Advances in Computer, Signals and Systems (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2024.080211
ISSN 2371-8838 Vol. 8 Num. 2

Change Detection in Images with Viewpoint Difference

Yaxin Dong[a], Yang Yang[b,*]

The Laboratory of Pattern Recognition and Artificial Intelligence, Yunnan Normal University, Kunming, China
[a]dongyaxin9830@163.com, [b]yyang_ynu@163.com
[*]Corresponding author

Keywords: Deep learning, change detection, optical flow estimation, viewpoint difference

Abstract: Change detection plays a crucial role in identifying differences between multi-temporal images captured over the same geographical area, with applications spanning various fields including urban planning, environmental monitoring, and disaster assessment. However, challenges persist in handling bitemporal images with viewpoint difference, affecting the performance of traditional change detection models. To address these challenges, this paper proposes a novel end-to-end optical flow alignment change detection (PFCD) model. The PFCD model integrates optical flow estimation technology, enabling direct change detection in images with viewpoint differences without the need for a separate image registration model. Through end-to-end training, the model achieves higher detection accuracy and faster processing speeds. Experimental results on the scattered garbage regions change detection dataset (SGRCD-VD) and the building change detection dataset (WHUCD-VD) validate the effectiveness of the model. On the SGRCD-VD test set, the PFCD model achieves an F1 score of 91.00%, while on the WHUCD-VD test set, it reaches 94.82%, demonstrating excellent performance in handling images with viewpoint differences. Additionally, the model exhibits advantages in processing speed and model parameter.

1. Introduction

Change detection aims to identify differences between multi-temporal images captured over the same geographical area. Its applications span various domains including urban planning, environmental monitoring, disaster assessment, and more. Over the past few decades, with the development of deep learning technologies, change detection has witnessed rapid progress, leading to the emergence of numerous change detection methods. However, despite the rapid advancement in change detection techniques, detecting changes in images with viewpoint differences remains a significant challenge for most change detection models.

When acquiring bitemporal images using devices such as small unmanned aerial vehicles (UAVs), vehicle-mounted cameras, or remote sensing satellites, differences in trajectory, camera viewpoint, device orientation, and complex geographical environments inevitably lead to viewpoint difference in the acquired bitemporal images. Traditionally, the impact of viewpoint differences has been overlooked, and many existing change detection networks have been designed without considering the viewpoint disparities between bitemporal images. This is because image registration models are typically employed to eliminate such disparities before performing change detection. However, this approach suffers from issues such as low detection accuracy and slow processing speed. Errors introduced during the process of image registration can propagate to the process of change detection, resulting in greater errors and worse performance. Moreover, both image registration and change detection models require feature extraction from images, resulting in slow processing speeds and making them unsuitable for applications requiring high-speed processing.

To address these issues, this paper proposes a new change detection network, PFCD. The model integrates optical flow estimation, enabling direct detection of changes in images with viewpoint differences without the need for separate image registration models. Through end-to-end training, the model achieves higher detection accuracy. By directly aligning pairs of feature maps where there is a difference in viewing angle, redundant feature extraction in the image is avoided, thus increasing the processing speed.

2. Related Work

This section elaborates on the rationale behind our proposed change detection method and the relevant work.

Change detection: Over the past few decades, a plethora of change detection methods have emerged. However, many of these methods failed to initially consider the impact of viewpoint differences in bitemporal images on change detection performance. For example, Fang et al.[1] introduced SNUNet, a densely connected siamese network for change detection. SNUNet addresses the issue of deep localization information loss by concatenating multi-level features. Similarly, Chen et al.[2] proposed BitCD, a bitemporal image transformer change detection network for remote sensing images. BitCD enhances context information of image features through a transformer and derives the change map from feature differences.

However, some methods that did consider viewpoint differences during design exhibited poor performance when handling images with significant viewpoint disparities. For instance, Sakurada et al.[3] presented a novel model for weakly supervised semantic scene change detection, which addresses camera viewpoint differences in vehicular images using correlation layers. Nevertheless, due to the lack of effective supervision for correcting viewpoint differences during network training, the performance of this network diminishes in bitemporal images with substantial viewpoint disparities.

Optical flow estimation: Optical Flow refers to the movement of light. It describes the motion of each pixel in an image and can be represented as a vector field, where each vector indicates the direction and speed of motion for the corresponding pixel. Pixel A is located at position (xt, yt) at time t, and it moves to position (x(t+1), y(t+1)) at time t+1. The motion vector of pixel A, (dx, dy) = (x(t+1), y(t+1)) - (xt, yt), represents the optical flow of pixel A.

In recent years, there has been rapid development in the field of optical flow estimation, driven by the advancement of convolutional neural networks (CNNs). Dosovitskiy et al.[4] were pioneers in this area, introducing CNNs to optical flow estimation and proposing two network structures named FlowNetSimple and FlowNetCorr. Ilg et al.[5] further advanced the field with FlowNet2, a model composed of multiple stacked instances of FlowNetSimple and FlowNetCorr. Additionally, Ranjan and Black[6] introduced SpyNet, which combines classical spatial-pyramid formulations with deep learning techniques for optical flow estimation. Subsequently, notable contributions have been made with networks like PWCnet[7] and LiteFlowNet[8]. These networks utilize correlation layers to estimate optical flow at each level of the feature pyramid and subsequently warp the features of each level based on the estimated optical flow, resulting in more compact and efficient networks.

3. Methodology

To address the challenges of poor performance and slow processing speed in change detection models when handling images with viewpoint disparities, we propose an optical flow alignment model called PFCD. The PFCD model, depicted in Figure 1, comprises three modules: the feature extraction module, the optical flow alignment module, and the change detection module. In the feature extraction module, we employ the ResNet[9] network as the backbone to extract feature maps of varying sizes, depths, and semantic information from the bitemporal images. The optical flow alignment module is responsible for aligning feature maps with viewpoint disparities to eliminate the existing differences. Subsequently, the change detection module detects changes in target objects from the aligned feature maps, ultimately generating feature difference maps. In this section, we will provide a detailed description of our proposed method and the training strategy.

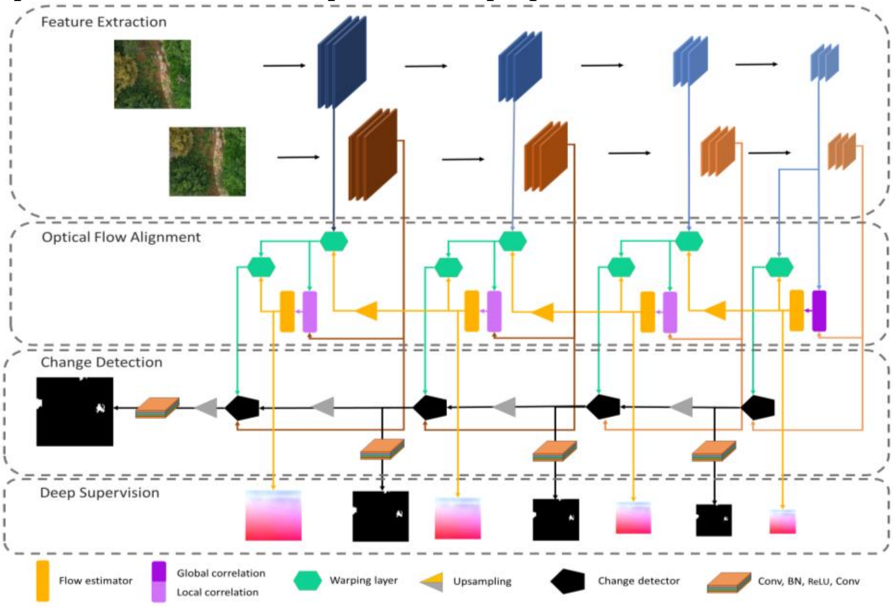


Figure 1: Details of PFCD network structure.

3.1. Feature Extraction Module

ResNet18[9] is composed of 18 convolutional layers and introduces a novel concept of residual connections, leading to remarkable performance across various tasks. Therefore, we opt to utilize ResNet18 as the backbone network for feature extraction in our model. Through the ResNet18[9] model, the feature extraction module conducts four 2x downsampling operations on the input image. The resulting feature maps after each downsampling step are utilized for subsequent module operations. This downsampling process effectively reduces the resolution of the feature maps, thereby decreasing the time required for the feature map alignment by the optical flow alignment module. Additionally, the feature maps of different sizes and depths encompass both rich semantic information from deeper layers and accurate positional information from shallower layers. This combination contributes significantly to enhancing the accuracy of the change detection task.

3.2. Optical Flow Alignment Module

The optical flow alignment module plays a pivotal role in overcoming the challenges faced by change detection models when processing images with viewpoint differences. The optical flow alignment module consists of global and local correlations, an optical flow estimator, and a warping layer. Its primary function is to align image feature maps, which is crucial for improving the accuracy and efficiency of change detection tasks.

To accomplish this, the module employs a multi-step process. Initially, it establishes dense correspondences between pairs of image feature maps by computing either local or global correlations. These correlations provide valuable information about the relationships between pixels, enabling the model to estimate optical flow more accurately.

The optical flow estimation process involves using the optical flow estimator composed of multiple layers of convolutional operations. This estimator analyzes the correlations obtained from the previous step to infer the displacement of pixels between the two images. By leveraging advanced convolution techniques, it effectively captures the intricate patterns and motion information present in the image data.

Once the optical flow is estimated, the next step involves applying a warping layer. This layer deforms the feature map of one image to align it with the corresponding feature map of the other image. Through this deformation process, the module ensures that features representing the same spatial locations in both images are brought into alignment, facilitating accurate comparison and analysis.

Overall, the optical flow alignment module integrates advanced techniques in dense correspondence establishment, optical flow estimation, and feature map alignment to effectively address the challenges posed by viewpoint differences in change detection tasks. Its robust functionality enhances the model's ability to detect and analyze changes in diverse environmental conditions, making it a valuable component in change detection systems.

3.3. Change Detection Module

The change detection module comprises multiple layers of convolution. Within this module, the feature map of image A is subtracted from that of image B, and the absolute value of the result is taken. Next, the small-sized feature difference map is upsampled by a factor of 2 and fused with the absolute value result. Then, the feature difference map undergoes processing through multiple layers of convolution. Finally, the feature difference map is further processed using two additional convolutional layers to generate the change map.

3.4. Loss Function

To improve the performance of the model in detecting image changes with viewpoint differences, we adopted a deep supervision strategy, which involves supervising the intermediate results of the model to enhance training effectiveness. In the optical flow alignment module, we supervised the optical flow generated from feature maps of different sizes using endpoint error loss, and then combined them proportionally. In the change detection module, we supervised the change map generated by the classifier from feature difference maps of different sizes using binary cross-entropy loss, and then combined them proportionally. Finally, we summed up the supervised losses in the optical flow alignment module and the change detection module as the loss function for model training.

4. Experiments

4.1. Datasets

The scattered garbage regions change detection dataset (SGRCD-VD) and the building change detection dataset (WHUCD-VD) with viewpoint differences are utilized for model validation in this study. Each dataset comprises 50,000 samples of size 256 × 256 × 3. The datasets are split into training, testing, and validation sets in a ratio of 70%, 15%, and 15%, respectively.

4.2. Implement details

The model is implemented using the PyTorch 1.8 framework and optimized with the Adam optimizer. Training is conducted on an NVIDIA GeForce RTX 3090 for 100 epochs. The model is evaluated after each training epoch, and the best-performing model on the validation set is selected to evaluate the test set. In model evaluation, F1-score will be used as the primary evaluation metric, while overall accuracy (OA), intersection over union (IoU), precision (P), and recall (R) will serve as auxiliary evaluation metrics.

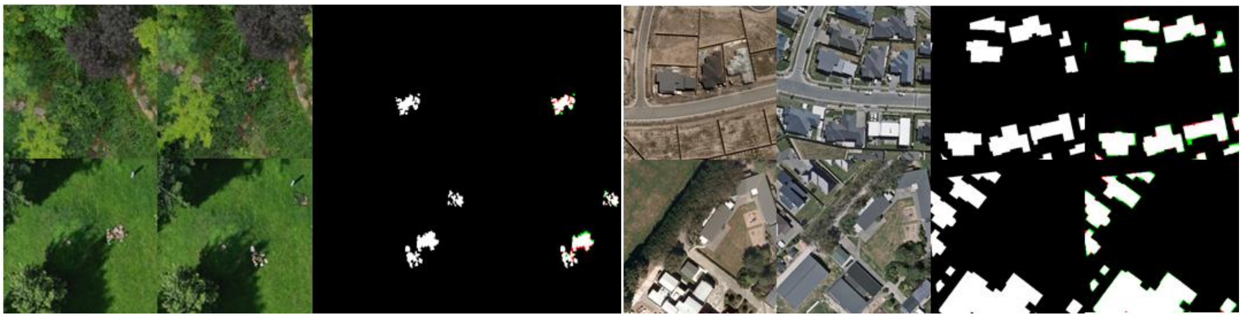
4.3. Results on SGRCD-VD and WHUCD-VD

From Table 1, it can be observed that the proposed PFCD model performs well on the SGRCD-VD and WHUCD-VD datasets. Specifically, on the SGRCD-VD test set, PFCD achieves an F1 score of 91.00%, while on the WHUCD-VD test set, PFCD achieves an F1 score of 94.82%. This demonstrates that the proposed PFCD model can effectively handle images with viewpoint differences and excels in both processing speed and model parameters. Notably, the model parameters are only 12.56M, while the image processing speed reaches 84.34 FPS.

Table 1: Results on SGRCD-VD and WHUCD-VD. All scores are expressed as percentages (%).

	F1	OA	IoU	P	R
SGRCD-VD	91.00	99.61	84.67	90.95	91.05
WHUCD-VD	94.82	99.27	90.54	95.67	93.99

From Figure 2, it can be seen that there is a significant viewpoint difference between the dual-temporal images, both in the SGRCD-VD and WHUCD-VD datasets. However, this does not affect the PFCD model, which can accurately detect changes occurring in images with viewpoint differences.



(a) SGRCD-VD (b) WHUCD-VD

Figure 2: Examples of predicted results. From left to right: image A, image B, ground truth, and OFACD. To enhance representation, we employ different colours: white for True Positives (TP), black for True Negatives (TN), red for False Positives (FP), and green for False Negatives (FN).

5. Conclusion

This paper proposes an end-to-end optical flow alignment change detection model, which can directly detect target changes in images with viewpoint differences. By introducing the optical flow module, the model addresses the issues of low detection accuracy and slow processing speed encountered by conventional change detection models when dealing with images with viewpoint differences. Finally, experiments on the SGRCD-VD and WHUCD-VD datasets with viewpoint differences demonstrate the suitability of PFCD for change detection tasks in images with viewpoint differences, with advantages such as low parameter count and high throughput.

References

[1] Fang S, Li K, Shao J, et al. SNUNet-CD: A densely connected Siamese network for change detection of VHR images [J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 1-5.

[2] Chen H, Qi Z, Shi Z. Remote sensing image change detection with transformers [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-14.

[3] Sakurada K, Shibuya M, Wang W. Weakly supervised silhouette-based semantic scene change detection[C]//2020 IEEE International conference on robotics and automation (ICRA). IEEE, 2020: 6861-6867.

[4] Dosovitskiy A, Fischer P, Ilg E, et al. Flownet: Learning optical flow with convolutional networks[C]//Proceedings of the IEEE international conference on computer vision. 2015: 2758-2766.

[5] Ilg E, Mayer N, Saikia T, et al. Flownet 2.0: Evolution of optical flow estimation with deep networks[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2462-2470.

[6] Ranjan A, Black M J. Optical flow estimation using a spatial pyramid network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4161-4170.

[7] Sun D, Yang X, Liu M Y, et al. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8934-8943.

[8] Hui T W, Tang X, Change Loy C L. A lightweight convolutional neural network for optical flow estimation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition .2018: 8981-8989.

[9] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.