

# Change Detection in Multisource VHR Images via Deep Siamese Convolutional Multiple-Layers Recurrent Neural Network

Hongruixuan Chen, Chen Wu<sup>✉</sup>, Member, IEEE, Bo Du<sup>✉</sup>, Senior Member, IEEE,  
Liangpei Zhang<sup>✉</sup>, Fellow, IEEE, and Le Wang<sup>✉</sup>

**Abstract**—With the rapid development of Earth observation technology, very-high-resolution (VHR) images from various satellite sensors are more available, which greatly enrich the data source of change detection (CD). Multisource multitemporal images can provide abundant information on observed landscapes with various physical and material views, and it is exigent to develop efficient techniques to utilize these multisource data for CD. In this article, we propose a novel and general deep siamese convolutional multiple-layers recurrent neural network (RNN) (SiamCRNN) for CD in multitemporal VHR images. Superior to most VHR image CD methods, SiamCRNN can be used for both homogeneous and heterogeneous images. Integrating the merits of both convolutional neural network (CNN) and RNN, SiamCRNN consists of three subnetworks: deep siamese convolutional neural network (DSCNN), multiple-layers RNN (MRNN), and fully connected (FC) layers. The DSCNN has a flexible structure for multisource image and is able to extract spatial-spectral features from homogeneous or heterogeneous VHR image patches. The MRNN stacked by long-short term memory (LSTM) units is responsible for mapping the spatial-spectral features extracted by DSCNN into a new latent feature space and mining the change information between them. In addition, FC, the last part of SiamCRNN, is adopted to predict change probability. The experimental results in two homogeneous data sets and one challenging heterogeneous VHR images data set demonstrate that the promising performances of the proposed network outperform several state-of-the-art approaches.

Manuscript received September 21, 2019; revised November 11, 2019; accepted November 24, 2019. Date of publication December 20, 2019; date of current version March 25, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61971317, 41801285, 61822113, and 41871243. (*Corresponding author: Chen Wu*)

H. Chen is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: qschr@whu.edu.cn).

C. Wu is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, School of Computer Science, Wuhan University, Wuhan 430072, China (e-mail: chen.wu@whu.edu.cn).

B. Du is with the School of Computer Science, Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430072, China (email: gunspace@163.com).

L. Zhang is with the Remote Sensing Group, State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: zlp62@whu.edu.cn).

L. Wang is with the Department of Geography, University at Buffalo, The State University of New York, Buffalo, NY 14261, USA (e-mail: lewang@buffalo.edu).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2019.2956756

**Index Terms**—Change detection (CD), deep siamese convolutional multiple-layers recurrent neural network, deep siamese convolutional neural network (DSCNN), heterogeneous images, long-short term memory (LSTM), multiple-layers recurrent neural network (MRNN), very-high-resolution (VHR) images.

## I. INTRODUCTION

REMOTE sensing images can provide a macro view of Earth observation and reflect types and attributes of objects within the observed area. Multitemporal images covering the same scene are able to reveal the dynamic changes in ground. Thus, change detection (CD) technology in multitemporal images has become more and more significant. CD is the process of identifying differences in the state of an object or phenomenon by observing it at different time [1]. Now, CD has played an important role in wide areas, such as land-use and land-cover (LULC) change analysis, ecosystem monitoring, urban expansion research, resource management, deforestation monitoring, and damage assessment [2]–[8].

As a common data source of CD, multispectral remote sensing image with medium- and low-resolution can provide a rough and macro view of the surface and a lot of CD methods were proposed based on it. Change vector analysis (CVA) [9] is a most commonly used method, which can provide change intensity and change direction. There are many extension methods based on CVA, such as parcel CVA (PCVA) [10] and robust CVA (RCVA) [11]. Gramm–Schmidt (GS) transformation [12] and principal component analysis (PCA) [13], as transformation methods, transform the images into a new feature space and select a part of new bands for CD. Based on canonical correlation analysis (CCA), Nielsen *et al.* [14] [15] proposed multivariate alteration detection (MAD) and iteratively reweighting multivariate alteration detection (IRMAD). Wu *et al.* [16] proposed a novel CD method based on slow feature analysis (SFA), which aims to find the most invariant component in multitemporal images to highlight changed regions. In addition to CD, SFA was also used in radiometric correction [17] and scene CD [18].

With the development of Earth observation technology, more and more new satellite sensors are designed to acquire very high resolution (VHR) images, such as SPOT, QuickBird, IKONOS, GaoFen (GF) *et al.* Compared with medium- and

low-resolution images, VHR images are able to provide abundant surface details and spatial distribution information. CD in multitemporal VHR images has crucial effects on the research of building CD, urban internal change analysis, and urban expansion research; therefore, it has attracted a lot of research interests [5], [18]–[21]. Furthermore, in addition to optical VHR images, multisource data from many different sensors can also be acquired, such as light detection and ranging (LiDAR) and SAR. These sensors of different sources can provide observation information with various physical mechanisms of the ground and can also solve the problem that the optical sensors are subjected to weather conditions, which greatly enrich the data source of CD. Nonetheless, due to the diverse physical mechanisms of multisource data, it is difficult to compare the acquired heterogeneous images, so that most of the existing research focuses on the CD in homogeneous images. With the gradual enrichment of multisource data, it is necessary to develop a general CD method, which can accurately detect the changes between homogeneous images and can also process the multisource data efficiently.

Recently, deep learning (DL) has been shown to be very successful in a variety of computer vision and remote sensing image interpretations [22], [23]. Compared with conventional methods, DL methods can utilize spatial-context information in VHR images more efficiently. So plenty of CD methods based on DL have been developed for multitemporal VHR images. In [24], a CD method for multitemporal SAR images are proposed by constructing deep neural network (DNN). In [25], a deep siamese convolutional network was designed for aerial image CD, which extract features by two weight-sharing convolutional branches and generate binary change map based on the feature difference of the last layer. Saha *et al.* [26] combined CVA with pretrained deep convolutional neural network (DCNN) and developed deep change vector analysis (DCVA) for VHR image CD. Lyu *et al.* [27] adopted recurrent neural network (RNN) to learn transferable change-rules between multitemporal homogeneous images. Going one step further, for learning unified features, Lyu *et al.* [27] utilized recurrent convolutional neural network (RCNN) to detect multiclass changes. By learning nonlinear features with DNN and highlighting changes via SFA, an unsupervised deep slow feature analysis (DSFA) model is designed by Du *et al.* [29] for CD. However, all of these above-mentioned methods can only be used in homogeneous images and is powerless for CD in heterogeneous images. In fact, even though DL has been applied to CD, there are currently only a few methods [30]–[32] that can be used for heterogeneous image CD not to mention a universal CD method that can perform well in both homogeneous and heterogeneous VHR images.

For the purpose of developing a general architecture for CD in both homogeneous and heterogeneous multitemporal VHR images, a deep siamese convolution multiple-layer RNN (MRNN) is designed by integrating the merits of both CNN and RNN. First, since CNN is skill at automatically extracting relevant features, several convolutional layers are stacked for designing a deep siamese convolutional

neural network (DSCNN), which can extract representative spatial-spectral features from multitemporal VHR images. Because the relationship between change intensity and feature difference of homogeneous images is usually non-linear, and features extracted from heterogeneous images are located in diverse feature spaces, MRNN are constructed by long-short term memory (LSTM) units to dig change information between homogeneous or heterogeneous spatial-spectral features. Distinguished from the RCNN proposed by Mou *et al.* [28], the DSCNN part of SiamCRNN has a flexible deep siamese structure for multisource data, and the MRNN in SiamCRNN aims to fully mine the change information between heterogeneous images. Most importantly, in [28], RCNN is developed only to detect changes in homogeneous low- and medium-resolution multispectral images, whereas SiamCRNN is specifically designed to process multisource VHR images.

The main contributions of this article are concluded as follows.

- 1) This article presents a powerful and general end-to-end network architecture, SiamCRNN, for CD in homogeneous and heterogeneous VHR image, which combine CNN and RNN to process images.
- 2) The DSCNN can extract spatial-spectral features from both homogeneous and heterogeneous VHR images, and its structure is flexible. For homogeneous images, the DSCNN is a pure-siamese network structure and has two weight-shared branches. This structure can push out change information. For heterogeneous images, the DSCNN is modified as a pseudo-siamese network structure and its two branches are designed for two heterogeneous VHR images, respectively, which can make the extracted spatial-spectral features more representative.
- 3) By stacking LSTM units, MRNN is designed to process the spatial-spectral features extracted by DSCNN, which can map the two homogeneous or heterogeneous spatial-spectral features into a new feature space, fully excavate the change information and generate a spatial-spectral-temporal feature with abundant change information.

The rest of the article is organized as follows. Section II elaborates the proposed SiamCRNN. Section III provides homogeneous data sets information, experiment setting, experimental results and discussion. In Section IV, the experiment of CD in heterogeneous VHR images is carried out. In the end, Section V draws the conclusion of our work in this article.

## II. METHODOLOGY

The overview of CD architecture based on the proposed SiamCRNN is shown in Fig. 1. First, multitemporal VHR images are preprocessed. Then multitemporal spatial-spectral high-dimensional features are extracted from VHR image patches by DSCNN. Next, the spatial-spectral features from two homogeneous or heterogeneous VHR images are mapped

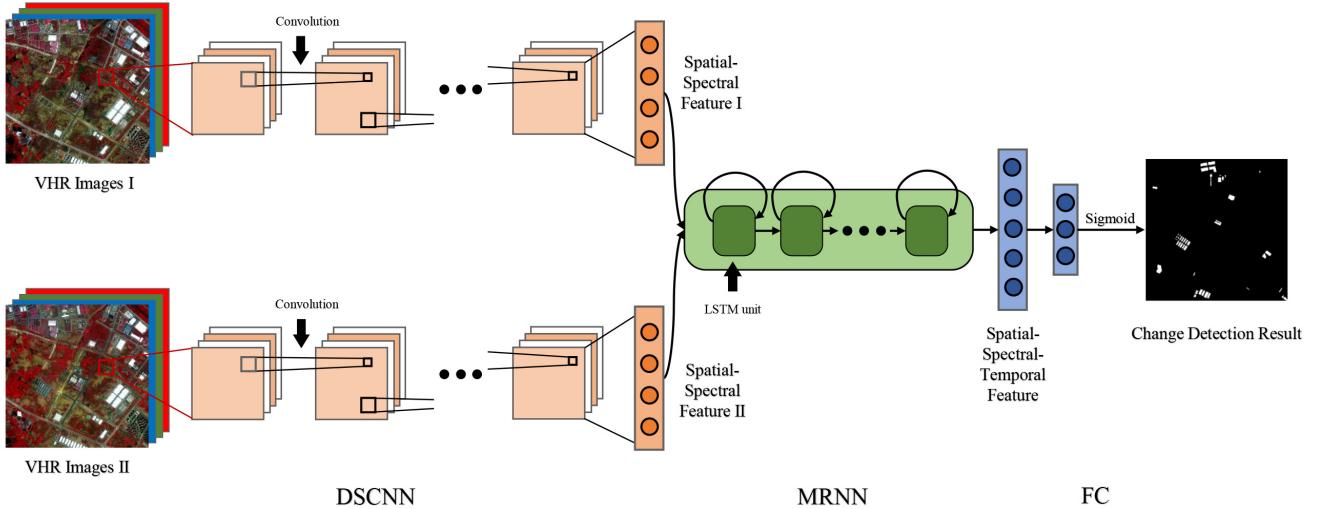


Fig. 1. Overview of the CD architecture based on the proposed SiamCRNN. SiamCRNN consists of three subnetworks: DSCNN, MRNN, and FC.

into same feature space and mined change information by MRNN. Finally, the CD results of multitemporal VHR images are predicted by FC. In this section, the proposed architecture is elaborated from four aspects: data preprocessing, DSCNN for spatial–spectral features extraction, MRNN for feature mapping and information mining, and change probability predicted by FC.

#### A. Data Preprocessing

The first step of our CD architecture is data preprocessing consisting of coregistration and radiometric correction. The specific processing procedure for homogeneous and heterogeneous VHR images is slightly different.

Image registration is the process of aligning two or more images of the same scene obtained at different times [33]. Given two multitemporal images, only when both images are aligned geometrically, can the CD between the corresponding areas be meaningful. In our CD architecture, there are four main steps in image registration: collecting matched point-pairs, establishing transform model, transforming images, and resampling. owing to the spatial resolutions of multitemporal heterogeneous VHR images are generally different, the resampling step is adopted to resample the relative higher resolution image to the relative lower resolution of reference image. Since the spatial resolution of multitemporal homogeneous VHR images is the same, there is no need to implement resampling. Through the above-mentioned four steps, the two given multitemporal VHR images are geometrically aligned.

Radiometric correction is adopted to eliminate the radiometric difference between multitemporal VHR images caused by different sun angle, light intensity, and atmospheric conditions. The specific method is radiometric relative normalization, which can normalize images with zero mean and unit variance. Given two multitemporal VHR spectral vectors, denoted by  $x^i = [x_1^i, x_2^i, \dots, x_N^i]$  and  $y^i = [y_1^i, y_2^i, \dots, y_M^i]$ , where  $i$  and  $j$  indicate the pixel number,  $N$  and  $M$  indicate the

dimension of the spectral band. The radiometric relative normalization can be formulated as follows:

$$\begin{cases} \hat{x}_n^i = \frac{(x_n^i - \mu_{x_n})}{\sigma_{x_n}} \\ \hat{y}_m^j = \frac{(y_m^j - \mu_{y_m})}{\sigma_{y_m}} \end{cases} \quad (1)$$

where  $\mu_{x_n}$  is the mean and  $\sigma_{x_n}$  is the standard deviation for band  $n$  of image  $X$ ,  $\mu_{y_m}$  is the mean and  $\sigma_{y_m}$  is the standard deviation for band  $m$  of image  $Y$ .

Nevertheless, heterogeneous VHR images are captured by different types of sensors with different characteristics. Thus, radiometric correction actually cannot eliminate the radiometric difference between heterogeneous VHR images. But in the proposed architecture, heterogeneous VHR images are still preprocessed by radiometric relative normalization. Because radiometric relative normalization is able to assign the value of various dimensions of the input heterogeneous VHR images to be in a similar range, which can facilitate the training of DNN. After radiometric correction, the radiometric difference between multitemporal VHR images caused by different conditions would be eliminated.

#### B. DSCNN for Spatial–Spectral Features Extraction

The VHR images could provide abundant ground details, texture information, and spatial distribution information. The convolutional layer of CNN is a suitable structure for simultaneously extracting spatial context features and spectral features, namely spatial–spectral features and coping with internal complexity caused by the extremely high resolution of VHR images. Considering an input image  $I$ , the convolutional layer of CNN could extract spatial–spectral features from  $I$  via several convolution kernels and get feature maps, i.e.,  $F = C(I)$

$$F^{(i)} = C^{(i)}(I) = a^{(i)}(W^{(i)} * I + b^{(i)}) \quad (2)$$

where  $F^{(i)}$  is the  $i$ th feature map obtained by the  $i$ th convolution kernel  $W^i$  and bias  $b^i$ ,  $*$  denotes convolution operation,

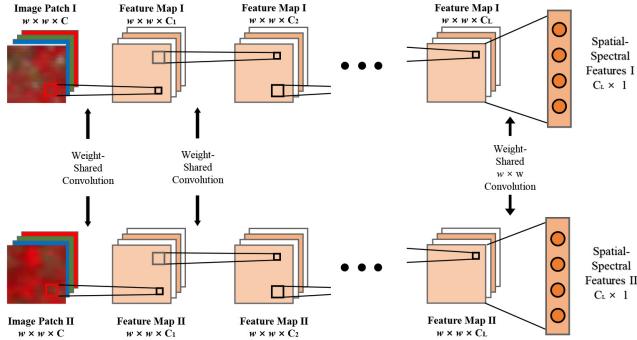


Fig. 2. Illustration of the DSCNN for homogeneous VHR images. The DSCNN for homogeneous VHR images is a pure siamese network, which means the convolution kernel size, kernel parameters, and the number of convolutional layers of two branches are all the same.

and  $a^i$  is the activation function of the  $i$ th convolution operation, which can introduce nonlinearity into network and enhance the fitting ability of network. Generally, the activation function in the one layer are all the same. In DCNN, the values of convolution kernel and bias would be learned by backpropagation (BP) algorithms [34].

Though single convolutional layer could extract spectral features and spatial context features in some extent, it still has two obvious drawbacks in processing VHR images. First, the single convolutional layer has a limited receptive field and could only extract single scale features. But in VHR images, except small and normal scale features, there also exists a few large-scale continuous features. Besides, the number of parameters of single convolutional layer is relatively less, which means it cannot extract meaningful and informative high dimensional features from VHR images. Therefore, in order to extract features from VHR images better, convolutional layers are stacked for developing a DCNN to extract high dimensional global features. Because CD involves two images, a novel DSCNN, consisting of two branches, is designed to extract spatial-spectral features from multitemporal VHR images. Given two multitemporal VHR images  $I^{t_1}$  and  $I^{t_2}$ , the outputs of DSCNN is expressed as follows:

$$\begin{cases} f_{ss}^{t_1} = S_1(I^{t_1}) = C_1^{L_1}(C_1^{(L_1-1)} \dots (C_1^1(I^{t_1})) \dots) \\ f_{ss}^{t_2} = S_2(I^{t_2}) = C_2^{L_2}(C_2^{(L_2-1)} \dots (C_2^1(I^{t_2})) \dots) \end{cases} \quad (3)$$

where  $f_{ss}^{t_1}$  and  $f_{ss}^{t_2}$  are the outputs of DSCNN,  $S_1$  and  $S_2$  are two branches of DSCNN,  $C_i^i$  is the  $i$ th convolutional layer of  $S_1$ ,  $L_1$  and  $L_2$  are numbers of the convolutional layers in two branches, respectively.

Fig. 2 shows the designed structure of DSCNN for CD in homogeneous VHR images. The DSCNN for homogeneous VHR images is a pure siamese network, namely  $S_1 \equiv S_2$ . So the two branches of DSCNN extract features from multitemporal image patches via the exactly same way because of weight-sharing, which can highlight the change information. Through several weight-sharing convolutional layers, two high-dimension spatial-spectral feature maps with size of  $w \times w \times C_L$  are extracted. Then the high dimensional feature

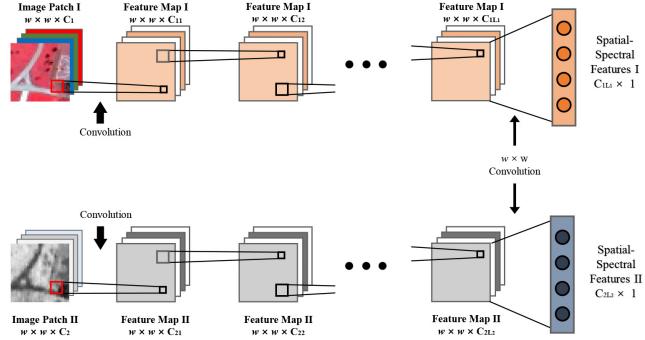


Fig. 3. Illustration of the DSCNN of SiamCRNN for heterogeneous VHR images. The two branches of DSCNN are designed for two heterogeneous VHR images, respectively, which can make DSCNN extract more representative spatial-spectral features.

maps are converted to the spatial-spectral vectors with size of  $C_L \times 1$  by the last layer with  $w \times w$  convolution kernel.

Because spatial resolution, range of digital number, statistical characteristics of heterogeneous VHR images could all be diverse, the structure of DSCNN for different heterogeneous VHR images-pairs would be different. Fig. 3 shows the illustration of DSCNN for CD in heterogeneous VHR images. For heterogeneous VHR images of different types, such as multitemporal optical-SAR VHR images, the DSCNN would be a pseudo-siamese network and the two branches of DSCNN would be designed with different convolution kernels and convolutional layer numbers, namely  $S_1 \not\equiv S_2$ . By designing the branches of DSCNN for different VHR images, the DSCNN would be able to extract more representative spatial-spectral features. Because of extracting features from multitemporal image patches, not the same with the conventional CNN, the DSCNN does not contain max-pooling layer or average-pooling layer to reduce the data size.

### C. MRNN for Feature Mapping and Information Mining

After the multitemporal high-dimensional features are extracted, the common operation in plenty of researches is calculating the difference feature to highlight change information, such as DSCN and DSFA. This common operation intuitively assumes that there is a linear relationship between change intensity and feature difference. Nevertheless, the relationship between change intensity and difference of features extracting from multitemporal images is usually nonlinear. Thus, there exist some pseudochanges which would be falsely detected. More importantly, the high-dimensional features extracted from heterogeneous images are located in diverse feature spaces, so calculating difference cannot highlight change information between multitemporal heterogeneous images and is completely not suitable for heterogeneous images CD. This is one of the reasons why DSCN, DSFA, and other similar methods can only be used for CD in homogenous VHR images.

Therefore, for the purpose of making the proposed method applicable for heterogeneous VHR images and mining change information from spatial-spectral high dimensional features better, RNNs are adopted to process the spatial-spectral

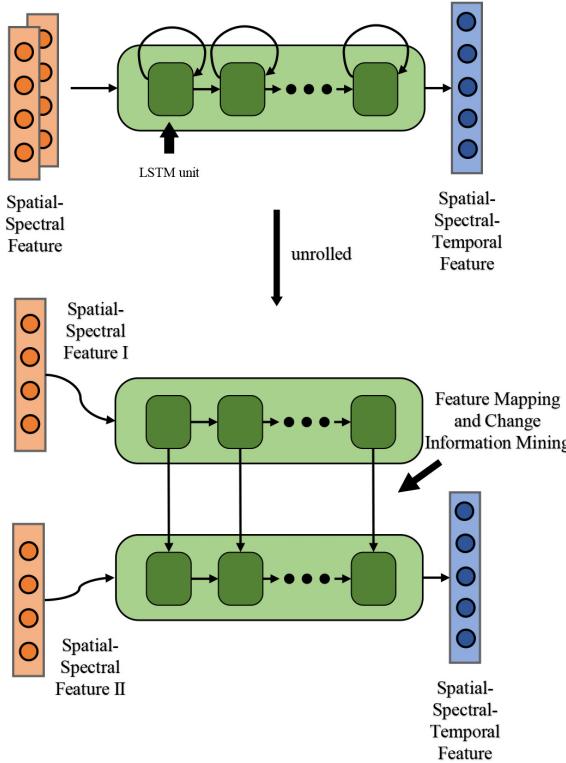


Fig. 4. Illustration of the multiple RNN of SiamCRNN. The multitemporal homogeneous or heterogeneous spatial-spectral features obtained by DSCNN is mapped into a new latent feature space and mined change information by multiple RNN layers.

features and extract change information. As a connectionist model, RNN is capable of capturing the dynamics of sequences via cycles in the network of nodes, which is suitable for digging change information between multitemporal images. Sometimes a single-layer RNN may not be able to extract change information well, especially for heterogeneous images. So single-layer RNNs are stacked and the MRNN is designed. As shown in Fig. 4, the spatial-spectral high-dimensional features extracted by the DSCNN are mapped into a new latent feature space and the nonlinear relationship between multitemporal features is modeled. Through modeled by multiple hidden layers, the multitemporal spatial-spectral features are fully mined and the spatial-spectral-temporal feature with abundant change information is generated.

What is more, the selection of RNN architecture has significant influence on the final performance of CD, since different architecture has diverse mining ability. In our method, the LSTM model [35] is chosen to construct the MRNN. The detailed graphic model of LSTM is shown in Fig. 5. Different from traditional RNN, LSTM has a cell state and several gates, which allows LSTM perform better than the traditional RNN on more challenging issues involving sequential time series, such as CD. The core of LSTM is the cell state  $c_t$ , and the information in  $c_t$  is controlled by three gates: forget gates, input gates, and output gates. Gate is a structure that selectively passes information, consisting of a sigmoidal function  $\sigma$  and

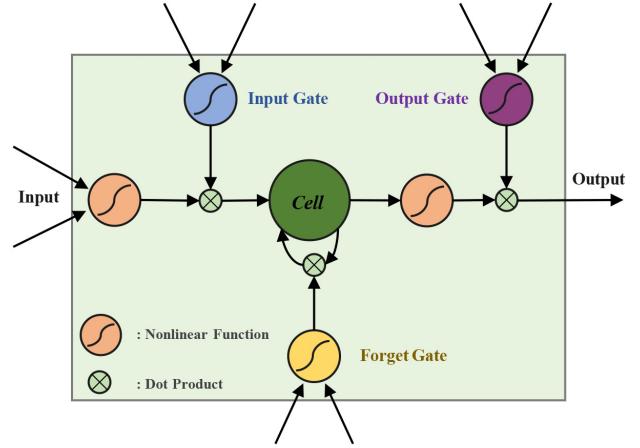


Fig. 5. Detailed graphic model of LSTM hidden unit in the proposed MRNN.

a dot product operation  $\odot$ . The definition of cell state, three gates, and the final output of LSTM are described as follows.

1) *Forget Gate*: The first step of LSTM is to decide what information would be discarded from the cell state. This decision is made through the forget gate  $f_t$ . The forget gate  $f_t$  is expressed as

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{(t-1)} + b_f). \quad (4)$$

2) *Input Gate*: Then input gate  $i_t$  decided what new information would be encoded into the cell state, which consists of two parts

$$\begin{cases} i_t = \sigma(W_{ix}x_t + W_{ih}h_{(t-1)} + b_i) \\ \tilde{c}_t = \tanh(W_{cx}x_t + W_{ch}h_{(t-1)} + b_c) \end{cases} \quad (5)$$

where  $i_t$  determines which values would be updated and a fully connected (FC) layer with  $\tanh$  activation function creates a new candidate, which would be encoded into the cell state.

3) *Cell State*: After calculating forget gate and input gate, the information stored in cell state could be updated via input gate and forget gate

$$c_t = f_t \odot c_{(t-1)} + i_t \odot \tilde{c}_t. \quad (6)$$

4) *Output Gate*: Finally, output gate  $o_t$  determines the output values of LSTM unit, which also consists of two parts

$$\begin{cases} o_t = \sigma(W_{ox}x_t + W_{oh}h_{(t-1)} + b_o) \\ h_t = o_t \odot \tanh(c_t) \end{cases} \quad (7)$$

where  $o_t$  determines which values would be output and an FC layer with  $\tanh$  activation function creates the final output  $h_t$ , or say spatial-spectral-temporal features.

As seen above, in CD task, the cell state would contain rich change information between multitemporal images, and the general way of feature mapping and the universal change-rules would be learned by the three doors.

#### D. Change Probability Predicted by FC

After inputting spatial-spectral features obtained by DSCNN, the change information between multitemporal VHR

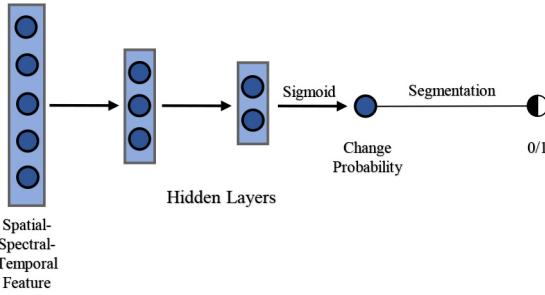


Fig. 6. Illustration of change probability of the center pixel in multitemporal VHR image patches predicted by FC layers. After change probability is predicted, threshold is set up to 0.5, and change probability is segmented to get binary value.

image patches are mined and spatial-spectral-temporal feature with abundant change information is obtained. Then the spatial-spectral-temporal feature generated by MRNN is entered into the last part of SiamCRNN: FC layers. As illustrated in Fig. 6, after mapped by several hidden layers, the spatial-spectral-temporal features are transformed into a real number, which represents the change probability of the center pixel in multitemporal VHR images patches. The above process can be expressed as the following formula:

$$p_i = FC(f_{sst}^i) \\ = a^L(W^L a^{(L-1)}(\dots a^1(W^1 f_{sst}^i + b^1)\dots) + b^L) \quad (8)$$

where  $p_i$  is the  $i$ th pixel change probability,  $f_{sst}^i$  is spatial-spectral-temporal features of the  $i$ th pixel,  $L$  is layer numbers of FC,  $W^l$  and  $b^l$  are parameters of the  $l$ th layer, and  $a^l$  is the activation function of the  $l$ th layer. In our SiamCRNN, the activation function of the last layer of FC is sigmoid. Owing to sigmoid function, the threshold segmentation step in the proposed method could be simplified and just adopt 0.5 as the threshold to get binary value.

### E. Optimization

The forward propagation of the entire SiamCRNN is described as the previous sections. Given two corresponding multitemporal VHR image patches, the final output of SiamCRNN is the change probability of the center pixel. By selecting appropriate training patch samples and defining loss function, the parameters of SiamCRNN can be updated via the BP algorithm. Suppose that we have a loss  $G$  that depends on the output of SiamCRNN  $p$  and the ground truth  $y$  via a loss function  $g$

$$G = g(p, y). \quad (9)$$

The ultimate goal of the BP algorithm is to use gradient descent to minimize the loss  $G$ .

Since FC is the last part of SiamCRNN, the parameters of FC are updated first in the BP algorithm. According to the chain rule, the loss gradient of the parameters in the last layer of FC can be expressed as the following formula:

$$\begin{cases} \frac{\partial G}{\partial W^{L_1}} = \frac{\partial G}{\partial a^{L_1}} \frac{\partial a^{L_1}}{\partial z^{L_1}} \frac{\partial z^{L_1}}{\partial W^{L_1}} \\ \frac{\partial G}{\partial b^{L_1}} = \frac{\partial G}{\partial a^{L_1}} \frac{\partial a^{L_1}}{\partial z^{L_1}} \frac{\partial z^{L_1}}{\partial b^{L_1}}. \end{cases} \quad (10)$$

where  $L_1$  is the layer numbers of FC,  $W^{L_1}$  and  $b^{L_1}$  are the weight and bias of the last layer,  $a^{L_1}$  is the output of the last layer, and  $z^{L_1} = W^{L_1}a^{(L_1-1)} + b^{L_1}$ . Based on the gradient descent method,  $W^{L_1}$  and  $b^{L_1}$  can be updated as

$$\begin{cases} W^{L_1} \leftarrow W^{L_1} - \alpha \frac{\partial G}{\partial W^{L_1}} \\ b^{L_1} \leftarrow b^{L_1} - \alpha \frac{\partial G}{\partial b^{L_1}} \end{cases} \quad (11)$$

where  $\alpha$  is the learning rate. Since the gradient loss of the remaining parameters can be derived according to the chain rule, for the sake of simplicity, we only elaborate on the gradient loss of the parameters in the last layer of FC, MRNN, and DSCNN.

After FC, the gradient loss propagates to the last layer of MRNN. The loss gradient of the parameters in the last layer of MRNN can be computed as

$$\frac{\partial G}{\partial W^{L_2}} = \delta_1 \left( \frac{\partial h_{(2)}^{L_2}}{\partial W^{L_2}} + \frac{\partial h_{(2)}^{L_2}}{\partial h_{(1)}^{L_2}} \frac{\partial h_{(1)}^{L_2}}{\partial W^{L_2}} \right) \quad (12)$$

where  $\delta_1$  is the gradient loss propagated from FC,  $L_2$  is the layer numbers of MRNN,  $W^{L_2}$  is the scalar parameters in the last layer,  $h_{(1)}^{L_2}$  and  $h_{(2)}^{L_2}$  are the hidden state at the first and second moments, respectively.

Similar to the FC, the gradient loss of the parameters in the last layer of one branch of DSCNN is

$$\begin{cases} \frac{\partial G}{\partial W^{L_3}} = \delta_2 \frac{\partial G}{\partial a^{L_3}} \frac{\partial a^{L_3}}{\partial z^{L_3}} \frac{\partial z^{L_3}}{\partial W^{L_3}} \\ \frac{\partial G}{\partial b^{L_3}} = \delta_2 \frac{\partial G}{\partial a^{L_3}} \frac{\partial a^{L_3}}{\partial z^{L_3}} \frac{\partial z^{L_3}}{\partial b^{L_3}} \end{cases} \quad (13)$$

where  $\delta_2$  is the gradient loss propagated from MRNN,  $L_3$  is the layer numbers of the branch,  $W^{L_3}$  and  $b^{L_3}$  are the weight and bias of the last layer,  $a^{L_3}$  is the output of the last layer, and  $z^{L_3} = W^{L_3}a^{(L_3-1)} + b^{L_3}$ .

After the BP is completed, the parameters of SiamCRNN are updated based on the gradient descent method, and then forward propagation is executed again. By repeating the progress of the forward propagation and BP, the parameters in SiamCRNN will be gradually updated until the training step is completed.

## III. EXPERIMENT OF CD IN HOMOGENEOUS VHR IMAGES

### A. Data Description

In the experiment of CD in homogeneous VHR images, two VHR image data sets are adopted to evaluate the proposed method. The first homogeneous VHR data set was acquired by GF-2 sensor on April 4, 2016 and September 1, 2016, covering the city of Wuhan, China, denoted as WH. The size of two multitemporal images is  $1000 \times 1000$  with red band, green band, blue band, and near infrared band. The spatial resolution of images is 4 m/pixel. The pseudocolor VHR images and ground truth of changes and nonchanges are shown in Fig. 7. In ground truth, changed area (red) contains 20 026 pixels, unchanged area (green) contains 484 143 pixels, and the remaining pixels are undefined. In WH data

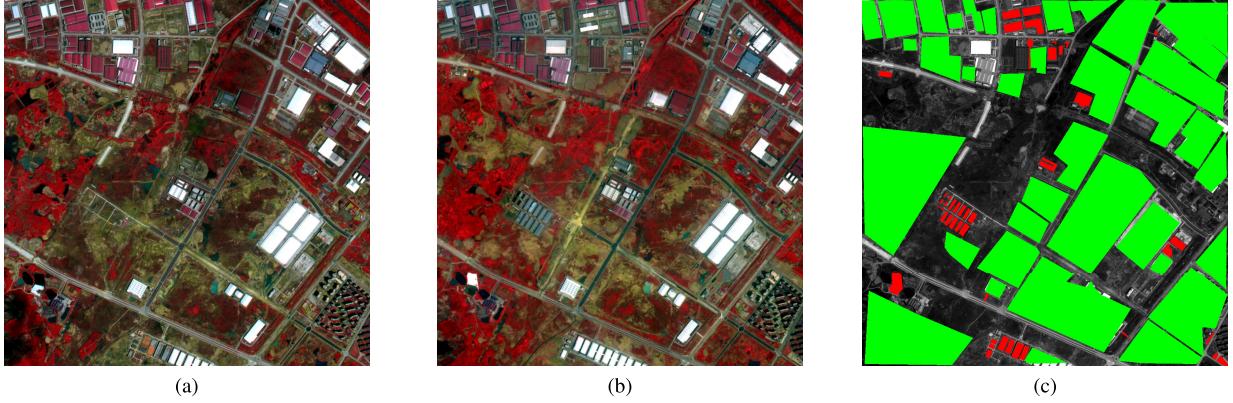


Fig. 7. Pseudocolor VHR images of (a) WH-1. (b) WH-2. (c) Ground truth, where red indicates changed region and green indicates unchanged region.

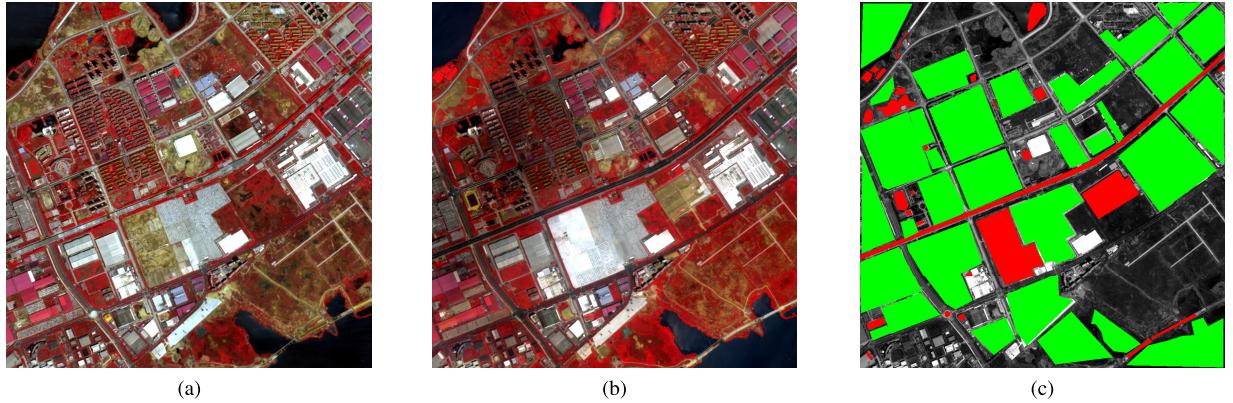


Fig. 8. Pseudo-color VHR images of (a) HY-1, and (b) HY-2. (c) Ground truth, where red indicates changed region and green indicates unchanged region.

set, the main changes between two images are groundwork before building over, new built-up regions and water change.

The second VHR data set was also acquired by GF-2 with a spatial resolution of 4 m/pixel, denoted as HY. The multitemporal images cover the Hanyang area of Wuhan city. The size of two images is  $1000 \times 1000$  with four bands. The pseudocolor VHR images and ground truth of changes and nonchanges are shown in Fig. 8. In ground truth, changed area (red) contains 59051 pixels, and unchanged area contains 416404 pixels. Owing to the rapid development of Hanyang, the study area shows obvious land-cover changes. The main differences between multitemporal images are construction of factory and railway, change of vegetation and water, and groundwork before building over. Compared with WH data set, the ground situation of WH data set is more complicated.

In both data sets, the change areas occupy only a few parts, which means there exists a heavy skewed-class problem between changed and unchanged classes. This situation brings greater challenge to CD. Besides, a small part of buildings in two data sets suffer from “overexposed” problem, which breaks the linear relationship of the digital numbers of unchanged regions between multitemporal images and cannot be eliminated by radiometric normalization. Therefore, “overexposed” problem makes accurate CD more difficult.

### B. Automatic Predetection

In order to train deep networks with high performance, it is often necessary to select a large number of representative annotated samples. However, it is undeniable that the manual selection of annotated samples is costly. In [29], [36], and [37], automatic predetection methods for homogeneous images based on difference method and threshold segmentation were proposed, which can partition pixels of images into three clusters: high probability to be changed, high probability to be unchanged, and need to be classified. Therefore, for the purpose of avoiding to select training samples manually and training the proposed SiamCRNN, the predetection method is implemented to select training samples automatically. Among the three clusters, the pixels with high probability of change and nonchange would be chosen as training samples. After training is completed, all the pixels are detected by the network.

### C. Experimental Setting

In the experiment involving homogeneous VHR images, the layer numbers of DSCNN is six and the size of convolution kernel in the former five layers is  $3 \times 3$ . The number of convolution kernel in each layer are 16, 16, 32, 32, 64, and 64. And all layers adopt rectified linear unit (ReLU) as the activation function. For the MRNN, the LSTM unit numbers of hidden layers are 128 and 64. For the FC, the neural numbers

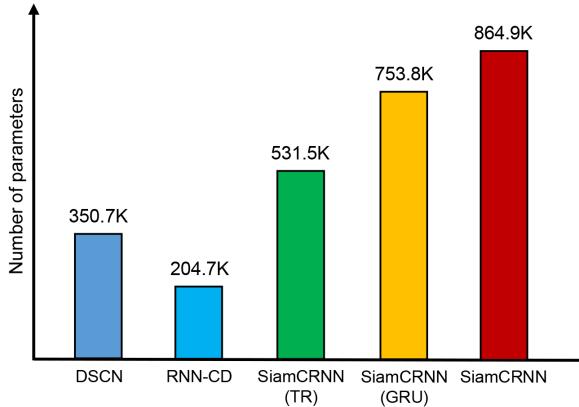


Fig. 9. Comparison of DSCN, RNN-CD, and SiamCRNN and its two variants in terms of model size.

of three layers are 64, 32, and 1. The parameters of DSCNN are initialized by he-normal way [38] and the others belonging to MRNN and FC are randomly initialized. The loss function of SiamCRNN apply weighted binary cross-entropy function for overcoming skew-class problem, and its definition is

$$L = w_p \hat{y} \log y + (1 - \hat{y}) \log(1 - y) \quad (14)$$

where  $w_p$  is the reciprocal of frequency of changed training samples. By means of giving a larger weight for changed class, the changed samples can play a more important role at the training phase. Adam optimizer is adopted to optimize loss function [39]. The parameters of Adam are chosen as recommended in [39]. Learning rate is set to 2e-4. As discussed in [24] and [36], the generally best choice of VHR image patch size is  $5 \times 5$ , in our experiments,  $5 \times 5$  is adopted as image patch size.

To validate the effectiveness of the proposed SiamCRNN, nine of the most widely used CD methods are used for comparison, namely CVA [9], MAD [14], IRMAD [15], SFA [16], iterative slow feature analysis (ISFA) [16], PCA-K-means [13], support vector machine (SVM), DSCN [25], and RNN-CD [27]. IRMAD and ISFA are iterative version of MAD and SFA. DSCN is a deep siamese convolutional network and performs well in aerial images. RNN-CD is an RNN-based CD network proposed in [27], which has shown a good performance in CD. What is more, to compare the fitting ability of different RNN architectures, two variants of the proposed method are developed by replacing LSTM with traditional RNN and gated recurrent unit (GRU), denoted as SiamCRNN(TR) and SiamCRNN(GRU). Among them, SVM, DSCN, and RNN-CD are supervised methods, we train them on the same samples as SiamCRNN. The kernel function of SVM adopts radial basis function (RBF). The optimal hyper-parameters  $C$  and  $\gamma$  are chosen by grid search. In our experiments, the hyperparameters of all unsupervised methods use the optimal values recommended in original references. K-means, Otsu, and fuzzy C-means (FCM) are adopted as segmentation methods, the best results of comparison methods are chosen as final results.

Fig. 9 shows the number of total trainable parameters in five deep learning methods, namely, DSCN, RNN-CD, and

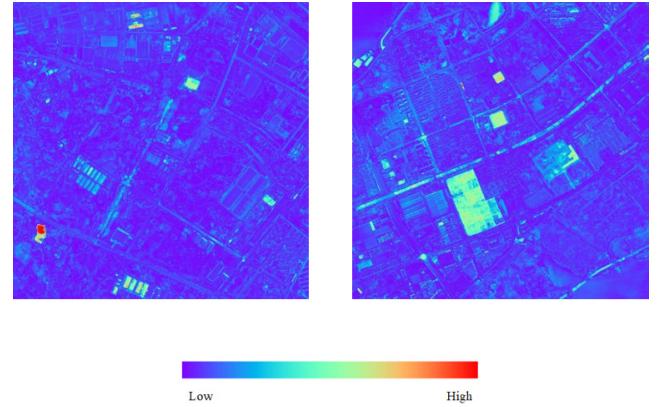


Fig. 10. Change intensity maps of WH and HY data sets obtained by predetection. The regions of warmer tones have larger probability of change.

SiamCRNN and its two variants. The number of parameters in DSCN and RNN-CD is in relatively small quantity because their architectures are relatively simple, which also means that their ability to extract features and mining change information is relatively weak. By contrast, the model size of SiamCRNN and its two variants are larger, since they consist of two different kinds of networks. Compared to the traditional RNN, both LSTM and GRU keep the original content and add the new components on top of it. Thus, the number of parameters in SiamCRNN and SiamCRNN(GRU) are more than SiamCRNN(TR). Besides, GRU utilizes only two gates to learn the general way of feature mapping and the universal change-rules, whereas LSTM utilizes three gates to do them. Therefore, the number of parameters in SiamCRNN is increased by about 15% compared to that in SiamCRNN(GRU). The CD problem could be regarded as a binary classification problem for change and nonchange. To evaluate the performance of the SiamCRNN and comparison methods, recall rate, precision rate, F1 score, overall accuracy (OA), and kappa coefficient (KC) are utilized. The following part discusses the experimental results in two homogeneous VHR images.

#### D. Experimental Results and Analysis

As introduced in Section II-A, the WH data set and HY data set are first preprocessed. Then predetection is implemented to compute change intensity maps (CIMs) and segment CIM to acquire training samples. As shown in Fig. 10, the areas of warmer tone in change intensity map have the greater probability of change. Based on CIM, the threshold segmentation is carried out and training samples are generated.

The binary change maps obtained by SiamCRNN and other comparison methods are shown in Fig. 11. As we can see, the results of MAD and SFA are unsatisfactory, numerous unchanged pixels are falsely detected as change. By means of iteratively reweighting, the binary change maps obtained by IRMAD and ISFA are cleaner in visual. However, it is obvious that some small changed areas are omitted. Besides, a part of the buildings is misclassified into change class, which implies that IRMAD and ISFA are confused

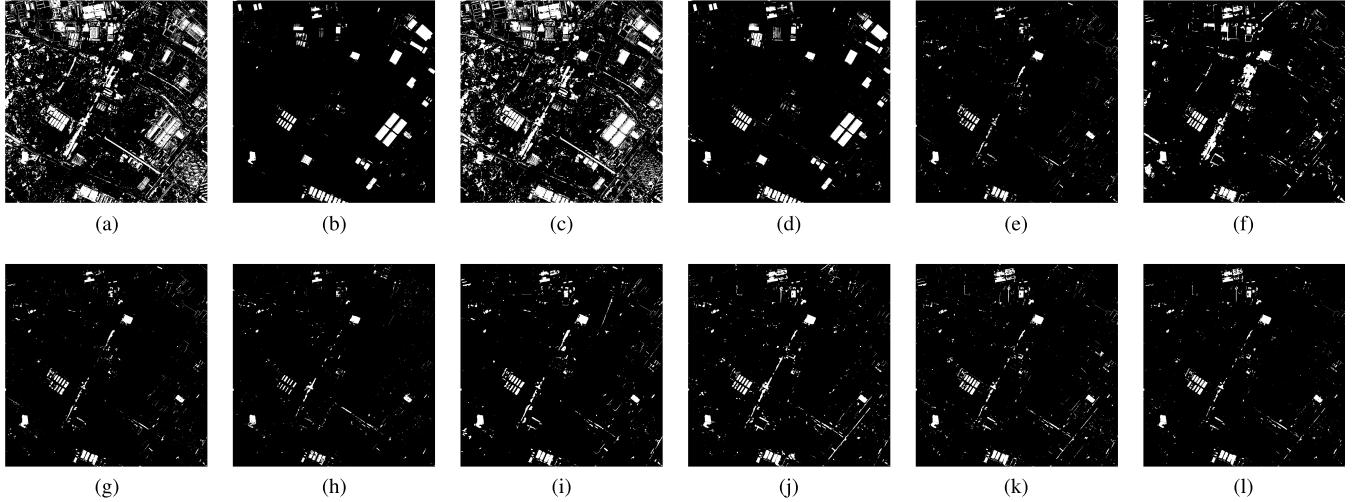


Fig. 11. Binary change maps obtained by the proposed method and comparison methods on WH data set. (a) MAD, (b) IRMAD, (c) SFA, (d) ISFA, (e) CVA, (f) PCA-K-means, (g) SVM, (h) DSCN, (i) RNN-CD, (j) SiamCRNN(TR) (k) SiamCRNN(GRU), and (l) SiamCRNN(LSTM).

by “overexposed” problem. By computing difference image and performing clustering, CVA acquires a relatively good result. But some small objects, such as road, are not detected correctly and a lot of margins of building are falsely detected as change. PCA-K-means utilizes spatial context information by transforming blocks, thus a majority of small objects are detected correctly. Nevertheless, some margins of building are still falsely detected. Based on training samples, a hyperplane is constructed by SVM to separate change and nonchange. But limited to relatively weak fitting ability of SVM, a part of groundwork before the building is over is not recognized. Through utilizing convolution operation and segmenting high-dimensional features, in the results obtained by DSCN, margins of building are detected correctly, and DSCN ignores the effects of “over-exposed” problem. However, many changes of build-up regions are not recognized. RNN-CD adopts RNN architecture to learn the change-rules and mine change information between multitemporal images. The binary change map obtained by RNN-CD is better than all the methods mentioned above. Nonetheless, without convolution layer, RNN-CD cannot utilize spatial context information effectively, which leads to omission of small object, internal fragmentation, and misdetection of margins.

Through extracted spatial-spectral features by DSCNN and mining changed information by MRNN, the proposed SiamCRNN and its two variants acquire the best binary change maps. Same with DSCN, the SiamCRNN and its two variants are completely immune to “overexposed” problem. Among the results obtained by them, the binary change map generated by SiamCRNN with LSTM achieves the best qualitative results with less omission and residual error.

The accuracy assessments of CD results on WH data set based on five evaluation criteria as described in Section IV-C is reported in Table I. As we observed in Fig. 11, IRMAD and ISFA only has a low F1 of 0.3651 and Kappa of 0.3289. ISFA is slightly better than IRMAD with F1 of 0.3888 and KC of 0.3530. Both IRMAD and ISFA has the same shortcomings in VHR images, because they are based on the central limit

TABLE I  
ACCURACY ASSESSMENT ON CD RESULTS OBTAINED BY DIFFERENT METHODS ON WH DATA SET

Method	Rec.	Pre.	OA	F1	KC
MAD	<b>0.8968</b>	0.1905	0.8446	0.3143	0.2662
IRMAD	0.5708	0.3651	0.9212	0.3651	0.3289
SFA	<b>0.8866</b>	0.1695	0.8230	0.2846	0.2335
ISFA	0.6408	0.2790	0.9200	0.3888	0.3530
CVA	0.5679	0.7682	0.9711	0.6530	0.6434
PCA-Kmeans	0.7303	0.5836	0.9686	0.6488	0.6325
SVM	0.5610	0.7596	0.9755	0.6454	0.6330
DSCN	0.4484	<b>0.7852</b>	0.9729	0.5708	0.5564
RNN-CD	0.6325	0.7001	0.9746	0.6646	0.6515
SiamCRNN(TR)	0.6983	0.7084	0.9766	0.7033	0.6912
SiamCRNN(GRU)	0.6949	0.7322	<b>0.9778</b>	<b>0.7131</b>	<b>0.7016</b>
SiamCRNN(LSTM)	0.6648	<b>0.8295</b>	<b>0.9813</b>	<b>0.7380</b>	<b>0.7285</b>

theorem, whereas the Gaussianity of VHR images is not obvious. Hence, they are not applicable for CD in VHR images. Although trained on the samples selected by automatic predetection, the SiamCRNN achieves the best result with OA of 0.9813, F1 of 0.7380, and KC of 0.7285. It means the proposed SiamCRNN can effectively extract spatial-spectral features by DSCNN, exhaustively mine change information by MRNN and fit the distributions of ground changes of VHR images from predetection samples. By contrast, the DSCN based on processing features with KC of 0.5564 and the RNN-CD lacking of extracting spatial context information effectively with KC of 0.6515 cannot compete with our approach. In addition, it is worth mentioning that the second and the third best results are two variants of our method, which demonstrate the effectiveness of the proposed architecture again.

Fig. 12 shows CD results obtained by the proposed method and comparison methods on HY data set. Similar to what we observed on WH data set, the misclassification and noise in

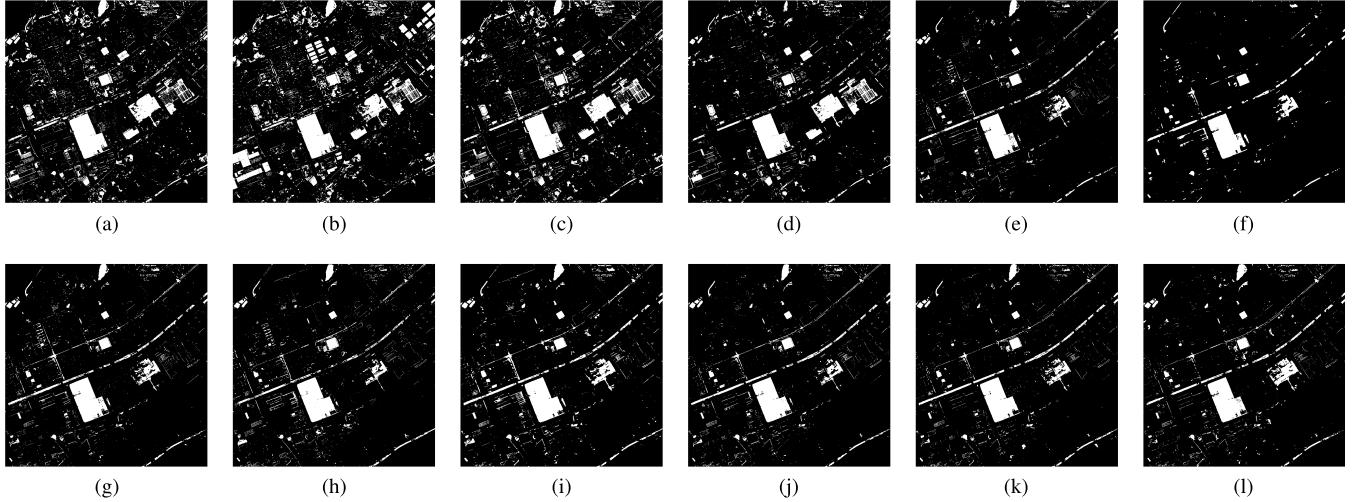


Fig. 12. Binary change maps obtained by the proposed method and comparison methods on HY data set. (a) MAD, (b) IRMAD, (c) SFA, (d) ISFA, (e) CVA, (f) PCA-K-means, (g) SVM, (h) DSCN, (i) RNN-CD, (j) SiamCRNN(FC), (k) SiamCRNN(GRU), and (l) SiamCRNN(LSTM).

the results obtained by MAD and SFA are obvious, where dozens of unchanged pixels are falsely detected as changes. Some important changes, such as railways, are detected incompletely. Induced by complexity of VHR images, the results obtained by IRMAD and ISFA are worse and have more misclassifications via iteratively reweighting. Besides, they are puzzled by “over-exposed” problem again. The CD results acquired by CVA and PCA-K-means are better in visual. Nonetheless, some margins of building are misclassified by CVA and a part of obvious change of build-up regions are missed by PCA-K-means. And the binary change maps obtained by SVM and DSCN is similar to that obtained by CVA, yet more edges are falsely detected.

Compared with the results obtained by SVM, more changes of groundwork before building over are missed by DSCN. In the binary change map generated by RNN-CD, a majority of change and unchanged regions are correctly classified. But due to not utilizing spatial context information efficiently, there exist a few falsely detected edges in the result. What is more, the fact it is slightly influenced by “over-exposed” problem. Once again, the SiamCRNN and its variants generate better qualitative results. Though a few changed regions are missed, most of unchanged regions are accurately classified. Furthermore, compared with its two variants, the changed regions are detected more completely and there is undeniable that our SiamCRNN outperforms the other methods in comparison and achieves best qualitative result.

Table II lists the quantitative analysis results on HY data set. The accuracy of MAD, SFA and their variants are unsatisfactory. The accuracy of CVA and PCA-K-means is higher, with OA of 0.9445 and 0.9405, F1 of 0.7480 and 0.7110, and KC of 0.7174 and 0.6794. And PCA-K-means acquires the highest precision rate. Because of falsely detecting more margins of building, the accuracy of DSCN is lower than CVA with OA of 0.9340, F1 of 0.7074, and KC of 0.6706. By constructing hyperplane in high-dimensional space to separate change and nonchange, the accuracy of SVM is slightly better than CVA with 0.7210 of KC. By contrast, RNN-CD, extracting

TABLE II  
ACCURACY ASSESSMENT ON CD RESULTS OBTAINED BY DIFFERENT METHODS ON HY DATA SET

Method	Rec.	Pre.	OA	F1	KC
MAD	0.7709	0.5989	0.9074	0.6741	0.6211
IRMAD	0.7110	0.4356	0.8497	0.5402	0.4565
SFA	<b>0.7876</b>	0.6313	0.9165	0.7008	0.6530
ISFA	0.7120	0.6481	0.9162	0.6786	0.6305
CVA	0.6631	0.8579	0.9445	0.7480	0.7174
PCA-Kmeans	0.5899	<b>0.8948</b>	0.9405	0.7110	0.6794
SVM	0.6898	0.8278	0.9436	0.7348	0.7210
DSCN	0.6428	0.7865	0.9340	0.7074	0.6706
RNN-CD	0.7222	0.8044	0.9437	0.7611	0.7293
SiamCRNN(TR)	0.6803	0.8490	0.9453	0.7554	0.7250
SiamCRNN(GRU)	0.7051	0.8734	0.9507	<u>0.7803</u>	<u>0.7528</u>
SiamCRNN(LSTM)	0.7228	<u>0.8843</u>	<b>0.9538</b>	<b>0.7954</b>	<b>0.7697</b>

information between multitemporal images based on RNN architecture, is better than SVM, with a higher OA of 0.9437, F1 of 0.7611 and KC of 0.7293. The variant, SiamCRNN(TR), does not show an outstanding performance in HY data set compared with RNN-CD and SVM, which demonstrate that the choice of RNN architectures affects the mining ability of the network. Finally, the proposed SiamCRNN achieves the highest accuracy with 0.9538 of OA, 0.7954 of F1 and 0.7697 of KC and one of the variant, SiamCRNN(GRU) achieves the second highest accuracy, which proves the effectiveness of our network in homogeneous VHR images CD once again.

For the purpose of further comparing change information mining capability of the three RNN architectures and proving the superiority of LSTM, the CD results of three networks in a certain local area of HY data set are shown in Fig. 13. SiamCRNN(TR) performs feature mapping and information mining only through FC layer of traditional RNN, thus the binary change map obtained by it is worst. In the result,

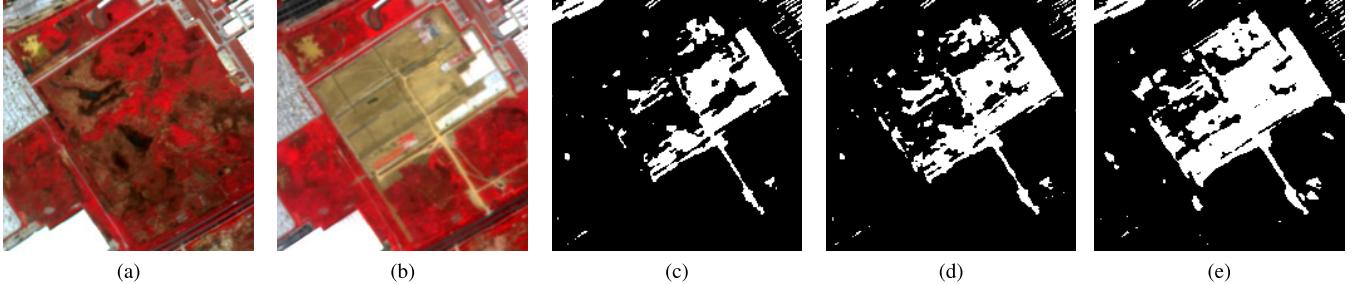


Fig. 13. Comparison of local area in binary change maps obtained by SiamCRNN and its two variants. (a) and (b) Multitemporal local areas. (c)–(e) Results obtained by SiamCRNN(TR), SiamCRNN(GRU), and SiamCRNN, respectively.

TABLE III  
COMPARISON OF RUNTIME OF DIFFERENT CD  
METHODS ON WH DATA SET

Method	Training time (s)	Inference time (s)
IRMAD	-	9.87
ISFA	-	13.00
SVM	5.01	68.89
DSCN	200.71	9.33
RNN-CD	47.67	3.12
SiamCRNN(TR)	197.76	6.17
SiamCRNN(GRU)	206.79	6.95
SiamCRNN(LSTM)	205.03	6.72

the internal fracture is severe, and the changed region with similar spectral features in the upper left corner is not detected. Compared with SiamCRNN(TR), the binary change map obtained by SiamCRNN(GRU) is slightly better. Depending on GRU, a variant of LSTM, the results of SiamCRNN(GRU) are more continuous and complete. However, limited to the mining capacity of GRU, the change region with similar spectral features in the upper left corner is almost not detected. In contrast, the binary change map detected by the proposed SiamCRNN is extremely complete and most of the changes in this area are detected, which imply that the general way of feature mapping and the universal change-rules can be effectively learned by the three doors and demonstrate the superiority of LSTM.

#### E. Runtime Analysis

We present the comparison of the runtime of IRMAD, ISFA, SVM, DSCN, RNN-CD, SiamCRNN(TR), SiamCRNN(GRU), SiamCRNN on WH data set in Table III. IRMAD, ISFA, and SVM are implemented with Python and run on CPU. DSCN, RNN-CD, SiamCRNN and its two variants are implemented with Python and run on GPU. The CPU used is Intel Core i7-8750H with a clock rate of 2.2 GHz. The GPU used is a single NVIDIA GeForce GTX 1060.

As shown in Table III, IRMAD and ISFA are unsupervised, so they do not require training. The inference time of ISFA and IRMAD are 9.87s and 13.00s, respectively. In the six supervised methods, SVM has the shortest training time, but the inference phase of SVM is time-consuming. It takes 68.89 s for SVM to infer the full CD result. The inference

phase of RNN-CD is the fastest, followed by the proposed SiamCRNN and its two variants. Though the training phase of the proposed method is relatively time-consuming, the speed of inference is fast, it only takes 6.72 s for SiamCRNN to infer the full CD result. Considering its good performance, the computational cost of SiamCRNN is acceptable.

## IV. EXPERIMENT OF CD IN HETEROGENEOUS VHR IMAGES

#### A. Data Description

Not only can be used for CD in homogeneous images, the SiamCRNN also can detect changes between multitemporal heterogeneous VHR images. In the experiment of CD in heterogeneous VHR images, LiDAR-Opt data set is adopted to evaluate the proposed method. This data set covers a part of downtown area in Buffalo, NY, USA. As shown in Fig. 14, LiDAR-Opt data set consists of multitemporal LiDAR image and optical VHR image. The airborne LiDAR intensity data [see Fig. 14(a)] was acquired in 2008 with spatial resolution of 0.956 m/pixel. The optical image [see Fig. 14(b)], captured in 2011, was provided by the New York State Digital Orthoimagery Program (DYSDOP). The spatial resolution of optical image is 0.3048 m/pixel and it contains four bands: red band, green band, blue band, and near-infrared band. After image registration, the spatial resolution of the two images is 1 m/pixel and the size is 800 × 1024. The ground truth [see Fig. 14(c)] is acquired manually by integrating some prior information with image interpretation based on the input images. In ground truth, changed area (red) contains 14561 pixels, and unchanged area contains 101858 pixels.

Note that the statistical properties in LiDAR and optical images are distinct, LiDAR-Opt data set is much more challenging than WH and HY data sets. Besides, we could observe that there exist a few shadows in the optical image, which makes accurate CD more difficult.

#### B. Experiment Setting

Owing to different statistical properties of LiDAR and optical images, the two branches of DSCNN are designed respectively (it is necessary to modify the structure of the DSCNN according to the type of image). The LiDAR image only contains one intensity band, and spectral features available for extraction are relatively fewer, thus the branch (called

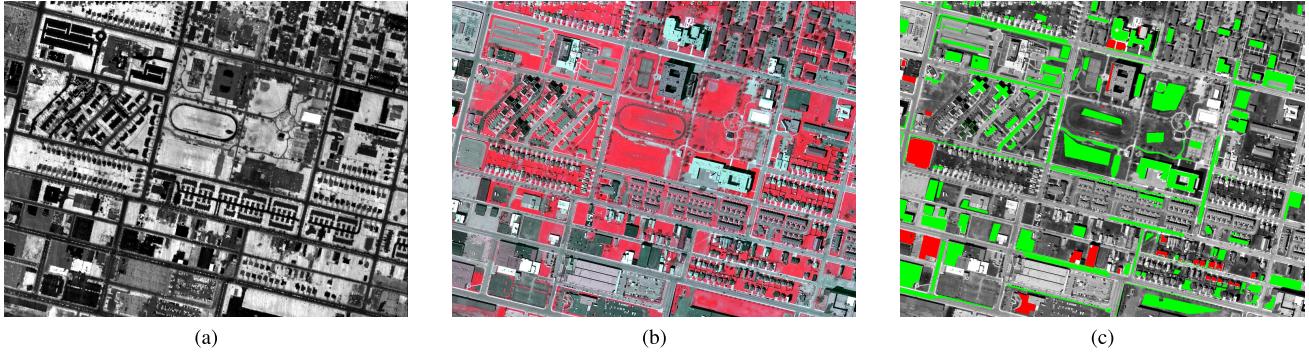


Fig. 14. Illustration of LiDAR-Opt data set. (a) Airborne LiDAR intensity image acquired in 2008. (b) Pseudocolor VHR optical image captured in 2011. (c) Ground truth, where red indicates change and green indicates nonchange.

LiDAR branch) used to process the LiDAR image is relatively simple, which consists of three convolution layers. The size of convolution kernel in the former two layers is  $3 \times 3$  and the number of convolution kernel in each layer are 16, 32, and 64. For optical VHR images, there are more spectral features available. Therefore, the branch (called optical branch) for feature extraction in optical VHR image is deeper and contains six convolution layers. The size of convolution kernel in the former five layers is  $3 \times 3$  and the number of convolution kernel per layer are 16, 16, 32, 32, 64, and 64. For two branches, ReLU is adopted as activation function. The LSTM unit numbers of hidden layers in the MRNN are 64 and 64. The neural numbers of FC are 64, 32, and 1. The weights of convolution kernel in DSCNN are initialized by he-normal way and the parameters of MRNN and FC are randomly initialized. The loss function also applies weighted binary cross-entropy function. Adam optimizer is adopted and the learning rate is set to 1e-4. Similar to heterogeneous experiments in [31],  $5 \times 5$  is adopted as the image patch size in our experiment.

For the purpose of demonstrating the superiority of SiamCRNN, eight representative methods are utilized for comparison: direct subtraction (DS), postclassification comparison (PCC), MAD [14], IRMAD [15], S-MAD [14], SVM, early fusion (EF)-DCNN, and DSCN [25]. PCC is a widely used framework for CD and can be used in heterogeneous images. MAD is based on CCA and is applicable for different view data [40], [41]. After learning two transformation matrices on training samples, S-MAD maps heterogeneous images into new feature space and minimizes the difference between unchanged samples to highlight the changed regions. SVM with RBF constructs hyperplane in high-dimensional space to demarcate changes between heterogeneous VHR images. The optimal hyperparameters  $C$  and  $\gamma$  are chosen by grid search. EF-DCNN is a deep convolutional neural network, which stacks multitemporal heterogeneous images in spectral channel and end-to-end output change probability. To prove the mapping and mining capability of MRNN, DSCN is also compared with our method. Because traditional RNN and its variants cannot process heterogeneous images with different bands directly, the RNN-CD cannot be adopted as a comparison method. K-means, Otsu, and FCM are adopted as segmentation methods, the best results of comparison methods are chosen as final results.

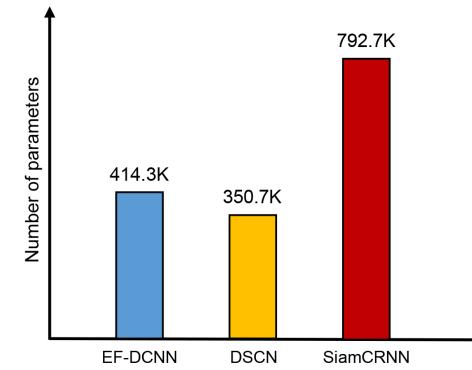


Fig. 15. Comparison of EF-DCNN, DSCN, and SiamCRNN in terms of model size.

The total trainable parameters in three deep-learning methods, namely EF-DCNN, DSCN, and SiamCRNN, are shown in Fig. 15. The model size of DSCN is the smallest due to its pure siamese network architecture. And compared with DSCN, the number of parameters in EF-DCNN is increased by about 18%. The number of parameters in SiamCRNN is the most, which implies that SiamCRNN has a more powerful fitting ability.

For training SiamCRNN and comparison methods, 4719 (0.57% of the all image patches) image patches are selected as training samples, which are different from the test samples in ground truth. Specifically, it contains 732 changed samples and 3987 unchanged samples.

Similar to the homogeneous experiment, recall rate, recall rate, precision rate, overall accuracy, F1 score and kappa coefficient are adopted as evaluation criteria.

### C. Experimental Results and Analysis

As elaborated in Section II-B, the LiDAR image and optical image are extracted by two different branches. Through several convolution layers, two high-dimension spectral-spatial features of LiDAR image and optical image are extracted. A portion of visualized feature images are illustrated in Fig. 16. It is obvious that the DSCNN is able to extract meaningful features, such as the object contours, vegetation, roads, shadows, and buildings. Totally, the feature images extracted by LiDAR and optical branch is representative and diverse.

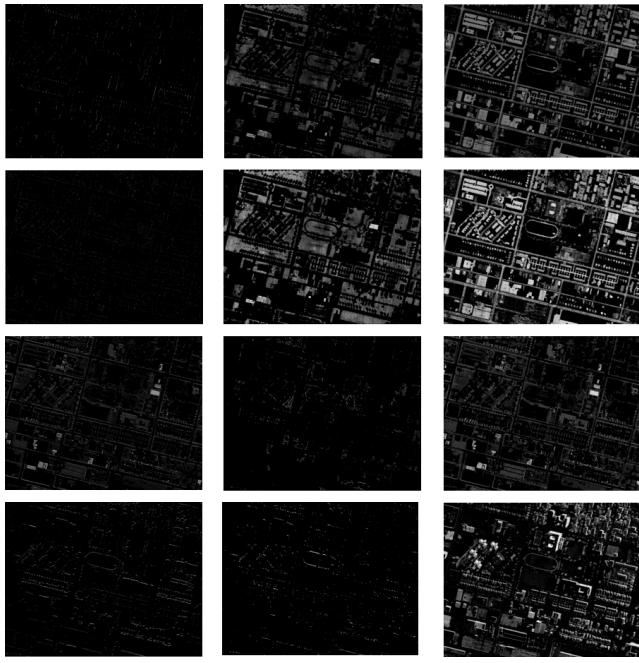


Fig. 16. Part of high dimensional spectral-spatial feature images extracted by DSCNN from LiDAR image and optical image. The images in the first and second rows are feature images extracted by the LiDAR branch. The images in the third and fourth rows are feature images extracted by optical branch.

Fig. 17 shows the CD results obtained by SiamCRNN and other comparison methods on the data set. In Fig. 17(a), it is obvious that the DS method has failed to detect changes between heterogeneous images. In the results obtained by DS, a large number of unchanged regions are divided into change and some important changed regions are not detected. By clustering two images independently and detecting change, the results detected by PCC is slightly better than DS. However, owing to distinct statistical properties of LiDAR images and optical images, a lot of unchanged regions are mistakenly detected as change.

Based on CCA, MAD maximizes the variance difference of projected features and detects change. In Fig. 17(c), compared with DS and PCC, more changed regions are recognized and fewer unchanged regions are mistakenly detected. Nonetheless, it is difficult to find correlation between multitemporal LiDAR image and optical image, the results obtained by MAD is still unsatisfactory. Therefore, it is not surprised that the binary change map detected by IR-MAD is worse than MAD. After iteratively reweighing, most of the regions in multitemporal images are detected as change. By training on samples, plenty of unchanged detected result of S-MAD is correctly classified. But a lot of change regions are not highlighted by CCA caused by distinct statistical properties of two images.

By constructing hyper-plane to separate change and non-change, the binary change map generated by SVM is better than all aforementioned methods. In the result of SVM, the main changes between the LiDAR image and optical image are detected and most of the unchanged regions are correctly classified. But due to utilizing spatial context infor-

mation inefficiently and relatively weak mining ability, some changed regions are fragmented and the shaded areas are falsely detected as change. In contrast, the EF-DCNN, extracting features and information by deep convolution structure, generates a better qualitative result [see Fig. 17(g)]. In the binary change map generated by EF-DCNN, a small part of unchanged regions are false recognized and changed regions are more complete. However, subjected to limitations on the CNN architecture, the shaded regions are misclassified as change. In Fig. 17(h), the results acquired by DSCN is not satisfactory, in which many changed regions are not recognized and plenty of unchanged regions are falsely detected as change. This means that the features obtained by DSCN are located in different feature space and the directly processing way of DSCN is not applicable for heterogeneous images. Eventually, as shown in Fig. 17(i), the SiamCRNN is the best to complete the detection task. In the result obtained by our SiamCRNN, most of the changed regions are detected precisely, and unchanged regions are correctly classified. What is more, our method is less affected by shadows.

The quantitative analysis of the proposed method and all competitors on LiDAR-Opt data set is reported in Table IV. As we observed in their binary change maps, DS and PCC get very low accuracy, with KC of 0.1649 and 0.2716. Compared with them, MAD and S-MAD achieve a slightly good result, with OA of 0.7072, F1 of 0.4354, and KC of 0.3031. Because of setting almost all regions into change, the recall rate of IR-MAD is the highest, however, its precision rate, OA, F1, and KC are very low. SVM, as a classic machine method, achieves a relatively high accuracy, with OA of 0.9391, F1 of 0.7589, and KC of 0.7240. The performance of EF-DCNN is better than SVM in all five evaluation criteria, which in one aspect implies the superiority of deep network architecture. The accuracy of DSCN is not good, with F1 of 0.4562 and KC of 0.3682, which means a purely deep siamese network structure is not suitable for processing heterogeneous VHR images. Finally, the proposed SiamCRNN achieves the best performance in LiDAR-Opt data set with a precision rate of 0.8738, OA of 0.9579, F1 of 0.8215, and KC of 0.7978. Compared with SVM, DSCN, and EF-DCNN, our method behaves prominent superiority in all five evaluation criteria, which demonstrate that the proposed SiamCRNN is applicable to heterogeneous VHR images. Note that only 0.57% of all image patches are selected for training, which means the SiamCRNN has good generalization capabilities. Therefore, if the type of input data does not change and the ground situations of the areas covered by input data are similar, the SiamCRNN does not need to train again.

As we introduced in Section II-C, MRNN is able to extract more change information. Therefore, we make a comparison between SiamCRNNs with different RNN layers. The number of layer ranges from 1 to 3. Caused by the self-connectionist architecture of RNN, when the number of layers is greater than 3, the network would be difficult to train and easier to overfitting. The performance comparison of SiamCRNNs with different RNN layers is shown in Fig. 18, where Fig. 18(a) shows the receiver operating characteristic (ROC) plots and Fig. 18(b) shows the histograms of AUC, OA, and KC values.

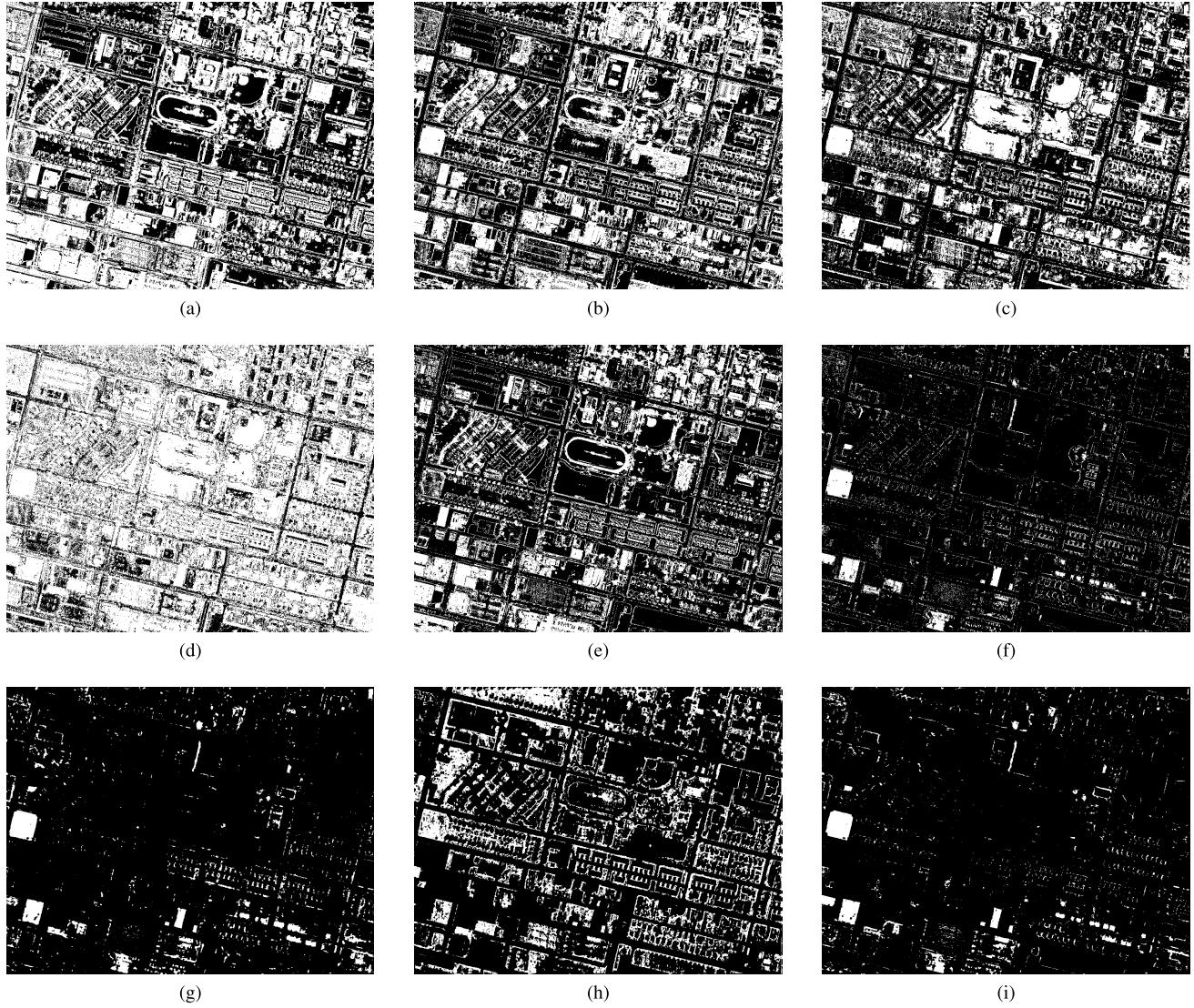


Fig. 17. Binary change maps obtained by the proposed method and comparison methods on LiDAR-Opt data set. (a) DS, (b) PCC, (c) MAD, (d) IRMAD, (e) S-MAD, (f) SVM, (g) EF-DCNN, (h) DSCN, and (i) SiamCRNN.

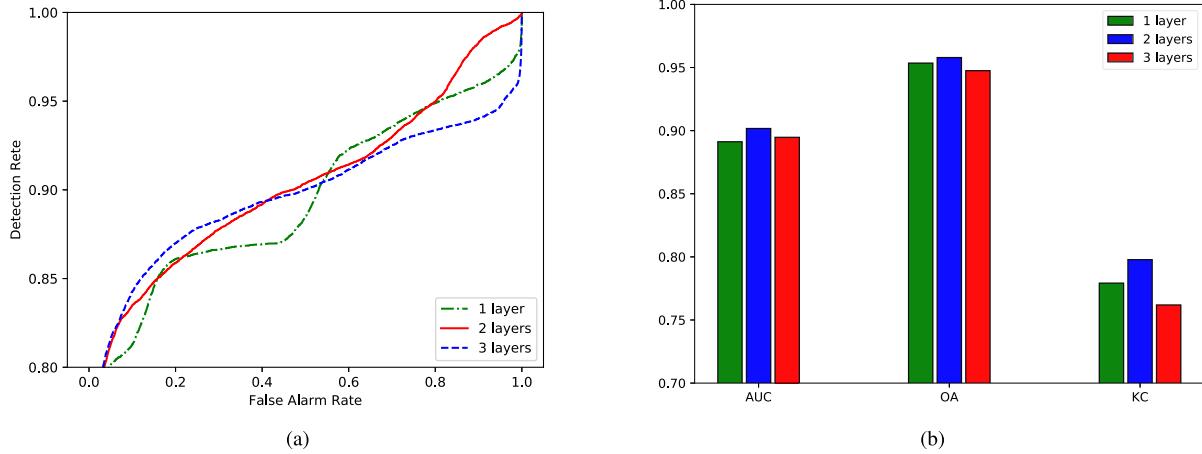


Fig. 18. Comparison between SiamCRNNs with different hidden layers in MRNN on LiDAR-Opt data set. (a) ROC plots. (b) Histograms of AUC, OA, and KC values.

The ROC plots of SiamCRNNs with different layers show that the ROC curve of SiamCRNN with 2 layers is more stable and almost envelops the other two curves, which indicate

that SiamCRNN with 2 layers has a better performance. In Fig. 18(b), it is obvious to find that as the number of RNN layers increases from 1 to 2, the capacity of network

TABLE IV  
ACCURACY ASSESSMENT ON CD RESULTS OBTAINED BY DIFFERENT METHODS ON LiDAR-OPT DATA SET

Method	Rec.	Pre.	OA	F1	KC
DS	0.7567	0.2097	0.6130	0.3284	0.1649
PCC	0.8042	0.2670	0.6776	0.4123	0.2716
MAD	<u>0.9028</u>	0.2869	0.7072	0.4354	0.3031
IR-MAD	<b>0.9780</b>	0.1674	0.3891	0.2859	0.0920
S-MAD	0.6091	0.2661	0.7410	0.3704	0.2377
SVM	0.7664	0.7515	0.9391	0.7589	0.7240
EF-DCNN	0.7679	<u>0.8063</u>	<u>0.9479</u>	<u>0.7867</u>	<u>0.7570</u>
DSCN	0.5149	0.4095	0.8465	0.4562	0.3682
SiamCRNN	0.7752	<b>0.8738</b>	<b>0.9579</b>	<b>0.8215</b>	<b>0.7978</b>

is significantly improved, which demonstrate that multiple RNN has more powerful mining ability. Nonetheless, when the number of RNN layers increases from 2 to 3, the performance of SiamCRNN is damaged. This is because MRNN of two layers with LSTM units has been able to learn the universal change-rules and mine enough change information. When the number of RNN layers is greater than 2, the network becomes difficult to train, and too many parameters would not make the network more capable of mining, but make the network easier to overfit. Therefore, for the number of RNN layers in MRNN, 2 is usually a suitable choice.

## V. CONCLUSION

In this article, a powerful and general end-to-end network, called deep siamese convolutional MRNN is proposed for CD in multitemporal homogeneous and heterogeneous VHR images. The SiamCRNN consists of three subnetworks: DSCNN, MRNN, and FC layers. The first part of SiamCRNN is DSCNN. For homogeneous VHR images, the DSCNN is designed as a pure-siamese network structure consisting of two weight-shared deep convolutional branches, which extracts spatial-spectral features from image patches in the exactly same way to highlight change information. For heterogeneous VHR images, the DSCNN is modified as a pseudo-siamese network structure and its two branches are designed for two heterogeneous VHR images to extract more representative spatial-spectral features. Then the two features are mapped into a new latent feature space and mined change information between them through MRNN constructed by LSTM units. The general way of feature mapping and the universal change-rules are learned by the three doors of LSTM. After change information is fully extracted, a spatial-spectral-temporal features with abundant change information is generated. Finally, the change probability is predicted by the FC.

In the experiment of CD in homogeneous VHR images with the challenging WH and HY data sets, the qualitative and quantitative results indicate that the SiamCRNN outperforms most widely used CD methods with better overall accuracy, F1 score and kappa coefficient. The experimental results of CD in heterogeneous VHR images with LiDAR-Opt

data set verify the effectiveness of SiamCRNN and also manifest the superiority of SiamCRNN over several existing methods with more prominent precision rate, overall accuracy, F1 score, and kappa coefficient. Both experimental results reveal that SiamCRNN is applicable for CD in homogeneous VHR images and heterogeneous VHR images with different statistical characteristics.

Our future work includes but is not limited to applying SiamCRNN or its variants for CD in hyper-spectral images.

## REFERENCES

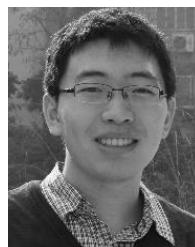
- [1] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989.
- [2] G. Xian, C. Homer, and J. Fry, "Updating the 2001 national land cover database land cover classification to 2006 by using Landsat imagery change detection methods," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1133–1147, 2009.
- [3] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Digital change detection methods in ecosystem monitoring: A review," *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1565–1596, 2004.
- [4] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Change detection techniques," *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2401, 2004.
- [5] H. Luo, C. Liu, C. Wu, and X. Guo, "Urban change detection based on Dempster-Shafer theory for multitemporal very high-resolution imagery," *Remote Sens.*, vol. 10, no. 7, pp. 20–22, 2018.
- [6] M. E. Zelinski, J. Henderson, and M. Smith, "Use of Landsat 5 for change detection at 1998 Indian and Pakistani nuclear test sites," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3453–3460, Aug. 2014.
- [7] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using VHR optical and SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2403–2420, May 2010.
- [8] C. Wu, B. Du, X. Cui, and L. Zhang, "A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion," *Remote Sens. Environ.*, vol. 199, pp. 241–255, Sep. 2017.
- [9] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.
- [10] F. Bovolo, "A multilevel parcel-based approach to change detection in very high resolution multitemporal images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 1, pp. 33–37, Jan. 2009.
- [11] F. Thonfeld, H. Feilhauer, M. Braun, and G. Menz, "Robust change vector analysis (RCVA) for multi-sensor very high resolution optical satellite data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 50, pp. 131–140, Aug. 2016.
- [12] C. E. Woodcock and J. B. Collins, "An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data," *Remote Sens. Environ.*, vol. 56, no. 1, p. 66, 1996.
- [13] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and K-means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, Oct. 2009.
- [14] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sens. Environ.*, vol. 64, no. 1, pp. 1–19, 1998.
- [15] A. A. Nielsen, "The regularized iteratively reweighted mad method for change detection in multi- and hyperspectral data," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 463–478, Feb. 2007.
- [16] C. Wu, B. Du, and L. Zhang, "Slow feature analysis for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2858–2874, May 2014.
- [17] L. Zhang, C. Wu, and B. Du, "Automatic radiometric normalization for multitemporal remote sensing imagery with iterative slow feature analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6141–6155, Oct. 2014.

- [18] C. Wu, L. Zhang, and B. Du, "Kernel slow feature analysis for scene change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2367–2384, Apr. 2017.
- [19] Y. Tang, X. Huang, and L. Zhang, "Fault-tolerant building change detection from urban high-resolution remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1060–1064, Sep. 2013.
- [20] X. Huang, W. Yuan, J. Li, and L. Zhang, "A new building extraction postprocessing framework for high-spatial-resolution remote-sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 2, pp. 654–668, Feb. 2017.
- [21] H. Yang, P. Wu, X. Yao, Y. Wu, B. Wang, and Y. Xu, "Building extraction in very high resolution imagery by dense-attention networks," *Remote Sens.*, vol. 10, no. 11, p. 1768, 2018.
- [22] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [23] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the Art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.
- [24] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 125–138, Jan. 2015.
- [25] 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 Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1845–1849, Oct. 2017.
- [26] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised deep change vector analysis for multiple-change detection in VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3677–3693, Jun. 2019.
- [27] H. Lyu, H. Lu, and L. Mou, "Learning a transferable change rule from a recurrent neural network for land cover change detection," *Remote Sens.*, vol. 8, no. 6, p. 506, 2016.
- [28] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, Feb. 2019.
- [29] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9976–9992, Dec. 2019. [Online]. Available: <http://arxiv.org/abs/1812.00645>
- [30] T. Zhan, M. Gong, X. Jiang, and S. Li, "Log-based transformation feature learning for change detection in heterogeneous images," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 9, pp. 1352–1356, Sep. 2018.
- [31] X. Niu, M. Gong, T. Zhan, and Y. Yang, "A conditional adversarial network for change detection in heterogeneous images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 45–49, Jan. 2019.
- [32] 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 Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 545–559, Mar. 2018.
- [33] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, 2003.
- [34] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [36] M. Gong, X. Niu, P. Zhang, and Z. Li, "Generative adversarial networks for change detection in multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2310–2314, Nov. 2017.
- [37] F. Gao, J. Dong, B. Li, and Q. Xu, "Automatic change detection in synthetic aperture radar images based on PCANet," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1792–1796, Dec. 2016.
- [38] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
- [39] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [40] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *Proc. 30th Int. Conf. Mach. Learn. (ICML)*, 2013, vol. 28, no. 3, pp. 2284–2292.
- [41] W. Wang, R. Arora, K. Livescu, and J. Bilmes, "On deep multi-view representation learning," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, vol. 2, 2015, pp. 1083–1092.



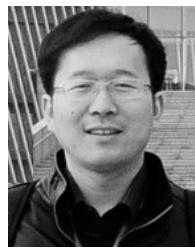
**Hongruixuan Chen** received the B.S. degree in surveying and mapping engineering from the School of Resources and Environmental Engineering, Anhui University, Hefei, China, in 2019. He is currently pursuing the M.S. degree with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote sensing, Wuhan University, Wuhan, China.

His research interests include deep learning and multitemporal remote sensing image change detection.



**Chen Wu** (M'16) received B.S. degree in surveying and mapping engineering from Southeast University, Nanjing, China, in 2010, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote sensing, Wuhan University, Wuhan, China, in 2015.

He is currently an Associate Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His research interests include multitemporal remote sensing image change detection (CD) and analysis in multispectral and hyperspectral images.



**Bo Du** (M'10–SM'15) received the B.S. and Ph.D. degrees in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2005 and 2010, respectively.

He is currently a Professor with the School of Computer, Wuhan University. He has more than 40 research articles published in the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS), the IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP), the IEEE JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND APPLIED REMOTE SENSING (JSTARS), and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (GRSL). Five of them are ESI hot papers or highly cited papers. His major research interests include pattern recognition, hyperspectral image processing, and signal processing.

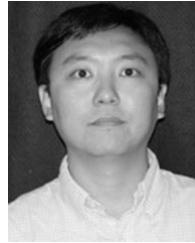
Dr. Du received the Best Reviewer Awards from the IEEE GRSS for his service to the IEEE JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND APPLIED REMOTE SENSING (JSTARS) in 2011 and ACM Rising Star Awards for his academic progress in 2015. He was the Session Chair for both International Geoscience and Remote Sensing Symposium (IGARSS) 2016 and the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS). He also serves as a Reviewer for 20 Science Citation Index (SCI) Magazines including IEEE TGRS, TIP, JSTARS, and GRSL.



**Liangpei Zhang** (M'06–SM'08–F'19) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He was a Principal Scientist with the China State Key Basic Research Project from 2011 to 2016 appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He is currently a Chang-Jiang Scholar Chair Professor appointed by the Ministry of Education of China with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University. He is an Institute for Scientific Information (ISI) highly cited author. He has authored or coauthored many research articles and five books. He holds 30 patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a fellow of the Institution of Engineering and Technology (IET). He was a recipient of the 2010 Best Paper Boeing Award, the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing (ASPRS), and the 2016 Best Paper Theoretical Innovation Award from the International Society for Optics and Photonics (SPIE). His research teams won the top three prizes of the IEEE GRSS 2014 Data Fusion Contest, and his students have been selected as the winners or finalists of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS) Student Paper Contest in recent years. He is the Founding Chair of the IEEE Geoscience and Remote Sensing Society (GRSS) Wuhan Chapter. He also serves as an Associate Editor or an Editor for more than ten international journals. He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.



**Le Wang** received the B.E. degree from the Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1996, the M.S. degree from Peking University, Beijing, China, in 1999, and the Ph.D. degree in environmental science, policy, and management from the University of California at Berkeley, Berkeley, CA, USA, in 2003.

He is currently a Professor of geography with The State University of New York, Buffalo, NY, USA. He has authored or coauthored more than 90 refereed journal articles, which have been cited 4280 times on Google Scholar. His work has been funded by NSF, USDA, and USGS. His research interests focus on remote sensing of coastal mangrove forests, invasive species, and urban environments.

Dr. Wang was the Chair of the AAG Remote Sensing Specialty Group from 2015 to 2017. He is currently an Editor of the *International Journal of Remote Sensing*, a Guest Editor of *Remote Sensing of Environment*, and an Editorial Board Member of *Annals of the American Association of Geographers*.