

Article Urban Building Change Detection in SAR Images Using Combined Differential Image and Residual U-Net Network

Lu Li ^{1,2}, Chao Wang ^{1,2}, Hong Zhang ^{1,*}, Bo Zhang ¹ and Fan Wu ¹

- ¹ Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China; lilushhf@sina.com (L.L.);
- wangchao@radi.ac.cn (C.W.); zhangbo@radi.ac.cn (B.Z.); wufan@radi.ac.cn (F.W.)
 ² University of Chinese Academy of Sciences Beijing 100049 China
- ² University of Chinese Academy of Sciences, Beijing 100049, China
 * Correspondence: zhanghong@radi.ac.cn; Tel.: +86-010-8217-8186

Received: 10 April 2019; Accepted: 4 May 2019; Published: 7 May 2019



Abstract: With the rapid development of urbanization in China, monitoring urban changes is of great significance to city management, urban planning, and cadastral map updating. Spaceborne synthetic aperture radar (SAR) sensors can capture a large area of radar images quickly with fine spatiotemporal resolution and are not affected by weather conditions, making multi-temporal SAR images suitable for change detection. In this paper, a new urban building change detection method based on an improved difference image and residual U-Net network is proposed. In order to overcome the intensity compression problem of the traditional log-ratio method, the spatial distance and intensity similarity are combined to generate a weighting function to obtain a weighted difference image. By fusing the weighted difference image and the bitemporal original images, the three-channel color difference image is generated for building change detection. Due to the complexity of urban environments and the small scale of building changes, the residual U-Net network is used instead of fixed statistical models and the construction and classifier of the network are modified to distinguish between different building changes. Three scenes of Sentinel-1 interferometric wide swath data are used to validate the proposed method. The experimental results and comparative analysis show that our proposed method is effective for urban building change detection and is superior to the original U-Net and SVM method.

Keywords: weighted function; color difference image; urban building change detection; synthetic aperture radar (SAR); residual U-Net

1. Introduction

Urban change detection is an essential remote sensing application that analyzes two or more remote sensing images that have been acquired over the same geographical area at different times to find changes that may have occurred between their acquisition dates [1]. According to the latest report provided by United Nations World Urbanization Prospects in 2014, 54% of the world's population resides in urban areas and this percentage is expected to increase to 66% by 2050 [2]. Building change is one of the most obvious landcover changes in the urbanization process; the evaluation of changes in buildings and reliable urban change information acquisition have become an urgent need in government management, economic construction, sociological research, and so on.

Recently, with the development of synthetic aperture radar (SAR) technology and short repeat-pass cycle, it has gradually become an effective approach for change detection. At present, many methods have been proposed for SAR image change detection [3–7]. As mentioned in reference [3], the procedure of change detection in SAR images can be divided into three steps: (i) Image processing; (ii) difference image (DI) generation; (iii) analysis of DI. Among these steps, difference image generation was a key



step for SAR change detection, and the simplest method was to play subtraction operation on the two SAR images. However, this method led to many fake changed pixels due to the speckle noise and then increased the false alarm rate in the following detection procedure. As references [8–10] mentioned, log-ratio operation was a better DI generation method due to its ability to transform multiplicative noise into additive noise and the fact that it is less sensitive to radiometric errors. But the disadvantage was that log-ratio operation could compress the changed pixels in high-level intensity [11], especially for building objects. To improve the description ability and noise robustness of DI, many neighbor-based and combination methods were proposed. Most acquired improved DI were mean-ratio DI [8], neighbor-based ratio (NR) DI [12], and improved neighborhood-based ratio DI [13]. By using the local spatial information and mean operation, these DIs improved the change detection accuracy. However, the common disadvantage of them was that the optimal window size of the neighborhood was difficult to determine, since there was no reference map or prior knowledge about the image. To solve this problem, Zhuang et al. [14] employed heterogeneity to adaptively select the spatial homogeneity neighborhood and used the temporal adaptive strategy to determine multi-temporal neighborhood windows. This way, the new DI could both suppress the negative influence of noise and preserve edge details. Meanwhile, Zhang et al. [15] proposed a linear weighted function to solve the high-level intensity compression in log-ratio operations. By using the logarithmic function in a dark changed area and saliency extraction in a bright changed area, the changed area could be well described. Bovolo et al. [1] combined multiscale image information to preserve change details by using wavelet decomposition under log-ratio images. Therefore, it can be concluded from the above-mentioned improvement methods that combining different DIs or using neighborhood information could help to improve change detection performance.

Change detection methods based on DI are mainly divided into three categories: (i) Threshold-based methods; (ii) unsupervised clustering methods; (iii) supervised learning methods. For thresholding methods, statistical model design and the thresholding estimation were important. K&I, as a minimum error thresholding method, has been used in change detection. For example, Bazi et al. [16] proposed an automatic change detection approach based on the generalized Gaussian model and a modified K&I criterion. Ghanbari et al. [17] proposed a statistical approach to obtain an image by measuring the similarity of two covariance matrices of the bitemporal polarimetric SAR and then applied the generalized K&I to the obtained image for change detection. The latter two methods are the most popular in current research [18–22]. Their core idea consists of two parts, namely, feature extraction and pixel-wise classification. Gao et al. [7] utilized semi-nonnegative matrix factorization of the SAR images to generate pixelwise feature vectors. Then, they detected the changed and unchanged regions based on a two-layer singular value decomposition (SVD) classifier. Celik et al. [22] used principal component analysis to get each pixel's features and then classified them using k-means clustering. Gong et al. [23] proposed an MRF-FCM unsupervised clustering method to improve the classification accuracy. By enhancing the spatial correlation with Markov Random Field (MRF), the consistency of detection results was improved. However, due to the limitation of useful information, clustering methods could only achieve binary change detection (change or unchanged detection).

In recent years, deep learning networks have been introduced into SAR change detection. Gong et al. [24] used LR difference image and both fuzzy C-means (FCM) and convolutional neural network (CNN) to achieve ternary change detection (positive change, negative change, and unchanged). The deep believe network (DBN) [25] and stacked autoencoder (SAE) [26,27] were used as feature extraction and speckle denoising in SAR change detection. However, the methods based on CNN or DBN were time-consuming. In 2014, full convolution network (FCN) [28] was successfully applied to optical image end-to-end semantics segmentation. The FCN could extract more hierarchical features at different convolution layers due to its skip connection operation, which could combine low-level details features (like line, edge, and texture features) with high-level semantic features (class or context information). In addition, this network was more effective than the traditional CNN or DBN network in remote sensing classification [29–31] or change detection [32,33]. Its up-sampling layer could output

a mapping result which had the same size as the input data. In 2015, inspired by the idea of FCN, Ronneberger et al. [34] proposed U-Net, which was proved to have a better semantic segmentation performance than FCN [30,35,36]. Once the network was trained well, an image patch could be put at the input layer with a classified or segmented map appearing at the output layer immediately. The U-shape construction and the concatenation operator enabled U-Net to combine the low-level details provided by the bitemporal SAR images and high-level semantic information provided by the difference image. However, the drawback of U-Net was that it could only predict on one scale and could not deal with multi-scale segmentation problems well. Thus, Zhang [37] introduced residual block into U-Net and proposed residual U-Net to extract road in optical remote sensing images. By using identity mapping of the residual block, the next convolutional unit obtained information propagation without degradation.

The above-mentioned deep learning methods or networks improved the accuracy of classification or change detection, but if we still used a common DI (as a subtraction image or log-ratio image), it was still impossible to detect changes in specific objects or land cover. The main reason for this was that those DIs did not provide the class information of the objects. Therefore, the goal of this paper was to generate a new DI which included an object's class information, which not only better distinguishes changed and unchanged areas, but also divides the urban building change into positive and negative changes. By introducing spatial and intensity information into DIs, we propose a new combination method based on the idea of NR difference image [12] and weighting function in reference [14]. Considering the complexity of urban environment and the small scale of building changes, the construction and classifier of the residual U-Net network are modified to distinguish between different building changes. The following content is organized as follows: Section 2 gives a detailed presentation of the proposed method. The study area and the test site are described in Section 3. Then, experiments and a detailed analysis are given in Section 4. Lastly, the discussion and conclusions are presented in Sections 5 and 6, respectively.

2. Methodology

Built-up areas in SAR images appear heterogeneous, with alternating brightness and darkness due to the double bounce reflection of buildings, the shadow effect, and multiple reflections [38,39]. As far as building change detection is concerned, it is necessary to add more information to DI and then apply image enhancement to enlarge differences between different building changes (e.g., water to buildings, land to buildings, buildings to land, buildings to water, and so on).

Based on the above analysis, the proposed urban building change detection method mainly includes the following two parts: (i) Difference map generation based on multiple DI fusion and (ii) change semantics segmentation with the residual U-Net network. A brief flowchart of the proposed method is given below in Figure 1. First, all of the SAR images were preprocessed by radiometric calibration, co-registration, enhanced frost filtering, and geocoding using ENVI SARscape software. Then, a neighborhood-constraint (NC) difference image was proposed by merging the traditional subtraction-based (SD) difference image and neighbor-based log-ratio difference image (NR) under a proposed weighted function (see Section 2.1). Next, a three-channel color difference image, different change classes could be identified easily by their colors. After that, a residual U-Net was constructed, and the color difference image was sent to the input layer to train the network under supervision. Finally, after several iterative trainings, a change map was obtained at the output layer of the residual U-Net.



Figure 1. The flowchart of the proposed urban building change detection method. The areas enclosed by the red-dashed lines are the innovations of this paper.

2.1. Neighborhood-Constraint-Based Difference Image Generation

results

Considering that the subtraction-based difference image (SD) preserves high-order change and NR reduces the influence of speckle, we merged the two difference images to generate a new difference image. The new difference image not only had a good description in the high-order change area, but also had a strong robustness to speckle noise. In order to preserve the texture and spatial structure of the object in the new difference image, each pixel in the two images was merged with its adjacent pixels under the constraints of Euclidean distance and the local intensity value. The mathematic expression of the method is as follows:

Suppose that p_i is a pixel in the NC, and that its m^2 neighbor (the length of square neighborhood) pixels are $p_i = \{p_{i,1}, p_{i,2}, \dots, p_{i,m^2}\}$; then, the Euclidean distance of each neighbor pixel is expressed by Equation (1):

$$dis_{i,j} = \sqrt{\left(p^{x}_{i,j} - p^{x}_{i}\right)^{2} + \left(p^{y}_{i,j} - p^{y}_{i}\right)^{2}}, j = 1, 2, \dots, m^{2}$$
(1)

Then, the distance constraint in p_i is presented by a weight matrix W_{dis}^i :

$$W_{dis}^{i} = \left\{ \exp^{\frac{-d_{i,1}}{2\sigma^{2}}}, \exp^{\frac{-d_{i,2}}{2\sigma^{2}}}, \dots, \exp^{\frac{-d_{i,m^{2}}}{2\sigma^{2}}} \right\}$$
(2)

In Equation (2), the adjacent pixels far away from the center pixel p_i have little influence on the results. Here, the exponential form is inspired by the expression of the Gaussian-weighted Euclidian

distance and the value of σ is 1 for the normal Gauss distribution. Then, the intensity constraint is expressed by the intensity similarity, which consists of Equations (3) and (4):

$$S_{i,j} = \sqrt{(I_{i,j}^{SD} - I_i^{SD})^2},$$
 (3)

$$W_{int}^{i} = \left\{ \exp^{\frac{-s_{i,1}}{2\sigma^{2}}}, \exp^{\frac{-s_{i,2}}{2\sigma^{2}}}, \dots, \exp^{\frac{-s_{i,m^{2}}}{2\sigma^{2}}} \right\}.$$
 (4)

In Equation (3), $I_{i,j}^{SD}$ and I_i^{SD} are the intensity values of $p_{i,j}$ and p_i in the SD, respectively. SD is used for the similarity constraint because the difference value in SD can describe the change degree in any intensity level without intensity compression. In Equation (4), speckle noise that has a small similarity with the center pixel p_i will have a small impact on the result, but pixels, which are similar to p_i , will bring a greater contribution to the result. Thus, W_{int}^i can be robust to speckle noise and give a good description at any intensity level. So, W_{dis}^i and W_{int}^i were both computed from the SD in a neighborhood region.

Then, in the new generated difference image NC, the intensity of p_i was computed by weighting the corresponding neighbor pixels' intensity values of NR with the above two weights, as Equation (5) depicts:

$$NC^{i} = \left\langle W_{dis}^{i} \odot W_{int}^{i}, I_{i}^{NR} \right\rangle$$
(5)

Here, the NC^i is a pixel in NC which has the same position as p_i . \odot means the Hadamard product of two matrices and $\langle \rangle$ means the dot product of two vectorized matrices. The variable I_i^{NR} is a matrix formed by neighborhood pixels and its size was the same as the dimensions of the Hadamard product results. Lastly, to keep pixels of NC in a consistent range, a normalizing processing by the sum of the weights was used to the output of Equation (5); the mathematical expression is $NC^i = NC^i / \sum (W_{dis}^i \odot W_{int}^i)$. In order to observe the results of the method intuitively, we also give an example of above SD, NR, and NC in the following Section 4.2.

2.2. Three-Channel Difference Image Generation

Although the NC difference image can give a better description of the urban building changed area, it is still hard to use for identification of the building change from other changes, which also have bright or dark intensity values. As we know, the original SAR image has distinctive information for different land covers. Thus, the original image information should be introduced and combined with NC difference images to generate a new three-channel color DI, named TC (triple-channel color DI). In TC, both the original class information and the change information can be acquired simultaneously. The specific operational approach is to stack the three images (bitemporal original images and a NC image) layer by layer.

It should be noted that the stacking order of each layer is not strictly restricted. Here, we used the post-temporal image as the first layer, the pre-temporal image as the second layer, and the NC difference image as the last layer. Based on this stacking order, one can easily distinguish the different changes of buildings visually according to the RGB color distribution. Section 4.2 shows an example of the generated three-channel color DI.

2.3. Residual U-Net Construction

The residual U-Net combines the strengths of residual learning and U-Net. The skip connections within the residual units and between the encoding and decoding paths of the network facilitate information propagations both in forward and backward computations [37]. Considering that building change areas are multi-scale in the SAR image, we employed the residual U-Net to achieve urban building change detection. However, the original residual U-Net is not suitable to directly use on building change detection. Because the building targets usually occupy a small proportion of the whole

SAR image, if image patches are sent to the deep residual U-Net directly, position and detail information of the small building target will be blurred or even lost significantly after several convolution and pooling operation [37]. In addition, the original network can only receive a grayscale image and output a binary segmentation image. So, it cannot be directly used for building change-type identification.

To solve this problem, we modified the original residual U network, which is shown in Figure 2. In this figure, the network consists of one input layer, five convolutional blocks, and one output layer. Four major layer types (convolutional layer, batch normalization layer, ReLU layer, and addition layer) are displayed in different colors. The image down-sampling or up-sampling processing is represented by the maxpooling operation or the up-sampling operation. The output information of each convolutional block is noted at the side and the convolutional blocks (1), (2), (4), and (5) have a residual block respectively, as shown in the right box. The components highlighted in red represent the modifications we made. Firstly, the number of channels in the input layer was changed to three and the number of layers of the third convolution block was reduced. The purpose of this was to reduce the loss of position and detail information. Next, the binary cross-entropy loss function was replaced by the category cross-entropy loss function. This loss function is always used when the number of samples in the difference classes is unbalanced. Generally, the number of positive building changes was not equal to that of negative building change and the unchanged. Therefore, the use of the category cross-entropy loss function in our network would be more suitable for the actual situation. Finally, because the results are divided into three classes, the sigmoid classifier of raw network was replaced by softmax for multiple classification.



Figure 2. The architecture of the revised residual U-Net network for building change detection. Dashed lines in blue represent the residual block plotted in the right blue box. The abbreviation Conv means convolution, BN means batch normalization, and ReLU means an activation function with the rectified linear unit.

3. Study Areas and Data

3.1. Nanjing City

The study area of this paper is located in Nanjing City, in the southwest part of Jiangsu Province, China, as shown in Figure 3a. Nanjing city is a fast-developing provincial city which plans to expand the building construction area to 652 km² by 2020, according to the statistical data provided by Nanjing government [40] in 2017. Seen from this blueprint, Nanjing City mainly consists of three parts: (i) Central district in red; (ii) metropolitan district in orange; (iii) city proper in green. From 2017 to the present, the construction of residential and commercial buildings are the main activities in the Pukou and Jiangning Districts and Chunxi Town. Pukou and Jiangning Districts are main developing living areas of the Central district, which mainly plans to construct office and residential buildings. Chunxi Town, as a new planned town in the south of Nanjing, will develop fast in the following several years. Therefore, we tested the proposed method by detecting the construction of new buildings and the changes in building removal in these areas.



Figure 3. The study area: **(a)** The administrative map of Nanjing City, Jiangsu Province, China. **(b)** The Sentinel-1 SAR image. The blue contour area was used as the test site and the red outline areas were used as the training samples.

3.2. Experimental Data

Sentinel-1 satellite produced C-band dual-polarization of the SAR data and its interferometric wide swath (IW) mode provided an SAR image with 250 km coverage range and 10 m sampling space. We chose Sentinel-1 IW images in 2017 and 2018 to perform the experiments. Table 1 lists the basic information of the images in the following experiments.

No.	Date	Satellite	Polarization	Data Type	Image Size
s1	2017/3/16	S1A	VH	GRD	$18,434 \times 11,991$
s2	2018/1/22	S1A	VH	GRD	$18,436 \times 11,992$
s3	2018/12/24	S1A	VH	GRD	$18,433 \times 11,991$

Table 1. Basic information on the experimental data.

As shown in Figure 3b, Nanjing City was selected as the test data, while the other three cities, Changzhou City, Wuxi City, and Yixing City, were selected for the training data. As the major cities of

Jiangsu Province, these cities have developed fast in recent years and their natural environments are similar to that of Nanjing. The specific information on the training and test data are listed in Table 2. In Table 2, 'positive pixels' means the number of pixels where buildings were constructed and 'negative pixels' indicates the number of pixels where the buildings were removed. The 'number of patches' counts the number and the shape (length and width) of the image patches. These patches were used as input data of the residual U-Net network. Since the proposed TC was a three-channel color image, each patch in the input layer was $224 \times 224 \times 3$.

Sample Type	City	Positive Pixels	Negative Pixels	Acres (Positive/Negative)	Number of Patches
Training	Changzhou Yixing	30,550 5500	7310 3150	2152/511 385/220	1288 416
	Wuxi	27,047	4939	1893/346	545
Test samples	Nanjing	28,711	8892	2009/622	1656

Table 2. Specific information on the training and test samples.

In addition, the ground truth of both training and test data were made by manual interpretation of optical images with similar imaging times of the SAR images on Google Earth. Figure 4 shows an example of building changes using optical and SAR images in Jiangning District. Figure 4a,b show the optical images of 6 March 2017 and 8 February 2018, respectively. Figure 4c,d are SAR VH images of 16 March 2017 and 22 January 2018, respectively. The green box marked A1 indicates the site where some buildings were demolished in 2018. This change represents a negative building change. Another three red boxes labeled A2–A4 shows the location of new buildings in 2018, which represent positive building changes.



Figure 4. An example of positive and negative building changes using Google Earth optical and SAR images in Jiangning District.

4. Experimental Results

4.1. Parameter Setting

In the proposed difference image NC, the neighbor size *m* and σ are the two main parameters that would influence the final TC difference image's quality. The value *m* determines the number of neighbor pixels participating in the calculation of the spatial weight matrix W_{dis}^i and the intensity of the similar degree weight W_{int}^i . As we know, the large neighbor window leads to more pixels, resulting in an ambiguity border of the change region, while the small neighbor window is less effective on reducing the influence of speckle noise. Generally, the 3 × 3, 5 × 5, and 7 × 7 neighbor windows perform well and there is little difference between 3 × 3 and 5 × 5. Here, we selected 3 × 3 as the

preset parameter value for building change detection. σ is a variance which influenced the range of the intensity level. The small (or large) value of σ results in a small (or large) intensity level range, which worsens the texture or intensity contrast. Thus, we followed the normal distribution with $\sigma = 1$ in our method.

As for the revised residual U-Net, the main parameters used in our experiment are summarized as follows: The training environment of the network is Ubuntu 16.0.4 with Core i7 CPU and NVIDIA GTX 1080Ti 12G GPU. The total number of samples in Table 2 is 2249. For each sample in Table 2, we obtained another nine new augmented samples by performing data augmentation (rotation with random angle, flip, wrap). Thus, we had 22,490 samples in the final training dataset. The learning rate, epoch, and batch size were 0.0005, 30, and 5, respectively.

Since the input of the network was $224 \times 224 \times 3$, we clipped the whole images into patches of the same size to input data format. In the training stage, we used a non-overlapping 224×224 sliding window to segment the whole training image into patches. The remaining boundary regions less than 224×224 were discarded. This method had no adverse effect on the training process. In the test stage, we divided the image into patches by half overlapping, which avoided changed areas to be separated into different patches. The results of the two adjacent image patches were processed by joint operation to ensure that the changed area and its true boundary remained unchanged as much as possible.

4.2. Result of the Proposed Difference Images

Figure 5 is an example of the four types of difference images. Seen from Figure 5, in SD and NR maps, the intensity contrast between the changing and unchanged area was low so that some of the changed buildings in the figure are blurred, which makes it difficult to distinguish building changes from the unchanged background. In order to see the difference among the four difference images clearly, we list two examples in the right two columns. By comparing the four difference image patches, we find that although NR had less speckle noise, it had a worse description on the changed areas, shown in Figure 5a,e. This is because that the log-ratio operation reduces the detection rate of high-level change areas. In the SD image, some changed pixels were submerged in noise which makes the shape of changed areas still incomplete. However, this drawback was improved in the proposed NC difference image. As shown in Figure 5c,g, both positive and negative changes were better described than the first two DIs, especially when the edges were well detected. However, in NC, it was still difficult to distinguish the different building change classes (unchanged, positive, and negative changes) from DI. Fortunately, this problem has been well addressed in the proposed TC, as shown in Figure 5d,h. Red areas represent positive building changes, while light green areas represent negative building changes. Compared with the previous three DIs, the last one more easily distinguished between building change areas and unchanged backgrounds.



Figure 5. The illustration of the four difference images by using bitemporal SAR image patches in Jiangning District. The two right columns show details of these difference images; (**e**–**f**) correspond to (**a**–**d**).

4.3. Validation of the Proposed Difference Images

To further verify the effectiveness of the proposed difference images NC and TC, we performed two validation experiments in this section. Firstly, to test if the training samples extracted from TC can separate positive and negative changes from other unchanged colored backgrounds, we introduced the Jeffries Matusita (JM) distance to measure the separability among the three classes. The JM distance is a distinguishing indicator that widely used in the remote sensing field [41–43], and its value ranges from 0 to 2. The larger the value is, the better separability is among the different categories. Its mathematical equation can be expressed by Equation (6):

$$JM(C_i, C_j) = \int \left[\int \left[\sqrt{p\left(\frac{x}{C_i}\right)} - \sqrt{p\left(\frac{x}{C_j}\right)} \right] \right]^2 dx$$
(6)

where $p\left(\frac{x}{C_i}\right)$ represents the probability that the pixel *i* belongs to the C_i class. Here, the probability density function (PDF) of the random selected pixels obeys normal distribution. In order to get reliable estimation results quickly, we obtained the PDFs of classes by using the commercial software ENVI. Then the JM separation degree was calculated by the following equations:

$$J = 2(1 - \exp(-B))$$
(7)

$$B = \frac{1}{8}(m_1 - m_2)^2 \frac{2}{\sigma_1^2 + \sigma_2^2} + \frac{1}{2} \ln \left[\frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1 \sigma_2} \right]$$
(8)

Here, m_1 and m_2 are mean values of the two classes and σ_1 and σ_2 are variance values of the two classes. By using the following two steps, the JM distance of any two classes can be computed.

In Table 3, both the separability value and the number of tested pixels are listed and the numbers of pixels are under the JM value. Seen from the table, the separative values between the two changed classes are almost at the maximum, so it is easy to distinguish between them. The JM value of positive change and the unchanged part is 1.8367, which indicates that the two classes are also easy to separate. The JM value between the negative change and the unchanged part is slightly lower than the last, 1.7383, which is because the same volume of scattering objects, such as trees and crops, have similar intensity changes as the negative building change. Although it seems that they are not easy to separate, the texture information between them is significantly different and the information can be obtained by the deep convolutional network. Therefore, these classes can be well separated by the proposed TC and the residual U-Net network.

Table 3. The JM separability among two changed classes and the unchanged class.

	Positive Change	Negative Change	Unchanged
Positive Change	_	_	_
Negative Change	1.9937 (1049, 3758)	—	_
Unchanged Part	1.8367 (4078, 3758)	1.7383 (4078, 1094)	_

To verify that the proposed NC has a better description on changed and unchanged areas, we used the OTSU [44] threshold method to detect changed and unchanged pixels in these difference images. Then, we used ROC to evaluate the performance of them according to the measurement used in Inglada's paper [8]. The statistical indices are false alarm rate (FAR) and detection rate (DR) and are used to make the ROC curve line. The definitions of FAR and DR are shown as

$$DR = PC/AC \tag{9}$$

$$FAR = (FU + FC)/(AC + AU)$$
(10)

Here, positive changed (PC) pixels means the number of correctly detected changed pixels in the difference image. All changed (AC) pixels means the total number of true changed pixels in the difference image. All unchanged (AU) pixels is the total number of true unchanged pixels in the difference image. False unchanged (FU) pixels is the number of incorrectly detected unchanged pixels; these are actually changed pixels. False changed (FC) pixels is contrary to FU. Using this method, a DI will have zero DR and FAR values when the detection threshold is 1 because all pixels are detected as unchanged. When the detection threshold is set to 0 all changed pixels are detected, but plenty of false changed pixels are also detected, so the FAR goes to 1. For performance evaluation, the larger coverage area of ROC curve under the first quadrant of coordinate axis, the better the description performance of this method.

The statistical ROC result is shown in Figure 6. It can be seen that our proposed NC outperformed the other two difference images. NR was better than SD, although it had a low detection rate when the threshold was large. This is because some high-level changed pixels are compressed by the log-ratio operation, so their intensity values are less than the threshold. However, with the threshold reduction, the false alarm rate was lower than SD when it had the same detection rate as SD.



Figure 6. ROC comparison of the three difference images under detection rate (DR) and false alarm rate (FAR) indices by using the OTSU thresholding segmentation method.

4.4. Results of Urban Building Change in Nanjing City

In this section, Nanjing city was taken as the test site to validate the effectiveness of the proposed method in two groups of experiments. Pukou District and Chunxi Town were selected as the first test sites and Jiangning District, shown in Figure 4, was selected as another test site to compare the performance of our method with U-Net and SVM results. The confusion matrix, overall accuracy (OA), and the Kappa coefficient were used to illustrate the accuracy of the building change detection results. All accuracy evaluation results were calculated by counting pixels in the detected results and the corresponding ground truth images. Since the number of unchanged pixels was much higher than that of the positive and negative changes, it was better to have a comparable number of control pixels for positive, negative, and unchanged pixels. In the following accuracy assessment, we randomly selected the same number of pixels for each class.

The building change results at Pukou District and Chunxi Town from 16 March 2017 to 22 January 2018 are shown in Figure 7. In the figure, the first two rows display the two pairs of bitemporal SAR images, detection results, and corresponding reference images. We also have six main change examples in the next six columns and show the specific change area by drawing the outline in each patch. Since the outlines were drawn by visual interpretation, they are just used to help readers to quickly find the location of the changed area in the image. For the first result (shown in the first column), many positive building changes occurred and few small-scale negative building changes happened. The detection results of our method roughly extracted the main changed areas and their change classes. For the second result (shown in the second column), the proposed method also detected main building changes well. However, there were also some omission errors on the result map compared with its reference image and most of the errors came from the negative change. The negative change in this area was mainly caused by small factory buildings or residential buildings, which were easily ignored due to the limitation of the image resolution. In fact, there were many small-scale building changes such as this in Nanjing City, which could have gone undetected on the result map; a performance evaluation would be influenced by these omission errors. Therefore, in order to eliminate the influence of resolution, we set a scale threshold to remove the tiny pieces belonging to the changed pixels. The threshold was set to 25 pixels, namely when a changed area was less than 25 pixels or the region is smaller than 5×5 , it would be removed and regarded as speckle noise. In addition, these tiny pieces in the reference image were also ignored, so we only focused on verifying the detection performance of the selected change area.



Figure 7. The building change results at Pukou District and Chunxi Town from 16 March 2017 to 22 January 2018. Pixels in red indicate newly constructed buildings and pixels in blue indicate removed buildings.

Table 4 shows the accuracy evaluation of change detection results in Pukou District and Chunxi Town by using the confusion matrices. In this test, we randomly selected 550 pixels for each class in the image patches (both results and references shown in Figure 7). According to the confusion matrices, we found that most of the building changes were correctly detected in both areas. There was no error in pixel classification between positive and negative changes. The OA values of the two sites were about 0.8518 and 0.83, respectively. According to the Kappa value, the first site had a better detection result than the second one. However, compared with the result of the second test site, it had a lower omission error. The high omission error in the Chunxi Town result was mainly caused by the small change areas, since this place is a new, developing town where small residential buildings or houses

are constructed. Therefore, we can assume that the proposed method was able to perform better in the large-scale change area with a low commission and omission error.

Prediction								
	Pukou District							
ath	Class	Positive Change	Negative Change	Unchanged	Total			
Ľ	Positive Change	479	0	71	550			
pui	Negative Change	0	458	92	550			
rot	Unchanged	24	40	486	550			
Ċ	Total	Total 503		649	1650			
	Kappa	0.8021		OA	0.8624			
	Chunxi Town							
Ground Truth	Class	Positive Change	Negative Change	Unchanged	Total			
	Positive Change	471	0	79	550			
	Negative Change	0	442	124	550			
	Unchanged	31	34	485	550			
	Total	502	476	672	1650			
	Kappa	0.7933		OA	0.8473			

Table 4. Accuracy assessment of the results in Figure 7.

Figure 8 shows comparison results of our method and the other two approaches in Jiangning District from 16 March 2017 to 22 January 2018. In the figure, the first two columns display the optical images from Google Earth and the third and fourth columns display the corresponding SAR images. The first result (A1–A7) is an example of the negative building change and the next three results show positive building changes where different buildings were built during this period. It is obvious that all compared methods performed well at the large-scale residential building change shown by D1–D7 and the proposed method performed best. This is because the buildings in this area presented a large bright patch by the high-density building distribution in SAR images. It is noted that SVM showed a comparable detection performance at the large-scale building change area, but it also generated many incorrect classification results (A7–D7). The reason is that SVM only used single-pixel information to classify the building change, while the spatial construction or texture information of the buildings were not involved. In contrast, U-Net had a lower commission error than SVM (A6-D6 and A7-D7) though it did not perform well at the border of the building change area. This indicates that the deep neural network could effectively reduce the probability of incorrect classification. Secondly, in comparing the results (A6–D6) of U-Net and the results (A5–D5) of our revised residual U-Net, we found that the latter method had a better detection border than the previous method. This was because the residual block shown in Figure 2 preserved more useful spatial information during multiple convolutional operations and provided more information for the final building change map-generation procedure.

In order to evaluate the three methods accurately, we used the area of Nanjing city in Figure 2 as the test data to detect the urban building changes during the period from 22 January 2018 to 24 December 2018. Since most of the building changes occurred in urban areas, we randomly selected 10,000 pixels for each class in the urban areas of Nanjing City and then used these pixels to evaluate the accuracy. Table 5 lists the overall accuracy evaluation results of all the methods and Table 6 lists the statistical results of the proposed method. Seen from Table 5, the optimal OA of the proposed method was close to 0.87, while SVM had the lowest OA. The differences in OA and Kappa between the three methods were close to 15% and 20%, respectively. Although the OA of U-Net was nearly up to 0.8, its kappa coefficient was only 0.7233, which means that the consistency between the detected result and the reference was low. From the 'commission' and 'omission' results, it can be seen that the detection error of SVM was more than 37%, mainly due to the poor classification ability of the linear classifier. When determining the class type of a pixel, it only used the vector information of a single

pixel, ignoring the important spatial neighborhood information. Compared with SVM, U-Net had better OA and Kappa values, but still had poor performance in negative building change detection. This was because negative changes always accounted for a very small proportion in the image and, because of the lack of residual blocks, it was difficult to obtain the negative change information in deeper layers.



Figure 8. Detailed results of different methods at the test site in Jiangning District. All positive change results and ground truths are masked in red and the negative results are masked in blue. The eight columns respectively show Google Earth images from 16 March 2017 (A1–D1) to 22 January 2018 (A2–D2), SAR images (A3–D3 and A4–D4), detection results of our method (A5–D5), U-Net (A6–D6), SVM (A7–D7), and ground truth maps (A8–D8).

Class	SVM		U-Net		Residual U-Net	
	Commission (%)	Omission (%)	Commission (%)	Omission (%)	Commission (%)	Omission (%)
Positive Change	25.44	16.60	14.65	22.84	5.49	9.74
Negative Change	37.30	42.19	11.75	23.70	7.24	12.20
Unchanged	29.13	16.23	35.33	14.80	19.97	12.10
OA	0.7041		0.7955		0.8865	
Kappa	0.6011		0.7233		0.8298	

Table 5. Accuracy assessment of the urban building change detection results.

In Table 6, about 974 positive changed and 1220 negative changed pixels were erroneously detected as unchanged types. This was because many small-scale building changes scattered in urban areas usually looked like small blocks in the image. These pixels could easily be classified as unchanged. In January 2018, some rough bare land areas had a near high backscatter coefficient, but dropped to a lower level in December. Therefore, some pseudo-negative changed pixels were detected. Additionally, we can observe three other findings from the confusion matrix in Table 6: (i) The omission error of positive change was caused by the misclassification of unchanged pixels; (ii) the commission error of unchanged class was caused by the misclassification of negative changed pixels; (iii) the number of false detected pixels between positive and negative changes is zero.

Prediction						
ith	Class	Positive Change	Negative Change	Unchanged	Total	
ound tru	Positive Change	9026	0	974	10,000	
	Negative Change	0	8780	1220	10,000	
	Unchanged	524	686	8790	10,000	
Gr	Total	9350	9166	11,484	30,000	

Table 6. Confusion matrix of the detection results of the proposed method in Nanjing City.

5. Discussion

From the experimental results in Sections 4.2–4.4, it can be seen that the proposed difference image generation method not only had a better description of building change areas than the traditional difference images, but also better distinguished the building changes from other land object changes. This was mainly due to the following reasons: (i) The proposed NC effectively enhanced the contrast between the homogeneous region (changes happened) and the heterogeneous region (noise dominated). Meanwhile, it had a denoising effect on the foreground target and better preserved the details of the building edge. The advantage of NC lies mainly in the proposed neighbor-based weighting function in Section 2.1. The spatial constraint of the function played an important role in contrast enhancement, while the neighbor intensity constraint preserved the texture information and the difference degree of the building change area. (ii) The proposed three-channel color difference image contained the original SAR image information and building changes information simultaneously. Due to dihedral angle scattering characteristics, buildings had brighter spatial textures and structural intensities which made them easily identifiable. Thus, the proposed approach can distinguish the urban building changes from unchanged areas.

The experimental results show that the method can be further improved in the following two aspects. Firstly, as shown by the experimental results in Figure 7, the performance of this method in small-scale building change detection needs to be further improved. The maxpooling layer used in the network may lose the position information in small changed areas. Thus, it is necessary to design a better pooling layer to preserve the position information of small changed areas. What is more, the proposed NC has a denoising effect on the foreground target (namely changed regions), which only seems to enhance the contrast between unchanged and changed areas. Thus, the weighting function we proposed can be further improved to compress the intensity level of unchanged pixels. Secondly, for the moment, this proposed method cannot be used for multiclass change detection. The proposed differential image can only be used to divide building changes into positive and negative changes from dual-time SAR images. Specific land cover change types cannot be well identified, which could explain what land types or objects have been changed into buildings. To this end, we plan to study the improvement scheme of multi-class classifier for multi-class labeled samples to realize multi-class land cover change detection.

6. Conclusions

In this paper, a neighbor-based color difference image and an improved residual U-Net were used to detect urban building changes in SAR images. Compared with the traditional subtraction or log-ratio difference image, the proposed difference image generated by the newly proposed weighting function better described the building change areas. The stacked color image effectively separated building changes from the unchanged areas. By reducing the depth of the original residual U-Net and employing the softmax classifier and the category cross-entropy loss function, it was apparent that this network was more suitable for urban building change detection. The experimental results in Nanjing City verified the validity and robustness of the proposed method. This method still needs some improvement for application in the small-scale building change areas, which is also the focus of our next work.

Author Contributions: L.L. was mainly responsible for design and implementation of algorithms, conceived the original manuscript, and conducted the experiments. C.W. supervised the experiments. H.Z. helped design the algorithms and also contributed to the organization of the paper and the experimental analysis. B.Z. and F.W. were also responsible for the experiments.

Funding: This research was funded by the National Natural Science Foundation of China (41331176 and 41371352) and the National Key Research and Development Program of China (2016YFB0501501).

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

References

- 1. Bovolo, F.; Bruzzone, L. A detail-preserving scale-driven approach to change detection in multitemporal SAR images. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2963–2972. [CrossRef]
- 2. United Nations. World Urbanization Prospects—The 2014 Revision. *Technical Report.* Available online: http://esa.un.org/unpd/wup/ (accessed on 25 October 2018).
- 3. Hu, Z. An unsupervised change detection approach based on K&I dual thresholds under the generalized gauss model assumption in SAR images. *Acta Geod. Cartogr. Sin.* **2013**, *42*, 116–122.
- 4. Su, L.; Gong, M.; Sun, B.; Jiao, L. Unsupervised change detection in sar images based on locally fitting model and semi-em algorithm. *Int. J. Remote Sens.* **2014**, *35*, 621–650. [CrossRef]
- Lu, J.; Li, J.; Chen, G.; Zhao, L.; Xiong, B.; Kuang, G. Improved pixel-based change detection accuracy using an object-based approach in multitemporal SAR flood images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2015, *8*, 3486–3496. [CrossRef]
- 6. Cui, B.; Zhang, Y.; Yan, L.; Cai, X. A SAR intensity images change detection method based on fusion difference detector and statistical properties. *ISPRS Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *4*, 439–443. [CrossRef]
- 7. Gao, F.; Liu, X.; Dong, J.; Zhong, G.; Jian, M. Change detection in SAR images based on deep semi-NME and SVD networks. *Remote Sens.* **2017**, *9*, 435. [CrossRef]
- 8. Inglada, J.; Mercier, G. A new statistical similarity measure for change detection in multitemporal sar images and its extension to multiscale change analysis. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1432–1445. [CrossRef]
- 9. Dekker, R. Speckle filtering in satellite SAR change detection imagery. *Int. J. Remote Sens.* **1998**, *19*, 1133–1146. [CrossRef]
- 10. Li, Y.; Martinis, S.; Plank, S.; Ludwig, R. An automatic change detection approach for rapid flood mapping in sentinel-1 sar data. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 123–135. [CrossRef]
- Zhang, Y.; Wang, C.; Wang, S.; Zhang, H.; Liu, M. SAR image change detection method based on visual attention. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 3078–3081.
- 12. Argenti, F.; Alparone, L. Speckle removal from SAR images in the undecimated wavelet domain. *IEEE Trans. Geosci. Remote Sens.* **2002**, 40, 2363–2374. [CrossRef]
- 13. Zhuang, H.; Fan, H.; Deng, K.; Yu, Y. An improved neighborhood-based ratio approach for change detection in SAR images. *Eur. J. Remote Sens.* **2018**, *51*, 723–738. [CrossRef]
- 14. Zhuang, H.; Fan, H.; Deng, K.; Yao, G. A spatial-temporal adaptive neighborhood-based ratio approach for change detection in SAR images. *Remote Sens.* **2018**, *10*, 1295. [CrossRef]
- 15. Zhang, Y.; Wang, S.; Wang, C.; Li, J.; Zhang, H. SAR image change detection using saliency extraction and shearlet transform. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4701–4710. [CrossRef]
- Bazi, Y.; Bruzzone, L.; Melgani, F. An unsupervised approach based on the generalized gaussian model to automatic change detection in multitemporal SAR images. *IEEE Trans. Geosci. Remote Sens.* 2005, 43, 874–887. [CrossRef]
- Ghanbari, M.; Akbari, V. Generalized minimum-error thresholding for unsupervised change detection from multilook polarimetric SAR data. In Proceedings of 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 1853–1856.
- 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 Photogramm. Remote Sens.* 2016, 116, 24–41. [CrossRef]

- 19. Liu, W.; Yang, J.; Zhao, J.; Shi, H.; Yang, L. An unsupervised change detection method using time-series of PolSAR images from Radarsat-2 and GaoFen-3. *Sensors* **2018**, *18*, 559.
- 20. Hao, M.; Zhang, H.; Shi, W.; Deng, K. Unsupervised change detection using fuzzy -means and MRF from remotely sensed images. *Remote Sens. Lett.* **2013**, *4*, 1185–1194. [CrossRef]
- Nielsen, A.A.; Vestergaard, J.S. Change detection in bi-temporal data by canonical information analysis. In Proceedings of the 8th International Workshop on the Analysis of Multitemporal Remote Sensing Images (Multi-Temp), Annecy, France, 22–24 July 2015; pp. 1–4.
- 22. Celik, T. Unsupervised change detection in satellite images using principal component analysis and k-means clustering. *IEEE Geosci. Remote Sens.* 2009, *6*, 772–776. [CrossRef]
- 23. Gong, M.; Su, L.; Jia, M.; Chen, W. Fuzzy clustering with a modified mrf energy function for change detection in synthetic aperture radar images. *IEEE Trans. Fuzzy Syst.* **2014**, 22, 98–109. [CrossRef]
- 24. Gong, M.; Yang, H.; Zhang, P. Feature learning and change feature classification based on deep learning for ternary change detection in SAR images. *Isprs Photogramm. Remote Sens.* **2017**, *129*, 212–225. [CrossRef]
- 25. 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]
- 26. Planinšič, P.; Gleich, D. Temporal change detection in SAR images using log cumulants and stacked autoencoder. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 297–301. [CrossRef]
- 27. Chini, M.; Pelich, R.; Hostache, R.; Matgen, P.; Lopez-Martinez, C. Towards a 20 m global building map from Sentinel-1 SAR data. *Remote Sens.* **2018**, *10*, 1833. [CrossRef]
- 28. Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef]
- 29. Wei, S.; Zhang, H.; Wang, C.; Wang, Y.; Xu, L. Multi-temporal SAR data large-scale crop mapping based on U-Net model. *Remote Sens.* **2019**, *11*, 68. [CrossRef]
- 30. Wang, Y.; He, C.; Liu, X.; Liao, M. A hierarchical fully convolutional network integrated with sparse and low-rank subspace representations for PolSAR imagery classification. *Remote Sens.* **2018**, *10*, 342. [CrossRef]
- Wang, Y.; Wang, C.; Zhang, H. Integrating h-a-α with fully convolutional networks for fully PolSAR classification. In Proceedings of the 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP), Shanghai, China, 18–21 May 2018; pp. 1–4.
- 32. Song, A.; Choi, J.; Han, Y.; Kim, Y. Change detection in hyperspectral images using recurrent 3d fully convolutional networks. *Remote Sens.* **2018**, *10*, 1827. [CrossRef]
- 33. Xu, Z.; Robert, W.; Li, N.; Zhang, H.; Zhang, L. A novel approach to change detection in SAR images with CNN classification. *J. Radar.* **2017**, *6*, 483–491.
- Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
- 35. Tang, H.; He, C. SAR image scene classification with fully convolutional network and modified conditional random field-recurrent neural network. *J. Comput. Appl.* **2016**, *36*, 3436–3441.
- Tanase, R.; Datcu, M.; Raducanu, D. A convolutional deep belief network for polarimetric SAR data feature extraction. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 7545–7548.
- Zhang, Z.; Liu, Q.; Wang, Y. Road extraction by deep residual U-Net. *IEEE Geos. Remote Sens. Lett.* 2018, 15, 749–753. [CrossRef]
- 38. Cao, H.; Zhang, H.; Wang, C.; Zhang, B. Operational built-up areas extraction for cities in china using Sentinel-1 SAR data. *Remote Sens.* **2018**, *10*, 874. [CrossRef]
- Han, J.; Deng, K.; Li, B. SAR image change detection methods based on glcm texture features. J. Geod. Geodyn. 2012, 32, 94–98.
- 40. Nanjing Bureau of Planning and natural Resources. Available online: http://ghj.nanjing.gov.cn/ztzl/ghbz/ztgh/ (accessed on 2 May 2019).
- 41. Dabboor, M.; Howell, S.; Shokr, M.; Yackel, J. The Jeffries–Matusita distance for the case of complex Wishart distribution as a separability criterion for fully polarimetric SAR data. *Int. J. Remote Sens.* **2014**, *35*, 6859–6873.
- 42. Wang, Y.; Qi, Q.; Liu, Y. Unsupervised segmentation evaluation using area-weighted variance and jeffries-matusita distance for remote sensing images. *Remote Sens.* **2018**, *10*, 1193. [CrossRef]

- 43. De Grandi, E.; Mitchard, E.; Hoekman, D. Wavelet based analysis of TanDEM-X and LiDAR DEMs across a tropical vegetation heterogeneity gradient driven by fire disturbance in indonesia. *Remote Sens.* **2016**, *8*, 641. [CrossRef]
- 44. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]



© 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/).