IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 9, 2019, accepted September 26, 2019, date of publication October 14, 2019,
date of current version November 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2947286

Remote Sensing Image Change Detection Based on Information Transmission and Attention Mechanism

RUOCHEN LIU, (Member, IEEE), ZHIHONG CHENG, LANGLANG ZHANG, AND JIANXIA LI
Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an 710071, China

Corresponding author: Ruochen Liu (ruochenliu@xidian.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61876141, Grant 61373111, Grant 61272279, Grant 61103119, and Grant 61203303, and in part by the Provincial Natural Science Foundation of Shaanxi of China under Grant 2019JZ-26.

ABSTRACT Change detection is one of the core issues of earth observation and has been extensively studied in recent decades. This paper presents a novel deep neural network architecture based on information transmission and attention mechanism. Existing methods rely on a simple mechanism for independently encoding bi-temporal images to obtain their representation vectors. In view of the fact that these methods do not make full use of the rich information between bi-temporal images, we introduce the information transmission module in the design of DNN structure for doing the transmission and interaction of information. In addition, we introduce the attention mechanism behind the information transmission module to give the corresponding attention weight to each temporal image feature so as to enhance the change information of the image, which noticeably improves final prediction. The proposed network is validated on real remote sensing image data sets. Both visual and quantitative analyses of the experimental results demonstrate competitiveness of the proposed method.

INDEX TERMS Remote sensing image, change detection, deep neural network, information transmission, attention mechanism.

I. INTRODUCTION

Change detection is an important branch in the field of image processing. It has a long history of research and evolved with the development of computer vision. Change detection is the process of quantitatively analyzing and determining surface changes from object or phenomenon during the different periods [1]. The goal of the change detection system is to assign a binary label to each pixel based on a pair or series of co-registered images of a given area taken at different times. A positive label indicates that the area corresponding to the pixel has changed. Changes can refer to urban expansion, vegetation change [2], disaster assessment and so on. Change detection is a great tool for mapping urban coverage, land use, video monitoring, and other types of multi-temporal analysis.

Generally speaking, change detection methods can be divided into three categories: 1) Pixel level: pixel is always the basic unit of image analysis and CD technology, and it statistically analyzes the position where the change information occurs [3]. But the disadvantage is that the efficiency is low, the spatial and other characteristics are not taken into account, and the anti-interference ability is poor (natural factors such as sun illumination angle and surface humidity) [4]. 2) Feature level: the manual feature information (edge, shape, contour, texture) is extracted from the original image by a certain algorithm, and then the comprehensive analysis and change detection of these feature information are performed. The feature level-based method has higher operational efficiency, and the judgment of feature attributes has higher credibility and accuracy, which reduces the interference of external factors on the results to some extent. However, on the one hand, some information will be lost in the process of feature extraction, so it is difficult to provide subtle information. On the other hand, it depends on the result of feature extraction, but feature extraction itself is difficult. 3) Object-based level: Based on the definition

The associate editor coordinating the review of this manuscript and approving it for publication was Joey Tianyi Zhou.

by [1], [5], OBCD can be defined as ''the process of identifying differences in geographic objects at different moments using object-based image analysis''. OBCD has become increasingly popular due to its significant advantages over pixel-based approach. Because it offers more possibilities for the extraction of highly distinctiveness features that can better highlight the changed regions [4]. It mainly detects certain specific objects (such as roads, houses and other objects with clear meanings), and detects changes based on image understanding and image recognition. It is close to the needs of users, the detection results can be directly applied, but the object extraction has some difficulties.The traditional image change detection methods based on feature level and object level classify the objects in the image by extracting features, such as minimum distance method, maximum likelihood method, gray level co-occurrence matrix, wavelet transform and so on. Traditional feature extraction work is complex and time consuming, however, change detection, especially disaster detection, requires more efficient feature extraction methods. With the development of artificial intelligence, deep learning has been introduced into the field of remote sensing image processing. Commonly used deep learning algorithms include Deep Belief Network (DBN) [6], Recurrent Neural Network (RNN), Convolutional Neural Network (CNN) and more. Among them, CNN is the most commonly used algorithm in the field of computer vision. It can achieve high-precision classification and has obvious advantages in processing 2-dimensional image data. CNN uses the original image as input, avoiding the complicated feature extraction process, and does not require excessive manual participation in the feature learning process. RNN is a kind of neural network for processing sequence data, so it is natural to use RNN to model the temporal relationship between bi-temporal images.

In this paper, we present a novel network architecture which combines CNN with bidirectional long short-term memory network (BiLSTM)[16]. The former is responsible for extracting the rich spectral-spatial features of bi-temporal images, while the latter is effective in analyzing the temporal dependence of bi-temporal images and transferring the features of images. The features extracted by CNN module can better realize information transmission through BiLSTM. We also propose for the first time to apply the common attention mechanism in text processing to the enhancement of change information in the production of image feature vectors.

In summary, the contributions of this paper are as follow:

1) We propose a novel end-to-end network framework that combines CNN and RNN with information transmission module.

2) We introduce attention mechanism into the change detection task for improving detection performance for the first time.

The rest of this paper is organized as follows. Section II discusses other studies and related works. Section III describes in detail the proposed network architecture for change detection. Section IV contains qualitative and quantitative comparisons with previous change detection methods. Finally, the conclusion of this letter is drawn in Section V.

II. RELATED WORK

In the early age of change detection, many existing methods aim to obtain the difference map first [7], including the traditional difference map generation operator [1], [8], [9] and deep neural network to generate the difference map [10], [11], but generating the difference map increases the time cost of change detection, and the quality of the difference map generation also affected the performance of the detection algorithms. Therefore, based on the existing methods, researchers propose end-to-end change detection method abandoning the step of generating difference map, and extracting abstract features of images directly from the original image using deep neural network. Reference [12] introduced deep learning into change detection for the first time, they learned the feature representation of the image through DBNs network and superimposed the learned feature vectors together. For such feature representation, the change information was easy to be detected by image difference. Reference [13] proposed a change detection method based on a deep Siamese CNN for optical aerial images, and the Siamese network is learned to extract features directly from the image pairs. Although the change detection method based on DBNs and CNN can extract rich image features, neither of them considers the temporality between bi-temporal images. For modeling temporal connection between bi-temporal images, almost all the mentioned approaches adopt simple strategies: either image differencing or stacking. In this regard, RNN, as an important branch of the deep learning family, is a natural candidate for dealing with temporal relationships between multiple time series data in change detection problems. Reference [14] made RNN-based network to solve the multispectral change detection task, in which, the joint spectral-temporal feature representation is learned from a bi-temporal image sequence using long short-term memory network (LSTM). Reference [15] proposed a novel network architecture, which is trained to learn a joint spectral-spatial-temporal feature representation in a unified framework for change detection of multi-spectral images. For this purpose, they combined CNN and RNN into an end-to-end network framework. The former is responsible for extracting the rich spectral-spatial features of bi-temporal images, while the latter is effective in analyzing the temporal dependence of bi-temporal images.

Since our algorithm is based on deep neural network(DNN), we should review DNN and attention mechanism briefly before introducing our algorithm. DNN consists of a range of networks such as CNN, RNN, etc. CNN and RNN play a very important role in feature extraction and delivery respectively.

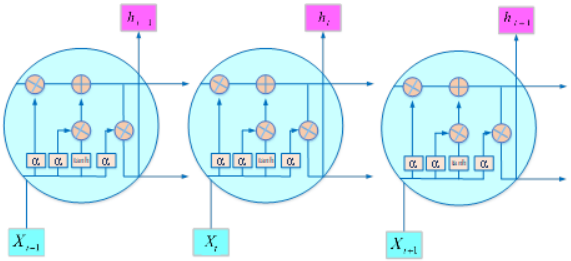


FIGURE 1. Internal Structure of LSTM Units.

A. CONVOLUTIONAL NEURAL NETWORK(CNN)

CNN [26] is a deep learning algorithm, which can receive input images, assign importance (learning weights and deviations) to all the objects contained in the images, and distinguish them. Compared with other classification algorithms, the pretreatment requirement of CNN is much lower. The function of CNN is to simplify the image into a more easily processed form without losing the features that are essential for good prediction.

The main core part is the convolution operation, and the formulated convolution operation can be defined as follows:

$$z_k^L = \sum_{i=1}^{N_{L-1}} X_i^{L-1} W_{ik}^L + b_k^L, \quad k = 1, 2, \ldots, N_L \tag{1}$$

$$O_{w,b}(x) = f\left(z_k^L\right) \tag{2}$$

where z_k^L represents the output of the layer L,N_{L-1} represents the number of feature maps in layer $L-1$, N_L represents the number of feature maps in layer L, X_i^{L-1} represents the $i-th$ feature map in layer $L-1$, W_{ik}^L represents the weight of the $k-th$ feature map from the $i-th$ feature map in layer $L-1$ to the L, b_k^L represents the bias of layer L, and $f(\cdot)$ represents the activation function, such as the Sigmoid activation function.

B. RECURRENT NEURAL NETWORK(RNN)

A recurrent neural network is a neural network that takes the output of the previous step as input and inputs it to the current step. The most important feature of RNN is the hidden state, which remembers some information about the sequence. Long Short-Term Memory (LSTM) [17], as one of the RNN models, allows the network to collect and save the relevant information and inject it into the model when necessary. In addition, LSTM can also help solve the problem of gradient disappearance during RNN training.

The internal structure of LSTM is shown in the Fig. 1. X_t refers to input vector to the LSTM unit and h_t refers to hidden state vector also known as output vector of the LSTM unit. Compared with the hidden layer operation of RNN, LSTM has a complex structure and a LSTM cell can be divided into an input gate,a forget gate and an output gate. These three gates are used to preserve relevant information for the later stages of the learning process.The input gate mainly acts on the cell state, which is responsible for selectively adding new information to the cell state,the forget gate selectively forgets the information in the cell state and the output gate controls the influence of long-term memory on the current input. The activation function of the LSTM gates is often the logistic function.

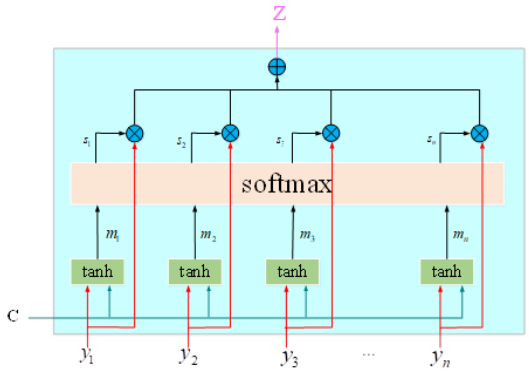


FIGURE 2. The Internal Structure of Attention Model.

C. ATTENTION MECHANISM

Attention mechanism [18] is a mechanism used to enhancements to an effect of RNN (LSTM or BiLSTM) based coding-decoding model. Attention mechanism is popular in many fields, such as image annotation, speech recognition, machine translation and so on. Attention mechanism can be combined with many existing network models and interpolated between layers. It can accept the output of the former layer of the network and the state representation of the context information, so that the context-related part of the output of the former layer can be screened out by the attention mechanism.

As we can see in the Fig. 2, attention model receives n inputs $y_1, y_2, \ldots, y_n$ and context C. The n input is the hidden state of the output from the front layer of the network. The output Z of the model is the arithmetic average of y_i, and its weight coefficient is the degree of correlation between each input y_i and context information. First of all, the aggregate value m_i of each input y_i and context C is calculated by tanh function. It is noteworthy that the calculation of each m_i is only related to its corresponding y_i, and does not depend on other inputs. The calculation formula is as follows:

$$m_i = \tanh\left(W_c C + W_y y_i\right) \tag{3}$$

where W_c and W_y are the corresponding weight matrices. Then the weighting coefficients s_i are calculated by using the softmax function. The output Z of the attention model is the average value of the weighted arithmetic of the input y_i and the weight coefficient s_i, and the calculation formula is as follows:

$$Z = \sum s_i y_i \tag{4}$$

Although the combination of CNN and RNN is a good established technology for remote sensing applications, the feature information between two temporal images is not fully taken into account in the whole feature extraction and

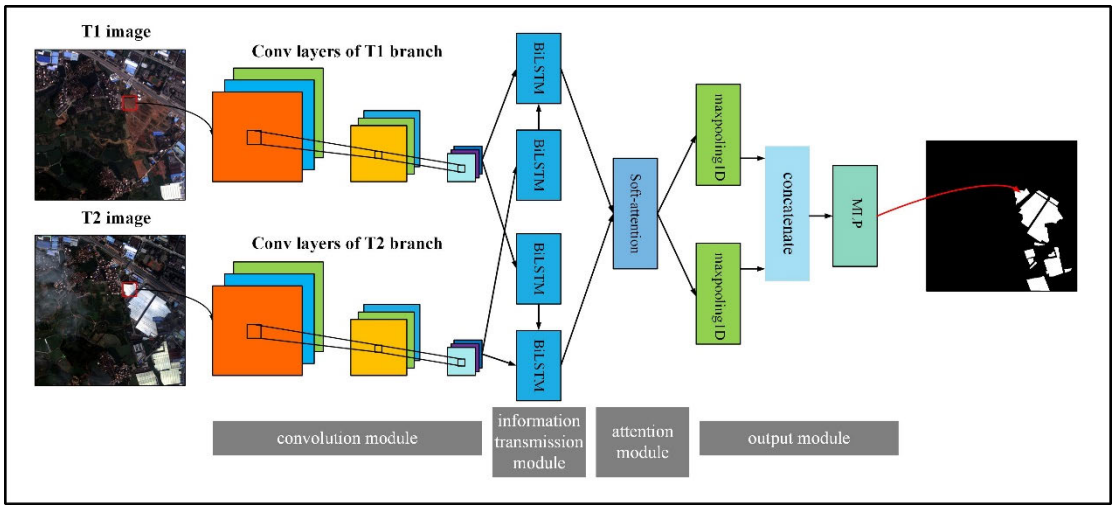


FIGURE 3. The overall framework of the proposed method.

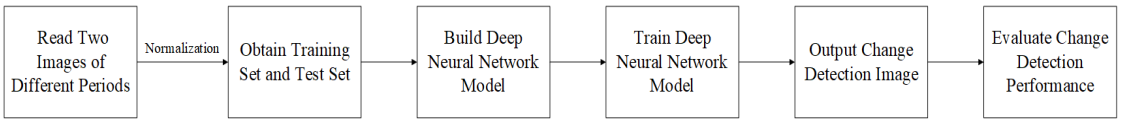


FIGURE 4. The flow chart of the proposed method.

sequential modeling process, and the transmission of information is lacking. In [15], the author directly regards the image features extracted by CNN at the previous moment as the hidden state of the RNN at the next moment and there is no interactive process.

On this foundation, we introduce an information transmission module, which is responsible for the transmission and interaction of features between bi-temporal images. We transmit information between the state identification of image processing features at the previous moment and the features at the next moment. BiLSTM can transmit and process the front and back information better than single-layer RNN. The state representation of the image features is obtained by the first BiLSTM, and the information are transmitted through the second BiLSTM. Besides, we integrate the attention mechanism behind the information transmission module to give the corresponding attention weight to each temporal image feature to enhance the change information of the image, which noticeably improves final prediction. As far as we know, it is the first time that attention mechanism has been introduced into the image change detection task.

III. PROPOSED FRAMEWORK

This section describes our network framework in detail. The architecture of the proposed method, as shown in Fig. 3, consists of four parts, including a convolution module, an information transmission module, an attention module, and an output module, from bottom to top. The convolution module is configured to perform bi-temporal image features extraction on the input image blocks by two independent convolutional neural networks. On the top of convolution module, the information transmission module uses the feature maps obtained by the convolution module as input to perform the transfer and interaction of the features of the bi-temporal images. The third part gives the corresponding attention weight to the hidden state of each branch image feature. The output module is two fully connected layers that are widely used in classification problems.We supply our overall framework with flow chart. In Fig.4, we use a certain proportion of the changed and unchanged parts as training set and test set. Generating change detection images by deep neural network and evaluating them. Although the network is made up of different modules, it can be trained end-to-end through the back-propagation algorithm.

A. SPECTRAL-SPATIAL FEATURE EXTRACTION VIA THE CONVOLUTION MODULE

Since it is time-consuming and difficult to extract image features manually, and the extracted image features are relatively simple, therefore, it directly resorts to the powerful feature extraction ability of CNN to extract spectral spatial features. CNN is an important branch of deep learning, which has attracted people's attention because it can automatically discover related contextual 2D spatial features and spectral features. Let I_1 and I_2 be two patches obtained from exactly the same position in the t_1 and t_2 images. Both of them have a shape of $9 \times 9 \times c$, where c is the number of channels.The reason why we choose image patch of 9×9 will be given

in Section IV. Let Y be the label that indicates the class to which the patch belongs (changed or not changed). The convolution module receives I_1 and I_2 as input, and has two separate but identical CNN branches (i.e., T1 branch and T2 branch (see Fig.3)), which process I_1 and I_2 in parallel, and the acquired image features are fed into the following information transmission module. Because of the size of the image patch is very small, we choose a convolutional filter with the receiving field 3×3 instead of using a larger filter, such as 5×5. We also use spatial padding in the convolution layers, and there is no pooling operation behind the convolutional layer. The convolution module has a depth of 2, and the reasons for not making use of more complex network structure similar to VGG, ResNet, etc., [19] which is typical in identification tasks, are as follows. First, since the spatial resolution of multispectral images is limited, it is necessary to make the input size small, thereby naturally reducing the depth of the network. Second, the size of our input image is very small, and we apply spatial padding in the convolution and discard the pooling layer. The large depth of the convolution module will lead to more misleading information introduced by the spatial padding. Third, in the change detection problem, smaller networks are obviously more efficient, and the real-time performance of the test can be implemented on a large scale.

B. INFORMATION TRANSMISSION MODULE

RNN is a type of neural network used to process sequence data. First of all, we need to clarify what is sequence data. Time series data refers to data collected at different points in time. This kind of data reflects the state or degree of a thing, phenomenon, etc that changes over time. Of course, it can be more than just time, such as a sequence of words. However, there is a characteristic of the total sequence data, that is, the latter data is related to the previous data. Reflected in image change detection task, RNN can well model temporal dependence between two input images. LSTM is a variant of RNN, and LSTM model can better capture long-distance dependencies. Because the LSTM can learn what to remember and what to forget through training process. However, there is still a problem with modeling with LSTM that cannot encode information from back to front. When we judge the category of a pixel at a certain location, we rely not only on the upstream information of the location but also on the downstream information. BiLSTM is a combination of forward LSTM and backward LSTM, which is often used to model context information in natural language processing tasks. We use BiLSTM to better capture bidirectional feature dependencies between bi-temporal images. Moreover, when modeling sequential dependencies, we cannot ignore the information between feature representations. Therefore, the additional information transfer module is integrated throughout the framework. The specific process is as follows.

As shown in Fig. 3, the information transmission module comprises a first BiLSTM and a second BiLSTM sequentially connected, and the first BiLSTM is used to obtain the state representation of each branch image features. Specifically, the extracted image features are reshaped adapted to the input format of BiLSTM, and then the image features of each branch are sent to the first BiLSTM of weight sharing to get the state representation. The second BiLSTM is used for image information transmission and interaction through each branch image feature and its state representation. The image features of each branch are input into second BiLSTM, and the initial state of second BiLSTM is initialized as the state representation of another branch image feature. In this way, the information transmission and interaction in the second BiLSTM are carried out, and the hidden state of each temporal image feature is output after interaction.

C. ATTENTION MODULE

In recent years, the combination of deep learning and visual attention mechanism is mostly focused on the use of mask to form attention mechanism. The principle of mask is to identify the key features in the image data by another layer of weight, and through learning and training, the deep neural network can learn the areas that need attention in every new picture. This idea has evolved into two different types of attention, one is soft attention, the other is hard attention. The key point of soft attention is that this kind of attention focuses more on the region or the channel, and it is the definite attention. After learning is completed, it can be directly generated through the network. Secondly, the most important point is that soft attention is differentiable, which is very important. Differential attention can be used to calculate gradients through neural networks, and to learn the weight of attention through forward propagation and backward feedback. The difference between hard attention and soft attention is that, first of all, hard attention pays more attention to point, that is, every point in the image is likely to extend attention. Meanwhile, hard attention is a random prediction process, with more emphasis on dynamic changes. The key point is that hard attention is a non-differentiable attention, and the training process is often completed through reinforcement learning.

In order to enhance the change information, we integrate the soft attention mechanism behind information transmission module. Specifically, the hidden state of each branch image features outputted by the information delivery module is used as an input of the attention module, then the corresponding attention weights are assigned to each temporal image features through the soft attention mechanism, and the weighted representation of each branch features is obtained.

D. OUTPUT MODULE

The output shape of the attention module is two dimensions, while the input shape of the fully connected layer is required to be one-dimensional. In order to adapt to the input format of the fully connected layer, we first perform GlobalMaxPooling1D operation on it. GlobalMaxPooling1D for temporal

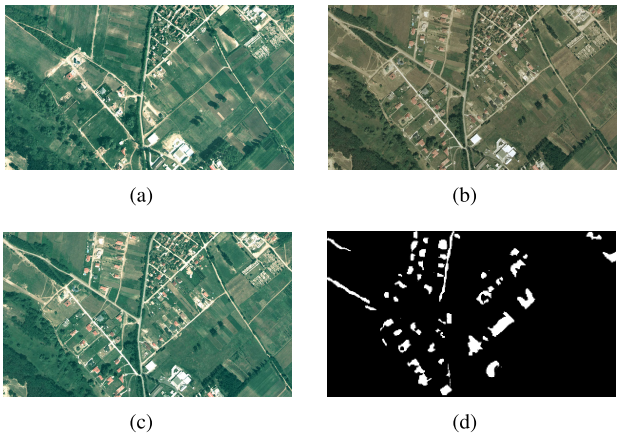


FIGURE 5. Optical aerial image pairs of SZADA/1 data set and radiation correction result. (a) Image acquired at t_1. (b) Image acquired at t_2. (c) Radiation correction result for image at t_2. (d) Ground truth.

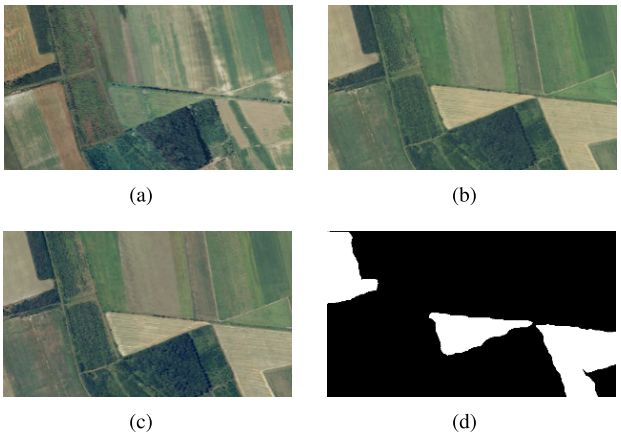


FIGURE 6. Optical aerial image pairs of TISZADOB/3 data set and radiation correction result. (a) Image acquired at t_1. (b) Image acquired at t_2. (c) Radiation correction result for image at t_2. (d) Ground truth.

data takes the max vector over the steps dimension. So a tensor with shape (*samples*, *steps*, *features*) becomes a tensor with shape (*samples*, *features*) after global pooling.

IV. EXPERIMENTAL STUDY

A. DATA SETS

In order to evaluate the performance of our proposed method, we choose two data sets for comparative experiments. One of them is SZTAKI AirChange Benchmark set [20], which is a ground truth collection for change detection in optical aerial images taken with a few years of time difference and in different seasonal conditions. The dataset contains three groups of optical aerial image pairs, named TISZADOB, SZADA and ARCHIEVE, in which respectively contains 5, 7, and 1 image pairs of size 952 × 640, resolution 1.5m/pixel and binary change masks (drawn by an expert). Given the apparent differences in radiation condition, color histogram matching is applied to the two co-registered images. Another is a data set for building change detection, called QuickBird data set, which is stored in Tiff image file format and captured in Guangdong Province, China, in 2015 and 2017, respectively. Each image contains four channels, namely, blue, green, red and near infrared in our experiments, three sets of images of the data set were used, each of which contained 512 × 512 pixels. And Figs.5-10 show different images in different data sets.

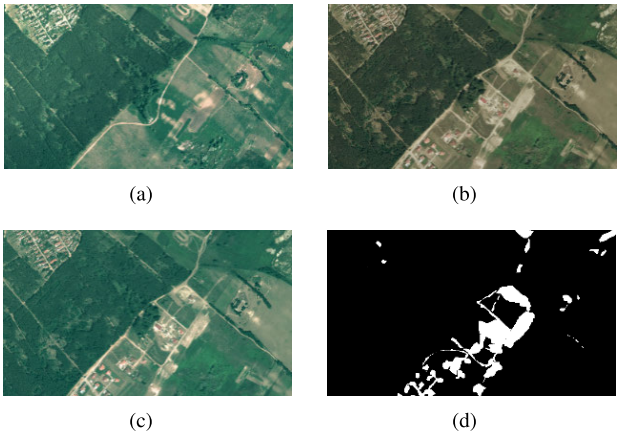


FIGURE 7. Optical aerial image pairs of SZADA/2 data set and radiation correction result. (a) Image acquired at t_1. (b) Image acquired at t_2. (c) Radiation correction result for image at t_2. (d) Ground truth.

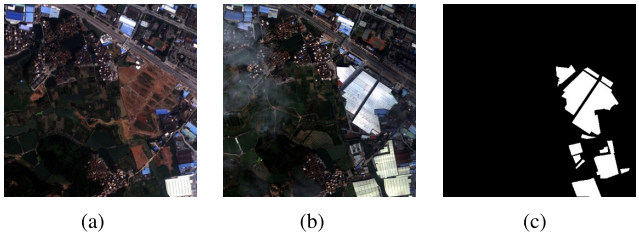


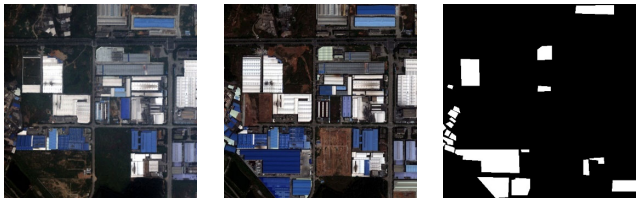
FIGURE 8. Building change detection data set 1. (a) Image acquired at 2015. (b) Image acquired at 2017. (c) Ground truth.

B. OPTIMIZATION AND MANAGEMENT OF TRAINING DETAILS

We choose Adam [21] with learning rate 0.0001 as an optimization algorithm, which can be seen as a modified Momentum+RMSProp algorithm, quite robust to the choice of hyper-parameters. Xavier initialization [22] is applied to initialize the weight of each layer of the network. The batch size of training data is set to 32, we train the network for 100 epochs until the overfitting occurs, where early-stopping plays an important role. Dropout [25] is also added to the network as a means of avoiding over-fitting by randomly zeroing the output of some nodes in the training phase. In this paper, dropout is set after each hidden layer, and about 20% of the output of nodes is randomly zeroed. What's more, we implement the proposed network based on the keras framework, and a single NVIDIA GTX 1070ti GPU with 8G memory is used for training and testing.

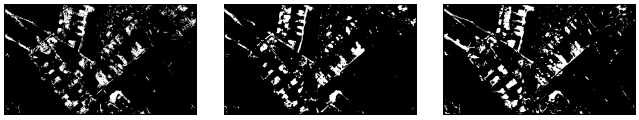
(a) (b) (c)

FIGURE 9. Building change detection data set 2. (a) Image acquired at 2015. (b) Image acquired at 2017. (c) Ground truth.



(a) (b) (c)

FIGURE 10. Building change detection data set 3. (a) Image acquired at 2015. (b) Image acquired at 2017. (c) Ground truth.



(a) (b) (c)

FIGURE 11. Experimental results by the proposed method and other methods on SZADA/1 data set. (a) Result by CNN. (b) Result by CNN-LSTM. (c) Result by our proposed method.

(a) (b) (c)

FIGURE 12. Experimental results by the proposed method and other methods on TISZADOB/3 data set. (a) Result by CNN. (b) Result by CNN-LSTM. (c) Result by our proposed method.

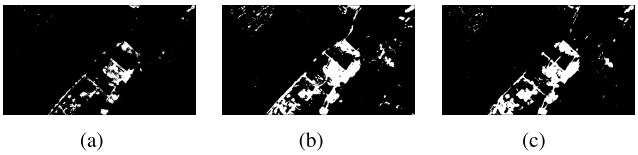
C. RESULTS AND EVALUATION

The precision(P), recall(R), and $F_1-score$ [23] are employed to quantitatively evaluate the performance of the proposed method, and the ROC curve [24] is analyzed graphically to evaluate the performance of the proposed method.

1) EXPERIMENTAL RESULTS OF SZTAKI AIRCHANGE BENCHMARK DATA SET

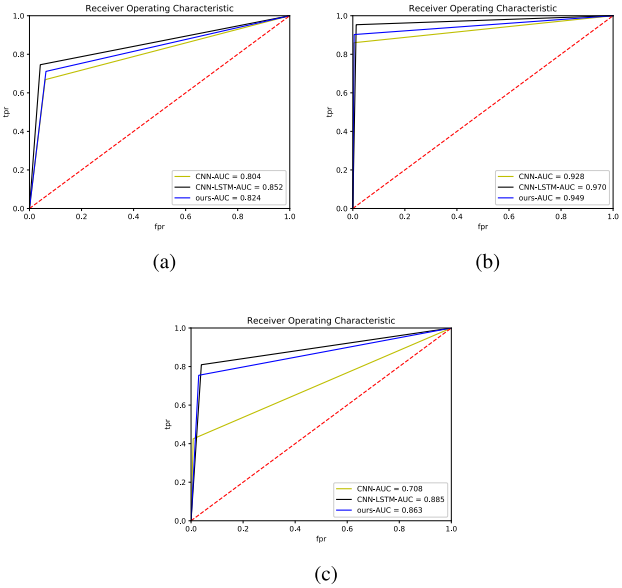
The experimental results on the SZTAKI AirChange Benchmark data set are shown in Figs. 11-14 and Table 1.

Among them, Fig. 11 shows the detection result of the CNN method has the most noise points, and the white area is not smooth. Compared with the other two methods, the detection result is rough. The main reason is that the CNN method only extracts the spectral-spatial features of the image for classification, but it does not consider the time dependence



(a) (b) (c)

FIGURE 13. Experimental results by the proposed method and other methods on SZADA/2 data set. (a) Result by CNN. (b) Result by CNN-LSTM. (c) Result by our proposed method.



(a) (b) (c)

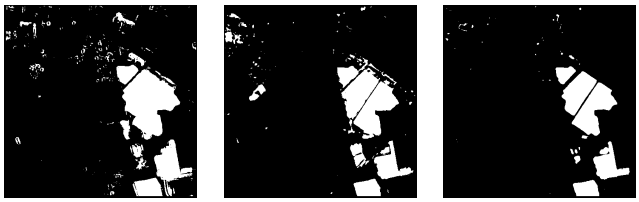
FIGURE 14. ROC curves of the three different methods for the SZTAKI AirChange Benchmark set. (a) ROC curves for SZADA/1 data set. (b) ROC curves for TISZADOB/3 data set. (c) ROC curves for SZADA/2 data set.

TABLE 1. Quantitative comparison among different methods on the SZTAKI AirChange Benchmark set.

Datasets	Metrics	CNN	CNN-LSTM	ours
SZADA/1	Precision(%)	40.6	52.6	41.0
	Recall(%)	66.8	74.6	71.1
	$F_1-score(\%)$	50.5	61.7	52.0
TISZADOB/3	Precision(%)	97.8	93.8	97.3
	Recall(%)	86.0	95.3	90.2
	$F_1-score(\%)$	91.5	94.5	93.6
SZADA/2	Precision(%)	79.1	60.1	65.4
	Recall(%)	42.5	80.9	75.5
	$F_1-score(\%)$	55.3	69.0	70.1

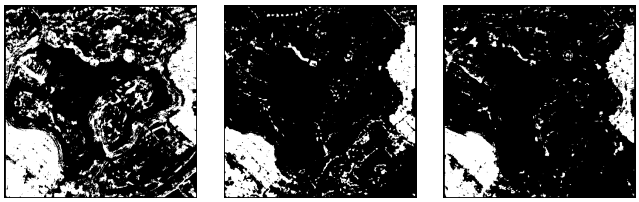
of the bi-temporal image, so it is sensitive to image noise, which leads to the occurrence of misjudgment. By comparing the change detection results of the CNN-LSTM method and the proposed method in Fig. 11, it is difficult to distinguish the detection results of the two methods visually. However, through the quantitative evaluation in Table 1, the performance of the CNN-LSTM method is better than that of the proposed method, but the detection accuracy of this sub-dataset is low on the whole.

Fig. 12 shows CNN method and our proposed method have many black holes, and there are many missed detection points, while the change area detected by CNN-LSTM



(a) (b) (c)

FIGURE 15. Experimental results by the proposed method and other methods on building change detection data set 1. (a) Result by CNN. (b) Result by CNN-LSTM. (c) Result by our proposed method.

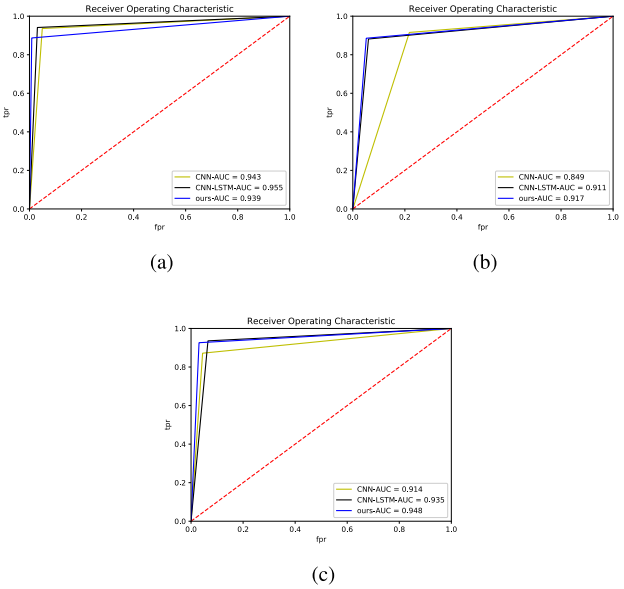


(a) (b) (c)

FIGURE 16. Experimental results by the proposed method and other methods on building change detection data set 2. (a) Result by CNN. (b) Result by CNN-LSTM. (c) Result by our proposed method.

(a) (b) (c)

FIGURE 17. Experimental results by the proposed method and other methods on building change detection data set 2. (a) Result by CNN. (b) Result by CNN-LSTM. (c) Result by our proposed method.



(a) (b) (c)

FIGURE 18. ROC curves of the three different methods for the QuickBird data set. (a) ROC curves for data set 1. (b) ROC curves for data set 2. (c) ROC curves for data set 3.

TABLE 2. Quantitative comparison among different methods on the QuickBird data set.

Datasets	Metrics	CNN	CNN-LSTM	ours
Dataset 1	Precision(%)	65.6	75.6	91.0
	Recall(%)	93.6	94.1	88.7
	$F_1 - score(\%)$	77.2	83.9	89.9
Dataset 2	Precision(%)	36.1	66.2	69.6
	Recall(%)	91.6	88.2	88.6
	$F_1 - score(\%)$	51.7	75.6	77.9
Dataset 3	Precision(%)	68.4	61.3	76.9
	Recall(%)	87.2	93.6	92.6
	$F_1 - score(\%)$	76.6	74.1	84.0

method is relatively smooth. The evaluation indicators in Table 1 are also superior to the other two methods.

Fig. 13 shows the detection result of CNN method are obviously worse than those of the other two methods, while the CNN-LSTM method has multi-detection phenomenon compared with our proposed method. In Table 1, the proposed method is superior to the two comparison methods in terms of comprehensive evaluation indicator $F_1 - score$.

In addition, Fig. 14 shows the ROC curve of the three methods on the SZTAKI AirChange Benchmark data set. It can be seen that the ROC curve of the CNN-LSTM method is close to the upper left corner, so the detection result is the best. Generally speaking, in the SZTAKI AirChange Benchmark data set, the detection performance of CNN-LSTM is similar to that of our proposed method, and slightly better than that of our proposed method. The performance of CNN method is the worst.

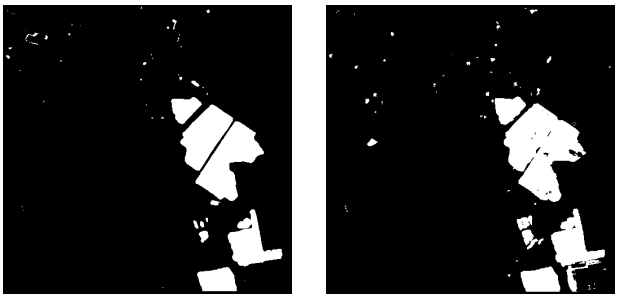
2) EXPERIMENTAL RESULTS OF QUICKBIRD DATA SET

The experimental results on the QuickBird data set are shown in Figs. 15-18 and Table 2.

In Fig. 15, the white noise point of the CNN method is the most, and there is a serious multi-detection phenomenon, and the details of the image are not well preserved. The CNN-LSTM method has obvious false alarm phenomenon in the upper left corner of the detection results compared with our proposed method. The change area detected by our proposed method is relatively smooth, mainly because the information transmission module and the attention mechanism are added when designing the network model, which significantly improves the detection performance. The proposed method is also much higher than the two comparison methods on the term of $F_1 - score$ in Table 2.

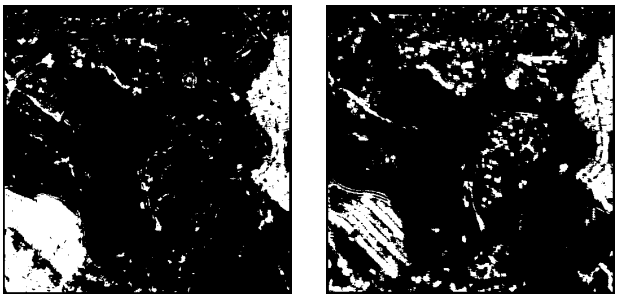
As shown in Fig. 16, the surface condition of the data set is very complex and the detection results of the three methods all have serious white noise. The proposed method is visually superior to the other two methods.

Fig. 17 shows all the three methods have white bar areas, leading to the occurrence of multi-detection. Compared with the reference image, the noise point of the proposed method is much less than the other two methods. The $F_1 - score$ of the proposed method in Table 2 is much higher than that of CNN and CNN-LSTM.



(a) (b)

FIGURE 19. Change detection results of building change detection data set 1 achieved by (a) Remain attention mechanism. (b) Remove attention mechanism.



(a) (b)

FIGURE 20. Change detection results of building change detection data set 2 achieved by (a) Remain attention mechanism. (b) Remove attention mechanism.

As shown in Fig. 18, we can see that the classification effect of the proposed method is better by observing the ROC curves of the three methods on the QuickBird data set. Overall, our proposed method has the best detection results on the QuickBird data set, followed by the CNN-LSTM method, and the CNN method has the worst effect.

3) EXPERIMENTAL RESULTS ON WHETHER ATTENTION MECHANISM EXISTS

In order to prove the validity of attention mechanism, we have carried out relevant experiments, and compared the results of retaining attention mechanism and eliminating attention mechanism. The experimental results show that attention mechanism is effective.

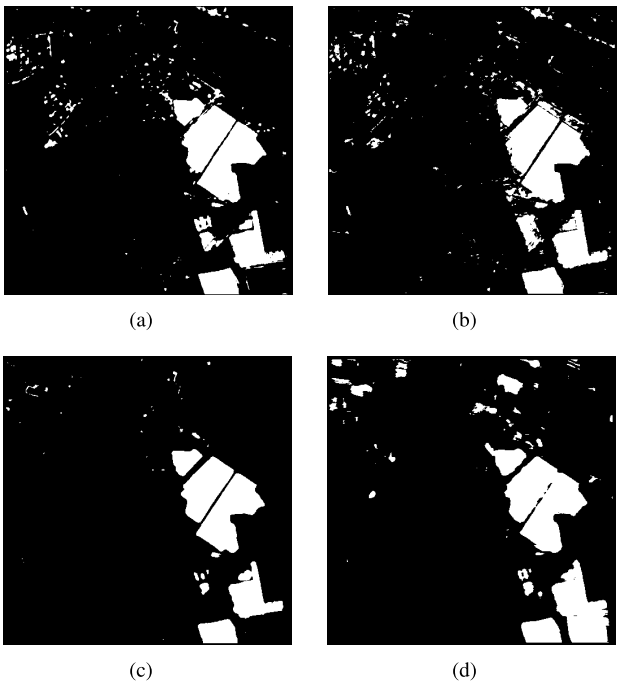
As shown in Fig. 19, compared with our proposed algorithm, images without attention mechanism do not perform well in noise suppression and image details preservation. In Fig. 20, white speckle noise is almost as much as the real part and it is obviously that attention mechanism suppresses noise. In Fig. 21, there are some changes that can not be detected without retained attention mechanism, and the white noise is more than our proposed algorithm. The proposed method is also much higher on the term of $F_1 - score$ in Table 3. Considering comprehensively, attention mechanism is effective in preserving image details, suppressing noise and detecting changing areas.

(a) (b)

FIGURE 21. Change detection results of building change detection data set 3 achieved by (a) Remain attention mechanism. (b) Remove attention mechanism.

TABLE 3. The influence of different patches on the experimental results.

Datasets	Metrics	Remain attention mechanism	Remove attention mechanism
Dataset 1	Precision(%)	91.0	82.5
	Recall(%)	88.7	88.3
	$F_1 - score(\%)$	89.9	85.3
Dataset 2	Precision(%)	69.9	52.7
	Recall(%)	88.6	64.1
	$F_1 - score(\%)$	77.9	57.8
Dataset 3	Precision(%)	76.9	63.6
	Recall(%)	92.6	89.6
	$F_1 - score(\%)$	84.0	74.4



(a) (b) (c) (d)

FIGURE 22. Change detection experimental results on QuickBird Dataset1. (a)5 × 5. (b)7 × 7. (c)9 × 9. (d)11 × 11.

4) EXPERIMENTAL RESULTS OF THE DIFFERENT PATCHES

Image patch is a very important parameter in our algorithm. In the experiment, in order to illustrate the effect of different image patches on the experimental results. We guaranteed that the network parameters were the same as previous

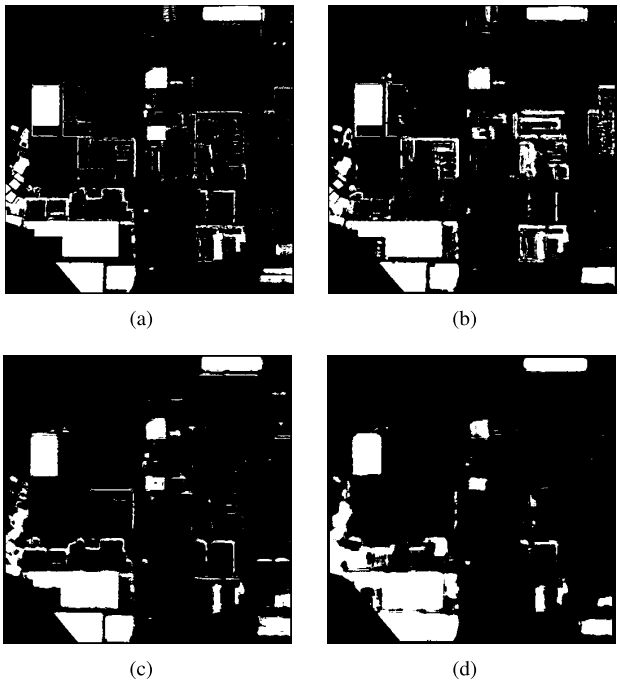


FIGURE 23. Change detection experimental results on QuickBird Dataset3. (a) 5 × 5. (b) 7 × 7. (c) 9 × 9. (d) 11 × 11.

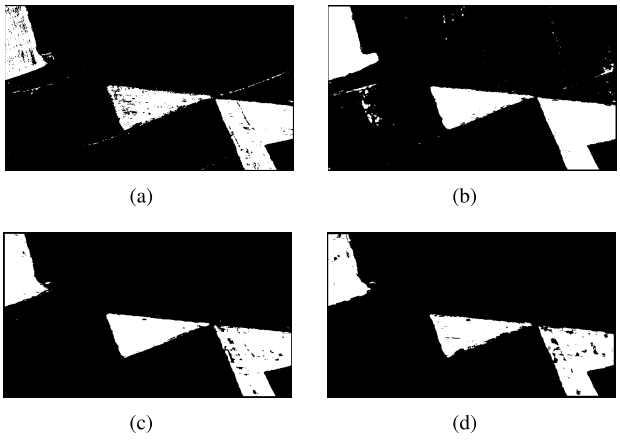


FIGURE 24. Change detection experimental results on TISZADOB/3. (a)5 × 5. (b) 7 × 7. (c) 9 × 9. (d) 11 × 11.

experiment, and sent the different image patches of 5, 7, 9 and 11 respectively. The experimental results are shown in Figs.22-25 and Table 4.

Fig.23 shows the change detection experimental results on QuickBird Dataset3.It can be seen that the image patch of the 5×5 and 7×7 both have serious white noise. For 11×11, although it has good resistance to noise, it is not good for image detail protection. The advantage of 9×9 image patch in comparison with 5×5 and 7×7 is that it performs well in anti-noise. While compared with 11×11, it protects the details of the image well.

As shown in Fig.25, the image patch of 9×9 is superior to the other three in noise resistance and image detail preservation. Considering comprehensively, we can conclude that the

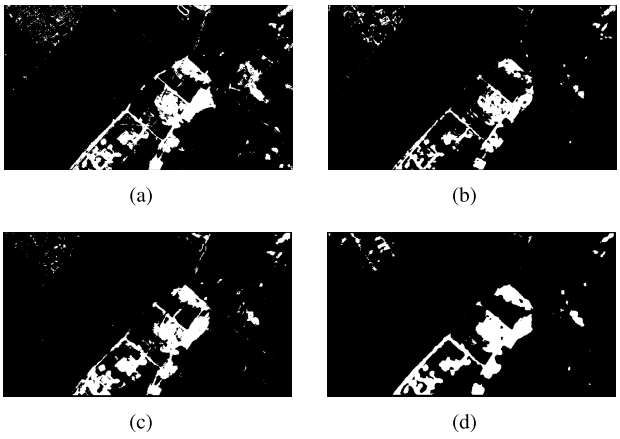


FIGURE 25. Change detection experimental results on SZADA/2. (a) 5 × 5. (b) 7 × 7. (c) 9 × 9. (d) 11 × 11.

TABLE 4. The influence of different patches on the experimental results.

Datasets	Metrics	5×5	7×7	9×9	11×11
TISZADOB/3	Precision(%)	96.2	93.8	97.3	98.5
	Recall(%)	88.4	95.3	90.2	85.7
	$F_1 - score(\%)$	92.1	94.5	93.6	91.6
SZADA/2	Precision(%)	55.7	67.7	65.4	63.3
	Recall(%)	76.0	63.1	75.5	73.9
	$F_1 - score(\%)$	64.3	65.3	70.1	68.2
QuickBird Dataset1	Precision(%)	78.3	76.6	91.0	66.7
	Recall(%)	92.2	95.5	88.7	89.3
	$F_1 - score(\%)$	84.7	85.0	89.9	76.3
QuickBird Dataset3	Precision(%)	75.4	66.1	76.9	68.3
	Recall(%)	93.1	94.0	92.6	90.7
	$F_1 - score(\%)$	83.3	77.6	84.0	78.0

image patch of 9×9 is better than the other three in noise resistance and image detail protection.

$F_1 - score$ as our main evaluation indicators, we can see that when the image patch is 9×9, three of the four data sets have the best results. Although there is one result that is not the best, it is not much different from the best one. Finally, according to the Figs.22-25 and Table 4, we choose the image patch of 9×9.

V. CONCLUSION

In this paper, we present a novel neural network architecture, which introduces information transmission and attention mechanism.Information transmission module can extract joint spectral-spatial-temporal features from bi-temporal multispectral images, which makes full use of the rich semantic information. In addition, in order to further improve the image change information, we add the attention mechanism to the design of the model framework. Moreover, it is end-to-end trainable. All these properties make it an excellent approach for multi-temporal remote sensing data analysis. The experimental results show that the proposed method can compete with the state-of-the-art methods and even performs better. Our future works will focus on new mechanisms for training adaptation, such as semi-supervised training for

training proposed models, usually with a small amount of labeled data with a large amount of unlabeled data.

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] A. A. Aly, A. M. Al-Omran, A. S. Sallam, M. I. Al-Wabel, and M. S. Al-Shayaa, "Vegetation cover change detection and assessment in arid environment using multi-temporal remote sensing images and ecosystem management approach," *Solid Earth*, vol. 7, no. 2, pp. 713–725, 2016.

[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] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogramm. Remote Sens.*, vol. 80, pp. 91–106, Jun. 2013.

[5] G. Chen, G. J. Hay, L. M. T. Carvalho, and M. A. Wulder, "Object-based change detection," *Int. J. Remote Sens.*, vol. 33, no. 14, pp. 4434–4457, 2012.

[6] Q. N. Zhao, J. Ma, M. Gong, H. Li, and T. Zhan, "Three-class change detection in synthetic aperture radar images based on deep belief network," *J. Comput. Theor. Nanosci.*, vol. 13, no. 6, pp. 3757–3762, Jun. 2016.

[7] H. Zhuang, H. Fan, K. Deng, and Y. Yu, "An improved neighborhood-based ratio approach for change detection in SAR images," *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 723–738, 2018.

[8] E. F. Lambin and A. H. Strahlers, "Change-vector analysis in multitemporal space: A tool to detect and categorize land-cover change processes using high temporal-resolution satellite data," *Remote Sens. Environ.*, vol. 48, no. 2, pp. 231–244, May 1994.

[9] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognit.*, vol. 19, no. 1, pp. 41–47, 1986.

[10] S. Ghosh, L. Bruzzone, S. Patra, F. Bovolo, and A. Ghosh, "A context-sensitive technique for unsupervised change detection based on Hopfield-type neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 778–789, Mar. 2007.

[11] X. L. Dai and S. Khorram, "Remotely sensed change detection based on artificial neural networks," *Photogram. Eng. Remote Sens.*, vol. 65, no. 10, pp. 1187–1194, 1999.

[12] M. Gong, T. Zhan, P. Zhang, and Q. Miao, "Superpixel-based difference representation learning for change detection in multispectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2658–2673, May 2017.

[13] 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.

[14] H. Lyu and H. Lu, "Learning a transferable change detection method by recurrent neural network," in *Proc. Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 5157–5160.

[15] 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.

[16] B. Plank, A. Søgaard, and Y. Goldberg, "Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss," 2014, *arXiv:1604.05529*. [Online]. Available: https://arxiv.org/abs/1604.05529

[17] A. Graves, "Generating sequences with recurrent neural networks," 2013, *arXiv:1308.0850*. [Online]. Available: https://arxiv.org/abs/1308.0850

[18] W. Yin, H. Schütze, B. Xiang, and B. Zhou, "ABCNN: Attention-based convolutional neural network for modeling sentence pairs," 2015, *arXiv:1512.05193*. [Online]. Available: https://arxiv.org/abs/1512.05193

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*. [Online]. Available: https://arxiv.org/abs/1512.03385

[20] C. Benedek and T. Szirányi, "Change detection in optical aerial images by a multilayer conditional mixed Markov model," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 10, pp. 3416–3430, Oct. 2009.

[21] D. P. Kingma and J. L. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, 2015, pp. 1–13.

[22] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *J. Mach. Learn. Res.*, vol. 9, pp. 249–256, May 2010.

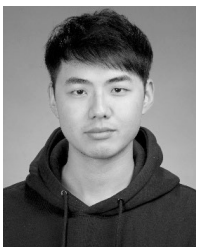
[23] N. Chinchor and B. Sundheim, "MUC-5 Evaluation metrics," in *Proc. 5th Message Understand. Conf. (MUC) Conf. Held Baltimore*, Jun. 1992, pp. 22–29.

[24] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

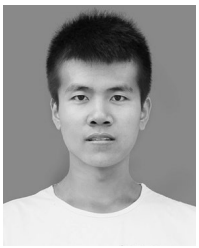
[25] A. Agarwal, S. Negahban, and M. J. Wainwright, "A simple way to prevent neural networks from overfitting," *Ann. Stat.*, vol. 40, no. 2, pp. 1171–1197, 2012.

[26] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 1, Dec. 2012, pp. 1097–1105.

RUOCHEN LIU received the Ph.D. degree from Xidian University, Xi'an, China, in 2005, where she is currently a Professor with the Intelligent Information Processing Innovative Research Team of the Ministry of Education of China. Her research interests are broadly in the areas of computational intelligence. Her areas of special interests include evolutionary computation, data mining, and deep learning.



ZHIHONG CHENG received the B.S. degree from the Qingdao University of Technology, Shandong, China, in 2018. He is currently pursuing the M.S. degree with Xidian University. His current research focuses on deep learing and NLP.



LANGLANG ZHANG received the B.S. degree from Xidian University, Xi'an, China, in 2016, where he is currently pursuing the M.S. degree. His current research focuses on remote sensing image change detection.

JIANXIA LI received the B.S. from Xidian University, Xi'an, China, in 2015, where she is currently pursuing the Ph.D. degree. Her current research focuses on dynamic multiobjective optimization.