the SC-2016 was converged after 5 epochs. This process took about 5.5 h. For the MS-FCN, the model pretrained on the SC-2016 was converged after 20 epochs. This process took about 1.5 h.

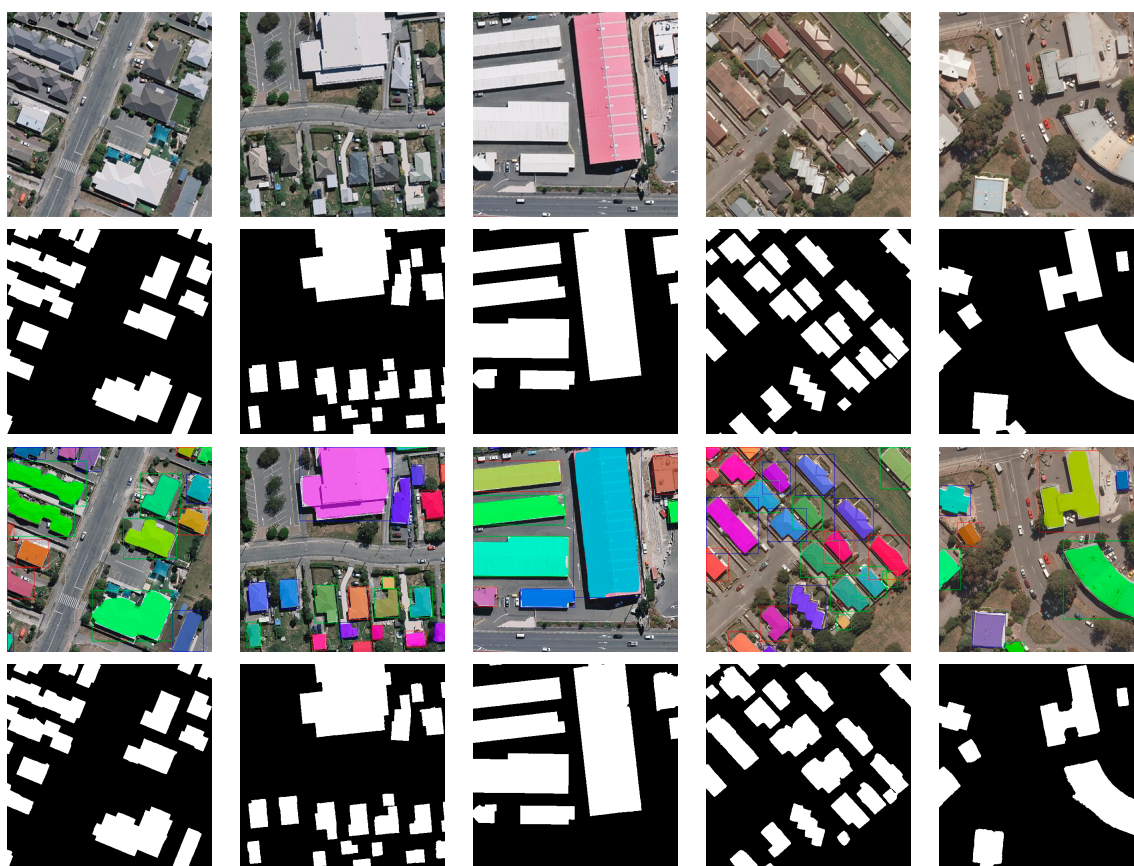
Due to the low accuracy of the extraction network at the edge of tiles, predictions from the Mask R-CNN and MS-FCN were made on the overlapped tiles. That is, when cutting the original large image of the TA-2016 and TA-2011, all the cropped tiles had 50% overlapped regions. After prediction, the edges of each tile were removed and then stitched to a seamless large image for evaluation. This strategy effectively avoids the edge effect that especially affects object instance detection methods such as the Mask R-CNN.

Building extraction accuracy of the two networks at the object and pixel levels are shown in Table 2. For TA-2011 and TA-2016, the AP of Mask R-CNN was 0.06 and 0.001 higher than the MS-FCN, respectively, and the IoU was 0.002 and 0.023 lower than the MS-FCN, respectively, indicating the Mask R-CNN could provide better object-level building extraction, but for pixel-level building classification, MS-FCN performs slightly better.

The examples of prediction results of the Mask R-CNN and the MS-FCN on TA-2016 and TA-2011 are shown in Figure 9. In comparison of the third and fourth lines, the MS-FCN is slightly better than the Mask R-CNN in extracting building edges.

Table 2. Building extraction accuracy of the Mask R-CNN and MS-FCN at the object and pixel levels in the test area.

Dataset	Method	Objected-Based						Pixel-Based		
		AP	Precision	Recall	TP+FP	TP	TP+FN	IoU	Precision	Recall
TA-2011	Mask R-CNN	0.833	0.892	0.930	9993	8916	9588	0.867	0.943	0.915
	MS-FCN	0.773	0.922	0.837	8702	8022	9588	0.869	0.934	0.925
TA-2016	Mask R-CNN	0.858	0.922	0.929	11,684	10,768	11,595	0.897	0.956	0.936
	MS-FCN	0.857	0.939	0.911	11,243	10,560	11,595	0.920	0.960	0.957

**Figure 9.** Examples of building extraction results of the Mask R-CNN and MS-FCN. From top to bottom: image, label, results of the Mask R-CNN, results of the MS-FCN. From left to right: columns 1–3 are from the TA-2016 data and columns 4–5 are from TA-2011.

3.3. Building Change Detection Results

Firstly, the binary building map obtained through the building extraction network was preprocessed by a simple filter. Buildings smaller than 500 pixels (corresponding to the ground $4.8 \times 4.8 \text{ m}^2$) were considered as false detections and are removed. Then, the bi-temporal classification map was divided into $1827 \ 512 \times 512$ tiles, with corresponding change labels as the ground truth.

We carried out three groups of tests. The first one did not use any change labels in the target study area (TA-2011 and TA-2016). We train our change detection network only on the automatically simulated data (SI-2016), from which 1800 tiles are used for training and 92 tiles for validation. Then, the model is applied to predict building changes in the test area (outside of the red box in Figure 10).

The second and third tests use half (green box) and full training samples (red box), respectively, to train the model, and therefore could be compared to other recent deep learning methods that require

training samples in the original images. Two recent methods are compared to our method. One is the FC-EF [62], which is an end-to-end change detection method based on the CNN and predicts changes from bi-temporal images directly. The other is a generative adversarial network (GAN)-based method [70] with the same end-to-end manner.

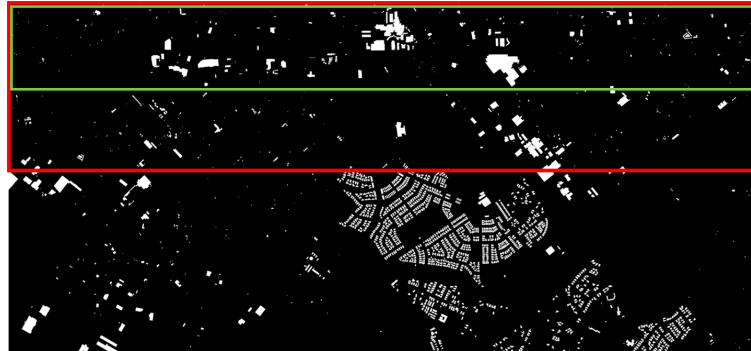


Figure 10. The change map of the study area (TA-2011 and TA-2016) with 2007 changed buildings. The red box contains training samples, and the rest are test samples including 1715 changed buildings. The green box expresses only half of the samples used for training.

In Table 3, in the case of training using only the simulation dataset, the AP (counted on changed building instances) reaches to 0.630 and 0.609 using the Mask R-CNN and MS-FCN building extraction, respectively; the IoU (counted on changed pixels) reaches 0.798. The FC-EF- and GAN-based change detection methods could not execute without real labels in the remote sensing images.

Table 3. Building change detection accuracy at the object level and pixel level under different training data: merely simulated data, half-change samples (within the green box in Figure 10) and full-change samples.

Dataset	Extraction Method	Objected-Based						Pixel-Based		
		AP	Precision	Recall	TP+FP	TP	TP + FN	IoU	Precision	Recall
simulated	Mask R-CNN	0.630	0.644	0.943	2511	1618	1715	0.798	0.856	0.922
	MS-FCN	0.609	0.659	0.896	2332	1537	1715	0.798	0.839	0.943
Half	Mask R-CNN	0.806	0.928	0.857	1584	1470	1715	0.773	0.952	0.804
	MS-FCN	0.793	0.881	0.880	1714	1510	1715	0.843	0.912	0.918
	FC-EF [62]	0.027	0.200	0.114	980	196	1715	0.261	0.516	0.346
	GAN [70]	0.023	0.135	0.127	1616	218	1715	0.232	0.538	0.290
Full	Mask R-CNN	0.814	0.910	0.883	1663	1514	1715	0.837	0.931	0.892
	MS-FCN	0.796	0.891	0.872	1679	1496	1715	0.830	0.938	0.878
	FC-EF [62]	0.254	0.519	0.462	1525	792	1715	0.502	0.767	0.593
	GAN [70]	/	/	/	/	/	/	/	/	/

When half of the samples were used for further training, the AP of our model was improved to 0.806 and 0.793, respectively, and the IoU changed to 0.773 and 0.843, respectively. In contrast, the FC-EF- and GAN-based methods obtained extremely poor results: only 2% AP and less than 26% IoU.

When all the training samples were used, the AP of our method was slightly improved to 0.814 and 0.796, respectively, and the IoU is improved to 0.83. The AP of FC-EF was improved from 0.02 to 0.25, indicating it requires a considerable amount of training samples to train an adequate model. However, even with enough change samples (about 300 changed buildings), it performed much worse than our method. Note that this area had undergone an earthquake in 2011, and plenty of the buildings were changed. Normally, it is even less possible to supply enough change samples to train a network

like the FC-EF. The GAN method could not converge with these samples. This indicates that the GAN-based method is unstable.

It should be noted that when our change detection network was trained directly (with random weights), its performance approached those pretrained on the simulated data. For example, with full training samples, the AP of the direct training strategy was 0.803 and 0.732, respectively.

It is concluded from Table 3 that, first, without any real training samples, our algorithm is dominant to other methods trained with adequate samples (incrementally by at least 40% AP and 30% IoU). Second, at the pixel level, without real change samples, our algorithm (0.798 IoU) approached the top performance (0.830 IoU); at the object level, the algorithm reached to the top performance with only a small amount of training samples (i.e., the performance did not improve with more samples). These two features are highly favorable in a practice where change samples are scarce or even unavailable.

Figure 11 lists four examples of building change detection results. Our change detection network with either the Mask R-CNN or the MS-FCN building extraction could detect changes with high accuracy. The results of our MS-FCN strategy were slightly better than the Mask R-CNN, as the later over-smoothed at building boundaries. Most of the changed buildings were missed by the FC-EF and GAN-based methods, and those detected changes were very noisy.

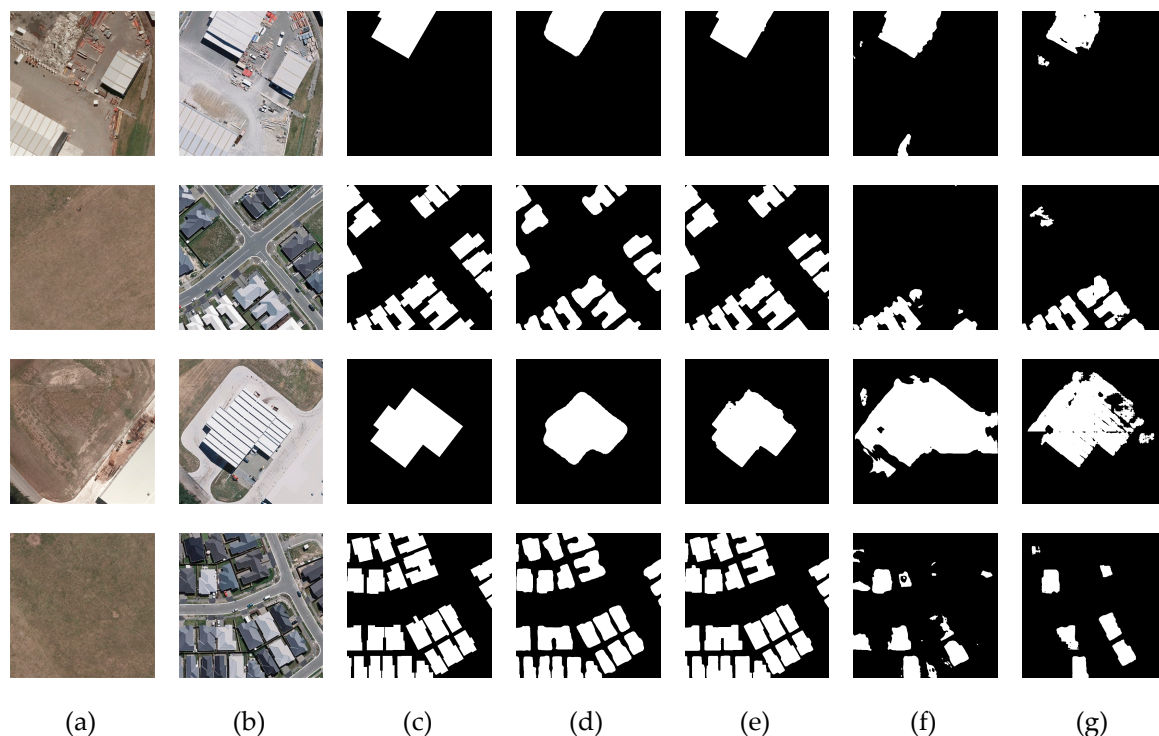


Figure 11. Comparison of change detection methods with half of the training samples. (a) Image 2011. (b) Image 2016. (c) Label. (d) Our method with the Mask R-CNN building extraction. (e) Our method with the MS-FCN building extraction. (f) FC-EF. (g) GAN-based change detection.

Figure 12 shows the results of different methods on the whole test area and clearly demonstrates our method to be much better than other methods even without any change samples. Right behind our method is the FC-EF with full training samples (Table 3); however, it only reaches 25% AP.

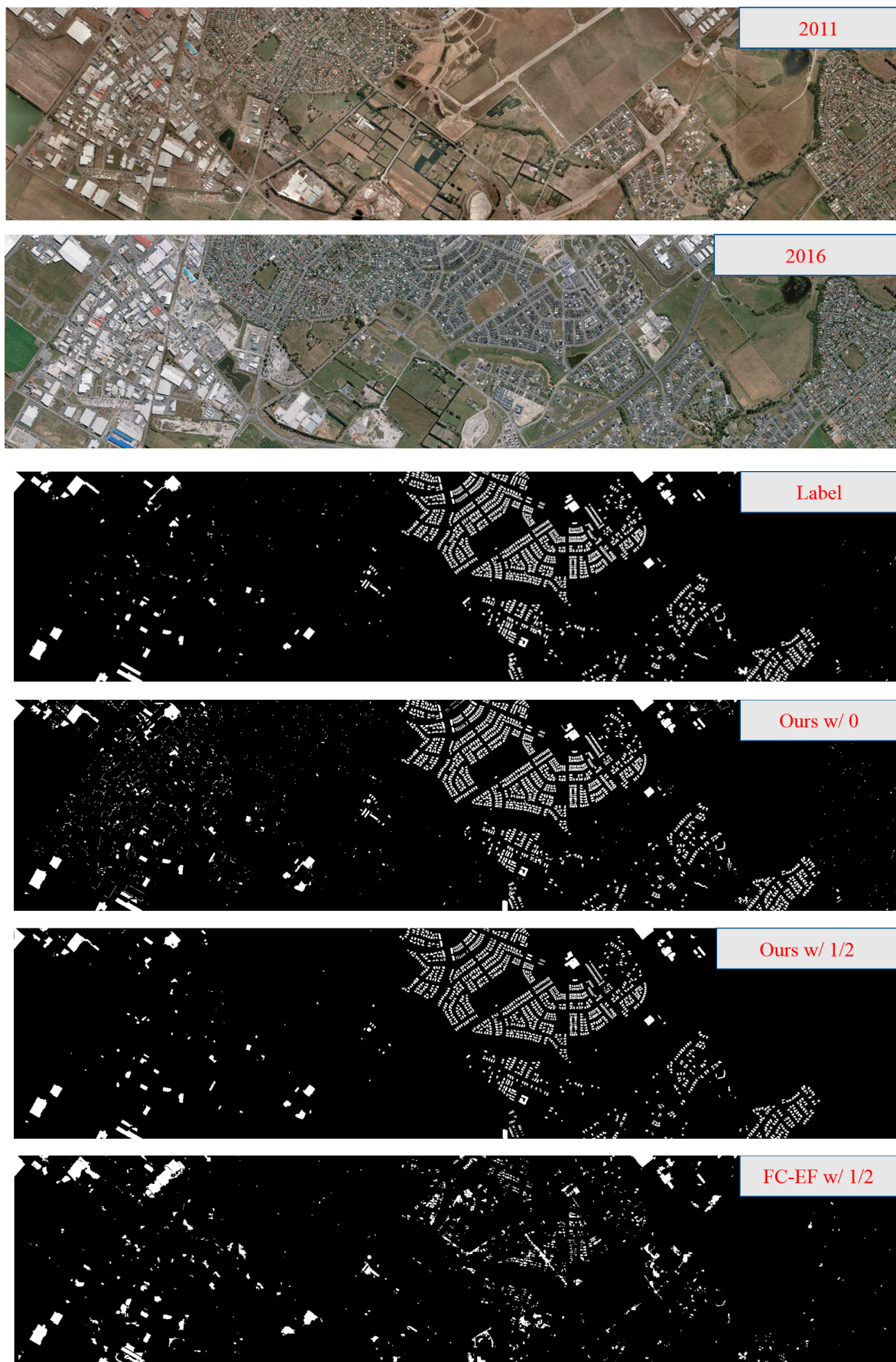


Figure 12. Cont.

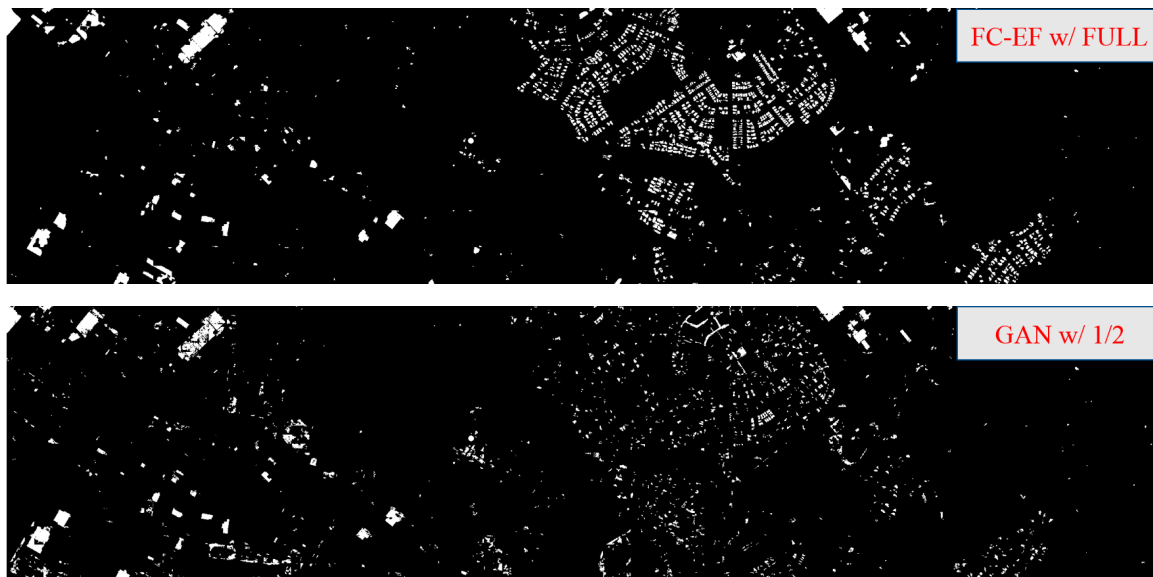


Figure 12. Change detection results of different methods on the whole test area. From top to bottom: 2011 and 2016 images; label with 1715 truly changed buildings; results of our method (Mask R-CNN) without change samples; our method (Mask R-CNN) fine-tuned on half of the change samples (green box in Figure 10); FC-EF trained on half of the samples; FC-EF trained on all of the samples; GAN-based method trained on half of the samples. We did not list the results of our methods with the MS-FCN and the results with all of the samples because the former looks the same as with the Mask R-CNN, and the latter looks the same as with half samples.

4. Discussion

In this section, we further discuss: (1) the advantages of our change detection network compared to the traditional methods with available building masks from the building extraction network, (2) the prerequisites of our method and (3) potential improvement of the framework.

(1) The advantages of our change detection network

Even after extracting the building mask, change detection at the object (building instance) level could still be extremely challenging with a traditional change detection method. As most of the object-based methods treat arbitrary groups of pixels as objects, we only compare our algorithm with our empirical designs (Table 4) at the building instance level. Table 4 shows that, although different empirical methods were tried, the accuracy they obtained was much lower than our change detection network. In addition, the parameters are unstable and data-specific. This is the reason that a self-training CNN is applied to the bi-temporal building classification map for building change detection.

Table 5 shows the comparison between our method and other conventional methods (i.e., an image-differencing-based method [24], an image-ratioing-based method [25] and the FLICM [73]) at the pixel level. The IoU score of all the methods was obviously lower than ours. Note that on binary maps the differencing and ratioing methods achieved the same results.

(2) The prerequisite of our method

The change detection framework depends on the accuracy of the building extraction network, which requires sufficient training data. However, we don't treat this as a shortcoming, as there are plenty of existing building datasets. Besides the open-source datasets such as the WHU [66], Inria [74] and OSM (open street map) [75], there are global building GIS maps in central or local government branches of surveying, mapping and city planning, which can be used in practice. The critical shortage is the change samples. This problem has been greatly reduced with our self-training strategy.

(3) The potential improvement of the framework

Although we only simulated changed samples on the SI-2016 area containing 1892 tiles, they enabled us to train a very good model at pixel and object levels. Especially at the pixel level, the self-trained model

approached the top performance achieved with adequate real change samples. Moreover, the self-training area can easily be extended, which would further improve the model's accuracy at the instance level without requiring real samples of changed buildings.

Table 4. Different methods to discover changed building instances from the bi-temporal building classification map produced by the Mask R-CNN. "Difference" indicates the direct differencing of the two classification maps; "Distance & IoU 1" indicates a threshold of a 20-pixel shift of center points of corresponding building masks, and 0.33 IoU is used to determine if the two masks of different times are the same buildings. "Distance & IoU 2" indicates a threshold of a 20-pixel shift of center points of corresponding building masks, and 0.5 IoU is used. "Erode & dilate" indicates we used a morphological erosion operation to eliminate small masks and alignment errors and a dilation operation to restore buildings. "Erode & intersect" indicates that we used an erosion operation followed by an intersection operation with a threshold of 0.5 IoU. "Our network" is the change detection network trained with simulated samples.

Method	AP	Precision	Recall
Difference	0.010	0.010	0.872
Distance & IoU 1	0.290	0.345	0.839
Distance & IoU 2	0.290	0.343	0.844
Erode & dilate	0.388	0.489	0.793
Erode & intersect	0.450	0.540	0.832
Our network	0.630	0.644	0.943

Table 5. Different methods to discover changed buildings at the pixel level from the bi-temporal building classification map produced by the Mask R-CNN. "Our network" is the change detection network trained with simulated samples.

Method	IoU	Precision	Recall
Differencing [24]	0.667	0.709	0.918
ratioing [25]	0.667	0.709	0.918
FLICM [73]	0.678	0.723	0.917
Our network	0.798	0.856	0.922

5. Conclusions

This paper proposes a new building change detection framework with a building extraction network and a self-trained building change detection network for VHR remote sensing images. The building extraction network provides highly accurate building classification maps. The building change detection network takes bi-temporal building classification maps as inputs and computes building change maps at the pixel and object levels. The network could be well trained through simulated changed buildings, and is robust to the registration errors caused by unavoidable parallax changes in the VHR images. The experimental results proved the distinctive superiority of the proposed algorithm compared with other recent methods. Without any real change labels, our change detection network dominated other methods trained with adequate samples. As the change labels are commonly scarce, the reduced demand of training samples makes our framework effective and applicable in practice. In this study we focused on building change detection, but the application of our framework can be easily adapted to detect changes of other land cover objects.

Author Contributions: S.J. led the research and wrote the paper; Y.S. performed the experiments; M.L. analyzed the results and revised the paper; Y.Z. edited the paper.

Funding: This work was supported by the National Key Research and Development Program of China, Grant No. 2018YFB0505003.

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

References

1. Singh, A. Review Article Digital change detection techniques using remotely-sensed data. *Inter. J. Remote Sens.* **1989**, *10*, 989–1003. [[CrossRef](#)]
2. Chen, G.; Hay, G.J.; Carvalho, L.M.; Wulder, M.A. Object-based change detection. *Int. J. Remote Sens.* **2012**, *33*, 4434–4457. [[CrossRef](#)]
3. Coops, N.C.; Wulder, M.A.; White, J.C. Identifying and describing forest disturbance and spatial pattern: data selection issues and methodological implications. In *Forest Disturbance and Spatial Pattern: Remote Sensing and GIS Approaches*; CRC Press (Taylor and Francis): Boca Raton, FL, USA, 2006; pp. 33–60.
4. Lunetta, R.S.; Johnson, D.M.; Lyon, J.G.; Crotwell, J. Impacts of imagery temporal frequency on land-cover change detection monitoring. *Remote. Sens. Environ.* **2004**, *89*, 444–454. [[CrossRef](#)]
5. Shalaby, A.; Tateishi, R. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. *Appl. Geogr.* **2007**, *27*, 28–41. [[CrossRef](#)]
6. Peiman, R. Pre-classification and post-classification change-detection techniques to monitor land-cover and land-use change using multi-temporal Landsat imagery: A case study on Pisa Province in Italy. *Int. J. Remote. Sens.* **2011**, *32*, 4365–4381. [[CrossRef](#)]
7. Ochoa-Gaona, S.; González-Espinosa, M. Land use and deforestation in the highlands of Chiapas, Mexico. *Appl. Geogr.* **2000**, *20*, 17–42. [[CrossRef](#)]
8. Green, K.; Kempka, D.; Lackey, L. Using remote sensing to detect and monitor land-cover and land-use change. *Photogramm. Eng. Remote Sens.* **1994**, *60*, 331–337.
9. Torres-Vera, M.A.; Prol-Ledesma, R.M.; García-López, D. Three decades of land use variations in Mexico City. *Int. J. Remote. Sens.* **2008**, *30*, 117–138. [[CrossRef](#)]
10. Jenson, J. Detecting residential land use development at the urban fringe. *Photogramm. Eng. Remote Sens.* **1982**, *48*, 629–643.
11. 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.
12. Koltunov, A.; Ustin, S. Early fire detection using non-linear multitemporal prediction of thermal imagery. *Remote. Sens. Environ.* **2007**, *110*, 18–28. [[CrossRef](#)]
13. Coops, N.C.; Gillanders, S.N.; Wulder, M.A.; Gergel, S.E.; Nelson, T.; Goodwin, N.R. Assessing changes in forest fragmentation following infestation using time series Landsat imagery. *For. Ecol. Manag.* **2010**, *259*, 2355–2365. [[CrossRef](#)]
14. Hame, T.; Heiler, I.; San Miguel-Ayaz, J. An unsupervised change detection and recognition system for forestry. *Int. J. Remote. Sens.* **2010**, *19*, 1079–1099. [[CrossRef](#)]
15. Wulder, M.A.; Butson, C.R.; White, J.C. Cross-sensor change detection over a forested landscape: Options to enable continuity of medium spatial resolution measures. *Remote. Sens. Environ.* **2008**, *112*, 796–809. [[CrossRef](#)]
16. Deer, P. *Digital Change Detection Techniques in Remote Sensing*; Defence Science and Technology Organization: Canberra, Australia, 1995.
17. Jenson, J. Urban/suburban land use analysis. In *Manual of Remote Sensing*; American Society of Photogrammetry, 1983; pp. 1571–1666. Available online: <https://ci.nii.ac.jp/naid/10003189509/> (accessed on 3 May 2019).
18. 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](#)]
19. Lu, D.; Mausel, P.; Brondizio, E.; Moran, E. Change detection techniques. *Int. J. Remote Sens.* **2004**, *25*, 2365–2401. [[CrossRef](#)]
20. Chen, X.; Vierling, L.; Deering, D. A simple and effective radiometric correction method to improve landscape change detection across sensors and across time. *Remote. Sens. Environ.* **2005**, *98*, 63–79. [[CrossRef](#)]
21. Du, Y.; Teillet, P.M.; Cihlar, J. Radiometric normalization of multitemporal high-resolution satellite images with quality control for land cover change detection. *Remote. Sens. Environ.* **2002**, *82*, 123–134. [[CrossRef](#)]

22. Huang, X.; Zhang, L.; Zhu, T. Building change detection from multitemporal high-resolution remotely sensed images based on a morphological building index. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 105–115. [[CrossRef](#)]
23. Xiao, P.; Yuan, M.; Zhang, X.; Feng, X.; Guo, Y. Cosegmentation for Object-Based Building Change Detection From High-Resolution Remotely Sensed Images. *IEEE Trans. Geosci. Remote. Sens.* **2017**, *55*, 1587–1603. [[CrossRef](#)]
24. Quarmby, N.A.; Cushnie, J.L. Monitoring urban land cover changes at the urban fringe from SPOT HRV imagery in south-east England. *Int. J. Remote. Sens.* **1989**, *10*, 953–963. [[CrossRef](#)]
25. Howarth, P.J.; Wickware, G.M. Procedures for change detection using Landsat digital data. *Int. J. Remote. Sens.* **1981**, *2*, 277–291. [[CrossRef](#)]
26. Ludeke, A.K.; Maggio, R.C.; Reid, L.M. An analysis of anthropogenic deforestation using logistic regression and GIS. *J. Environ. Manag.* **1990**, *31*, 247–259. [[CrossRef](#)]
27. Lunetta, R.S.; Elvidge, C.D. *Remote Sensing Change Detection*; Taylor & Francis: Abingdon, UK, 1999; Volume 310.
28. Richards, J. Thematic mapping from multitemporal image data using the principal components transformation. *Remote. Sens. Environ.* **1984**, *16*, 35–46. [[CrossRef](#)]
29. Kauth, R.J.; Thomas, G. The tasselled cap—A graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. In Proceedings of the LARS Symposia, West Lafayette, IN, USA, 29 June–1 July 1976; p. 159.
30. 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.* **2007**, *45*, 218–236. [[CrossRef](#)]
31. 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](#)]
32. 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 (JURSE 2011), Munich, Germany, 11–13 April 2011; pp. 329–332.
33. Im, J.; Jensen, J.R. A change detection model based on neighborhood correlation image analysis and decision tree classification. *Remote. Sens. Environ.* **2005**, *99*, 326–340. [[CrossRef](#)]
34. Bouziani, M.; Goïta, K.; He, D.-C. Automatic change detection of buildings in urban environment from very high spatial resolution images using existing geodatabase and prior knowledge. *ISPRS J. Photogramm. Remote. Sens.* **2010**, *65*, 143–153. [[CrossRef](#)]
35. Yuan, F.; Sawaya, K.E.; Loeffelholz, B.C.; Bauer, M.E. Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing. *Remote. Sens. Environ.* **2005**, *98*, 317–328. [[CrossRef](#)]
36. Serpico, S.B.; Moser, G. Weight Parameter Optimization by the Ho–Kashyap Algorithm in MRF Models for Supervised Image Classification. *IEEE Trans. Geosci. Remote. Sens.* **2006**, *44*, 3695–3705. [[CrossRef](#)]
37. Wiemker, R. An iterative spectral-spatial Bayesian labeling approach for unsupervised robust change detection on remotely sensed multispectral imagery. In Proceedings of the Transactions on Rough Sets VII, Kiel, Germany, 10–12 September 1997; Volume 1296, pp. 263–270.
38. Melgani, F.; Bazi, Y. Markovian Fusion Approach to Robust Unsupervised Change Detection in Remotely Sensed Imagery. *IEEE Geosci. Remote. Sens. Lett.* **2006**, *3*, 457–461. [[CrossRef](#)]
39. Pijanowski, B.C.; Pithadia, S.; Shellito, B.A.; Alexandridis, K.; Alexandridis, K. Calibrating a neural network-based urban change model for two metropolitan areas of the Upper Midwest of the United States. *Int. J. Geogr. Inf. Sci.* **2005**, *19*, 197–215. [[CrossRef](#)]
40. Liu, X.; Lathrop, R.G. Urban change detection based on an artificial neural network. *Int. J. Remote. Sens.* **2002**, *23*, 2513–2518. [[CrossRef](#)]
41. Nemmour, H.; Chibani, Y. Multiple support vector machines for land cover change detection: An application for mapping urban extensions. *ISPRS J. Photogramm. Remote. Sens.* **2006**, *61*, 125–133. [[CrossRef](#)]
42. Huang, C.; Song, K.; Kim, S.; Townshend, J.R.; Davis, P.; Masek, J.G.; Goward, S.N. Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote. Sens. Environ.* **2008**, *112*, 970–985. [[CrossRef](#)]
43. Hall, O.; Hay, G.J. A Multiscale Object-Specific Approach to Digital Change Detection. *Int. J. Appl. Earth Obs. Geoinf.* **2003**, *4*, 311–327. [[CrossRef](#)]

44. Lefebvre, A.; Corpetti, T.; Hubert-Moy, L. Object-Oriented Approach and Texture Analysis for Change Detection in Very High Resolution Images. In Proceedings of the IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, MA, USA, 7–11 July 2008.
45. Fisher, P. The pixel: a snare and a delusion. *Int. J. Remote Sens.* **1997**, *18*, 679–685. [[CrossRef](#)]
46. De Chant, T.; Kelly, N.M. Individual Object Change Detection for Monitoring the Impact of a Forest Pathogen on a Hardwood Forest. *Photogramm. Eng. Remote. Sens.* **2009**, *75*, 1005–1013. [[CrossRef](#)]
47. Conchedda, G.; Durieux, L.; Mayaux, P. An object-based method for mapping and change analysis in mangrove ecosystems. *ISPRS J. Photogramm. Remote. Sens.* **2008**, *63*, 578–589. [[CrossRef](#)]
48. Du, S.; Zhang, Y.; Qin, R.; Yang, Z.; Zou, Z.; Tang, Y.; Fan, C. Building Change Detection Using Old Aerial Images and New LiDAR Data. *Remote. Sens.* **2016**, *8*, 1030. [[CrossRef](#)]
49. Liu, H.; Yang, M.; Chen, J.; Hou, J.; Deng, M. Line-Constrained Shape Feature for Building Change Detection in VHR Remote Sensing Imagery. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 410. [[CrossRef](#)]
50. Gong, M.; Niu, X.; Zhang, P.; Li, Z. Generative Adversarial Networks for Change Detection in Multispectral Imagery. *IEEE Geosci. Remote. Sens. Lett.* **2017**, *14*, 2310–2314. [[CrossRef](#)]
51. Zhang, P.; Gong, M.; Su, L.; Liu, J.; Li, Z. Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images. *ISPRS J. Photogramm. Remote. Sens.* **2016**, *116*, 24–41. [[CrossRef](#)]
52. Mou, L.; Bruzzone, L.; Zhu, X.X. Learning Spectral-Spatial-Temporal Features via a Recurrent Convolutional Neural Network for Change Detection in Multispectral Imagery. *IEEE Trans. Geosci. Remote. Sens.* **2019**, *57*, 924–935. [[CrossRef](#)]
53. Gong, M.; Zhan, T.; Zhang, P.; Miao, Q. Superpixel-Based Difference Representation Learning for Change Detection in Multispectral Remote Sensing Images. *IEEE Trans. Geosci. Remote. Sens.* **2017**, *55*, 1–16. [[CrossRef](#)]
54. 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](#)]
55. Gong, M.; Zhao, J.; Liu, J.; Miao, Q.; Jiao, L. Change Detection in Synthetic Aperture Radar Images Based on Deep Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 125–138. [[CrossRef](#)]
56. Khan, S.H.; He, X.; Porikli, F.; Bennamoun, M. Forest Change Detection in Incomplete Satellite Images With Deep Neural Networks. *IEEE Trans. Geosci. Remote. Sens.* **2017**, *55*, 5407–5423. [[CrossRef](#)]
57. Ding, A.; Zhang, Q.; Zhou, X.; Dai, B. Automatic recognition of landslide based on CNN and texture change detection. In Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, China, 11–13 November 2016; pp. 444–448.
58. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–8 December 2012; pp. 1097–1105.
59. Hinton, G.E.; Osindero, S.; Teh, Y.-W. A Fast Learning Algorithm for Deep Belief Nets. *Neural Comput.* **2006**, *18*, 1527–1554. [[CrossRef](#)]
60. Zhang, F.; Du, B.; Zhang, L. Saliency-Guided Unsupervised Feature Learning for Scene Classification. *IEEE Trans. Geosci. Remote. Sens.* **2015**, *53*, 2175–2184. [[CrossRef](#)]
61. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montréal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
62. 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.
63. Nemoto, K.; Imaizumi, T.; Hikosaka, S.; Hamaguchi, R.; Sato, M.; Fujita, A. Building change detection via a combination of CNNs using only RGB aerial imageries. *Remote Sens. Tech. Appl. Urban Environ.* **2017**, *10431*, 23.
64. Zhang, Z.; Vosselman, G.; Gerke, M.; Tuia, D.; Yang, M.Y. Change Detection between Multimodal Remote Sensing Data Using Siamese CNN. *arXiv* **2018**, arXiv:1807.09562.
65. El Amin, A.M.; Liu, Q.; Wang, Y. Zoom out CNNs features for optical remote sensing change detection. In Proceedings of the 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, China, 2–4 June 2017; pp. 812–817.

66. Ji, S.; Wei, S.; Lu, M. Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set. *IEEE Trans. Geosci. Remote. Sens.* **2019**, *57*, 574–586. [[CrossRef](#)]
67. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
68. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Lecture Notes in Computer Science*; Springer Nature: Switzerland, 2015; Volume 9351, pp. 234–241. Available online: https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28 (accessed on 3 May 2019).
69. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)]
70. Lebedev, M.A.; Vizilter, Y.V.; Vygolov, O.V.; Knyaz, V.A.; Rubis, A.Y. Change detection in remote sensing images using conditional adversarial networks. *ISPRS - Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2018**, *XLII-2*, 565–571. [[CrossRef](#)]
71. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the Topics in Artificial Intelligence, Munich, Germany, 8–14 September 2018; pp. 833–851.
72. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the Transactions on Rough Sets VII, Zurich, Switzerland, 6–12 September 2014; Volume 8693, pp. 740–755.
73. Chatzis, V.; Krinidis, S. A Robust Fuzzy Local Information C-Means Clustering Algorithm. *IEEE Trans. Image Process.* **2010**, *19*, 1328–1337.
74. Dalal, N. *Finding People in Images and Videos*; Institut National Polytechnique de Grenoble-INPG: Grenoble, France, 2006.
75. OpenStreetMap. Available online: <https://www.openstreetmap.org> (accessed on 30 April 2019).



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).