Deep Difference Representation Learning for Multi-spectral Imagery Change Detection

Hui Zhang[1, a], Puzhao Zhang[2, b]

[1]Department of Integrated Circuit Design and Integrated System,

School of Microelectronics, Xidian University, Xi'an, 710071, China

[2]Key Laboratory of Intelligent Perception and Image Understanding of Ministry of

Education of China, Xidian University, Xi'an, 710071, China

[a]email: zhsoxidian@163.com, [b]email:puzhaozhang@126.com

Abstract. Change detection is an ongoing hot topic in multi-spectral imagery applications, how to exploit the available spectral information effectively for change detection is still an open question. Considering the noise interference and redundancy of multi-spectral imagery, it is important and necessary to learn more abstract and robust feature from raw spectrums for change detection application. In this paper, a deep difference representation learning model is proposed for multi-spectral change detection. In this model, two stacked denoising autoencoders are established, one for learning more abstract features from raw spectrums blocks, and the other for learning difference representations from the stacked change feature. The former is used to weaken noise interference and reduce redundancy, while the latter has the ability to highlight changes and suppress unchanged pixels. The experimental results on real multi-spectral data demonstrate the feasibility, effectiveness and robustness of the proposed deep difference representation learning model on multi-spectral change detection task.

Introduction

Change detection is an ongoing hot research topic in remote sensing field, and its objective is to highlight the difference between two or more images acquired over the same area at different times. It has been widely used in many practical applications such as disaster management [1], urban planing [2], environment monitoring [3] and land-use transitions [4]. Though change detection on remote sensing imagery has been studied for several decades, multi-spectral image change detection remains an open problem, due to the inadequate utilization of spectral information and the limited ability of the conventional change detection methods on change extraction. In the literature, many change detection techniques have been developed over the last two decades, which can be divided into two categories, pre-classification spectral change detection and post-classification change detection [5]. In post-classification change detection, bi-temporal images are first independently classified and labelled, and then the changes are extracted through the direct comparison between these two classified maps. Post-classification change detection techniques shift the difficulties in change detection techniques rely on the principle that earth surface changes result in persistent changes in the spectral signature. These techniques transform two or more raw image into a new single-band or multi-bands one in which the changes are highlighted [4].

Most of the proposed change detection techniques are be- longs to spectral change identification category, such as image differencing, image ratioing, image stacking [6], vegetation index differencing, principal-component analysis (PCA) [4] and change vector analysis (CVA) [7] etc. The arithmetical operations such as differencing and ratioing are sensitive to the atmosphere, lighting and seasonal variation. Compared to other change detection techniques, image stacking is simpler but more promising in improving interclass difference and highlighting changes [6]. However, the distribution of change vectors built by stacking bi-temporal raw spectrums is

relatively complicated and hard to estimate. Fig. 1 shows the distribution of the change vectors generated by mapping the stacked raw spectrums into 2-D principal component space. It is necessary to learn a better difference representation from the stacked spectral vectors, which should have the ability to highlight main changes and suppress meaningless ones, and the learned difference representation would be promising for change detection task.
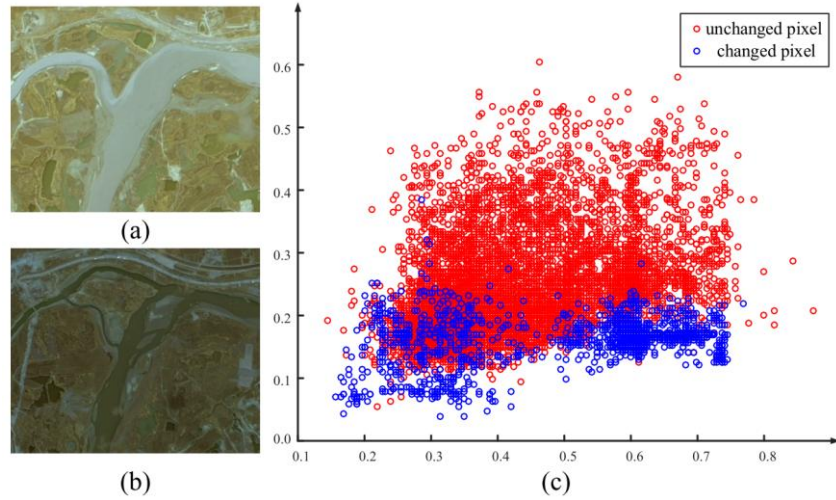

Fig. 1. Distribution of change feature generated by mapping the stacked bi-temporal raw spectrums into 2-D principal component space (10000 samples chosen in random). (a) Multi-spectral image at t_1. (b) Multi-spectral image at t_2. (c) Distribution of stacked raw spectrums. Red 'o' stands for the unchanged pixels and blue 'o' denotes the changed pixels.
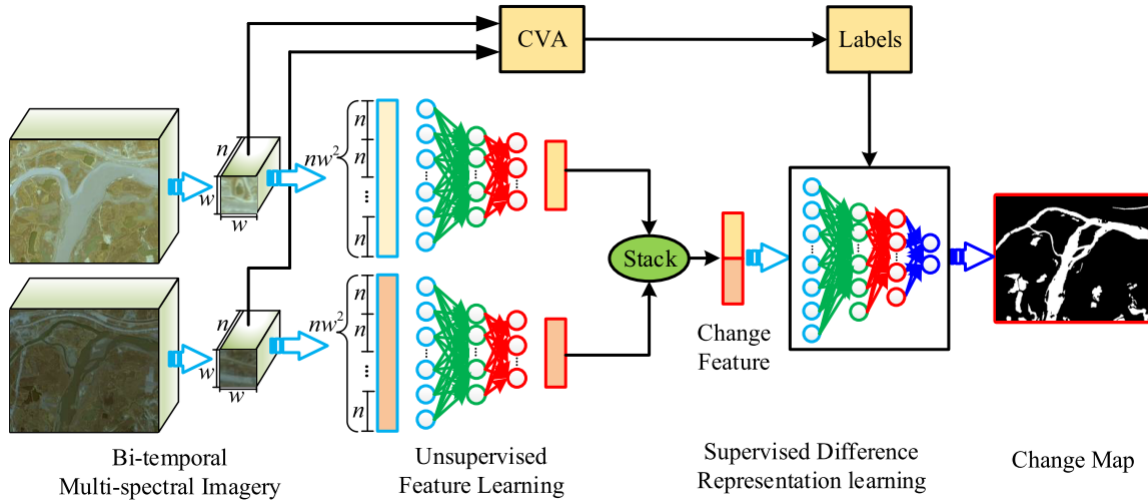

Fig. 2. Flow chart of the proposed difference representation learning model. Bi-temporal spectral blocks are fed into the first SDAE for unsupervised feature learning, and the features learned by the first SDAE are stacked to form a change feature. And then the change features and its corresponding labels (produced by CVA on raw spectrums) are fed into the second SDAE for supervised difference representation learning. After training phrase, the testing samples are fed into this system, and it outputs the final change maps.

Recently, deep learning has become a hot research point, which has the ability to treat with raw data and automatically discover the representations desired for specific application such as image classification and change detection. For change detection tasks, higher-level difference representation has the ability to highlight main changes by suppressing meaningless variations or noise interference. In this paper, we proposed a deep difference representation learning model for

multi- spectral change detection task, in which two stacked denoising autoencoders (SDAE) [8] are established, one for feature representation learning from the raw spectrums, and the other for difference representation learning from the stacked change features. In difference representation learning, the targets are obtained by using CVA on the raw bi-temporal multi-spectral images, and then the second SDAE are trained on the randomly selected samples (each sample contains difference representations and labels). Finally, the testing samples are fed into this deep model to generate the corresponding labels and the final change map can be obtained.

The rest of this paper is organized as follows. Section II introduces SDAE and the proposed difference representation learning model in details. Section III presents the experimental results and analysis. At last, Section IV draws the conclusions of this paper.
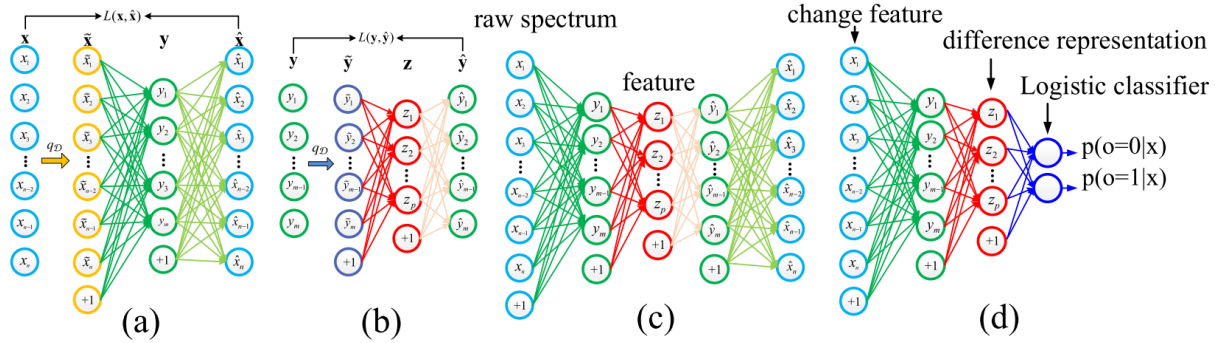

Fig. 3. Illustration on DAE, FLN and DLCN. (a) The first DAE with structure n-m. (b) The second DAE with structure m-p. (c) FLN (n-m-p): Spectral block feature learning network. (d) DLCN (n-m-p-2): Difference representation learning and classification network.

Proposed Method

Denoising Autoencoder. A traditional autoencoder consists of an encoder and a decoder. The encoder transforms an input vector into a higher-level representation, and the decoder can map the representation back to the input data. When a denoising function is imposed in the traditional autoencoder, then this autoencoder becomes denoising autoencoder (DAE) [8]. Fig. 3 shows that two DAEs are stacked into SDAEs (FLN and DLCN). As shown in Fig. 3(a), firstly, the raw input data $\mathbf{x}$ is corrupted into $\tilde{\mathbf{x}}$ through a stochastic mapping $\tilde{\mathbf{x}} \sim q_D(\tilde{\mathbf{x}} | \mathbf{x})$. And then, the corrupted input $\tilde{\mathbf{x}}$ is mapped into the hidden representation through

$$\mathbf{y} = f_\theta(\tilde{\mathbf{x}}) = \text{sig}(\mathbf{W}\tilde{\mathbf{x}} + \mathbf{b}) \tag{1}$$

Its parameter set is $\theta = \{\mathbf{W}, \mathbf{b}\}$, where $\mathbf{W}$ is a $d' \times d$ weight matrix and $\mathbf{b}$ is a bias vector of dimension d'. In addition, $\text{sig}(x) = 1/(1 + \exp(-x))$ is called sigmoid function. Finally, the decoder seeks to map the resulting hidden representation $\mathbf{y}$ back to a reconstruction $\hat{\mathbf{x}}$, the estimation of the clear input $\mathbf{x}$, in its input space by

$$\hat{\mathbf{x}} = g_\theta(\mathbf{y}) = \text{sig}(\mathbf{W}'\tilde{\mathbf{x}} + \mathbf{b}') \tag{2}$$

Its parameter set is $\theta' = \{\mathbf{W}', \mathbf{b}'\}$, where $\mathbf{W}'$ is a $d \times d'$ weight matrix and $\mathbf{b}$ is a bias vector of dimension d. The parameters θ and θ' of the DAE can be obtained by minimizing the average reconstruction error as follows, making $\hat{\mathbf{x}}$ to be as close as possible to the uncorrupted input $\mathbf{x}$:

$$\theta^*, \theta'^* = \arg\min_{\theta, \theta'} \frac{1}{N} \sum_{i=1}^{N} L(\mathbf{x}^{(i)}, \hat{\mathbf{x}}^{(i)})$$
$$= \arg\min_{\theta, \theta'} \frac{1}{N} \sum_{i=1}^{N} L(\mathbf{x}^{(i)}, g_{\theta'}(f_\theta(\mathbf{x}^{(i)}))) \tag{3}$$

Where L can be a loss function $L(\mathbf{x}, \hat{\mathbf{x}}) = \| \mathbf{x} - \hat{\mathbf{x}} \|^2$, and $N = w \times h$ is the total number of training examples. The optimization of DAE is achieved by stochastic gradient descent (SGD). Like this, the

representation $\mathbf{y}$ is fed into next DAE for learning a higher-level representation $\mathbf{z}$. Finally, a fine-tuning procedure is used to improve the performance of the entire network, which treats all layers of DAEs as a single model.

Difference Representation Learning Model. Fig. 2 shows the flow chart of the proposed difference representation learning and classification model. In this model, two deep models (SDAEs) are established, one for unsupervised feature learning, and the other for supervised difference representation learning and classification. For convenience, the former deep network is called feature learning network (FLN) while the latter one is denoted as difference learning and classification network (DLCN), as shown in Fig. 3(c) and (d) respectively. FLN and DLCN have different functions due to their difference in three aspects: training samples, network structure and training mode. FLN is trained on raw spectrums blocks in an unsupervised way, while DLCN is trained on the stacked change features in a supervised way. Additionally, at the end of DLCN, a soft-max layer is added for supervised global fine-tuning, i.e., the soft-max layer is used for the classification of the learned difference representation. Though the difference learning procedure is supervised, the entire system is still unsupervised due to that the labels for supervised training are obtained in an unsupervised way. The coarse labels are obtained by applying CVA on the bi-temporal raw spectrums, it is worth noting that CVA is an unsupervised technique, which highlights changes and identifies the change types by analyzing the change magnitudes and directions of the considering bi-temporal spectral blocks [7]. Due to the limited accuracy of CVA technique, deep model such as SDAE is introduced to improve the change detection results.

In the proposed difference learning and classification framework, the bi-temporal corresponding blocks (of size $w \times w \times n$) are unfolded into two spectral vectors with dimension nw^2, respectively. Then these bi-temporal spectral vectors are mixed together, and a certain number of randomly selected training examples are fed to FLN for its optimization. Through DLN, the spectral vectors are transformed into feature space, in which the noise interference is suppressed and the redundancy is reduced. Next, the bi-temporal features are stacked into change features, and these change features and their corresponding labels (obtained by CVA) are fed into DLCN for its optimization. In the optimization of DLCN, layer-wise pre-training is adopted to initialize the entire deep network, and then a supervised fine-tuning is used to globally fine-tune this deep network for change detection tasks. After training phrase, the testing samples are input into this system, and the corresponding change detection results will be output.
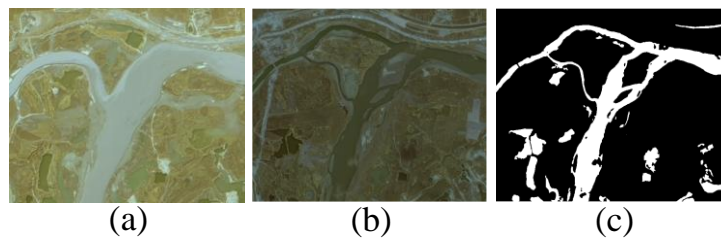

Fig. 4. Weihe river dataset. (a) Multi-spectral imagery (4-band) acquired by GF-1 satellite on Aug. 19, 2013. (b) Multi-spectral imagery (4-bands) acquired by GF-1 satellite on Aug. 29, 2015. (c) The ground truth map.

Experimental Results

Experimental Dataset, Evaluation and Settings. Fig.4 shows the Weihe river dataset used in our experiments, of size 718 592, in which two bi-temporal images are taken by GF-1 satellite on Aug. 19, 2013 and Aug. 29, 2015 respectively over the same scene in Xi'an city, China. Each image is generated by fusing PAN (2m) and multi-spectral (3.5m) images with 4-bands. The ground truth map is established by manual drawing according to true changes happened between them. In our experiments, false positive (FP), false negative (FN), the percentage of correct classification (PCC) and Kappa coefficient is used to quantitatively evaluate the performance of the proposed

method. Unsupervised and supervised fine- tuning can be selected for specific tasks. In our experiments, after many practical tests, w is set to be 3, the structure of FLN is set to be 36-100-20 and that of DLCN is set to be 40-100-4-2.

Results and Analysis. Fig. 5 shows the feature maps learned by FLN from raw spectrums blocks, the first row shows the feature maps of the image acquired on Aug. 19, 2013, while the second row shows those of the image acquired on Aug. 29, 2015. In Fig. 5, each column, such as (a), (b) and (c), represents the activation of bi-temporal multi-spectral images on a certain neuron, which gives us a chance to look into the function of each neuron. From these feature maps, we can find that these neurons seem to have the ability of objects detection based on raw spectrums, for example, the neuron (b) and (e) detect main river areas, while the neuron (d) detects the silt in the river way. The objects detection function of deep model would be very helpful for highlighting changes in change detection application.
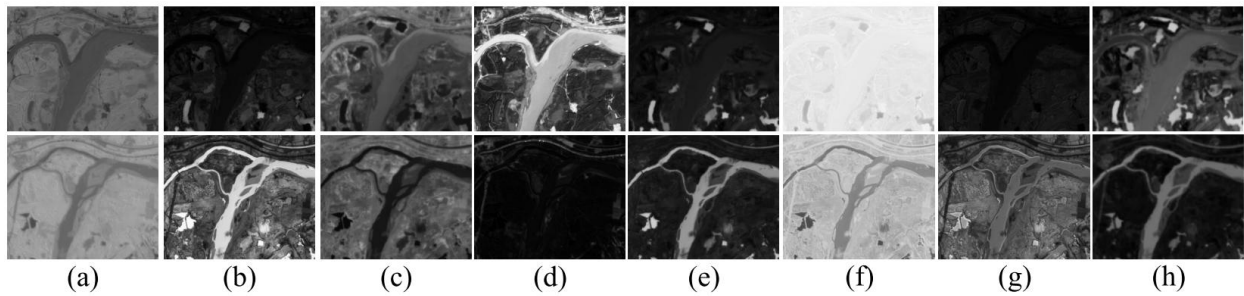

(a)　　(b)　　(c)　　(d)　　(e)　　(f)　　(g)　　(h)

Fig. 5.　Feature maps of bi-temporal images. The first row shows the feature maps of the image acquired on Aug. 19, 2013, while the second row shows the feature maps of the image acquired on Aug. 29, 2015. (a)-(h) shows the activation of bi-temporal images on neuron (a)-(h), respectively.

Fig. 6 shows the distributions of change vector between before and after supervised difference representation learning. As shown in Fig. 6(a), many blue circles appear among the red ones, which means that it is difficult to distinguish them clearly, but its distribution looks much better than that of raw spectrums, shown in Fig. 1. Fig. 6(b) shows the distribution of difference representation (DR), in which only several changed pixels appear among the unchanged ones, and the distribution looks interesting and easy to distinguish changed pixels from unchanged ones. This distribution comparison indicates the necessary of unsupervised feature learning and supervised difference representation learning.
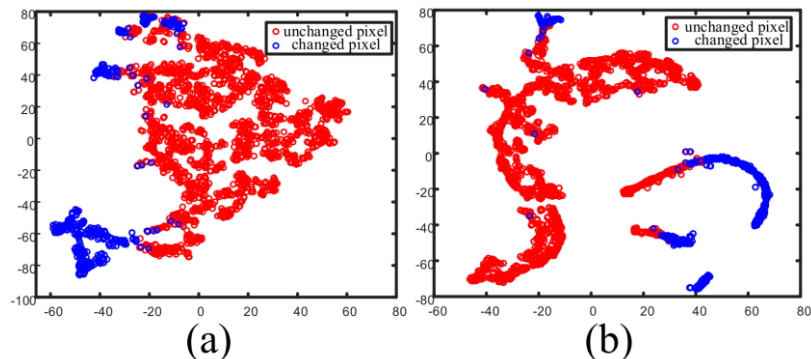

Fig. 6. Distribution comparison of change vectors before and after difference representation learning (DRL). (a) Before DRL. (b) After DRL.

Fig. 7 shows the feature maps of difference representation learned from change feature, the first row shows the difference representation maps learned by FLN in an unsupervised way, while the second row shows the difference representation maps learned by DLCN in a supervised fashion. Both rows of difference representation maps seem to highlight main changes and suppress

unchanged areas, and these maps look clear and less noisy, showing some merits of object-based change detection techniques. However, some unchanged areas are highlighted by FLN, but suppressed by DLCN, which indicates that supervised difference learning model can highlight changes more accurately by introducing knowledge, such as the initial labels generated by CVA techniques.
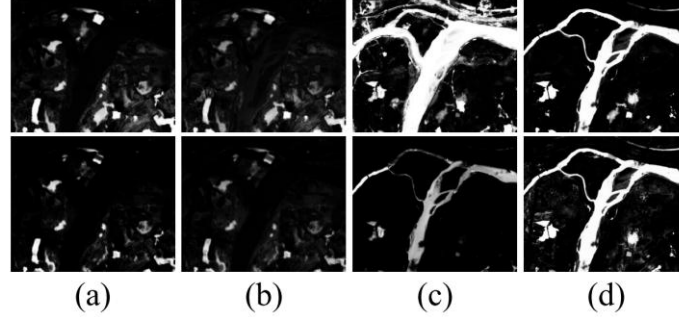

(a)   (b)   (c)   (d)

Fig. 7. Feature maps of difference representation learned from change features. The first row shows the difference representation maps learned in an unsupervised way, while the second row shows the difference representation maps learned in a supervised way. (a)-(d) shows the activation of change feature maps on the difference representation neuron (a)-(d), respectively.

Fig. 8 shows the change maps produced by different meth- ods, in which four state-of-the-art methods such as CVA [7], PCA [4], IR-MAD [9] and SDAE CVA [7] are treated as the comparison ones. Unsupervised CVA is an effective method for change detection, and the magnitude of the spectral change vectors is used to highlight changes in our experiments. Difference PCA is simple in implementation but effective in identifying meaningful changes. IR-MAD is a classic unsupervised method for change detection, and SDAE CVA means to apply CVA technique on the change feature stacked by the features learned by FLN from bi-temporal spectral blocks. DRL represents the proposed difference representation learning method.
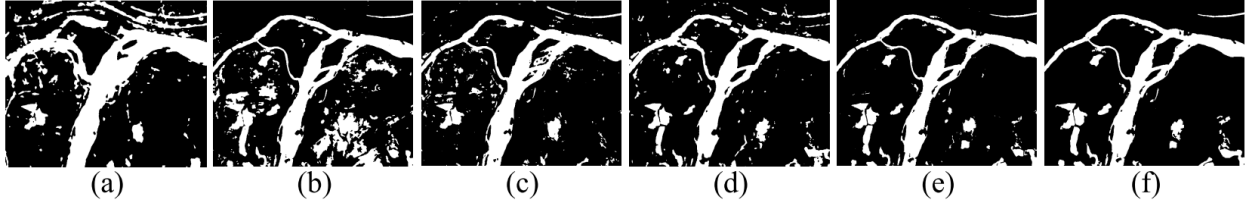

(a) (b) (c) (d) (e) (f)

Fig. 8 shows the change maps produced by different methods. (a) CVA. (b) PCA. (c) IR-MAD. (d) SDAE_CVA. (e) DRL. (f) The ground truth.

TABLE I. QUANTITATIVE COMPARISON ON CHANGE MAPS

methods	FP	FN	PCC(%)	Kappa
CVA	56839	14961	83.10	0.5186
PCA	33151	**1086**	91.94	0.7602
IR-MAD	18315	12853	92.66	0.7514
SDAE_CVA	16807	9815	93.73	0.7893
DRL(ours)	**2684**	5216	**98.13**	**0.9342**

As shown in Fig. 8 (a) and (c), CVA and IR-MAD highlight some unchanged regions but miss some main changes. PCA detects main changes but the result looks noisy and exists many FP pixels. SDAE CVA performs much better than CVA, which indicates the importance and necessary of SDAE-based unsupervised feature learning. Compared with the ground truth maps, DRL achieves the best visual results. Table I summarizes the FP, FP, FN, PCC and Kappa values, showing the quantitative comparison among the state-of-the-art change detection methods. Compared with CVA, PCA, IR-MAD and SDAE CVA, DRL achieves the best accuracy, with higher PCC and Kappa,

lower FP and FN values. Both the visual and quantitative comparisons indicate the feasibility, effectiveness and robustness of the proposed method. FLN-based unsupervised feature learning weakens the noise interference and reduces the redundancy of multi-spectral imagery, while DLCN-based supervised difference learning highlights main changes and suppresses the unchanged regions, producing a robust change map. The proposed method also shows up some merits of object-based change detection techniques, such as good change structure, region continuity and less noise

Conclusion

In this paper, we propose a deep difference representation learning model for multi-spectral imagery change detection, in which two SDAEs are established. one is used for feature learning, weakening noise interference and reducing redundancy of multi-spectral data, and the other for difference representation learning to highlight changes and suppress unchanged pixels. The proposed deep difference learning model has the ability to automatically learn a robust difference representation from the stacked change features for change detection task. The experimental results also have demonstrated the feasibility, effectiveness and robustness of the proposed deep difference learning model on multi-spectral imagery change detection task. In the near future, multiple types of changes will be considered to detect from multi-spectral imagery.

References

[1] B. Gokaraju, A. C. Turlapaty, D. A. Doss, et al. Change detection analysis of tornado disaster using conditional copulas and Data Fusion for cost-effective disaster management[C]. 2015 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). IEEE, 2015.

[2] Yuan Y, Meng Y, Lin L, et al. Continuous Change Detection and Classification Using Hidden Markov Model: A Case Study for Monitoring Urban Encroachment onto Farmland in Beijing[J]. Remote Sensing, 2015, 7(11):15318-15339.

[3] Sakieh Y, Gholipour M, Salmanmahiny A. An integrated spectral-textural approach for environmental change monitoring and assessment: analyzing the dynamics of green covers in a highly developing region[J]. Environmental Monitoring & Assessment, 2016, 188(4):1-19.

[4] J. S. Deng, K. Wang, Y. H. Deng, et al. PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data[J]. International Journal of Remote Sensing, 2008, 29(16):4823-4838.

[5] Ye, Su, Chen, Dongmei, Yu, Jie. A targeted change-detection procedure by combining change vector analysis and post-classification approach[J]. ISPRS Journal of Photogrammetry & Remote Sensing, 2016, 114:115-124.

[6] Ding K, Huo C, Xu Y, et al. Sparse Hierarchical Clustering for VHR Image Change Detection[J]. IEEE Geoscience & Remote Sensing Letters, 2015, 12(3):577-581.

[7] Bovolo F, Marchesi S, Bruzzone L. A Framework for Automatic and Unsupervised Detection of Multiple Changes in Multitemporal Images[J]. IEEE Transactions on Geoscience & Remote Sensing, 2012, 50(50):2196-2212.

[8] Vincent P, Larochelle H, Lajoie I, et al. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion[J]. Journal of Machine Learning Research, 2010, 11(6):3371-3408.

[9] Nielsen A A. The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data.[J]. IEEE Transactions on Image Processing, 2007, 16(2):463-78.