Change Detection Based on Deep Siamese Convolutional Network for Optical Aerial Images

Yang Zhan, *Student Member, IEEE*, Kun Fu, Menglong Yan, Xian Sun, Hongqi Wang, and Xiaosong Qiu

***Abstract*— In this letter, we propose a novel supervised change detection method based on a deep siamese convolutional network for optical aerial images. We train a siamese convolutional network using the weighted contrastive loss. The novelty of the method is that the siamese network is learned to extract features directly from the image pairs. Compared with hand-crafted features used by the conventional change detection method, the extracted features are more abstract and robust. Furthermore, because of the advantage of the weighted contrastive loss function, the features have a unique property: the feature vectors of the changed pixel pair are far away from each other, while the ones of the unchanged pixel pair are close. Therefore, we use the distance of the feature vectors to detect changes between the image pair. Simple threshold segmentation on the distance map can even obtain good performance. For improvement, we use a k-nearest neighbor approach to update the initial result. Experimental results show that the proposed method produces results comparable, even better, with the two state-of-the-art methods in terms of F-measure.**

***Index Terms*— Change detection, deep convolutional network, optical aerial images, siamese network.**

I. Introduction

AERIAL images nowadays can be easily taken by many aerial platforms, especially with the development of unmanned aerial vehicles. Automatic change detection in these images is an important field of remote sensing, because manual evaluation is time-consuming, tedious, and even boring. This letter focuses on processing the aerial images, which were taken in different seasons and lighting conditions, to extract the change regions.

In the literature, change detection algorithms can be summarized into two categories. First, they may follow postclassification comparison [1], which classify the input images separately into many predefined classes and then obtain the change regions by comparing the labels pixel by pixel. Second, the other kind of methods is called postcomparison analysis [2]–[7]. These methods all follow the procedures: first deriving a similarity-feature map [e.g., a difference image (DI)] from the multitemporal images and then analyzing the feature map to segment changed and unchanged areas. Image arithmetical operations (e.g., differencing [2] and ratioing [3]) and image transformation [7] can be used to obtain the feature map. The algorithms to analyze the feature map include thresholding, clustering [2], and other advanced methods, such as an extreme learning machine (ELM) [4] and an Markov random field [5], [6]. In [4], ELM is applied as the classification model for change detection, which is trained using some labeled samples obtained by a preclassification schema. Benedek and Szirányi [5] proposed a conditional multilayer mixed MRF (CXM) model for change detection, which is a three-layered MRF using two weak features: global intensity co-occurrence statistics and block correlation. In [6], a multilayer MRF model (L_3MRF) is proposed. The L_3MRF model uses the similar structure to the CXM [5] but different features: the modified histogram of oriented gradients and gray-level difference features.

As described earlier, the features used by conventional change detection algorithms are almost hand-crafted, which are weak in image representation. Recently, deep neural networks (DNNs) are learned to extract features directly from the input images. The features are more abstract and robust. In literature, some change detection methods [3], [7], [8] using DNN have been proposed and demonstrate good performance. In [3], a DNN based on the restricted Boltzmann machine is learned to classify the DI for synthetic aperture radar (SAR) images. Gao *et al.* [8] design a preclassification schema to obtain some labeled samples of high accuracy and then use these samples to train PCANet as the classification model for change detection. In [7], a symmetric convolutional coupling network (SCCN) is applied to detect changes between optical and SAR images. The method extracts features first and then transforms the features into a consistent feature space, where a different map can be obtained. The network is learned by optimizing a coupling function, which is only defined on the point of view of the unchanged pixels. In [7], the method demonstrated superiority over several existing approaches, but the limitation is that the method only considers the unchanged pixels.

The methods described earlier can also be distinguished to be unsupervised [2]–[4], [7], [8] or supervised [1], [5], [6]. Unsupervised methods do not use the ground truth (GT) data. Thus, they usually rely on some prior assumptions, such as that unchanged regions should demonstrate a smaller pixelwise difference [7]. However, the feature

Manuscript received May 27, 2017; revised July 17, 2017; accepted August 6, 2017. Date of publication August 30, 2017; date of current version September 25, 2017. This work was supported by the National Natural Science Foundation of China under Grant 41301493. *(Corresponding author: Kun Fu.)*

Y. Zhan is with the Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China, and also with the University of Chinese Academy of Sciences, Beijing 100190, China.

K. Fu, M. Yan, X. Sun, and H. Wang are with the Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China (e-mail: fukun@mail.ie.ac.cn).

X. Qiu is with the State Key Laboratory of Transducer Technology, Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China, and also with the University of Chinese Academy of Sciences, Beijing 100190, China.

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

Digital Object Identifier 10.1109/LGRS.2017.2738149

statistics in optical images may be multimodal and strongly overlapping [5]; therefore, it may be challenging for the unsupervised methods in some situation. On the other hand, if training data with GT data are available, they can provide significant additional information for classification. Thus a supervised method will be introduced here.

In this letter, we propose a novel model for change detection in optical aerial images, which is based on the supervised deep siamese convolutional neural network (CNN) [9]. Our main contributions are summarized as follows.

1) Welearn a siamese CNN to extract features directly from the images pixel by pixel. The extracted features, which are suitable for change detection, demonstrate a unique property: the feature vectors associated with changed pixel pairs are far away from each other in the feature space, whereas the ones of unchanged pixel pairs are close.
2) In order to distinguish the changed and unchanged pixels more effectively and reduce the influence of imbalance data (i.e., the numbers of changed and unchanged pixels vary greatly) in change detection, we use a weighted contrastive loss [10], in which not only the unchanged pixels but also the changed ones are considered as the objective function when training the network.

The rest of this letter is organized as follows. The proposed method is described in Section II. The experimental results and discussions are presented in Section III. Finally, the conclusion of this letter is drawn in Section IV.

II. Proposed Method

A. Siamese Convolutional Neural Network

In the proposed method, the siamese CNN is used for feature extraction. Because the conventional CNN takes one image patch as input, we use a siamese version of CNN to extract features from image pairs.

In [9], in order to compare image patches, three types of CNN models are presented: siamese, pseudosiamese, and 2-channel. Both the siamese and pseudosiamese models have two branches with the same architectures. Each branch takes one of the two patches as input. The difference is that the two branches in the siamese model share the same set of weights, while in the pseudosiamese model do not. In the 2-channel network, the two patches are considered as a 2-channel image, which is directly fed to the network. In fact, all of the above-mentioned three models can be adopted in our feature extraction task. However, the siamese one is chosen because of the following consideration. The siamese network has better explanation in the feature extraction task. The two branches of the siamese network share the same weights, which means that they extract features from the two patches using the same approach. As the two input images we used for change detection are both optical images, in other word homogeneous, which are captured by the same type of sensor (both optical sensors) and have similar characteristics, it is natural to extract features in the same way.

Fig. 1 shows the designed siamese CNN model. In the network, each branch follows CNN's architectures while there

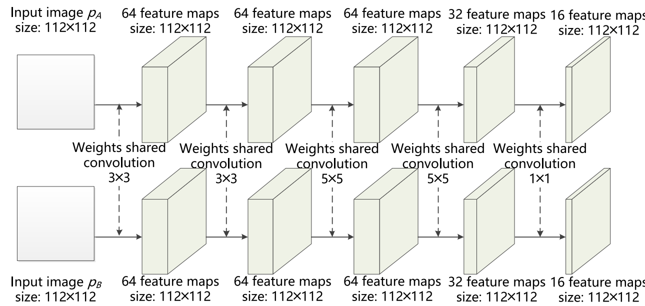

Fig. 1. Illustration of the designed siamese CNN model.

are also some differences from the conventional CNN. In the conventional CNN, there exist three types of layers: convolutional, pooling, and fully connected layers. The convolutional layers can extract the hierarchical features from the input image. The pooling layers' functionalities consist of receptive field enlargement and dimensionality reduction, which means to reduce the size of the output feature maps. The fully connected layers are used as a classifier, which outputs the probabilities predicting the input image to each class. Because our goal is to extract features pixel by pixel, the pooling and fully connected layers are not yet used in our designed network.

Another important issue in the network design is the selection of kernel size of each convolutional layer. We apply a series of kernels with gradually increasing size to the convolutional layers except the last one. This strategy, which ensures the enlargement of the receptive field, can serve as functional substitutes for the pooling layer. The kernel size of the last convolutional layer, which acts as a selector of feature maps, is 1. In this letter, we select the above-mentioned parameters experimentally.

B. Contrastive Loss Function

During training the designed network, we need a proper objective function whose optimization can produce good performance in the feature extraction task. The purpose of training is to find a parametric function G_w, such that after the image transformation by the function, the feature points of the unchanged pixel pairs are close to each other in the feature space and the ones of changed pixel pairs are far away. As [10] presented, the contrastive loss can produce the above-mentioned function when it gets a minimum value.

Let $X = \{x(i, j)|1 \leq i \leq h, 1 \leq j \leq w\}$ be an aerial image, with a size of $h \times w \times c$, where h and w are spatial dimensions and c is the channel dimension. Thus, $x(i, j)$ is the intensity vector at location (i, j) in the image, with c dimensions, which is equal to the number of channels of the input image. In our proposed method, we use all of the RGB channels of the available images, so c is 3. Let X_1 and X_2 be two coregistered aerial images. Let $G_w(X) = \{G_w(X)_{i,j}|1 \leq i \leq h, 1 \leq j \leq w\}$ be the output feature tensor, where X is the input image. The size of the feature tensor is $h \times w \times d$, where d is the feature dimension. Thus, $G_w(X)_{i,j}$ is the feature vector of the pixel whose location is (i, j) in the image X. Let $D_w(X_1, X_2) = \{D_w(X_1, X_2)_{i,j}|1 \leq i \leq h, 1 \leq j \leq w\}$ be

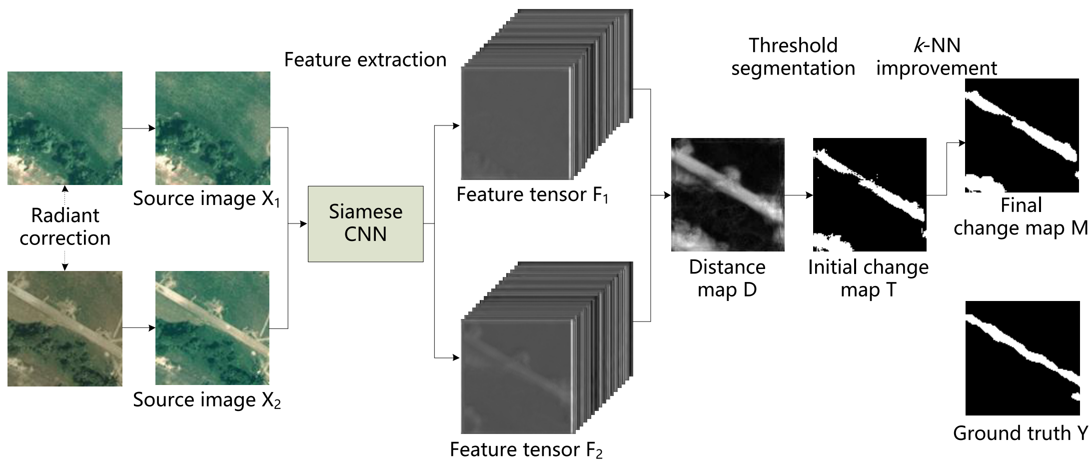

Fig. 2. Schematic of the proposed siamese CNN-based change detection method.

the distance map between X_1 and X_2, where $D_w(X_1, X_2)_{i,j}$ is the parameterized distance function to be learned between the pixels $x_1(i, j)$ and $x_2(i, j)$. Define $D_w(X_1, X_2)_{i,j}$ be the Euclidean distance between the feature vector $G_w(X_1)_{i,j}$ and $G_w(X_2)_{i,j}$ [10]. That is

$$D_w(X_1, X_2)_{i,j} = \|G_w(X_1)_{i,j} - G_w(X_2)_{i,j}\|_2. \quad (1)$$

To shorten notation, $D_w(X_1, X_2)_{i,j}$ is written as $D_{i,j}$. Then, the loss function in its general form is

$$\begin{aligned}\ell(W) &= \sum_{k=1}^{P} L(W, (Y, X_1, X_2)^k)\\ &= \sum_{k=1}^{P}\sum_{i,j}\left(1 - y_{i,j}^k\right)L_U\left(D_{i,j}^k\right) + y_{i,j}^k L_C\left(D_{i,j}^k\right) \quad (2)\end{aligned}$$

where $Y = \{y(i, j)|1 \le i \le h, 1 \le j \le w\}$ is a binary GT map assigned to the input image pair, with a size of $h \times w$, $y(i, j) = 0$ if the corresponding pixel pair is deemed to be unchanged, and $y(i, j) = 1$ if it is deemed to be changed. $(Y, X_1, X_2)^k$ is the kth labeled training sample pair, L_U and L_C are the partial loss functions for a pair of unchanged and changed pixels, respectively, and P is the number of training sample pairs.

L_U and L_C must be designed, such that $D_{i,j}$ would produce a low value for a pair of unchanged pixels and a high value for the changed pixel pair when L gets the minimum value [10]. In [10], L_U and L_C are defined as follows:

$$L_U\left(D_{i,j}^k\right) = \frac{1}{2}\left(D_{i,j}^k\right)^2 \quad (3)$$

$$L_C\left(D_{i,j}^k\right) = \frac{1}{2}\left\{\max\left(0, m - D_{i,j}^k\right)^2\right\} \quad (4)$$

where $m > 0$ is a margin. Changed pixel pairs contribute to the loss function only if their parameterized distance is within this margin [10]. m is set to 1.

In the proposed method, we optimize a weighted contrastive loss during training the network. For change detection, the numbers of changed and unchanged pixel pairs vary greatly (e.g., changed pixel pairs are usually fewer). Then, it is necessary to weight the loss differently based on the true class (changed and unchanged). This is termed *class balancing*. Thus, the final loss function is

$$\begin{aligned}L(W, (Y, X_1, X_2)^k) = &\sum_{i,j}\left(1 - y_{i,j}^k\right)\frac{1}{2}\left(D_{i,j}^k\right)^2 w_U\\ &+ y_{i,j}^k\frac{1}{2}\left\{\max\left(0, m - D_{i,j}^k\right)^2\right\}w_C \quad (5)\end{aligned}$$

where w_U and w_C denote the weights for unchanged and changed pixel pairs, respectively. We verify the effectiveness of the weighted loss function experimentally.

We use *average frequency balancing*, where the weights w_U and w_C assigned to the partial loss functions L_U and L_C are given by

$$w_U = \frac{f_{\text{avg}}}{f_U} \quad (6)$$

$$w_C = \frac{f_{\text{avg}}}{f_C} \quad (7)$$

where f_U and f_C are the frequencies of unchanged and changed pixel pairs, respectively, while f_{avg} is the average class frequency. The class frequencies can be computed on all of the training image pairs. In fact, the average class frequency is 0.5, since there are only two classes, changed and unchanged, for change detection. This balancing implies that if the changed pixel pairs are fewer than unchanged ones, then the weight is larger than 1, which can result in more contribution to the loss.

C. Detailed Detection Scheme

The schematic of the proposed method is shown in Fig. 2. First, histogram matching is used for radiant correction to the two coregistered images. Then, the processed images X_1 and X_2 are fed to the trained siamese CNN and two feature tensors $G_w(X_1)$ and $G_w(X_2)$ (denoted as F_1 and F_2) are obtained on the top of the network. Then, a distance map $D_w(X_1, X_2)$ (for convenience, denoted as D) can be generated by (1). In the distance map D, the larger the value is, the more likely changed the corresponding pixel pair is. Therefore, an initial change map (denoted as T) can be obtained by using simple

TABLE I

COMPARISON OF LOSS FUNCTION, WEIGHTED LOSS FUNCTION, AND k-NN IMPROVEMENT; MEASURES ARE WITH RESPECT TO THE CHANGED CLASS

Method	SZADA/1			TISZADOB/3		
	Precision (%)	Recall (%)	F-rate (%)	Precision (%)	Recall (%)	F-rate (%)
no-weighted loss	29.3	50.5	37.1	63.7	48.4	55.0
weighted loss	40.0	56.2	46.8	87.1	82.7	84.8
weighted loss with k-NN	**41.2**	**57.4**	**47.9**	**88.3**	**85.1**	**86.7**

threshold segmentation on D. The threshold getting the best performance can be computed on the training sample pairs.

After threshold segmentation, an improvement (denoted as M) for the initial change map through a k-nearest neighbor (k-NN) approach [11] can be obtained.

III. EXPERIMENTS

A. Implementation Details

1) Data Set: In order to train the proposed network and evaluate the method, we employ the SZTAKI AirChange Benchmark Set. This data set has already been used in [5], [6], [12], and [13]. The data set contains three sets of registered optical aerial image pairs taken with large time difference and in different seasonal conditions. The data sets, which are named SZADA, TISZADOB, and ARCHIEVE, contain 7, 5, and 1 image pairs, respectively. The associated GTs are manually labeled by an expert. The size of each image with 1.5-m/pixel resolution is 952×640 pixels. In our proposed method, image pairs from SZADA and TISZADOB data sets are used for training and testing.

Considering that the training of the network is processed on the data set, the proposed method can be suitable for the optical aerial images, which have similar feature statistics to the images in the data set.

2) Training and Testing Data Construction: Because the designed network does not contain the fully connected layers, which have fixed dimensions on input and output data, the size of image pairs for testing and training can be different. To construct testing clips, we crop the top-left corner of each image pair to 784×448. The rest of the region is used for training data construction, where the regions are cropped to 112×112 overlappingly. To augment training data, each cropped training pair is rotated by 90°, 180°, and 270° and also horizontally and vertically flipped. Note that the image regions used for testing and training are not overlapped, which can ensure the validity of the experimental results. Finally, we get 12 testing image pairs with a size of 784×448 and 3744 training image pairs with a size of 112×112.

3) Optimization: We implement the proposed siamese network using the Caffe [14] framework. The stochastic gradient descent (SGD) with momentum is applied for optimization. In the training procedure, the learning rate is always set to 0.001, and the momentum and the weight decay are set to 0.9 and 0.0005, respectively. The weights of each convolutional layer are initialized with the Msra algorithm [15]. The mini-batch size is set to 32. The trained network is finally obtained after around 10k SGD iterations, when the training loss no longer decreases. The training and testing are performed in a single NVIDIA GTX 1080 GPU with 8G memory.

4) Computational Time: Although training the network in the proposed method needs a lot of time, the inference is not time-consuming. In our implementation, the training takes about 2 h and 16 min, whereas the inference on a sample image pair needs around 50 s. What is more, the training process can be finished offline.

B. Results and Evaluation

To evaluate the proposed method, we compare it to two other state-of-the-art methods (CXM [5] and SCCN [7]). In [13], some comparative tests have been provided between the CXM model [5] and another method. In order to refer to the further results in [13], we perform evaluations based on two test image pairs, because most of the comparison results demonstrated in [13] are relevant to these two image pairs. To keep here all validation figures relevant, we limit our forthcoming experiments to these two image pairs. One image pair is selected from the SZADA data set and another from the TISZADOB data set.

1) Evaluation Measures: In order to evaluate the performance of the proposed method, we calculate the precision (Pr) and recall (Rc) rates from the point of view of the changed class, and calculate the F-measure rate (Fr) as the harmonic mean of Pr and Rc.

2) Experiments on the Loss Function and k-NN Improvement: In order to verify the validity of the weighted contrastive loss function, we use the contrastive loss function with or without weighting as the objective function when training the network and with other experimental conditions remaining the same. The comparison results are shown in Table I. We can see here a promising improvement by using the weighted contrastive loss function for both of the testing image pairs, which achieves the performance improvement of around 9.7% and 29.8% of F-measure, respectively. In Table I, we also provide the improvement of the k-NN approach, which confirms the effectiveness of the update method. The update method shows an overall improvement at least 1% in terms of F-measure over the initial change detection results.

3) Comparison of the Proposed Method and Two Other Methods: We compare our proposed method to the CXM model [5] and the SCCN [7], in which the default parameters are used. In CXM, $K = 5$, $z = 17$, and $\rho = \varphi^i = 1$, $i \in \{g, c, *, \nu\}$ are used. In SCCN, the sizes of all convolution kernels are set to 3×3 and 3 coupling layers, with 20 feature maps specified. We set $\lambda = 0.1$ as mentioned in [7]. The quantitative results are listed in Table II, and a few qualitative change mask examples are shown in Figs. 3 and 4. We can see that our proposed method obtains comparable and even better results in terms of F-measure. From the results, we can find some advantage of our method that it achieves higher Pr while slightly smaller Rc.

TABLE II

COMPARISON WITH TWO METHODS ON TWO IMAGE PAIRS; MEASURES ARE WITH RESPECT TO THE CHANGED CLASS

Dataset	Metrics	CXM	SCCN	Ours
SZADA/1	Precision (%)	36.5	24.4	**41.2**
	Recall (%)	**58.4**	34.7	57.4
	F-rate (%)	44.9	28.7	**47.9**
TISZADOB/3	Precision (%)	61.7	**92.7**	88.3
	Recall (%)	**93.4**	79.8	85.1
	F-rate (%)	74.3	85.8	**86.7**

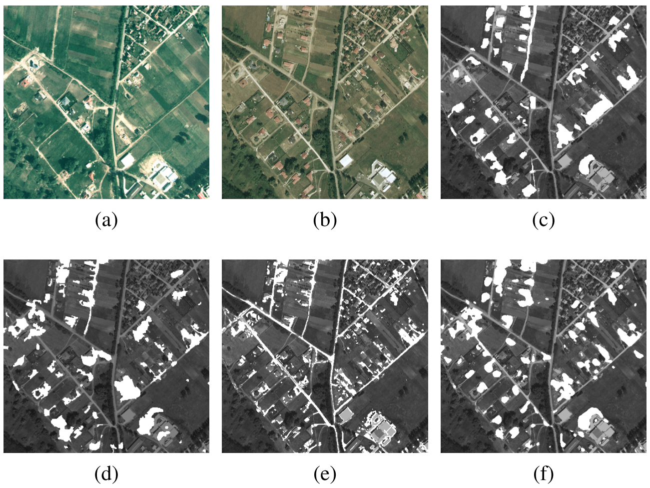

Fig. 3. Results by the proposed method, the SCCN, and the CXM model for an image segment from the SZADA data set. (a) and (b) display the two images. Changes are displayed with white color. (c) GT. (d) Result by CXM. (e) Result by SCCN. (f) Result by the proposed method.

Compared with the CXM model based on Figs. 3 and 4, the proposed method produces more smoothening effect. We owe the good performance to the extracted features by the deep siamese CNN. The feature is more robust than the weak feature used by the CXM model. The proposed method also outperforms the SCCN on the two data sets. On the SZADA/1 data set, atmospheric and light variations and the complicated feature statistics may lead to invalidity to the assumption, which is used in the unsupervised SCCN method, and therefore result in the worse performance. It is evident that the proposed supervised method can obtain significant additional information for the change detection from the training data compared with the unsupervised SCCN method. On the TISZADOB/3 data set, the SCCN, which might be suitable for this kind of images, achieves higher Pr than the proposed method does, since in the images, the changed regions are larger and the unchanged regions show similar features. However, our method outperforms it in terms of F-measure.

IV. CONCLUSION

In this letter, a supervised change detection method based on the deep siamese convolutional network has been proposed for optical aerial images. The designed siamese network, which is trained using the weighted contrastive loss, extracts hierarchical features from the input image pairs. The extracted features are more abstract and robust than the hand-crafted features. Then, we use the distance of the feature vectors to detect changes between the image pair. The experimental results have shown that our proposed method achieves comparable

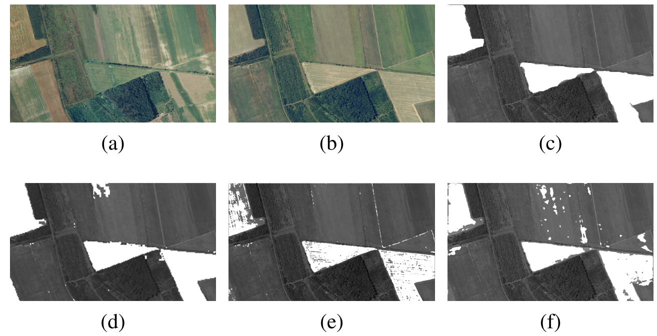

Fig. 4. Results by the proposed method, the SCCN, and the CXM model for an image segment from the TISZADOB data set. (a) and (b) display the two images. Changes are displayed with white color. (c) GT. (d) Result by CXM. (e) Result by SCCN. (f) Result by the proposed method.

and even better performance in terms of F-measure on the testing image pairs than two other state-of-the-art methods do. Our future work includes but is not limited to avoiding the postprocessing by improving the siamese network, or the threshold segmentation.

REFERENCES

[1] P. Zhong and R. Wang, "A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3978–3988, Dec. 2007.

[2] 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.

[3] J. Zhao, M. Gong, J. Liu, and L. Jiao, "Deep learning to classify difference image for image change detection," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jul. 2014, pp. 411–417.

[4] F. Gao, J. Dong, B. Li, Q. Xu, and C. Xie, "Change detection from synthetic aperture radar images based on neighborhood-based ratio and extreme learning machine," *J. Appl. Remote Sens.*, vol. 10, no. 4, p. 046019, 2016.

[5] 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.

[6] P. Singh, Z. Kato, and J. Zerubia, "A multilayer Markovian model for change detection in aerial image pairs with large time differences," in *Proc. IEEE 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 924–929.

[7] 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.*, to be published.

[8] 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.

[9] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 4353–4361.

[10] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Sep. 2006, pp. 1735–1742.

[11] A. Touazi and D. Bouchaffra, "A k-nearest neighbor approach to improve change detection from remote sensing: Application to optical aerial images," in *Proc. IEEE 15th Int. Conf. Intell. Syst. Design Appl.*, Dec. 2015, pp. 98–103.

[12] C. Benedek and T. Szirányi, "A mixed Markov model for change detection in aerial photos with large time differences," in *Proc. IEEE 19th Int. Conf. Pattern Recognit.*, Sep. 2008, pp. 1–4.

[13] C. Benedek, M. Shadaydeh, Z. Kato, T. Szirányi, and J. Zerubia, "Multilayer Markov random field models for change detection in optical remote sensing images," *J. Photogram. Remote Sens.*, vol. 107, pp. 22–37, Sep. 2015.

[14] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd Int. Conf Multimedia (ACM)*, 2014, pp. 675–678.

[15] 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.*, Feb. 2015, pp. 1026–1034.