remote sensing

MDPI

Article

The Spectral-Spatial Joint Learning for Change Detection in Multispectral Imagery

Wuxia Zhang [1,2] and Xiaoqiang Lu [1,*]

[1] Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China; wuxiazhang100@126.com
[2] University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: luxiaoqiang@opt.ac.cn

Received: 22 December 2018; Accepted: 21 January 2019; Published: 24 January 2019

Abstract: Change detection is one of the most important applications in the remote sensing domain. More and more attention is focused on deep neural network based change detection methods. However, many deep neural networks based methods did not take both the spectral and spatial information into account. Moreover, the underlying information of fused features is not fully explored. To address the above-mentioned problems, a Spectral-Spatial Joint Learning Network (SSJLN) is proposed. SSJLN contains three parts: spectral-spatial joint representation, feature fusion, and discrimination learning. First, the spectral-spatial joint representation is extracted from the network similar to the Siamese CNN (S-CNN). Second, the above-extracted features are fused to represent the difference information that proves to be effective for the change detection task. Third, the discrimination learning is presented to explore the underlying information of obtained fused features to better represent the discrimination. Moreover, we present a new loss function that considers both the losses of the spectral-spatial joint representation procedure and the discrimination learning procedure. The effectiveness of our proposed SSJLN is verified on four real data sets. Extensive experimental results show that our proposed SSJLN can outperform the other state-of-the-art change detection methods.

Keywords: multispectral imagery; spectral-spatial representation; Siamese CNN; feature fusion; discrimination learning; change detection

1. Introduction

Change detection is the process of aiming at the identification of the differences in multitemporal remote sensing data. The multitemporal data generally includes two or more multispectral images. These images are usually taken in the same geographical region at two or more different times using the same or different sensors [1,2]. The attention on change detection studies in the remote sensing domain is focused on the geographical location and type of changes identification. It has been applied to many fields, such as urban expansion monitoring [3–7], land cover change [8–10], damage assessment and defoliation [11], resource management and forest mortality [12].

Recently, a large number of different change detection methods have been proposed and have achieved very promising results. One important branch of the change detection method is based on post-classification approaches. Post-classification methods first classify two images separately for generating independent class maps and then label the changed pixels by comparing these two class maps [9,13,14]. These methods do not consider the correlation between bi-temporal images, but the correlation between bi-temporal images is very useful for the change detection task [15,16].

Image algebra based change detection methods fully consider the correlation between bi-temporal images. Change vector analysis (CVA) [15] is a classic change detection method based on image

algebra approaches. The change information is measured by the differences of Euclidean distance between pixels, but it contains a lot of noise. Some image transformation based methods are proposed to alleviate the impact of noise on the detection performance. These methods are primarily intended to learn new transformed feature representation [17–19]. In the new learned feature space, the change information is enlarged and unchanged information is suppressed. These transformation-based models contain principal component analysis (PCA) [20], iteratively reweighed multivariate alteration detection (IRMAD) [21], and slow feature analysis (SFA) [16].

Although the above-mentioned methods have made outstanding contributions to the development of change detection domain and achieved good experimental results, most of these traditional methods use hand-crafted features that are weak in the image representations [22]. It is noteworthy that the good generalization ability of hierarchical features on the high level is very helpful for human brains in the object recognition [23].

Fortunately, the deep neural network provides a new way to extract abstract, robust and high-level features [24–26]. Deep neural networks have achieved great success in the remote sensing field. Recently, researchers have shown the increasing interests in deep learning based approaches [24–29]. In general, deep learning based change detection methods consist of two steps: the first and also most crucial one aims to extract effective feature representations via the deep neural network; the second step is the discrimination procedure. Its purpose is to determine whether the areas are changed by obtained features. The effectiveness of feature representation directly affects the change detection accuracy. Good feature representations allow unchanged pairwise samples to be closer in the feature space and changed pairwise samples to be farther [30]. Therefore, good feature representations can help to improve the change detection accuracy. Liu et al. [31] propose a symmetric convolutional coupling network (SCCN) that is verified on the optical and SAR images. The SCCN is based on Deep Belief Network (DBN), so it is an unsupervised layerwise feature learning method. Zhan et al. [22] present a novel supervised change detection method for optical aerial images. This method uses the framework of the siamese convolutional network. The siamese convolutional network is trained by the weighted contrastive loss function. Gong et al. [27] propose an unsupervised change detection method that incorporates superpixel-based change features extraction and difference representation learning model by neural networks. Although the superpixel-based extracted features contain spectral, textual, and spatial information, but they are only used to obtain the labels of training samples and are not employed to finally detect changed areas. The input of difference representation learning network is the vectorization of the patch, it loses the spatial information. So the features used in determining changed areas do not include spatial information. The above-mentioned methods are summarized in Table 1.

Table 1. The change detection methods.

Branch	Category	Method
Conventional methods	Post-classification based methods	Demir et al. [9], Muñoz-Marí et al. [13], Volpi et al. [14]
	Image algebra based methods	Image difference, Image ratio, CVA [15]
	Image trasnformation based methods	SFA [16], PCA [20], IRMAD [21]
Deep learning based methods	Deep neural network based methods	Zhan et al. [22], Gong et al. [27], Liu et al. [31]

Although the aforementioned methods have got promising results, they do not take the spectal-spatial information into account and explore the underlying information of their corresponding fusion features at the same time. To address the above-mentioned problems, we propose a *Spectral-Spatial Joint Learning Network* (SSJLN). The proposed network is an end-to-end deep neural network. As shown in Figure 1, our proposed network contains three parts: spectral-spatial joint representation, feature fusion, and discrimination learning.

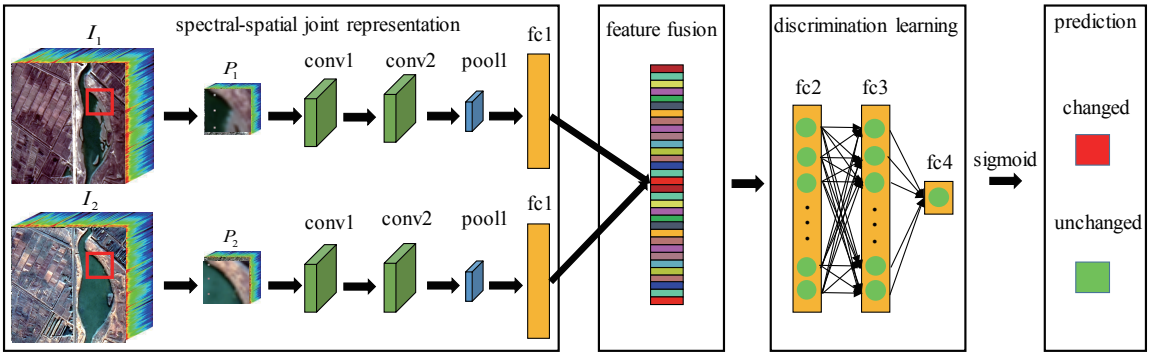


Figure 1. Flowchart of our proposed spectral-spatial joint learning network.

First, the spectral-spatial joint representation is learned from the deep network that is similar to *Siamese Convolutional Neural Network* (S-CNN). Second, the spectral-spatial features are fused to represent the difference information, which will help to recognize the changed areas. Third, the discrimination learning part is proposed to further explore the underlying information of fused features, which is implemented by several fully connected layers.

The contributions of our paper can be summarized as follows:

1. We propose an end-to-end network that jointly learns spectral-spatial representations for the change detection task, named spectral-spatial joint learning network (SSJLN).
2. We present the discrimination learning part that uses the fused features obtained from the spectral-spatial joint representation part as input. The discriminative learning can further explore and analyze the underlying information of fused features.
3. We build a new loss function that integrates the loss of the spectral-spatial joint representation part and the loss of the discriminative learning part to better learn spectral-spatial representations.

The organization of this paper is as follows. Section 2 will briefly discuss the related work of S-CNN. In Section 3, our proposed spectral-spatial joint learning network will be described in detail, which consists of the network architecture, how to train the network and how to predict the testing samples by the trained model. In Section 4, the effectiveness of our proposed method is verified on four real data sets. Section 5 is the conclusion.

2. Background

Recently, S-CNN is widely used to extract patch-pairwise features and applied to many fields, such as hyperspectral image classification [32], image patch comparison [33], change detection [22], object tracking [34] and person re-identification [35].

S-CNN has two branch networks and the two branches have the same architecture and weights. Each branch includes a series of convolutional, Rectified Linear Unit (ReLU) and max-pooling layers. Each patch in the patch pair is used as input for each branch. Then, the features provided by two branches are fed to the contrastive loss function defined in Equation (2).

p_1 and p_2 stand for the pairwise patch. $f(p_1)$ and $f(p_2)$ represent features provided by the two CNN branches. The Euclidean distance between features $f(p_1)$ and $f(p_2)$ corresponding to the input patch p_1 and p_2 are computed as

$$D = \|f(p_1) - f(p_2)\|_2. \tag{1}$$

The margin-based contrastive loss function is defined as follows:

$$\mathcal{L} = \frac{1}{2}(1-l)D^2 + \frac{1}{2}lmax(m-D,0)^2, \tag{2}$$

where l denotes a binary label that means whether the input pairwise patch p_1 and p_2 is dissimilar ($l = 1$) or similar ($l = 0$). $m > 0$ is the margin for dissimilar pairs. D represents the distance between

features $f(p_1)$ and $f(p_2)$. When the loss function $\mathcal{L}$ obtains the minimum value, the value of D is small for similar pairs and large for dissimilar pairs [36]. When the distance between dissimilar pairs is smaller than the margin m, dissimilar pairs can affect the value of the loss function $\mathcal{L}$. Moreover, when the distance between dissimilar pairs is larger than the margin m, dissimilar pairs will not contribute to the loss function [22,30]. Hence, the purpose of this loss function is to make similar pairs closer in the feature space and push dissimilar pairs apart.

3. Our Method

The network architecture will be first described in detail in this section. Then, the presented new loss function will be discussed. Finally, we will give more details of training and prediction.

3.1. Network Architecture

As shown in Figure 1, the proposed SSJLN is an end-to-end deep network architecture, which uses the supervised information to directly guide both the spectral-spatial joint representation procedure and the discrimination learning procedure. The extracted features from the spectral-spatial joint representation part are fused, then fused features are fed to the discrimination learning part to explore the underlying information. SSJLN contains several types of layers that are commonly used in deep learning network for computer vision.

SSJLN contains three parts: spectral-spatial joint representation, feature fusion, and discrimination learning as shown in Figure 1. We will discuss the spectral-spatial joint representation, feature fusion, and discrimination learning separately in the following.

3.1.1. Spectral-Spatial Joint Representation Part

The spectral-spatial joint representation part is similar to S-CNN using two tower structure sharing the same parameters.

Obviously, the input of the spectral-spatial joint representation part is a series of corresponding pairwise patches $p_{1,i}$ and $p_{2,i}$, $i = 1, \ldots, N$ with the size of 5×5, where $p_{1,i}$ and $p_{2,i}$ are taken from I_1 and I_2, respectively. N represents the number of pixels in the image I_1 or I_2. There are two reasons to use the patch as input. First, the isolated pixel is very sensitive to the noise [31]. Second, the patch considers both the spectral information of the pixel to be detected and its' adjacent pixels [37]. In other word, using the patch as input takes both the spectral and spatial information into account. The patch size selected by users is an important factor affecting the detection performance, which will be verified in Section 4.4.1. Using the patch as input is based on the assumption that the adjacent pixels generally belong to the same class in the spatial space.

More specifically, the spectral-spatial joint representation part contains 2 convolutional layers (conv1 and conv2) and 1 pooling layer (pool1) and 1 fully-connected layer (FC1). Both the convolution and fully-connected layers use ReLU as the activation function [38]. The aim of this part is to learn spectral-spatial joint representations.

3.1.2. Feature Fusion Part

Generally, the learned pairwise features are fused to detect changes because the correlation between the pairwise features is very useful for the change detection task [15,16]. There are many fusion strategies, but the most common ones are only two strategies: stack and difference. In [27], the stacked strategy is employed. In [15,39], the difference strategy is used.

The difference features carry useful information for the change detection problem [39,40]. The pixel with small value is likely to be unchanged and with large value tends to be changed. Therefore, the difference features are very informative [39]. Moreover, in the feature space that learned from the spectral-spatial joint representation part, the distance between two learned features is smaller for unchanged pairs and larger for changed pairs. Hence, the difference strategy is used as the feature fusion strategy in our paper.

Suppose that f_1 and f_2 are the transforming functions for p_1 and p_2, respectively, where p_1 is from I_1 and p_2 is from I_2. The difference feature (DF) can be calculated as:

$$DF = F(p_1, p_2) = f_1(p_1) \ominus f_2(p_2), \tag{3}$$

where $\ominus$ is the difference operator. $f_1(p_1)$ and $f_2(p_2)$ represent the extracted features. Since the difference feature is informative, the difference feature can better represent the changed areas. The difference feature with small value is prone to be unchanged and this with high value is likely to be changed.

Since the deep neural network can better represent the nonlinear relationship between input and output, S-CNN in which the two branches share with the same parameters is adopted to transform p_1 and p_2 into features $f_1(p_1)$ and $f_2(p_2)$, respectively in our paper. In other words, p_1 and p_2 are fed to the two branches of S-CNN to obtain spectral-spatial joint representations $f_1(p_1)$ and $f_2(p_2)$.

3.1.3. Discrimination Learning Part

Although fused features are very useful to detect changed areas, the underlying information of fused features can also improve the discrimination of fused features. Hence, the discrimination learning part uses fused features as input to learn underlying information of fused features. Hence, the abstract features are extracted by the discrimination learning part, which is helpful for classification [41]. The effectiveness of the discrimination learning is verified and demonstrated in Section 4.4.3.

The discrimination learning part is modeled by two fully-connected layers with ReLU nonlinearity (FC2-FC3) and one fully-connected layer with sigmoid nonlinearity and cross-entropy (FC4). The output value of FC4 is in [0, 1]. These values are non-negative and can be added up to one. Hence, they can be seen as the probability estimation of the network that the pairwise patches are changed or unchanged, respectively.

3.2. Loss Function

We present a new loss function that takes both the losses of the spectral-spatial joint representation part and the discrimination part into account. Hence, the overall loss function $L_{overall}$ contains two parts: L_1 and L_2, which is defined as follows:

$$L_{overall} = \omega_1 L_1 + \omega_2 L_2, \tag{4}$$

where ω_1 and L_1 are the weight and the loss of the spectral-spatial joint representation part in SSJLN. ω_2 and L_2 represent the weight and the loss of the discrimination learning part in SSJLN.

L_1 is calculated by the contrastive loss function as shown in Equation (5). The aim of the contrastive loss function is to make similar pairs closer in the feature space and push dissimilar pairs apart.

$$L_1 = \frac{1}{2}(1-l)D^2 + \frac{1}{2}lmax(m-D,0)^2, \tag{5}$$

where l denotes a binary label that means whether the input pairwise patch p_1 and p_2 is dissimilar ($l = 1$) or similar ($l = 0$). D represents the distance between feature vectors $f(p_1)$ and $f(p_2)$. $m > 0$ is the margin for dissimilar pairs.

L_2 is computed by the cross-entropy loss function that is defined as follows:

$$E = -\frac{1}{n}\sum_{i=1}^{n}[y_i log\hat{y}_i + (1-y_i)log(1-\hat{y}_i)], \tag{6}$$

where the value of y_i is 0 or 1, which represents the label information of input pair. 1 indicates that the input pair is changed. $\hat{y}_i$ indicates the probability that the prediction of the input patch pair is positive. $\hat{y}_i$ is calculated by the sigmoid function.

Since the input of the discrimination learning part is obtained by fusing the spectral-spatial joint features, the intermediate features acquired from FC2 is also important for classification [42]. However, many methods only focus on the final output features and ignore the important effect of the intermediate features. The extra information provided by the intermediate features can enhance the classification performance [42]. Hence, L_2 contains two part: E_{FC3} and E_{FC2}. E_{FC3} is the cross-entropy loss computed on the FC3 layer. E_{FC2} is also the cross-entropy loss that is calculated on the FC2 layer. L_2 is defined as follows:

$$L_2 = E_{FC3} + \lambda E_{FC2}, \tag{7}$$

where λ is a parameter that trades off the E_{FC3} term and the E_{FC2} term. The value of λ indicates that which term gives a more contribution to the loss L_2. Generally, features obtained from FC3 is more abstract than those of FC2, therefore, λ should not be larger than 1. The effectiveness of the proposed new loss function is verified and demonstrated in Section 4.4.4.

3.3. Training and Prediction

The SSJLN is trained in a supervised manner, as shown in Figure 1. Since the training data set may be quite large, a stochastic approximation of this objective is employed in practice. Training data is randomly divided into several mini-batches. The forward propagation is performed on the current mini-batch, and then the output and loss are calculated. Back propagation is then employed to calculate the gradients on this batch, and the network weights are updated [6]. The weights are updated by the stochastic gradient descent (SGD) strategy.

Although our proposed SSJLN contains three parts: spectral-spatial joint representation, feature fusion, and discrimination learning, the proposed SSJLN is trained in the unified framework. So it is very easy to predict the testing pairs in our proposed SSJLN. The testing pairwise patches are directly fed to the trained SSJLN. The obtained prediction is the detection result.

4. Experiments

In this section, several experiments are performed on the four multispectral data sets to prove the effectiveness of our proposed SSJLN. First, we give a detailed description of four real multispectral data sets. Second, comparison methods and evaluation metrics are described in detail. Third, the influence of important parameters or steps on the detection results will be analyzed. Finally, we show experimental results on four data sets and analyze the experimental results in detail.

4.1. Datasets

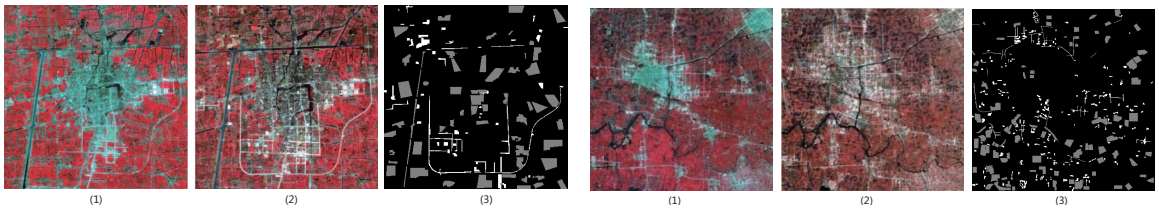
- *ETM+ Data* : Taizhou and Kunshan data sets are collected by the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) sensor. The images acquired by ETM+ detectors have 6 spectral bands (including bands 1–5 and 7), and its spatial resolution is 30 m. The ground truth of Taizhou and Kunshan data sets are obtained from [43].

  The Taizhou date set consisting of two images with the size of 400×400 is acquired from Taizhou city, China. One of two images is collected on 17 March 2000, and the other is collected on 6 February 2003. The types of change included in the Taizhou data set is mainly the urban expansion. Figure 2a shows the pseudocolor images and ground truth of the Taizhou date set.

  The Kunshan date set containing two images with the size of 800×800 is acquired from Kunshan city in 2000 and 2003 at the same time. Figure 2b shows the pseudocolor images and ground truth of the Kunshan data set. Compared with Taizhou data set, it can be clearly seen that Kunshan data has more complex texture because the main change types of Kunshan data set contain both urban expansion and farmland changes.
- *GF-1 Data*: The Minfeng and Hongqi Canal data sets are both acquired by GF-1 satellite. Both of them consist of two images, one of which is collected on 9 December 2013, and the other one is

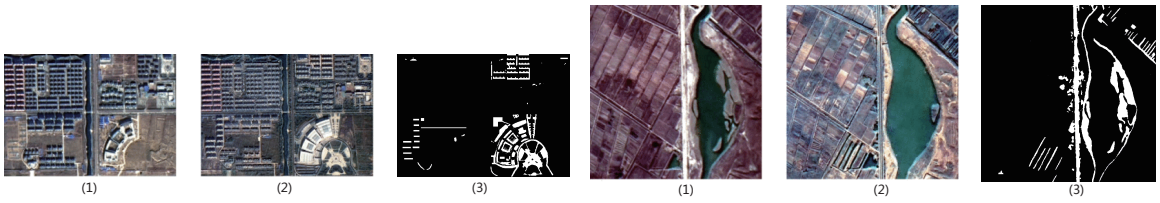
collected on 16 October 2015. The spatial resolution of the acquired images is 2 m, and the images include 4 spectral bands.

The pseudocolor images and ground truth of Minfeng data set shown in Figure 3a display the changes of buildings around Minfeng lake. The size of images in Minfeng data set is 651 × 461. The pseudocolor images and ground truth of Hongqi Canal data set show the changes of Hongqi Canal riverway near the Xijiu village, as shown in Figure 3b. The size of images in Hongqi data set is 539 × 543.



(**a**) Taizhou (**b**) Kunshan

Figure 2. Pseudocolor images of Taizhou and Kunshan data sets in (1) 2000. (2) 2003. (3) ground truth. The background is black, the unchanged samples are gray, and the changed samples are white.



(**a**) Minfeng (**b**) Hongqi

Figure 3. Pseudocolor images of Minfeng and Hongqi data sets in (1) 2013. (2) 2015. (3) ground truth. The unchanged samples is black, and the changed samples are white.

4.2. Competitors

The proposed SSJLN is compared with the state-of-the-art methods as follows:

- CVA [15] is a classic unsupervised change detection algorithm in the field of remote sensing. CVA detects the changed areas by using the magnitude of the difference vectors.
- Iteratively reweighed MAD (IRMAD) [21] gives high weights to the observations that vary little with time through an iterative strategy compared with multivariate alteration detection (MAD) and is widely used in the change detection domain.
- SCCN [31] projects the two input images into a feature space with an unsupervised deep network including one convolutional layer and several coupling layers, which has two sides. By calculating the distance in the learned feature space, the difference map is generated to detect the changed areas.
- S-CNN+Threshold [22] uses S-CNN to extract abstract and robust features with the weighted contrastive loss. The distance between pairwise features is calculated and then is used by the threshold strategy to generate the changed map.
- S-CNN+SVM [32] uses S-CNN to learn the deep learning features which can separate different classes. Then features extracted from S-CNN are employed to train a linear SVM classifier.
- Superpixel-Based Difference Representation Learning (SDRL) [27] is an unsupervised change detection method that incorporates superpixel-based change features extraction and difference representation learning model by neural networks. The network is trained to learn the difference representations.

4.3. Evaluation Criteria

In order to testify the robustness and validity of our proposed SSJLN, some quantitative analysis is essential. First, overall accuracy (OA) is used to assess the overall effectiveness of the change detection methods.

$$OA = (TP + TN)/(TP + TN + FP + FN), \tag{8}$$

where TP refers to true positive, TN stands for true negative, FP stands for false positive and FN represents false negative.

Second, Kappa Coefficient (KC) is used to measure the classification accuracy, which is calculated by

$$KC = (OA - PRE)/(1 - PRE), \tag{9}$$

where

$$PRE = \frac{(TP + FP)(TP \dot{+} FN) + (FN + TN)(FP \dot{+} TN)}{(TP + TN + FP + FN)^2}. \tag{10}$$

Finally, the AUC value (the area under the receiver operating characteristics (ROC) curves) provides a numerical accuracy measure. The true positive rate (TPR) represents sensitivity. The false positive rate (FPR) stands for the probability of false alarm. With TPR as Y-axis and FPR as X-axis, the ROC curve can be drawn at various threshold settings. The TPR and FPR are calculated by

$$TPR = TP/(TP + FN), \quad FPR = FP/(FP + TN). \tag{11}$$

The AUC value is 1, which represents the perfect test.

The larger values of OA, Kappa and AUC mean the better detection performance.

4.4. Parameter Settings

First, we describe the parameter settings related to the SSJLN structure, such as the number of neurons, the kernel size and the stride of each layer. The detailed layer parameter settings of our proposed SSJLN are illustrated in Table 2.

Table 2. Layer parameters of SSJLN.

Names	Types	Input Dim.	Output Dim.	KS	S
conv1	C	5*5*BN	4*4*32	2*2	1
conv2	C	4*4*32	4*4*64	2*2	1
pool1	MP	4*4*64	2*2*64	2*2	2
FC1	FC	2*2*64	128	-	-
FC2	FC	128	128	-	-
FC3	FC	128	128/96	-	-
FC4	FC	128/96	1	-	-

Table 2 shows the detailed configuration of the layers, where the dimensions of the input and output are expressed by height*width*depth. KS: kernel size for convolution and pooling layers. BN: band number, it is 6 for ETM+ data and 4 for GF-1 data. S: stride. Layer types: C: convolution, MP: max-pooling, FC: fully-connected. The number of neurons in FC3 is 128 for EMT+ data and 96 for GF-1 data.

Then, the parameters used in the training and testing phases are described and discussed. In the training phase, the equal number of changed and unchanged pairwise samples are used to address the problem of bias towards the unchanged decision because the number of unchanged pairwise samples is very large in the changed detection domain. The number of changed pairwise patches is 1000, and the number of unchanged pairwise patches is also 1000 for our proposed SSJLN and the other comparison methods such as SCCN and SDRL. All remaining samples are used for testing.

The mini-batch strategy is employed in the training phase and the batch size is set to 32 empirically. The value of the learning rate in our proposed method is 10^{-4} experimentally. The model can achieve the optimal results for four real data sets after running 400 iterations or convergence.

For the proposed new loss function, four parameters ω_1, ω_2 in Equation (4), m in Equation (5) and λ in Equation (7) should be considered. The weights ω_1 and ω_2 are set to 1 empirically. In our experiment, m is set to 0.5. λ is set to 0.5 experimentally. We perform 20 iterations on EMT+ and GF-1 data sets. The changed and unchanged pairwise patches are randomly selected from the changed and unchanged sample set, respectively for each iteration. The best detection result of 20 iterations is selected as the final detection result. For instance, the batch size is 32, and 16 positive and 16 negative pairwise patches are fed to stochastic gradient descent in each training iteration.

Finally, the effects of the key parameter like the patch size and important steps such as the difference strategy, the discrimination learning and the loss function are discussed in detail.

4.4.1. Effect of Patch Size

The patch is treated as the basic unit in our proposed SSJLN because the patch not only considers the spectral information of the pixel to be detected but also the spectral information of the pixels surrounding to the pixel to be detected. The value of the patch size n determines how much local information the patch can contain. Hence, the parameter n is an important factor affecting the detection accuracy, which is determined by users.

The spatial resolution of multispectral images is generally not too high. If the value of n selected by users is too large, the types of land cover contained in the same patch will probably not be the same. We do the experiment to verify this conclusion. In the experiment, the values of patch size n we select are $3, 5, 7, 9$, respectively. Figure 4 shows the effect of different patch size n on OA, Kappa coefficient and AUC values of the detection result for four different data sets. For the sake of simplicity, we only use E_{FC3} to calculate the loss.

It can be clearly observed from Figure 4 that the values of all the quantitative indicator, especially OA and Kappa for Taizhou and Kunshan date sets are the highest when $n = 5$. When $n = 3$, the local information included in the patch is not sufficient. However, the detection accuracy of $n > 5$ does not exceed that of $n = 5$, because the impact of the newly included pixels in the 7×7 patch are no contribution or negative contribution to the detection accuracy compared with the 5×5 patch. However, for Minfeng and Hongqi canal date sets, the detection performance changes with the accompanying changes in the value of n, but it does not show the regular distribution. Although $n = 3$ is the best selection for Minfeng and Hongqi canal date sets, for the sake of simplicity, we still select $n = 5$ as the value of the patch size.

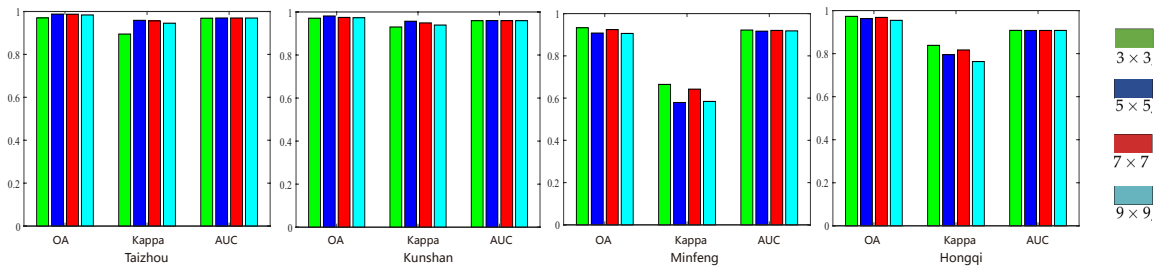


Figure 4. OA, Kappa and AUC values comparision of difference patch sizes on the four different data sets. Green, blue, red and sky blue colors denote patch size 3×3, 5×5, 7×7 and 9×9, respectively.

4.4.2. Effect of Difference Strategy

In the feature fusion procedure, the difference strategy is adopted to fuse extracted features from the spectral-spatial joint representation part. The contrast experiment is exploited to compare the difference strategy and stacked strategy. The stacked strategy based method has the same network architecture as our proposed SSJLN. The only difference is that the fusion strategy is replaced by stack

strategy. In other words, the output dimension after the feature fusion part is 128, while the output dimension of the stacked strategy based method is 256. For the sake of simplicity, E_{FC3} is only used to calculate the loss in this comparative experiment.

The change detection accuracy on four difference data sets is illustrated in Figure 5, where green color denotes stacked strategy based method and red color denotes our proposed SSJLN. All OA, Kappa and AUC values in Figure 5 acquired from SSJLN are higher than those acquired from stacked strategy based method for four different data sets. Especially, the increase in kappa coefficients is much higher than that of AUC value on all four data sets. This can prove that the difference features have better discriminative ability compared with stacked features.

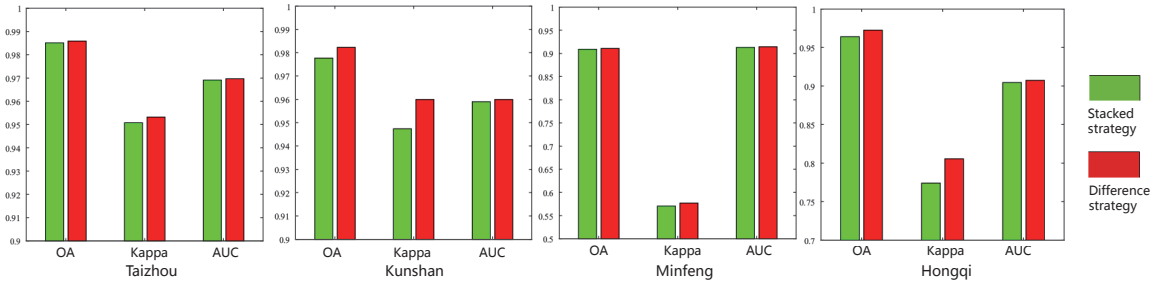


Figure 5. OA, Kappa and AUC values comparision of four different data sets. Green color denotes the stacked strategy based method and red color denotes our proposed SSJLN that uses the difference strategy.

4.4.3. Effect of Discrimination Learning

The aim of the discrimination learning part is to explore the underlying information of fused features for better discrimination. We do a comparative experiment to better verify the effect of the discrimination learning part on detection accuracy. SSJLN-NDL means SSJLN does not contain the discrimination learning part. For the sake of simplicity, the cross-entropy loss is only used in this comparative experiment.

Figure 6 shows the OA, Kappa and AUC values of SSJLN-NDL and SSJLN, where green color denotes SSJLN-NDL and red color denotes SSJLN. All OA, Kappa and AUC values in Figure 6 acquired from SSJLN are higher than those acquired from SSJLN-NDL on four different data sets, especially Kunshan data set. This is because the discrimination learning part in the SSJLN can better explore the underlying information of fused features that can enhance the discrimination. Therefore, it can prove that the underlying information learned from the discrimination learning part can indeed improve the detection accuracy.

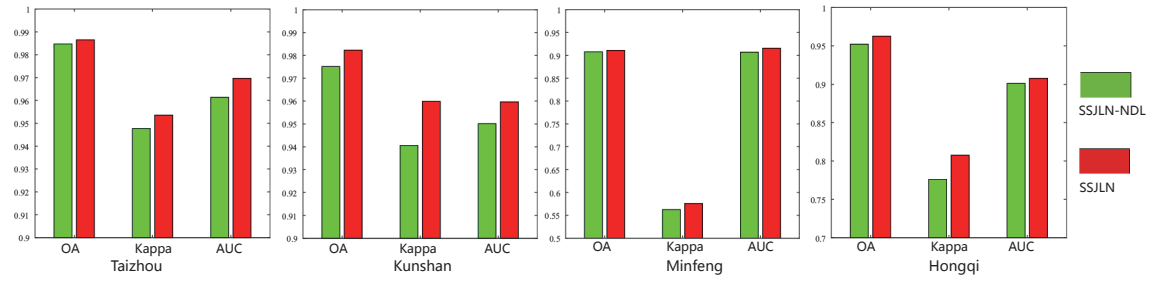


Figure 6. OA, Kappa and AUC values comparision of four different data sets. Green color denotes SSJLN-NDL method that does not contain the discrimination learning part and red color denotes SSJLN.

4.4.4. Effect of Loss Function

We do a comparative experiment to verify the effect of the losses of final output features E_{FC3}, the intermediate features E_{FC2} and the spectral-spatial joint representation part L_1 on the detection accuracy for four different data sets.

The OA, Kappa coefficient and AUC values of the detection result for four different data sets are shown in Figure 7, where green color denotes E_{FC3}, blue color represents $L_2 = E_{FC3} + 0.5 * E_{FC_2}$ and red color denotes $L_{overall} = L_1 + L_2$. It can be obviously seen that E_{FC2} and L_1 have a positive effect on improving the detection accuracy.

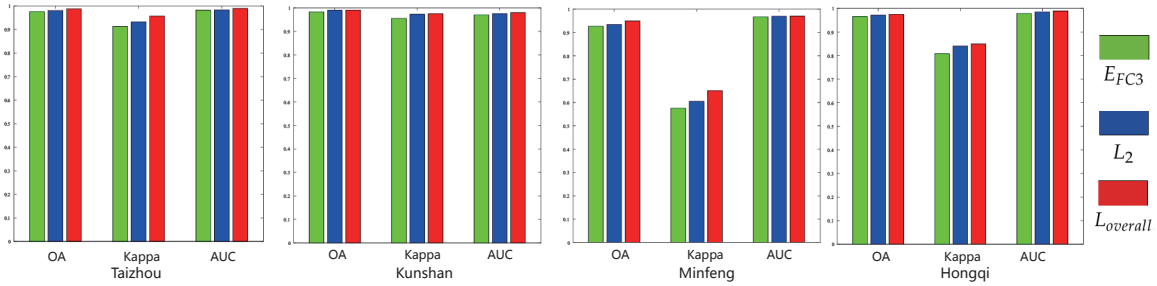


Figure 7. OA, Kappa and AUC values comparision of four different data sets. Green, blue and red colors denote E_{FC3}, L_2 and $L_{overall}$, respectively.

4.5. Detection Performance

4.5.1. Results of EMT+ Data

The EMT+ data includes Taizhou and Kunshan data sets. We will analyze the results of Taizhou and Kunshan data sets from qualitative and quantitative aspects.

The binary change maps of our proposed SSJLN and all other comparison methods on Taizhou and Kunshan data sets are shown in Figures 8 and 9, respectively. It can be clearly observed that the detection results from CVA and IRMAD methods shown in (a) and (b) of Figures 8 and 9 are very poor, and many unchanged pixels are incorrectly labeled as changed pixels. That's because these two methods use manual features that rely on data and prior information and are unsupervised.

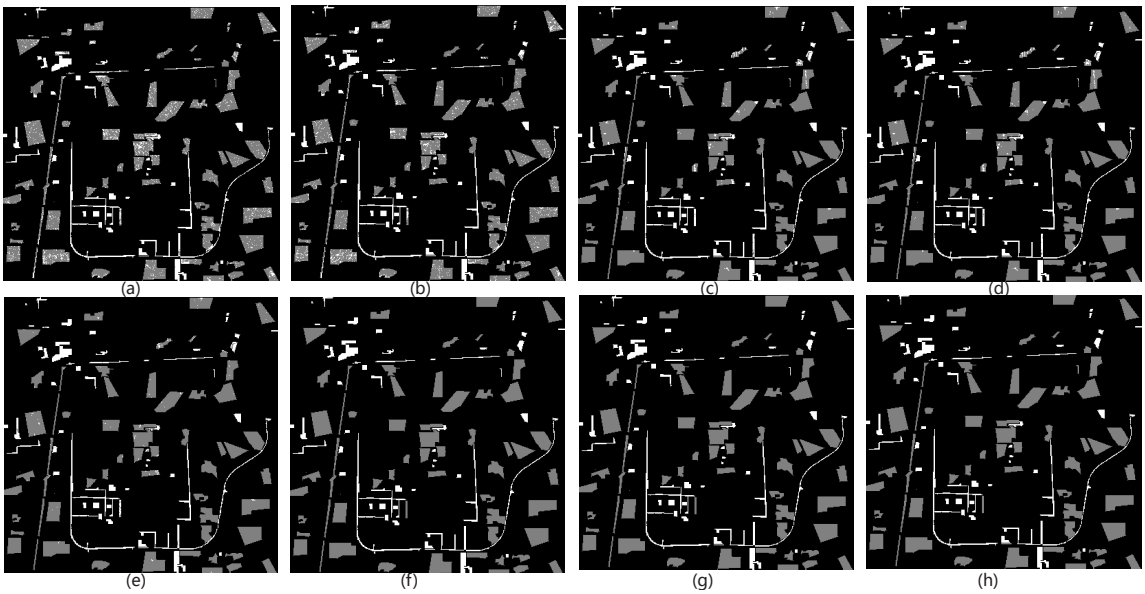


Figure 8. Binary change maps of Taizhou data set. (**a**) CVA. (**b**) IRMAD. (**c**) SCCN. (**d**) S-CNN+Threhold. (**e**) S-CNN+SVM. (**f**) SRDL. (**g**) Our proposed SSJLN. (**h**) Ground truth. The background is black, unchanged samples are gray, and changed samples are white.

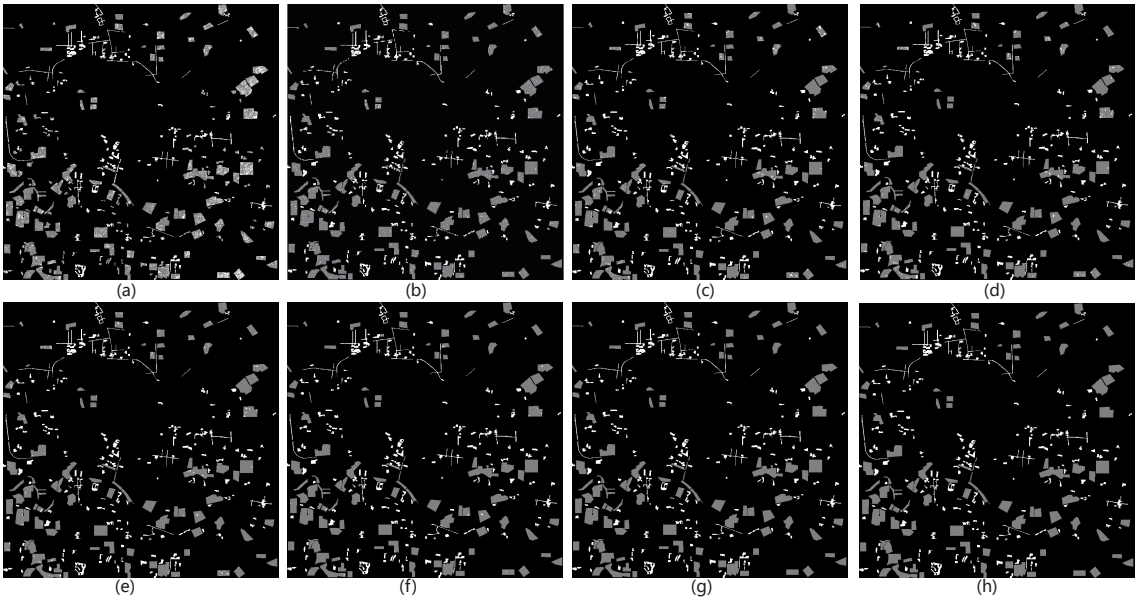


Figure 9. Binary change maps of Kunshan data set. (**a**) CVA. (**b**) IRMAD. (**c**) SCCN. (**d**) S-CNN+Threhold. (**e**) S-CNN+SVM. (**f**) SRDL. (**g**) Our proposed SSJLN. (**h**) Ground truth. The background is black, unchanged samples are gray, and changed samples are white.

The binary change maps obtained from SCCN, S-CNN+Threshold, S-CNN+SVM, SDRL and our proposed SSJLN are much better than those of CVA and IRMAD methods visually because of using deep learning network to extract deep features, as shown in (c), (d), (e), (f) and (g) of Figures 8 and 9. SCCN is an unsupervised method that does not use the label information of pixels, and S-CNN+Threshold and S-CNN+SVM are supervised methods. The number of unchanged pixels detected as changed pixels in Figures 8e and 9e is much less than those in (c) and (d) of Figures 8 and 9. Hence, S-CNN+SVM can surpass SCCN and S-CNN+Threshold because SVM is more robust and effective than the threshold strategy.

Different from the comparison methods, SDRL and our proposed SSJLN use fused features to better represent changes. SDRL adopts the stacked strategy, while our proposed SSJLN uses the difference strategy. As shown in Figure 5, the difference strategy is better than the stacked strategy. Moreover, our proposed SSJLN can extract spectral-spatial joint representation, while SDRL loses the spatial information because of using the vectorization of the patch as input. From Figures 8g and 9g, we can observe that the result obtained from our proposed SSJLN virtually eliminates the false positives from the changed areas.

For the quantitative comparison, OA, Kappa coefficients and AUC values are computed and summarized in Table 3. Our proposed SSJLN can get the highest detection accuracy compared with those of CVA, IRMAD, SCCN, S-CNN+Threshold, S-CNN+SVM and SDRL. The OA, Kappa coefficients and AUC values obtained by SSJLN on Taizhou data set are 0.9875, 0.9570 and 0.9897, respectively. The OA, Kappa coefficients and AUC values obtained by SSJLN on Kunshan data set are 0.9907, 0.9753 and 0.9799, respectively. The qualitative analysis shown in Figures 8 and 9 are consistent with the quantitative analysis presented in Table 3.

Table 3. Over accuracy, Kappa coefficients, and AUC value over state-of-the-art methods on Taizhou, Kunshan, Minfeng and Hongqi data sets.

Data	Metric	CVA	IRMAD	SCCN	SCCN+Threshold	SCCN+SVM	SDRL	Ours
Taizhou	OA	0.8123	0.9654	0.9714	0.9724	0.9826	0.9854	**0.9875**
	KAPPA	0.7804	0.9549	0.9061	0.9089	0.9412	0.9501	**0.9570**
	AUC	0.7213	0.9019	0.9463	0.9424	0.9559	0.9799	**0.9897**
Kunshan	OA	0.7021	0.9327	0.9654	0.9412	0.9759	0.9830	**0.9907**
	KAPPA	0.6711	0.9196	0.9103	0.9163	0.9369	0.9553	**0.9753**
	AUC	0.5852	0.8506	0.9299	0.9298	0.9439	0.9607	**0.9799**
Minfeng	OA	0.7381	0.8376	0.8683	0.8720	0.8914	**0.9551**	0.9494
	KAPPA	0.2253	0.5221	0.5108	0.5130	0.5401	**0.6577**	0.6506
	AUC	0.7023	0.7411	0.7551	0.7603	0.8393	**0.9710**	0.9705
Hongqi	OA	0.7307	0.9419	0.9468	0.9422	0.9555	0.9741	**0.9746**
	KAPPA	0.2344	0.6902	0.7282	0.7124	0.7632	0.8464	**0.8490**
	AUC	0.8022	0.8627	0.8796	0.8788	0.8969	0.9881	**0.9889**

4.5.2. Results of GF-1 Data

The GF-1 data includes Minfeng and Hongqi data sets. The qualitative and quantitative analysis of GF-1 data will be described in detail below.

The binary change maps of all the methods on Minfeng and Hongqi Canal data sets are shown in Figures 10 and 11, respectively. Unsupervised CVA and IRMAD methods perform poorly on both Minfeng and Hongqi Canal data sets, as shown in (a) and (b) of Figures 10 and 11. The CVA method has a high false positive rate for these two data sets, which means many unchanged pixels are incorrectly labeled as changed pixels. On the contrary, the IRMAD method has a high false negative rate for these two data sets, which shows changed pixels are detected as unchanged pixels by mistake.

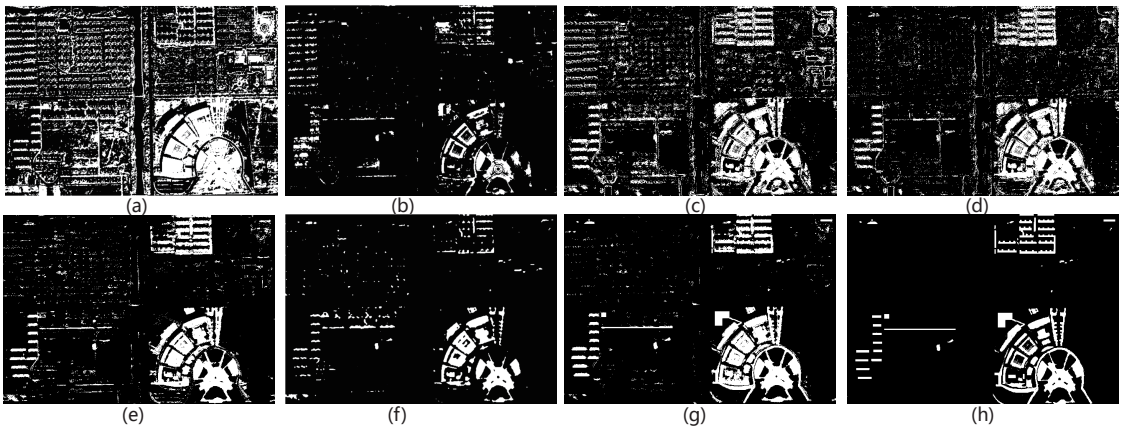


Figure 10. Binary change maps of Minfeng data set. (**a**) CVA. (**b**) IRMAD. (**c**) SCCN. (**d**) S-CNN+Threhold. (**e**) S-CNN+SVM. (**f**) SRDL. (**g**) Our proposed SSJLN. (**h**) Ground truth. The unchanged samples are black and changed samples are white.

Although SCCN and S-CNN+Threshold detect most of the changed areas, as shown in (c) and (d) of Figures 10 and 11, there are a lot of noise spots, especially on the left half of the binary change map for Minfeng data set and the farmland along the riverway for Hongqi Canal data set. Since SVM is more robust and effective than the threshold strategy, the detection result from the S-CNN+SVM shown in Figures 10e and 11e is better than those from SCCN and S-CNN+Threshold. The detection result of SRDL shown in Figures 10f and 11f is better than S-CNN+SVM, but it detects many changed pixels as unchanged pixels by mistake.

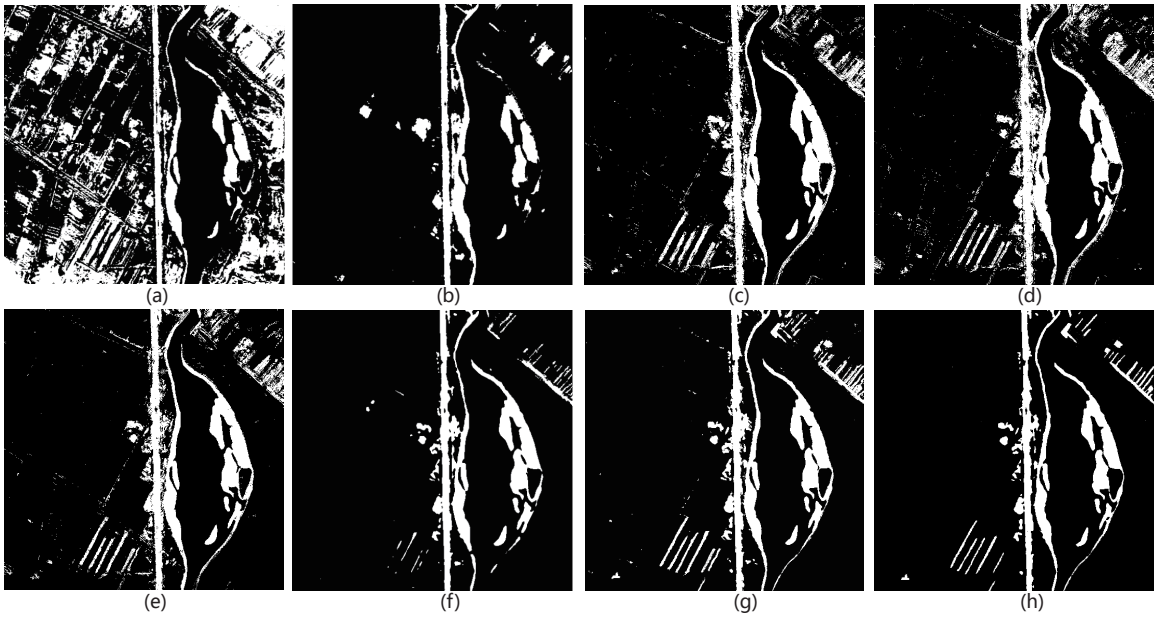


Figure 11. Binary change maps of Hongqi data set. (**a**) CVA. (**b**) IRMAD. (**c**) SCCN. (**d**) S-CNN+Threhold. (**e**) S-CNN+SVM. (**f**) SRDL. (**g**) Our proposed SSJLN. (**h**) Ground truth. The unchanged samples are black and changed samples are white.

It is noteworthy that our proposed SSJLN achieves the better results than those of SRDL. The number of noise spots on the areas at the bottom and top-right corners of the change detection map on Minfeng data set is reduced, as shown in Figure 10g. The number of isolated pixels on the farmland along the riverway significantly decreases, as shown in Figure 11g. It can be clearly seen that the binary change map obtained from our proposed SSJLN is more consistent with the ground truth compared with other comparison methods.

The quantitative analysis is applied to our proposed SSJLN and other comparison algorithms by calculating the OA, Kappa coefficients and AUC values as listed in Table 3. Our proposed SSJLN can obtain a higher OA, Kappa coefficients and AUC values than other comparison methods on the Hongqi Canal data set. But SSJLN has poorer performance than SRDL on the Minfeng data set. The reason is that the training samples of SRDL are more discriminative than those of our proposed SSJLN. We randomly select the changed and unchanged pairwise patches from the changed and unchanged sample set, while SRDL applies a voting rule in the process of selecting samples to obtain the changed and unchanged pairwise patches with discriminative properties. In addition, the pre-classification results of SDRL are acquired from features that consider spectral, texture and spatial information, which will result in more accurate classification results. Furthermore, Minfeng data set has more complex texture than Hongqi data set. The OA, Kappa coefficients and AUC values of the proposed method and all comparison methods on the Minfeng data set are worse than those on the Hongqi data set. Minfeng data set is more challenging for the change detection task. However, the difference between OA, Kappa coefficient and AUC values of SDRL and our proposed SSJLN on Minfeng data set is very small. Our proposed SSJLN outperforms SDRL on three other data sets.

The OA, Kappa coefficients and AUC values obtained by SSJLN on Minfeng data set are 0.9494, 0.6506 and 0.9705, respectively. The OA, Kappa coefficients and AUC values obtained by SSJLN on Hongqi Canal data set are 0.9746, 0.8490 and 0.9889, respectively. It can be concluded that our proposed SSJLN can surpass other comparison methods based on qualitative and quantitative analysis.

5. Conclusions

In this paper, we propose a spectral-spatial joint learning network to not only consider the spatial information but also explore the underlying information of the fused features, which contains three

parts: spectral-spatial joint representation, feature fusion, and discrimination learning. Moreover, we present a new loss function that both consider the losses of the spectral-spatial joint representation part and the discrimination learning part. The effectiveness of the proposed SSJLN is verified on four real data sets in multispectral images. Our proposed SSJLN can exhibits better performance compared with other methods. Although using the patch as input can both consider the spectral and spatial information, the contributions of all pixels in the patch are the same. In fact, the pixel under test at the center of the patch should give more contributions than other surrounding pixels during the procedure of feature extraction. Hence, we plan to focus on the influence of the central pixel in the patch when learning deep features in the future work. We will pay more attention to the influence of the sample imbalance on the supervised changed detection model as well as the generalization ability from the long-term perspective.

Author Contributions: All authors made contributions to proposing the method, doing the experiments and analyzing the result. All authors are involved in the preparation and revision of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 61761130079, in part by the Key Research Program of Frontier Sciences, CAS under Grant QYZDY-SSW-JSC044, in part by the National Natural Science Foundation of China under Grant 61772510, and in part by the Young Top-notch Talent Program of Chinese Academy of Sciences under Grant QYZDB-SSW-JSC015.

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

References

1. Radke, R.J.; Andra, S.; Al-Kofahi, O.; Roysam, B. Image change detection algorithms: A systematic survey. *IEEE Trans. Image Process.* **2005**, *14*, 294–307. [CrossRef] [PubMed]
2. Lu, D.; Mausel, P.; Brondizio, E.; Moran, E. Change detection techniques. *Int. J. Remote Sens.* **2004**, *25*, 2365–2401. [CrossRef]
3. Ridd, M.K.; Liu, J. A comparison of four algorithms for change detection in an urban environment. *Remote Sens. Environ.* **1998**, *63*, 95–100. [CrossRef]
4. 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]
5. Ban, Y.; Yousif, O.A. Multitemporal spaceborne SAR data for urban change detection in China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1087–1094. [CrossRef]
6. 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**, *4*, 3248–3261. [CrossRef]
7. Jae, H.; Chang, L. Urban change detection between heterogeneous images using the edge information. *J. Korean Soc. Surv. Geod. Photogramm. Cartogr.* **2015**, *33*, 259–266.
8. Huang, X.; Friedl, M.A. Distance metric-based forest cover change detection using MODIS time series. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *29*, 78–92. [CrossRef]
9. Demir, B.; Bovolo, F.; Bruzzone, L. Updating land-cover maps by classification of image time series: A novel change-detection-driven transfer learning approach. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 300–312. [CrossRef]
10. Jin, S.; Yang, L.; Danielson, P.; Homer, C.; Fry, J.; Xian, G. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. *Remote Sens. Environ.* **2014**, *29*, 78–92. [CrossRef]
11. Gueguen, L.; Hamid, R. Toward a generalizable image representation for large-scale change detection: Application to generic damage analysis. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3378–3387. [CrossRef]
12. Rokni, K.; Ahmad, A.; Selamat, A.; Hazini. S. Water feature extraction and change detection using multitemporal landsat imagery. *Remote Sens.* **2014**, *6*, 4173–4189. [CrossRef]
13. Muñoz-Marí, J.; Bovolo, F.; Gómez-Chova, L.; Bruzzone, L.; Camp-Valls, G. Semisupervised one-class support vector machines for classification of remote sensing data. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3188–3197. [CrossRef]

14. Volpi, M.; Tuia, D.; Bovolo, F.; Kanevski, M.; Bruzzone, L. Supervised change detection in VHR images using contextual information and support vector machines. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *20*, 77–85. [CrossRef]
15. Bovolo, F.; Bruzzone, L. A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 218–236. [CrossRef]
16. Wu, C.; Du, B.; Zhang, L. Slow Feature Analysis for Change Detection in Multispectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2858–2874. [CrossRef]
17. Wu, C.; Du, B.; Zhang, L. A subspace-based change detection method for hyperspectral images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 815–830. [CrossRef]
18. Zhang, Z.; Tian, Z.; Ding, M.; Basu, A. Improved robust kernel subspace for object-based registration and change detection. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 791–795. [CrossRef]
19. Ma, J.; Zhou, H.; Zhao, J.; Gao, Y.; Jiang, J.; Tian, J. Robust feature matching for remote sensing image registration via locally linear transforming. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6469–6481. [CrossRef]
20. Deng, J.; Wang, K.; Deng, Y.; Qi, G. PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. *Int. J. Remote Sens.* **2008**, *29*, 4823–4838. [CrossRef]
21. Canty, M.J.; Nielsen, A. Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation. *Remote Sens. Environ.* **2008**, *112*, 1025–1036. [CrossRef]
22. Zhan, Y.; Fu, K.; Yan, M.; Sun, X.; Wang, H.; Qiu, X. Change Detection Based on Deep Siamese Convolutional Network for Optical Aerial Images. *Remote Sens. Environ.* **2017**, *14*, 1845–1849. [CrossRef]
23. Dicarlo, J.J.; Zoccolan, D.; Rust, N.C. How does the brain solve visual object recognition? *Neuron* **2012**, *73*, 415–434. [CrossRef] [PubMed]
24. Zhang, H.; Gong, M.; Zhang, P.; Su, L.; Shi, J. Feature-level change detection using deep representation and feature change analysis for multispectral imagery. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1666–1670. [CrossRef]
25. Zhu, X.; Tuia, D.; Mou, L.; Xia, G.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]
26. Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]
27. 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]
28. Wu, K.; Zhong, Y.; Wang, X.; Sun, W. A Novel Approach to Subpixel Land-Cover Change Detection Based on a Supervised Back-Propagation Neural Network for Remotely Sensed Images With Different Resolutions. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1750–1754. [CrossRef]
29. Hou, B.; Wang, Y.; Liu, Q. Change Detection Based on Deep Features and Low Rank. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2418–2422. [CrossRef]
30. Melekhov, I.; Kannala, J.; Rahtu, E. Siamese network features for image matching. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 378–383.
31. Liu, J.; Gong, M.; Qin, K.; Zhang, P. A deep convolutional coupling network for change detection based on heterogeneous optical and radar images. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 545–559. [CrossRef]
32. Liu, B.; Yu, X.; Zhang, P.; Yu, A.; Fu, Q.; Wei, X. Supervised deep feature extraction for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 1909–1921. [CrossRef]
33. Zagoruyko, S.; Komodakis, N. Learning to compare image patches via convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4353–4361.
34. Bertinetto, L.; Valmadre, J.; Henriques, J.F.; Vedaldi, A.; Torr, P. Fully-convolutional siamese networks for object tracking. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, 850–865.
35. Varior, R.R.; Haloi, M.; Wang, G.P. Gated siamese convolutional neural network architecture for human re-identification. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 791–808.

36. Hadsell, R.; Chopra, S.; LeCun, Y. Dimensionality reduction by learning an invariant mapping. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006; Volume 2, pp. 1735–1742.
37. Zhao, W.; Wang, Z.; Gong, M.; Liu, J. Discriminative Feature Learning for Unsupervised Change Detection in Heterogeneous Images Based on a Coupled Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 7066–7080. [CrossRef]
38. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 807–814.
39. Zanetti, M.; Bruzzone, L.A. Theoretical Framework for Change Detection Based on a Compound Multiclass Statistical Model of the Difference Image. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 1129–1143. [CrossRef]
40. Bovolo, F.; Marchesi, S.; Bruzzone, L. A framework for automatic and unsupervised detection of multiple changes in multitemporal images. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *50*, 2196–2212. [CrossRef]
41. Mou, L.; Ghamisi, P.; Zhu, X. Deep recurrent neural networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3639–3655. [CrossRef]
42. Tian, Y.; Fan, B.; Wu, F. L2-Net: Deep Learning of Discriminative Patch Descriptor in Euclidean Space. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; Volume 1, pp. 661–669.
43. Lu, X.; Yuan, Y.; Zheng, X. Joint dictionary learning for multispectral change detection. *IEEE Trans. Cybern.* **2017**, *47*, 884–897. [CrossRef] [PubMed]