A Coarse-to-Fine Change Detection Framework for Remote Sensing Sparse Cultivated Land

Yuan Hu, Yifan Zhang, Mingyang Ma and Shaohui Mei
Northwestern Polytechnical University, Xi'an 710072, P.R. China
E-mail: huyuan1999hy@mail.nwpu.edu.cn, yifanzhang@nwpu.edu.cn, mamingyang@mail.nwpu.edu.cn, meish@nwpu.edu.cn

Abstract—**Remote sensing (RS) images contain rich geographic information. For specific application scenarios like cultivated land, it is necessary to select areas of interest to reduce data scale and focus on detailed features. In this article, an innovative coarse-to-fine change detection framework (CFCD) for sparse cultivated land is proposed to address these problems. Coarse screening module (CSM) first removes irrelevant low-difference image pairs, and then fine detection module (FDM) accurately locate change areas in remaining images. Experimental results show that two coarse screening methods can take out many disturbed images, and provide strong support for subsequent fine detection methods to achieve performance improvement.**

I. Introduction

In recent years, the rapid development of RS technology has driven the accumulation of relevant data, which provides strong support for the diversified application of Remote Sensing (RS) images. Taking change detection (CD) for example, its core idea is to accurately identify and analyze changes in the surface cover by jointly comparing two (or more) images acquired on the same geographical area at different times [1]. At present, it is widely utilized in various fields such as deforestation investigation [2], urban planning [3], disaster assessment [4] and other fields, demonstrating extremely high practical value and application prospects.

As the scarce land resource in the world, cultivated land not only serves as the lifeblood and main carrier for agricultural production activities, but also functions as the core factor in promoting food production and regional sustainable development [5]. The amount of adequate cultivated land is crucial for maintaining food security and fostering harmonious coexistence between humanity and nature. However, the ecological degradation caused by the change of cultivated land has become a prominent global issue [6]–[8]. On one hand, the distribution of cultivated land is deeply affected by natural conditions. On the other hand, the expansion of industrialization and urbanization has led to continuous compression and cutting of limited cultivated land, deepening the degree of sparsity and fragmentation. Therefore, it is urgent to implement protection policies for cultivated land.

How to quickly detect changes in cultivated land is a key step in effectively protecting the cultivated land. Under the background that the RS data acquired by a single satellite data center is dramatically increasing at a speed of several terabytes per day, traditional manual field investigation methods, as well as certain automated approaches, cannot meet the demands for direct real-time monitoring of large-scale images [9]. As the convolutional neural network (CNN) is introduced to CD, deep learning methods for cultivated land have attracted more attention from researchers. Liu *et al.* [10] designed a CNN-transformer network with multiscale context aggregation (MSCANet), which exploits context aggregate connections to fuse and aggregate features across different levels, fulfilling efficient and effective cultivated land CD. To address the problem of inadequate utilization of feature information, Miao *et al.* [11] proposed a Siamese network based on full-scale connected UNet (SNUNet3+), which combines the spatial and channel squeeze and excitation (scSE) attention mechanism and deep supervision modules to detect changes in cultivated land. In [12], A transformer-based multiscale feature fusion change detection network (M-Swin), which can capture the change information in small building through hierarchical windows and integrate the multiscale feature obtained from different windows to cope with the "scale gap" challenge.

Even though the existing mainstream methods have made better detection performance, they heavily rely on high-quality and well-produced standardized datasets, failing to fully consider the common problem of "non-focused change areas" in original large-sized RS images, which are far from practical application. In addition, the sparse characteristics of cultivated land also need to be further excavated.

In order to deal with the above-mentioned problems, a coarse-to-fine change detection (CFCD) framework is proposed in this paper for sparse cultivated land. The overall idea of this method is "coarse screening-fine detection". Firstly, differences between the two original bitemporal image pairs are analyzed and quantified through the method of coarse feature representation, and the image pairs with relatively small differences are discarded. Then the retained image pairs are finely detected by employing deep learning methods based on Siamese network. Finally, the patches of cultivated land change areas are located. The major contributions of this article are listed as follows.

- Two coarse feature representation methods are proposed for cultivated land CD, which aim to quickly identify areas with obvious changes and reduce the amount of data on subsequent processing.
- An efficient coarse screening-fine detection method is designed for cultivated land CD task based on large scale images. The proposed method can directly process

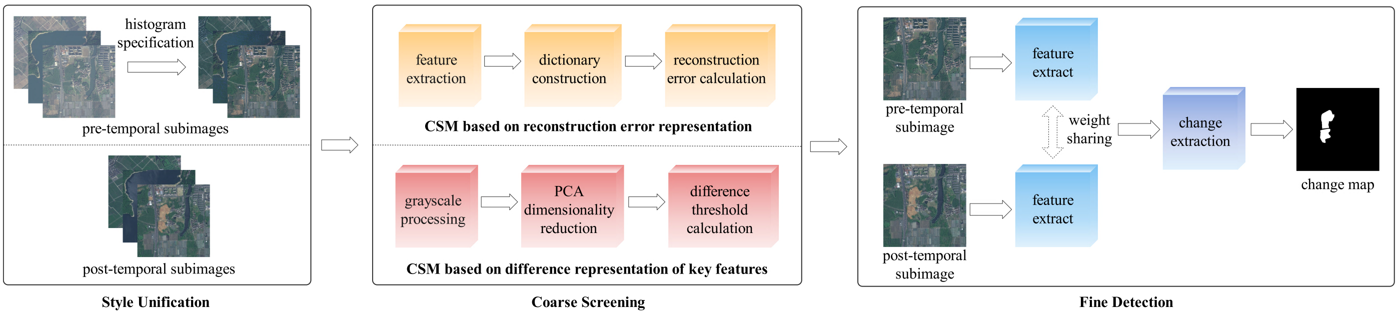

Fig. 1. Framework of the proposed CFCD

original images and significantly improve detection accuracy before being used in existing state-of-the-art (SOTA) methods.
- Extensive comparison and experiments with four SOTA CD models on ShangYu CLCD Datasets are conducted, which have proved the effectiveness of the proposed method.

II. Methodology

Except for the necessary preprocessing such as radiometric correction and geometric registration, which can truly reflect changes in land surface conditions, the original large-sized bitemporal images are directly cropped into unified small-sized subimages without any additional steps to form an image pair.

The proposed CFCD framework in this paper can rapidly remove low difference image pairs through coarse screening module (CSM), and extract the specific change areas from the remaining image pairs using fine detection module (FDM), so as to generate the final detection results. Framework of the proposed CFCD is shown in Fig 1.

A. Coarse Screening Module (CSM)

In large-scale RS images, the area of cultivated land changes usually occupies only a small part of them, and is sparsely distributed. The large presence of non-research areas not only consumes computing resources but also introduces noise, which greatly affects the accuracy of CD. To settle these problems, two coarse screening methods are designed in CSM modules of the proposed CFCD.

1) CSM based on reconstruction error representation (CSM-RER): Compressive Sensing (CS) theory manifests that even if only a small number of measured values are obtained, the original image can be effectively restored and reconstructed from sparse or sparsely represented images through specific measurement matrices and optimization algorithms [13], [14]. The proposed CSM based on reconstruction error representation adopts feature extraction algorithm to extract the same N-dimensional features from all image pairs. The features of the i-th post temporal subimage $\mathbf{y_i}$ is selected as the object to be reconstructed, and the corresponding features of pre-temporal subimage and neighboring subimages are selected to construct a dictionary $\mathbf{D_i}$ for sparse reconstruction, as shown in Fig 2.

The reconstruction error e_i is calculated using

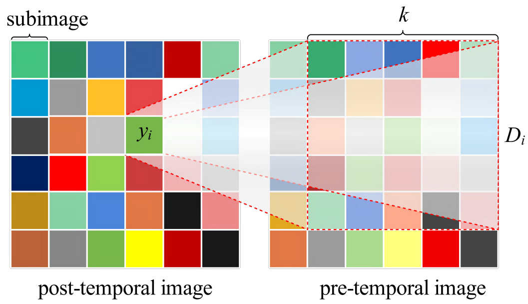

Fig. 2. CSM based on reconstruction error representation

$$e_i = \frac{\|\mathbf{y_i} - \mathbf{D_i} \cdot \mathbf{x_i}\|_2}{\|\mathbf{y_i}\|_2} < \varepsilon \tag{1}$$

where $\|\cdot\|_2$ represents the ℓ_2-norm, $\mathbf{y_i} \in \mathbb{R}^{N\times 1}$, $\mathbf{D_i} \in \mathbb{R}^{N\times M}$, ε is the error threshold. The number of words in the dictionary $M = k^2 \le K$, with K the number of subimages cropped from each original image. Greater error indicates more likely changes, that is, the image pairs with smaller values can be discarded by sorting according to the reconstruction error.

2) CSM based on difference representation of key features (CSM-DRKF): Principal component analysis (PCA) is a technique that reduces the dimensionality of original data while maximizing the preservation of its intrinsic information [15]. As shown in Fig 3, grayscale processing is first performed on all subimages and two matrices (i.e. $\mathbf{A}$, $\mathbf{B}$) are composed of pre-temporal and post-temporal images. Then, to highlight the key information and eliminate the disturbing information, PCA is used to reduce the dimensionality of the two image matrices to N-dimension. Finally, the differences are calculated and sorted based on (2)-(4), and the image pairs with small differences are removed.

$$\mathbf{C} = \sum_{i=1}^{N}\sum_{j=1}^{K} c_{ij} = \mathbf{A} - \mathbf{B} \tag{2}$$

$$\mathbf{D} = \left(\sum_{i=1}^{N}|c_{i1}|, \sum_{i=1}^{N}|c_{i2}|, \cdots, \sum_{i=1}^{N}|c_{ij}|\right) \tag{3}$$

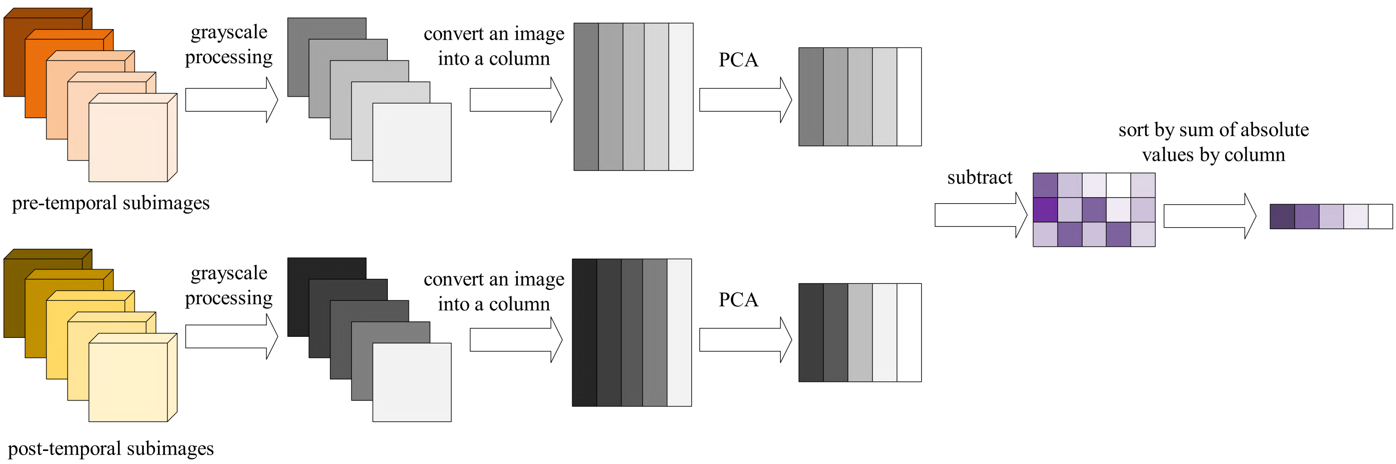

Fig. 3. CSM based on difference representation of key features

$$d_k = \sum_{i=1}^{N} |c_{ij}| < \delta \tag{4}$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{R}^{N \times K}$, $\mathbf{D} = (d_1, d_2, \cdots, d_k) \in \mathbb{R}^{1 \times K}$, δ is the difference threshold.

B. Fine Detection Module (FDM)

Siamese network is a deep learning architecture used to solve the task based on similarity comparison, consisting of two or more identical subnetworks that share weights and parameters [16], as shown in Fig 4.

Each sub network receives an input sample and extracts features through a series of convolutional layers (or other types of layers). At present, many mainstream CD methods are proposed based on Siamese network structure, which have exhibited powerful feature extraction and comparison capabilities. In FDM of the proposed CFCD framework, the following four typical methods can be incorporated.

FC-Siam-conc [20] is a feature-concatenation method, which extracts multi-level features of bi-temporal images from shared Siam structures, and concatenate them to detect changes. The skipping feature connection can capture subtle changes with richer feature expressions, which is particularly important for CD in cultivated land.

Unlike FC-Siam-conc, FC-Siam-diff [20] is a feature-difference method, which extracts features by using the Siamese connection and continuously superimposes residuals during deconvolution to achieve favorable CD. This approach focuses on capturing different information between images, which can reduce the pseudo-changes caused by non-substantive factors such as light and seasonal changes to a certain extent, and is more sensitive to identification in change areas.

BIT [21] is a feature fusion method, which integrates Siamese tokenizer and transformer encoder-decoder structure to achieve more meaningful context-information to obtain the change map. By introducing transformer structure, it is possible to globally obtain the dependency relationships in RS images, which is helpful to identify the complex patterns of cultivated land changes.

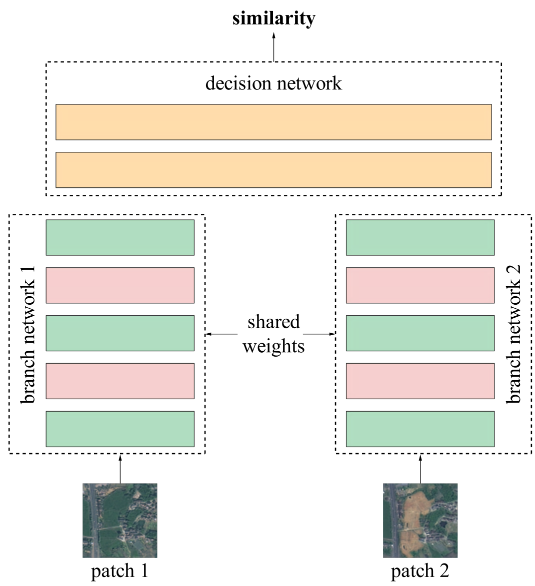

Fig. 4. Siamese network structure

DTCDSCN [22] is an attention-based method, which utilizes change information in spatial and channel information to extract more contextual features. It combines two tasks: CD and semantic segmentation. Through the shared feature extraction layer, the network can extract the features of cultivated land changes and the semantic information of cultivated land at the same time, which is obviously helpful for subtle change detection in cultivated land.

III. Experiments and Analysis

A. ShangYu CLCD Datasets

Many publicly datasets have been used for CD, typically including various typical scenarios such as urban areas, plains,

TABLE I
EXPERIMENTAL RESULTS ON CSM BASED ON RECONSTRUCTION ERROR REPRESENTATION (CSM-RER)

Dataset	Number of Images (K)	Number of Changes	Number of Words (k^2)	Error Threshold (ε)	Images Removal Rate
ShangYu-GF1b	5520	142	3×3	7.5264×10^{-2}	36.11%
			5×5	6.4046×10^{-2}	36.59%
			7×7	5.6378×10^{-2}	**37.07%**
			9×9	5.0041×10^{-2}	36.01%
ShangYu-BJ2	1596	168	3×3	5.5320×10^{-2}	15.60%
			5×5	5.2397×10^{-2}	15.66%
			7×7	4.8918×10^{-2}	15.73%
			9×9	4.6291×10^{-2}	**15.79%**

TABLE II
EXPERIMENTAL RESULTS ON CSM BASED ON DIFFERENCE REPRESENTATION OF KEY FEATURES (CSM-DRKF)

Dataset	Number of Images (K)	Number of Changes	Dimension (N)	Difference Threshold (δ)	Images Removal Rate
ShangYu-GF1b	5520	142	256	79.7307	45.65%
			512	123.2878	**46.59%**
			1024	187.5146	46.20%
ShangYu-BJ2	1596	168	256	74.2009	18.98%
			512	112.6452	19.49%
			1024	187.1340	**19.80%**

TABLE III
EXPERIMENTAL RESULTS ON FDM AND CFCD

Method	FDM					CFCD (FDM + CSM-DRKF)				
	Precision	Recall	F1	IoU	OA	Precision	Recall	F1	IoU	OA
FC-Siam-conc	40.71	46.00	58.58	37.39	63.69	76.88	63.98	71.92	56.14	82.10
FC-Siam-diff	52.07	50.24	42.80	44.21	63.77	69.51	44.99	56.32	49.20	75.28
BIT	52.62	52.98	53.66	46.24	64.02	72.93	43.33	54.36	47.33	77.13
DTCDSCN	32.15	43.29	41.78	38.27	52.03	64.06	49.05	55.34	43.03	76.73

cultivated land, and mountainous areas. However, the availability of datasets for cultivated land remains extremely restricted, greatly limiting the innovative application of CD methods. To address this issue, two large-scale optical remote sensing images of GF-1b and BJ-2 taken by Zhejiang Institute of Surveying and Mapping Science and Technology are obtained as the new dataset for detecting changes in cultivated land, with the sizes of 40598×35178 and 21243×19041 pixels, respectively, and spatial resolution ranged from 0.8 to 2 m. These images were uniformly cropped to 512×512 pixels, resulting in a total of 7116 image pairs (i.e. ShangYu-GF1b obtains 5520 pairs, ShangYu-BJ2 obtains 1596 pairs), forming ShangYu CLCD Datasets. The types of change targets mainly include buildings and structures. These samples are randomly separated for training, test and validation in the ratio of 6:2:2.

B. Implementation Details and Evaluation Metrics

Obvious stylistic differences exist between pre-temporal and post-temporal images of datasets, caused by different lighting and atmospheric conditions during images shooting. To reduce its impact on the accuracy of subsequent CD, histogram specification is used to unify the style [17]–[19].

The purpose of coarse screening is to exclude as many images as possible there are no changes, without affecting the precision of fine detection. Therefore, the effectiveness of two methods is measured with zero missed detection. This shows that error threshold (ε) and difference threshold (δ) are based on the image with the first missed detection, meaning it is less than this value and the change is negligible. As for CSM based on reconstruction error representation (CSM-RER), feature extraction algorithm is applied to each subimage to extract 360 features (i.e. $N = 360$). And k is set to be 3, 5, 7, 9 respectively.

As for CSM based on difference representation of key features (CSM-DRKF), PCA is applied to each subimage and N is set to be 256, 512, 1024 respectively.

Five evaluation metrics are employed for quantitative assessment of the proposed CD method, including precision (Pre), recall (Rec), F1-score (F1), intersection over union (IoU), and overall accuracy (OA). They range from 0 to 1, and higher values indicate better performance.

$$\mathrm{Pre} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{5}$$

$$\mathrm{Rec} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{6}$$

$$F1 = \frac{2\mathrm{Pre} \times \mathrm{Rec}}{\mathrm{Pre} + \mathrm{Rec}} \tag{7}$$

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{8}$$

$$\mathrm{OA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{9}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

C. Comparative Experiments

1) Comparison Experiments on Different CSMs: Table I and II lists evaluated metrics for different values of k and N on different CSMs. With the increasing of k or N, images removal rate shows an upward trend indicating that the better the quality of the processed dataset, the more obvious the removal effect of non-relevant changing images. In ShangYu-GF1b dataset, images removal rate reaches its maximum as 37.07% and 46.59% when k=7 or N=512, respectively. In ShangYu-BJ2 dataset, when k=9 or N=1024, images removal rate is 15.79% and 19.80%. Generally, both CSM-DRKF and CSM-RER can achieve better screening effects. In ShangYu CLCD Dataset, CSM-DRKF outperform CSM-RER as it can remove more invalid images.

2) Comparison Experiments on FDM and CFCD: Based on the results of CSM-DRKF, all images retained in ShangYu-GF1b and ShangYu-BJ2 datasets are combined to form new ShangYu CLCD Dataset, and comparative experiments of FDM are conducted. Most of the existing FDM methods directly detect images in the dataset, but the proposed CFCD architecture first coarsely screens images and then performs fine detection through any CD method, which reduces computational complexity and significantly improves accuracy. Table III illustrates quantitative evaluations of the proposed CFCD architecture employing four different SOTA CD methods. It can be observed that, compared to the results of ShangYu CLCD Datasets, the proposed CFCD framework can achieve better evaluation metrics in CD performance, with significant improvements in all five metrics. For example, the F1/IoU/OA of the proposed CFCD framework exceeds previous SOTA by 41.9/33.4/36.6% for ShangYu CLCD Datasets, respectively. It can also be observed from the Table III, the Precision, Recall, IoU, and OA values of traditional FC-Siam-conc method are lower than BIT. While the evaluation metrics of the proposed CFCD framework have been significantly improved because of the preprocessing procedure, exceeding those of BIT, which is 31.56/17.19/17.63/22.02%, respectively. In addition, although the ability of DTCDSCN to detect changes performed the worst on ShangYu CLCD Dataset, Precision, F1 and OA indicators have significantly improved after integration into the CFCD framework. These quantitative and qualitative comparisons indicate the superiority of the proposed CFCD framework in improving the detection performance of existing SOTA methods.

IV. Conclusions

In this paper, a CFCD framework based on dual modules (CSM and FDM) is proposed, which makes it possible to utilize the original RS images directly for detection. In CSM, a variety of non-cultivated land changes are quickly identified and screened, which greatly reduces the amount of data. In FDM, deep learning methods based on Siamese network are adopted which can specifically extract and locate the attention areas, to achieving the final CD results. Experimental results illustrate the effectiveness and competitiveness of the proposed CFCD framework. In future work, sparse feature extraction is also a direction worth further exploring.

Acknowledgment

This work was supported in part by the Zhejiang Province "Pioneering Soldier" and "Leading Goose" Research and Development Project under Grant 2023C01027.

References

[1] L. Bruzzone and F. Bovolo, "A novel framework for the design of change-detection systems for very-high-resolution remote sensing images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 609–630, 2012.

[2] T. Hame, I. Heiler, and J. San Miguel-Ayanz, "An unsupervised change detection and recognition system for forestry," *International journal of remote sensing*, vol. 19, no. 6, pp. 1079–1099, 1998.

[3] L. Yang, G. Xian, J. M. Klaver, and B. Deal, "Urban land-cover change detection through sub-pixel imperviousness mapping using remotely sensed data," *Photogrammetric Engineering & Remote Sensing*, vol. 69, no. 9, pp. 1003–1010, 2003.

[4] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using vhr optical and sar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2403–2420, 2010.

[5] G. Jiang, M. Wang, Y. Qu, D. Zhou, and W. Ma, "Towards cultivated land multifunction assessment in china: Applying the "influencing factors-functions-products-demands" integrated framework," *Land Use Policy*, vol. 99, p. 104 982, 2020.

[6] W. B. Meyer and B. L. Turner, "Human population growth and global land-use/cover change," *Annual review of ecology and systematics*, vol. 23, no. 1, pp. 39–61, 1992.

[7] R. Lal, "Restoring soil quality to mitigate soil degradation," *Sustainability*, vol. 7, no. 5, pp. 5875–5895, 2015.

[8] H. Xie, Y. Zhang, Z. Wu, and T. Lv, "A bibliometric analysis on land degradation: Current status, development, and future directions," *Land*, vol. 9, no. 1, p. 28, 2020.

[9] Y. Ma, H. Wu, L. Wang, *et al.*, "Remote sensing big data computing: Challenges and opportunities," *Future Generation Computer Systems*, vol. 51, pp. 47–60, 2015.

[10] M. Liu, Z. Chai, H. Deng, and R. Liu, "A cnn-transformer network with multiscale context aggregation for fine-grained cropland change detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 4297–4306, 2022.

[11] L. Miao, X. Li, X. Zhou, *et al.*, "Snunet3+: A full-scale connected siamese network and a dataset for cultivated land change detection in high-resolution remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[12] J. Pan, Y. Bai, Q. Shu, Z. Zhang, J. Hu, and M. Wang, "M-swin: Transformer-based multi-scale feature fusion change detection network within cropland for remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[13] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[14] Y. Tsaig and D. L. Donoho, "Extensions of compressed sensing," *Signal processing*, vol. 86, no. 3, pp. 549–571, 2006.

[15] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[16] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a" siamese" time delay neural network," *Advances in neural information processing systems*, vol. 6, 1993.

[17] A. K. Jain, *Fundamentals of digital image processing*. Prentice-Hall, Inc., 1989.

[18] K. R. Castleman, *Digital image processing*. Prentice Hall Press, 1996.

[19] D. Coltuc, P. Bolon, and J.-M. Chassery, "Exact histogram specification," *IEEE Transactions on Image processing*, vol. 15, no. 5, pp. 1143–1152, 2006.

[20] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Urban change detection for multispectral earth observation using convolutional neural networks," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, Ieee, 2018, pp. 2115–2118.

[21] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021.

[22] Y. Liu, C. Pang, Z. Zhan, X. Zhang, and X. Yang, "Building change detection for remote sensing images using a dual-task constrained deep siamese convolutional network model," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 5, pp. 811–815, 2020.