



# Article A Novel End-to-End Unsupervised Change Detection Method with Self-Adaptive Superpixel Segmentation for SAR Images

Linxia Ji <sup>1</sup>, Jinqi Zhao <sup>2</sup>,\* and Zheng Zhao <sup>1</sup>

- <sup>1</sup> Institute of Photogrammetry and Remote Sensing, Chinese Academy of Surveying & Mapping (CASM), Beijing 100830, China; jilinxia0825@gmail.com (L.J.); zhaozheng@casm.ac.cn (Z.Z.)
- <sup>2</sup> The School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China
- \* Correspondence: masurq@cumt.edu.cn; Tel.: +86-13212715319

Abstract: Change detection (CD) methods using synthetic aperture radar (SAR) data have received significant attention in the field of remote sensing Earth observation, which mainly involves knowledge-driven and data-driven approaches. Knowledge-driven CD methods are based on the physical theoretical models with strong interpretability, but they lack the robust features of being deeply mined. In contrast, data-driven CD methods can extract deep features, but require abundant training samples, which are difficult to obtain for SAR data. To address these limitations, an end-to-end unsupervised CD network based on self-adaptive superpixel segmentation is proposed. Firstly, reliable training samples were selected using an unsupervised pre-task. Then, the superpixel generation and Siamese CD network were integrated into the unified framework to train them end-to-end until the global optimal parameters were obtained. Moreover, the backpropagation of the joint loss function promoted the adaptive adjustment of the superpixel. Finally, the binary change map was obtained. Several public SAR CD datasets were used to verify the effectiveness of the proposed method. The transfer learning experiment was implemented to further explore the ability to detect the changes and generalization performance of our network. The experimental results demonstrate that our proposed method achieved the most competitive results, outperforming seven other advanced deep-learning-based CD methods. Specifically, our method achieved the highest accuracy in OA, F1-score, and Kappa, and also showed superiority in suppressing speckle noise, refining change boundaries, and improving detection accuracy in a small area change.

**Keywords:** change detection; deep learning; synthetic aperture radar (SAR); superpixel segmentation; Siamese network; transfer learning

# 1. Introduction

Natural and human activities have a continuous impact on Earth's resources and environment. The accurate detection of changes is of great significance in resource and environmental protection [1], agricultural survey [2], urban renewal [3,4], forest resource management [5], and other applications of Earth observation. Remote sensing Earth observation has the advantages of large-scale and periodic observations. By using image processing and pattern recognition techniques, change information can be identified from multi-temporal remote sensing data.

The synthetic aperture radar (SAR) is an advanced active remote sensing technology characterized by its penetration ability, all-weather and all-time work, wide coverage, and other advantages. Therefore, SAR images provide crucial data support for acquiring ground information in harsh environments and are widely used in remote sensing change detection (RSCD).

To the best of our knowledge, the change detection (CD) methods that use multitemporal SAR images can be divided into two categories: traditional knowledge-driven



Citation: Ji, L.; Zhao, J.; Zhao, Z. A Novel End-to-End Unsupervised Change Detection Method with Self-Adaptive Superpixel Segmentation for SAR Images. *Remote Sens.* 2023, *15*, 1724. https://doi.org/10.3390/rs15071724

Academic Editor: Dusan Gleich

Received: 3 February 2023 Revised: 14 March 2023 Accepted: 20 March 2023 Published: 23 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). methods and data-driven methods, whereby both of which involve supervised and unsupervised designs. Supervised methods require a large number of labeled samples as prior knowledge, which are difficult to obtain for SAR images. Therefore, our research is focused on designing an automatic and efficient unsupervised CD method for SAR data.

In the early stages of SAR CD method research, the majority of the studies focused on developing pixel-based change detection (PBCD) methods. These methods mainly involve three steps: image preprocessing, difference image (DI) generation, and DI analysis. Specifically, image preprocessing includes speckle noise filtering, radiometric calibration, geometric correction, registration, etc., which generate comparable multi-temporal images with less noise. The traditional knowledge-driven methods mainly focus on DI generation and DI analysis. In the step of DI generation, the most straightforward method is based on image algebra, such as image difference [6] and image ratio [7]. The more complex approaches include methods based on image transformation, such as principal component analysis (PCA) [8,9] and change vector analysis (CVA) [10,11]. Other methods include texture analysis [12], edge-based detection [13,14], machine learning [15,16], GIS analysis [17], and mixed techniques. After the ideal DI is extracted, thresholding [18], clustering [19], or other advanced methods [20,21] are used to analyze DI, and then the binary change map is finally obtained.

However, the PBCD method is sensitive to speckle noise, thus detecting change with false alarms, holes, and jagged boundaries. To solve these problems, object-based change detection (OBCD) methods are proposed. These methods [22–24] segment pixels into image objects and take them as study units, which can smooth holes and improve the boundary detection accuracy. However, the proper setting of scale parameters in OBCD is complex and mechanical, and improper settings may cause important small changes to be missed. Moreover, the performance of OBCD methods depends on the accuracy of the segmentation algorithm. As a compromise, the CD method based on superpixel segmentation [25] has become a popular choice to generate uniform and homogeneous regions with the ability to perceive semantic information.

For the object or superpixel segmentation of multi-temporal images, three strategies are generally used [26,27]: (1) Only one phase image is segmented, and the other image is directly stacked with this segmentation result to perform CD. This strategy will cause missed and false detection. (2) Multi-temporal images are segmented independently, which often produces sliver polygons due to the inconsistency of the segmentation. As a result, it is difficult to perform CD analysis, and segmentation error will propagate to the CD analysis step. (3) The multi-temporal images are segmented simultaneously by stacking them; this manner with low computational efficiency often leads to over-segmentation and boundary fragmentation. Therefore, a better segmentation strategy still needs to be further investigated.

Due to the continuous development of satellite sensors, a large number of accumulated images provides opportunities for data-driven deep learning change detection (DLCD) methods. Deep learning automatically extracts high-level features from images [28], which has been proven to be an effective feature-learning technique [29–31]. The end-to-end DLCD method can directly obtain CD results from multi-temporal images. Moreover, the extracted deep features are robust to speckle noise [32]. To our best knowledge, according to the strategy of fusing multi-temporal information, the DLCD methods involve three types: (1) Early fusion [33], where multi-temporal information is fused before being inputted into the network. In order to increase the information, other useful manual features can be added for different tasks. (2) Multi-temporal information fusion based on the Siamese network [34,35], where the multi-temporal images are input into different branches of the Siamese network, respectively, to learn the correlation or difference in the multi-temporal information. (3) Multi-temporal information fusion based on the recurrent neural network (RNN) [36], where RNN is used to mine the dependence of sequential images acquired at different times. Among these three strategies, the Siamese network has proven to be more specific for CD and has great potential to improve detection accuracy [37,38]. Currently, the DLCD methods generally input the sampling patches (patch-based) or the whole image (image-based) into the network. The patch-based method with low computational efficiency loses a considerable amount of spatial context information. However, image-based fully convolutional network methods, such as FCN [39], Unet [40], DeepLab [41], and their variants, can accept any size of input image (if the computing memory allows), utilizing global context information to generate the dense pixel-wise prediction. These methods are efficient and accurate and have become the mainstream network in DLCD fields [25].

Furthermore, existing DLCD methods mainly take a pixel as the basic analysis unit, which limits DLCD to perceive object boundaries and model semantic information. Therefore, researchers have proposed the hybrid method combining deep learning with the object-based method or superpixel segmentation. For example, ref. [42] proposed an objectbased method that used a convolutional neural network (CNN) to extract change features, which achieved higher accuracy and computational efficiency. In [43], a CD method combining a neural network and the extraction of superpixel-level change features was proposed, which can obtain a robust and high-contrast CD result. The authors in [44] proposed a CD method combining superpixel segmentation and a graph neural network. Bi-temporal superpixel maps were generated via simple linear iterative clustering (SLIC) [45], and the superpixel-level change features were extracted to generate the graph. However, the above methods isolate the superpixel or object generation from the deep network training. The generated segmentation cannot be dynamically adjusted during the training, thus greatly limiting the performance of CD and failing to obtain the global optimal solution.

To solve the above problems, a novel end-to-end unsupervised CD method combining the superpixel segmentation network and Siamese deep convolutional network is proposed. Two weight-sharing superpixel sampling networks (SSNs) [46] are introduced in series with a Siamese deep convolutional CD network, and the overall framework still follows the Siamese architecture (Figure 1). Firstly, two SSNs are used to generate superpixel and deep features containing segmentation information with less noise. Then, the Unet-based Siamese CD network is used to extract multi-scale change information. The proposed method can train the superpixel segmentation part and CD network end-to-end under a unified framework and finally obtain the global optimal parameters. During the training process, the task-specific loss function promotes the adaptive attachment of the superpixel to the change boundary. The main contributions of this paper are as follows:

- (1) This study combines knowledge-driven and unsupervised learning to propose an endto-end CD network. The incorporation of superpixel segmentation information is an interesting practice of integrating prior knowledge into the deep learning technique. The generated superpixels in our proposed method can be adjusted adaptively, which ensures better consistency in the superpixel segmentation of unchanged areas and closer segmentation to change boundaries in changed areas for the bi-temporal data.
- (2) This study is the first to explore the ability of the network to detect changes, which is crucial for the generalization performance of CD networks. We designed transfer learning experiments between homogeneous data and even heterogeneous data to explore the ability to detect changes and generalization performance. This information is of great importance for the development of DLCD for SAR images with no or limited samples in the future.
- (3) The proposed method is unsupervised and is friendly to SAR data with extremely limited labeled samples. Preprocessed SAR images of different sizes can be input into our network to obtain the change map with high accuracy. Furthermore, this method has the potential to be extended to more complex sequential image processing.



**Figure 1.** The overall framework of the proposed method. The input of the network is the SAR intensity image pairs and the output is the predicted probability. The joint loss function is used to backpropagate and the binary change map is finally obtained. Training steps see Algorithms 1.

_	
ł	Algorithms 1: Training steps of the proposed method.
J	( <b>nput:</b> Bi-temporal SAR images $I_1$ and $I_2$
(	<b>Dutput:</b> The binary change map
ľ	Normalize the bi-temporal SAR images $I_1$ and $I_2$
f	f <b>or</b> Number of training iterations <b>do</b>
1	. Two weight-sharing SSNs take normalized bi-temporal SAR images, respectively, as input.
2	2. Two SSNs' output pixel–superpixel associations $Q_1$ and $Q_2$ and high-level features $F_{pix_1}$ and
ł	$F_{pix_2}$ for different times.
З	3. K-dimensional $F_{pix_1}$ and $F_{pix_2}$ are fed into the Siamese CD network.
4	<ol> <li>The Siamese CD network outputs the predicted probability map P.</li> </ol>
5	5. Calculate the weighted cross entropy loss $L_{CE}$ and dice loss $L_{Dice}$ as Equations (7) and (9).
6	5. For both of the different times, calculate task-specific reconstruction loss $L_{rec}$ as Equation (10)
а	and take the positional pixel features $I_{xy}$ of input to calculate compactness loss $L_{cpt}$ as Equation
(	
7	7. Calculate the joint loss <i>L</i> as Equation (15).
8	3. Update the parameters of networks based on the joint loss <i>L</i> .
ł	End for
_	
2.	
2	1. Unsupervised Change Detection Workflow
	The proposed method works in an unsupervised manner. The first step was to genera
re	liable training samples via an unsupervised pre-task to train the network.
	Given two coregistered SAR intensity images $I_1 = \{I_1(i, j), 1 \le i \le A, 1 \le j \le N\}$
ar	and $I_2 = \{I_2(i, j), 1 \le i \le A, 1 \le j \le B\}$ , acquired at different times. $t_1$ and $t_2$ , over the second seco

and  $I_2 = \{I_2(i, j), 1 \le i \le A, 1 \le j \le B\}$ , acquired at different times,  $t_1$  and  $t_2$ , over the same geographic area, the log-ratio operator [7] was used to generate the DI. Previous studies have proven that for SAR images, the ratio operator is not only more robust toward calibration errors, but can suppress multiplicative noise [47]. The log-ratio operator takes the logarithm to the ratio image and further converts the residual multiplicative noise into additive noise which is easier to process, as shown in Equation (1).

$$LR = \log\left(\frac{I_1}{I_2}\right) = \log I_1 - \log I_2 \tag{1}$$

where *LR* is the log-ratio DI, and *log* represents the natural logarithm.

After the DI was obtained, the hierarchical FCM (HFCM) [48] clustering algorithm was used to classify the DI into three clusters: changed class  $\Omega_c$ , unchanged class  $\Omega_u$ , and uncertain class  $\Omega_i$ . Pixels belonging to  $\Omega_c$  and  $\Omega_u$  could be considered to be reliable samples with a high probability of being changed or unchanged. Although these "uncertain" pixels were not used as training samples, their semantic information could still be utilized because we used the image-based fully convolutional network.

The above unsupervised design allowed us to perform CD even with only one target image pair. However, only a few pixels in this image pair were selected as samples and other pixels need to be further classified, so training samples are extremely rare. Therefore, the data augmentation technique was used to augment samples to prevent overfitting [49,50]. We applied random crops and rotations in multiples of 90° (90°, 180°, and 270°) to the normalized bi-temporal images and the generated pseudo-label map with a 50% probability. Then, we used these samples to train the network and finally obtained the binary change map. Although the training of the network was supervised, the selection of training samples was unsupervised, so the whole CD flow was essentially unsupervised, as shown in Figure 2.



Figure 2. The flowchart of unsupervised change detection for SAR image.

## 2.2. Superpixel Sampling Networks (SSNs)

The SSN [46] is the first end-to-end deep superpixel segmentation network that can be easily integrated with the downstream deep network to generate task-specific superpixels. This paper aims to use SSN to generate CD-specific superpixels that adhere better to the change boundary, reduce the influence of speckle noise, and refine the boundary of the final change map.

Figure 3 shows the overall architecture of the SSN, which consists of two parts: (1) the CNN-based feature extractor, where a deep network was used to extract features for superpixel segmentation to replace manually designed features. (2) Differentiable SLIC (DSLIC), where the features from the feature extractor were fed into the DSLIC to implement the superpixel segmentation. Given an image to be segmented as the input of SSN, we could obtain the pixel–superpixel soft association  $Q \in R^{n \times m}$  and the high dimensional features  $F_{pix} \in R^{n \times c}$ . Q could be used to realize the mutual mapping between pixel feature representation  $P \in R^{n \times c}$  and superpixel feature representation  $S \in R^{m \times c}$ , where n is the number of pixels, m represents the number of superpixels, and c is the number of channels.



**Figure 3.** Superpixel sampling networks (SSNs). The SSN consists of a feature extractor and the differentiable SLIC. Input: SAR image; output: the pixel–superpixel associations Q and high-level feature  $F_{pix}$  from the feature extractor. The arrows in the black circle illustrate upsampling.

## 2.2.1. The Input of the SSN

In [46], the SSN was used for RGB image segmentation, and its input was 5-dimensional scaled XYLab features, including three-channel CIELAB color and two-channel positional features (x, y). The position and color scales are expressed as  $\gamma_{pos}$  and  $\gamma_{color}$ , respectively. The value of  $\gamma_{color}$  was selected by experiences or trial, while the value of  $\gamma_{pos}$  was determined by the number of superpixels:

$$\gamma_{pos} = \eta \, max \left( \frac{m_w}{n_w}, \, \frac{m_h}{n_h} \right) \,, \tag{2}$$

 $m_w$ ,  $m_h$ ,  $n_w$ , and  $n_h$  represent the initial number of superpixels and the number of pixels along the image width (*w*) and height (*h*), respectively.  $\eta$  is an empirical constant, set to 2.5 in [46].

# 2.2.2. Feature Extractor Based on CNN

As shown in Figure 3, the feature extractor is a common CNN-based network, consisting of a series of  $3 \times 3$  convolutional layers, batch normalization (BN) layers, and rectified linear unit (ReLU) nonlinear layers. After the second and fourth convolution layers, two 2  $\times$  2 max-pooling layers were, respectively, used for downsampling to expand the receptive field. Skip connections were used to fuse the multi-scale information from shallow and deep layers. The output channel of each hidden layer (i.e., base channel) was set to 64. The feature channel of the output layer was set to (K - 1). Then, the one-channel input and (K - 1)-channel output were concatenated to produce the final *K*-dimensional pixel features. This feature extractor can also be replaced by other networks. The resulting *K*-dimensional features will be fed into the DSLIC and downstream Siamese CD networks. The pixel-superpixel association  $Q \in \mathbb{R}^{n \times m}$  will be iteratively updated.

#### 2.2.3. Differentiable SLIC

The core of differentiable SLIC is to replace the non-differentiable nearest neighbor operation of SLIC [45] with a distance soft association  $Q \in \mathbb{R}^{n \times m}$  defined by a Gaussian radial basis function. The initialization strategy of the SSN is to divide the image into regular grids as initial superpixels; the clustering algorithm can use soft k-means or others. For the pixel feature representation  $P \in \mathbb{R}^{n \times c}$  and the superpixel feature representation  $S \in \mathbb{R}^{m \times c}$ , the soft association between the pixel *i* and the superpixel *j* can be calculated at the *t*th iteration as follows:

$$Q_{ij}^{t} = e^{-D(P_{i}, S_{j}^{t-1})} = e - \|P_{i}, S_{j}^{t-1}\|^{2},$$
(3)

where *D* denotes distance computer. The new superpixel centers were computed using the weighted sum of pixel features:

$$S_{j}^{t} = \frac{\sum_{i=1}^{n} Q_{i,j}^{t} P_{i}}{\sum_{i=1}^{n} Q_{i,j}^{t}}$$
(4)

For convenience, denote column-normalized  $Q^t$  as  $Q^t$ ; then, rewrite the update of superpixel centers as

$$S^{t} = \hat{Q}^{t^{-1}} P \tag{5}$$

Through Equation (5), one can realize the mapping from pixel to superpixel representation, and the inverse mapping from superpixel to pixel representation can be achieved through Equation (6):

$$P = Q^t S^t \tag{6}$$

where  $Q^t$  is the row-normalized  $Q^t$ . In the calculation of  $Q \in \mathbb{R}^{n \times m}$ , only 9 superpixels surrounding the pixel are considered to improve the computational efficiency, that is, m = 9. This simplification is similar to the nearest neighbor searching of SLIC. The interactive update of Q and S was realized by Equations (3) and (4). It is worth noting that P was updated in the continuous learning of the model rather than in the iteration.

## 2.3. End-to-End Change Detection Network with SSN

## 2.3.1. Overall Framework

As shown in Figure 1, SAR images acquired at  $t_1$  and  $t_2$  were, respectively, fed into two weight-shared SSNs to obtain the pixel–superpixel associations  $Q_1$  and  $Q_2$  and high-level features  $F_{pix_1}$  and  $F_{pix_2}$ , which were set to be passed, respectively, to the two branches of the following Siamese CD network. This study used two Siamese CD networks with different designs, which we will introduce in the next section. The joint loss combining multiple loss function with different roles was calculated between the predicted probability map and the label map, which required  $Q_1$  and  $Q_2$  as well as the positional features of the input. Finally, the binary change map was obtained. It is worth noting that  $F_{pix_1}$  and  $F_{pix_2}$  were features generated specifically for superpixel segmentation; that is to say, these features were averaged according to the segmented superpixels to compress the noise.

#### 2.3.2. Siamese CD Network

In paper [34], three fully convolutional neural network (FCNN) architectures were proposed for the CD of Earth observation data, and two of these Siamese networks were used as our CD network. We connected the SSN in series with these two Siamese networks, respectively, to verify the effectiveness of the proposed method.

FC-Siam-conc

The first Siamese network (Figure 4a) is a fully convolutional network based on the decoder–encoder architecture. Ignoring one branch of this network, we can find that the backbone is actually a shallow version of U-net [29]. This structure uses the Siamese network as its encoder to process images acquired at different times through two branches with shared weights. This design can fully mine bi-temporal information to generate bi-temporal high-level features. The multi-level features from the two branches of the encoder are concatenated with the output of the corresponding scale decoding layer using two skip connections. The purpose of this design is to use this decoder to mine the correlation and difference between the bi-temporal information. This structure is named the full convolutional Siamese connection (FC-Siam-conc).



**Figure 4.** Two Siamese networks based on the encoder–decoder structure. (a) FC-Siam-conc; (b) FC-Siam-diff. The top subfigure is the implementation details of the model. Two types of residual blocks (Res-I and Res-II) and decoder modules (Dec-I and Dec-II) are used.  $X_e$  and  $X'_e$  represent the input features from the last residual block and output features, respectively, in the encoding stage.  $X_d$  and  $X'_d$  have a similar denotation to  $X_e$  and  $X'_e$  but they belong to the decoding stage.  $X_{e_{1,2}}$  denotes the bi-temporal features extracted from the encoding module by using skip connections. Orange arrows illustrate weight sharing.

• FC-Siam-diff

The second Siamese network (Figure 4b) differs from FC-Siam-conc only in that it uses one skip connection to obtain the absolute value of the difference in the bi-temporal features from two encoding streams for each scale. This difference feature is then concatenated with the output of the decoding layer. This design mines the multi-scale change information by adding the robust and explicit difference, which is more specific to CD and is named FC-Siam-diff.

These two Siamese networks have the same backbone, using  $3 \times 3$  convolutional layers and  $2 \times 2$  max-pooling layers for downsampling as well as using the BN layer to speed up model convergence. Each block uses residual connections to mitigate gradient vanishing. The skip connections between the encoder and the decoder are used to fuse multi-scale information. The implementation details are shown in Figure 4.

#### 2.4. Loss Function

The number of samples in the changed set  $\Omega_c$  and the unchanged set  $\Omega_u$  is often highly unbalanced, and usually  $\Omega_c$  has fewer samples. If this situation is not considered, the detection accuracy for the changed class will be reduced. Therefore, we use the weighted cross entropy (CE) function to deal with this problem, defined as follows:

$$L_{CE}(\boldsymbol{P},\boldsymbol{G}) = \frac{1}{N} \sum_{i} \left[-\omega_c \boldsymbol{G}_i log(\boldsymbol{P}_i) - \omega_u (1 - \boldsymbol{G}_i) log(1 - \boldsymbol{P}_i)\right],$$
(7)

where *P* and *G* represent the predicted change map and the ground truth, respectively, *i* is pixel index, and  $\omega_c$  and  $\omega_u$  represent the weights of the changed and unchanged class, respectively. *N* is the number of pixels excluding ignored pixels.

Dice loss is also an appropriate choice to further reduce the problem of sample imbalance. Dice similarity is defined as Equation (8), which can measure the similarity of the predicted change map P and ground truth G, and its value ranges from 0 to 1, i.e., [0, 1]. The dice loss is defined as in Equation (9):

$$Dice = \frac{2|P \cap G|}{|P| + |G|} , \qquad (8)$$

$$L_{Dice}(\boldsymbol{P},\boldsymbol{G}) = 1 - Dice \tag{9}$$

To generate task-specific and more compact superpixels, we use a combination of a task-specific reconstruction loss and compactness loss to train the SSN, as in paper [46]:

$$L_{rec}(\boldsymbol{G},\boldsymbol{Q}) = L_{CE}(\boldsymbol{G},\boldsymbol{G}^*) = L_{CE}\left(\boldsymbol{G},\widetilde{\boldsymbol{Q}}\widehat{\boldsymbol{Q}}^T\boldsymbol{G}\right), \tag{10}$$

$$L_{cpt}(\boldsymbol{I}_{xy},\boldsymbol{Q}) = \|\boldsymbol{I}_{xy} - \bar{\boldsymbol{I}}_{xy}\|_{2}$$
(11)

where  $I_{xy}$  represents the positional pixel features of the input, firstly map the  $I_{xy}$  into the superpixel space to obtain  $S^{xy}$  through Equation (12):

$$\boldsymbol{S}^{xy} = \boldsymbol{\hat{Q}}^T \boldsymbol{I}_{xy} \tag{12}$$

Then, the pixel is endowed with the absolute index of the superpixel through hard association rather than soft association Q to obtain  $\overline{I}_{xy}$ :

$$\bar{I}_{xy} = S_j^{xy} \mid H_i = j \tag{13}$$

$$H_i = \underset{j \in \{1, \dots, m\}}{\operatorname{argmax}} Q_{ij} \tag{14}$$

 $\overline{I}_{xy}$  also belongs to the pixel space. As shown in Equation (11), we calculate the L2 norm of  $I_{xy}$  and  $\overline{I}_{xy}$ .

Finally, the joint loss function is defined as in Equation (15), and we used it to train our CD networks.

$$L = L_{CE}(P, G) + L_{Dice}(P, G) +\lambda_1 [L_{Rec}(G, Q_1) + L_{Rec}(G, Q_2)] +\lambda_2 [L_{cpt}(I_{xy}, Q_1) + L_{cpt}(I_{xy}, Q_2)]$$
(15)

where  $\lambda_1$  and  $\lambda_2$  are the weight factors. The first term is the main component of the loss function, which penalizes the overall network learning. The second and last terms are calculated from different times, which encourages the network to simultaneously mine bitemporal information as much as possible to generate task-specific and compact superpixels. In addition, Algorithms 1 provides the training steps of our proposed method.

# 3. Results

# 3.1. Datasets and Evaluation Criteria

Several public CD datasets were used in our experiment, including the Ottava dataset, Sulzberger dataset, Yellow River dataset, and San Francisco dataset, whereby all of which comprise single-polarization SAR images and are commonly used in published papers. In addition, we also collected an optical CD dataset, the Mexico dataset.

• Ottava dataset: two images were acquired via the Radarsat-1 satellite over Ottawa in May 1997 and August 1997, and the change was caused by the summer flooding (Figure 5).



**Figure 5.** Ottava dataset. (**a**) Image acquired in May 1997; (**b**) image acquired in August 1997; (**c**) ground truth.

• **Sulzberger dataset** [51]: This dataset was acquired via the Envisat satellite on 11 and 16 March 2011. Both images show the process of sea ice breakup (Figure 6).



**Figure 6.** Sulzberger dataset. (**a**) Image acquired on 11 March 2011; (**b**) image acquired on 16 March 2011; (**c**) ground truth.

• Yellow River dataset: These two images were acquired via the Radarsat-2 satellite in June 2008 and June 2009 at the estuary of the Yellow River in Dongying, Shandong Province (Figure 7). It is worth noting that the two images are single-look and four-look, respectively. As a result, they are affected by noise to different degrees. Four typical change areas were selected: Farmland-A, Farmland-B, inland water, and coastline dataset.



**Figure 7.** Yellow River dataset. (**a**) Image acquired in June 2008; (**b**) image acquired in June 2009; (**c**) ground truth.

• San Francisco dataset [52]: This dataset was captured via the ERS-2 satellite in August 2003 and May 2004 (Figure 8).



**Figure 8.** San Francisco dataset. (**a**) Image acquired in August 2003; (**b**) image acquired in May 2004; (**c**) ground truth.

• **Mexico dataset:** This dataset consists of two optical images captured via Landsat-7 in Mexico City in April 2000 and May 2002, respectively. They were extracted from ETM+ images in band 4, the near infrared (NIR) band. This dataset shows the destruction of vegetation after a forest fire in Mexico city (Figure 9).



**Figure 9.** Mexico dataset. (**a**) Image acquired in April 2000; (**b**) image acquired in May 2002; (**c**) ground truth.

Refer to Table 1 for more information about the experimental datasets.

Table 1. Details of the experimental datasets.

Items	Ottawa	Sulzberger	Yellow River		San Francisco	Mexico
Satellite	Radarsat-1	Envisat	Radar	Radarsat-2		Landsat-7
Acquisition time	May 1997– August 1997	11 March 2011– 16 March 2011	June 2008– June 2009		August 2003– May 2004	April 2000– May 2002
Band	Band C	Band C	Band C		Band C	NIR (0.775–0.900 μm)
Size	290 × 350	$256 \times 256$	Farmland-A: Farmland-B: Inland water: Coastline:	$306 \times 291$ 257 × 289 291 × 444 450 × 280	$256 \times 256$	512 × 512
Reasons for change	Flood	Sea ice breakup	Environmental change		Unknown	Forest fire

Five evaluation indicators were introduced to quantitatively evaluate the method, including overall accuracy (OA), precision (Pre), recall, F1-score (F1), and Kappa coefficient. Specifically, OA is the ratio between the pixels correctly predicted against the sum of all pixels. Precision corresponds to the proportion of the number of pixels correctly predicted as the changed class and the total number of pixels predicted as the changed class. The F1-score combines precision (Pre) and recall and is often used to evaluate the binary classification accuracy. Recall reflects the percentage of pixels correctly predicted as the changed class and the total changed pixels of the ground truth. They are calculated as follows:

$$OA = (TP + TN) / (TF + FP + TN + FN)$$
(16)

$$Pre = TP/(TP + FP) \tag{17}$$

$$F1 \ score = \frac{2 \times Pre \times Recall}{2}$$
(18)

$$Pre + Recall$$
 (10)

$$Recall = TP/(TP + FN)$$
<sup>(19)</sup>

where *TP* and *TN* are the number of true positives and true negatives. *FP* and *FN* are the number of false positives and false negatives. The Kappa coefficient can measure the overall

consistency between the predicted map and the ground truth, and its value has higher reference reliability for the CD with sample imbalance, which is calculated as follows:

$$P_e = \frac{(TP + FP) \times (TP + FN) + (FN + TN) \times (FP + TN)}{(TP + TN + FP + FN)}$$
(20)

$$Kappa = \frac{OA - P_e}{1 - P_e} \tag{21}$$

## 3.2. Experimental Setting

We implemented the proposed method in PyTorch v1.8, and the training was driven by the NVIDIA Quadro RTX 8000 GPU produced by Lenovo in Beijing, China.

The hyperparameter setting is shown in Table 2. To facilitate the reader in obtaining information, we divide the hyperparameters into four parts: deep learning universal hyperparameters, adjustable parameters for the feature extractor, differentiable SLIC in SSN, and hyperparameters in the loss function. The initial learning rate was set to 0.001 and halved for every 100 epochs using Lr\_scheduler. L2 regularization and the aforementioned data enhancement strategy were used to mitigate overfitting. The batch size could only be set to one. The crop size was adjusted according to the size of the image, and the number of superpixels in the SSN was adjusted according to crop size. For the feature extractor of the SSN, we set the base channel to 64 and the output channel K to 20 for our data, which could be adjusted based on the complexity of the data. The number of iterations for differentiable SLIC in the SSN was set to 10 for both the training and prediction. In the loss function,  $\omega_c$  and  $\omega_u$  were set to 0.4 and 0.6, and  $\lambda_1$  and  $\lambda_2$  were set to 0.0001 and 1.0. In addition, this study only involved a single-polarization SAR, and when we tried to add two-channel positional features (x and y) into the input, the model was difficult to converge. We suspect that the model mistakenly believed that the positional feature was more important than the original image, as the original image only has one channel while the positional feature has two channels. We also tried adding positional features before feeding the differentiable SLIC of the SSN, which allowed the model to converge. However, the results of 100 experiments showed that adding positional features reduced the detection accuracy by 1-2%. We infer that the high-level features extracted by the deep network are highly effective for superpixel segmentation, and adding original and primary positional features may diminish the advancement of these features, resulting in a decrease in accuracy. Therefore, in this study, only normalized single-channel SAR data were fed into the SSN without positional features. Apparently, the scalers,  $\gamma_{pos}$ , and  $\gamma_{color}$ , as well as  $\eta$ , are not required to be set.

To verify the effectiveness of the proposed method, we connected the SSN to two Siamese CD networks, FC-Siam-conc and FC-Siam-diff, to obtain SSN-Siam-conc and SSN-Siam-diff networks. We provide the running time of SSN-Siam-diff under the above hardware conditions and experimental settings for the readers' reference. The time to train 300 epochs is about 90 s, and the prediction time for an image size of  $290 \times 350$  is about 0.08 s.

# 3.3. Enhancement Effect in Series with SSN

Two SAR datasets were used to verify the enhancement effect in series with the SSN, i.e., the Ottawa (Figure 5) and Sulzberger (Figure 6) datasets. For each network and dataset, we conducted 100 experiments and recorded the accuracy matrix of the best model for each experiment. We present the best results of 100 experiments in the visualized and quantitative CD results. Furthermore, the superpixel generation results are presented to explore the reasons for the superior performance of our method.

 	User are available	Setting		
Items	nyperparameter –	Ottawa	Sulzberger	
	Initial learning rate Optimizer	0.0 Ad	01 am	
Deep learning universal hyperparameters	Lr_scheduler Regularization Batch size	StepLR (step_size = L2 regula	= 100, gamma = 0.5) arization	
	Crop size	256	196	
(SSN) Feature extractor	Base channel	64		
hyperparameters	channel K	20		
(SSN) Differentiable SLIC	Num superpixel	256	196	
hyperparameters	Num iterations	10		
Loss function	$\omega_c, \omega_u$ $\lambda_1, \lambda_2$	(0.4, 0.6) (0.0001, 1.0)		

# Table 2. Hyperparameter setting.

# 3.3.1. CD Results

Figure 10 shows two groups of results obtained by connected and unconnected SSNs to FC-Siam-conc and FC-Siam-diff for the Ottawa (first row) and Sulzberger (second row) datasets. The results of column 1 and column 2 are compared, corresponding to FC-Siam-diff. Similarly, the results of column 3 and column 4 are compared, corresponding to FC-Siam-conc and SSN-Siam-conc. SSN-Siam-diff and SSN-Siam-conc win by a large margin. In the areas marked by the red box, the change boundary obtained by the network with the SSN is closer to the ground truth, and very small change areas are also detected more accurately. The quantitative results are shown in Tables 3 and 4. After being connected with the SSN in series, every accuracy indicator of the network was improved, and the improvement effect is particularly significant for the Ottawa dataset. These results suggest that our proposed method with SSN not only has a good boundary-preserving ability, but also improves the detection accuracy of small area changes.



**Figure 10.** Comparison of CD results from networks with or without SSN in the Ottawa dataset (first row) and Sulzberger dataset (second row). (a) FC-Siam-diff; (b) SSN-Siam-diff; (c) FC-Siam-conc; (d) SSN-Siam-conc; and (e) ground truth. The areas marked by red boxes deserve more attention.

Network	OA (%)	Pre (%)	Recall (%)	F1 (%)	Kappa (%)
FC-Siam-diff	97.47	91.67	92.42	92.04	90.54
SSN-Siam-diff	98.42	93.00	97.35	95.13	94.19
FC-Siam-conc	98.01	94.83	92.47	93.64	92.46
SSN-Siam-conc	98.87	95.48	97.49	96.48	95.81

Table 3. Quantitative results of Ottawa dataset for models with SSN or without.

Table 4. Quantitative results of Sulzberger dataset for models with SSN or without.

Network	OA (%)	Pre (%)	Recall (%)	F1 (%)	Kappa (%)
FC-Siam-diff	97.71	97.03	90.88	93.85	92.45
SSN-Siam-diff	98.36	95.48	96.00	95.74	94.72
FC-Siam-conc	98.28	96.42	94.57	95.48	94.42
SSN-Siam-conc	98.77	96.20	97.45	96.82	96.05

Figure 11 shows the distribution of accuracy indicators of 100 experiments for each network and two datasets in the form of a boxplot. The boxplots of both datasets show that all accuracy indicators of the networks with the SSN are significantly higher than the original networks without the SSN. It is worth noting that in 100 experiments of the networks without the SSN, many outliers with very low accuracy appear, while for the networks with the SSN, the outliers are always much higher than the overall detection accuracy, which indicates that the incorporation of superpixel segmentation has great potential to improve the accuracy and is more stable, since its detection accuracy is always maintained at a high level.



Figure 11. Cont.



**Figure 11.** Accuracy indicator distribution of 100 experiments. (**a**,**b**) refer to the boxplot of FC-Siamdiff vs. SSN-Siam-diff and FC-Siam-conc vs. SSN-Siam-conc for the Ottawa dataset, respectively. Similarly, (**c**,**d**) correspond to the Sulzberger dataset.

# 3.3.2. Superpixel Segmentation Results for Bi-Temporal SAR Images

We cropped a  $256 \times 256$  area on the Ottawa dataset to display the superpixel segmentation results from the SSN-Siam-diff network in Figure 12. The first column is the bi-temporal SAR images, and the middle column corresponds to their superpixel generation results. The third column shows the results where the pixel value is replaced by the mean value of the superpixel to which this pixel belongs. The Ottawa dataset contains winding coastlines and some narrow rivers, which are challenges for superpixel segmentation. For example, in the areas marked in the red box, there are narrow streams of water or land with complex boundaries. In these areas, it is difficult for the superpixel segmentation to perfectly adhere to the boundaries, but the superpixel generation network in our method performs very well. The segmentation results for different times demonstrate that our method fully mines bi-temporal information and generates high-quality superpixels for both bi-temporal data. These high-quality results can be attributed to our proposed end-to-end unified framework for obtaining global optimal solutions, as well as Siamese structures, and the adaptive adjustments of superpixels. High-quality superpixel segmentation lays a foundation for boundary optimization and detail preservation for the final change map.

## 3.4. Transfer Learning Experiments

Deep learning relies heavily on a large number of training data. In the RSCD field, although the Earth observation data have been considerably enriched, the available labeled CD data are scarce, especially for the SAR data. Transfer learning is an important tool to solve the problem of insufficient training samples. We think that the transfer learning ability or generalization performance of the CD network is positively correlated with its ability to detect changes. The better the ability of the model to detect changes, the better the generalization performance of the other CD datasets. Research on how to design a model with strong transfer learning ability is helpful to fully utilize multi-source CD datasets.

Therefore, in this section, we designed the transfer learning experiment to explore the ability of the model to "learn how to detect change information", and in comparison, to examine whether the ability and generalization performances of our proposed method are enhanced. In this part, the pre-trained models were obtained based on Ottawa dataset training, which still followed the unsupervised CD flow as mentioned above. Then, these pre-trained models were used on other SAR CD datasets, even optical CD datasets, that are never seen during the training process. This is a simple "parameter sharing" type of transfer learning.



**Figure 12.** Superpixel segmentation results from the SSN-Siam-diff network for the Ottawa dataset. The first row corresponds to the image captured at  $t_1$ . (a)  $t_1$  image; (b) superpixel generation for  $t_1$  image; (c) the pixel value is replaced by the mean value of the superpixel to which this pixel belongs. Similarly, (d–f) correspond to the results of the image captured at  $t_2$ . The areas marked by red boxes deserve more attention.

#### 3.4.1. Transfer Learning for SAR Dataset

FC-Siam-conc

SSN-Siam-conc

Figure 13 shows the results of applying pre-trained models to other SAR datasets, and Tables 5–7 provide the quantitative results. For the San Francisco dataset (Figure 13, first row), the results obtained by SSN-Siam-diff are smoother with less noise and holes than FC-Siam-diff. These results prove that the proposed method can effectively compress noise and holes, thus improving the CD accuracy, which indicates that the proposed network can extract more robust deep features. The comparison between SSN-Siam-conc and FC-Siam-conc also confirmed these conclusions. Among the four networks, only SSN-Siam-conc can detect the narrow change area marked by the red box, which indicates that the design of "conc" (Figure 4a) seems to exceed the design of "diff" (Figure 4b), but whether this is the truth will be discussed briefly in Section 3.4.2.

Network	OA (%)	Pre (%)	Recall (%)	F1 (%)	Kappa (%)
FC-Siam-diff	96.62	76.83	75.43	76.12	74.30
SSN-Siam-diff	98.10	85.61	88.26	86.92	85.89

Table 5. Quantitative results of transfer learning experiment for San Francisco dataset.

66.93

91.91

95.61

98.97

Table 6.	Quantitative	results of tra	insfer learning	; experiment for	Yellow	River-	-Farmland-A	dataset.
----------	--------------	----------------	-----------------	------------------	--------	--------	-------------	----------

76.35

93.85

71.33

92.87

68.97

92.32

Network	OA (%)	Pre (%)	Recall (%)	F1 (%)	Kappa (%)
FC-Siam-diff	98.69	93.93	83.17	88.22	87.53
SSN-Siam-diff	<b>98.88</b>	<b>97.43</b>	<b>83.30</b>	<b>89.81</b>	<b>89.22</b>
FC-Siam-conc	98.80	89.30	90.63	89.96	89.32
SSN-Siam-conc	<b>99.01</b>	<b>95.20</b>	<b>87.67</b>	<b>91.28</b>	<b>90.75</b>

**Figure 13.** Change detection results of transfer learning experiment for SAR datasets, including the San Francisco (first row), Farmland-A (second row) and Farmland-B datasets. (**a**) FC-Siam-diff; (**b**) SSN-Siam-diff; (**c**) FC-Siam-conc; (**d**) SSN-Siam-conc; and (**e**) ground truth. The areas marked by red boxes deserve more attention.

Table 7. Quantitative results of transfer learning experiment for Yellow River–Farmland-B dataset.

Network	OA (%)	Pre (%)	Recall (%)	F1 (%)	Kappa (%)
FC-Siam-diff	94.05	83.21	82.73	82.97	79.37
SSN-Siam-diff	<b>95.20</b>	<b>89.68</b>	<b>82.04</b>	<b>85.69</b>	<b>82.81</b>
FC-Siam-conc	94.78	90.46	78.49	84.05	80.94
SSN-Siam-conc	<b>95.01</b>	<b>89.09</b>	<b>81.50</b>	<b>85.12</b>	<b>82.13</b>

The accurate segmentation of farmland change boundaries is a challenge for CD. As for the Yellow River Farmland-A dataset (Figure 13, second row), compared with FC-Siam-diff, the results obtained by SSN-Siam-diff show that the change boundary segmentation is unbroken and continuous (for example, the area marked with the red box), which is closer to the ground truth and has less noise. The false connectivity between the independent change components is also significantly reduced. FC-Siam-conc has the worst problem of false boundary connectivity, which is probably related to the design of "conc". However, SSN-Siam-conc significantly improves this problem and compresses the noise, and it achieved the best performance both in terms of visual presentation and accuracy indicators.

The results of the Farmland-B dataset (Figure 13, third row) strongly demonstrate the robustness of the proposed method toward speckle noise. Furthermore, FC-Siam-diff and SSN-Siam-diff can successfully detect the slender change area marked by the red box. SSN-Siam-conc yields a less noisy result than FC-Siam-conc, but neither of them detected the change in the red box. Therefore, it seems difficult to judge which design of "diff" or "conc" has more advantages in transfer learning.

3.4.2. Comparison of Generalization Performance between Conc and Diff Models

According to the above experimental results, it seems difficult to judge which design is better, diff or conc. Initially, we found that the "conc" pre-trained models failed to detect changes in some datasets, which caught our attention. Therefore, we tried to exchange the order of the bi-temporal images, that is, exchange the input of the two branches in Figure 4a. In this way, the "conc" model obtained completely different results from before.

FC-Siam-diff and FC-Siam-conc do not connect to SSN, have lower computational cost, and can also clearly state this problem. Therefore, Figure 14 shows the CD results of exchanging the input sequence of FC-Siam-diff and FC-Siam-conc applied to the Yellow River coastline (first row), Farmland-A (second row), and inland water (third row) datasets.



**Figure 14.** CD results of swapping the input sequence of bi-temporal images for FC-Siam-diff and FC-Siam-conc. (**a**,**b**) correspond to the results of FC-Siam-diff switching the input sequence. Similarly, (**c**,**d**) correspond to the results of FC-Siam-conc. (**e**) Ground truth. These three datasets are Yellow River coastline (first row), Farmland-A (second row), and inland water (third row).

As the results for the coastline dataset show, FC-Siam-diff can effectively detect changes regardless of the input order. However, when the input order is based on image acquisition time, FC-Siam-conc does not work at all, and no valid information is detected (Figure 14d, first row). However, after switching the order of inputs, FC-Siam-conc can effectively detect the change (Figure 14c, first row). The results for Farmland-A also show the same information as above. When FC-Siam-conc is used in the inland water dataset, only a partial change can be detected in both orders. These two different change components shown in Figure 14c,d (third row) correspond, respectively, to the positive and negative change in the water. These results indicate that the "conc" model can only learn to detect changes consistent with the change in the training data. As in the Ottawa dataset, the flood in the image acquired at  $t_2$  has faded compared with the image acquired at  $t_1$ , and the water body shows a negative change. As a result, the "conc" pre-trained models can only detect the change similar to the negative change in water. However, the "diff" model overcomes this problem. We infer that FC-Siam-diff and SSN-Siam-diff add explicit difference guidance to the model, which makes the model more specific to the CD and gives the model the ability to detect changes even with few samples. Therefore, from these results, the "diff" model shows better generalization performance.

When performing with few training samples, it is important to consider whether the model can detect both positive and negative changes. However, many existing studies have ignored this problem. The methods in many published papers are based on few training data; for example, a small area is always clipped from a large image and marked

as training data. This small area may only contain positive changes or negative changes, or the samples of the two types of change may be extremely unbalanced. Therefore, designing a model that is robust in all three conditions as above can be considered a useful generable CD approach.

## 3.4.3. Transfer Learning for Optical Dataset

We further applied these pre-trained models to the optical CD data, which was more challenging because their imaging mechanisms are completely different. Figure 15 shows the CD results and the superpixel generated by the SSN-Siam-diff of the Mexico dataset. In terms of SSN-Siam-diff vs. FC-Siam-diff and SSN-Siam-conc vs. FC-Siam-conc, the networks with SSN were better at retaining the change details, such as in areas marked by red circles. Therefore, these results show that the proposed method exhibits strong generalization ability even when transferred to heterogeneous data, and the ability to compress noise and refine boundaries is maintained.



**Figure 15.** Change detection results and superpixel generation of transfer learning experiment for Mexico (optical) datasets. (a) FC-Siam-diff; (b) SSN-Siam-diff; (c) FC-Siam-conc; (d) SSN-Siam-conc; (e) ground truth; (f)  $t_1$  image; (g)  $t_2$  image; (h,i) superpixel generation via SSN-Siam-diff for  $t_1$  and  $t_2$  images. (j) The pixel value of the  $t_2$  image is replaced by the mean value of the superpixel to which this pixel belongs. The areas marked by red circles deserve more attention.

Speckle noise does not exist in optical images, so the superpixels generated by SSN-Siam-diff for the bi-temporal images of the Mexico dataset (Figure 15h–i) are tighter, smoother, and less broken than the superpixels of the SAR images. This makes it easier to observe and analyze the superpixel generation. Figure 15i shows that the pixel value of the  $t_2$  image is replaced by the mean value of the superpixel to which this pixel belongs. As can be seen from Figure 15i,j, the generated superpixel in the changed region better fits the change boundary, such as the area marked by the red circle, while the superpixel boundary in the unchanged region has a better consistency for bi-temporal images. In conclusion, the proposed method has a promising prospect in making full use of multisource heterogeneous data to complete complex CD tasks.

# 3.5. Comparison with Other Methods

To verify the superiority of the proposed method, we compared our method with seven classical and advanced DLCD algorithms, including DBN [32], PCANet [48], CNN [53], LR-CNN [54], DCNet [55], SAFNet [56], and RUSACD [57].

In DBN [32], a pre-task was used to select reliable training samples, and the deep belief network (DBN) was used to detect changes in SAR images. In PCANet [48], a SAR CD algorithm based on PCANet and more robust toward speckle noise was presented. In

CNN [53], the CNN based on patch sampling was used for SAR image CD for the first time. In LR-CNN [54], a more advanced method was adopted to select training samples, and local restricted CNN (LRCNN) was proposed to detect changes in polarized SAR data. DCNet [55] is a channel-weighting-based deep cascade network, which has achieved competitive detection accuracy. SAFNet [56], like our proposed method, is a Siamese CD network with adaptive fusion for bi-temporal SAR images. RUSACD [57] adopted a multi-scale superpixel reconstruction method to generate DI, and then used a clustering algorithm to select training samples and designed a model based on the convolutional wavelet neural network and deep convolutional generative adversarial network to detect small area changes in SAR images.

The results of DBN, CNN, DCNet, SAFNet, and RUSACD were extracted from the original paper with optimal accuracy. PCANet was implemented using the default optimal parameters provided in the original paper. Since the original LR-CNN considered polarization information, we modified the LR-CNN to make it suitable for single-polarization SAR images.

Figure 16 shows the results of different methods used for the Ottawa dataset. Table 8 lists the quantitative evaluation indicators; the best results are in bold font and the secondbest are underlined. We observed that many change pixels were missing for PCANet, LR-CNN, and DCNet. SAFNet and RUSACD obtained competitive results, but the detection accuracy for the small area change is poor, such as the part marked with the red circles. It can be seen from the visual results that the proposed SSN-Siam-conc shows the most abundant details, which are closer to the ground truth, and it achieved the highest values of OA, Recall, F1-score, and Kappa. The proposed SSN-Siam-diff also achieved very good performance. For this dataset, DBN obtained the second-best result in terms of evaluation indicators, but the visual map contained more noise than the other methods, which may be related to the significant limitation in the use of neighborhood information. In conclusion, these results suggest that the proposed method can effectively improve the accuracy of change boundary segmentation and compress speckle noise to a certain extent. Compared with the state-of-the-art methods, it is highly competitive and has considerable potential for exploitation.



**Figure 16.** Change detection results of different methods on the Ottawa dataset. (**a**) DBN; (**b**) PCANet; (**c**) CNN; (**d**) LR-CNN; (**e**) DCNet; (**f**) SAFNet; (**g**) RUSACD; (**h**) SSN-Siam-diff (ours); (**i**) SSN-Siam-conc (ours); and (**j**) ground truth. The areas marked by red circles deserve more attention.

Network	OA (%)	Pre (%)	Recall (%)	F1 (%)	Kappa (%)
DBN	98.83	96.45	96.14	96.29	95.59
PCANet	98.67	95.36	93.07	94.20	93.06
CNN	98.67	96.41	95.13	95.77	95.00
LR-CNN	96.25	99.49	76.65	86.59	84.45
DCNet	98.30	95.67	93.45	94.55	93.54
SAFNet	98.60	94.62	96.67	95.64	94.81
RUSACD	98.13	92.06	96.51	94.24	93.12
SSN-Siam-diff (ours)	98.42	93.00	<u>97.35</u>	95.13	94.19
SSN-Siam-conc (ours)	98.87	95.48	97.49	96.48	95.81

Table 8. Quantitative results of different methods used for the Ottawa dataset.

Figure 17 shows the results of the Farmland-A dataset, and Table 9 shows the quantitative results. It is worth noting that for this dataset, the results of SSN-Siam-diff and SSN-Siam-conc were generated by the transfer learning experiment, that is, generated by the pre-trained models trained by the Ottawa dataset. For DBN, PCANet, CNN, and LR-CNN, the results contain a lot of speckle noise and false positive pixels. DCNet is effective at suppressing noise but has large areas of false positive detection. In the results of RUSACD, we can observe a lot of false connectivity between different change components. The proposed SSN-Siam-conc and SSN-Siam-diff achieved very good performance. Both of these two networks contain relatively little noise, and the continuity and details of the change boundary are well maintained. In particular, SSN-Siam-conc won out of all of the methods and achieved the best results with regard to OA, F1-score, and Kappa. SAFNet also achieved very competitive results and was robust against speckle noise. We infer that this is due to the design of the Siamese network and the appropriate bi-temporal information fusion strategy. In conclusion, our results are the most competitive, although we present the results obtained in the transfer learning experiment, which suggests that our proposed method significantly enhances the ability to compress noise and preserve boundaries and generalization performance.



**Figure 17.** Change detection results of different methods used for the Farmland-A dataset. (a) DBN; (b) PCANet; (c) CNN; (d) LR-CNN; (e) DCNet; (f) SAFNet; (g) RUSACD; (h) SSN-Siam-diff (ours); (i) SSN-Siam-conc (ours); and (j) ground truth.

Network	OA (%)	Pre (%)	Recall (%)	F1 (%)	Kappa (%)
DBN	98.56	87.71	88.00	87.85	86.92
PCANet	96.14	61.89	<u>90.65</u>	73.55	71.55
CNN	98.59	92.31	83.10	87.46	87.09
LR-CNN	98.27	81.26	91.97	86.28	85.36
DCN	98.71	90.34	87.51	88.91	88.33
SAFNet	98.94	92.89	88.96	90.88	90.32
RUSACD	98.67	92.49	84.12	88.10	87.65
SSN-Siam-diff (ours)	98.88	97.43	83.30	89.81	89.22
SSN-Siam-conc (ours)	99.01	<u>95.20</u>	87.67	91.28	90.75

Table 9. Quantitative results of different methods used for the Farmland-A dataset.

#### 4. Discussion

The experimental results show that our method outperforms several existing advanced CD methods. It can not only compress speckle noise effectively and refine the change boundary, but it also has good generalization ability. Furthermore, some important information still deserves to be discussed.

First of all, the proposed method is a combination of superpixel segmentation and a deep CD network, as well as a successful practice of combining prior knowledge and deep learning techniques. In addition, it is difficult for existing methods to balance noise compression and detail preservation, while our proposed method achieves a balance between the two because we can obtain globally optimal parameters.

This study provides a better segmentation strategy for the superpixel generation of multi-temporal images to perform CD. On the one hand, different branches process data acquired at different times to ensure the independence of superpixel generation for multi-temporal data. On the other hand, weight sharing and using the task-specific loss function result in an implicit correlation between the bi-temporal data to generate superpixels. This correlation enables the bi-temporal information to be fully mined and combined, which not only ensures the segmentation consistency of the unchanged area, but also results in the segmentation of the changed area to better fit the change boundary.

It is worth noting that the generated superpixels are not used directly in this article, but the advanced features generated by the SSN for the superpixel generation are fed into the flowing CD network. The higher the quality of the visible superpixel generation, the more accurate the segmentation information contained in this feature, and thus the more the performance of the downstream CD network can be improved.

Lastly, this paper is the first to investigate the CD network's ability to detect changes, which is closely related to the generalization performance. When only a few samples with a single change type (including only positive or negative changes) are available, which is often the case when using SAR images, if the model has a strong ability to detect changes, it can effectively identify both positive and negative changes. However, this issue has never been discussed in published papers. Future research could pay more attention to the ability to detect the change when designing CD models, which can promote the full utilization of multi-source CD datasets, so as to design models with a larger capacity to solve more complex CD tasks.

In conclusion, the proposed method is unsupervised, and it is easy to apply to other more complex data to perform CD, such as fully polarized SAR data, which is worth studying in the future. However, due to a lack of validation data, the performance of the proposed method has not been demonstrated in heterogeneous scattering cases such as buildings in SAR images. In addition, our study provides some heuristic information for many tasks involving time series image data processing.

# 5. Conclusions

In this paper, a novel end-to-end unsupervised CD method combining the superpixel segmentation and the Siamese CD network was proposed for SAR images. Firstly, the pseudo-training samples were selected using an unsupervised pre-task. Then, under the unified framework, the superpixel segmentation network and CD network were trained end-to-end to obtain the global optimal parameters. The superpixel segmentation network generates task-adaptive superpixels and outputs features containing accurate semantic information. The Siamese CD network based on U-net was used to mine multi-scale change information. The design of the Siamese structure and the use of the joint loss function enabled the multi-temporal information to be fully mined and combined to obtain change information. Several public CD datasets were used to verify the effectiveness and robustness of our proposed method. In addition, the transfer learning experiment was designed to explore the generalization performance of the network. The experimental results prove that the proposed method performs well in compressing noise, refining boundaries, and improving the CD accuracy for small area changes. Furthermore, this paper explores the ability of the network to detect changes for the first time, which deserves further attention in future research. It would also be interesting to extend this method to CD of more complex remote sensing data or sequential data in the future.

Author Contributions: Conceptualization, J.Z. and Z.Z.; methodology, L.J.; software, L.J. and Z.Z.; validation, L.J.; formal analysis, L.J. and J.Z.; investigation, L.J. and J.Z.; resources, Z.Z.; writing—original draft preparation, L.J.; writing—review and editing, J.Z.; visualization, L.J.; supervision, J.Z. and Z.Z.; project administration, Z.Z.; funding acquisition, Z.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key R&D Program of China, grant number 2022YFB3902605 and 2022YFB3901604; the Fundamental Research Funds for the Central Universities, grant number XJ2021005601; the Natural Science Foundation of China (NSFC), grant number 41901286; the Jiangsu Provincial Double-Innovation Doctor Program (2022), grant number 2022140923061; and CSAM FUNDING, grant number AR2206.

Data Availability Statement: Not applicable.

**Acknowledgments:** The author would like to thank the providers of public datasets and codes for their efforts. They would also like to thank the editors and anonymous reviewers for their helpful comments and suggestions.

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

# References

- 1. Brisco, B.; Schmitt, A.; Murnaghan, K.; Kaya, S.; Roth, A. SAR polarimetric change detection for flooded vegetation. *Int. J. Digit. Earth* **2013**, *6*, 103–114. [CrossRef]
- Bruzzone, L.; Serpico, S.B. An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* 1997, 35, 858–867. [CrossRef]
- 3. Yousif, O.; Ban, Y. Improving SAR-Based Urban Change Detection by Combining MAP-MRF Classifier and Nonlocal Means Similarity Weights. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 4288–4300. [CrossRef]
- 4. Hu, H.; Ban, Y. Unsupervised Change Detection in Multitemporal SAR Images Over Large Urban Areas. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2014, 7, 3248–3261. [CrossRef]
- 5. Hame, T.; Heiler, I.; San Miguel-Ayanz, J. An unsupervised change detection and recognition system for forestry. *Int. J. Remote Sens.* **1998**, *19*, 1079–1099. [CrossRef]
- SINGH, A. Review Article Digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* 1989, 10, 989–1003. [CrossRef]
- 7. Dekker, R.J. Speckle filtering in satellite SAR change detection imagery. Int. J. Remote Sens. 1998, 19, 1133–1146. [CrossRef]
- 8. Yousif, O.; Ban, Y. Improving Urban Change Detection From Multitemporal SAR Images Using PCA-NLM. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2032–2041. [CrossRef]
- Hui, Z.; Wang, J.G. A SAR Image Change Detection Algorithm Based on Principal Component Analysis. J. Electron. Inf. Technol. 2008, 30, 1727–1730.
- 10. Liu, S.; Du, Q.; Tong, X.; Samat, A.; Bruzzone, L.; Bovolo, F. Multiscale Morphological Compressed Change Vector Analysis for Unsupervised Multiple Change Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2017, *10*, 4124–4137. [CrossRef]

- 11. Saha, S.; Bovolo, F.; Bruzzone, L. Building Change Detection in VHR SAR Images via Unsupervised Deep Transcoding. *IEEE Trans. Geosci. Remote Sens.* 2021, 59, 1917–1929. [CrossRef]
- 12. Gong, M.; Li, Y.; Jiao, L.; Jia, M.; Su, L. SAR change detection based on intensity and texture changes. *ISPRS J. Photogramm. Remote Sens.* 2014, 93, 123–135. [CrossRef]
- 13. Rowe, N.C.; Grewe, L.L. Change detection for linear features in aerial photographs using edge-finding. *IEEE Trans. Geosci. Remote Sens.* 2001, *39*, 1608–1612. [CrossRef]
- 14. Ma, X.; Liu, S.; Hu, S.; Geng, P.; Liu, M.; Zhao, J. SAR image edge detection via sparse representation. *Soft Comput.* **2018**, 22, 2507–2515. [CrossRef]
- Ma, W.; Yang, H.; Wu, Y.; Xiong, Y.; Hu, T.; Jiao, L.; Hou, B. Change Detection Based on Multi-Grained Cascade Forest and Multi-Scale Fusion for SAR Images. *Remote Sens.* 2019, 11, 142. [CrossRef]
- Mastro, P.; Masiello, G.; Serio, C.; Pepe, A. Change Detection Techniques with Synthetic Aperture Radar Images: Experiments with Random Forests and Sentinel-1 Observations. *Remote Sens.* 2022, 14, 3323. [CrossRef]
- 17. Manzoni, M.; Monti-Guarnieri, A.; Molinari, M.E. Joint exploitation of spaceborne SAR images and GIS techniques for urban coherent change detection. *Remote Sens. Environ.* **2021**, 253, 112152. [CrossRef]
- 18. Kittler, J.; Illingworth, J. Minimum error thresholding. *Pattern Recognit*. **1986**, *19*, 41–47. [CrossRef]
- 19. Ghosh, A.; Mishra, N.S.; Ghosh, S. Fuzzy clustering algorithms for unsupervised change detection in remote sensing images. *Inf. Sci.* **2011**, *181*, 699–715. [CrossRef]
- 20. Krinidis, S.; Chatzis, V. A Robust Fuzzy Local Information C-Means Clustering Algorithm. *IEEE Trans. Image Process.* **2010**, *19*, 1328–1337. [CrossRef]
- Gou, S.; Yu, T. Graph based SAR images change detection. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 2152–2155.
- 22. Walter, V. Object-based classification of remote sensing data for change detection. *ISPRS J. Photogramm. Remote Sens.* 2004, *58*, 225–238. [CrossRef]
- Desclée, B.; Bogaert, P.; Defourny, P. Forest change detection by statistical object-based method. *Remote Sens. Environ.* 2006, 102, 1–11. [CrossRef]
- 24. Bontemps, S.; Bogaert, P.; Titeux, N.; Defourny, P. An object-based change detection method accounting for temporal dependences in time series with medium to coarse spatial resolution. *Remote Sens. Environ.* **2008**, *112*, 3181–3191. [CrossRef]
- Zhang, H.; Lin, M.; Yang, G.; Zhang, L. ESCNet: An End-to-End Superpixel-Enhanced Change Detection Network for Very-High-Resolution Remote Sensing Images. *IEEE Trans. Neural Netw. Learn. Syst.* 2021. online ahead of print. [CrossRef]
- Sui, H.; Feng, W.; Wenzhuo, L.I.; Sun, K.; Chuan, X.U. Review of Change Detection Methods for Multi-temporal Remote Sensing Imagery. Wuhan Daxue Xuebao Xinxi Kexue BanGeomatics Inf. Sci. Wuhan Univ. 2018, 43, 1885–1898.
- 27. Chen, G.; Hay, G.J.; Carvalho, L.M.T.; Wulder, M.A. Object-based change detection. *Int. J. Remote Sens.* 2012, 33, 4434–4457. [CrossRef]
- Ullah, Z.; Usman, M.; Latif, S.; Gwak, J. Densely attention mechanism based network for COVID-19 detection in chest X-rays. *Sci. Rep.* 2023, 13, 261. [CrossRef]
- 29. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* 2015, arXiv:1505.04597.
- Ullah, Z.; Usman, M.; Jeon, M.; Gwak, J. Cascade multiscale residual attention CNNs with adaptive ROI for automatic brain tumor segmentation. *Inf. Sci.* 2022, 608, 1541–1556. [CrossRef]
- Ullah, Z.; Usman, M.; Gwak, J. MTSS-AAE: Multi-task semi-supervised adversarial autoencoding for COVID-19 detection based on chest X-ray images. *Expert Syst. Appl.* 2023, 216, 119475. [CrossRef]
- 32. 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] [PubMed]
- Daudt, R.C.; Le Saux, B.; Boulch, A.; Gousseau, Y. Urban Change Detection for Multispectral Earth Observation Using Convolutional Neural Networks. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 2115–2118.
- Caye Daudt, R.; Le Saux, B.; 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.
- Chen, J.; Yuan, Z.; Peng, J.; Chen, L.; Huang, H.; Zhu, J.; Liu, Y.; Li, H. DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2021, 14, 1194–1206. [CrossRef]
- 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]
- Shi, C.; Zhang, Z.; Zhang, W.; Zhang, C.; Xu, Q. Learning Multiscale Temporal—Spatial—Spectral Features via a Multipath Convolutional LSTM Neural Network for Change Detection With Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* 2022, 60, 1–16. [CrossRef]
- Ji, L.; Zhao, Z.; Huo, W.; Zhao, J.; Gao, R. Evaluation of Several Fully Convolutional Networks in Sar Image Change Detection. In Proceedings of the ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences; Copernicus GmbH: Hannover, Germany, 2022; Volume X-3-W1-2022, pp. 61–68.

- 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]
- Fang, S.; Li, K.; Shao, J.; Li, Z. SNUNet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images. *IEEE Geosci. Remote Sens. Lett.* 2022, 19, 1–5. [CrossRef]
- Wang, Y.; Gao, L.; Hong, D.; Sha, J.; Liu, L.; Zhang, B.; Rong, X.; Zhang, Y. Mask DeepLab: End-to-end image segmentation for change detection in high-resolution remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* 2021, 104, 102582. [CrossRef]
- 42. Liu, T.; Yang, L.; Lunga, D. Change detection using deep learning approach with object-based image analysis. *Remote Sens. Environ.* **2021**, 256, 112308. [CrossRef]
- 43. 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, 2658–2673. [CrossRef]
- 44. Sun, Y.; Lei, L.; Guan, D.; Kuang, G. Iterative Robust Graph for Unsupervised Change Detection of Heterogeneous Remote Sensing Images. *IEEE Trans. Image Process.* **2021**, *30*, 6277–6291. [CrossRef]
- 45. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [CrossRef]
- 46. Jampani, V.; Sun, D.; Liu, M.-Y.; Yang, M.-H.; Kautz, J. Superpixel Sampling Networks. arXiv 2018, arXiv:1807.10174.
- 47. Rignot, E.J.M.; van Zyl, J.J. Change detection techniques for ERS-1 SAR data. *IEEE Trans. Geosci. Remote Sens.* **1993**, *31*, 896–906. [CrossRef]
- Gao, F.; Dong, J.; Li, B.; Xu, Q. Automatic Change Detection in Synthetic Aperture Radar Images Based on PCANet. *IEEE Geosci. Remote Sens. Lett.* 2016, 13, 1792–1796. [CrossRef]
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of the Advances in Neural Information Processing Systems*; Pereira, F., Burges, C.J., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2012; Volume 25.
- Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Stoyanov, D., Taylor, Z., Carneiro, G., Syeda-Mahmood, T., Martel, A., Maier-Hein, L., Tavares, J.M.R.S., Bradley, A., Papa, J.P., Belagiannis, V., et al., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 3–11.
- 51. Gao, F.; Wang, X.; Gao, Y.; Dong, J.; Wang, S. Sea Ice Change Detection in SAR Images Based on Convolutional-Wavelet Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1240–1244. [CrossRef]
- 52. Gao, F.; Dong, J.; Li, B.; Xu, Q.; Xie, C. Change detection from synthetic aperture radar images based on neighborhood-based ratio and extreme learning machine. *J. Appl. Remote Sens.* **2016**, *10*, 046019. [CrossRef]
- 53. Li, Y.; Peng, C.; Chen, Y.; Jiao, L.; Zhou, L.; Shang, R. A Deep Learning Method for Change Detection in Synthetic Aperture Radar Images. *IEEE Trans. Geosci. Remote Sens.* 2019, *57*, 5751–5763. [CrossRef]
- 54. Liu, F.; Jiao, L.; Tang, X.; Yang, S.; Ma, W.; Hou, B. Local Restricted Convolutional Neural Network for Change Detection in Polarimetric SAR Images. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 818–833. [CrossRef]
- 55. Gao, Y.; Gao, F.; Dong, J.; Wang, S. Change Detection from Synthetic Aperture Radar Images Based on Channel Weighting-Based Deep Cascade Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 4517–4529. [CrossRef]
- Gao, Y.; Gao, F.; Dong, J.; Du, Q.; Li, H.-C. Synthetic Aperture Radar Image Change Detection via Siamese Adaptive Fusion Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2021, 14, 10748–10760. [CrossRef]
- Zhang, X.; Su, H.; Zhang, C.; Gu, X.; Tan, X.; Atkinson, P.M. Robust unsupervised small area change detection from SAR imagery using deep learning. *ISPRS J. Photogramm. Remote Sens.* 2021, 173, 79–94. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.