IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-2010-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IGARSS52108.2023.10281681

M2-CDNET: A MULTI-SCALE AND MULTI-LEVEL NETWORK FOR REMOTE SENSING IMAGE CHANGE DETECTION

Qiong Wu [1], Zhi Zheng [1], Yi Wan [1], [*], Yongjun Zhang [1]

[1] School of Remote Sensing and Information Engineering, Wuhan University, 430079, China
[*]Corresponding author: yi.wan@whu.edu.cn

ABSTRACT

Change detection plays a crucial role in environmental monitoring and earth observation tasks, leveraging the abundant data acquired by remote sensing platforms. While learning-based methods have shown promise in strictly registered datasets, their practical applicability in real-world scenarios remains challenging. This paper addresses the limitations of existing methods by proposing M2-CDNet, a novel approach that integrates the U-Net architecture with the multi-scale fusion (MSF) strategy, deformable convolutions, and multi-scale outputs. Experiments on the public and self-collected datasets demonstrate that M2-CDNet achieves superior accuracy-efficiency trade-offs compared to state-of-the-art methods. Moreover, M2-CDNet shows better robustness against image projection bias and registration errors.

Index Terms— Change Detection (CD), accuracy-efficiency trade-off, multi-scale features

1. INTRODUCTION

With the advancement of remote sensing (RS) technology, RS platforms have significantly enhanced their capacity to acquire vast data. These abundant datasets have become invaluable resources for environmental monitoring, specifically in detecting changes occurring on the Earth's land surface [1]. Change detection plays a crucial role in many earth observation tasks, including land-use monitoring, disaster detection, and resource management [2][3]. Over the past few decades, change detection has been extensively investigated using various methods, including numerous learning-based methods. However, most deep learning-based methods excel in strictly registered datasets but struggle in real-world practical applications. Apart from the domain generalization ability, we have identified two additional factors that can hinder the practical applicability of these methods. The first factor is the trade-off between model accuracy and efficiency. Many models prioritize accuracy improvements at the expense of efficiency, greatly restricting their usefulness in time-critical tasks. Secondly, it has been observed that most methods underperform when applied to practical production images due to the adverse effects of image projection bias and registration errors.

To address these challenges, we developed M2-CDNet, a novel approach that leverages the U-Net architecture as its backbone. M2-CDNet incorporates several key components, including a multi-scale fusion (MSF) strategy, deformable convolutions, and multi-scale outputs [4]. We conducted evaluations of M2-CDNet on both a public VHR-Dataset and a dataset we collected ourselves [5]. The experimental results demonstrated that M2-CDNet achieves superior accuracy-efficiency trade-offs compared to state-of-the-art (SOTA) methods. Furthermore, M2-CDNet exhibits significant advantages in mitigating the adverse effects caused by image projection bias and registration errors.

2. RELATED WORK

Early approaches to automatic modeling of change detection heavily relied on hand-crafted features [6][7]. In recent years, a plethora of deep models based on neural networks have emerged in the field. Among these models, the fully convolutional neural network has gained significant popularity [8][9]. Many researchers designed various strategies to enhance the accuracy of change detection result. Daudt et al. employed a siamese structure for efficient feature extraction, concatenating the extracted features from the bitemporal images and using typical decoder to predict a change mask [10]; Lei et al. embedded the pyramid pooling into the FCN backbone to overcome the drawbacks of global pooling strategies, enlarging the network's receptive field [11]; Zheng et al. designed a cross-layer block to better correlate multi-scale features and multi-level information.

However, it needs to be noticed that these methods are only trained and tested for strictly registered datasets, so they will inevitably face challenges in practical applications. To address this problem, Zheng et al. over-emphasize the effectiveness while ignoring model's ability to resist projection errors. Esfandiari et al. employed patch-wise co-registration with the digital surface model for image registration [12]. Since this method necessitates additional data preprocessing, it may result in error accumulation. In this paper, we proposed M2-CDNet to directly modify the current networks and enhance learning-based methods' robustness.

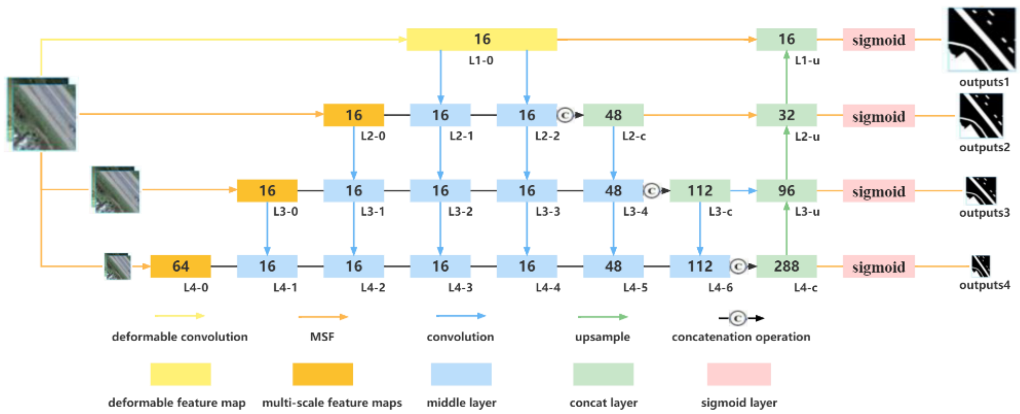


Figure 1. Architecture overview of the proposed M2-CDNet.

3. METHODOLOGY

The detailed structure of M2-CDNet is as shown in Figure 1. Since the multi-scale feature extraction strategy is proven to be effective for accuracy-efficiency trade-offs, we modified MSF structure in EMS-CDNet for the feature extraction module of M2-CDNet. Multi-scale features are extracted by the first two branches of MSF module using dilation convolutions with different rates and strides. The last two branches first resize raw images to their 1/8 and 1/16 counterparts and then transform the resized images into feature domain by 1x1 convolutions. Since the deformable convolution has ability to learn the offsets between given features, we introduced it into M2-CDNet to alleviate pixel displacement between images caused by image projection bias and registration errors. Deformable convolution adds 2D offsets to the regular grid sampling locations in the standard convolution. It enables free form deformation of the sampling grid. In addition, deformable convolution only introduces a few extra parameters to learn the offset and thus it can improve M2-CDNet's robustness to image projection bias and registration errors without efficiency loss. To refine the results and further alleviate the image registration errors, we proposed multi-scale outputs structure, outputting four results from different scales. It makes sure that M2-CDNet takes account of multi-scale features when calculating loss function value. Features extracted from different scales convey dissimilar information. As the scale shrinks, the side effects of image projection bias and registration errors are gradually weakened. The loss function of M2-CDNet derived from CLNet, which is

$$E_i = E_{bcei} + 0.5 * E_{dci}$$

Since M2-CDNet has four outputs, we combined them for the overall loss function, which is as follows:

$$E = \lambda_1 E_1 + \lambda_2 E_2 + \lambda_3 E_3 + \lambda_4 E_4$$

Where i represents the order of output results; E_i is the loss value of i output result; E_{bcei} and E_{dci} represent Binary cross-entropy loss and dice coefficient loss of i output result. λ_1, λ_2, λ_3 and λ_4 in the overall loss function (E) was set as 0.3, 0.5, 0.7 and 1.0 respectively.

4. EXPERIMENTAL RESULTS

We conducted experiments on VHR-Dataset and our collected dataset to assess M2-CDNet's performance. We selected FC-EF, SNUNet-CD and Unet++_MSOF as the comparison methods and use f1-score (F1), overall accuracy (OA), mean intersection over union (mIoU) and inference time (T) for accuracy and efficiency evaluation [13][14].

Figure 2(A) and Table 1 displayed the experimental results on VHR-dataset. FC-EF had the worst results, while SNUNet-CD and Unet++_MSOF achieved superior performance. However, M2-CDNet achieved the highest F1 value among these methods. As for OA and mIoU, M2-CDNet obtained similar values to Unet++_MSOF while ensuring efficiency. Our experimental results not only demonstrated that M2-CDNet achieved SOTA performance on the semantic change detection tasks, but also indicated the effectiveness of M2-CDNet for complex feature representation.

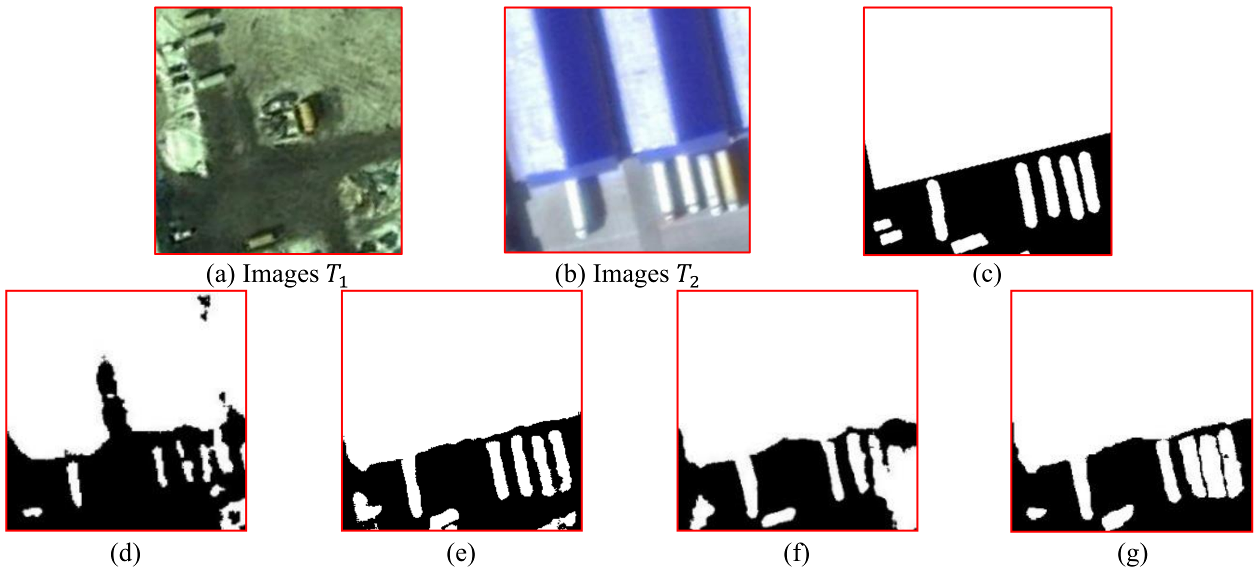
Figure 2(B) and Table 1 displayed the experimental results on our collected images. The results of all the compared methods significantly decreased compared to their performance on public dataset. However, M2-CDNet outperformed the other compared methods, especially in F1. Our experimental results demonstrated that M2-CDNet was more robust to image registration errors, which verified the effectiveness of introducing deformable convolution and multi-scale outputs.

5. CONCLUSION AND FUTURE WORK

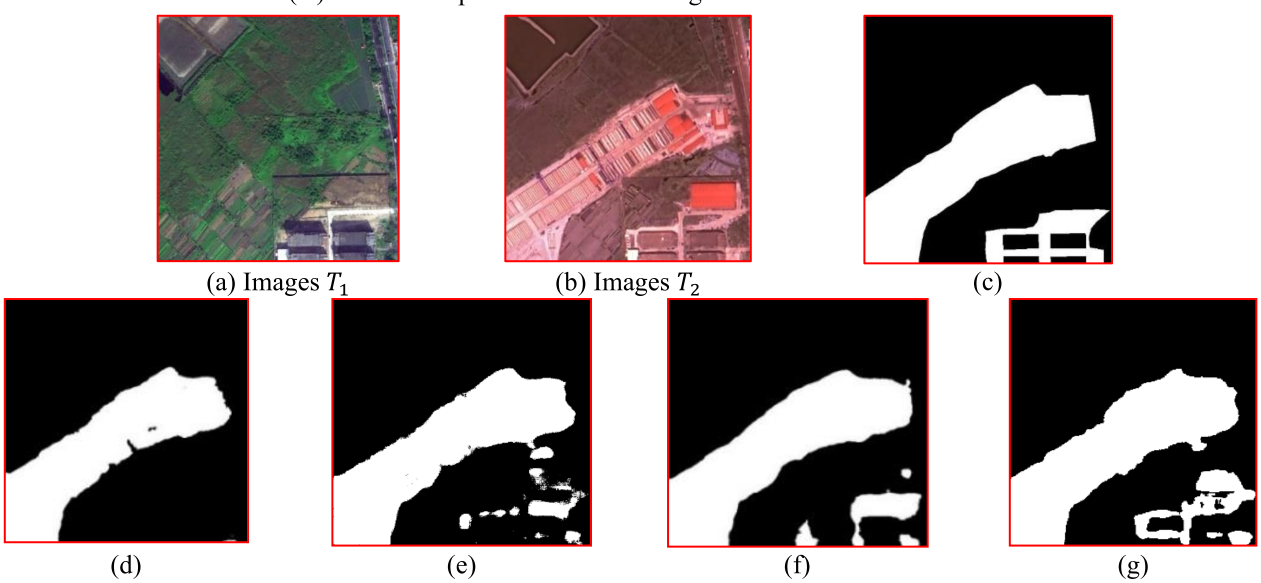
In this paper, we propose the M2-CDNet network to enhance the ability of deep learning change detection methods for practical applications. Our method integrates the MSF strategy, deformable convolution, and a multi-scale output structure into the network. This allows our method to leverage multi-scale features and overcome pixel displacement to a certain extent.

Table 1. Quantitative analysis of two datasets (Best results are emphasized in bold)

Dataset	VHR Dataset				Our Collected Images			
Methods	*F1*	*OA*	*mIoU*	*Time(s)*	*F1*	*OA*	*mIoU*	*Time(s)*
FC-EF	0.810	0.956	0.820	656	0.550	**0.970**	0.670	227
SNUNet-CD	0.911	0.967	0.902	**583**	0.534	0.961	0.662	**210**
Unet++_MSOF	0.910	**0.978**	**0.907**	3342	0.670	0.960	0.730	403
M2-CDNet	**0.927**	0.959	0.905	617	**0.725**	0.945	**0.754**	239



(A) Visual comparisons for the testing area in the VHR dataset.



(B) Visual comparisons for the testing area in the collected images.

Figure 2. Visual comparisons for the testing area in two datasets. (a) images T_1; (b) images T_2; (c)ground truth change map; (d)-(g) change maps of FC-EF, SNUNet-CD/16, Unet++_MSOF and M2-CDNet, where the changed pixels are labeled in white and the unchanged pixels are in black.

The experimental results demonstrate that M2-CDNet achieves improved accuracy-efficiency trade-offs compared to state-of-the-art methods. Furthermore, our method exhibits greater robustness to image projection bias and registration errors.

For future work, we aim to further enhance the model's performance in practical applications. We will focus on strengthening its capabilities to handle various real-world challenges and improve its adaptability in different scenarios.

6. ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China under project number 42030102 and the Fund for Innovative Research Groups of the Hubei Natural Science Foundation under project number 2020CFA003.

7. REFERENCES

[1] Shafique, Ayesha, et al. "Deep learning-based change detection in remote sensing images: A review." Remote Sensing 14.4 (2022): 871.

[2] Zheng, Zhi, et al. "EMS-CDNet: an efficient multi-scale-fusion change detection network for very high-resolution remote sensing images." International Journal of Remote Sensing 43.14 (2022): 5252-5279.

[3] Zheng, Zhi, et al. "CLNet: Cross-layer convolutional neural network for change detection in optical remote sensing imagery." ISPRS Journal of Photogrammetry and Remote Sensing 175 (2021): 247-267.

[4] Dai, Jifeng, et al. "Deformable convolutional networks." Proceedings of the IEEE international conference on computer vision. 2017.

[5] Lefebvre, Antoine, Thomas Corpetti, and Laurence Hubert-Moy. "Object-oriented approach and texture analysis for change detection in very high resolution images." IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium. Vol. 4. IEEE, 2008.

[6] Dellinger, Flora, et al. "Change detection for high resolution satellite images, based on SIFT descriptors and an a contrario approach." 2014 IEEE Geoscience and Remote Sensing Symposium. IEEE, 2014.

[7] Padrón-Hidalgo, José A., et al. "Kernel anomalous change detection for remote sensing imagery." IEEE Transactions on Geoscience and Remote Sensing 57.10 (2019): 7743-7755.

[8] Bu, Shuhui, et al. "Mask-CDNet: A mask based pixel change detection network." Neurocomputing 378 (2020): 166-178.

[9] Chen, Baohua, et al. "Building change detection with RGB-D map generated from UAV images." Neurocomputing 208 (2016): 350-364.

[10] Daudt, Rodrigo Caye, Bertr Le Saux, and Alexandre Boulch. "Fully convolutional siamese networks for change detection." 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018.

[11] Lei, Tao, et al. "Landslide inventory mapping from bitemporal images using deep convolutional neural networks." IEEE Geoscience and Remote Sensing Letters 16.6 (2019): 982-986.

[12] Esfandiari, Morteza, et al. "Building Change Detection in Off-Nadir Images Using Deep Learning." 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. IEEE, 2021.

[13] Fang, Sheng, et al. "SNUNet-CD: A densely connected Siamese network for change detection of VHR images." IEEE Geoscience and Remote Sensing Letters 19 (2021): 1-5.

[14] Peng, Daifeng, Yongjun Zhang, and Haiyan Guan. "End-to-end change detection for high resolution satellite images using improved UNet++." Remote Sensing 11.11 (2019): 1382.