Deep Siamese Multi-scale Convolutional Network for Change Detection in Multi-temporal VHR Images

Hongruixuan Chen
State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing
Wuhan University
Wuhan, P.R. China
Qschrx722RS@163.com

Chen Wu*
(Corresponding Author)
State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing
Wuhan University
Wuhan, P.R. China
chen.wu@whu.edu.cn

Bo Du
School of Computer Science, and Collaborative Innovation Center of Geospatial Technology
Wuhan University
Wuhan, P.R. China
gunspace@163.com

Liangpei Zhang
State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing
Wuhan University
Wuhan, P.R. China
zlp62@whu.edu.cn

***Abstract*—In this letter, we propose a powerful multi-scale feature convolution unit for change detection. The proposed unit is able to extract multi-scale features in the same layer. Based on the proposed unit, two novel deep Siamese convolution networks, deep Siamese multi-scale convolutional network (DSMS-CN) and deep Siamese multi-scale fully-convolutional network (DSMS-FCN), are designed for unsupervised and supervised change detection in multi-temporal very high resolution (VHR) images. For unsupervised change detection, we implement automatic pre-detection to obtain training patch samples, and the DSMS-CN fits the statistical distribution of changed and unchanged ground from patch samples for change detection through multi-scale feature extraction module and deep Siamese architecture. For supervised change detection, an end-to-end deep network DSMS-FCN is trained in any size of multi-temporal VHR images, and directly output the binary change map. The experimental results with a GF data set and an open change detection data set confirm that the two proposed architectures perform better than the state-of-the-art methods.**

***Keywords*—Change detection, multi-scale feature convolution, deep Siamese convolution network, multi-temporal images, very high resolution images.**

I. Introduction

Change detection (CD) is the process of identify differences in the state of an object or phenomenon by observing it at different time [1]. And change detection is playing a vital role in land-use and land-cover (LULC) change, forest or vegetation change, urban expansion research and damage assessment.

Multi-spectral image is the most commonly used data source for CD, and a number of methods based on it were proposed. Change detection analysis (CVA) generates a difference image (DI) and clusters DI to achieve the change result. Principal component analysis (PCA), as a dimension reduction methods, transforms the images into a new feature space and select a part of new bands for change detection. In [2] and [3], multivariate alteration detection (MAD) and its iterative version IRMAD are proposed, which extract change objects by maximizing difference of projection feature. Based on slow feature analysis (SFA) algorithm, Wu et al. [4] proposed a SFA change detection method. The method aims to find the most invariant component in multi-temporal images and find the optimal feature space, in which the change object would be highlighted.

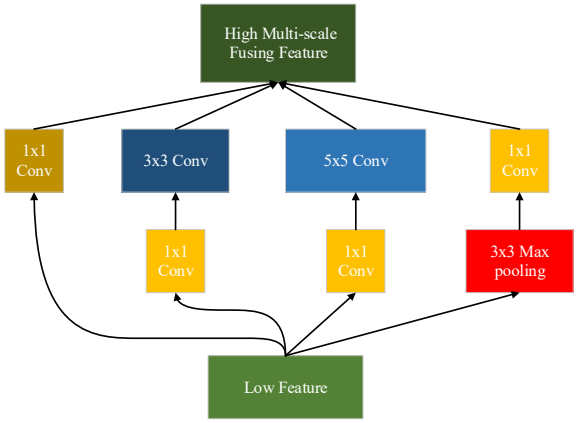
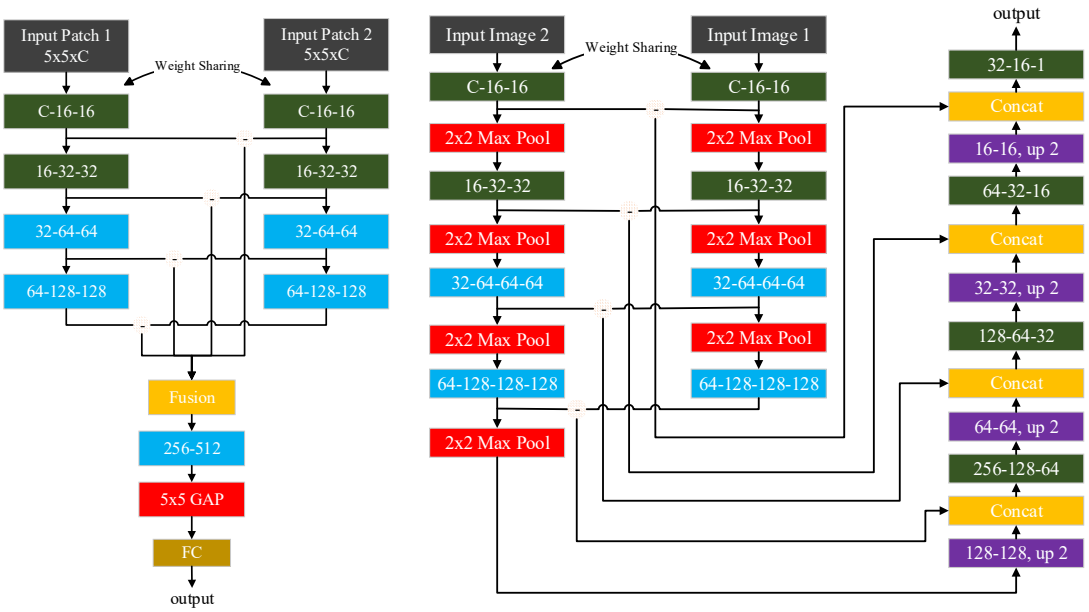


Fig. 1. The proposed multi-scale feature convolution unit. Unlike conventional single convolution units, this unit can extract multi-scale features in parallel by four ways in the same layer.

Nowadays, with the development of satellite sensors, very-high-resolution (VHR) images are more available by various types of sensors, such as QuickBird, IKONOS, SPOT and Worldview. And VHR images can provide abundant ground details and spatial distribution information, which have crucial effects on the research of urban change analysis and building detection. Nevertheless, due to internal complexity caused by very high resolution of VHR images, conventional multi-spectral change detection methods may not be applicable for VHR images. Recently, deep learning (DL) has achieved significant performances in many domains, as well as remote sensing image processing. And convolutional neural network (CNN), as a classical DL

978-1-7281-4615-7/19/$31.00 ©2019 IEEE



(a) DSMS-CN (b) DSMS-FCN

Fig. 2. Illustration of the two proposed architectures for change detection. Color legend: green means conventional 3x3 convolution module, blue means multi-scale feature convolution module MFCU, red means 2x2 max pooling layer and purple means transpose convolution.

architecture, is expert in extracting multi-level information, which is suitable for extracting spectral information and spatial context information in VHR images. In [5], a deep symmetric network is proposed for change detection of VHR heterogeneous images, and a convolutional layer plays a role of feature extraction. Yang et al. [6] introduce a deep Siamese convolutional network for aerial image change detection, which extract features by two weight sharing convolutional branches and generate binary change map based on the feature difference of the last layer. Two fully convolutional Siamese architectures are first proposed in [7], which are trained end-to-end on change detection dataset and have achieved good performances. All of these method only adopt 3x3 convolution kernel as feature extraction module. Though 3x3 convolution kernel could extract spectral features and spatial context features in some extents, it still has some powerlessness in complex ground situations of VHR images. No research has attempted to use other sizes of convolution kernels or even multiple kernels for change detection in VHR image. Inspired by "network in network" structure [8] and Inception network [9], we propose a multi-scale feature convolution unit (MFCU) extracting multi-scale spectral features and spatial context features in the same layer, which is suitable for VHR images. Adopting the MFCU as the basic feature extraction module, two methods are designed for unsupervised and supervised change detection.

The rest of this letter is organized as follows. Section II describes our methods in detail. Section III contains quantitative and qualitative comparisons with the state-of-the-art change detection methods. In the end, Section IV draws the conclusion of our work in this letter.

II. Proposed Methods

A. Multi-scale Feature Convolution Unit

For the purpose of extracting multi-scale features from VHR images, multi-scale feature convolution unit (MFCU) is proposed. As shown in Fig.1, the MFCU is a "network in network" structure, and it extracts multi-scale features in parallel by four ways, namely 1x1 convolution kernel, 3x3 convolution kernel, 5x5 convolution kernel and 3x3 max pooling. The 1x1 convolution kernel focuses on extracting the features of pixels itself. The 3x3 convolution kernel extracts the features in a neighborhood. The 5x5 convolution kernel extracts features in a larger range, which is suitable for some large-scale continuous objects. And, max pooling is responsible for extracting the salient features. At last, the four type features are fused to obtain the multi-scale features. It should be noted that the 1x1 convolution before 3x3 convolution and 5x5 convolution is a bottleneck design [9], which can reduce the parameters of network and make network easier to train. Compared with conventional single convolution unit, the MFCU extracts multi-scale features, which improves the feature abstraction ability of network, but does not significantly increase parameters of network.

B. Deep Siamese Multi-scale Convolutional Network

Using multi-scale space convolution unit as multi-scale feature extraction module, we design two novel Deep Siamese Multi-scale Convolutional Network for unsupervised and supervised change detection with multi-temporal VHR images, respectively.

The first proposed network is deep Siamese multi-scale convolutional network (DSMS-CN). The DSMS-CN (Fig. 2(a)) has two components: feature extraction network and change judging network. The feature extraction layer is a Siamese network and its two branches extract features from two patches using a same way because of weight sharing. The former two normal convolutional modules in each branch transform the spatial and spectral information into high dimension features, and the two latter MFCU modules extract abundant multi-scale features from high dimension features. Then, the absolute differences of multiple-layer features are fused and inputted into change judging network. In change judging network, a MFCU layer extracts multi-scale difference feature, and a global average pooling layer (GAP) replace fully connected layer to generate feature vector, which can make network more robust and reduce overfitting [9]. Lastly, the changed result is obtained by a fully connected layer.

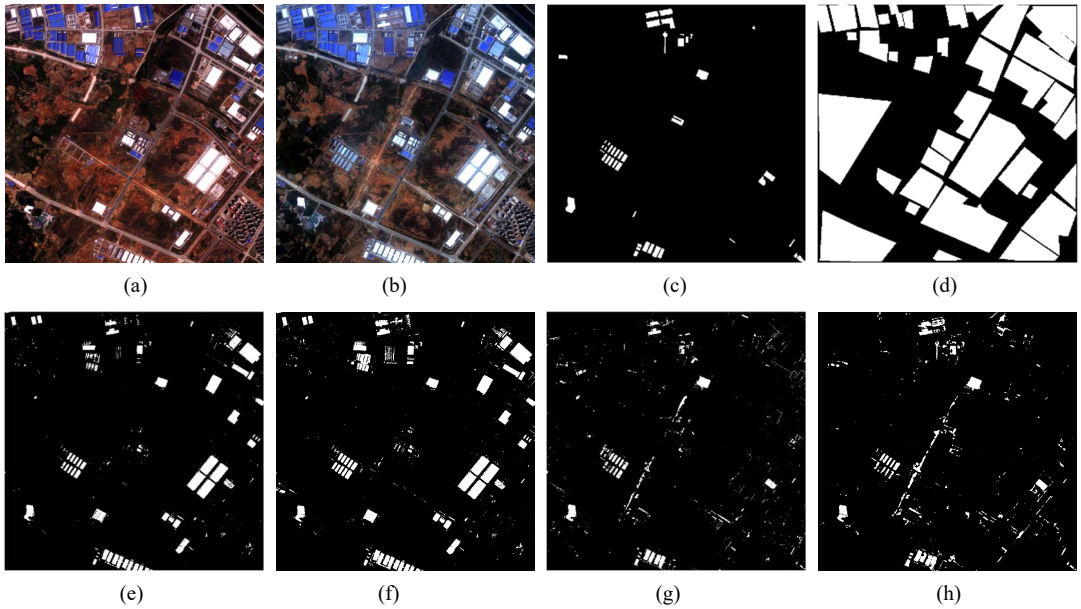


Fig. 3. Visualized results obtained by proposed methods, CVA, IRMAD and ISFA on the GF data set. (a) and (b) are multi-temporal images. (c) is changed ground truth. (d) is unchanged ground truth. (e) IRMAD. (f) ISFA. (g) CVA. (h) DSMS-CN. IRMAD and ISFA are iterative variants of MAD and SFA.

Another proposed network is deep Siamese multi-scale fully-convolutional network (DSMS-FCN). Same as the general FCN, the DSMS-FCN (Fig. 2(b)) consists of two parts: an encoder and a decoder. The encoder layers have two equal weight sharing branches, and the features of multi-temporal images are extracted in a same approach. Each branch has four max pooling and four subsampling layers, and the latter two subsampling layers consist of multi-scale feature convolution module. Based on the concept of skip connections in the U-Net [10], the features of subsampling layer and upsampling layer at the same scale are concatenated during upsampling, which can produce accurate binary changed map with precise boundaries. The motivation for using the absolute value of the difference between the two branches to be concatenated with the features of the upsampling layer is that change detection is trying to detect differences between multi-temporal images.

C. *Unsupervised & Supervised Change Detection Methods*

Based on the two aforementioned deep Siamese multi-scale feature convolution networks, we propose unsupervised and supervised change detection architectures.

In unsupervised change detection method, the DSMS-CN is adopted. The pre-classification is the first step. The main purpose of this step is to find pixels which have extremely high changed or unchanged probabilities. CVA is first adopted to generate difference image (DI) of multi-temporal images. Then, fuzzy c-means clustering (FCM) is implemented to partition DI into three clusters: ω_c , ω_{uc} , and ω_{tbc} .The pixels belonging to ω_c and ω_{uc} are reliable pixels which have high changed or unchanged probabilities. And the pixels in ω_{tbc} are uncertain and need to be classified. Finally, the neighborhood area of pixels in ω_c and ω_{uc} are chosen as training patches for training network. After training process is completed, the pixels in ω_{tbc} are categorized by network. The whole process is completely unsupervised.

In supervised architecture, the DSMS-FCN is directly trained end-to-end on change detection datasets without any pre-training. The inputs of the DSMS-FCN are two complete multi-temporal images, and the output is a binary change map. Unlike unsupervised architecture and majority of recent patch-based approach [6], the DSMS-FCN is able to process images of any sizes and do not require sliding patch-window, therefore the accuracy and speed of inference could be improved.

TABLE I. ACCURACY ASSESSMENT ON CHANGE DETECTION RESULTS OF DIFFERENT METHODS ON GF DATA SET

Method	Overall Accuracy	Kappa
MAD	84.46	26.62
IRMAD	92.12	32.89
SFA	82.30	23.35
ISFA	92.00	35.30
CVA	97.11	64.34
DSMS-CN	**97.89**	**69.73**

III. EXPERIMENT

A. *Unsupervised Change Detection*

To evaluate our DSMS-CN and proposed unsupervised methods, we apply our methods on a GF data set. The three unsupervised methods used for comparison are CVA, MAD [2], SFA [4], and their iterative versions [3, 4].

As shown is Fig. 3, the result obtained by IRMAD misclassify a lot of building roofs into changed class, while some slightly changed pixels are classified as unchanged class, thus its kappa coefficient is only 32.89. The result of ISFA is slightly better than IRMAD, where its kappa coefficient is 35.30. However, ISFA still has similar shortcomings with IRMAD. This is because both the MAD and SFA are based on the central limit theorem, whereas the Gaussianity of VHR images is not obvious. Therefore, they are not suitable for change detection of the VHR images covering a small region, even though they can achieve outstanding results in low- and medium- resolution images. Compared with MAD and SFA,

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

Fig. 4. Visualized results obtained by proposed methods on the Air Change Dataset. (a) and (b) are multi-temporal images. (c) is the ground truth. (d) is results obtained by multi-scale FC-EF. (e) is results obtained by the proposed DSMS-FCN.

TABLE II. ACCURACY ASSESSMENT ON CHANGE DETECTION RESULTS OF DIFFERENT METHODS ON ACD

Method	Rec.	Pre.	OA	F1
DSCN	41.2	57.4	NA	47.9
CXM	36.5	58.4	NA	44.9
SCCN	22.4	34.7	NA	28.7
FC-EF	43.57	62.65	93.08	51.40
FC-Siam-conc	40.93	65.61	92.46	50.41
FC-Siam-Diff	41.38	**72.38**	92.40	52.66
MS-FCEF	44.70	65.85	**93.25**	53.25
DSMS-FCN	**44.77**	72.31	93.18	**55.30**

CVA directly producing a difference image and performing clustering achieve a relative good result. Nevertheless, changes in a few buildings and small objects, such as roads, are not recognized and plenty of margins of buildings are misclassified as changed class. Although the training samples are selected by CVA and FCM, the DSMS-CN achieves the best result with OA of 97.89 and kappa coefficient of 69.73. It means that our proposed DSMS-CN can effectively fit the distributions of ground changes in VHR images based on pre-classification samples, and achieve a better performance.

B. Supervised Change Detection

For the purpose of training the proposed DSMS-FCN and evaluate the method, we employ an open available VHR images data set: Air Change Dataset (ACD), which has already been used in [6, 7, 11]. We adopted the data split that was proposed in [6] and [7]: the top-left 784x448 corner of the Szada-1 were cropped for testing, and the rest of the images were used for training. The methods used for comparison were DSCN [6], CXM [11], SCCN [5] and three fully convolutional architectures proposed in [7], using the values in [6] and [7]. Table 2 contains the accuracy assessment of the proposed DSMS-FCN, multi-scale FC-EF and other state-of-the-art methods. The FC-EF, FC-Siam-conc and FC-Siam-Diff are three architectures proposed in [7]. The multi-scale FC-EF is a variation of the FC-EF. Its conventional convolutional unit is replaced by our MFCU. And Fig. 4 is an illustration of our results on this dataset.

The results obtained on the Szada-1 show the superiority of the DSMS-FCN, which outperforms all the other methods and achieves the best recall metric and F1 rate. Utilizing the MFCU, each metric of multi-scale FC-EF is better than FC-EF, and this architecture achieves the best overall accuracy.

IV. CONCLUSION

In this letter, a powerful multi-scale feature convolution unit is presented, which is different from conventional convolution only extracting single-scale feature in one layer, and able to extract features at multi-scale in the same layer by a "network in network" structure. Based on the unit, two deep Siamese convolution network is designed for unsupervised and supervised change detection of VHR images. The DSMS-CN, used for unsupervised change detection and trained on pre-classification samples generated by CVA and FCM, outperforms state-of-the-art unsupervised methods. And the DSMS-FCN, as a fully convolutional architecture, is responsible for supervised change detection. In the experiment with Air Change Dataset, compared with three fully convolution network, a patch-based method and other state-of-the-art methods, our architecture delivers better performance and MFCU also exhibits powerful feature extraction capability.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 61601333, 61822113 and 41801285.

REFERENCES

[1] Singh A, "Digital change detection techniques using remotely sensed data, " International Journal of Remote Sensing, vol. 10, no. 6, pp. 989–1003, 1998.

[2] A. A. Nielsen, K. Conradsen, J. J. Simpson, "Multivariate Alteration Detection (MAD) and MAF Postprocessing in Multispectral, Bitemporal Image Data: New Approaches to Change Detection Studies, " in Remote Sensing of Environment, vol. 64, no. 1, pp. 1-19, 1998.

[3] A. A. Nielsen, "The Regularized Iteratively Reweighted MAD Method for Change Detection in Multi- and Hyperspectral Data," in IEEE Transactions on Image Processing, vol. 16, no. 2, pp. 463-478, Feb. 2007.

[4] C. Wu, B. Du and L. Zhang, "Slow Feature Analysis for Change Detection in Multispectral Imagery," in IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 5, pp. 2858-2874, May 2014.

[5] J. Liu, M. Gong, K. Qin and P. Zhang, "A Deep Convolutional Coupling Network for Change Detection Based on Heterogeneous Optical and Radar Images," in IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 3, pp. 545-559, March 2018.

[6] 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," in IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 10, pp. 1845-1849, Oct. 2017.

[7] R. Caye Daudt, B. Le Saux and A. Boulch, "Fully Convolutional Siamese Networks for Change Detection," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, 2018, pp. 4063-4067.

[8] Min Lin, Qiang Chen, and Shuicheng Yan. "Network in network, "CoRR, abs/1312.4400, 2013.

[9] C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 1-9. doi: 10.1109/CVPR.2015.7298594

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

[11] C. Benedek and T. Sziranyi, "Change Detection in Optical Aerial Images by a Multilayer Conditional Mixed Markov Model," in IEEE Transactions on Geoscience and Remote Sensing, vol. 47, no. 10, pp. 3416-3430, Oct. 2009.