SCL: Towards Domain Generalization via Single-Temporal Multimodal Contrastive Learning for Remote Sensing Change Detection

Model	LEVIR-CD			WHU-CD		
	F1	IoU	OA	F1	IoU	OA
BiT [4]	89.31	80.68	98.85	83.98	72.39	99.26
ChangeFormer [2]	90.40	82.48	99.04	85.61	75.14	98.84
STNet [8]	87.26	70.40	99.10	82.23	69.95	98.89
TinyCD [1]	91.05	83.57	99.10	91.48	84.30	99.32
USSFC-Net [7]	91.04	83.55	99.09	**92.20**	**85.54**	99.27
ChangeStar [14]	90.82	83.19	99.02	86.25	75.98	95.26
SCL(Ours)	**91.14**	**83.72**	**99.11**	92.18	85.50	**99.33**

Table 1. **Supervised change detection.** Comparison with the state-of-the-art change methods on the LEVIR-CD [3] and WHU-CD [6] test.

1. Supervised Change Detection Results.

We fine-tune SCL on the LEVIR-CD [3] and WHU-CD [6] datasets. Tab. 1 presents the supervised results. The SCL change detection performance improves significantly after supervised fine-tuning on each dataset. This demonstrates the potential of SCL as a unified model for change detection.

Additionally, to compare our approach with recent multimodal and zero-shot methods, we train them on the dataset of Changen [15] to assess their generalization performance. Due to space limitation, we present these experimental details in supplementary materials.

2. Comparison with Unsupervised Detectors.

Unsupervised methods, as like SCL, effectively address the dependency on annotated data in RSCD. However, they generally suffer from low accuracy. As shown in Tab 2, our comparison with the latest unsupervised change detection methods demonstrates that SCL, even without any training on the target dataset, achieves considerably higher accuracy, highlighting its significant advantage.

Qualitative results of SAIN. Fig. 1 illustrates the superiority of our single-temporal training strategy compared with single-temporal supervised learning (STAR) [14] used in ChangeStar [14]. It is evident that our generated images closely resemble with the changes in real world, including changes between different objects and color change between the same object. Additionally, our method avoids the unnatural changes, such as overlapping buildings that do not represent actual changes (as shown in the red region of the Fig. 1), which can occur in STAR [14]. This further validates the effectiveness of our approach.

Images of SAIN generated. To further demonstrate the superiority of our generation strategy, we have included additional examples of the images we generated in Fig. 2. As shown in Fig. 2 (a), the images generated by SAIN exhibit mutual changes between different categories, such as forests transforming into grasslands, forests transforming into bodies of water, forests transforming into barren lands, and changes in buildings, among others. In Fig. 2 (b), the generated examples also capture seasonal changes, such as forests and fields in different seasons. Fig. 2 (c) showcases variations within the same category, such as changes between dense and sparse forest vegetation. These changes closely resemble real-world transformations. Compared to STAR [14], SAIN strategy aligns well with real-world scenarios.

Methods	F1	IoU	OA
DCVA [10]	52.89	-	84.75
SCM-CD [12]	62.80	53.67	88.80
PUCD-SAM [9]	46.80	30.70	91.10
SCL (zero-shot, ours)	**85.19**	**74.20**	**98.59**

Table 2. Comparison with unsupervised change detection on LEVIR-CD [3].

Source Model	Architecture	F1	IoU
FarSeg	Base	67.43	51.51
	SCL+	72.62	57.01
Changer	Base	65.59	48.79
	SCL+	74.34	59.16

Table 3. Ablation experiment of image encoder replacement. We employ the image encoder with an change detection head of Changer [11] as the base architecture. The image encoders are sourced from FarSeg [13] and Changer [11].

∗Equal contribution †Corresponding author

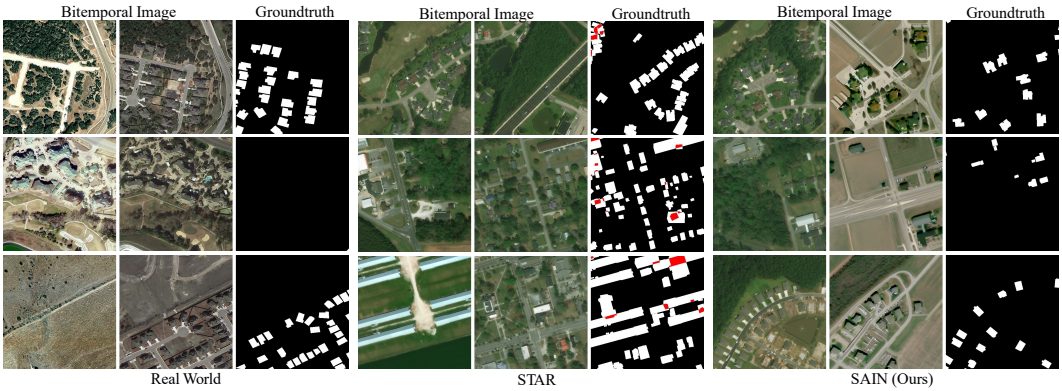

Figure 1. Qualitative analysis result of single-temporal image in SAIN. The red area is the area where the buildings overlap in the STAR [14], which is impossible in real world. SAIN generates images that encompass both pseudo-change within the same category and changes between different categories.

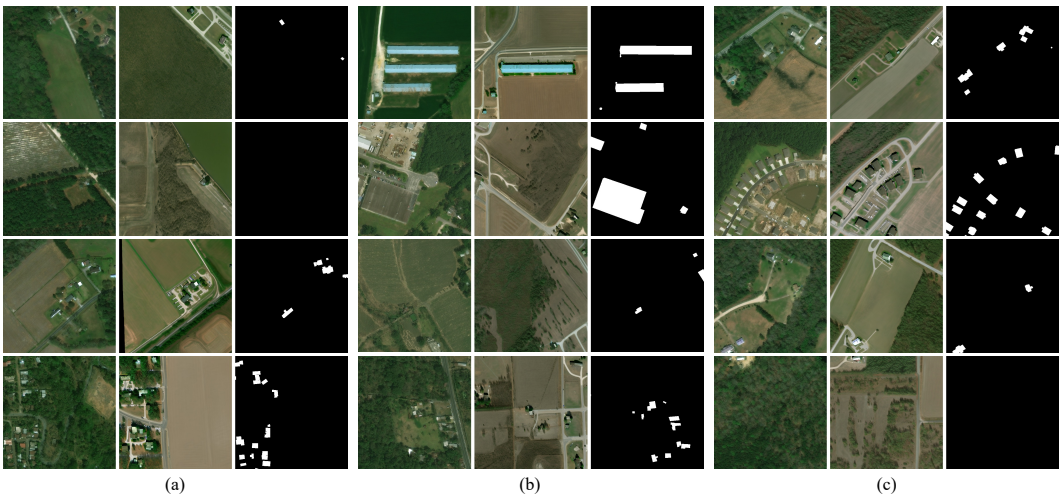

Figure 2. Image cases generated by SAIN on xView2 [5].

3. Ablation Experiment of Image Encoder:

To verify whether the excellent domain generalization ability of SCL is solely attributed to the image encoder of CLIP, and to validate the contribution of the visual-text interaction architecture proposed by SCL, we conducted replacement experiments on image encoder of SCL. Specifically, we substitute the original image encoder with those sourced from FarSeg [13] (remote sensing semantic segmentation) and Changer [11] (change detection). Subsequent fine-tuning experiments were conducted accordingly. As observed in the Tab. 3, incorporating these encoders into our network architecture achieved a significant improvement compared to using them individually for change object detection. This observation underscores the enhanced efficacy of the proposed SCL architecture.

References

[1] Alessandro Ferrari Andrea Codegoni, Gabriele Lombardi. Tinycd: A (not so) deep learning model for change detection. *arXiv preprint arXiv:2207.13159*, 2022. 1

[2] Wele Gedara Chaminda Bandara and Vishal M. Patel. A transformer-based siamese network for change detection. In *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 207–210, 2022. 1

[3] Hao Chen and Zhenwei Shi. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sensing*, 12(10), 2020. 1

[4] Hao Chen, Zipeng Qi, and Zhenwei Shi. Remote sensing image change detection with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022. 1

[5] Ritwik Gupta, Richard Hosfelt, Sandra Sajeev, Nirav Patel, Bryce Goodman, Jigar Doshi, Eric Heim, Howie Choset, and Matthew Gaston. xbd: A dataset for assessing building damage from satellite imagery, 2019. 2

[6] Shunping Ji, Shiqing Wei, and Meng Lu. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1):574–586, 2019. 1

[7] Tao Lei, Xinzhe Geng, Hailong Ning, Zhiyong Lv, Maoguo Gong, Yaochu Jin, and Asoke K. Nandi. Ultralightweight spatial–spectral feature cooperation network for change detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–14, 2023. 1

[8] Xiaowen Ma, Jiawei Yang, Tingfeng Hong, Mengting Ma, Ziyan Zhao, Tian Feng, and Wei Zhang. Stnet: Spatial and temporal feature fusion network for change detection in remote sensing images. *arXiv preprint arXiv:2304.11422*, 2023. 1

[9] Youngtack Oh, Minseok Seo, Doyi Kim, and Junghoon Seo. Prototype-oriented unsupervised change detection for disaster management, 2023. 1

[10] Sudipan Saha, Francesca Bovolo, and Lorenzo Bruzzone. Unsupervised deep change vector analysis for multiple-change detection in vhr images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6):3677–3693, 2019. 1

[11] Zhe Li Sheng Fang, Kaiyu Li. Changer feature interaction is what you need for change detection. *Arxiv*, 2023. 1, 2

[12] Xiaoliang Tan, Guanzhou Chen, Tong Wang, Jiaqi Wang, and Xiaodong Zhang. Segment change model (scm) for unsupervised change detection in vhr remote sensing images: a case study of buildings, 2023. 1

[13] Zhuo Zheng, Yanfei Zhong, Junjue Wang, and Ailong Ma. Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4096–4105, 2020. 1, 2

[14] Zhuo Zheng, Ailong Ma, Liangpei Zhang, and Yanfei Zhong. Change is everywhere: Single-temporal supervised object change detection in remote sensing imagery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15193–15202, 2021. 1, 2

[15] Zhuo Zheng, Shiqi Tian, Ailong Ma, Liangpei Zhang, and Yanfei Zhong. Scalable multi-temporal remote sensing change data generation via simulating stochastic change process. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21818–21827, 2023. 1