Change detection based on Faster R-CNN for high-resolution remote sensing images

Qing Wang, Xiaodong Zhang, Guanzhou Chen, Fan Dai, Yuanfu Gong & Kun Zhu

To cite this article: Qing Wang, Xiaodong Zhang, Guanzhou Chen, Fan Dai, Yuanfu Gong & Kun Zhu (2018) Change detection based on Faster R-CNN for high-resolution remote sensing images, Remote Sensing Letters, 9:10, 923-932, DOI: 10.1080/2150704X.2018.1492172

To link to this article: https://doi.org/10.1080/2150704X.2018.1492172

Published online: 22 Aug 2018.
Change detection based on Faster R-CNN for high-resolution remote sensing images

Qing Wang\(^1\)\(^2\), Xiaodong Zhang\(^1\), Guanzhou Chen\(^\circ\)\(^1\), Fan Dai\(^1\), Yuanfu Gong\(^1\) and Kun Zhu\(^1\)

\(^1\)State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China; \(^2\)School of Geosciences, Yangtze University, Wuhan, China

ABSTRACT

Change detection is of great significance in remote sensing. The advent of high-resolution remote sensing images has greatly increased our ability to monitor land use and land cover changes from space. At the same time, high-resolution remote sensing images present a new challenge over other satellite systems, in which time-consuming and tiresome manual procedures must be needed to identify the land use and land cover changes. In recent years, deep learning (DL) has been widely used in the fields of natural image target detection, speech recognition, face recognition, etc., and has achieved great success. Some scholars have applied DL to remote sensing image classification and change detection, but seldomly to high-resolution remote sensing images change detection. In this letter, faster region-based convolutional neural networks (Faster R-CNN) is applied to the detection of high-resolution remote sensing image change. Compared with several traditional and other DL-based change detection methods, our proposed methods based on Faster R-CNN achieve higher overall accuracy and Kappa coefficient in our experiments. In particular, our methods can reduce a large number of false changes.

ARTICLE HISTORY

Received 19 February 2018
Accepted 8 June 2018

1. Introduction

Change detection is the process of identifying differences in the state of an object or phenomenon by observing it at different times (Singh 1989). In remote sensing applications, we compare and analyze the changes between different remote sensing images and related data in different periods through image processing and mathematical models (Tong et al. 2015). In recent years, change detection has important applications in land and resource management, agricultural and forestry monitoring, natural disaster monitoring and assessment (Tian et al. 2013; Stramondo et al. 2006). The main tasks of using remote sensing images to detect land surface change are to answer three questions: Whether is the land surface changed? What is it changed for? And what is the change trajectory (Zhou 2011)? This letter mainly discusses the change detection based on the bi-temporal remote sensing images, and therefore it does not involve the change trajectory. Many scholars have put forward many different methods of remote sensing image change.
inspection, mainly including algebra, classification, transformation, advanced models, geographic information system (GIS) approaches, visual analysis, etc. (Lu et al. 2004).

With the increasing application of remote sensing change detection, some traditional methods are very limited, which is mainly reflected in two aspects. First, with the variety of remote sensing data source, the improvement of the spatial resolution of images and the enrichment of image details, traditional methods, such as image algebra, are easily influenced by the effects of changing season, satellite sensor, solar elevation and atmospheric conditions and reduce change detection accuracy (Pacifici et al. 2007). Second, although some methods, such as object-based image analysis and transformation, reduce false changes by extracting geometric and textural features, it requires time-consuming and tiresome manual procedures. They easily erase the advantage of an automated change detection technique.

Deep learning (DL) can automatically learn deep features from raw data, so that it can adapt to different situations of remote sensing image change detection, and avoid many shortcomings of traditional methods (Hinton and Salakhutdinov 2006). In recent years, deep learning has been widely applied in remote sensing image change detection (Ball, Anderson, and Chan 2017; Zhu et al. 2018). For example, Zhang proposes sparse autoencoder-based method to synthetic aperture radar (SAR) images change detection (Zhang et al. 2016). Lyu makes use of a recurrent neural network (RNN) based on long short-term memory (LSTM) to solve the multispectral change detection task (Lyu, Lu, and Mou 2016) and applies a RNN-based transfer-learning approach to detect annual urban dynamics using Landsat data. (Lyu et al. 2018). Gong compares the traditional change detection method with restricted boltzmann machine-based (RBM) method applied to SAR images which can further improve the detection performance (Gong et al. 2017). Zhan proposes a change detection method based on a deep siamese convolutional network for optical aerial images, and the siamese network is learned to extract features directly from the image pairs (Zhan et al. 2017). Most of these DL-based change detection methods have image processing by pixel or pixel neighborhood. These methods are not verified on high-resolution remote sensing images. Convolutional neural networks (CNN) originally designed natural image classification, which could directly process two-dimensional image data. It was proved to be an effective method to deal with classification of remote sensing images (Maggiori et al. 2016). Faster region-based convolutional neural networks (Faster R-CNN) is a region-based object detection method (Ren et al. 2017), by which the geometric and contextual spatial features of targets can be learnt. In the natural image target detection, Faster R-CNN achieved rapid extraction and target recognition in the complex scene. We try to apply this method to the change detection of high-resolution remote sensing images and regard the “changed” in remote sensing images as the target to be detected, while the “unchanged” is regarded as a background.

In summary, the contributions of this letter are as follow:

(1) We propose a new high-resolution remote sensing image change detection framework based on Faster R-CNN.

(2) Two kinds of high-resolution remote sensing image change models based on Faster R-CNN are proposed and compared with the traditional methods and other DL-based methods.
2. Methodology

In this section, we first introduce the basic principle and training process of Faster R-CNN, and then propose two kinds of remote sensing image change detection models based on Faster R-CNN.

2.1. Faster R-CNN architectures

Faster R-CNN has two tasks of detecting the location and category of the target. The former obtains the bounding box of the target and it is described by four parameters, the latter gets the category of the target in the bounding box. The basic frame of Faster R-CNN is shown in Figure 1.

The shared convolutional layers are the remaining parts of the CNN removing the output layer, and the main function is to extract features and get feature maps. In this letter, the shared convolution layers use a 50-layer Residual Net (ResNet-50) (He et al. 2016). Region Proposal Network (RPN) is a full convolution network. It uses a small network to slide and scan on the feature maps obtained by CNN to get low dimensional vectors, and predict the location and category of targets through two fully connected layers. The recommended area is end-to-end training through backpropagation and stochastic gradient descent (Lecun et al. 2014).

In this letter, we use the $3 \times 3$ sliding window. Each location corresponds to 3 scales with box areas of $128^2$, $256^2$ and $512^2$ pixels, and 3 aspect ratios of 1:1, 1:2 and 2:1 on the original map. We get the first 300 regions that are scored above 0.7 as proposals. Region of Interest (ROI) layer is responsible for unifying the characteristics of proposed regions into the same size by the method of down-sampling, and then return it to the exact target accurate location and category through fully connected layers. Compared with the traditional CNN, Faster R-CNN has two output layers, and its loss function $L$ is defined as equation (1):

$$L(p, k^*, t, t^*) = L_{cls}(p, k^*) + \lambda k^* \geq 1 || L_{loc}(t, t^*)$$  \hspace{1cm} (1)

![Figure 1. Framework of Faster R-CNN.](image-url)
Here, $L_{\text{cls}}$ is the loss function of the softmax layer. $p$ is the probability of each proposed region being the target, and $k^*$ is a label for the category. $L_{\text{loc}}$ is the loss function of bounding box regression layer. $t^*$ is a vector representing the 4 parameterized coordinates of the ground-truth box, $t$ is a vector representing the 4 parameterized coordinates of the predicted bounding box. The two terms ($L_{\text{cls}}$ and $L_{\text{loc}}$) are weighted by a balancing parameter $\lambda$, and they are roughly equally weighted when the value of $\lambda$ is 10.

### 2.2. Training Faster R-CNN

The shared convolutional layers are initialized by pretraining resnet-50 model for ImageNet classification (Russakovsky et al. 2014), others randomly initialized. Our training samples are taken from two real color remote sensing images of different time and different sensors. We unify the image coordinate reference system and resample to 0.5m spatial resolution.

According to the experimental task, we manually select three types of change samples, including change from vegetation to building, from bare-land to building and from farmland to road. For each type about 500 samples are selected, the length and width of most samples are in 200 – 800 pixels, and the sample size can be inconsistent. In addition, we augment the sample. Data augmentation is a method to enhance generalizability of networks without costing extra training time (Allen 1974). In our framework, it is achieved by randomly rotating by between $0^\circ$ and $45^\circ$ and vertically and horizontally transforming $0$ – $30\%$ for all training data. The number of sample is expanded to seven times the original one. In the end, there are 10,563 samples for training.

### 2.3. Change detection model based on Faster R-CNN

After we preprocess two different time images, geographic coordinate system and image spatial resolution are unified with second time images as a quasi. Two kinds of remote sensing image change detection models based on Faster R-CNN are proposed in this letter. The first one is to merge the images of two different times, and then send them to Faster R-CNN for change detection (MFRCNN). Another one is to subtract the first-date image from the second-date image, pixel by pixel and get the different image, and then send it to Faster R-CNN for change detection (SFRCNN). The training samples of the two models should be processed correspondingly. The results are a series of rectangular regions, and then the rectangular intersecting regions with the same change type are merged. Finally, the snakes (Kass, Witkin, and Terzopoulos 1988) model is used to segment the exact change area.

### 3. Experiments

In order to assess the effectiveness of the proposed framework, we challenge multi-temporal and multi-spatial-resolution change detection problems on two real datasets. Change detection accuracies are compared with previous methods including change vector analysis (CVA) (Bovolo and Bruzzone 2006), object-based change detection method (OBCD) (Descle, Bogaert, and Defourny 2006), post-classification comparison
(PCC) (Colditz et al. 2012). At the same time, we compare with other DL-based methods including stacked autoencoder (SAE) (Schlkopf, Platt, and Hofmann 2006), CNN (Hu et al. 2015) and long short-term memory (LSTM) (Ordez and Roggen 2016). Their basic process is to merge the bi-temporal images first, then we select neighborhood of size $3 \times 3$ for each position pixel and get a 1D vector of length 54, and finally send the corresponding deep learning model for change detection.

### 3.1. Datasets description

In this letter, two datasets are selected for comparison experiment. The first dataset, shown in Figure 2, is fetched from Anhui Province, China in 2013, with the size of $1669 \times 1368$ m. The two different time images have both been registered. The first-date image shown in Figure 2(a) is 1m resolution true color image from the IKONOS satellite in March 2013, and the second-date image shown in Figure 2(b) is 0.5m resolution unmanned aerial vehicle (UAV) true color image in June 2013. The main types of land cover in the region include buildings, bare-land, vegetation, and roads. The main change is the appearance of the new building. The ground truth, shown in Figure 2(c), is created by integrating prior information with artificial interpretation based on the input images in Figure 2(a) and (b).

The second dataset, shown in Figure 3, is fetched from Hubei Province, China, with the size of $615 \times 586$ m. The two different time images have also been registered. The first time image shown in Figure 3(a) is 1m resolution true color image from the Chinese GaoFen-2 satellite in November 2015, the second-date image shown in Figure 3(b) is 0.5 m resolution true color image from the Pleiades-1B satellite in November 2016. The background of the experiment is the 2015–2016 land use change survey in Hubei, China. November was the harvest season for rice in Hubei. Most of the rice had been harvested in the first-date image, but not in the second-date image. Most of the farmlands in first-data image have bare soil feature, and most of the farmlands in the second-data image have vegetation feature. The type of land use did not change. The main types of land use in the region include buildings, farmlands, vegetation, and roads. There are three main types of land use changes: change from vegetation to buildings (change A, red zone in Figure 3(c)), change from bare-land to

![Figure 2](image-url)  
**Figure 2.** Bi-temporal images (after co-registration) fetched from Anhui, China with a WGS-84 projection and a coordinate range of $30°52'53.84N - 30°53'38.86N, 118°17'5.93E - 118°18'8.13E$. (a) The image acquired from IKONOS on March 2013. (b) The image acquired from UAV on June 2013. (c) The ground truth image. (The red and black indicate the changed regions and the unchanged regions.).
buildings (change B, yellow zone in Figure 3(c)) and change from farmlands to roads (change C, blue zone in Figure 3(c)). The ground truth, shown in Figure 3(c), was created by integrating prior information with artificial interpretation based on the input images in Figure 3(a) and (b).

3.2. Evaluation criteria

The results of change detection can also be regarded as a special classification result. Whether it is detected to be changed or not can be seen as two classification problems, and detecting change from one to another can be seen as a multi-class classification problem. The most common accuracy assessment elements include overall accuracy (OA), producer’s accuracy (PA), user’s accuracy (UA) and kappa coefficient (Lu et al. 2004). This letter uses these indexes to evaluate the accuracy.

3.3. Experiment result

In the first experiment, the change detection results of traditional methods like CVA, OBCB and PPC are shown in Figure 4(a), (b) and (c), and the change detection results of other DL-based methods like SAE, CNN and LSTM are shown in Figure 4(d), (e) and (f). Based on Faster-RCNN, the change detection results are shown in Figure 4(g) and (h). By the way, because most of the buildings in the images are rectangular, we do not need segmentation. In Figure 4, there are more false changes in the first three traditional methods compared to the last five DL-based methods because of the difference in sensor. Compared with other three DL-based methods, our proposed methods further reduce false changes, especially in the left bottom of Figure 4. The quantitative analysis on PA, UA, OA and Kappa coefficient of different methods is shown in Table 1. As shown in Table 1, the DL-based method achieves better change detection results than the traditional method. The two methods based on Faster-RCNN have improved OA and kappa coefficient, and SFRCNN has better performance.
In the second experiment, we need to detect the type of land use change. Therefore, we give up the tedious traditional methods and directly compare different methods based on deep learning. The change detection results of methods such as SAE, LSTM and CNN are shown in Figure 5(a), (b) and (c). Based on Faster-RCNN, change detection results are shown in Figure 5(d) and (e). Because of the land use change detection, the farmlands with bare soil feature and the farmlands with vegetation feature are regarded as being unchanged. In this respect, we can see that our proposed method is better. According to the result of accuracy comparison shown in Table 2, OA and kappa coefficient are all improved based on the Faster-RCNN, and SFRCNN has better performance.

Table 1. Change detection results of different methods obtained using the first dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Changed PA</th>
<th>Unchanged PA</th>
<th>OA</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVA</td>
<td>0.8233</td>
<td>0.9685</td>
<td>0.9651</td>
<td>0.5054</td>
</tr>
<tr>
<td>PPC</td>
<td>0.8058</td>
<td>0.9669</td>
<td>0.9631</td>
<td>0.4909</td>
</tr>
<tr>
<td>OBCD</td>
<td>0.9972</td>
<td>0.9534</td>
<td>0.9544</td>
<td>0.4832</td>
</tr>
<tr>
<td>SAE</td>
<td>0.6473</td>
<td>0.9952</td>
<td>0.9729</td>
<td>0.5602</td>
</tr>
<tr>
<td>CNN</td>
<td>0.6451</td>
<td>0.9802</td>
<td>0.9721</td>
<td>0.5397</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.7413</td>
<td>0.9814</td>
<td>0.9755</td>
<td>0.6020</td>
</tr>
<tr>
<td>MFRCNN</td>
<td>0.7262</td>
<td>0.9880</td>
<td>0.9820</td>
<td>0.6403</td>
</tr>
<tr>
<td>SFRCNN</td>
<td>0.6674</td>
<td>0.9955</td>
<td>0.9880*</td>
<td>0.7116</td>
</tr>
</tbody>
</table>

*The figure in bold in each table mean the optimum results of all.
4. Discussion and conclusion

Although high-resolution remote sensing images have a great potential for change detection, the effects of changing season, satellite sensor, solar elevation and atmospheric conditions also reduce the achievable accuracy in change detection output. Traditional methods require time-consuming and tiresome manual procedures to reduce false changes. For example, in our first experiment, we normalize the color of the image before using the CVA method, and the post-classification process is used when using the PPC method. The OBCD method requires multiple attempts to segment parameters to achieve the best results of extracting spectral, geometric and textural features. The method based on deep learning can extract deep features from the training samples and be more intelligent. It not only improves the degree of automation, but also improves the accuracy of change detection. At present, most of the deep learning-based change detection methods deal with images in 1D pixel vector. For high-resolution images, it is obvious that some features such as geometric and contextual spatial features are lost. In this letter, Faster R-CNN is applied to the detection of high-resolution remote sensing image change. This region-based CNN method can extract the 2D deep features of the region. Compared with previous DL-based methods, this method can further improve change detection accuracy. In addition, we propose two change detection models based on Faster R-CNN: MFRCNN and SFRCNN. Two experiments show that SFRCNN works much better than MFRCNN. This indicates that although DL-based method can extract more features, the image difference method can help to improve its accuracy. Of course MFRCNN framework is simpler and more automated. There are

<table>
<thead>
<tr>
<th>Method</th>
<th>Change A</th>
<th>Change B</th>
<th>Change C</th>
<th>Unchanged</th>
<th>OA</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAE</td>
<td>0.1921</td>
<td>0.4113</td>
<td>0.6977</td>
<td>0.9683</td>
<td>0.9321</td>
<td>0.6431</td>
</tr>
<tr>
<td>UA</td>
<td>0.2001</td>
<td>0.4839</td>
<td>0.6701</td>
<td>0.9696</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>0.5942</td>
<td>0.5176</td>
<td>0.4974</td>
<td>0.9801</td>
<td>0.9304</td>
<td>0.5944</td>
</tr>
<tr>
<td>UA</td>
<td>0.5129</td>
<td>0.4259</td>
<td>0.7852</td>
<td>0.9520</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSTM</td>
<td>0.5104</td>
<td>0.4652</td>
<td>0.6532</td>
<td>0.9785</td>
<td>0.9411</td>
<td>0.6690</td>
</tr>
<tr>
<td>UA</td>
<td>0.7115</td>
<td>0.4702</td>
<td>0.7605</td>
<td>0.9629</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFRCNN</td>
<td>0.7099</td>
<td>0.5433</td>
<td>0.7289</td>
<td>0.9863</td>
<td>0.9577</td>
<td>0.7592</td>
</tr>
<tr>
<td>UA</td>
<td>0.4748</td>
<td>0.6039</td>
<td>0.9870</td>
<td>0.9702</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFRCNN</td>
<td>0.5560</td>
<td>0.5543</td>
<td>0.7633</td>
<td>0.9928</td>
<td>0.9650</td>
<td>0.7942</td>
</tr>
<tr>
<td>UA</td>
<td>0.7518</td>
<td>0.7726</td>
<td>0.9945</td>
<td>0.9968</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Change detection results obtained by applying different methods to the second dataset. (a) SAE. (b) CNN. (c) LSTM. (d) MFRCNN. (e) SFRCNN. (The red stands for Change A, the yellow for Change B, the blue for Change C, and the black for the Unchanged.)
two points in our follow-up work. One is to study how our models affect deep feature extraction, and the other is to improve our models to directly get the precise change area.

Acknowledgments

The authors would like to acknowledge the funding from the LIESMARS Special Research Funding. The authors would like to acknowledge the funding from the Fundamental Research Funds for the Central Universities. The authors would also like to thank the developers in the CDTStudio, GDAL, QGIS, Keras and Theano developer communities for their open source projects.

Funding

This work is supported in part by the funding from the LIESMARS Special Research Funding and in part by the Fundamental Research Funds for the Central Universities.

ORCID

Guanzhou Chen http://orcid.org/0000-0003-0733-9122

References


