END-TO-END CHANGE DETECTION USING A SYMMETRIC FULLY CONVOLUTIONAL NETWORK FOR LANDSLIDE MAPPING

*Tao Lei[1], Qi Zhang[1], Dinghua Xue[1], Tao Chen[2], Hongying Meng[3], Asoke K. Nandi[*3]*

[1]School of Electronical and Information Engineering, Shaanxi University of Science and Technology, Xi'an 710021, P. R. China
[2]Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China
[3]Department of Electronic and Computer Engineering, Brunel University London, Uxbridge, Middlesex, UB8 3PH, United Kingdom

ABSTRACT

In this paper, we propose a novel approach based on a symmetric fully convolutional network within pyramid pooling (FCN-PP) for landslide mapping (LM). The proposed approach has three advantages. Firstly, this approach is automatic and insensitive to noise because multivariate morphological reconstruction (MMR) is used for image preprocessing. Secondly, it is able to take into account features from multiple convolutional layers and explore efficiently the context of images, which leads to a good tradeoff between wider receptive field and the use of context. Finally, the selected pyramid pooling module addresses the drawback of single-scale pooling employed by convolutional neural network (CNN), fully convolutional network (FCN), U-Net, etc. Experimental results show that the proposed FCN-PP is effective for LM, and it outperforms state-of-the-art approaches in terms of four metrics, *Precision*, *Recall*, *F-score*, and *Accuracy*.

Index Terms— Convolutional neural network (CNN), change detection, landslide mapping (LM), Multivariate morphological reconstruction (MMR).

1. INTRODUCTION

Landslides are important geomorphic agents that occur over a wide variety of spatial and temporal scales in many mountains, and thus reshape the landscape and change local topography [1]. Landslide mapping (LM) focuses on outlining slide boundaries, neglecting the wealth of information revealed by internal deformation features. Traditional LM approaches rely on visual interpretation of aerial photographs and intensive field surveys, which are highly labor-intensive and time-consuming for mapping of large areas. With the rapid progress of machine learning and remote sensing technology, a large number of advanced approaches used for LM have been proposed in recent years. Most of them depend on change detection that aims to detect the changed information of target at areas by analyzing the multi-temporal images acquired in different time of the same geographical area [2]. The popular ones can be roughly divided into three categories: threshold-based approaches [3, 4], approaches based on feature extraction and feature classification [5-7], and deep learning approaches [8-10].

The approaches in the first category can generate landslide areas by computing one or more thresholds used for the difference image of a pair of bi-temporal images. However, they are sensitive to noise and have a low robustness for different landslide images.

The second category of approaches is often composed of two parts, feature extraction and feature classification. Most of them use unsupervised learning algorithms, such as k-means, FCM, GMM, etc. [11] to achieve change detection for a difference image of bi-temporal images. Because the difference image includes lots of noise caused by imaging devices or illumination, some image pre-processing and post-processing operations are necessary to improve LM results. Li et al. [5] employed edge-based level set evolution (ELSE) and region-based level set evolution (RLSE) to track initial change detection profiles, leading to better landslide candidate areas and LM results. Furthermore, they also proposed change detection based on Markov random field (CDMRF) for LM [6]. However, the performance of ELSE, RLSE, and CDMRF seriously depend on the quality of difference images and parameter selection. To reduce much dependencies, Lei et al. [7] employed morphological reconstruction and a fast clustering approach to distinguish changed and unchanged areas for LM. The method provides better LM results than ELSE, RLSE, and CDFCM.

Based on deep learning technologies, Gong et al. [12] proposed a change detection approach using a deep neural network for synthetic aperture radar images. As the proposed network architecture only included a few hidden layers, and it adopted full connection without using convolutional operation, the context of images is utilized inefficiently. Moreover, it is difficult to train the network due to the full connection. To

obtain better results, Liu et al. [13] proposed a new deep convolutional coupling network that is fully unsupervised without using any labels. To apply deep learning technologies to landslide recognition, Ding et al. [14] used convolutional neural network (CNN) and texture change detection to recognize landslides. Because CNN employs multiple pooling layers and a fully connectional layer to achieve classification tasks, the final result is coarse and has a low recognition accuracy.

To address the aforementioned issues, we propose a symmetric fully convolutional network within pyramid pooling (FCN-PP) that is able to learn better image features to improve LM results. Before applying the FCN-PP, a multivariate morphological reconstruction (MMR) [15] is performed on training or testing images. The proposed approach has two contributions. The first is that we design a powerful deep convolutional network that is able to tradeoff the use of context and the localization accuracy. The second is that we propose an end-to-end change detection approach that requires only a small number of training samples but provides high recognition accuracy for LM.

2. METHODOLOGY

The proposed approach is composed of two parts. The first part is image preprocessing using MMR [15]. Because a pair of bi-temporal high-resolution remote sensing images are captured at different time, different imagery environmental factors lead to a poor difference image that usually includes many falsely changed areas or misses truly changed areas. Therefore, it is necessary to implement image filtering for preprocessing. The second part is to construct a deep convolutional network that is able to utilize the context efficiently and provides high localization accuracy for LM.

2.1. The network structure of FCN-PP

In fact, change detection-based LM can be considered as an image segmentation task. Because the final output only includes changed and unchanged areas, the task is a pixel-level binary classification, i.e., a binary segmentation task. The typical use of deep convolutional networks is on classification tasks. However, pixel-level classification task (semantic segmentation) is more complex due to the requirement of localization.

Although CNN is able to achieve effective image classifications, it provides a poor result on image segmentation since it employs global pooling and misses the spatial information of images. Fully convolutional networks (FCN) [16] overcomes the problem by taking into account the features from multiple convolutional layers, which results in better localization and the use of context. Consequently, FCN provides better image segmentation results than CNN. Inspired by the idea of FCN, we presented a fully convolutional network within pyramid pooling (FCN-PP) to obtain better LM results. The

proposed FCN-PP is able to capture wider receptive field, it eventually overcomes the drawback of single-scale pooling. Fig. 1 shows the proposed network structure.
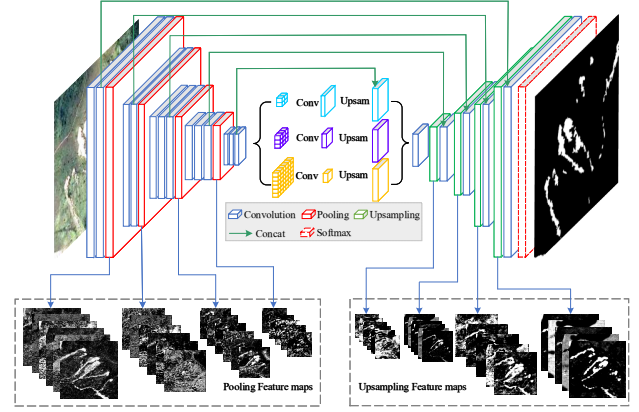

Fig. 1: Structure of FCN-PP. It is a symmetric fully convolutional network that includes four pooling layers and four upsampling layers.

In Fig. 1, the FCN-PP similarly yields a U-shape architecture that includes four pooling layers and corresponding four upsampling layers. It also has a elegant architecture structuring because pooling layers on the left and upsampling layers on the right are symmetric. Furthermore, the pyramid pooling module is integrated into FCN-PP to overcome the global pooling problem.

We choose a three-level pyramid pooling module that includes three different scales (convolutional kernels: 5×5, 10×10, and 15×15; strides: 5, 10, and 15), where the first scale (5×5) is marked by cyan color, the second and the third scale (10×10, 15×15) marked by purple and yellow, respectively. Then, we use 1×1 convolution to reduce the dimension of the three different-size feature maps, and then implement the operation of upsampling. Here, bilinear interpolation is used for upsampling to obtain feature maps of the same size as the original feature map. The final output of pyramid module is a fusion result of multi-level feature maps. The average pooling is chosen in the pyramid module as it provides better global information than the max-pooling.

2.2. The analysis of FCN-PP

To verify the validity of the pyramid pooling module for landslide feature learning, we extract the outputs from convolutional layer at different scales. Fig. 2 shows that a large convolutional kernel means a wider received field that is helpful for global feature representation, while, a small convolutional kernel means a narrower received field that is helpful for local feature representation. We combine different scaled features to achieve a stronger feature representation than single-scale feature learning. As landslide areas have a serious spatial un-

certainty, it is difficult to learn effective landslide features. The pyramid pooling module is able to address this difficulty and it is suitable for feature learning of landslide areas.
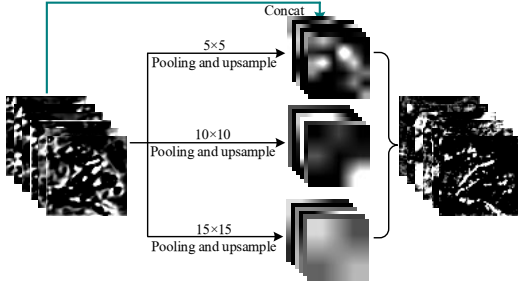

Fig. 2: Pyramid pooling module.

In Fig. 2, the input is the same as feature maps of FCN, which are pooled by different-size convolutional kernels. Then these results are upsampled and fused with the input feature maps, leading to a final output with more accurate localization and better semantic information than input feature maps of FCN. Consequently, the proposed FCN-PP achieves a tradeoff between the use of context and the localization accuracy, making a significant improvement over the CNN approach in LM. Moreover, the elegant architecture leads to the requirement of a small number of training samples.

3. EXPERIMENTS

The proposed FCN-PP is compared with two groups of popular approaches, i.e., traditional approaches (ELSE [5], RLSE [5], CDMRF [6], and CDFFCM [7]) and deep learning approaches (CNN [14], FCN [16], and U-Net [17]), for LM. The first group of comparative approaches is implemented with MATLAB 2017b and the second group is implemented with PyTorch on a workstaion with Intel Xeon CPU E5-1620v4, 3.5GHz, 4 cores, 64GB RAM, double NVIDIA GTX 1080 GPU. In our experiments, the parameter values of the first group of comparative approaches follow the original papers. In the proposed FCN-PP, the stochastic gradient descent with a constant learning rate of 1×10^{-5}, weight decay of 5×10^{-4}, momentum of 0.99, mini-batch size of 4, and epochs of 30 were used to train the proposed network. In addition, the structure element in MMR is a disk of size 1×1. In order to avoid the problem of overfitting, the corresponding parameters before pyramid pooling module are initialized by the pretrained parameters of the convolutional layers of VGG-16.

3.1. Data description

Five pairs of bi-temporal images on A-E areas in Hong Kong, were captured by the Zeiss RMK TOP 15 Aerial Survey Camera System at a flying height of approximately 2.4km on December 2007 and on November 2014, respectively [3]. Due

to the geometrical resolution of bi-temporal images is 0.5m, the captured images have a large size. We cropped A-E areas to obtain five interesting areas. The size of A-E areas are 750×950, 1252×2199, 923×593, 1107×743, 826×725, respectively. Because it is impossible to build a large dataset of bitemporal landslide images, we built a small dataset that is considered as training set in this paper. To distinguish the training data and testing data, we first cropped three typical areas from A-C areas, where different kinds of landslide occur. The rest areas in A-C and all areas in D-E are considered as the training data. As a result, we obtained 139 training images that are overlapped. To increase the training data, each image is rotated, horizontally and vertically flipped, sheared, and scaled of its original size but the final image size is the original size. Finally, we got 1,112 training image pairs of size 473×473 and 3 testing image pairs.
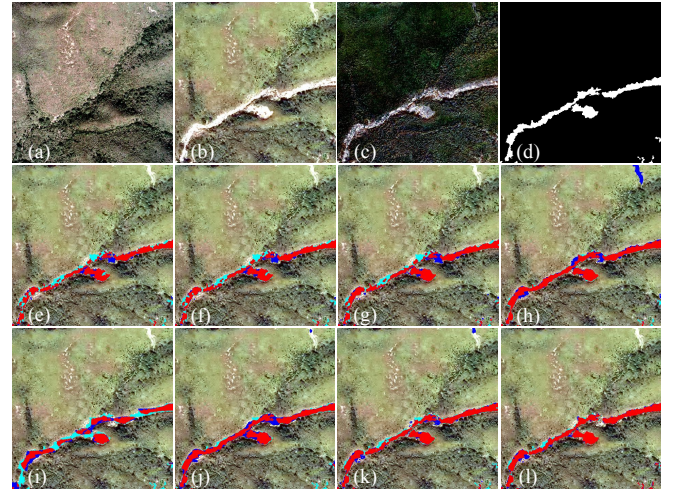

Fig. 3: Comparison of results on the A-area using different approaches, where the red areas are true-detected landslides, the cyan areas are false-detected landslides, and the blue areas are true-missed landslides. (a) The pre-event image. (b) The post-event image. (c) The difference image. (d) The ground-truth. (e) ELSE. (f) RLSE. (g) CDMRF. (h) CDFFCM. (i) CNN. (j) FCN. (k) U-Net. (l) FCN-PP.

3.2. Results and analysis

Due to the limitation of length of the paper, we only present the comparative results of A-area and B-area. Figs. 3 (a-b) and Figs. 4 (a-b) show two pairs of bi-temporal images. We can see that the landslide areas are simple and continuous in Fig. 3d but they are complex and discontinuous in Fig. 4d. Therefore, it is more difficult to extract the landslide areas in Fig. 4b than in Fig. 3b by using traditional approaches. Fig. 3c and Fig. 4c show the difference images of pre-event and post-event images. It is clear that each difference image includes lots of noises that influence the de-
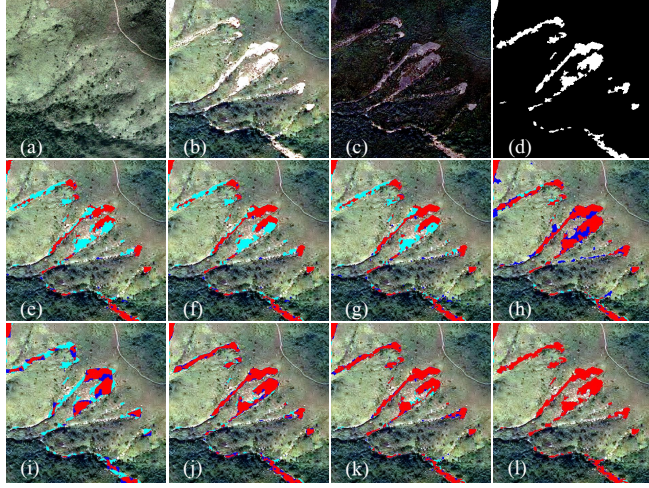
Fig. 4: Comparison of results on the B-area using different approaches, where the red areas are true-detected landslides, the cyan areas are false-detected landslides, and the blue areas are true-missed landslides. (a) The pre-event image. (b) The post-event image. (c) The difference image. (d) The ground-truth. (e) ELSE. (f) RLSE. (g) CDMRF. (h) CDFFCM. (i) CNN. (j) FCN. (k) U-Net. (l) FCN-PP.

Table 1. QUANTITATIVE RESULTS FOR A-AREA. LARGER VALUES ARE BETTER FOR $Precision$, $Recall$, $F\text{-}score$, AND $Accuracy$. THE BEST VALUES ARE IN BOLD.

Methods	$Precision$	$Recall$	$F\text{-}score$	$Accuracy$
ELSE	86.79	74.02	79.90	66.53
RLSE	86.61	75.87	80.89	67.91
CDMRF	84.89	79.15	81.92	69.38
CDFFCM	79.30	91.01	84.75	73.54
CNN	75.47	62.16	68.17	51.71
FCN	79.10	88.45	83.51	71.69
U-Net	85.65	83.92	84.78	73.58
FCN-PP	**89.84**	**90.32**	**90.08**	**81.95**

Table 2. QUANTITATIVE RESULTS FOR B-AREA. LARGER VALUES ARE BETTER FOR $Precision$, $Recall$, $F\text{-}score$, AND $Accuracy$. THE BEST VALUES ARE IN BOLD.

Methods	$Precision$	$Recall$	$F\text{-}score$	$Accuracy$
ELSE	91.24	51.29	65.67	48.89
RLSE	91.14	56.00	69.37	53.10
CDMRF	87.42	60.01	71.17	55.24
CDFFCM	73.52	88.42	80.28	67.06
CNN	71.44	51.44	59.82	42.67
FCN	83.94	84.39	84.17	72.66
U-Net	87.88	87.93	87.91	78.42
FCN-PP	**96.03**	**95.62**	**95.82**	**91.99**

tection of the true landslides areas. Figs. 3 (e-h) and Figs. 4 (e-h) show landslide areas detected by four conventional approaches, ELSE, RLSE, CDMRF, and CDFFCM. Because ELSE, RLSE, and CDMRF employ general image segmentation models to achieve LM, they are sensitive to noise. The detected landslide areas include lots of discontinuous areas that are continuous in ground truths. Although CDFFCM addresses the problem by using image filtering and improved FCM algorithm that incorporates spatial information of images into its objective function, some false landslide areas are detected as shown in Fig. 3h and Fig. 4h. Compared with unsupervised learning approaches, CNN is able to capture the semantic information of landslide areas, but the detected areas are coarse as shown in Fig. 3i and Fig. 4i. FCN provides better results than CNN and four unsupervised learning approaches. However, the detail of landslide areas are smoothed in Fig. 3i and Fig. 4i, since the global pooling is adopted by FCN. The proposed FCN-PP provides the best results that are similar to ground truths.

For quantitative evaluation of the proposed FCN-PP, we compare experimental results with ground truths according to four performance indices: $Precision$, $Recall$, $F\text{-}score$, $Accuracy$ [18, 19]. The experimental results are shown in Tables 1-2. It can be seen that $Precision$ and $Recall$ are inconsistent for the evaluation of results. The ELSE, RLSE, and CDMRF obtain high $Precision$ but low $Recall$ values, while the CDFFCM obtains low $Precision$ but high $Recall$ values. Considering that $F\text{-}score$, and $Accuracy$ evaluate the overall performance of an approach, CDFFCM is superior to

ELSE, RLSE, and CDMRF. Amongst the existing approaches based on deep learning, CNN provides low performance indices, FCN is superior to CNN but worse than FCN. What is abundantly clear that the proposed FCN-PP obtains the best performance indices as shown in Tables 1-2. From Figs. 3-4 and Tables 1-2, we can see that FCN-PP can provide excellent LM results, since the deep convolutional features overcomes the spatial uncertainty of landslides and the pyramid pooling addresses the difficulty of accurate localization of landslides.

4. CONCLUSION

In this work, we have proposed a fully convolutional network within pyramid pooling (FCN-PP) for LM. The FCN-PP adopts multiple layer connection to incorporate the low- and high-dimensional features into the final feature map. The pyramid pooling module is integrated into the FCN-PP and is able to efficiently exploit the spatial multiscale features of landslide areas, which addresses the drawback of single-scale pooling and thus outputs a better feature map with stronger feature representation capability than CNN, FCN, and U-Net. Experimental results show that the proposed FCN-PP generates satisfactory LM results without hard-tuning parameters, the FCN-PP clearly outperforms state-of-the-art approaches for LM.

5. REFERENCES

[1] F. Dai, C. Lee, and Y. Ngai, "Landslide risk assessment and management: an overview," *Eng Geol.*, vol. 64, no. 1, pp. 65-87, 2002.

[2] M. Gong, L. Su, M. Jia, and W. Chen, "Fuzzy clustering with a modified MRF energy function for change detection in synthetic aperture radar images," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 1, pp. 98-109, 2014.

[3] Z. Lv, W. Shi, X. Zhang, and, J. Benediktsson, "Landslide inventory mapping from bitemporal high-resolution remote sensing images using change detection and multiscale segmentation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 5, pp. 1520-1532, 2018.

[4] Y. Jin, "Change detection of enhanced, no-changed and reduced scattering in multi-temporal ERS-2 SAR images using the two-thresholds EM and MRF algorithms,"*IEEE Geosci. Remote Sens. Symp.*, vol. 6, pp. 3994-3997, 2005.

[5] Z. Li, W. Shi, S. Myint, P. Lu, and Q. Wang, "Semi-automated landslide inventory mapping from bitemporal aerial photographs using change detection and level set method," *Remote Sens. Environ.*, vol. 175, pp. 215-230, 2016.

[6] Z. Li, W. Shi, P. Lu, L. Yan, Q. Wang, and Z. Miao, "Landslide mapping from aerial photographs using change detection-based Markov random field," *Remote Sens. Environ.*, vol. 187, pp. 76-90, 2016.

[7] T. Lei, D. Xue, Z. Lv, S. Li, Y. Zhang, and A. K. Nandi, "Unsupervised Change Detection Using Fast Fuzzy Clustering for Landslide Mapping from Very High-Resolution Images," *Remote Sens.*, vol. 10, no. 9, pp. 1381, 2018.

[8] Q. Wang, Z. Yuan, Q Du, and X. Li, "GET-NET: A general end-to-end two-dimensional CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, 2018. Doi: 10.1109/TGRS.2018.2849692.

[9] X. Niu, M. Gong, T. Zhan, and Y. Yang, "A Conditional Adversarial Network for Change Detection in Heterogeneous Images," *IEEE Trans. Geosci. Remote Sens. Lett.*, 2018. Doi: 10.1109/LGRS.2018.2868704.

[10] M Gong, T Zhan, P Zhang, et al, "Superpixel-based difference representation learning for change detection in multispectral remote sensing images[J]," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2658-2673, 2017.

[11] T. Lei, X. Jia, Y. Zhang, L. He, H. Meng, and A. K. Nandi, "Significantly fast and robust fuzzy c-means clustering algorithm based on morphological reconstruction and membership filtering," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 3027 - 3041, 2018.

[12] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Trans. Neural Netw. Learn Syst.*, vol. 27, no. 1, pp. 125-138, 2016.

[13] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE Trans. Neural Netw. Learn Syst.*, vol. 29, no. 3, pp. 545-559, 2018.

[14] A. Ding, Q. Zhang, X. Zhou, and B. Dai, "Automatic recognition of landslide based on CNN and texture change detection," in *Proc. 31st Youth Acad. Annu. Conf. Chinese Assoc. Autom.*, 2016, pp. 444-448.

[15] T. Lei, Y. Zhang, Y. Wang, S. Liu, and Z. Guo, "A conditionally invariant mathematical morphological framework for color images," *Inform. Sci.*, vol. 387, pp. 34-52, 2017.

[16] J. Long, E. Shelhamer, and T. Darrell,"Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2015, pp. 3431-3440.

[17] O. Ronneberger, P. Fischer, and T. Brox,"U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Conf. Med. Image Comput. Comput. -Assist. Intervent.*, 2015, pp. 234-241.

[18] N. Goyette, P. M. Jodoin, F.M. Porikli, J. Konrad, and P. Ishwar,"A novel video dataset for change detection benchmarking," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4663-4679, 2014.

[19] X. Zhang, P. Xiao, X. Feng, X. Feng, and M. Yuan,"Separate segmentation of multi-temporal high-resolution remote sensing images for object-based change detection in urban area," *Remote Sens Environ.*, vol. 201, pp. 243-255, Nov. 2017.