Research Article

A Change Detection Method for Remote Sensing Images Based on Coupled Dictionary and Deep Learning

Weiwei Yang ⓘ, Haifeng Song ⓘ, Lei Du ⓘ, Songsong Dai ⓘ, and Yingying Xu ⓘ

School of Electronics and Information Engineering, Taizhou University, Taizhou 318000, Zhejiang, China

Correspondence should be addressed to Haifeng Song; isshf@tzc.edu.cn

With the rapid development of remote sensing technology, change detection (CD) methods based on remote sensing images have been widely used in land resource planning, disaster monitoring, and urban expansion, among other fields. The purpose of CD is to accurately identify changes on the Earth's surface. However, most CD methods focus on changes between the pixels of multitemporal remote sensing image pairs while ignoring the coupled relationships between them. This often leads to uncertainty about edge pixels with regard to changing objects and misclassification of small objects. To solve these problems, we propose a CD method for remote sensing images that uses a coupled dictionary and deep learning. The proposed method realizes the spatial-temporal modeling and correlation of multitemporal remote sensing images through a coupled dictionary learning module and ensures the transferability of reconstruction coefficients between multisource image blocks. In addition, we constructed a differential feature discriminant network to calculate the dissimilarity probability for the change area. A new loss function that considers true/false discrimination loss, classification loss, and cross-entropy loss is proposed. The most discriminating features can be extracted and used for CD. The performance of the proposed method was verified on two well-known CD datasets. Extensive experimental results show that the proposed method is superior to other methods in terms of precision, recall, F_1-score, IoU, and OA.

1. Introduction

In 1858, the first remote sensing image was taken over Paris, France. Human's cognition of the Earth has entered a new stage from local to large-scale surface observation as a result of remote sensing technology on various platforms. In recent years, with the rapid development of sensor and information technology, it has become possible to acquire and store large amounts of remote sensing images. Remote sensing images provide extensive Earth observational data that are characterized by high precision, wide range, and short period, offering important data support for the military, science, agriculture, environment, resources, transportation, and urban management [1]. Researching the surface ecological environment using these remote sensing images has great societal and scientific significance [2].

Remote sensing image change detection (CD) is an important application of remote sensing image interpretation that aims to compare and analyze two (or more) remote sensing images taken at different times in the same region and obtain change information about the ground object. This technology has been widely used in land planning [3], urban expansion [4], disaster assessment [5], agricultural investigation [6], scene classification [7], and ecological environment monitoring [8], among other fields. Traditional artificial visual interpretation methods mainly detect changes in the obtained remote sensing images using the visual interpretation experience, which has low efficiency and is easily affected by subjective conditions. Therefore, such methods cannot meet current needs for dynamic monitoring of the environment. With the development of remote sensing image CD technology, although new CD methods continue to emerge, many new problems and challenges have arisen due to improvements in imaging ability and different properties of sensors. The main problems and challenges are as follows. First, the continuous

improvement of spatial and spectral resolution of remote sensing images offers abundant ground object information, which significantly increases redundancy and noise. Second, with the continuous development of sensor technology, various new data sources are emerging. This poses a new challenge to the universality of existing remote sensing image processing methods. How to accurately and quickly extract useful information from massive datasets for specific tasks has become an urgent problem. Third, the applications for remote sensing images continuously involve more and more fields. Thus, a single model can no longer complete most of the complex tasks required. Therefore, the existing image classification and CD models must start with a practical application and fully combine the advanced technologies in various fields to meet the needs of different industries. However, for a variety of complex tasks, it is difficult for human beings to obtain an effective and comprehensive model due to limited prior knowledge of the task. Identifying new methods to solve the aforementioned problems has increasingly attracted researchers' attention.

With the rapid development of computer theory and technology, artificial intelligence technology has appeared in image segmentation, image classification, CD, and other related research areas. In particular, the recent emergence of deep learning methods as a powerful driving force of artificial intelligence technology has greatly stimulated the enthusiasm of various research fields to utilize deep learning to solve existing problems. The outstanding deep learning model UNet is widely used in segmentation and achieved good performance in multiobject [9, 10] and multiscale [11] remote sensing image segmentation. In respect of remote sensing image classification, the DRL-GAN is proposed by literature [12]; the network can not only enhance the image representation ability, but also identify tiny objects from remote sensing images. As for remote sensing image translation, the attention mechanism is introduced to the generative adversarial network architecture, and the SPA-GAN model is proposed by literature [13]. With regard to remote sensing image dehazing, considering the CNN's excellent performance in image dehazing, the FD-GAN is proposed by literature [14]; the model can produce more realistic and natural images with less color distortion and fewer artifacts. Concerning remote sensing image object detection, an architecture with ESRGAN, EEN, and detection network is proposed by literature [15]; the quality of remote sensing images is enhanced by using EESRGAN, and the detection performance is improved by backpropagating the loss into the EESRGAN. From the above achievement, we can discover that the rich hierarchical architecture of deep learning networks is conducive to adopting shallow to deep approaches. The target task is analyzed in depth and thoroughly, and the understanding of high-level and high-dimensional features is gradually formed through low-level information analysis. The strong feature learning and expression abilities of deep learning models can be used to realize automatic robust learning if the model is able to effectively utilize a large amount of existing information. A large number of training models based on deep learning methods within the category of supervised learning can

provide sufficient reference for massive multisource datasets, greatly improving the utilization of data and the timeliness of a variety of remote sensing tasks, including CD.

At present, deep learning methods have been used to carry out a great deal of research in the field of remote sensing image target extraction. The theory underlying such methods should also be applicable to the field of CD. High-level semantic features such as spectrum, space, and texture can be extracted from multitemporal images using deep learning to establish a nonlinear relationship between ground object features in multitemporal images. Remote sensing CD models based on deep learning are different from traditional remote sensing CD models in that deep learning can identify the features of data through representation learning, which can reduce the heavy dependence on manual work during feature extraction in remote sensing image CD. Therefore, researching remote sensing image CD methods based on deep learning can not only improve the accuracy of remote sensing image CD, but also offer a new automatic and intelligent processing method for realizing current application requirements.

The main contributions of the paper can be summarized as follows:

(1) In this paper, a change detection framework based on coupled dictionary learning is proposed to map multisource and multitemporal data from low-dimensional space to high-dimensional space and then estimate its change.

(2) A differential feature extraction network is proposed to construct a new reconstruction loss by integrating true-false discrimination loss, classification loss, and cross-entropy loss, which is conducive to training the coupled dictionary with stronger representation ability.

(3) The multisource remote sensing image change detection is realized by combining the advantages of model driven method (coupled dictionary learning) and data driven method (differential feature extraction network), and data from different observation spaces can be modeled uniformly.

The rest of this paper is organized in four sections. Section 2 is a description for related works of the change detection task we are dealing with. Section 3 provides details on the proposed model. Section 4 deals with the experimental results and discussion. Finally, Section 5 summarizes the observations and points out some probable future works to complete this paper.

2. Related Works

CD refers to the process of observing the same object on the Earth's surface at different times to determine its different states. The task of CD is to answer three questions about object change: (1) Has it changed? (2) Where has the change occurred? (3) What has changed? With the increasing resolution of sensors, the acquisition period of remote sensing images has gradually shortened, and a large number of CD methods based on remote sensing images have been proposed.

In recent years, the emergence of deep learning technology has provided new concepts for CD in remote sensing images and shown strong ability to solve the problems associated with traditional methods. First, compared with traditional methods, deep learning methods do not require a large amount of existing knowledge. Second, no manual intervention is required during the training process, which greatly improves the automation of remote sensing image CD. Third, since deep neural networks have strong learning and feature expression abilities, they are more suitable for processing high-dimensional data than traditional methods. Therefore, remote sensing image CD methods based on deep learning methods are attracting increasing attention. Many deep learning methods, such as NN [16], FCN [17], DenseNet [18], UNet [19], DeepLab [20], and SegNet [21], have been introduced for remote sensing image applications and have achieved extremely high accuracy. Some research results have even approached or attained the accuracy of human vision systems. Previous literature [22] has used an end-to-end recurrent neural network with a long short-term memory (LSTM) network to learn change rules of land cover. Literature [23] trained a deep belief network to classify changing and unchanging regions in SAR images, while literature [24] proposed an end-to-end two-dimensional convolutional neural network for hyperspectral image CD. Other studies [25, 26] have researched the use of generative adversarial networks to detect the change areas of multispectral remote sensing images and other cross-domain remote sensing images. Literature [27] proposed a deep Siamese convolutional network (DSCN) to detect changing regions with contrast loss in images. A symmetric convolutional coupled network (SCCN) based on a DSCN was proposed in literature [28] to detect changes between optical and SAR images. In sum, existing CD methods based on deep learning can autonomously learn high-dimensional features of the changed region without manual intervention and perform CD according to the truth value.

Most CD methods analyze multitemporal remote sensing data obtained by the same sensor or same type of sensor. However, with the development of remote sensing imaging technology, there are more and more multisource remote sensing data covering the same area. The imaging mechanisms and results of these multisource data are quite different, making it difficult to detect changes using traditional CD technology. Although there may be major differences in multisource data, they ultimately constitute different descriptions of the same scene. The imaging makes comparisons difficult in the original low-dimensional space, but having the same ground object makes multisource data comparable in high-dimensional space. On the other hand, although rich data are provided by sensors, these data are often the result of a combination of factors. Therefore, it is necessary to identify the key relevant information to explain and assist in decision-making. In recent years, sparse representation has increasingly been used in remote sensing image CD and data fusion. In 2021, in order to solve the problem that the accuracy of CD is easily affected by speckle noise, a method based on sparse representation and capsule network was proposed in literature [29]. First, sparse features of difference images are extracted using SR. Second, a lightweight capsule network (L-CapsNet) is constructed to mine the spatial relationships between features. Finally, a change map (CM) is generated. Experimental results show that this method can reduce the influence of speckle noise and obtain stronger robust features. An image fusion method based on sparse representation has also been proposed in the literature [30]. The method is based on the assumption that NDVI and HRVI with different resolutions have the same sparse coefficients in a particular dictionary. The application of this method in multisource NDVI CD was verified using GF-1 and GF-2 satellite datasets.

As mentioned above, multisource image pairs are different descriptions of the same object. The data representation and correlation in different observation space constitute a challenging task. Due to the difference of imaging mechanism, multisource images have completely different imaging results. This encouraged us to try to learn two independent and coupled dictionaries to relate different observation spaces. The dictionary can be coupled by forcing multiple image block pairs to share the same sparse representation. Inspired by the above idea, a coupled dictionary learning method is proposed for multisource image change detection. The goal of coupled dictionary learning is to match the elements from different observation spaces, so that they have one-to-one correspondence in the same position, and realize the mapping of multisource data. Through coupling dictionary learning, invariant image block pairs tend to have lower reconstruction errors, while variable image block pairs have higher reconstruction errors. This can suppress the unchanged area, highlight the changed area, and obtain a robust change map. However, the randomly selected training sample may contain some changing training samples, and these changing samples may interfere with the learning of the coupled dictionary. Therefore, we design a new loss function to automatically select training samples with low reconstruction errors and exclude training samples with changes in the training set, so as to build a more robust coupled dictionary. Finally, the multisource data is mapped to a high-dimensional feature space to enhance the variation of multisource data.

3. Proposed Method

Firstly, we give a general introduction to the proposed model in this paper; the input of the model is the original remote sensing images, and the output of the model is the CD results of the corresponding pixel. Then, the data preprocessing method, coupled dictionary learning method, and differential feature extraction network are introduced in detail. At the same time, a loss function for CD is introduced.

3.1. Overview of the Proposed Method. As shown in Figure 1, the method proposed in this paper consists of five steps: input, preprocessing, coupled dictionary learning, CD, and output. First, the input consists of multitemporal and multisource remote sensing image data. Second, in the preprocessing step, local image blocks of two different
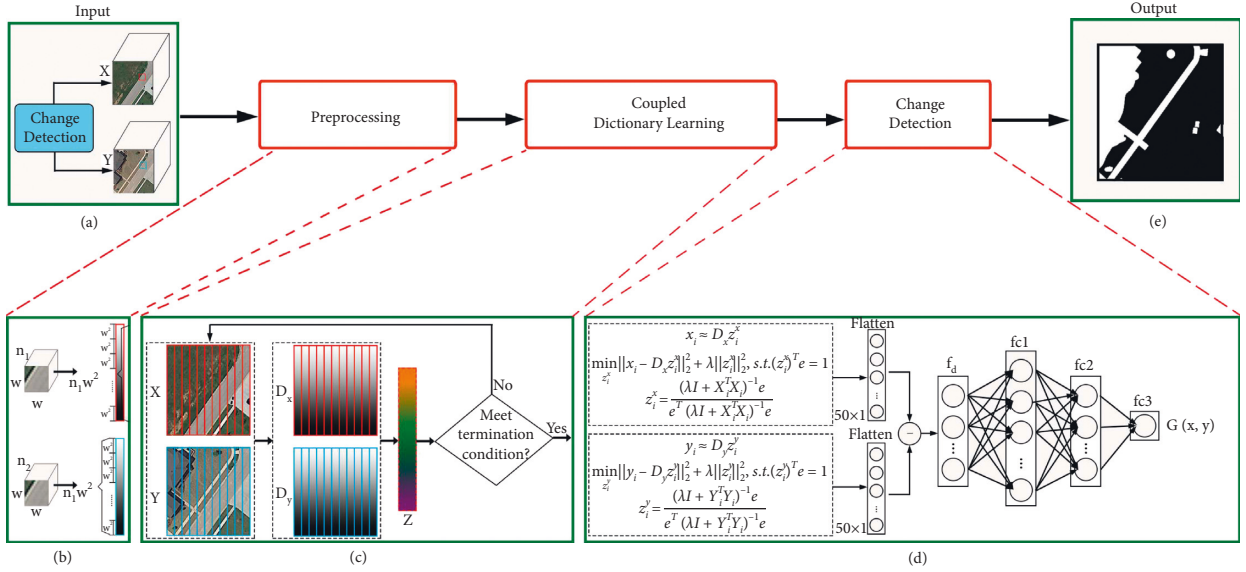
FIGURE 1: Overview of the proposed method.

periods data obtained by different sensors on the same scene are expanded into vector representations. Third, each image block pair is stacked as a training sample for coupled dictionary learning and forced to share the same reconstruction coefficient. After the coupled dictionary is learned, it can be determined whether the current coupled dictionary meets the termination condition. If not, the training sample with low reconstruction error is chosen as the next generation of the training set to learn the dictionary pairs with higher coupled degrees (and, if necessary, the process is repeated). Fourth, based on the coupled dictionary learned in the second step, the reconstruction coefficient of each image block pair is calculated and input into the difference feature extraction network to estimate the change degree. Finally, the final binary CD result is produced as output.

3.2. Data Preprocessing.

The concept of sparsity first appeared in the 1970s in the context of the efficient coding hypothesis in neurobiology. Today, this idea is widely used in the field of image processing. The following preprocessing should be carried out before attempting remote sensing shadow CD.

Let $X = \{x_i\}_{i=1}^N$ and $Y = \{y_i\}_{i=1}^N$ represent n pairs of image blocks. Each pair (x_i, y_i) is randomly sampled from two multisource remote sensing images taken at different times in the same geographical location. In the proposed method, two image blocks (x_i, y_i) at the same position are regarded as a pair of image blocks. As shown in Figure 1(b), an image block of size $w \times w \times n_t$ $(t = 1, 2)$ can be represented as R^{m_t}, where $m_t = n_t w^2$. x_i and y_i can be regarded as column vectors with dimensions m_1 and m_2, respectively.

3.3. Coupled Dictionary Learning.

The typical problems related to sparse representation are expressed as in the following equation:

$$\min_{D, \alpha_i} \sum_{i=1}^N \left\| x_i - D\alpha_i \right\|_2^2 + \lambda \left\| \alpha_i \right\|_1,$$

$$\text{s.t. } \left\| d_j \right\|^2 \leqslant 1, j = 1, 2, \ldots, k. \tag{1}$$

The ith image block is represented as x_i; $D = [d_1, d_2, \ldots, d_n] \in R^{m \times d}$ represents the dictionary corresponding to the image; α_i represents the sparse coefficient, which is obtained by sparse representation of the ith image block x_i according to dictionary D; λ is the regular parameter; and $\lambda \|\alpha_i\|_1$ ensures the sparsity of vector α_i. Finally, the constraint on $\|d_i^2\|^2$ ensures that each dictionary atom is homogeneously distributed in order to avoid too-large values that will increase the time necessary for optimization.

The coupled dictionary learning method for remote sensing images proposed in this paper is an extension of the above methods. Let $D_x = [d_1^x, d_2^x, \ldots, d_n^x] \in R^{m_1 \times d}$ and $D_y = [d_1^y, d_2^y, \ldots, d_n^y] \in R^{m_2 \times d}$ contain d elements (respectively, $d_i^x \in R^{m_1}$ and $d_i^y \in R^{m_2}$). The two images are represented by $X = x_{i}{}_{i=1}^n$ and $Y = y_{i}{}_{i=1}^n$. Image blocks x_i and y_i share the same reconstruction coefficient. The sparse representation of the two images can be combined into an equation as follows:

$$\min_{D_x, D_y, \alpha_i} \sum_{i=1}^N \left\{ \frac{1}{2} \left\| x_i - D_x \alpha_i \right\|_2^2 + \lambda \left\| \alpha_i \right\|_1 \right\}$$

$$+ \left\{ \frac{1}{2} \left\| y_i - D_y \alpha_i \right\|_2^2 + \lambda \left\| \alpha_i \right\|_1 \right\}, \tag{2}$$

$$\text{s.t. } \left\| d_i^2 \right\|^2 \leqslant 1, i = 1, 2, \ldots, k.$$

Let $\overline{X} = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$, $\overline{D} = \begin{bmatrix} D_x \\ D_y \end{bmatrix}$; (2) can be written as follows:

$$\min_{D,\alpha_i} \sum_{i=1}^{N} \left\| \overline{X} - \overline{D}\alpha_i \right\|_2^2 + \lambda \left\| \alpha_i \right\|_1, \tag{3}$$

$$\text{s.t. } \left\| \overline{d_j} \right\|^2 \leqslant 1, \, j = 1, 2, \ldots, k,$$

where $\overline{d_j}$ is the ith element of the coupled dictionary $\overline{D}$.

After the above combination, the complex equations are simplified into sparse representation objective functions, and the corresponding optimization process is also simplified. Equation (3) is not a convex function, which is not convenient for optimization. However, when one of the values is fixed, the above equation becomes a convex function that can be optimized. Therefore, the alternate iteration method is used to optimize it. The specific calculation method is as follows:

(1) Optimizing coupled sparse coefficient α_i. According to the above analysis, dictionary $\overline{D}$ must be fixed before coupled sparse coefficient α_i is optimized. For the first iteration, $\overline{D}$ is unknown. The dictionary is initialized by randomly selecting data in the training set or randomly generating a matrix. After dictionary $\overline{D}$ is determined, the objective function becomes a sparse coding problem, and the sparse coefficients are calculated according to the categories. When the ith sparse coefficient α_i is calculated, the remaining sparse coefficients that do not belong to the $\alpha_i \, (i \neq j)$ are fixed. The following equation can be used to solve the sparse coefficient α_i:

$$\min_{\alpha_i} \left\| \overline{x_i} - \overline{D_x}\alpha_i \right\|_2^2 + \left\| \overline{x_i} - \overline{D_x}\alpha_i^i \right\|_2^2$$
$$+ \sum_{j=1, j \neq i}^{c} \left\| \overline{D_j}\alpha_i^j \right\|_2^2 + \lambda \left\| \alpha_i \right\|_1. \tag{4}$$

The iterative projection method [31] is used to solve (4).

(2) Optimizing dictionary $\overline{D}$. In this optimization process, the sparse coefficient α_i is a known condition. The method of updating the dictionary is similar to the previous procedure. When the dictionary $\overline{D_i}$ belonging to the ith category is computed, the rest of the dictionaries $\overline{D_j}, \, (j \neq i)$ are fixed. The objective function can be simplified to the following equation:

$$\min_{\overline{D_i}} \left\| \overline{x_i} - \overline{D_x}\alpha_i - \sum_{j=1, j \neq i}^{c} \overline{D_x}\alpha_i \right\|_2^2$$
$$+ \left\| \overline{x_i} - \overline{D_x}\alpha_i^i \right\|_2^2 + \sum_{j=1, j \neq i}^{c} \left\| \overline{D_j}\alpha_j^i \right\|_2^2. \tag{5}$$

$\overline{x_i}$ represents the sparse coefficient obtained by sparse representation of training image set $\overline{X}$ on dictionary $\overline{D}$, which belongs to the ith category.

3.4. Differential Feature Extraction Network.
After obtaining the coupled dictionary D_x and D_y, the multisource remote sensing image data is mapped to the high-dimensional feature space. In this space, changing image blocks will have different reconstruction coefficients, while unchanging image blocks will have similar reconstruction coefficients.

The change degree of each image block pair can be measured by comparing the reconstruction coefficients α_i^x and α_i^y of the ith pair of images x_i and y_i with their coupled dictionaries D_x and D_y. The calculation method for α_i^x and α_i^y is shown in (6), (7), and (8).

$$\begin{cases} x_i \approx D_x \alpha_i^x, \\ y_i \approx D_x \alpha_i^y. \end{cases} \tag{6}$$

α_i^x and α_i^y can be solved as in

$$\begin{cases} \min_{z_i^x} \left\| x_i - D_x \alpha_i^x \right\|_2^2 + \lambda \left\| \alpha_i^x \right\|_2^2, \\ \\ \min_{z_i^y} \left\| y_i - D_y \alpha_i^y \right\|_2^2 + \lambda \left\| \alpha_i^y \right\|_2^2. \end{cases} \tag{7}$$

λ is the regular parameter. Equation (7) can be solved as in

$$\begin{cases} \alpha_i^x = \dfrac{\left(\lambda I + X_i^T X_i\right)^{-} 1e}{e^T \left(\lambda I + X_i^T X_i\right)^{-} 1e}, \\ \\ \alpha_i^y = \dfrac{\left(\lambda I + Y_i^T Y_i\right)^{-} 1e}{e^T \left(\lambda I + Y_i^T Y_i\right)^{-} 1e}. \end{cases} \tag{8}$$

e is a vector of all 1s in R^d, $X_i = [d_1^x - x_i, d_2^x - x_i, \ldots, d_d^x - x_i] \in R_1^m \times d$, $Y_i = [d_1^y - y_i, d_2^y - y_i, \ldots, d_d^y - y_i] \in R_2^m \times d$, and I is the unit matrix of $d \times d$.

Since the goal of CD is to detect the differences between two images taken at different times, the difference feature vectors are calculated after the feature vectors α_i^x and α_i^y of the two images are obtained using the above method. The difference feature vector α_d is calculated as in

$$\alpha_d = \alpha_i^x - \alpha_i^y. \tag{9}$$

In order to learn the dissimilarity probability between two images from the difference feature vectors, a difference feature extraction network is designed after the difference feature vectors are obtained. The network consists of three fully connected layers, f_1, f_2, f_3, as shown in Figure 1(d). The value of its hidden layer can be calculated using the following equation:

$$h^{(k)} = \text{ReLu}\left(W^{(k)} h^{(k-1)} + b^{(k)}\right). \tag{10}$$

$h^{(k)}$ and $h^{(k-1)}$ are the output matrices of the kth and $(k-1)^{\text{th}}$ layers. W_k is the weight that is input to the hidden layer, $b^{(k)}$ is the bias, and because we need is a binary value of 0 or 1, the ReLU $(\cdot)$ is selected as the activation function. When the information is transmitted to the last fully connected layer $fc3$, a value ranging from 0 to 1 is output, indicating the probability of dissimilarity.

For the above differential feature discriminant network, it is expected that the generated dissimilarity probability can be judged as the correct category label and that the obtained dissimilarity probability will be large for the changing category and small for the unchanging category. Therefore,

cross-entropy loss is added to its loss function. Its loss function consists of three parts, as shown in the following equation:

$$L_G = \min\{L_{\text{adv}} + L_{\text{cls}} + L_E\}. \tag{11}$$

L_{adv} represents true/false loss, L_{cls} represents category loss, and L_E represents cross-entropy loss. They are calculated as in the following equations:

$$L_{\text{adv}} = E_{X,Y\sim p(X,Y)}[\log(1 - D(G(X,Y)))], \tag{12}$$

$$L_{cls} = E[\log(C = c|y_{\text{fake}})], \tag{13}$$

$$L_E = \frac{1}{N}\left[\sum_{i=1}^{N} y^i y^i_{\text{fake}} + \left(1 - y^i\right)\log\left(1 - y^i_{\text{fake}}\right)\right]. \tag{14}$$

N is the number of training samples, and the value of y^i is 0 or 1, which indicates the label of the input neighborhood image block pairs. 1 indicates that they have changed, and 0 indicates that they have not changed. y^i_{fake} represents the probability of dissimilarity of input neighborhood image block pairs.

For the optimize algorithm, SGD algorithm is a famous parameter optimization method proposed, which is widely used in model parameter optimization of deep learning. However, its disadvantage is that it has a slow descent speed and may oscillate near the minimum value [32]. Therefore, SGD method based on first-order momentum is proposed [33]. This method updates each parameter with the same learning rate. However, deep neural network often contains a large number of parameters. For the parameters which are updated frequently, the model has accumulated a lot of knowledge about them; we hope that they are affected by a single sample, so the learning rate should be slower. However, for parameters that are occasionally updated, the model has little accumulated knowledge, and we hope that more knowledge can be learned from each random sample, so the learning rate should be higher. Therefore, an optimization method named AdaGrad based on second-order momentum is proposed [34]. This method performs well for sparse data. However, it may reduce the learning rate

monotonously to zero, thus ending the training process prematurely. Adam [35] optimization method solves the above problems completely. This method integrates first-order momentum and second-order momentum, and the coefficient is used to control the first-order momentum and second-order momentum, which is more conducive to large-scale remote sensing image optimization. Therefore, the Adam algorithm is used to optimize loss function L_G. During the training process, all parameters of the network are fixed first, and they are updated. When L_G training tends toward stability, this indicates that the network has reached a convergence state and the training process can be stopped.

4. Experiment and Discussion

4.1. Dataset. The first dataset is the CDD [36], shown in Figure 2. It consists of seven pairs of images with 4725 × 2700 resolution obtained from Google Earth. A total of 16,000 image pairs (10,000 in the training set, 3,000 in the test set, and 3,000 in the validation set) with a spatial resolution of 3 − 100 cm were collected from the seven pairs of images.

The second dataset is the LEVIR-CD dataset [37], shown in Figure 3. The dataset contains 637 pairs of CD 1024 × 1024 images with a spatial resolution of 0.5 m. The images were collected from 20 different locations in Texas between 2002 and 2018. This dataset helps train more efficient CD models while focusing on interesting changes and reducing the impact of other irrelevant changes on the model.

4.2. Evaluation Metrics. For CD, it is necessary to prove the effectiveness and robustness of the proposed method using quantitative analysis. For performance evaluation purposes, CD results are usually presented as binary images, wherein white pixels represent changed pixels and black pixels represent unchanged pixels. In this experiment, precision, recall, F_1-score, OA, and IoU were used to evaluate the proposed methods. The calculation methods are shown in the following equations:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{15}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{16}$$

$$F1 - \text{score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}, \tag{17}$$

$$OA = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{18}$$

$$\text{IoU} = \frac{(\text{TP}/(\text{TP} + \text{FP} + \text{FN}) + \text{TN}/(\text{TN} + \text{FN} + \text{FP}))}{2}. \tag{19}$$

Figure 2: The samples of CDD.

Figure 3: The samples of LEVIR-CD dataset.

4.3. Experimental Setting. For the above two remote sensing image CD datasets, experiments were carried out to verify the accuracy and effectiveness of the proposed method. For the CDD, the training dataset contained 10,000 images, the test dataset and validation dataset contained 3,000 images each, and the resolution of each image was 256×256. For the LEVIR-CD dataset, the training dataset contained 445 images, the test dataset contained 128 images, the validation dataset contained 64 images, and the resolution of each image was 1024×1024.

The CPU of the computer used in this experiment was an Intel Xeon Silver 4212, 2.2 GHz, with a memory size of 512 GB. The GPU was an NVIDIA Tesla V100, 32 GB. The CUDA version was 10.2. The operating system was Microsoft Windows Server 2016 Datacenter. The *Python* version was 3.6.10, and the deep learning framework was PyTorch version 1.9.0.

During the training process, due to the low resolution of the CDD, its batch size was set to 16, while due to the high resolution of the LEVIR-CD dataset, its batch size was set to 2 in consideration of the capacity of the system. The initial

TABLE 1: Ablation experiments results for the LEVID-CD dataset.

Methods	Precision	Recall	F_1-score	IoU	OA
Sparse representation	0.724	0.715 8	0.711	0.718	0.854
+ CDL	0.796	0.785	0.801	0.763	0.895
+ CDL + DFE	0.812	0.806	0.828	0.783	0.913
+ CDL + DFE + L_G	0.883	0.861	0.872	0.860	0.985

TABLE 2: Comparison experiment results of CDD.

Methods	Precision	Recall	F_1-score	IoU	OA
FC-EF	0.815	0.761	0.771	0.698	0.941
FCN-PP	0.826	0.805	0.871	0.737	0.954
FC-Siam-Conc	0.844	0.825	0.825	0.751	0.957
FC-Siam-Diff	0.858	0.836	0.837	0.771	0.958
CDNet	0.827	0.817	0.822	0.797	0.964
CD-UNet++	0.876	0.859	0.868	0.828	0.967
GANs	0.881	0.856	0.865	0.847	0.974
Proposed method	0.883	0.861	0.872	0.860	0.985

learning rate was set to 1e-3 and was adjusted automatically by the system according to the training situation. The number of epochs was set to 100. The K-SVD algorithm was used to learn the coupled dictionary. The differential feature extraction network was composed of three fully connected layers (fc_1, fc_2, and fc_3). The range of the numbers in the hidden layer of the differential feature extraction network was determined by empirical formulas and Kolmogorov theorem [38]. The number of nodes was set to 150, 100, and 1, respectively.

4.4. Ablation Analysis.

Ablation analysis was used to illustrate the contribution of the proposed method's component. We assessed the contribution of the component to the model by comparing the model that contains it with the one that does not. Instead of subtracting, we gradually added modules such as coupled dictionary learning (CDL), differential feature extraction (DFE), and L_G loss function (L) onto our proposed baseline network to verify its performance. Compared with LEVID-CD dataset, CDD is a large-scale remote sensing dataset. Therefore, the CDD is used for ablation analysis.

Table 1 shows the results of the ablation experiment. The last row in the table represents the performance of our proposed model. The first row represents the performance of the baseline model (sparse representation). The next row in Table 1 indicates that deploying the coupled dictionary learning module, differential feature extraction, and L_G loss function to the baseline model leads to a significant improvement in all the metrics. Adding the coupled dictionary learning module to the model results in a slight improvement in all the metrics except the recall rate. The F_1-score is improved due to the addition of differential feature extraction. We then introduce the L_G loss function, which raises all the metrics except slightly. This variation can be attributed to L_G loss function, which enhances the representation ability for the proposed method.

4.5. Results and Discussion.

To evaluate the performance and effectiveness of the proposed method, we compared it with several classical methods.

For the CDD, we used seven state-of-the-art (SOTA) methods: FC-EF [39], FCN-PP [40], FC-Siam-Conc [39], FC-Siam-Diff [39], CDNet [41], CD-UNet++ [42], and GANs [43]. The experimental results and visual comparisons are shown in Table 2 (some of the experimental results were obtained from [44]) and Figures 4 to 7.

In this experiment, four typical change areas, road, car, single object, and multiple objects, were selected to evaluate the performance of the proposed method. As can be seen from Figures 4 to 7, the proposed method consistently achieved the best performance. The boundary of the proposed method's CM is clearer than those of the other methods, and the misclassification area of the proposed method is smaller than those of the other methods. In addition, the change maps obtained by the proposed method are almost consistent with the reference CM.

Meanwhile, the quantitative results in terms of the five quantitative evaluation metrics listed in Table 2 show that the proposed method achieved excellent performance in terms of precision, recall, F_1-score, IoU, and OA. The OA of the proposed method was 0.883, which is 0.02 higher than that of the GANs method (which had the best precision among the seven SOTA CD methods). The accuracy value was also significantly improved compared with the other methods.

For the LEVID-CD dataset, the experimental results and visual comparisons with the seven SOTA methods (FC-EF [39], FC-Siam-Diff [39], FC-Siam-Conc [39], CD-UNet++ [42], DASNet [45], IFN [46], and GANs [43]) are shown in Table 3 (some of the experimental results were obtained from [47]) and Figures 8 to 11.

In this experiment, four typical change areas, single-object change, more-object change, much-more-object change, and dense-object change, were selected to evaluate the performance of the proposed method. As can be seen in Figures 8 to 11, the proposed method achieved the best
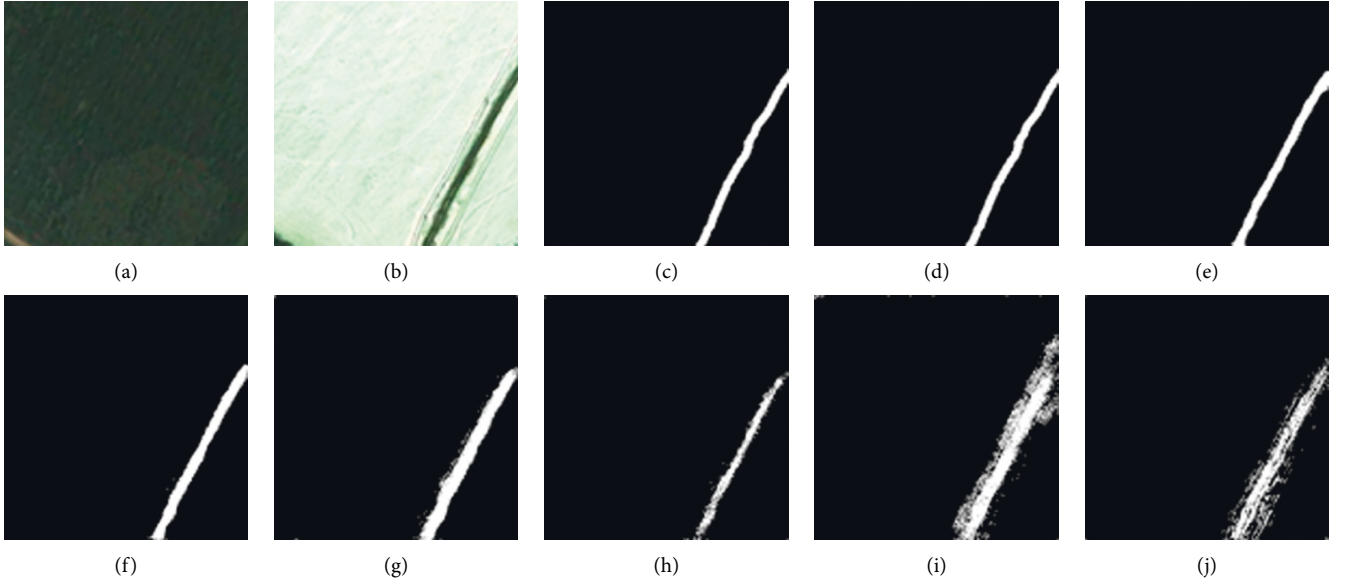
FIGURE 4: Qualitative comparison of different approaches on CDD for road change. (a) T1 image (#0257). (b) T2 image (#0257). (c) Reference change map. (d) Ours. (e) CD-UNet++. (f) CDNet. (g) FC-Siam-Diff. (h) FC-Siam-Conc. (i) FCN-PP. (j) FC-EF.
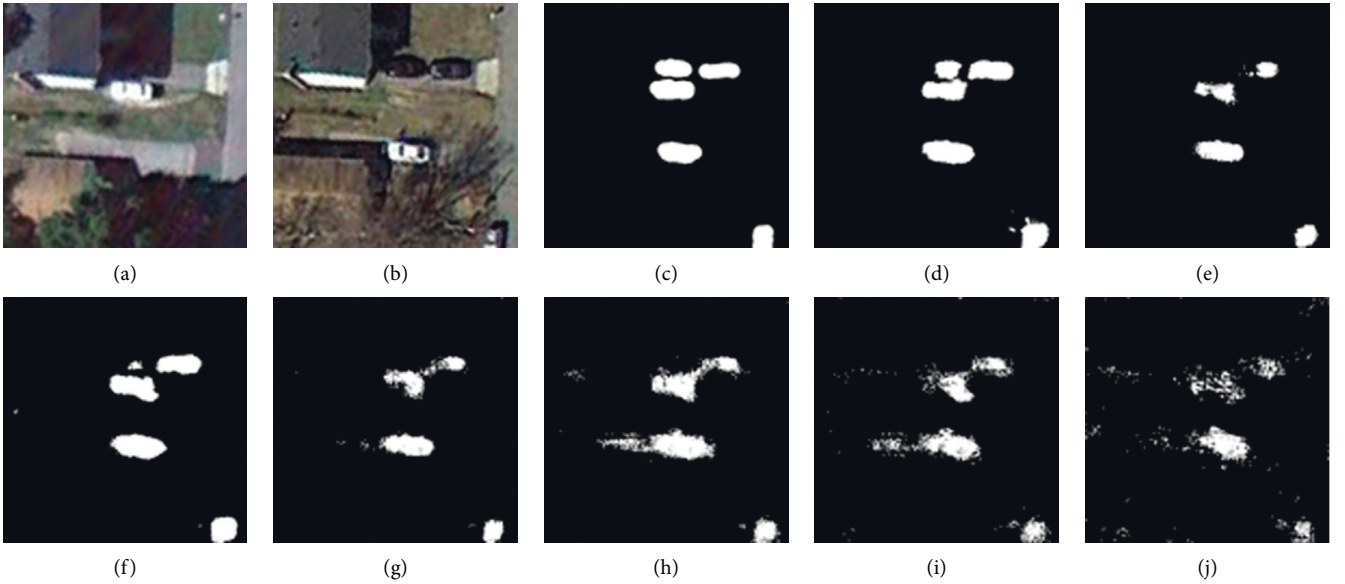

FIGURE 5: Qualitative comparison of different approaches on CDD for car change. (a) T1 image (#0327). (b) T2 image (#0327). (c) Reference change map. (d) Ours. (e) CD-UNet++. (f) CDNet. (g) FC-Siam-Diff. (h) FC-Siam-Conc. (i) FCN-PP. (j) FC-EF.

performance. As the number of objects for CD increases, the proposed method maintains good recognition performance. The proposed method can generate an accurate CM even if the size or number of objects changes.

Meanwhile, the quantitative results on the four quantitative evaluation metrics listed in Table 3 show that the proposed method achieved excellent performance in terms of precision, recall, F_1-score, IoU, and OA. Compared with the seven SOTA CD methods, the precision, recall, F_1-score, IoU, and OA of the proposed method were 0.877, 0.889, 0.873, 0.852, and 0.991, respectively, being 0.031, 0.006, 0.012, 0.005, and 0.004 higher than their counterparts for the GANs method (the second-best method).
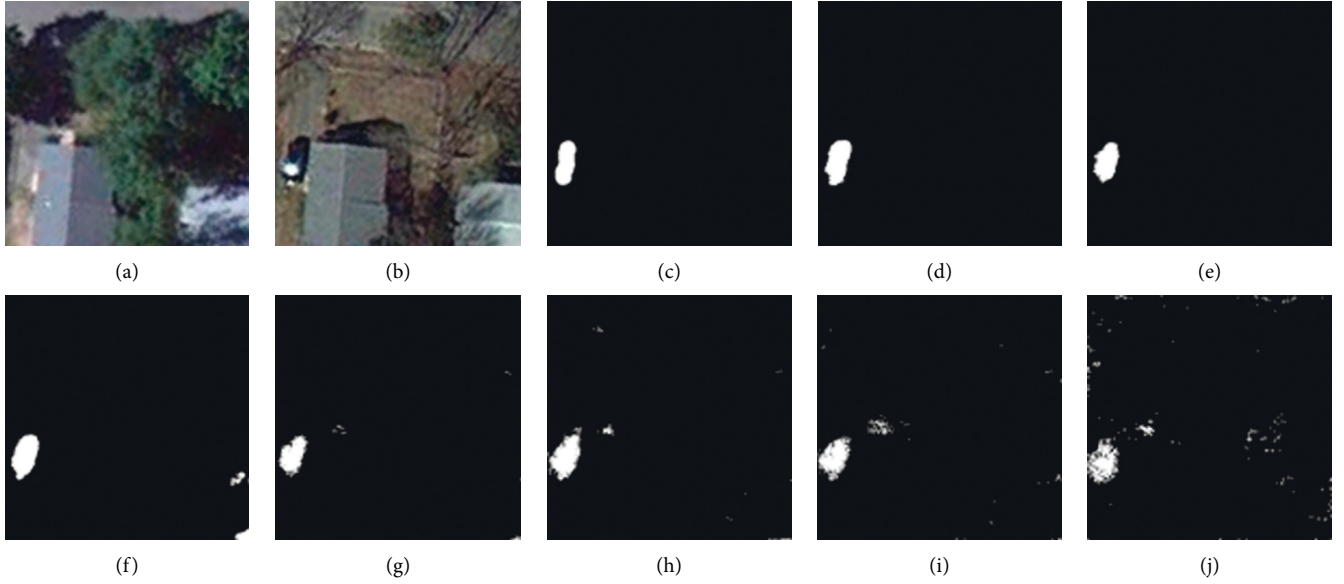
Figure 6: Qualitative comparison of different approaches on CDD for single-object change. (a) T1 image (#0628). (b) T2 image (#0628). (c) Reference change map. (d) Ours. (e) CD-UNet++. (f) CDNet. (g) FC-Siam-Diff. (h) FC-Siam-Conc. (i) FCN-PP. (j) FC-EF.
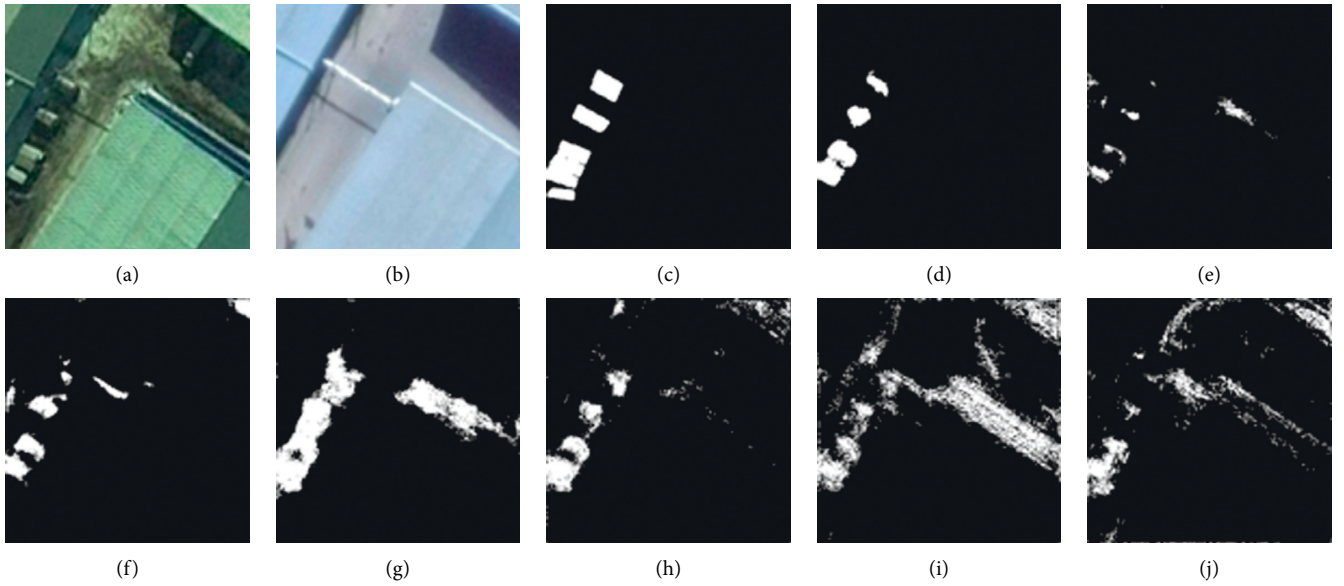

Figure 7: Qualitative comparison of different approaches on CDD for multiple-object change. (a) T1 image (#0283). (b) T2 image (#0283). (c) Reference change map. (d) Ours. (e) CD-UNet++. (f) CDNet. (g) FC-Siam-Diff. (h) FC-Siam-Conc. (i) FCN-PP. (j) FC-EF.

Table 3: Comparison experiment results of the LEVID-CD dataset.

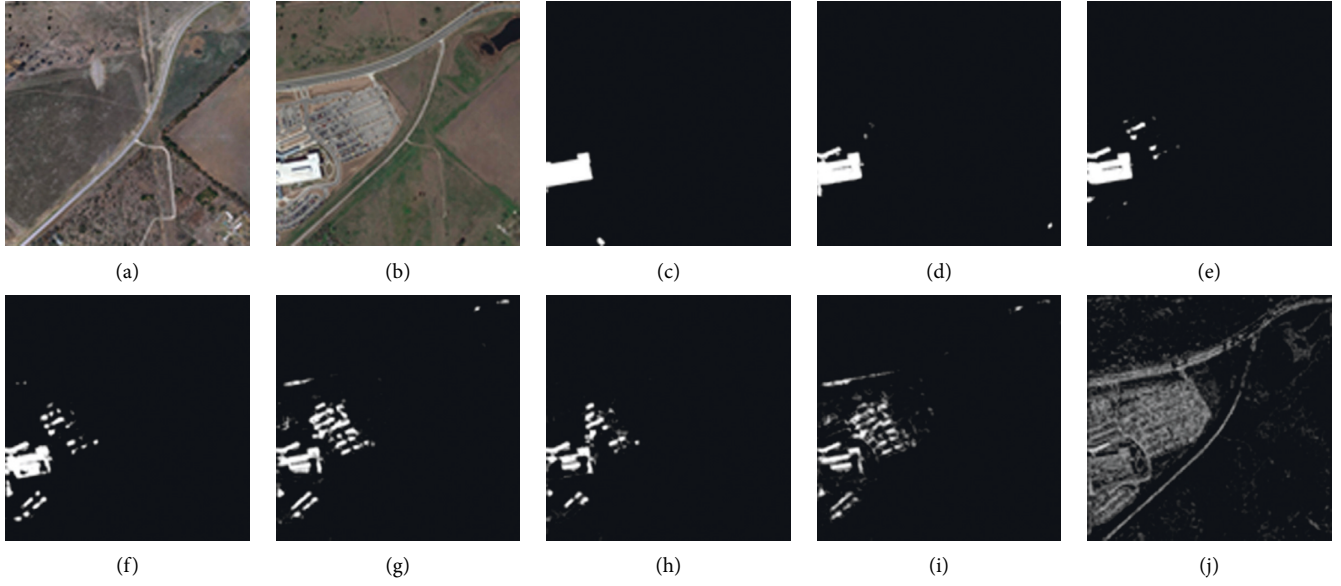| Methods	Precision	Recall	F_1-score	IoU	OA
FC-EF	0.489	0.859	0.623	0.790	0.938
FC-Siam-Diff	0.542	0.732	0.631	0.793	0.960
FC-Siam-Conc	0.604	0.766	0.682	0.822	0.963
CD-UNet++	0.790	0.841	0.815	0.828	0.980 5
DASNet	0.774	0.899	0.828	0.836	0.980 6
IFN	0.796	0.880	0.836	0.839	0.982
GANS	0.846	0.883	0.861	0.847	0.987
Proposed method	0.877	0.889	0.873	0.852	0.991

FIGURE 8: Qualitative comparison of different approaches on LEVID-CD dataset for single-object change. (a) T1 image (#0104). (b) T2 image (#0104). (c) Reference change map. (d) Ours. (e) IFN. (f) DASNet. (g) CD-UNet++. (h) FC-Siam-Conc. (i) FC-Siam-Diff. (j) FC-EF.
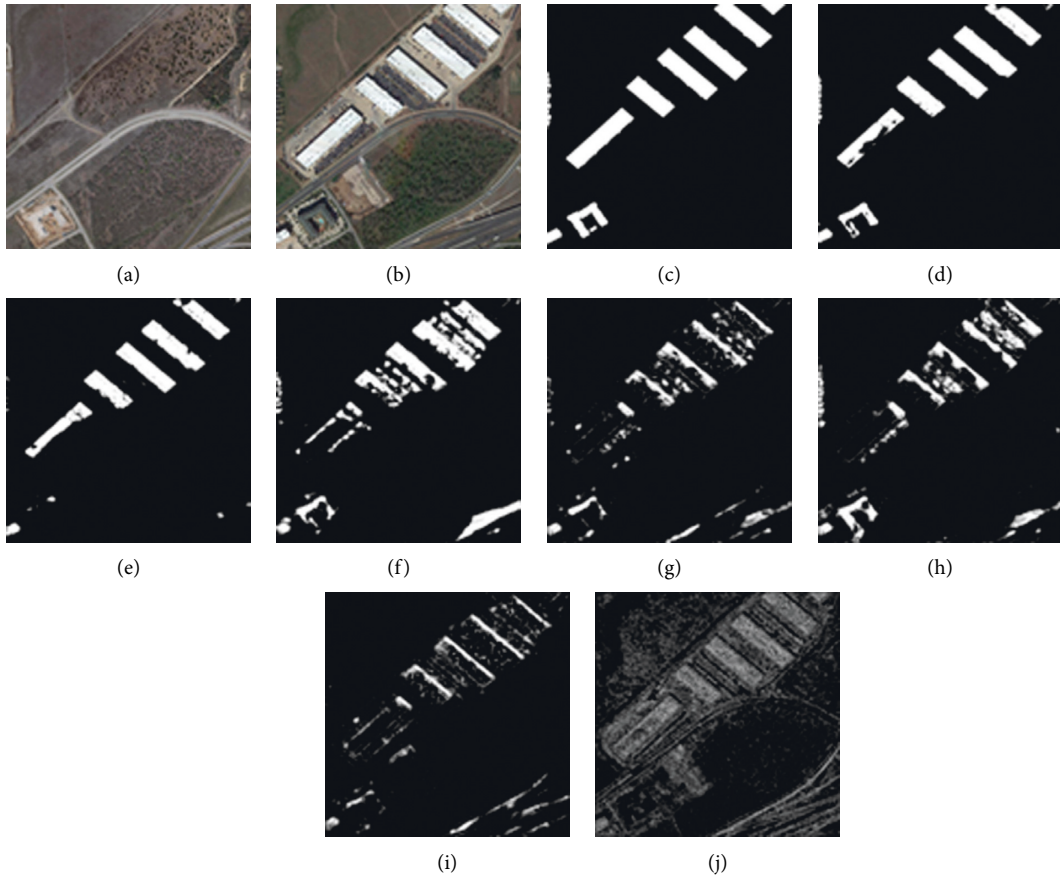

FIGURE 9: Qualitative comparison of different approaches on LEVID-CD dataset for more-object change. (a) T1 image (#0102). (b) T2 image (#0102). (c) Reference change map. (d) Ours. (e) IFN. (f) DASNet. (g) CD-UNet++. (h) FC-Siam-Conc. (i) FC-Siam-Diff. (j) FC-EF.
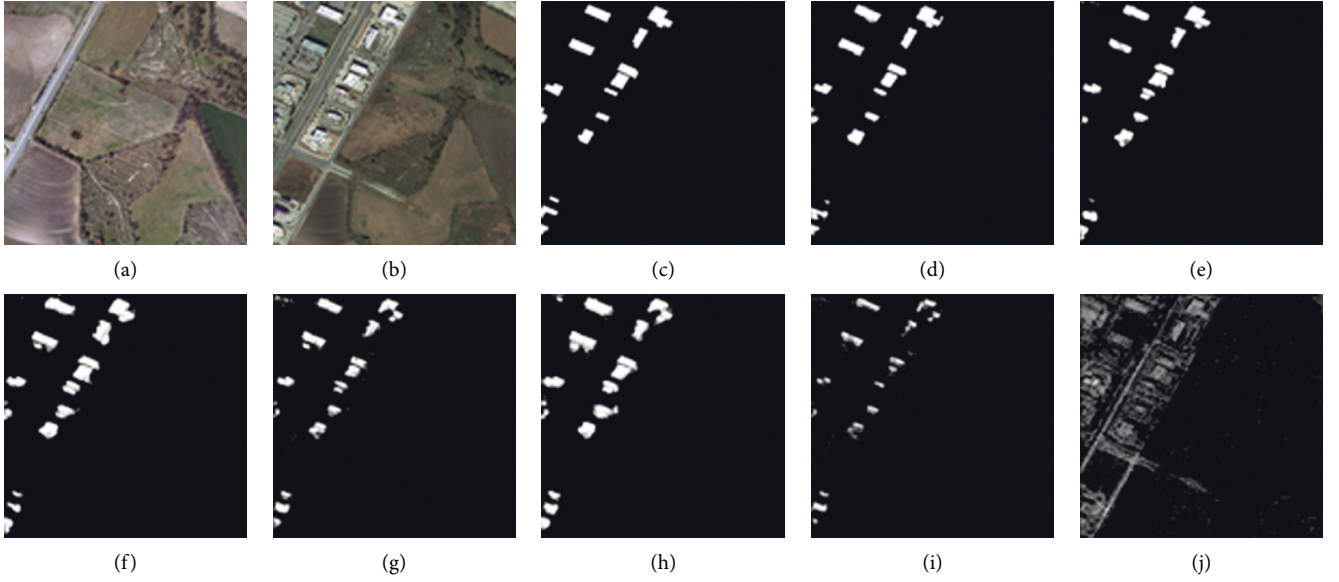
FIGURE 10: Qualitative comparison of different approaches on LEVID-CD dataset for much-more-object change. (a) T1 image (#0054). (b) T2 image (#0054). (c) Reference change map. (d) Ours. (e) IFN. (f) DASNet. (g) CD-UNet++. (h) FC-Siam-Conc. (i) FC-Siam-Diff. (j) FC-EF.
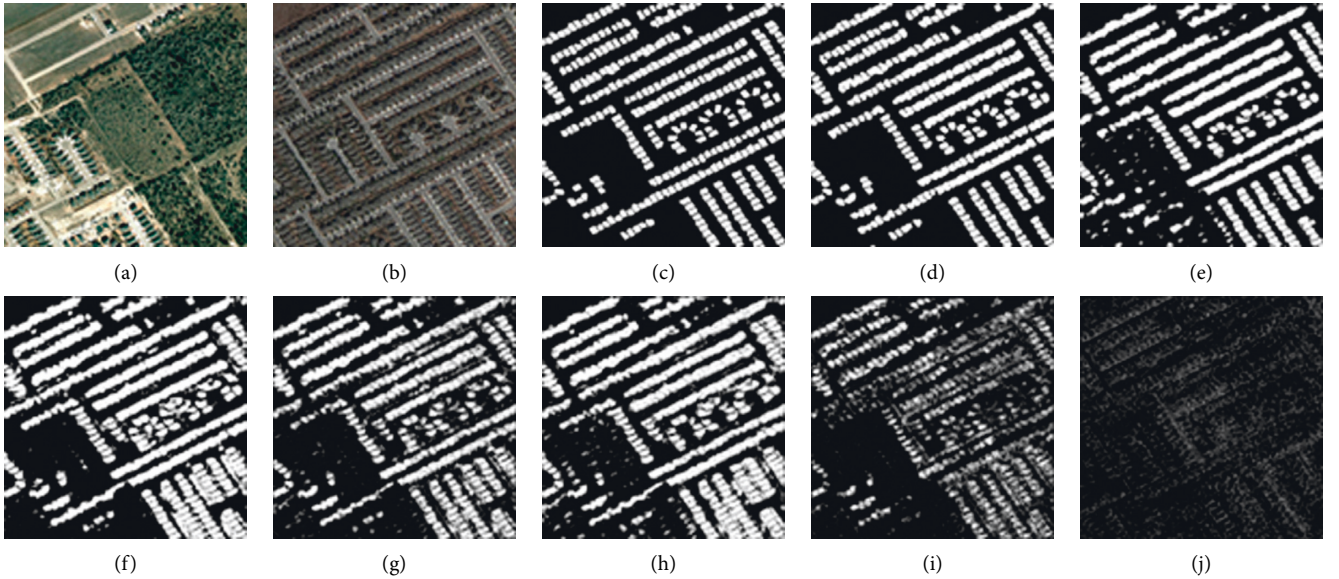

FIGURE 11: Qualitative comparison of different approaches on LEVID-CD dataset for dense-object change. (a) T1 image (#0045). (b) T2 image (#0045). (c) Reference change map. (d) Ours. (e) IFN. (f) DASNet. (g) CD-UNet++. (h) FC-Siam-Conc. (i) FC-Siam-Diff. (j) FC-EF.

5. Conclusions

In this paper, we proposed a CD method for multisource remote sensing images using a coupled dictionary and deep learning. To reveal the coupled relationship between multisource remote sensing images, a coupled dictionary module was proposed. This module maps multisource remote sensing images from low-dimensional to high-dimensional space to calculate the reconstruction coefficient. The changing image blocks will have different reconstruction coefficients, while unchanging image blocks will have similar reconstruction coefficients. To more accurately describe the degree of change in an object, we constructed a new image block for reconstruction loss in a differential feature discriminant network, which measures the change degree of each image block pair by considering true/false discrimination loss, classification loss, and cross-entropy loss. Compared with other CD methods, the proposed method can effectively alleviate the uncertainty of edge pixel CD and realize satisfactory results. In the future, we plan to focus on CD for large remote sensing images and extend the proposed method to practical applications.

The method that we proposed has its limitations. The location where has changed in multisource remote sensing images can be detected by the proposed method. But what has changed in multisource remote sensing images cannot be detected. To address this problem, hyperspectral images containing a lot of spectral information can be used to detect what has changed. In addition, cloud and fog occlusion in remote sensing image can also affect change detection. Therefore, it is necessary to preprocess remote sensing images before using the proposed method for change detection. These limitations indicate that although a highly efficient change detection method is proposed, it is still a challenging task to generalize the method to practical production, and this will be the direction of our future efforts.

Data Availability

The underlying data supporting the results of our study can be found in [36] and [37].

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Weiwei Yang, Haifeng Song, Lei Du, Songsong Dai, and Yingying Xu contributed equally to this work.

Acknowledgments

References

[1] J. A. Benediktsson, J. Chanussot, and W. M. Moon, "Very high-resolution remote sensing: challenges and opportunities [point of view]," *Proceedings of the IEEE*, vol. 100, no. 6, pp. 1907–1910, 2012.

[2] L. Khelifi and M. Mignotte, "Deep learning for change detection in remote sensing images: comprehensive review and meta-analysis," *IEEE Access*, vol. 8, pp. 126385–126400, 2020.

[3] M. Kamel, "Monitoring of land use and land cover change detection using multi-temporal remote sensing and time series analysis of qena-luxor governorates (QLGs), Egypt," *Journal of the Indian Society of Remote Sensing*, vol. 48, no. 12, pp. 1767–1785, 2020.

[4] K. Kundu, P. Halder, and J. K. Mandal, "Urban change detection analysis during 1978–2017 at Kolkata, India, using multi-temporal satellite data," *Journal of the Indian Society of Remote Sensing*, vol. 48, no. 11, pp. 1535–1554, 2020.

[5] S. Derakhshan, S. L. Cutter, and C. Wang, "Remote sensing derived indices for tracking urban land surface change in case of earthquake recovery," *Remote Sensing*, vol. 12, no. 5, Article ID 895, 2020.

[6] D. Cao, J.-z. Feng, L.-y. Bai et al., "Delineating the rice crop activities in Northeast China through regional parametric synthesis using satellite remote sensing time-series data from 2000 to 2015," *Journal of Integrative Agriculture*, vol. 20, no. 2, pp. 424–437, 2021.

[7] H. Song and W. Yang, "GSCCTL: a general semi-supervised scene classification method for remote sensing images based on clustering and transfer learning," *International Journal of Remote Sensing*, 2021.

[8] M. E. Bauer, "Remote sensing of environment: history, philosophy, approach and contributions, 1969–2019," *Remote Sensing of Environment*, vol. 237, Article ID 111522, 2020.

[9] A. Abdollahi, B. Pradhan, and A. Alamri, "RoadVecNet: a new approach for simultaneous road network segmentation and vectorization from aerial and google earth imagery in a complex urban set-up," *GIScience and Remote Sensing*, vol. 58, no. 7, pp. 1151–1174, 2021.

[10] A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty, and A. Alamri, "Multi-object segmentation in complex urban scenes from high-resolution remote sensing data," *Remote Sensing*, vol. 13, no. 18, Article ID 3710, 2021.

[11] X. Qi, K. Li, P. Liu, X. Zhou, and M. Sun, "Deep attention and multi-scale networks for accurate remote sensing image segmentation," *IEEE Access*, vol. 8, pp. 146627–146639, 2020.

[12] Y. Xi, W. Jia, J. Zheng et al., "DRL-GAN: dual-stream representation learning GAN for low-resolution image classification in UAV applications," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1705–1716, 2021.

[13] H. Emami, M. M. Aliabadi, M. Dong, and R. B. Chinnam, "SPA-GAN: spatial attention GAN for image-to-image translation," *IEEE Transactions on Multimedia*, vol. 23, pp. 391–401, 2021.

[14] Y. Dong, Y. Liu, H. Zhang, S. Chen, and Y. Qiao, "FD-GAN: generative adversarial networks with fusion-discriminator for single image dehazing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 10729–10736, New York, NY, USA, Febraury 2020.

[15] J. Rabbi, N. Ray, M. Schubert, S. Chowdhury, and D. Chao, "Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network," *Remote Sensing*, vol. 12, no. 9, 2020.

[16] S. Dai, "Quantum cryptanalysis on a multivariate cryptosystem based on clipped hopfield neural network," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–5, 2021.

[17] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017.

[18] H. Song, W. Yang, H. Yuan, and H. Bufford, "Deep 3d-multiscale densenet for hyperspectral image classification based on spatial-spectral information," *Intelligent Automation and Soft Computing*, vol. 26, no. 4, pp. 1441–1458, 2020.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, Munich, Germany, 2015.

[20] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[21] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[22] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.

[23] F. Samadi, G. Akbarizadeh, and H. Kaabi, "Change detection in SAR images using deep belief network: a new training approach based on morphological images," *IET Image Processing*, vol. 13, no. 12, pp. 2255–2264, 2019.

[24] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: a general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 3–13, 2019.

[25] M. Gong, Y. Yang, T. Zhan, X. Niu, and S. Li, "A generative discriminatory classified network for change detection in multispectral imagery," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 1, pp. 321–333, 2019.

[26] X. Niu, M. Gong, T. Zhan, and Y. Yang, "A conditional adversarial network for change detection in heterogeneous images," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 1, pp. 45–49, 2019.

[27] Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu, "Change detection based on deep siamese convolutional network for optical aerial images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1845–1849, 2017.

[28] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 3, pp. 545–559, 2018.

[29] S. Wang, Y. Wang, Y. Liu, and L. Li, "SAR image change detection based on sparse representation and a capsule network," *Remote Sensing Letters*, vol. 12, no. 9, pp. 890–899, 2021.

[30] M. Zhang, Y. Chen, S. Li, and X. Tian, "Sparse representation-based image fusion for multi-source NDVI change detection," in *Proceedings of the IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, pp. 692–695, Waikoloa, HI, USA, September 2020.

[31] L. Rosasco, S. Mosci, M. Santoro, A. Verri, and S. Villa, *Iterative Projection Methods for Structured Sparsity Regularization*, Massachusetts Institute of Technology, Cambridge, MA, USA, Mit-Csail-Tr-2009-050, 2009.

[32] A. Jentzen, B. Kuckuck, A. Neufeld, and P. von Wurstemberger, "Strong error analysis for stochastic gradient descent optimization algorithms," *IMA Journal of Numerical Analysis*, vol. 41, no. 1, 492 pages, 2021.

[33] Y. Liu, Y. Gao, and W. Yin, "An improved analysis of stochastic gradient descent with momentum," *Advances in Neural Information Processing Systems*, vol. 2020, Article ID 07989, 2020.

[34] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, pp. 2121–2159, 2011.

[35] D. P. Kingma and J. L. Ba, "Adam: a method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, CA, USA, May 2015.

[36] M. A. Lebedev, Y. V. Vizilter, O. V. Vygolov, V. A. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2, pp. 565–571, 2018.

[37] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sensing*, vol. 12, no. 10, Article ID 1662, 2020.

[38] K. Ye, J. Wang, H. Gao, L. Yang, and P. Xiao, "Optimization of lapping process parameters of CP-Ti based on PSO with mutation and BPNN," *International Journal of Advanced Manufacturing Technology*, vol. 117, no. 9-10, pp. 2859–2866, 2021.

[39] R. Caye Daudt, B. Le Saux, and A. Boulch, "Fully convolutional siamese networks for change detection," in *Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 4063–4067, Athens, Greece, October 2018.

[40] T. Lei, Y. Zhang, Z. Lv, S. Li, S. Liu, and A. K. Nandi, "Landslide inventory mapping from bitemporal images using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 6, pp. 982–986, 2019.

[41] P. F. Alcantarilla, S. Stent, G. Ros, R. Arroyo, and R. Gherardi, "Street-view change detection with deconvolutional networks," *Autonomous Robots*, vol. 42, no. 7, pp. 1301–1322, 2018.

[42] D. Peng, Y. Zhang, and H. Guan, "End-to-end change detection for high resolution satellite images using improved UNet++," *Remote Sensing*, vol. 11, no. 11, Article ID 1382, 2019.

[43] J. Liu, K. Chen, G. Xu et al., "Semi-supervised change detection based on graphs with generative adversarial networks," in *Proceedings of the IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 74–77, Yokohama, Japan, July 2019.

[44] L. Di, W. Liejun, C. Shuli, L. Yongming, and D. C. A. N. Anyu, "A combined attention network for remote sensing image change detection," *Information*, vol. 12, pp. 1–16, 2021.

[45] J. Chen, Z. Yuan, J. Peng et al., "DASNet: dual attentive fully convolutional siamese networks for change detection in high-resolution satellite images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1194–1206, 2021.

[46] C. Zhang, P. Yue, D. Tapete et al., "A deeply supervised image fusion network for change detection in high resolution bitemporal remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 183–200, 2020.

[47] X. Hou, Y. Bai, Y. Li, C. Shang, and Q. Shen, "High-resolution triplet network with dynamic multiscale feature for change detection on satellite images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 177, pp. 103–115, 2021.