

# MMCD: Memory-Based Multimodal Change Detection

Limeng Zhang<sup>1</sup>, Zenghui Zhang<sup>1</sup>, Juanping Wu<sup>1</sup>, Weiwei Guo<sup>2</sup>, Tao Zhang<sup>1</sup> and Wenxian Yu<sup>1</sup>

<sup>1</sup>Shanghai Jiao Tong University, <sup>2</sup>Tongji University.



SHANGHAI JIAO TONG  
UNIVERSITY

## 1. Abstract

Single-modal change detection methods based on optical or SAR images face challenges such as weather degradation and noise interference, while multimodal methods struggle with domain gaps. Inspired by temporal memory mechanism for video segmentation, this paper introduces memory into change detection and proposes MMCD. By treating change detection as a temporal problem and modeling remote sensing images as video sequences, MMCD integrates historical optical images with current SAR images to enhance detection accuracy. Experimental results on OSCD dataset show that MMCD achieves state-of-the-art performance in multimodal change detection, demonstrating its effectiveness.

## 2. Motivation

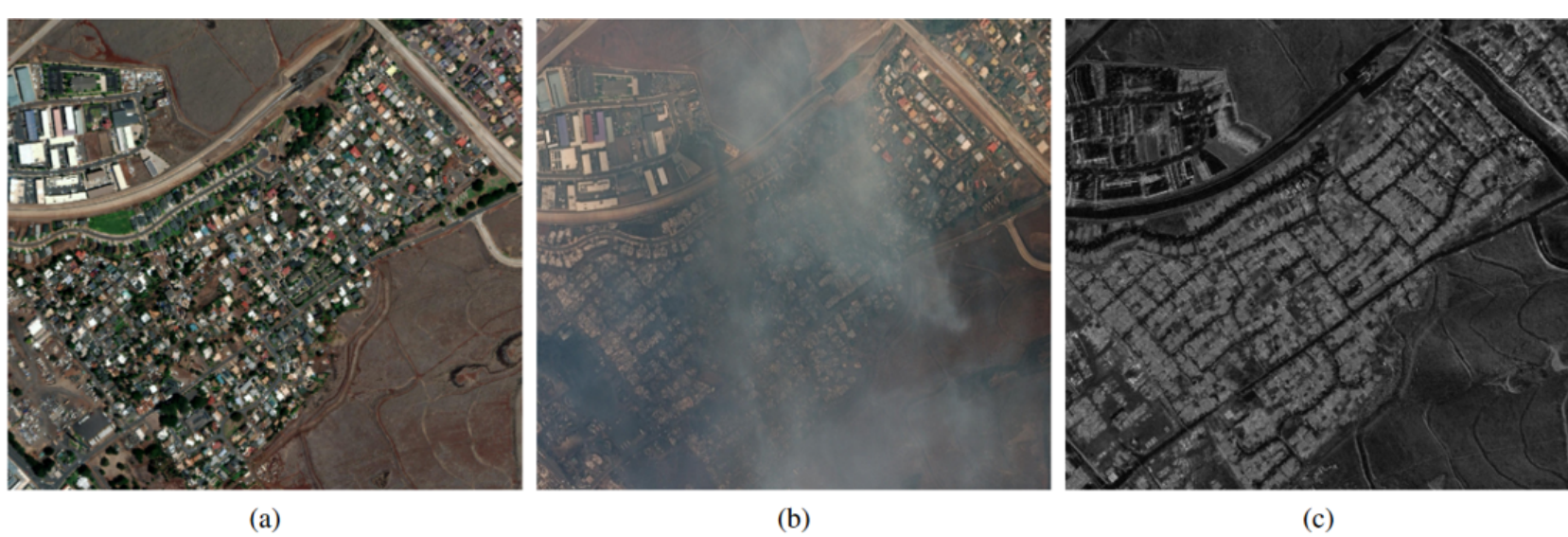


**Figure 1:** An example of the wildfire occurring in Maui, Hawaii, USA, August 2023. (a) Pre-event optical imagery (Maxar). (b) Postevent optical image (Maxar) with land-cover features obscured by smoke from wildfires. (c) Post-event SAR imagery (Capella Space) unaffected by smoke, clearly showing disaster area.

Change detection involves identifying and analyzing changes in Earth’s surface or target objects by comparing remote sensing images from different time points. At present, change detection primarily relies on single-source optical or SAR images to detect temporal changes. However, optical images are susceptible to degradation from lighting conditions and adverse weather, whereas SAR images can be affected by noise and loss of detail, both of which reduce the accuracy of change detection.

Compared to using only pre and post-time SAR images or pre-time optical and post-time SAR images, this paper introduces a method that leverages the concept of memory to enhance multimodal change detection accuracy by utilizing as much available data as possible, including both historical optical and SAR images.

## 3. Methods

While significant progress has been made in single-source optical image-based change detection, adverse weather and poor lighting often limit the availability of post-time optical data, making SAR images the primary alternative. However, the lack of detailed features in SAR images reduces detection accuracy. To address this, we incorporate historical optical images to supplement SAR data and enhance detail.

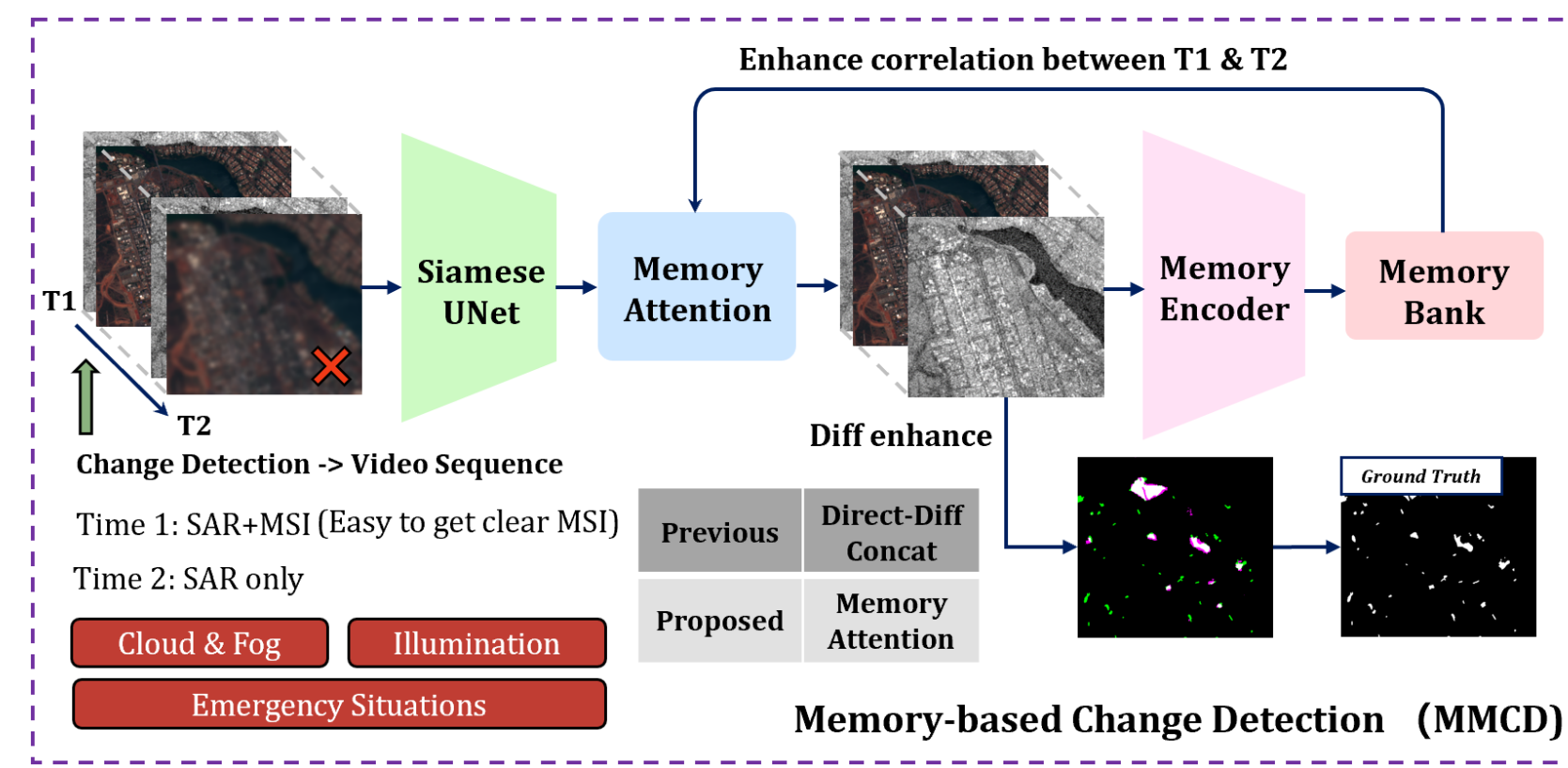


**Figure 2:** Framework of the proposed MMCD. Introduce concept of memory into the multimodal change detection by utilizing historical information when post-time optical images are unusable due to degradation.

The proposed memory-based multimodal change detection (MMCD) framework, illustrated in Figure 2, consists of a Memory Encoded Module (MEM) and a Difference Enhancement Module (DEM). It employs a SiameseUNet feature extractor, comprising three U-Net models: one for optical data and two for SAR data. The optical and SAR U-Nets have independent weights, while the two SAR U-Nets share weights. T1 and T2 denote pre-time and post-time, respectively.

## 4. Experiment

We evaluated the proposed method on the Onera Satellite CD (OSCD) dataset, which is a popular multimodal change detection dataset. It includes an optical dataset from Sentinel-2 with a multi-sensor version incorporating Sentinel-1 SAR images.

For the loss function, since the number of changed pixels is typically much smaller compared to unchanged pixels, there is a significant class imbalance between pixel categories. This imbalance causes models to favor the majority class while neglecting the minority class during training. To address this issue, the loss function in this paper combines Dice and Focal loss.

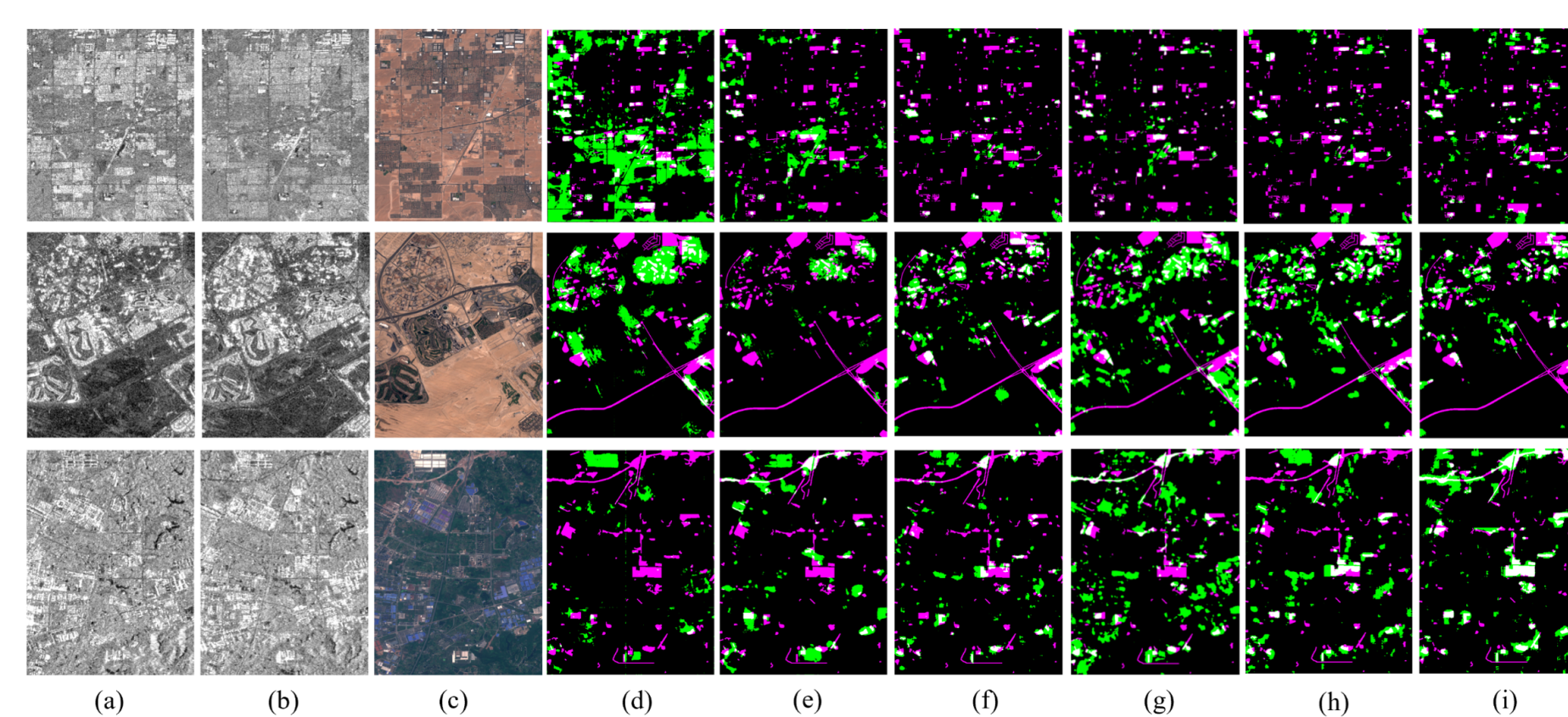


**Figure 3:** Change detection result in missing modality scenarios for Las Vegas (first row), Dubai (second row) and Chongqing (last row): (a) SAR image captured at T1. (b) SAR image captured at T2. (c) Optical image captured at T1. (d) Result of EF-CNN (e) Result of FC-Siam-conc. (f) Result of MSCD. (g) Result of Siamese. (h) Result of SmaDS. (i) Result of the proposed MMCD. The colors white, green, magenta, and black denote TP, FP, FN and TN, respectively.

Figure 3 gives the visualized results. The higher number of false positive (FP) predictions (indicated by more green areas) suggests that the model is more aggressive in its predictions. However, this aggressiveness may cause false changes due to the significant domain gap. Conversely, frequent false negative (FN) predictions indicate weaker change detection capability.

| Method           | Prec         | Rec          | Acc          | F1           |
|------------------|--------------|--------------|--------------|--------------|
| EF-CNN           | 0.180        | 0.218        | 88.58        | 0.197        |
| EF-UNet          | 0.220        | 0.390        | 89.70        | 0.277        |
| Siamese-CNN      | 0.189        | 0.272        | 89.68        | 0.223        |
| FC-Siam-conc     | 0.202        | 0.357        | 89.96        | 0.258        |
| FC-Siam-diff     | 0.220        | 0.321        | 90.71        | 0.261        |
| Fres-UNet        | 0.225        | 0.364        | 88.60        | 0.278        |
| MSCD             | 0.364        | 0.237        | 91.91        | 0.287        |
| DS-UNet          | 0.256        | 0.334        | 91.53        | 0.290        |
| Siamese-UNet     | 0.220        | <b>0.470</b> | 89.10        | 0.303        |
| Siamese-UNet-TTA | 0.270        | 0.410        | 91.28        | 0.321        |
| SmaDS            | 0.341        | 0.327        | 91.26        | 0.334        |
| MMCD (proposed)  | <b>0.384</b> | 0.378        | <b>93.66</b> | <b>0.381</b> |

**Table 1:** Change detection results of pre-SAR, post-SAR and pre-optical images.

According to evaluation results in Table 1. It can be seen that the proposed method achieves a significant performance improvement in multimodal change detection. We also conducted preliminary experiments to explore the optimal number of temporal images for enhancing performance.

Additionally, using two SAR images from the same modality slightly surpasses the performance of combining optical and SAR images, likely due to distributional differences between modalities. However, additional temporal images may confuse the model’s temporal reasoning, underscoring the need for more sophisticated designs. Thus, the proposed approach, which leverages three images, effectively enhances performance and addresses current limitations, offering a promising direction for future research.

## 5. Conclusions

In conclusion, this paper proposes MMCD, a memory-based multimodal change detection method that leverages historical information to enhance accuracy. Unlike traditional approaches relying on feature concatenation or differencing, MMCD models change detection as a temporal problem, improving temporal correlations through memory attention. Experiments confirm its state-of-the-art performance. Future work may explore additional historical images and optimize foundation model for further improvement.

## 6. References

[1] Sudipan Saha et al., “Supervised change detection using pre-change optical-SAR and post-change SAR data,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, 2022.

## 7. Acknowledgement

This work was conducted as part of the Dragon 6 cooperation programme between the European Space Agency (ESA) and the Ministry of Science and Technology (MOST) of China. The authors acknowledge the support and coordination provided by the Dragon Programme Office.