remote sensing

Article

Novel Spatial–Spectral Channel Attention Neural Network for Land Cover Change Detection with Remote Sensed Images

Xu Yang [1], Zhiyong Lv [2,*], Jón Atli Benediktsson [3] and Fengrui Chen [4]

[1] College of Geological Engineering and Geomatics, Chang'an University, Xi'an 710054, China
[2] School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China
[3] Faculty of Electrical and Computer Engineering, University of Iceland, IS 107 Reykjavik, Iceland
[4] Key Laboratory of Geospatial Technology for Middle and Lower Yellow River Regions, Henan University, Ministry of Education, Kaifeng 475001, China
* Correspondence: zhiyonglyu@xaut.edu.cn

Abstract: Land cover change detection (LCCD) with remote-sensed images plays an important role in observing Earth's surface changes. In recent years, the use of a spatial-spectral channel attention mechanism in information processing has gained interest. In this study, aiming to improve the performance of LCCD with remote-sensed images, a novel spatial-spectral channel attention neural network (SSCAN) is proposed. In the proposed SSCAN, the spatial channel attention module and convolution block attention module are employed to process pre- and post-event images, respectively. In contrast to the scheme of traditional methods, the motivation of the proposed operation lies in amplifying the change magnitude among the changed areas and minimizing the change magnitude among the unchanged areas. Moreover, a simple but effective batch-size dynamic adjustment strategy is promoted to train the proposed SSCAN, thus guaranteeing convergence to the global optima of the objective function. Results from comparative experiments of seven cognate and state-of-the-art methods effectively demonstrate the superiority of the proposed network in accelerating the network convergence speed, reinforcing the learning efficiency, and improving the performance of LCCD. For example, the proposed SSCAN can achieve an improvement of approximately 0.17–23.84% in OA on Dataset-A.

Keywords: change detection; deep learning; attention module; remote-sensed images

check for updates

Citation: Yang, X.; Lv, Z.; Atli Benediktsson, J.; Chen, F. Novel Spatial–Spectral Channel Attention Neural Network for Land Cover Change Detection with Remote Sensed Images. *Remote Sens.* **2023**, *15*, 87. https://doi.org/10.3390/rs15010087

Academic Editors: Olga Sykioti, Gangyao Kuang and Xin Su

Received: 18 November 2022
Revised: 13 December 2022
Accepted: 20 December 2022
Published: 23 December 2022

1. Introduction

Obtaining land cover change detection (LCCD) with bitemporal remote-sensed images is important for monitoring geological disasters [1,2], evaluating the health of ecosystems [3], assisting the determination of urban development [4], capturing forest large-scale deformation [5], and land-use management [6]. To date, various LCCD methods have been developed and applied in practical applications, such as change detection with serial long-term Landsat images [7], change detection with synthetic aperture radar images [8], high-resolution optical images [9], hyperspectral images [10,11], change detection with heterogeneous images [12–16], and pixel/object-based change-detection approaches [17]. However, given the uncertainty factors involved in capturing a pair of bitemporal images for LCCD, including imaging atmospheric conditions and phenological differences, achieving LCCD with bitemporal remote-sensed images remains a challenge, and improvements are required for practical applications [18].

In recent years, deep-learning techniques have achieved profound success in the domain of remote-sensing image applications [19–21], especially LCCD [22–24]. In the process of distinguishing between changed and unchanged areas, deep-learning-based LCCD methods automatically discover and learn complicated, hierarchical, and nonlinear features from a raw dataset, with the motivation of overcoming the limitations of traditional methods [25]. For example, Yang et al. [26] proposed an unsupervised change-detection approach based on the guidance of time distance, which is effective for change detection in irregularly collected images. LCCD with heterogeneous remote-sensed images that are acquired by different remote-sensed sensors is extremely popular in practical applications [21]. For example, Lv et al. [27] proposed a simple but effective neural network for change detection with heterogeneous remote-sensed images. Consequently, deep-learning-based LCCD methods have also become increasingly investigated, attracting major attention and yielding good results in the remote-sensing sector.

One of the most popular deep-learning techniques for LCCD with remote-sensed images is the convolutional neural network (CNN)-based approach. Daudt et al. [28] proposed two classical Siamese extensions of fully CNN networks for change detection, which is a simple but effective approach on any available change-detection dataset. Inspired by the work of Daudt et al. [28], many similar studies based on CNN and a Siamese structure have been promoted for LCCD [29–31]. In addition, learning robust features by means of CNN for LCCD is helpful to address the uncertain difference in bitemporal remote-sensed images. For instance, Zhang et al. [32] used CNN to learn the deep features from remote-sensed images and performed transfer learning to compose a two-channel network and subsequently generate multiscale and multi-depth difference feature maps for change detection. Mou et al. [33] learned spectral-spatial-temporal features by using a recurrent CNN for change detection with multispectral images. From the abovementioned literature review, CNNs have been widely used as a basic but classical artificial neural network; the motivation for CNN being used for LCCD lies in learning deep features for smoothing pseudo changes and reducing noise from bitemporal images [34,35].

Apart from the classical CNN-based approaches for LCCD, various modified CNNs have been promoted to enhance the performance of LCCD. For example, multiscale feature extraction with CNN was proposed to explore other targets of various shapes and sizes for LCCD [36–38]. Enhancing a sample is also simple and intuitive in the performance improvement of CNN-based LCCD methods, such as Gao et al.'s [39] proposed convolutional-wavelet neural network (CWNN). Moreover, Ji et al. [40] suggested creating simulated samples and coupling them with CNN to achieve building change detection. Numerous studies have also recommended fusing the traditional image processing approach with CNN owing to its merits of improving the performance of LCCD. For example, Zhang et al. [41] promoted a deep supervised information fusion network (DSIFN) for change detection with bitemporal images. Lee et al. [42] developed a local similarity Siamese network for achieving LCCD in an urban area. Wu et al. [43], aiming to avoid the requirement of annotated samples to train a deep-learning network, proposed an unsupervised model called the kernel PCA convolutional mapping network (KPCA_CMN) for LCCD with VHR remote-sensed images. The abovementioned reviewed literature indicates that various methods based on CNNs have been developed and widely applied for achieving LCCD. However, no method can be marked as "good" or "bad." An improvement space for LCCD may still exist in practical applications [25,34,44].

In recent years, the attention mechanism embedded in neural networks has become popular for achieving LCCD tasks [34]. In these approaches, attention modules are embedded in the different neural networks to strengthen the network by paying extra attention to the changed area. For example, Liu et al. [45] introduced a dual attention module to exploit interdependencies between spectral channels and spatial positions, subsequently enhancing feature representation for capturing changes. Fang et al. [46] embedded a channel attention module in a densely connected Siamese network for change detection. Shi et al. [47] proposed a deeply supervised attention metric-based network (DSAMNet) for LCCD with aerial images. Liu et al. [48] promoted a super resolution-based change-detection network with a stacked attention module. Many of the aforementioned studies indicate that the motivation of the attention mechanism is for the network to enhance its focus on small but important parts of the data. For the specific task of detecting land cover change with remote-sensed images, a changed area can be regarded as an important part of the whole area. However, various networks for LCCD are centered on using novel attention mechanism-based neural networks. Nonetheless, learning which part of the data is more important than others depends on the network itself and the training progress.

In this study, a novel spatial-spectral channel attention neural network (SSCAN) is proposed to achieve the LCCD task for bitemporal remote-sensed images. The proposed network has two objectives: (i) to improve the detection performance with remote-sensed images and (ii) to construct a simple but effective training strategy for enhancing the learning progress. On this basis, the proposed SSCAN is designed as follows: First, an encoder-decoder subnetwork is embedded for extracting deep features from bitemporal images. Then, pre- and post-event images are fed into the spatial channel attention module (SCAM) and convolution block attention module (CBAM), respectively, to amplify the change magnitude between change areas and reduce the change magnitude between unchanged areas. Finally, an optimized training strategy is constructed to train the network and obtain a trained module for prediction. The main contributions of this study can be summarized as follows:

(i) A novel neural network, SSCAN, is proposed to enhance LCCD with remote-sensed images, including images with very high spatial resolution and low median resolution. Moreover, the results obtained by the proposed SSCAN are superior to those of other approaches under limited training sample scenarios.
(ii) In the proposed SSCAN, two typical modules, namely, SCAM and CBAM, are combined to extract the different deep features from the pre- and post-event images, respectively. The aims of this combined operation lie in amplifying the change magnitude between changed areas and reducing the change magnitude between unchanged areas, which is beneficial for smoothing noise and enhancing change-detection performance.
(iii) A batch-size dynamic adjustment (BDA) strategy is promoted to train the proposed SSCAN. This strategy can improve the convergence speed of the proposed neural network. If we assume that gradual learning is more effective than stepwise learning, then training a neural network with BDA also allows for learning to gradually progress.

The remainder of this article is organized as follows: Section 2 presents the details of our proposed neural network. Section 3 describes the comparative experiments that were conducted to verify the performance and superiority of the proposed neural network. The conclusion is drawn in Section 4.

2. Materials and Methods

This section provides a brief overview of the proposed SSCAN. Then, it presents a detailed description for each main part of the proposed SSCAN.

2.1. Overview

The proposed SSCAN (Figure 1) consists of three parts: convolutional autoencoders (CAEs), SCAM, and CBAM. On the basis of the backbone of U-Net, CAEs are used to extract deep features from bitemporal images. Then, SCAM and CBAM are applied crossly to amplify the change magnitude between bitemporal images in terms of changed areas. Subsequently, a cross-entropy loss function and a dynamic batch-size adjustment are used to train the proposed model. Finally, the Argmax function is used to determine the label of each pixel and output a binary change-detection map.

2.2. CAEs

CAEs are designed for deep feature extraction in our proposed SSCAN because they can obtain robust features without the need for additional training; furthermore, robust features are beneficial for subsequent change detection [49,50]. Apart from the broad applications of autoencoders, the input size of CAEs is flexible, and the structure of CAEs is similar to the classical autoencoder (i.e., they are symmetrical). In our proposed SSCAN, the encoder and decoder consist of a series of convolutional layers, and the number of convolutional layers depends on the inputs of the network.

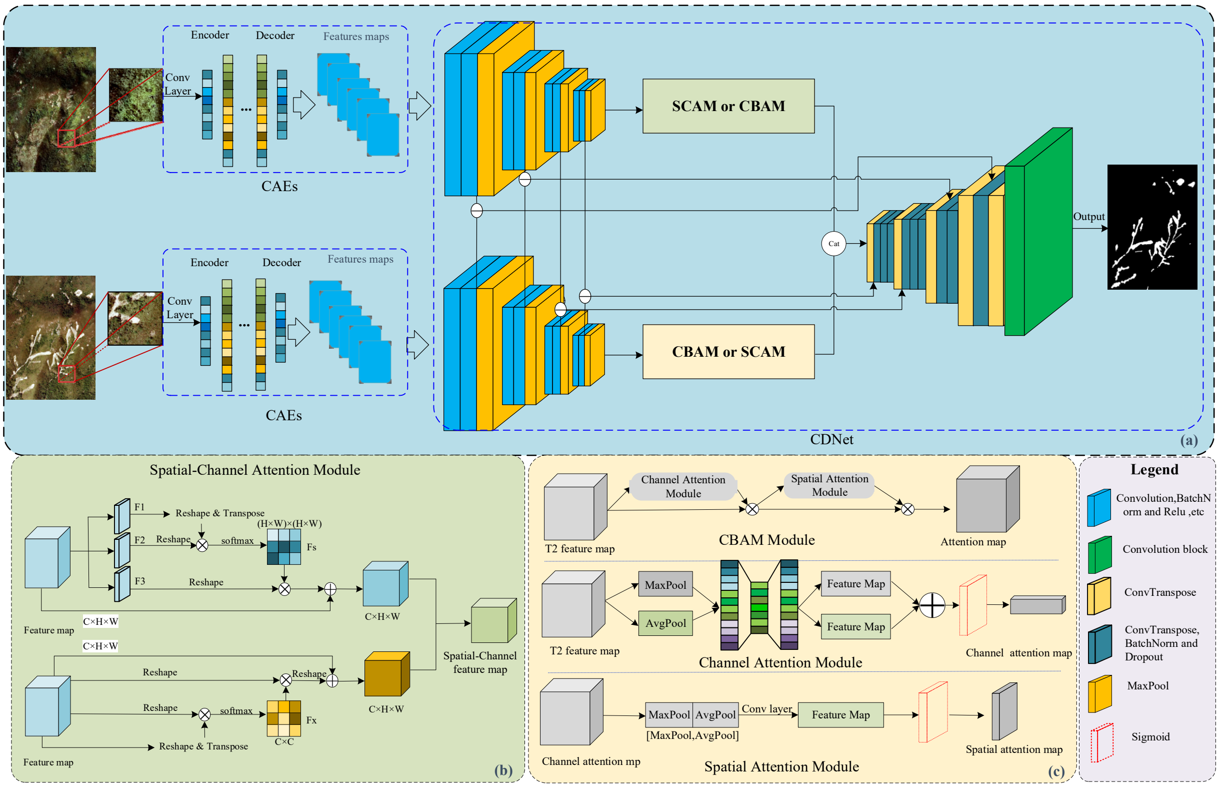

Figure 1. (**a**) Flowchart of the proposed SSCAN; (**b**) SCAM, and (**c**) CBAM.

2.3. Different Attention Modules

After extracting the deep features from bitemporal images with CAEs, different attention modules are designed in our proposed SSCAN network to refine the deep feature maps and capture the change areas. Attention plays an important role in human perception [51], and one of the important properties of attention mechanisms is capturing the structure of the whole image and focusing on the target area [52]. In our proposed SSCAN network, the motivation for using different attention modules for each data item of bitemporal images lies in amplifying the change magnitude between changed areas while simultaneously reducing the change distance between unchanged areas. The details of this scheme can be described as follows:

SCAM: This module was first used for scene segmentation in [53]. As illustrated in Figure 1b, SCAM contains two branches, namely, spatial attention and channel attention. In the promoted module, a latent feature $f \in R^{C\times H\times W}$ is first fed into the convolution layer. Then, the convolutional layer creates three new feature maps denoted as $f_A \in R^{C\times H\times W}$, $f_B \in R^{C\times H\times W}$, and $f_C \in R^{C\times H\times W}$. The new feature is reshaped into $f_A \in R^{C\times N}$, $f_B \in R^{C\times N}$, and $f_C \in R^{C\times N}$, where $N = W \times H$. Consequently, a matrix is multiplied between $f_B \in R^{C\times N}$ and $f_C \in R^{C\times N}$, and a softmax layer is applied to generate the spatial attention feature map $f_S \in R^{N\times N}$. Each value of f_S is calculated by $f_{S_{j,i}} = \frac{\exp\left(f_{B_i} \cdot f_{C_j}\right)}{\sum_{i=1}^{i=N} \exp\left(f_{B_i} \cdot f_{C_j}\right)}$, where $f_{S_{j,i}}$ measures the relationship between the i^{th} position and the j^{th} position; a greater $f_{S_{j,i}}$ means a tighter correlation between them. Furthermore, $f_C \in R^{C\times N}$ is multiplied with a matrix to reshape the result, $R^{C\times N\times W}$. Finally, a scale parameter α is employed to multiply $f_C \in R^{C\times N\times W}$ and sum up the original feature $f \in R^{C\times H\times W}$ to obtain the final feature $E_f \in R^{C\times H\times W}$.

$$E_{f_j} = \alpha \sum_{i=1}^{i=N} f_{S_{j,i}} f_{C_i} + f_j \tag{1}$$

where α is a hyperparameter initialized as 0; it gradually learns to assign a dynamic weight. According to Equation (1), the final feature E_f is constructed by a position weight map (f_S) and original feature (f); therefore, it can give an overview of the global context and selectively aggregate context according to the spatial attention map. The changed and unchanged areas gradually achieve similar gains based on the guide of the training sample set. Consequently, the homogeneity of the intraclass (changed or unchanged areas) is improved.

CBAM: If SCAM processes the pre-event image in our proposed SSCAN, then CBAM processes the latent feature of the post-event image. As shown in Figure 1c, given a latent feature $f \in R^{C\times H\times W}$ as input, CBAM is sequentially composed of a channel attention submodule and a spatial attention submodule. Here, $M_C \in R^{C\times 1\times 1}$ and $M_S \in R^{1\times H\times W}$ symbolize the channel and spatial attention maps, respectively. The whole attention process can be expressed as follows:

$$\begin{aligned} f' &= M_c(f) \otimes f \\ f'' &= M_s(f') \otimes f' \end{aligned} \tag{2}$$

where $\otimes$ denotes elementwise multiplications, and $M_C(\cdot)$ is given by

$$\begin{aligned} M_c(f) &= \sigma(MLP(AvgPool(f))) + MLP(MaxPool(f)) \\ &= \sigma\left(W_1\left(W_0\left(f^c_{avg}\right)\right) + W_1(W_0(f^c_{max}))\right) \end{aligned} \tag{3}$$

where σ is the sigmoid function, and W_1 and W_0 are the weights of the multilayer perceptron. In our proposed approach, W_1 and W_0 are shared by both inputs and the ReLu activation function. Distinct from the motivation of the channel attention submodule for exploiting the interchannel relationships of input features, the spatial attention submodule concentrates on mining the interspatial relationship of features. $M_C(\cdot)$ is given by

$$\begin{aligned} M_s(f) &= \sigma\left(f^{3\times 3}([AvgPool(f); MaxPool(f)])\right) \\ &= \sigma\left(f^{3\times 3}\left(\left[f^s_{Avg}; f^s_{max}\right]\right)\right) \end{aligned} \tag{4}$$

where σ is the sigmoid function, and $f^{3\times 3}$ denotes a convolution operation with a 3×3 filter. Two 2D maps ($f^s_{avg} \in R^{1\times H\times W}$ and $f^s_{max} \in R^{1\times H\times W}$) are generated by pooling operations. Then, these maps are concatenated and convolved by a basic convolutional layer to produce a spatial attention map $M_s(f)$.

Therefore, channel attention focuses on "what" is meaningful in an input image, while spatial attention concentrates on providing "where" the informative part of the input image is located. In the employed SCAM and CBAM, channel attention and spatial attention are complementary to each other in different ways, thereby enhancing learning performance. In the proposed SSCAM network, SCAM and CBAM are used many times to process pre- and post-event images. In theory, if input image pairs depict an unchanged area, then SCAM and CBAM will focus on the same position and similar information, and the change magnitude will be narrowed. By contrast, if input image pairs depict a changed area, then the change magnitude will be amplified with the different attention maps from SCAM and CBAM.

2.4. Loss Function and Training Progress

Training a neural network with a suitable function is important in performance improvement. The most popularly used loss function for binary change detection contains the softmax loss function, contrastive loss function, and cross-entropy loss function. In our proposed SSCAN, the cross-entropy loss function is selected for training because it can measure the similarity between two probability distributions. Meanwhile, change detection with remote-sensed images can be used for measuring the change magnitude between the pixel distributions of bitemporal images. The cross-entropy loss can be formulated as follows:

$$Loss = \frac{1}{N}\sum_{n=1}^{n=N}[y_n \log \hat{y}_n + (1 - y_n \log(1 - \hat{y}_n))] \tag{5}$$

where N is the total number of training samples, y_n is equal to 0 or 1 (unchanged or changed status, respectively), and $\hat{y}_n$ is the corresponding prediction for an unchanged or changed label.

Batch size is one of the most important hyperparameters used for tuning modern deep-learning systems, and many studies have demonstrated its obvious effect on learning performance [54–56]. In further optimizing the training progress of the proposed SSCAN, a simple but effective strategy called BDA is applied in the training of the proposed network. In BDA, $b_{i \times base} = b_i + b_i \times 0.5$, where b_i is the batch size for the *ith* epoch, $b_{i \times base}$ is the next $(i \times base)$ epoch in the iterative training progress, b_i is initialized with a small value of 4, and the default value of *base* is close to 30. The BDA formula can be explained intuitively, as smaller batch sizes allow the model to start learning before viewing all of the data. Then, the batch size is increased gradually to allow the model to capture the global information and converge to the global optima.

3. Experiment

Two experiments were designed to verify the performance of the proposed SSCAN: (i) an experimental investigation of the superiority of the proposed approach over seven cognate and state-of-the-art methods, namely, FC_EF [28], FC_Conc [28], FC_Diff [28], CWNN [39], DSIFN [41], KPCA_CMN [43], and DSAMNet [47], and (ii) an ablation experiment with the view of promoting the widespread use of the proposed SSCAN. The details of the experiments can be summarized as follows.

3.1. Dataset Description

Four pairs of remote-sensed images for change detection were used in the experiment (Figure 2). The details of these datasets are presented below, and the description for each dataset is summarized in Table 1.

Table 1. Descriptive summary for each dataset.

Data	Size	Spatial Resolution	Change Events	Acquisition Date	Location
Dataset-A	1117 × 803	0.5 m/pixel	Landslide change	September 2017 and October 2019	Hong Kong, China
Dataset-B	694 × 754	0.5 m/pixel	Landslide change	September 2017 and October 2019	Hong Kong, China
Dataset-C	1250 × 950	0.62 m/pixel	Land-use change	June 2000 and December 2005	JiNan City, ShanDong Province, China
Dataset-D	400 × 400	30.0 m/pixel	Crop change	August 2001 and August 2002	Liaoning, China

First, we considered two pairs of remote-sensed images with very high spatial resolution and denoted them as Dataset-A and Dataset-B. As shown in Figure 2, these datasets represent the aerial orthophotos with a resolution of 0.5 m/pixel, and they depict landslide change events in Lantau Island, Hong Kong, China. The landslide occurred in a mountain area covered by forested and outcrop rock, a situation that typically hinders landslide inventory mapping with binary change-detection techniques.

The third dataset (Dataset-C) refers to a land-use change event that occurred in a countryside area in JiNan City, ShanDong Province, China. These images were acquired by the QuickBird satellite with a resolution of 0.62 m/pixel. As shown in Figure 2, the pre- and post-event images differ considerably in phenology seasons, which may cause pseudochange in the change-detection results.

The fourth dataset (Dataset-D) depicts the land-cover change events in a crop area. The two scenes were acquired by Landsat-7 Enhanced Thematic Mapper Plus (ETM+) in August 2001 and August 2002 in Liaoning Province of China. The ground reference map was obtained manually (Figure 2d).

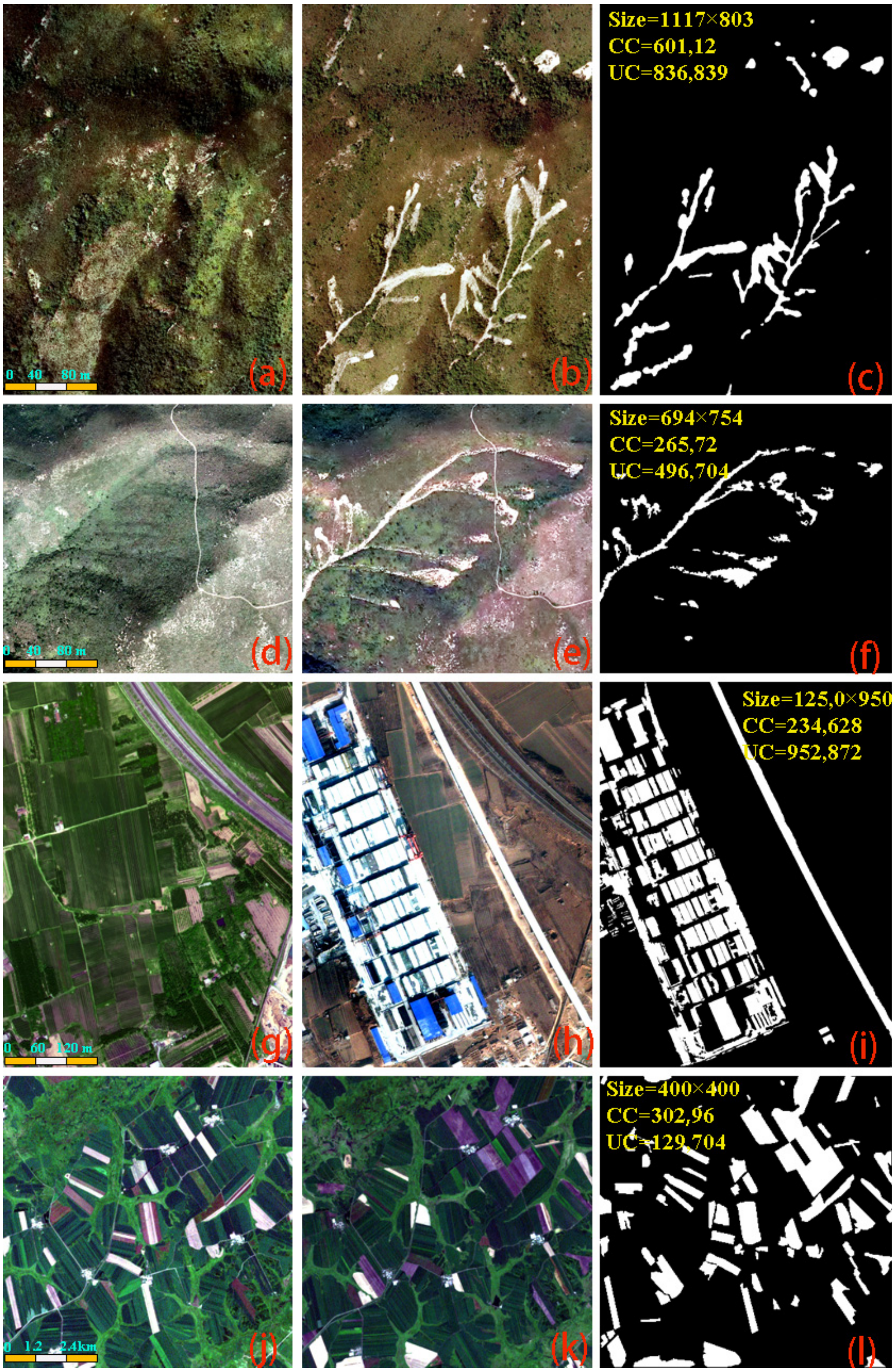

Figure 2. Testing datasets: (**a**–**c**) are pre-event image, post-event image, and the ground reference map for Dataset-A, respectively; (**d**–**f**) are pre-event image, post-event image, and the ground reference map for Dataset-B, respectively; (**g**–**i**) are pre-event image, post-event image, and the ground reference map for Dataset-C, respectively; (**j**–**l**) are pre-event image, post-event image, and the ground reference map for Dataset-D, respectively.

In addition, seven popular measurements were adopted to quantitatively evaluate the performance of each approach for comparison: overall accuracy (OA), average accuracy (AA), kappa coefficient (Ka), false alarm (FA), missing alarm (MA), total error (TE), precision, and F-score. Further details about these measurements can be read from [18].

3.2. Parameter Optimization and Training Samples

Seven classical and widely used change-detection methods were selected for comparison. The parameters of each approach can be detailed as follows: FC_EF [28] (epoch = 300, lr = 0.0004, lr_decay = 0.00005, batch_size = 8), FC_Siam_Conc [28] (epoch = 300, lr = 0.0004, lr_decay = 0.00005, batch_size = 8), FC_Siam_Diff [28] (epoch = 300, lr = 0.0004, lr_decay = 0.00005, batch_size = 8), CWNN [39] (Sam_num = 6000, Pos_num = 1000, epoch = 50, batch_size = 50, batch_size = 7), IFN [41], KPCA_CMN [43] (Sam_num = 100), DSAMNet [47] (epoch = 300, lr = 0.0004, lr_decay = 0.00005, batch_size = 8). In addition, a widely used approach named stochastic gradient descent was adopted for parameter optimization.

The training samples were prepared as follows: First, the bitemporal remote-sensed images and the corresponding ground reference map was divided into $n \times n$ image blocks, where n was equal to 16, 28, or 56. Second, half of the divided image block pairs were randomly selected for training a deep-learning module, whereas the other half of the divided image blocks were used to evaluate the performance of the trained deep-learning modules.

3.3. Visual Performance and Quantitative Comparison

On the basis of the abovementioned parameter settings, visual performance and quantitative comparisons were conducted.

Figures 3 and 4 show the results of applying the proposed method on Dataset-A and Dataset-B, respectively. The analysis of visual performance, which concentrated on the presentation of the detection results in terms of false alarm (cyan in the results) and missed alarm (red in the results), indicates the advantages of using the proposed SSCAN in landslide inventory mapping with change-detection techniques. For example, our proposed approach had the fewest false alarms among the compared methods. Moreover, the noise in the detection map achieved by our proposed approach was less than that of other methods. The corresponding quantitative results in Tables 2 and 3 support the visual observation conclusion. For example, apart from AA and MA for Dataset-A in Table 2, the proposed SSCAN has the best accuracy performance in terms of AA, Ka, FA, TE, precision, and F1-score.

Table 2. Comparison of other methods with the proposed approach on Dataset-A, ka $\in$ [0,1]; other values are presented as percentages (%).

Methods	OA	AA	KA	FA	MA	TE	Precision	F-Score
FC_EF [28]	98.94	97.16	0.92	0.79	4.89	1.06	89.65	92.3
FC_Siam_Conc [28]	98.61	98.17	0.90	1.32	2.33	1.39	84.18	90.42
FC_Siam_Diff [28]	96.85	92.92	0.77	2.54	11.62	3.15	71.44	79.01
CWNN [39]	75.27	70.54	0.17	24.00	34.93	24.73	16.30	26.07
IFN [41]	98.91	**98.87**	0.92	1.08	**1.19**	1.09	86.79	92.41
KPCA_CMN [43]	97.47	95.60	0.82	2.23	6.57	2.52	75.02	83.22
DSAMNet [47]	97.83	89.78	0.82	0.93	19.51	2.17	86.24	83.26
Proposed SSCAN	**99.11**	95.91	**0.93**	**0.40**	7.78	**0.89**	**94.35**	**93.27**

Apart from comparing the proposed approach with other methods in terms of achieving landslide inventory mapping tasks (Dataset-A and Dataset-B), the proposed approach was also investigated on Dataset-C, which represents the land-use change events. Figure 5 shows the comparative results of the different methods. Some of these could not obtain satisfactory detection results due to the large phenological difference between the bitemporal images. CWNN [39] and KPCA_CMN [43] incorrectly detected several pixels as changed areas, while FC_Siam_Diff [28] missed a substantial amount of changed area (red parts in Figure 5c). The proposed SSCAN clearly outperformed the other methods. Table 3 summarizes the quantitative comparative results, which convincingly support the visual comparative conclusion.

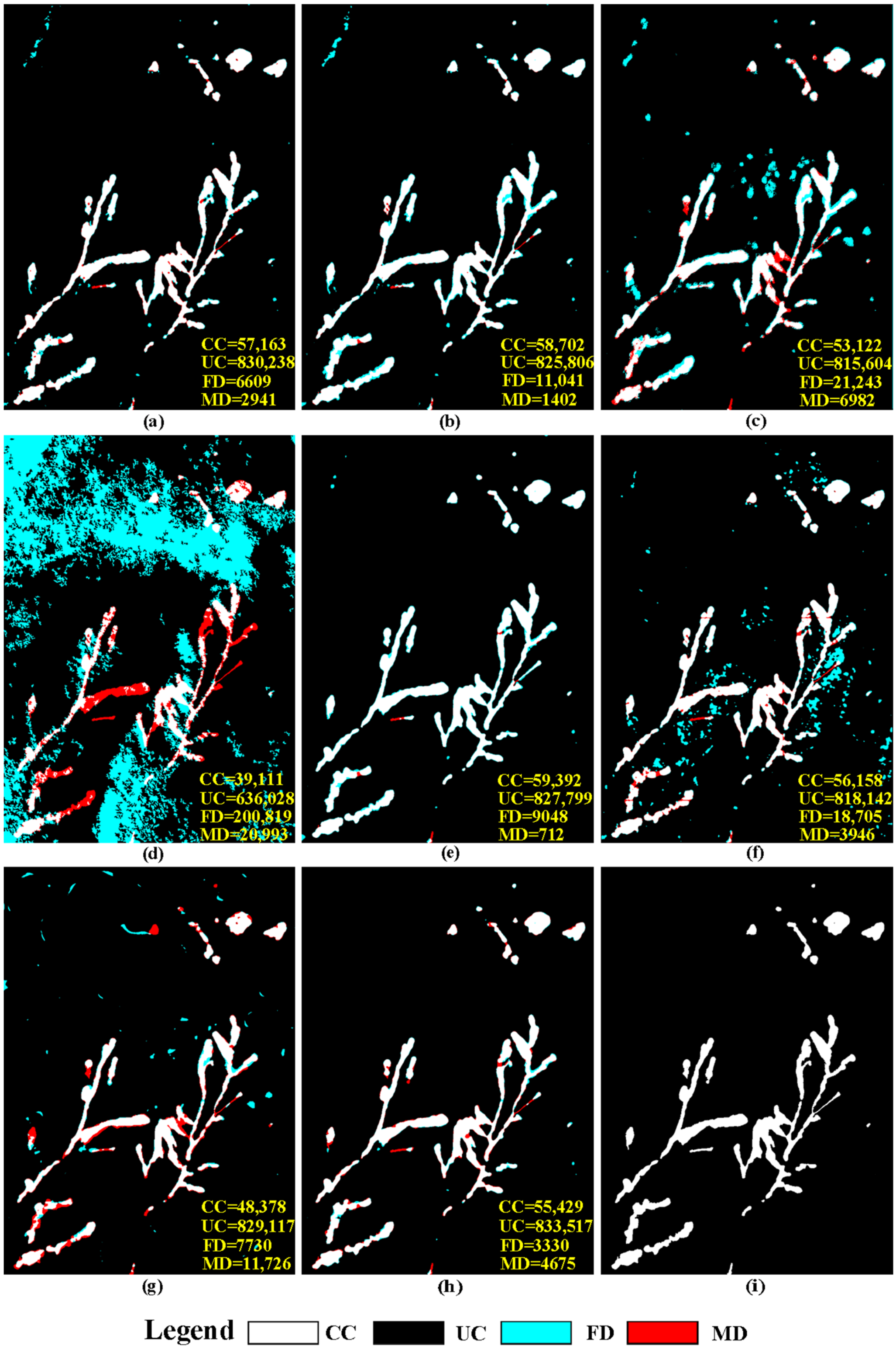

Figure 3. Binary change-detection map acquired using different methods on Dataset-A: (**a**) FC_EF [28], (**b**) FC_Siam_Conc [28], (**c**) FC_Siam_Diff [28], (**d**) CWNN [39], (**e**) IFN [41], (**f**) KPCA_CMN [43], (**g**) DSAMNet [47], (**h**) proposed method, and (**i**) ground truth. (CC: correct change; UC: unchanged; FD: false detection; MD: missed detection).

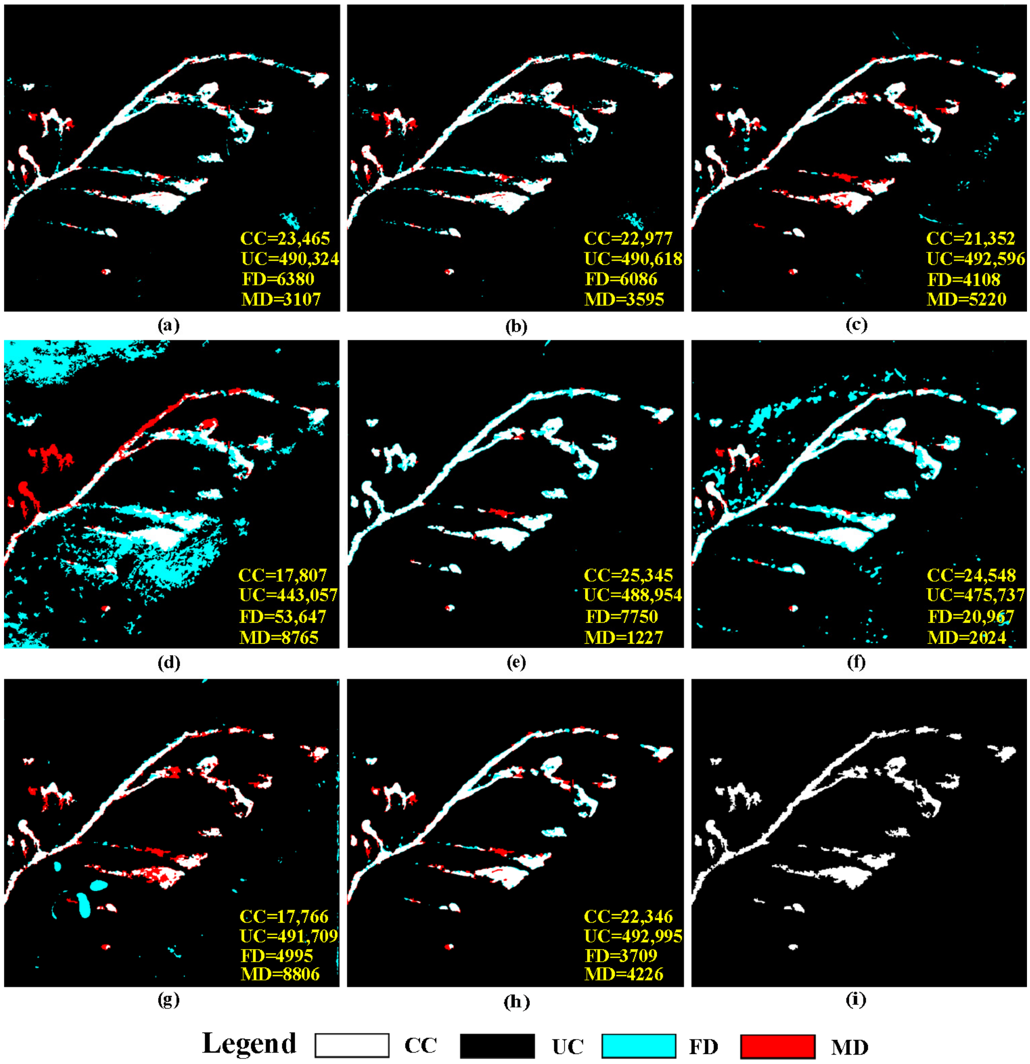

Figure 4. Binary change-detection map acquired using different methods on Dataset-B: (**a**) FC_EF [28], (**b**) FC_Siam_Conc [28], (**c**) FC_Siam_Diff [28], (**d**) CWNN [39], (**e**) IFN [41], (**f**) KPCA_CMN [43], (**g**) DSAMNet [47], (**h**) proposed method, and (**i**) ground truth. (CC: correct change; UC: unchanged; FD: false detection; MD: missed detection).

Table 3. Comparison of other methods with the proposed approach on Dataset-B, ka ∈ [0,1]; other values are presented as percentages (%).

Methods	OA	AA	KA	FA	MA	TE	Precision	F-Score
FC_EF [28]	98.19	**93.51**	0.82	1.28	11.69	1.81	78.62	83.18
FC_Siam_Conc [28]	98.15	92.62	0.82	1.23	13.53	1.85	79.06	82.60
FC_Siam_Diff [28]	98.22	89.76	0.81	0.82	19.64	1.78	83.86	82.07
CWNN [39]	88.07	78.11	0.312	10.80	32.99	11.93	24.92	36.33
IFN [41]	98.28	**96.91**	**0.84**	1.56	**4.62**	1.72	76.58	84.95
KPCA_CMN [43]	95.61	94.08	0.66	4.22	7.62	4.39	53.93	68.11
DSAMNet [47]	97.36	82.93	0.71	1.01	33.14	2.64	78.05	72.02
Proposed SSCAN	**98.48**	91.67	**0.84**	**0.75**	15.90	**1.52**	**85.76**	84.92

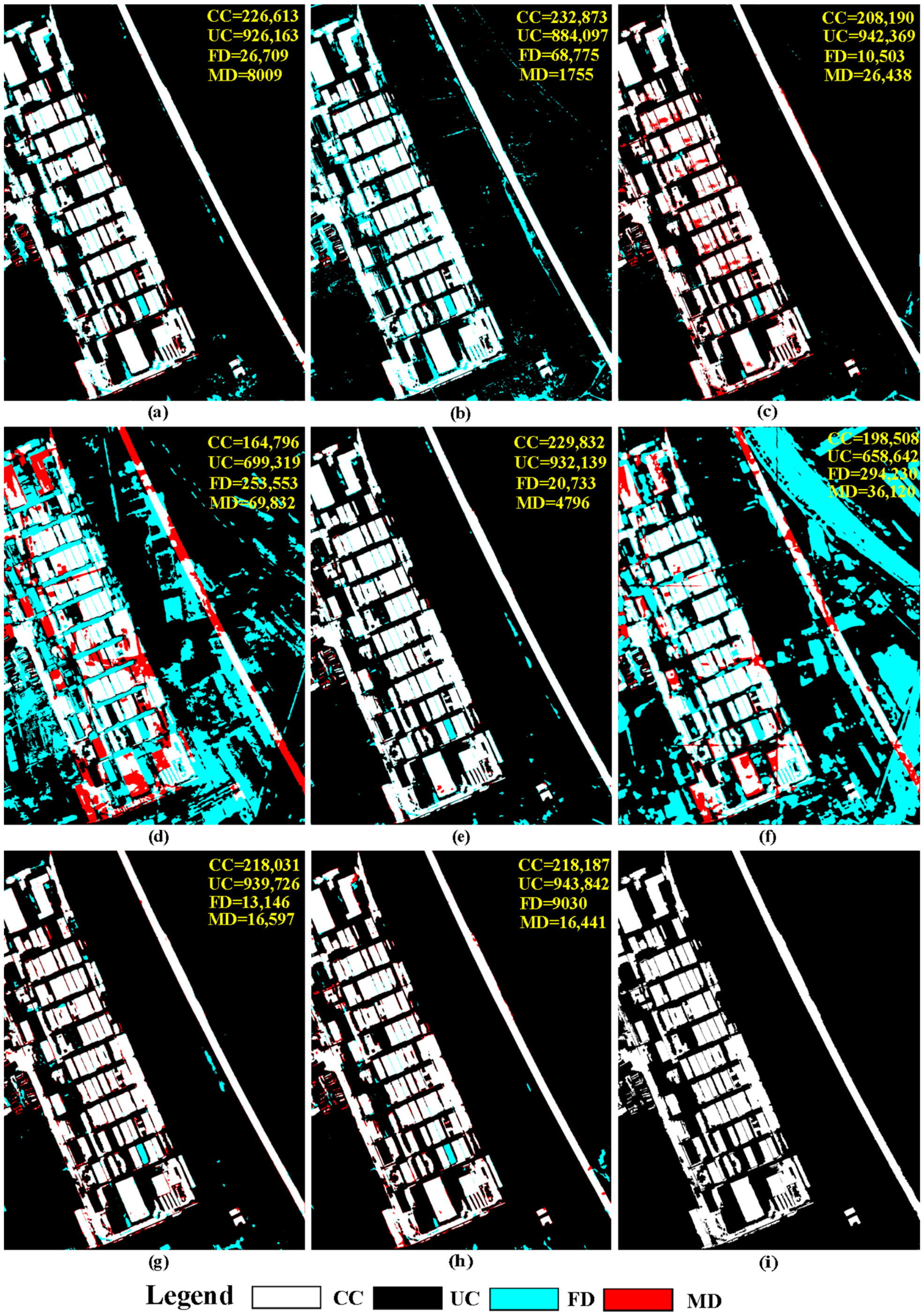

Figure 5. Binary change-detection map acquired using different methods on Dataset-C: (**a**) FC_EF [28], (**b**) FC_Siam_Conc [28], (**c**) FC_Siam_Diff [28], (**d**) CWNN [39], (**e**) IFN [41], (**f**) KPCA_CMN [43], (**g**) DSAMNet [47], (**h**) proposed method, and (**i**) ground truth. (CC: correct change; UC: unchanged; FD: false detection; MD: missed detection).

The proposed SCAN and state-of-the art methods were also compared for land-use change detection by using remote-sensed images with low median resolutions. As shown in Figure 6, our proposed approach outperformed the other methods in terms of noise, false detection pixels, and missed detection pixels, among others. The quantitative results were summarized in Tables 4 and 5.

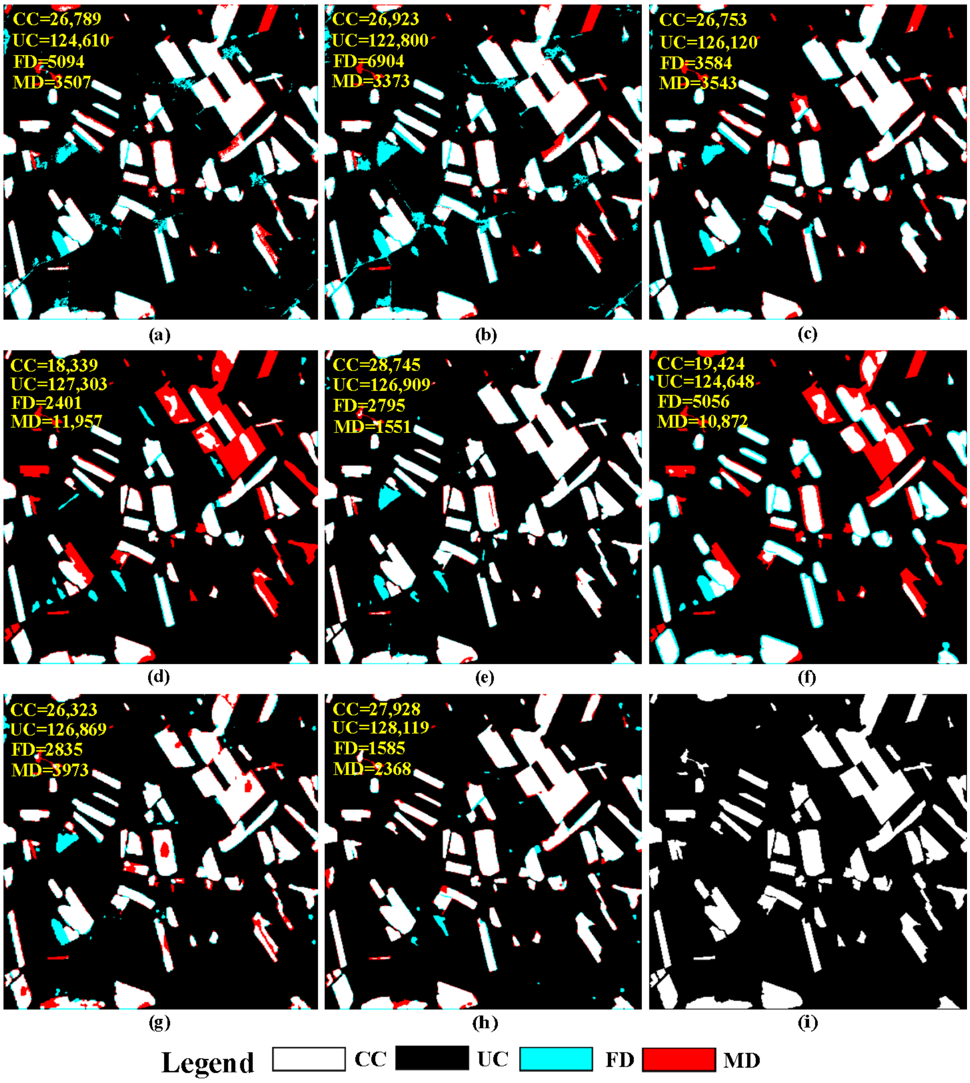

Figure 6. Binary change-detection map acquired using different methods on Dataset-D: (**a**) FC_EF [28], (**b**) FC_Siam_Conc [28], (**c**) FC_Siam_Diff [28], (**d**) CWNN [39], (**e**) IFN [41], (**f**) KPCA_CMN [43], (**g**) DSAMNet [47], (**h**) proposed method, and (**i**) ground truth. (CC: correct change; UC: unchanged; FD: false detection; MD: missed detection).

Table 4. Comparison of other methods with the proposed approach on Dataset-C, ka ∈ [0,1]; other values are presented as percentages (%).

Methods	OA	AA	KA	FA	MA	TE	Precision	F-Score
FC_EF [28]	97.08	**96.89**	0.91	2.80	3.41	2.92	89.46	92.89
FC_Siam_Conc [28]	94.06	96.02	0.83	7.22	**0.75**	5.94	77.20	86.85
FC_Siam_Diff [28]	96.89	93.81	0.90	1.10	11.27	3.11	95.20	91.85
CWNN [39]	72.77	71.81	0.34	26.61	29.76	27.23	39.39	50.48
IFN [41]	97.85	**97.89**	0.93	2.18	2.04	2.15	91.73	94.74
KPCA_CMN [43]	72.18	76.86	0.38	30.88	15.39	27.82	40.29	54.58
DSAMNet [47]	97.50	95.77	0.92	1.38	7.07	2.51	94.31	93.61
Proposed SSCAN	**97.86**	96.02	**0.93**	**0.95**	7.01	**2.15**	**96.03**	**94.48**

Table 5. Comparison of other methods with the proposed approach on Dataset-D, ka ∈ [0,1]; other values are presented as percentages (%).

Methods	OA	AA	KA	FA	MA	TE	Precision	F-Score
FC_EF [28]	94.62	92.25	0.83	3.93	11.58	5.38	84.02	86.17
FC_Siam_Conc [28]	93.58	91.77	0.80	5.32	11.13	6.43	79.59	83.97
FC_Siam_Diff [28]	95.55	92.77	0.86	2.76	11.69	4.45	88.19	88.25
CWNN [39]	91.03	79.34	0.67	**1.85**	39.47	8.97	**88.42**	71.87
IFN [41]	97.28	**96.36**	0.91	2.16	**5.12**	2.72	91.14	92.97
KPCA_CMN [43]	90.04	80.11	0.65	3.90	35.89	9.96	79.35	70.92
DSAMNet [47]	95.74	92.35	0.86	2.19	13.11	4.26	90.28	88.55
Proposed SSCAN	**97.53**	95.48	**0.92**	**1.22**	7.82	**2.47**	**94.63**	**93.39**

The aforementioned comparative studies entailing four pairs of remote-sensed images related to real land-cover change events indicate the superiority of the proposed SSCAN in terms of visual performance and quantitative observations.

3.4. Ablation Experiment and Discussion

We first investigated how the BDA in our proposed SSCAN network would influence the detection performance of each dataset. Figure 7 shows the detection accuracies, with BDA and without BDA, of the proposed SSCAN and their comparison with those of other methods in terms of OA, AA, Ka, FA, MA, and TE. The bar chart trends in Figure 7 clearly show that the proposed SSCAN with the BDA strategy offers a better approach in improving detection performance on all datasets and enhancing the accuracy measurements.

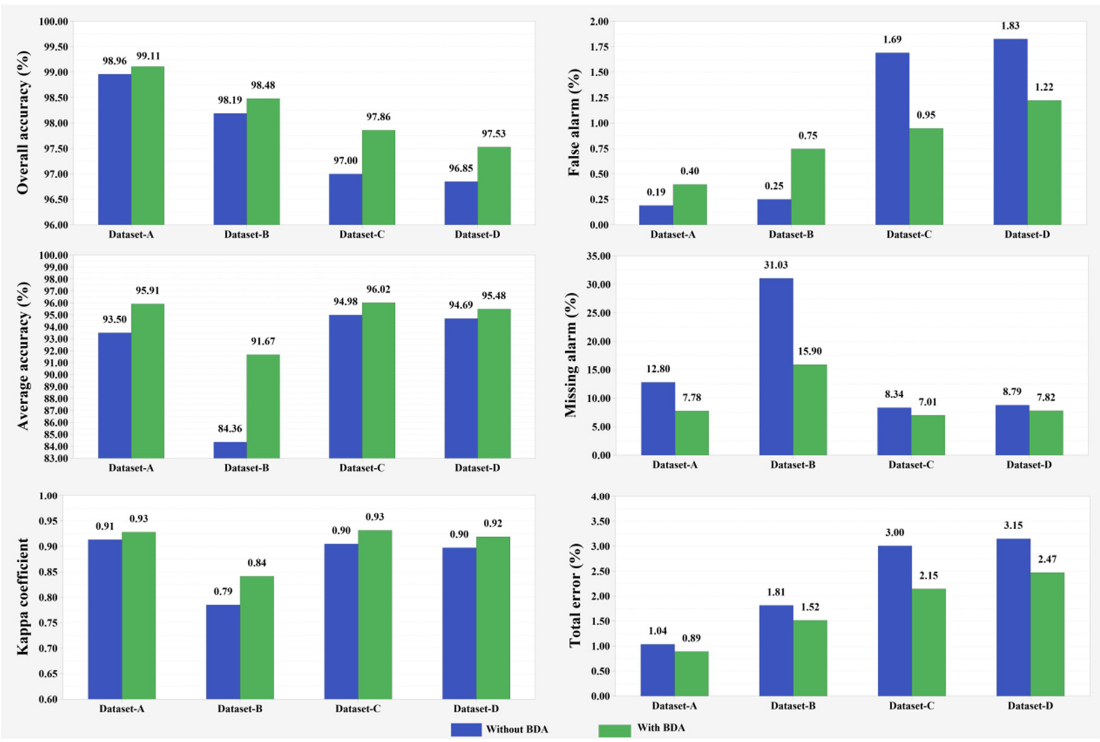

Figure 7. Bar chart comparisons of the proposed SSCAN with and without the suggested BDA strategy.

The loss value is commonly used as a parameter to reflect the learning performance of a neural network. Here, the advantage of using our proposed SSCAN for change detection was further demonstrated by investigating the relationship between the loss value and epoch of SSCAN. As shown in Figure 8, the loss value of SSCAN sharply decreased to a lower value compared with those of other methods, and then it is gradually maintained with increasing training epochs. Thus, the proposed SSCAN had a better learning performance at the same training time for a given dataset. Detailed comparisons of the training time are summarized in Table 6. All of the experiments were conducted on a computer with RTX 2080 Ti GPU, 64G DDR, and Intel Core i7 CPU specifications.

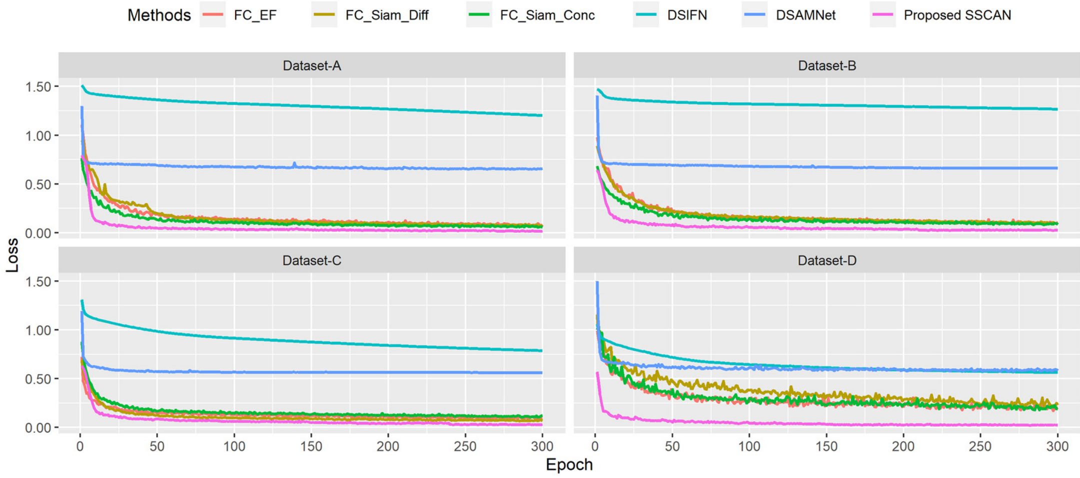

Figure 8. Relationship between loss value and epoch with respect to the application of the proposed SSCAN on each dataset.

Table 6. Training time summary for each approach and dataset (in seconds.)

Method	Dataset-A	Dataset-B	Dataset-C	Dataset-D
FC_EF [28]	817.94	501.76	1545.08	248.72
FC_Siam_Conc [28]	1157.99	705.45	1600.87	325.4
FC_Siam_Diff [28]	2554.42	1114.83	2562.74	384.41
CWNN [39]	/	/	/	-
IFN [41]	195620.47	7749.83	20520.47	1587.93
KPCA_CMN [43]	/	/	/	/
DSAMNet [47]	1526.7904	856.24	1752.26	683.81
Proposed	4508.47	3054.7	4823.6	913.86

Notes: CWNN [39] and KPCA_CMN [43] do not require a training process.

4. Conclusions

In this study, a novel SSCAN was proposed to improve the performance of LCCD with remote-sensed images. To achieve the objective, two attention modules, namely, the SCAM and CBAM, were employed to process pre- and post-events, respectively. Then, the BDA strategy was developed to train the proposed SSCAN. The proposed SSCAN was implemented with four pairs of remote-sensed images, allowing for the depiction of real land-cover change events. The proposed SSCAN achieved better performance and higher detection accuracies than the state-of-the-art methods. The advantages of the proposed SSCAN can be briefly summarized as follows:

(i) Advanced change-detection results were obtained by SSCAN, especially for the three real land-cover change events with four pairs of remote-sensed images, including high-resolution and low-median resolution images. The experimental results indicate that the SSCN outperformed seven widely used LCCD methods, namely, FC_EF [28], FC_Siam_Conc [28], FC_Siam_Diff [28], CWNN [39], IFN [41], KPCA_CMN [43], and DSAMNet [47], in the visual observation and quantitative evaluation criteria.

(ii) The quick and effective learning performance of the proposed SSCAN may be achieved and easily promoted in practical engineering applications. The findings on the relationship between the loss value and epoch indicate the quick learning effect of SSCAN. In other words, SSCAN can be easily trained, and the convergence speed of obtaining the optical model is rapid. These characteristics are acceptable and even preferred in practical applications.

The proposed SSCAN is a promising neural network for achieving LCCD tasks with remote-sensed images. In our future studies, we plan to collect large-area datasets with other types of change-detection methods and apply the proposed network to them to further test its robustness and adaptability.

Author Contributions: Conceptualization, methodology, and writing by X.Y.; validation and experiments by Z.L.; writing—review and editing by J.A.B.; and F.C. provided the fund for publication. All authors have read and agreed to the published version of the manuscript.

Funding: National Natural Science Foundation of China (42271385), the State Key Laboratory of Rail Transit Engineering Informatization (Grant No. SKLKZ22-01), and Open Fund of Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions (Henan University), Ministry of Education (Grant No. GTYR202206).

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to privacy restrictions.

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

References

1. Li, Z.; Shi, W.; Lu, P.; Yan, L.; Wang, Q.; Miao, Z. Landslide mapping from aerial photographs using change detection-based markov random field. *Remote Sens. Environ.* **2016**, *187*, 76–90. [CrossRef]
2. Wu, Y.; Ding, H.; Gong, M.; Qin, A.; Ma, W.; Miao, Q.; Tan, K.C. Evolutionary multiform optimization with two-stage bidirectional knowledge transfer strategy for point cloud registration. *IEEE Trans. Evol. Comput.* **2022**. [CrossRef]
3. Baker, C.; Lawrence, R.L.; Montagne, C.; Patten, D. Change detection of wetland ecosystems using landsat imagery and change vector analysis. *Wetlands* **2007**, *27*, 610–619. [CrossRef]
4. Hegazy, I.R.; Kaloop, M.R. Monitoring urban growth and land use change detection with gis and remote sensing techniques in daqahlia governorate egypt. *Int. J. Sustain. Built Environ.* **2015**, *4*, 117–124. [CrossRef]
5. Desclée, B.; Bogaert, P.; Defourny, P. Forest change detection by statistical object-based method. *Remote Sens. Environ.* **2006**, *102*, 1–11. [CrossRef]
6. Rußwurm, M.; Korner, M. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 11–19.
7. Zhu, Z. Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 370–384. [CrossRef]
8. Hachicha, S.; Chaabane, F. On the sar change detection review and optimal decision. *Int. J. Remote Sens.* **2014**, *35*, 1693–1714. [CrossRef]
9. Lv, Z.; Li, G.; Jin, Z.; Benediktsson, J.A.; Foody, G.M. Iterative Training Sample Expansion to Increase and Balance the Accuracy of Land Classification From VHR Imagery. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 139–150. [CrossRef]
10. Liu, S.; Marinelli, D.; Bruzzone, L.; Bovolo, F. A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 140–158. [CrossRef]
11. Wei, C.; Zhao, P.; Li, X.; Wang, Y.; Liu, F. Unsupervised change detection of vhr remote sensing images based on multi-resolution markov random field in wavelet domain. *Int. J. Remote Sens.* **2019**, *40*, 7750–7766. [CrossRef]
12. Niu, X.; Gong, M.; Zhan, T.; Yang, Y. A conditional adversarial network for change detection in heterogeneous images. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 45–49. [CrossRef]
13. Mercier, G.; Moser, G.; Serpico, S.B. Conditional copulas for change detection in heterogeneous remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1428–1441. [CrossRef]
14. Sun, Y.; Lei, L.; Li, X.; Sun, H.; Kuang, G. Nonlocal patch similarity based heterogeneous remote sensing change detection. *Pattern Recognit.* **2021**, *109*, 107598. [CrossRef]
15. Wu, Y.; Li, J.; Yuan, Y.; Qin, A.; Miao, Q.-G.; Gong, M.-G. Commonality autoencoder: Learning common features for change detection from heterogeneous images. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 4257–4270. [CrossRef] [PubMed]
16. Wu, Y.; Zhang, Y.; Fan, X.; Gong, M.; Miao, Q.; Ma, W. Inenet: Inliers estimation network with similarity learning for partial overlapping registration. *IEEE Trans. Circuits Syst. Video Technol.* **2022**. [CrossRef]

17. Chen, G.; Hay, G.J.; Carvalho, L.M.; Wulder, M.A. Object-based change detection. *Int. J. Remote Sens.* **2012**, *33*, 4434–4457. [CrossRef]
18. Lv, Z.; Wang, F.; Cui, G.; Benediktsson, J.A.; Lei, T.; Sun, W. Spatial-Spectral Attention Network Guided with Change Magnitude Image for Land Cover Change Detection Using Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–12. [CrossRef]
19. Wu, X.; Hong, D.; Chanussot, J. Convolutional neural networks for multimodal remote sensing data classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–10. [CrossRef]
20. Cheng, G.; Yang, C.; Yao, X.; Guo, L.; Han, J. When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative cnns. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2811–2821. [CrossRef]
21. Lv, Z.; Huang, H.; Li, X.; Zhao, M.; Benediktsson, J.A.; Sun, W.; Falco, N. Land cover change detection with heterogeneous remote sensed images: Review, progress, and perspective. *Proc. IEEE* **2022**. [CrossRef]
22. ZhiYong, L.; Liu, T.; Benediktsson, J.A.; Falco, N. Land cover change detection techniques: Very-high-resolution optical images: A review. *IEEE Geosci. Remote Sens. Mag.* **2021**, *10*, 44–63.
23. Wen, D.; Huang, X.; Bovolo, F.; Li, J.; Ke, X.; Zhang, A.; Benediktsson, J.A. Change detection from very-high-spatial-resolution optical remote sensed images: Methods, applications, and future directions. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 68–101. [CrossRef]
24. Cheng, G.; Wang, G.; Han, J. Isnet: Towards improving separability for remote sensing image change detection. *IEEE Trans. Geosci. Remote Sens.* **2022**. [CrossRef]
25. Shafique, A.; Cao, G.; Khan, Z.; Asad, M.; Aslam, M. Deep learning-based change detection in remote sensed images: A review. *Remote Sens.* **2022**, *14*, 871. [CrossRef]
26. Yang, B.; Qin, L.; Liu, J.; Liu, X. Utrnet: An unsupervised time-distance-guided convolutional recurrent network for change detection in irregularly collected images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [CrossRef]
27. Lv, Z.; Huang, H.; Gao, L.; Benediktsson, J.A.; Zhao, M.; Shi, C. Simple multiscale unet for change detection with heterogeneous remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2022**. [CrossRef]
28. Daudt, R.C.; 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), IEEE, Athens, Greece, 7–10 October 2018; pp. 4063–4067.
29. 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.* **2020**, *14*, 1194–1206. [CrossRef]
30. Wang, Z.; Peng, C.; Zhang, Y.; Wang, N.; Luo, L. Fully convolutional siamese networks based change detection for optical aerial images with focal contrastive loss. *Neurocomputing* **2021**, *457*, 155–167. [CrossRef]
31. Zhan, Y.; Fu, K.; Yan, M.; Sun, X.; Wang, H.; Qiu, X. Change detection based on deep siamese convolutional network for optical aerial images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1845–1849. [CrossRef]
32. Zhang, M.; Shi, W. A feature difference convolutional neural network-based change detection method. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7232–7246. [CrossRef]
33. 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.* **2018**, *57*, 924–935. [CrossRef]
34. Jiang, H.; Peng, M.; Zhong, Y.; Xie, H.; Hao, Z.; Lin, J.; Ma, X.; Hu, X. A survey on deep learning-based change detection from high-resolution remote sensing images. *Remote Sens.* **2022**, *14*, 1552. [CrossRef]
35. Shi, W.; Zhang, M.; Zhang, R.; Chen, S.; Zhan, Z. Change detection based on artificial intelligence: State-of-the-art and challenges. *Remote Sens.* **2020**, *12*, 1688. [CrossRef]
36. Yu, X.; Fan, J.; Chen, J.; Zhang, P.; Zhou, Y.; Han, L. Nestnet: A multiscale convolutional neural network for remote sensing image change detection. *Int. J. Remote Sens.* **2021**, *42*, 4898–4921. [CrossRef]
37. Hou, X.; Bai, Y.; Li, Y.; Shang, C.; Shen, Q. High-resolution triplet network with dynamic multiscale feature for change detection on satellite images. *ISPRS J. Photogramm. Remote Sens.* **2021**, *177*, 103–115. [CrossRef]
38. Huang, R.; Zhou, M.; Zhao, Q.; Zou, Y. Change detection with absolute difference of multiscale deep features. *Neurocomputing* **2020**, *418*, 102–113. [CrossRef]
39. 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]
40. Ji, S.; Shen, Y.; Lu, M.; Zhang, Y. Building instance change detection from large-scale aerial images using convolutional neural networks and simulated samples. *Remote Sens.* **2019**, *11*, 1343. [CrossRef]
41. Zhang, C.; Yue, P.; Tapete, D.; Jiang, L.; Shangguan, B.; Huang, L.; Liu, G. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 183–200. [CrossRef]
42. Lee, H.; Lee, K.; Kim, J.H.; Na, Y.; Park, J.; Choi, J.P.; Hwang, J.Y. Local similarity siamese network for urban land change detection on remote sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4139–4149. [CrossRef]
43. Wu, C.; Chen, H.; Du, B.; Zhang, L. Unsupervised change detection in multitemporal vhr images based on deep kernel pca convolutional mapping network. *IEEE Trans. Cybern.* **2021**, *52*, 12084–12098. [CrossRef] [PubMed]
44. Mandal, M.; Vipparthi, S.K. An empirical review of deep learning frameworks for change detection: Model design, experimental frameworks, challenges and research needs. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 6101–6122. [CrossRef]

45. Liu, Y.; Pang, C.; Zhan, Z.; Zhang, X.; Yang, X. Building change detection for remote sensing images using a dual-task constrained deep siamese convolutional network model. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 811–815. [CrossRef]
46. 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.* **2021**, *19*, 1–5. [CrossRef]
47. Shi, Q.; Liu, M.; Li, S.; Liu, X.; Wang, F.; Zhang, L. A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–16. [CrossRef]
48. Liu, M.; Shi, Q.; Marinoni, A.; He, D.; Liu, X.; Zhang, L. Super-resolution-based change detection network with stacked attention module for images with different resolutions. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–18. [CrossRef]
49. Zhang, P.; Gong, M.; Su, L.; Liu, J.; Li, Z. Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *116*, 24–41. [CrossRef]
50. Zhang, X.; Shi, W.; Lv, Z.; Peng, F. Land cover change detection from high-resolution remote sensing imagery using multitemporal deep feature collaborative learning and a semi-supervised chan–vese model. *Remote Sens.* **2019**, *11*, 2787. [CrossRef]
51. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
52. Zhang, H.; Dana, K.; Shi, J.; Zhang, Z.; Wang, X.; Tyagi, A.; Agrawal, A. Context encoding for semantic segmentation. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7151–7160.
53. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–19 June 2019; pp. 3146–3154.
54. Masters, D.; Luschi, C. Revisiting small batch training for deep neural networks. *arXiv* **2018**, arXiv:1804.07612.
55. Qian, X.; Klabjan, D. The impact of the mini-batch size on the variance of gradients in stochastic gradient descent. *arXiv* **2020**, arXiv:2004.13146.
56. Lin, T.; Kong, L.; Stich, S.; Jaggi, M. Extrapolation for large-batch training in deep learning. In Proceedings of the International Conference on Machine Learning, Baltimore, MA, USA, 17–23 June 2020; pp. 6094–6104.