

# BUILDING-GUIDED PSEUDO-LABEL LEARNING FOR CROSS-MODAL BUILDING DAMAGE MAPPING

Jiepan Li

*Wuhan University*430072 Wuhan, China  
jiepanli@whu.edu.cn

He Huang

*Wuhan University*430072 Wuhan, China  
huang\_he@whu.edu.cn

Yu Sheng

*Wuhan University*430072 Wuhan, China  
shengyu@whu.edu.cn

Yujun Guo

*Wuhan University*430072 Wuhan, China  
yujunguo@whu.edu.cn

Wei He\*

*Wuhan University*430072 Wuhan, China  
weihe1990@whu.edu.cn

**Abstract**—Accurate building damage assessment using bi-temporal multi-modal remote sensing images is essential for effective disaster response and recovery planning. This study proposes a novel Building-Guided Pseudo-Label Learning Framework to address the challenges of mapping building damage from pre-disaster optical and post-disaster SAR images. First, we train a series of building extraction models using pre-disaster optical images and building labels. To enhance building segmentation, we employ multi-model fusion and test-time augmentation strategies to generate pseudo-probabilities, followed by a low-uncertainty pseudo-label training method for further refinement. Next, a change detection model is trained on bi-temporal cross-modal images and damaged building labels. To improve damage classification accuracy, we introduce a building-guided low-uncertainty pseudo-label refinement strategy, which leverages building priors from the previous step to guide pseudo-label generation for damaged buildings, reducing uncertainty and enhancing reliability. Experimental results on the 2025 IEEE GRSS Data Fusion Contest dataset demonstrate the effectiveness of our approach, which achieved the highest mIoU score (54.28%) and secured first place in the competition. The source code will be available at [GPL](#).

**Index Terms**—Building damage mapping, cross-modal, change detection.

## I. INTRODUCTION

Building damage assessment [1] is crucial for post-disaster response and recovery planning, providing essential information for emergency management, resource allocation, and reconstruction efforts. With the increasing availability of remote sensing (RS) data [2], bi-temporal multi-modal imagery, such as optical and synthetic aperture radar (SAR) images, has become a key resource for damage assessment [3]. Optical images capture high-resolution visual details under normal lighting conditions [4, 5], while SAR images provide robust structural information regardless of weather and illumination [6]. However, integrating these modalities for accurate damage mapping poses significant challenges due to differences in imaging geometry (*e.g.*, perspective and illumination), resolution disparities, inconsistent spatial coverage, and modality-specific noise artifacts such as speckle in SAR, all of which hinder accurate pixel-wise feature alignment.

Corresponding author: Wei He.

To advance research in this field, the 2025 IEEE GRSS Data Fusion Contest introduced the "All-Weather Building Damage Mapping" challenge [7]. This competition focuses on assessing building damage using bi-temporal multi-modal images, with a dataset consisting of pre-disaster optical images and post-disaster SAR images at submeter resolution. The dataset is annotated with four classes: background, intact buildings, damaged buildings, and destroyed buildings. A major challenge in this task lies in the substantial differences between optical and SAR imagery, which complicate feature alignment and damage classification. During the evaluation, models are tested on unseen image pairs to generate damage assessments, with performance measured by the mean Intersection over Union (mIoU) metric.

To address these challenges, we propose a Building-Guided Pseudo-Label Learning Framework that leverages building priors to enhance damage classification. Our framework consists of two key stages: building extraction and damaged building change detection. In the first stage, we train a series of building extraction models using pre-disaster optical images and building labels. To further refine building segmentation results, we incorporate multi-model fusion and test-time augmentation strategies to generate pseudo-probabilities for buildings, followed by a low-uncertainty pseudo-label training method. We train a change detection model using bi-temporal cross-modal images and damaged building labels based on the reliable building priors from the first stage. To improve detection accuracy, we introduce a building-guided low-uncertainty pseudo-label refinement strategy, which leverages building priors from the first stage to guide pseudo-label generation for damaged buildings. This reduces uncertainty in training labels and enhances the model's ability to distinguish different levels of building damage.

We evaluate our approach on the official test set of the 2025 IEEE GRSS Data Fusion Contest. Experimental results demonstrate the effectiveness of our framework, achieving the highest mIoU score of 54.28%, ranking first place in the competition. These findings highlight the potential of our method for improving building damage mapping accuracy in real-world disaster scenarios. The main contributions of this

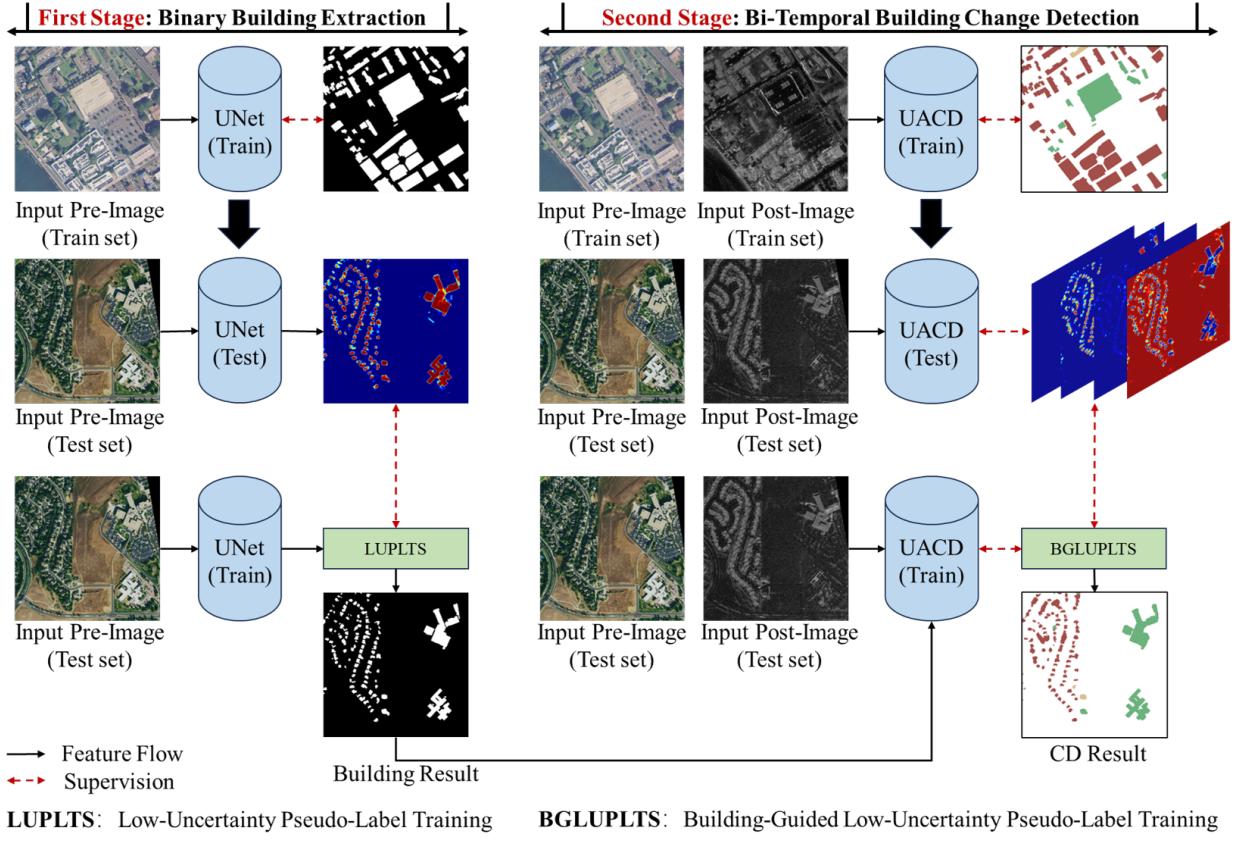


Fig. 1. Architecture of our proposed Cross-Modal Building Damage Mapping Framework.

paper are summarized as follows:

- 1) We propose a Building-Guided Pseudo-Label Learning Framework that integrates building priors to improve cross-modal building damage mapping.
- 2) We introduce a low-uncertainty pseudo-label refinement strategy to enhance label reliability and reduce the impact of modal discrepancies.
- 3) Our framework achieves state-of-the-art performance, ranking first in the 2025 IEEE GRSS Data Fusion Contest, demonstrating its effectiveness in real-world disaster assessment.

The remainder of this paper is organized as follows. Section 2 details our proposed framework, including building segmentation, damaged building change detection, and pseudo-label refinement. Section 3 and Section 4 present experimental results and the corresponding analysis. Finally, Section 5 concludes the paper and discusses future research directions.

## II. METHODOLOGY

### A. Overview

As illustrated in Fig. 1, our building-guided pseudo-label learning framework comprises two main stages. In the first stage, we train multiple building extraction models using pre-disaster optical images and their corresponding building annotations. To improve the quality of segmentation results, we employ a multi-model fusion strategy and test-time augmentation to generate pseudo-probability maps, which are

subsequently refined using a low-uncertainty guided training strategy.

In the second stage, we train a change detection model (UACD [8]) using bi-temporal and cross-modal images, along with damaged building annotations. To further enhance detection performance, we introduce a building-guided pseudo-label refinement approach, leveraging building priors from the first stage. This method helps reduce the uncertainty of the label and strengthens the model's ability to accurately differentiate between the varying levels of damage to the building.

### B. Binary Building Extraction

This stage involves three major components: (1) Building Extraction Model Training, (2) Pseudo-Label Generation, and (3) Low-Uncertainty Pseudo-Label Training. Initially, we train multiple segmentation models using labeled pre-disaster optical images. Next, ensemble learning and test-time augmentation are applied to produce reliable pseudo-probability maps. Finally, a selective training scheme guided by uncertainty is adopted to refine the predictions. Each step is detailed below.

*1) Building Extraction Model Training:* Following previous works [9, 10], modeling long-range dependencies is crucial for accurate segmentation in RS imagery. To this end, we adopt three variants of the Pyramid Vision Transformer [11] (PVT-v2-b2, PVT-v2-b3, and PVT-v2-b4) as encoders, each coupled with a Feature pyramid network(FPN [12])-style decoder for dense prediction.

The official dataset is partitioned into 80% for training and 20% for validation. Only pre-disaster optical images are used as input, with binary labels supervising the training via standard cross-entropy loss. We retain the best-performing model from each configuration based on validation performance, resulting in three models:  $M_1$ ,  $M_2$ , and  $M_3$ .

**2) Pseudo-Label Generation:** Using the trained models ( $M_1$ ,  $M_2$ , and  $M_3$ ), we generate pseudo-labels for pre-disaster images in the test set through an ensemble strategy augmented with test-time data augmentation. Given an input image  $I_1$ , we first apply horizontal flipping to obtain its augmented version  $I_2$ . The three models process both  $I_1$  and  $I_2$ , and the outputs are averaged separately. The final pseudo-probability map is calculated by combining the original and flipped predictions as follows:

$$\begin{aligned} P_1 &= \frac{M_1(I_1) + M_2(I_1) + M_3(I_1)}{3}, \\ P_2 &= \frac{M_1(I_2) + M_2(I_2) + M_3(I_2)}{3}, \\ P_f &= \frac{P_1 + De(P_2)}{2}, \end{aligned} \quad (1)$$

where  $De(\cdot)$  denotes the inverse horizontal flip operation to align  $P_2$  with the original spatial orientation.

**3) Low-Uncertainty Pseudo-Label Training:** To mitigate the influence of noisy pseudo-labels, we design a low-uncertainty guided training strategy. From the pseudo-probability map  $P_f$ , we construct two types of supervision signals: a hard label  $G_h$  and a soft label  $G_s$ , defined as follows:

$$\begin{aligned} G_h(x) &= \begin{cases} 1, & \text{if } P_f(x) \geq 0.5, \\ 0, & \text{otherwise,} \end{cases} \\ G_s(x) &= [G_h(x), P_f(x)], \end{aligned} \quad (2)$$

where  $x$  represents a spatial location, and  $[\cdot]$  denotes channel-wise concatenation.

During training, the model prediction  $O$  for an input  $I$  is used to compute a pixel-wise entropy map, representing prediction uncertainty. This entropy is normalized to  $E_n \in [0, 1]$ . Pixels with  $E_n < 0.3$  are considered reliable and supervised using the hard label  $G_h$  via cross-entropy loss. For uncertain pixels, we employ a self-training scheme by aligning the model outputs with  $G_s$  through Kullback–Leibler (KL) divergence.

### C. Bi-Temporal Building Change Detection

This stage follows a similar structure to binary building extraction, consisting of three key steps: (1) Change Detection Model Training, (2) Pseudo-Label Generation, and (3) Building-Guided Low-Uncertainty Pseudo-Label Training.

**1) Change Detection Model Training:** We adopt the UACD [8] architecture, removing its uncertainty modeling components to reduce computational overhead while retaining the core bi-temporal fusion mechanism. The model is trained on the image pairs from the official training split and evaluated on the Phase 1 validation set. The top five models are selected for ensemble construction to generate pseudo-labels for the test set in the next stage.

**2) Pseudo-Label Generation:** We generate pseudo-labels  $P_{cd}$  for the bi-temporal image pairs in the test set using the ensemble of top-performing models combined with test-time augmentation. The procedure mirrors that of the binary building extraction stage. Each pseudo-label  $P_{cd}$  is a four-channel probability map, where each channel corresponds to the predicted probability of a specific damage category.

**3) Building-Guided Low-Uncertainty Pseudo-Label Training:** To address noise in the pseudo-labels, we propose a building-guided low-uncertainty training strategy. Given the pseudo-probability map  $P_{cd}$ , we define a hard label  $G_h^{cd}$  and a soft label  $G_s^{cd}$  as:

$$\begin{aligned} G_h^{cd}(x) &= \arg \max_c P_{cd}^{(c)}(x), \\ G_s^{cd}(x) &= [G_h^{cd}(x), P_{cd}(x)], \end{aligned} \quad (3)$$

where  $x$  denotes a pixel location and  $[\cdot]$  indicates channel-wise concatenation.

During training, bi-temporal features extracted by the encoder (PVT-v2-b2) are modulated using the binary building map obtained in the first stage. This guidance helps enhance building-related features and preserves structural consistency. Given a bi-temporal input pair  $(I_{pre}, I_{post})$ , the model generates an output  $O^{cd}$ , from which a pixel-wise entropy map is derived to estimate prediction uncertainty. The normalized entropy  $E_n^{cd} \in [0, 1]$  is used to distinguish between reliable and uncertain regions.

Pixels with  $E_n^{cd} < 0.3$  are treated as confident and supervised using cross-entropy loss with hard labels  $G_h^{cd}$ . In contrast, for high-uncertainty pixels, we apply a self-regularization scheme based on KL divergence with soft labels  $G_s^{cd}$ .

## III. EXPERIMENTS

### A. Experimental Settings

All experiments were conducted using PyTorch 1.8.1 (CUDA 11.1) on an NVIDIA RTX 3090 GPU (24 GB). Models were trained for 100 epochs with the AdamW [13] optimizer ( $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ) and a cosine annealing schedule starting at  $10^{-4}$ . The loss function combines cross-entropy and KL divergence with a fixed 2 : 1 weight ratio, supervising low- and high-uncertainty regions, respectively. All images were kept at their original  $1024 \times 1024$  resolution. For building extraction, we ensembled PVT-v2-b2 and PVT-v2-b3 models trained with our Low-Uncertainty Pseudo-Label framework, and applied horizontal flip-based test-time augmentation (TTA). For change detection, we used a single PVT-v2-b2 model trained with the Building-Guided strategy, also evaluated with horizontal flip TTA.

### B. Visual Comparison

As illustrated in Fig. 2, we present two representative examples to demonstrate the superiority of our proposed method. It is evident that the compared change detection method, ChangeMamba [14], fails to identify many damaged buildings. Although UACD achieves more accurate detection, it still misclassifies some damaged buildings as destroyed ones. In

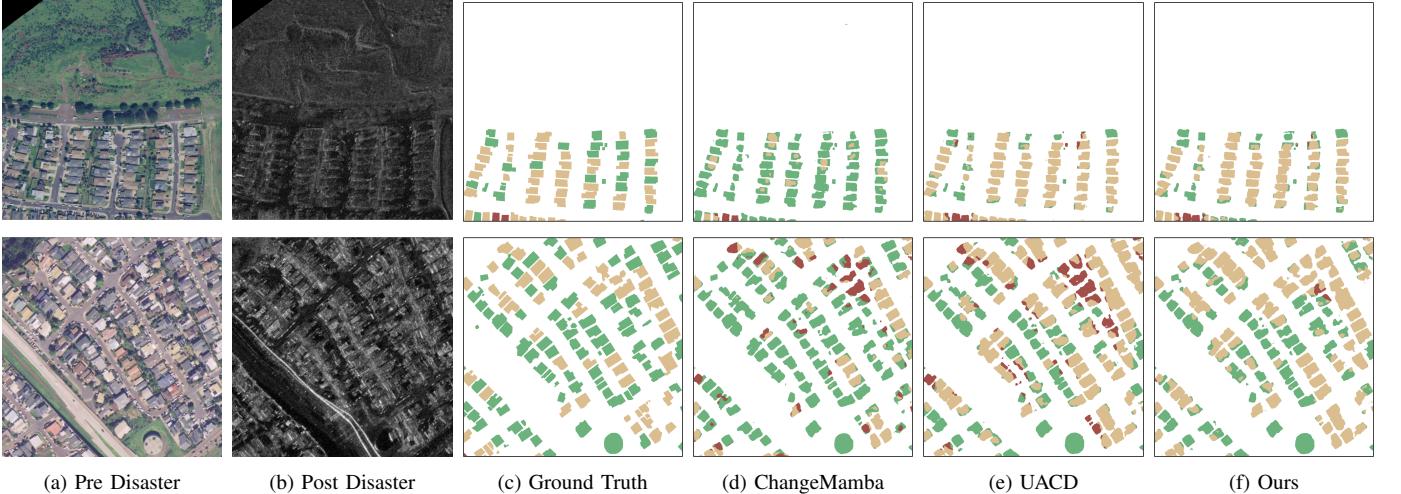


Fig. 2. Visual comparison between our proposed framework and the compared methods. Notably, white, green, yellow, and red denote the background, intact buildings, damaged buildings, and destroyed buildings, respectively.

TABLE I  
QUANTITATIVE RESULTS ON THE DEVELOPMENT PHASE AND TEST PHASE.

Method	Development Phase IoU (%)				mIoU
	Background	Intact	Damaged	Destroyed	
CGNet	96.722	79.519	31.000	70.022	69.316
ChangeMamba	96.737	80.022	38.181	71.464	71.601
UACD	96.756	80.076	39.575	73.148	72.389
Ours (w/o bg)	96.811	80.700	44.618	74.105	74.105
Ours (with bg)	<b>96.845</b>	<b>80.958</b>	<b>45.321</b>	<b>74.637</b>	<b>74.440</b>

Method	Test Phase IoU (%)				mIoU
	Background	Intact	Damaged	Destroyed	
Ours (with bg)	<b>97.995</b>	<b>67.304</b>	<b>9.748</b>	<b>42.086</b>	<b>54.283</b>

bg denotes the use of the proposed building-guided strategy.

contrast, our building-guided pseudo-label learning strategy effectively leverages building information and utilizes low-uncertainty pseudo labels to guide the model in learning richer damage-aware representations.

### C. Quantitative Comparison

As shown in Tab. I, we compare the performance of our proposed framework with several state-of-the-art (SOTA) change detection methods, including Change Guiding Network (CGNet[15]), ChangeMamba [14], and the Uncertainty-aware Change Detection Framework (UACD [8]). We also evaluate the effectiveness of our building-guided pseudo-label learning strategy by comparing the results with and without its incorporation. The results clearly demonstrate that our method significantly outperforms the competing approaches in terms of Intersection over Union (IoU) across all four categories.

In the development phase, where validation scenes are similar to training data, all methods achieve relatively strong performance in mapping damaged buildings. However, during testing, with differing disaster scenes, most methods see a performance drop. Our framework, however, maintains high accuracy and robustness, helping our team achieve first place in the 2025 IEEE GRSS Data Fusion Contest.

### IV. DISCUSSION

To evaluate the effectiveness of the low-uncertainty pseudo-label training strategy, we visualize the uncertainty associated

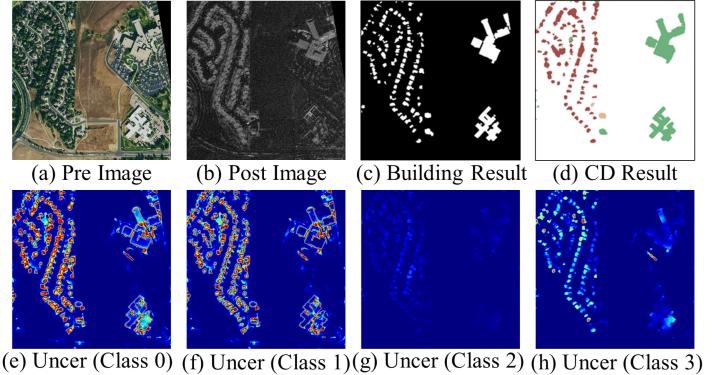


Fig. 3. Visualization of pseudo-label uncertainty (uncer).

with the pseudo-labels. As shown in Fig. 3, the pseudo-labels exhibit considerable uncertainty, particularly along building boundaries. However, our strategy selectively leverages the reliable, low-uncertainty predictions to guide the model in learning more accurate and damage-aware representations.

Distinguishing partially damaged buildings is difficult due to their blend of intact features and damage, along with challenges like segregated datasets, varying damage patterns, and resolution differences. As seen in the first phase, without isolated validation and test scenarios, the “damaged” class had significantly lower IoU, reflecting the model’s struggle in diverse disaster environments. Our uncertainty-guided approach, incorporating building priors, showed slight improvement. However, to further enhance detection, generative methods to synthesize more partially damaged samples could provide the model with better training data for realistic damage patterns.

### V. CONCLUSION

In this paper, we proposed a building-guided low-uncertainty pseudo-label training framework for cross-modal bi-temporal damaged building mapping. Our method achieved first place in Track 2 of the 2025 IEEE GRSS Data Fusion Contest. In future work, we plan to further explore strategies for improving the generalization ability of damaged building mapping across diverse disaster scenes.

## REFERENCES

- [1] Z. Zheng, Y. Zhong, J. Wang, A. Ma, and L. Zhang, “Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters,” *Remote Sensing of Environment*, vol. 265, p. 112636, 2021.
- [2] H. Huang, W. He, H. Zhang, Y. Xia, and L. Zhang, “Stfdiff: Remote sensing image spatiotemporal fusion with diffusion models,” *Information Fusion*, p. 102505, 2024.
- [3] H. Chen, N. Yokoya, and M. Chini, “Fourier domain structural relationship analysis for unsupervised multimodal change detection,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 198, pp. 99–114, 2023.
- [4] Y. Guo, W. He, Y. Xia, and H. Zhang, “Blind single-image-based thin cloud removal using a cloud perception integrated fast fourier convolutional network,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 206, pp. 63–86, 2023.
- [5] J. Li, H. Huang, W. He, H. Zhang, and L. Zhang, “Overcoming the uncertainty challenges in flood rapid mapping with multi-source optical data,” in *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2024, pp. 780–784.
- [6] H. Huang, J. Li, W. He, H. Zhang, and L. Zhang, “Overcoming the uncertainty challenges in flood rapid mapping with sar data,” in *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2024, pp. 767–771.
- [7] H. Chen, J. Song, O. Dietrich, C. Broni-Bediako, W. Xuan, J. Wang, X. Shao, Y. Wei, J. Xia, C. Lan *et al.*, “Bright: A globally distributed multimodal building damage assessment dataset with very-high-resolution for all-weather disaster response,” *arXiv preprint arXiv:2501.06019*, 2025.
- [8] J. Li, W. He, Z. Li, Y. Guo, and H. Zhang, “Overcoming the uncertainty challenges in detecting building changes from remote sensing images,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 220, pp. 1–17, 2025.
- [9] J. Li, W. He, W. Cao, L. Zhang, and H. Zhang, “Uanet: An uncertainty-aware network for building extraction from remote sensing images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–13, 2024.
- [10] J. Li, Y. Wei, T. Wei, and W. He, “A comprehensive deep-learning framework for fine-grained farmland mapping from high-resolution images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–15, 2025.
- [11] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, “Pvt v2: Improved baselines with pyramid vision transformer,” *CVM*, vol. 8, no. 3, pp. 415–424, 2022.
- [12] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.
- [13] I. Loshchilov and F. Hutter, “Fixing weight decay regularization in adam,” 2018.
- [14] H. Chen, J. Song, C. Han, J. Xia, and N. Yokoya, “Changemamba: Remote sensing change detection with spatiotemporal state space model,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–20, 2024.
- [15] C. Han, C. Wu, H. Guo, M. Hu, J. Li, and H. Chen, “Change guiding network: Incorporating change prior to guide change detection in remote sensing imagery,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 8395–8407, 2023.