Master's Thesis Presentation for the 145th Dissertation Evaluation

A Study on Efficient Remote Sensing Change Detection with Adaptive Frequency Masking for Pseudo-Change Suppression

유사 변화 억제를 위한 적응형 주파수 마스킹 기반 효율적인 원격 탐사 변화 탐지에 관한 연구

Kyoungmin Shin

shinkm1104@cau.ac.kr

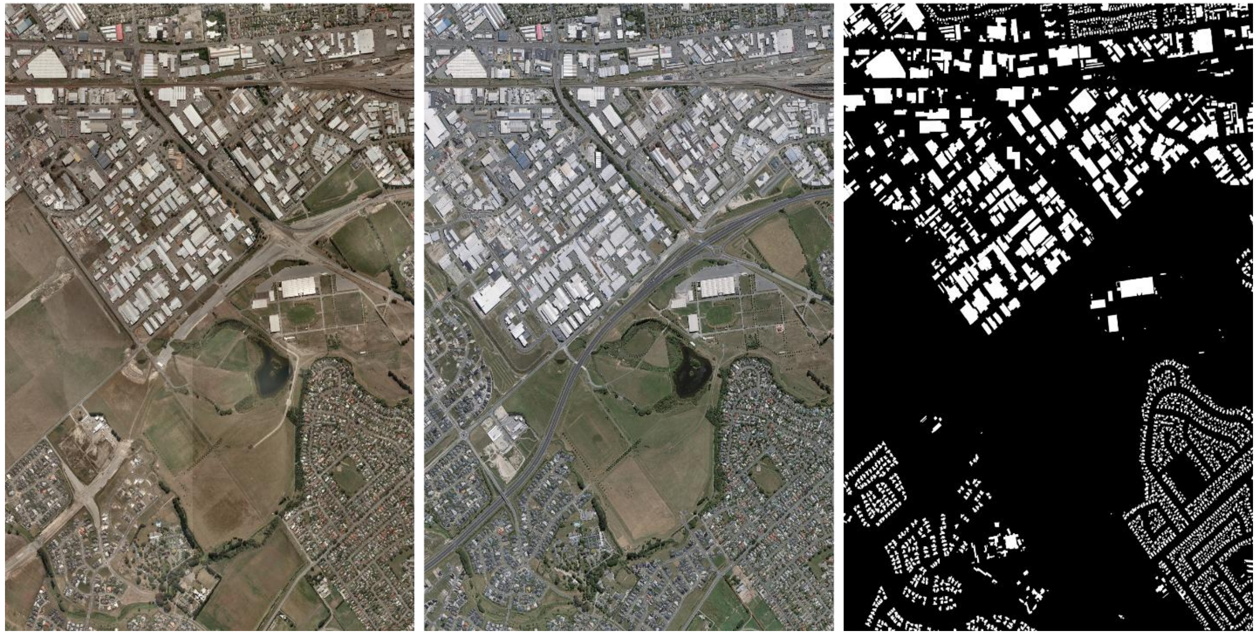
8th April, 2026

Contents

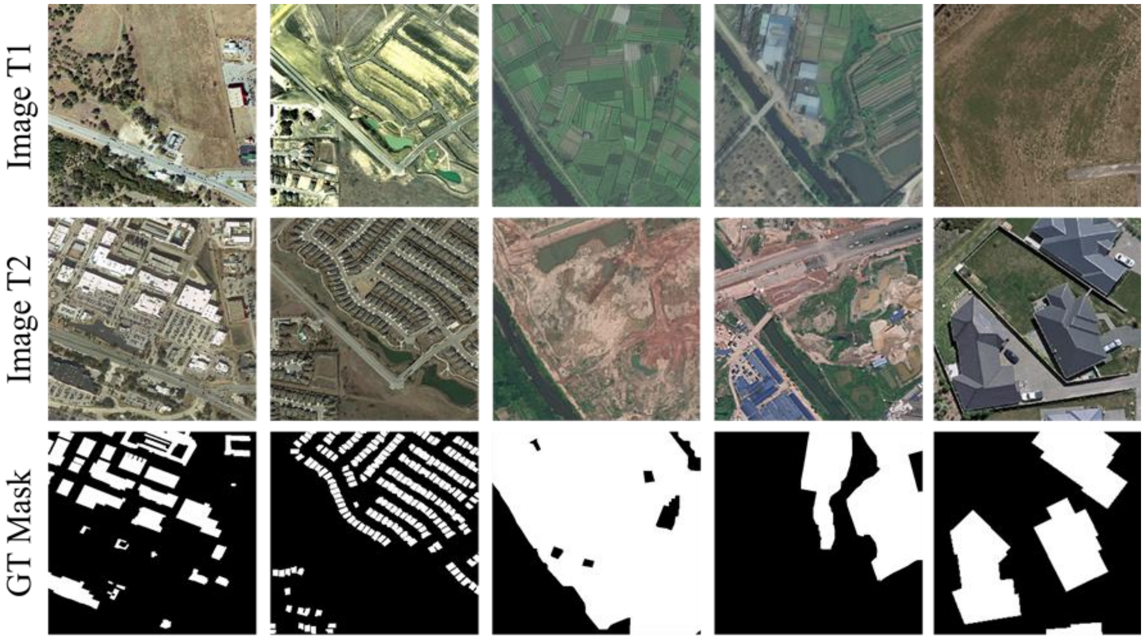
Introduction

Change Detection in Remote Sensing Images

- Remote Sensing Change Detection (RS-CD) is the process of identifying observable differences between images of the same geographical area captured at different times, focusing on notable changes rather than subtle variations [1,2].
- RS-CD has been widely used in disaster monitoring, urban planning, and environmental management, where it provides important information for analyzing dynamic environments and supporting decision-making [3].



[Figure 1] Example of Large-Scale RS-CD



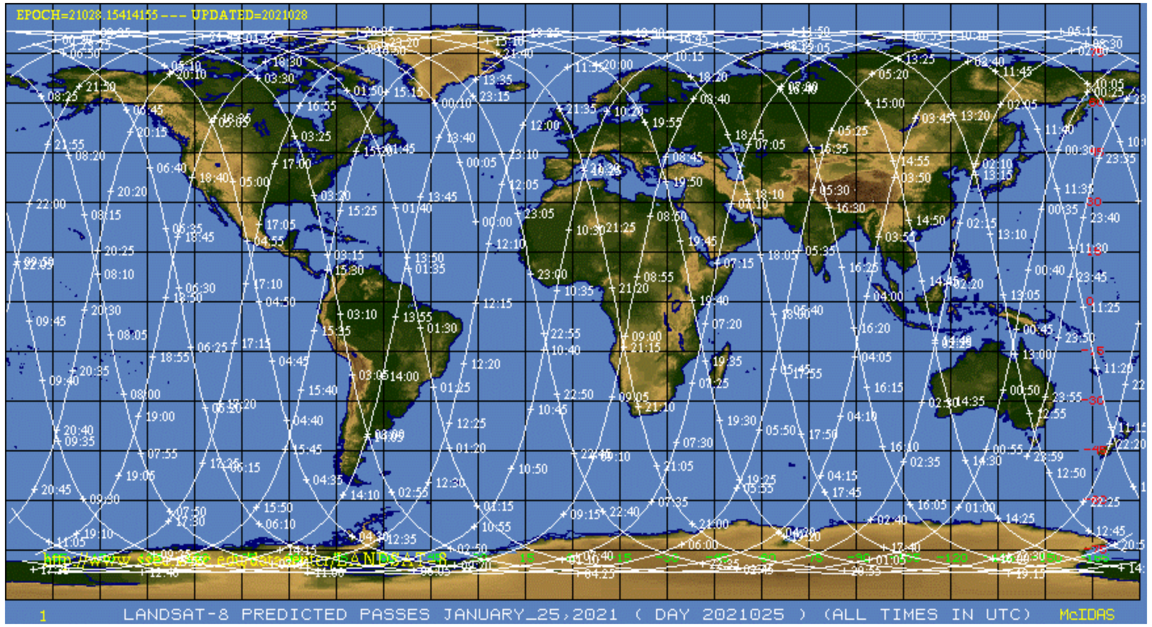
[Figure 2] Examples of application-dependent change definitions

[1] G. Cheng et al., "Change Detection Methods for Remote Sensing in the Last Decade: A Comprehensive Review," Remote Sensing, vol. 16, p. 2355, 2024.
[2] M. Hussain, D. Chen, A. Cheng, H. Wei and D. Stanley, "Change Detection from Remotely Sensed Images: From Pixel-Based to Object-Based Approaches," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 80, pp. 91-106, 2013.
[3] D. Hong et al., "Deep Learning for Change Detection in Remote Sensing: A Comprehensive Survey," IEEE Geoscience and Remote Sensing Magazine, vol. 13, no. 3, pp. 164-189, 2025.

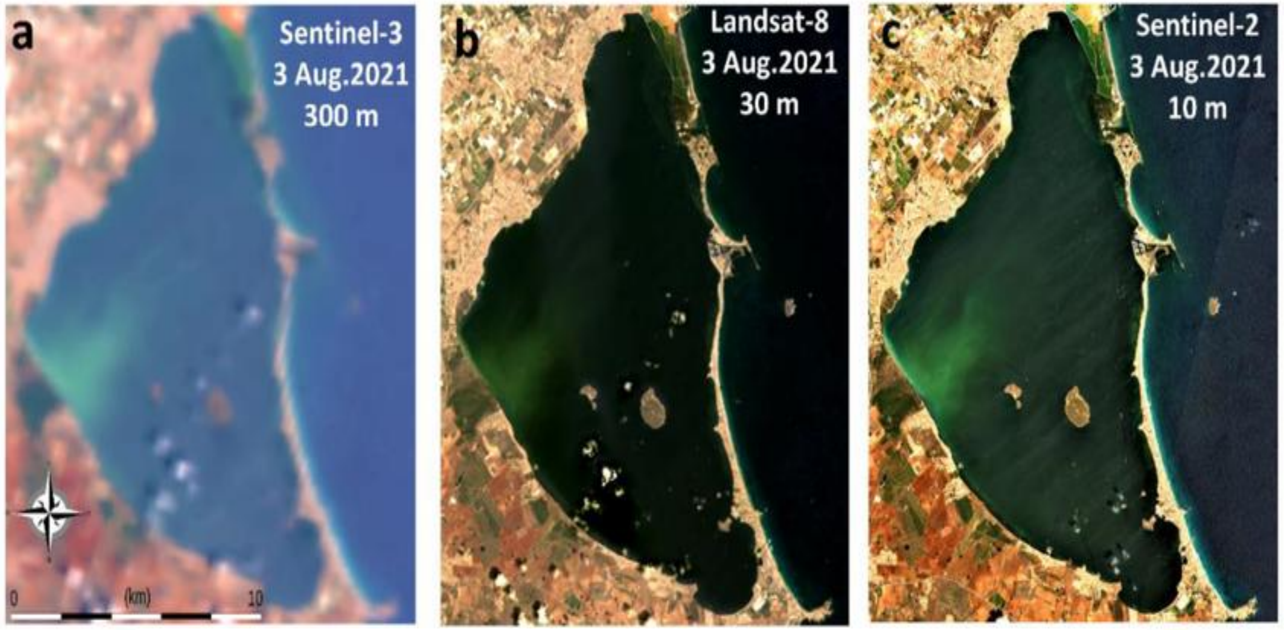
Introduction

Data Acquisition Constraints and Requirements for CD

- Continuous acquisition of high-resolution satellite imagery enables periodic observation, yet slow inference speeds lead to the accumulation of unprocessed data, impeding rapid time-series analysis and response [4].
- Furthermore, orbital constraints often prevent acquiring images from the same sensor, which may necessitate the use of heterogeneous imagery from multiple satellites with varying spectral and spatial resolutions for CD [5].



[Figure 3] Revisit Mismatch Caused by Orbital Constraints

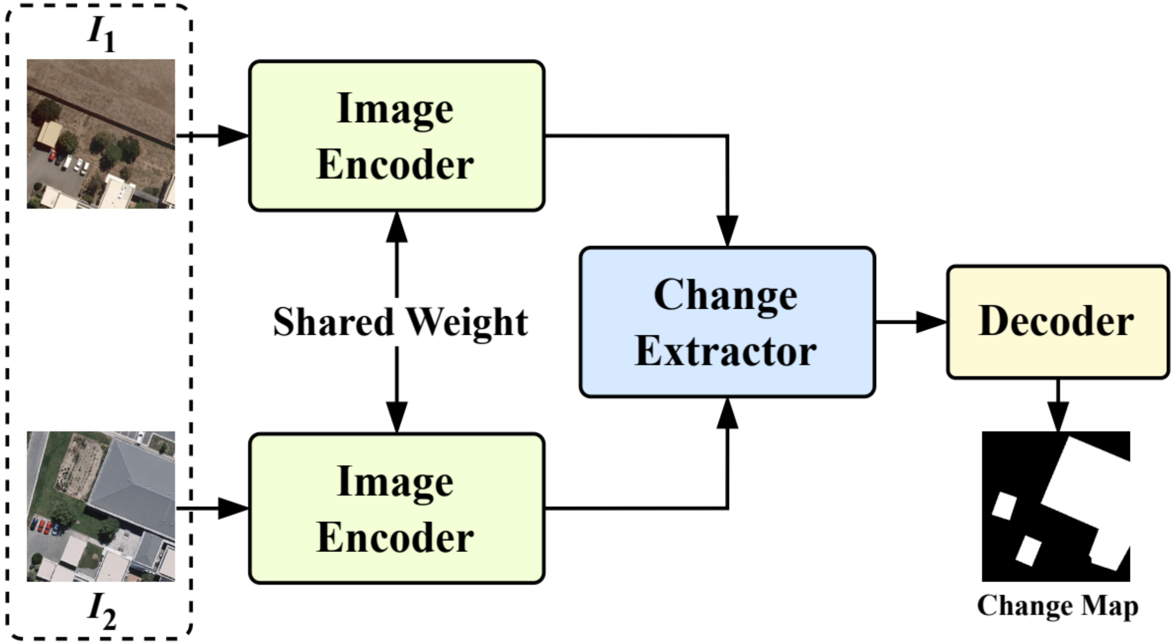


[Figure 4] Cross-Sensor Appearance Inconsistency in RS-CD [6]

[4] Z. Zhu, "Change Detection Using Landsat Time Series: A Review of Frequencies, Preprocessing, Algorithms, and Applications," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 130, pp. 370-384, 2017.
[5] Yang et al., "Edge-guided hierarchical network for building change detection in remote sensing images," Applied Sciences, vol. 14, no. 13, p. 5415, 2024.
[6] I. Caballero et al., "Use of the Sentinel-2 and Landsat-8 Satellites for Water Quality Monitoring: An Early Warning Tool in the Mar Menor Coastal Lagoon," Remote Sensing, vol. 14, p. 2744, 2022.

Objective and Standard Procedure

- CD aims to detect and highlight differences between images acquired at different times, focusing on notable changes while disregarding minor variations, and requires rapid inference to meet the mission time constraints of real-world satellite operations [4].
- The standard procedure of RS-CD typically involves extracting features from bi-temporal images through an encoder, integrating temporal differences using a change extractor, and generating the final change map via a decoder to localize detected changes [1].



[Figure 5] Standard RS-CD Pipeline

[Figure 6] Example of RS-CD Results

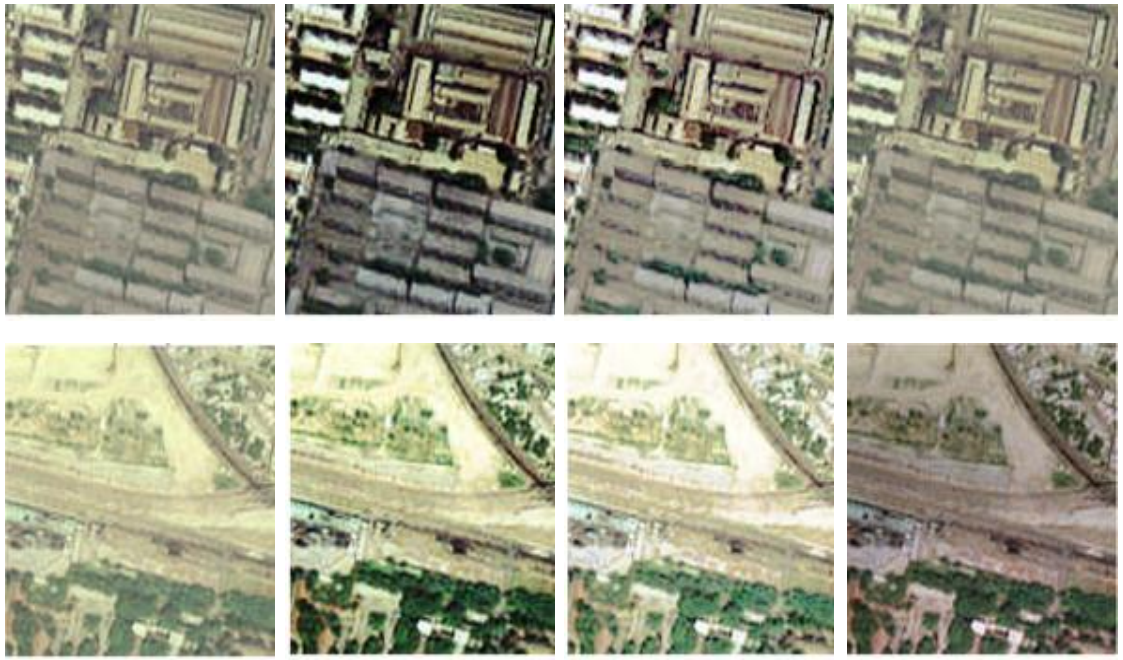
[1] G. Cheng et al., "Change Detection Methods for Remote Sensing in the Last Decade: A Comprehensive Review," Remote Sensing, vol. 16, p. 2355, 2024.
[4] Z. Zhu, "Change Detection Using Landsat Time Series: A Review of Frequencies, Preprocessing, Algorithms, and Applications," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 130, pp. 370-384, 2017.

Introduction

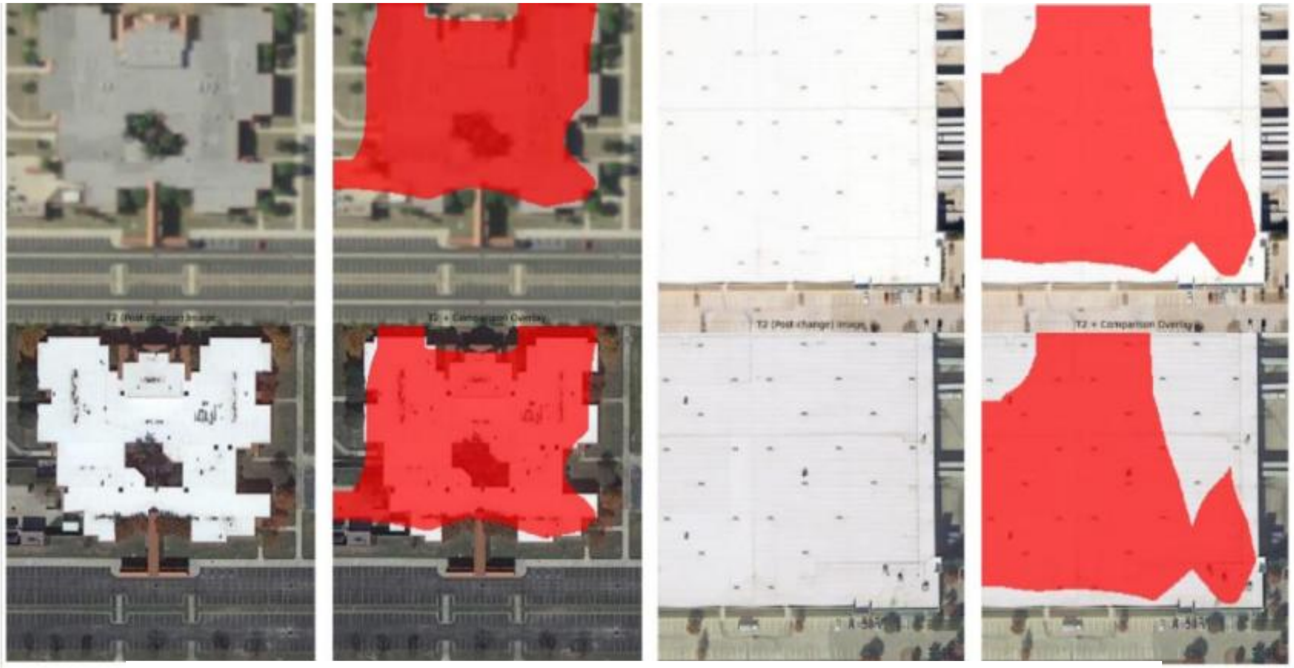
Ideal Goal and Challenge

- The ideal goal is to accurately identify genuine semantic changes by effectively suppressing non-semantic pseudo-changes caused by environmental conditions, seasonal variations, and sensor-specific characteristics [1].

- Efficient backbones fail to distinguish pseudo-changes distributed across the entire image because their small kernels only capture local details, missing the global context needed to recognize such widespread environmental patterns.



[Figure 7] Environmental Variations Causing Pseudo-Change



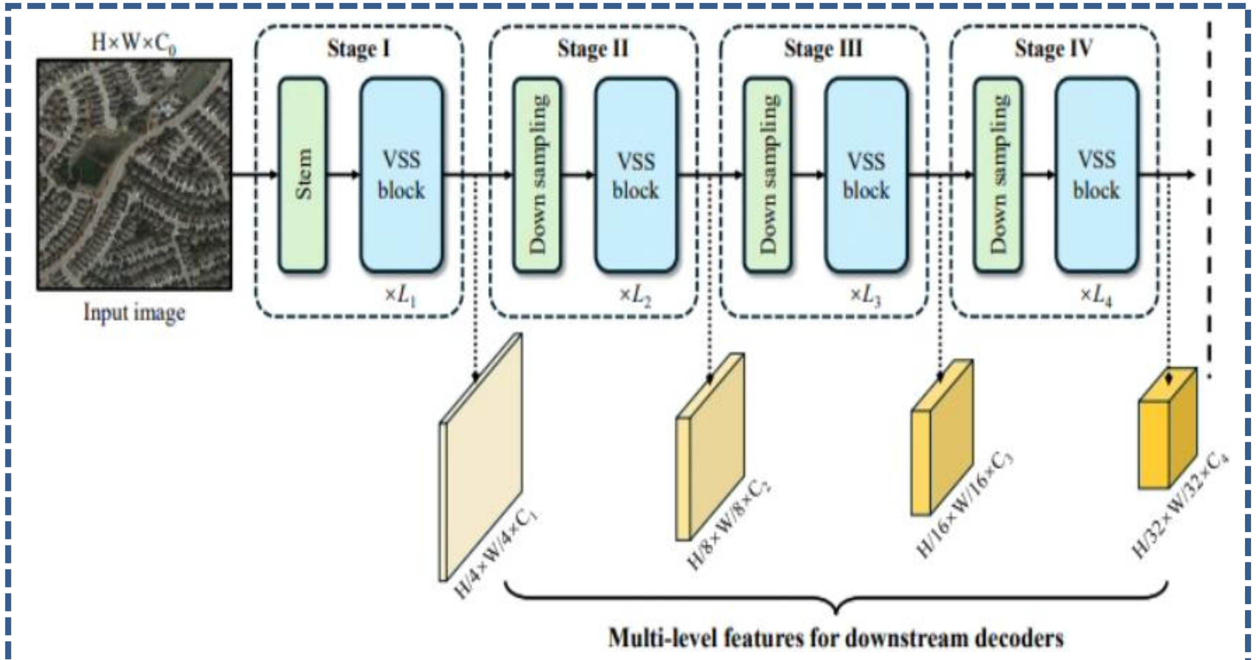
[Figure 8] Vulnerability of Efficient Backbones to Pseudo-Change

[1] G. Cheng et al., "Change Detection Methods for Remote Sensing in the Last Decade: A Comprehensive Review," Remote Sensing, vol. 16, p. 2355, 2024.

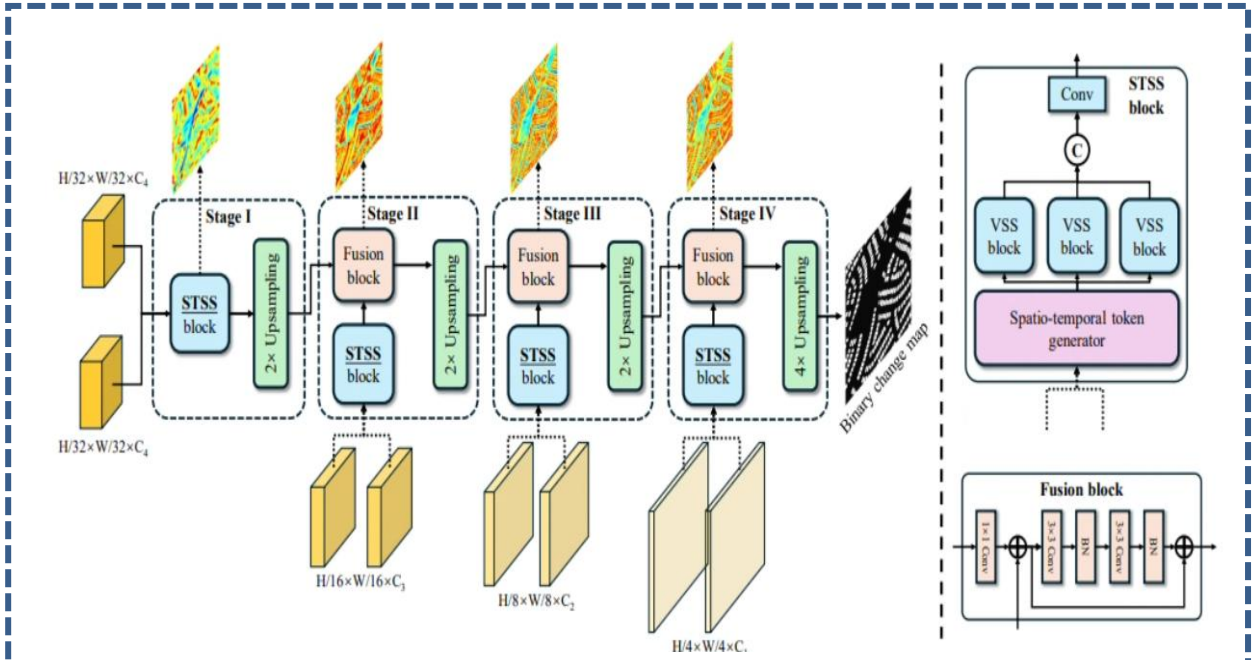
Related Work

ChangeMamba: Remote Sensing Change Detection With Spatio-Temporal State Space Model [7]

- ChangeMamba introduces the MambaBCD framework which utilizes a Visual Mamba encoder to learn global spatial context and Spatio-Temporal State Space (STSS) blocks to capture complex spatio-temporal relationships [7].
- Specialized Mamba encoders and STSS blocks introduce high architectural complexity and computational overhead, resulting in lower operational efficiency compared to models utilizing standard, lightweight 2D CNN backbones.



[Figure 17] The Overview of ChangeMamba Encoder



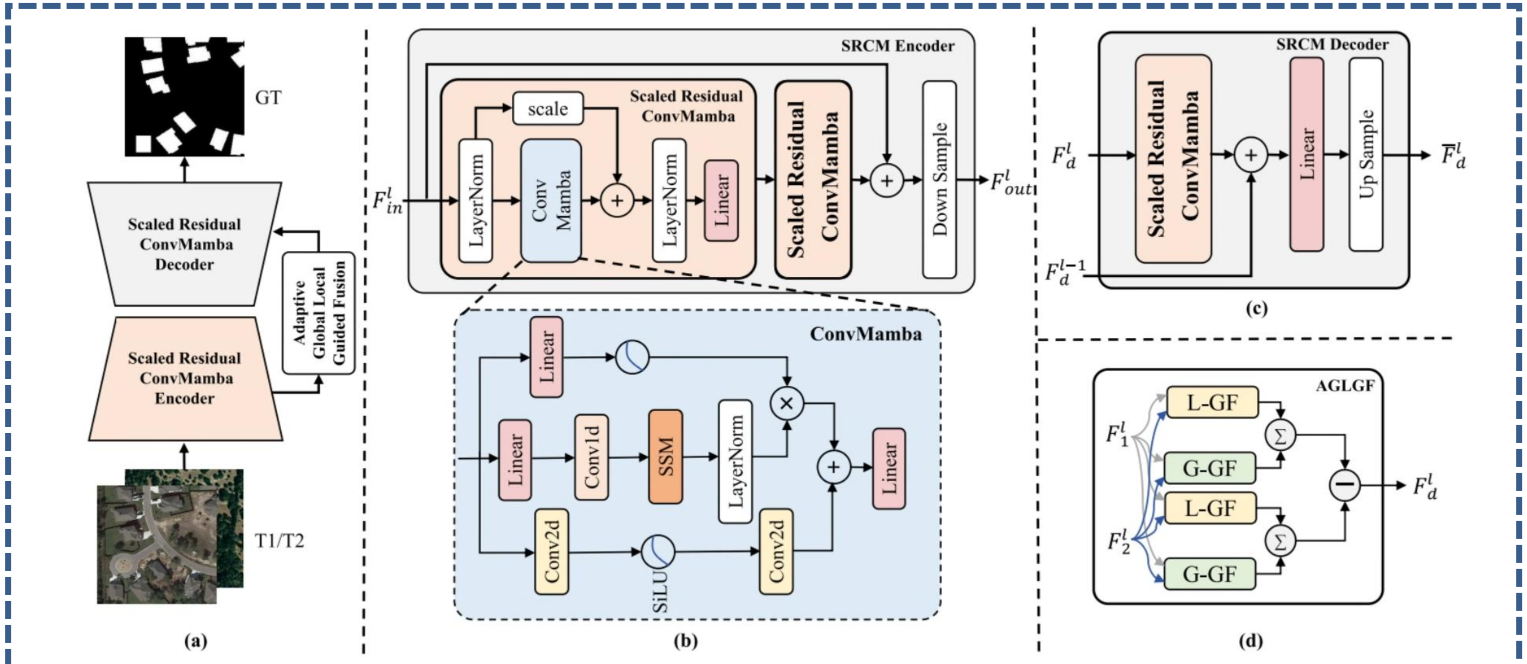
[Figure 18] The Overview of ChangeMamba Decoder

[7] Chen et al. "Changemamba: Remote sensing change detection with spatio-temporal state space model." IEEE Transactions on Geoscience and Remote Sensing (2024).

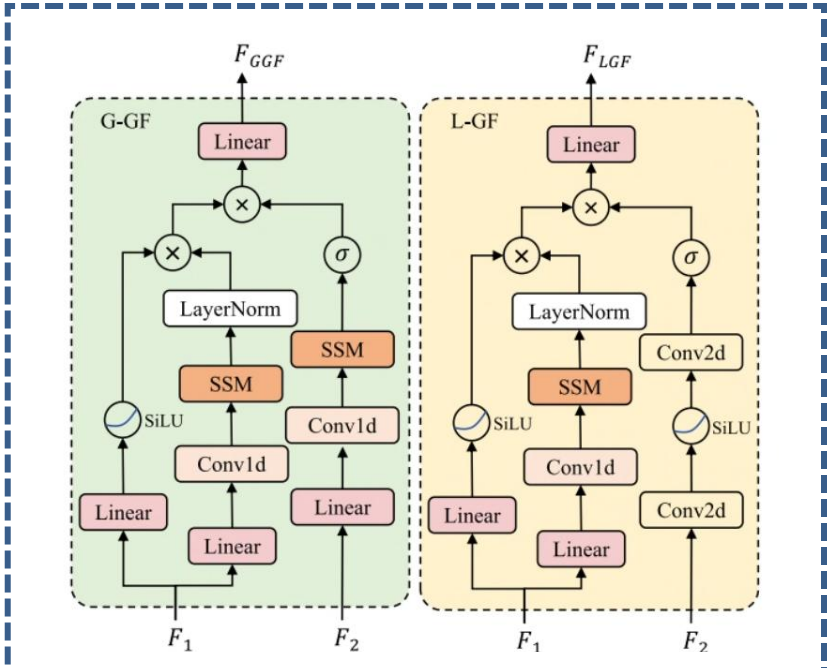
Related Work

CDMamba: Incorporating Local Clues Into Mamba for Remote Sensing Image Binary Change Detection [8]

- CDMamba constructs a backbone using Scaled Residual ConvMamba (SRCM) blocks to combine Mamba’s global modeling with convolutional local clues, employing Adaptive Global-Local Guided Fusion (AGLGF) for effective interaction [8].
- Structural complexity of multi-branch SRCM blocks and the direct processing of original-resolution images lead to excessive computational costs and low operational efficiency despite its lightweight design goal.



[Figure 15] The Architecture of CDMamba



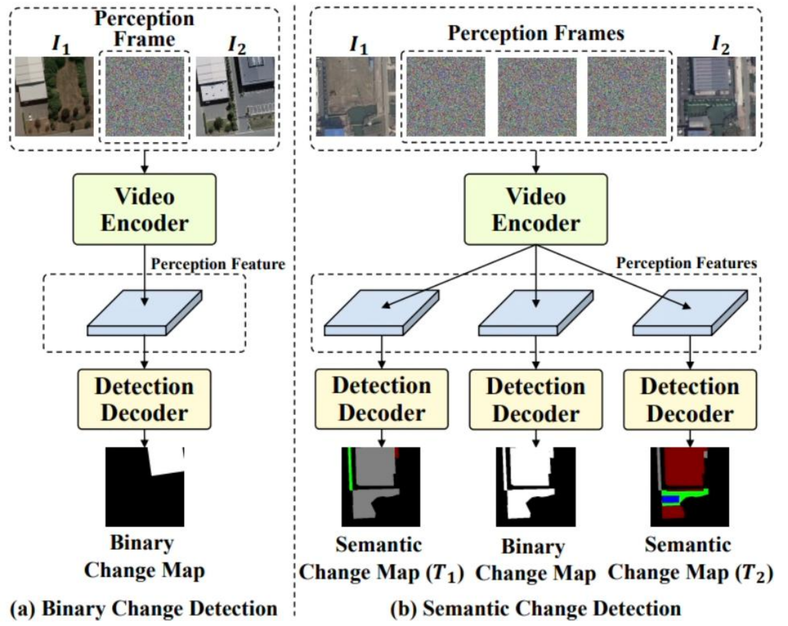
[Figure 16] The Overview of AGLGF

[8] Zhang et al. "CDMamba: Incorporating local clues into mamba for remote sensing image binary change detection." IEEE Transactions on Geoscience and Remote Sensing (2025).

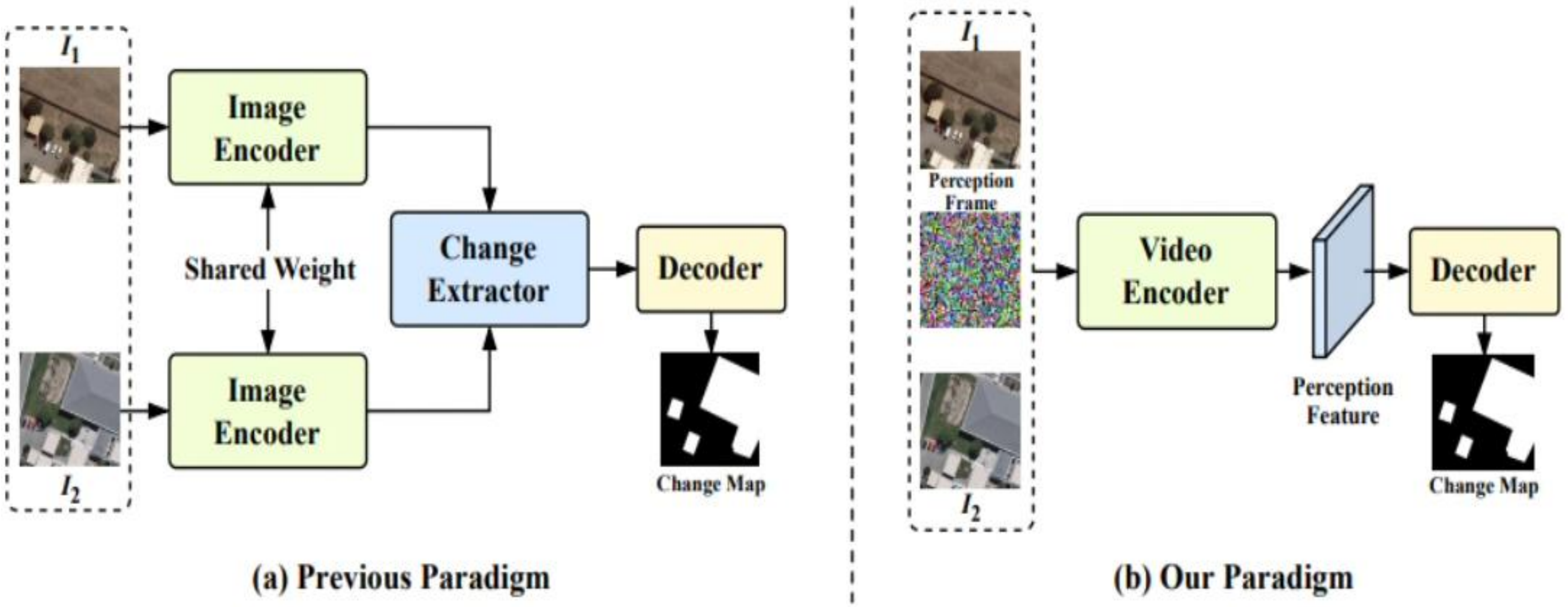
Related Work

Change3D: Revisiting Change Detection and Captioning from A Video Modeling Perspective [9]

- Change3D reconceptualizes bi-temporal change detection as video modeling by inserting learnable perception frames into a unified framework, enabling the video encoder to directly extract differences without complex, task-specific change extractors [9].
- Temporal modeling and 3D operations in video encoders demand higher memory bandwidth and hardware resources compared to standard 2D convolution layers and 2D cheap backbone models.



[Figure 9] The Overview of Change3D



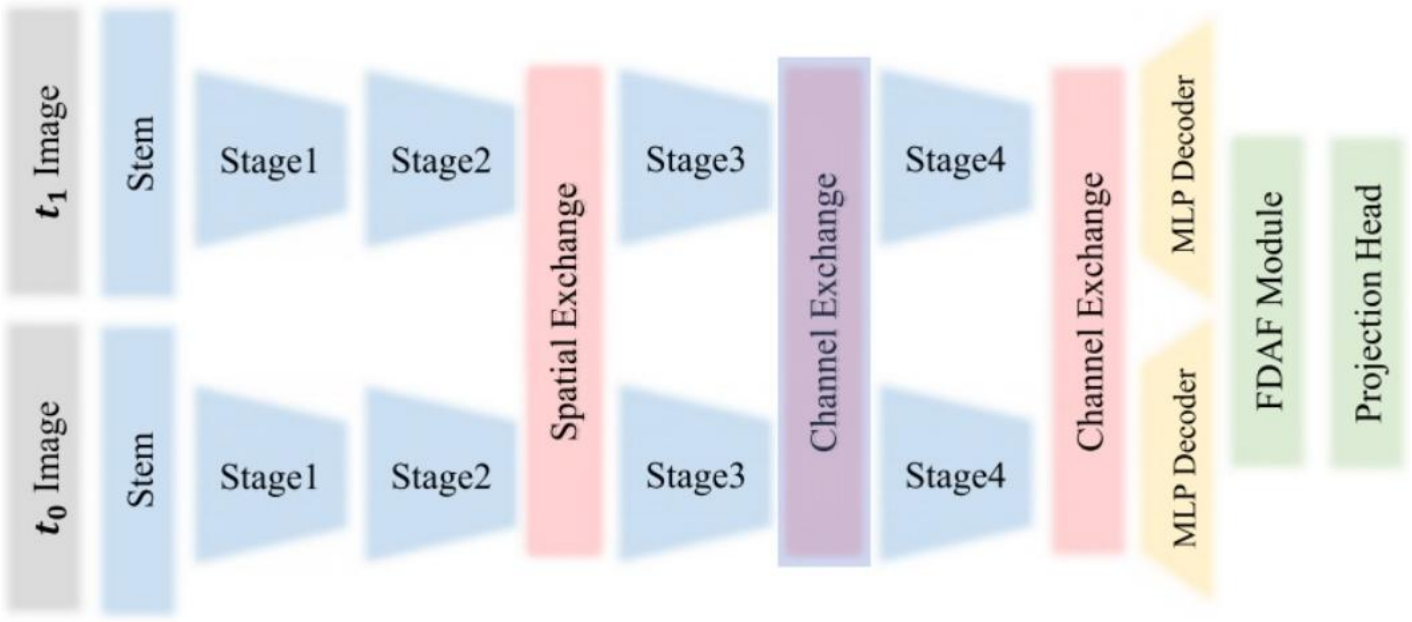
[Figure 10] Comparison Change3D with Standard Procedure

[9] Zhu et al. "Change3D: Revisiting Change Detection and Captioning from A Video Modeling Perspective." Proceedings of the Computer Vision and Pattern Recognition Conference. 2025.

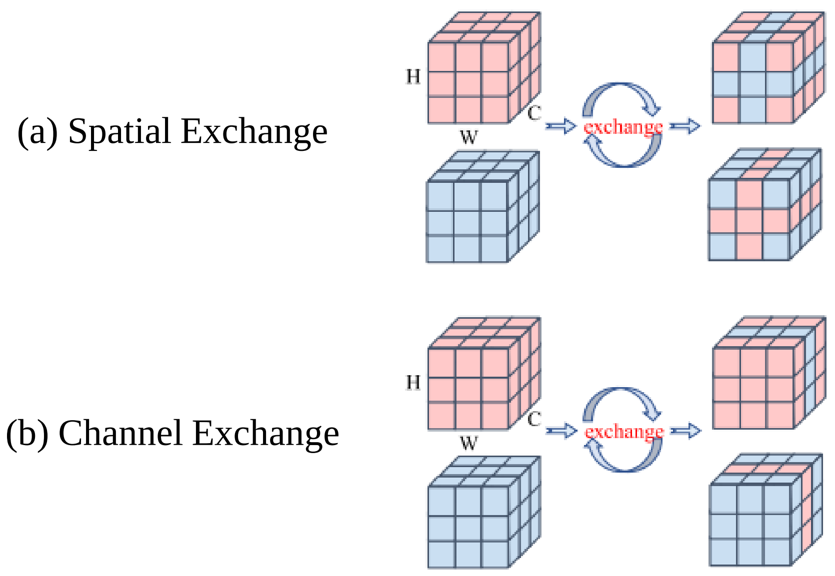
Related Work

Changer: Feature Interaction is What You Need for Change Detection [10]

- Changer facilitates bi-temporal interaction by incorporating Aggregation-Distribution (AD) for channel-wise reweighting and Feature Exchange (Ex) for zero-cost spatial/channel communication between encoders [10].
- Changer strengthens bi-temporal interaction through additional feature exchange modules, this interaction-centered design increases structural complexity without directly suppressing global stylistic pseudo-changes.



[Figure 11] The Overview of Changer-Ex



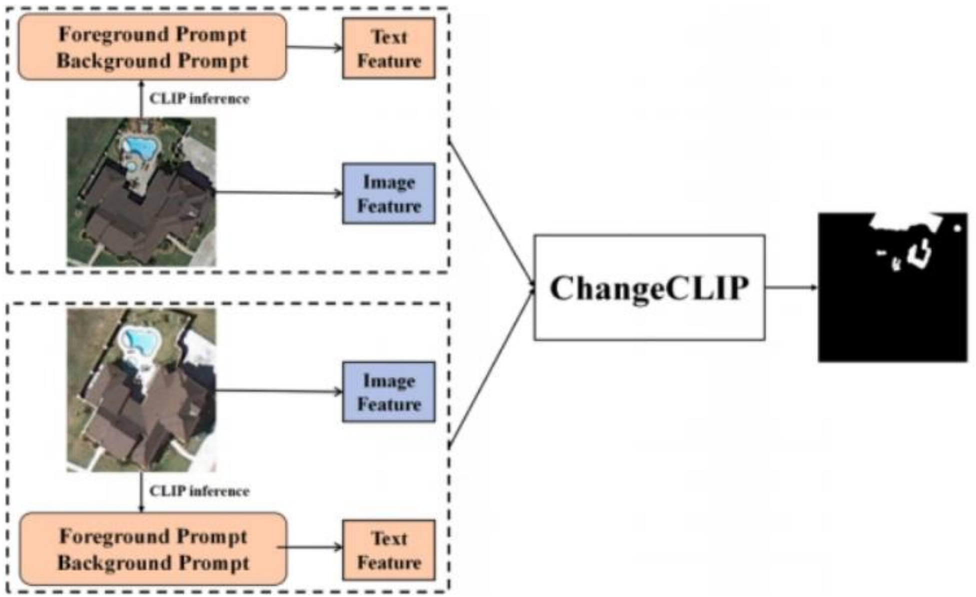
[Figure 12] Diagram of Spatial & Channel Exchange

[10] Fang et al. "Changer: Feature interaction is what you need for change detection." IEEE Transactions on Geoscience and Remote Sensing 61 (2023): 1-11.

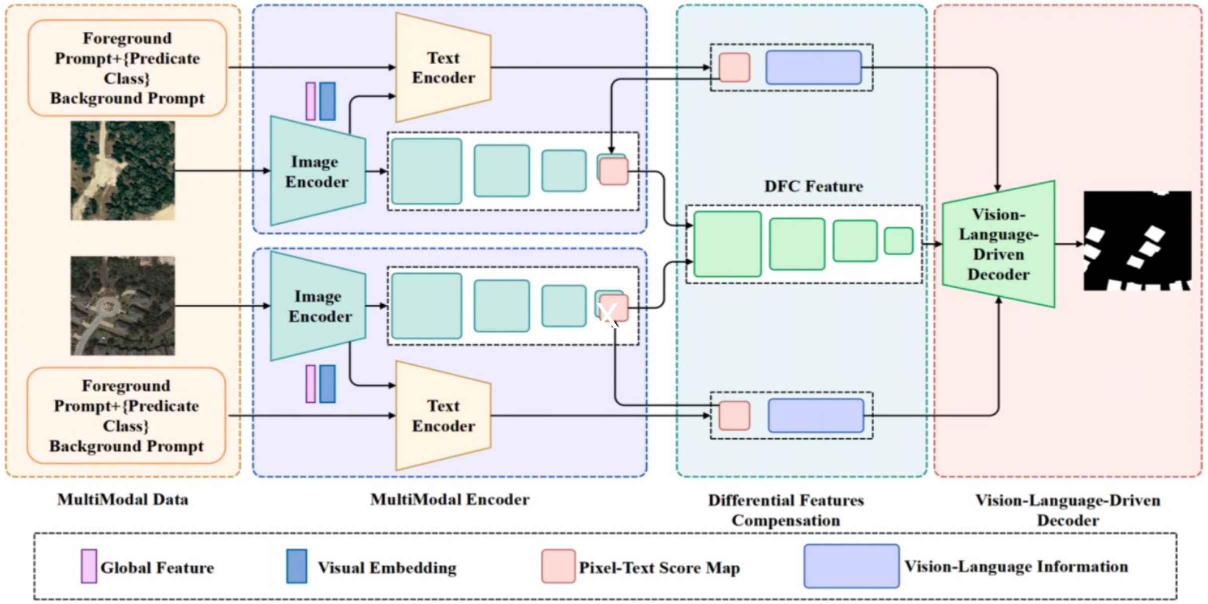
Related Work

ChangeCLIP: Remote sensing change detection with multimodal vision-language representation learning [11]

- ChangeCLIP leverages multimodal representation learning by encoding 56 land-cover category prompts with CLIP and integrating these textual priors with visual features through a vision-language-driven decoder [11].
- Dependency on dual encoders for image and text along with the heavy computational burden of generating multi-level image-text score maps limits the operational efficiency required for rapid data processing.



[Figure 13] The Overview of ChangeCLIP



[Figure 14] The Main Idea of ChangeCLIP

[11] Dong et al. "ChangeCLIP: Remote sensing change detection with multimodal vision-language representation learning." ISPRS Journal of Photogrammetry and Remote Sensing 208 (2024): 53-69.

Related Work

Common Drawback and Hypothesis

- Reliance on complex specialized architectures for spatio-temporal modeling leads to excessive computational overhead and high latency, significantly reducing operational efficiency compared to standard CD models.
- Global stylistic pseudo-changes are characterized by distinct statistical patterns in the frequency domain, and adaptively masking these noise-equivalent components suppresses environmental interference to facilitate accurate semantic detection.

[Table 1] Summary of recent RS-CD methods, outlining their key contributions and a common drawback

Method	Venue	Summary Regarding the Challenge	Drawback
ChangeMamba [7]	TGRS 2024	ChangeMamba employs Mamba for long-range modeling; however, its sequential spatial scanning is inefficient for capturing image-wide statistical patterns, resulting in high computational costs relative to accuracy.	Reliance on complex specialized architectures for spatio-temporal modeling leads to excessive computational overhead and high latency, significantly reducing operational efficiency compared to standard CD models.
CDMamba [8]	TGRS 2025	CDMamba integrates State Space Models to bridge global and local contexts, but its complex scanning process imposes a heavy computational burden without a direct way to suppress large-scale environmental noise.	
Change3D [9]	CVPR 2025	Change3D treats bi-temporal images as video sequences to capture temporal flow, but its complex interaction layers incur high computational costs, hindering overall operational efficiency.	
Changer [10]	TGRS 2023	Changer utilizes feature exchange mechanisms for alignment; however, it remains computationally demanding and fails to effectively suppress widespread pseudo-changes within the limited receptive fields of spatial kernels.	
ChangeCLIP [11]	ISPRS 2024	ChangeCLIP incorporates large-scale multimodal priors for semantic consistency, yet its reliance on massive foundation architectures leads to excessive memory overhead and deployment inefficiency.	

Proposed Method

Notation and Definition

- RS-CD dataset is defined as $\mathcal{D} = \{(I_i^{T1}, I_i^{T2}, M_i\}_{i=1}^{N}$, where I_i^{T1} and $I_i^{T2} \in \mathbb{R}^{H\times W\times 3}$denote the bi-temporal images of the i-th sample acquired at times T_1 and T_2, and $M_i \in \{0,1\}^{H\times W}$is the corresponding Ground Truth (GT) change mask.
- The objective of RS-CD is to learn a mapping function f that transforms the bi-temporal input (I^{T1}, I^{T2}) into a predicted change mask $\widehat{M} = f(I^{T1}, I^{T2})$, which accurately segments the target semantic changes M while excluding non-change regions.

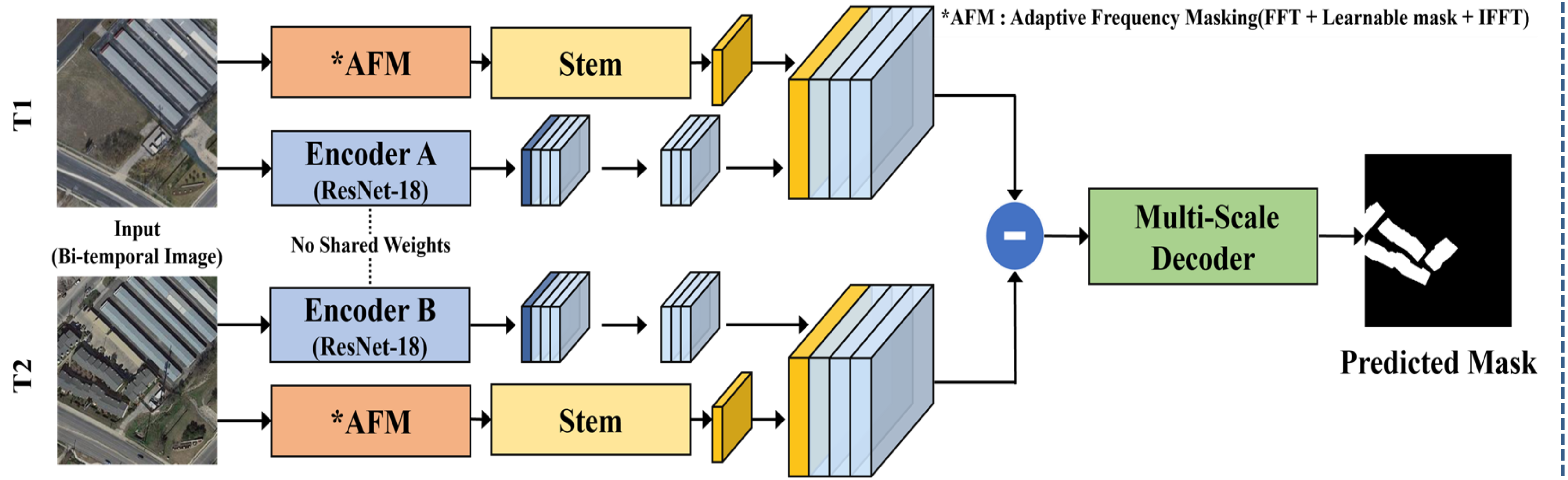
[Table 2] Notations and definitions used in the formulation of the RS-CD

Symbol	Name	Description
I_i^{T1}, I_i^{T2}	Bi-temporal images	RGB images acquired at times T_1, T_2
M	GT mask	Binary mask where 1 indicates a change and 0 indicates non-change
$\mathcal{D}$	CD dataset	$\mathcal{D} = \{(I_i^{T1}, I_i^{T2}, M_i\}_{i=1}^{N}$
$\mathcal{F}, \mathcal{F}^{-1}$	FFT / IFFT	2D Fast Fourier Transform and its inverse transform.
X^T	Shifted spectrum	X^T = fftshift($\mathcal{F}(I^T)$)
$\mathcal{M}_{amp}^T, \mathcal{M}_{pha}^T$	AFM masks	Learnable channel-specific amplitude and phase masks for temporal branch T
$\tilde{I}^T$	AFM-refined image	Refined image reconstructed from masked amplitude and phase by inverse FFT
F_ℓ^T	Backbone feature	Backbone feature at encoder stage$\ell \in \{2, 3, 4\}$
Δ_ℓ	Difference feature	Absolute bi-temporal difference feature used by the decoder

Proposed Method

Procedural Overview

- The proposed model incorporates an AFM module that transforms each temporal input into the frequency domain, applies separate amplitude and phase masks, and reconstructs frequency-refined representations to suppress pseudo-change.
- The model then computes absolute bi-temporal differences at four scales and aggregates them through a multi-scale decoder, where the AFM-adapter branch provides the shallowest decoder-side input instead of replacing the internal Stage-1 backbone computation.

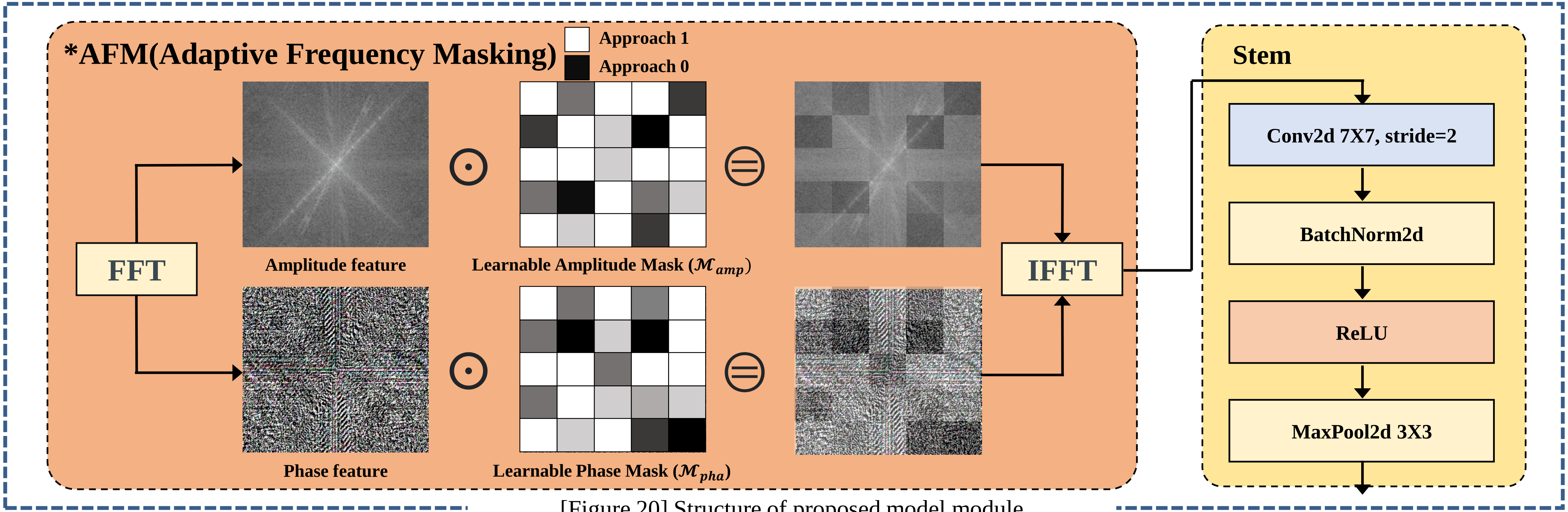


[Figure 19] Overall structure of proposed model framework

Proposed Method

Details

- The AFM module transforms the input into the frequency domain via FFT and independently filters amplitude and phase with learnable spectral masks to suppress pseudo-change, subsequently reconstructing the refined representation via IFFT.
- The stem block maps the frequency-refined image to a stage-1-scale shallow feature at 1/4 resolution, providing the shallowest decoder-side input rather than replacing the internal Stage-1 backbone computation.



[Figure 20] Structure of proposed model module

Proposed Method

Pseudo Algorithm

- The proposed framework applies AFM directly to each temporal image in the frequency domain and maps the refined image to a stage-1-scale shallow feature, providing the first decoder-side input for pseudo-change suppression.
- It then computes Δ_1 from the AFM-based shallow branch and $\Delta_2, \Delta_3, \Delta_4$ from backbone Stages 2–4, and aggregates these four difference features through a multi-scale decoder for final mask prediction.

Algorithm Proposed RS-CD Framework via Adaptive Frequency Masking(AFM)

Algorithm Full Framework with Adaptive Frequency Masking(AFM)

Input: Bi-temporal images $I^{T_1}, I^{T_2} \in \mathbb{R}^{H\times W\times 3}$
Parameters: Spectral masks $M^{T1}_{amp}, M^{T1}_{Pha}, M^{T2}_{amp}, M^{T2}_{Pha}$
Function: G: shallow adapter, E^T : temporal encoder, Dec: multi-scale decoder

1 **for each** $T \in \{T_1, T_2\}$ **do**
2 $X^T \leftarrow fftshift(\mathcal{F}(I^T))$
3 $A^T \leftarrow |X^T + \epsilon|,\ P^T \leftarrow \angle(X^T + \epsilon)$
4 $\bar{\mathcal{M}}^T_{amp} \leftarrow \sigma(M^T_{amp}), \bar{\mathcal{M}}^T_{pha} \leftarrow \sigma(M^T_{pha})$
5 $\tilde{A}^T \leftarrow A^T \odot \bar{\mathcal{M}}^T_{amp}, \tilde{P}^T \leftarrow P^T \odot \bar{\mathcal{M}}^T_{pha})$
6 $\tilde{X}^T \leftarrow polar(\tilde{A}^T, \tilde{P}^T)$
7 $\tilde{I}^T \leftarrow Re(\mathcal{F}^{-1}(ifftshift(\tilde{X}^T)))$
8 $Z^T \leftarrow G(\tilde{I}^T)$
9 $F^T \leftarrow E^T(I^T)$
10 **end for**
11 $\Delta_1 \leftarrow |Z^{T1} - Z^{T2}|$
12 **for** $i \leftarrow 2\ to\ 4$ **do**
13 $\Delta_i \leftarrow |F_i^{T_1} - F_i^{T2}|$
14 **end for**
15 $\widehat{M} \leftarrow Dec(\Delta_1, \ldots, \Delta_4)$
16 $\boldsymbol{return}\ \widehat{M}$

Line 1 → Process each temporal image independently
Line 2 → Convert the image to the shifted frequency spectrum
Line 3 → Separate the spectrum into amplitude and phase
Line 4 → Generate learnable amplitude and phase masks
Line 5 → Apply the masks to the spectral components
Line 6 → Reconstruct the filtered spectrum
Line 7 → Recover the refined image in the spatial domain
Line 8 → Produce a shallow AFM feature through the adapter
Line 9 → Extract deeper backbone features from the original image
Line 10 → Finish feature extraction for both temporal inputs
Line 11 → Compute the AFM-based first difference feature
Line 12 → Traverse backbone stages 2 to 4
Line 13 → Compute backbone-based difference features
Line 14 → Finalize all deeper difference features
Line 15 → Decode the four difference features into the final mask
Line 16 → Return the predicted binary change mask

Experimental Settings

Dataset

- Six benchmark datasets categorized into Urban, Agricultural, and Disaster domains to verify its generalization capability and consistent effectiveness across diverse semantic transitions.
- To ensure consistency, all images are converted into non-overlapping 256 × 256 patches and then partitioned into training, validation, and test sets with a 6:2:2 ratio, totaling over 38,000 bi-temporal samples across all benchmarks.

[Table 3] Overview of RS-CD Datasets

Dataset	Year	Original Image Size	Converted Image Size	# of Images (train/validation/test)	Domain
LEVIR-CD+	2022	1024 × 1024	256 × 256	15,760 (9456/3152/3152)	Building Growth & Decay
WHU-CD	2019	21243 × 15354, 11265 × 15354		7,680 (4608/1536/1536)	Building Footprint Changes
CLCD	2022	512 × 512		2,400 (1440/480/480)	Cropland & Land-cover Transitions
S2Looking	2021	1024 × 1024		10,000 (6000/2000/2000)	Off-nadir Building Changes
CaBuAr	2022	5490 × 5490		1,424 (854/285/285)	Fire-induced Damage Assessment
SEN1Floods11	2020	Various sizes		1,612 (968/322/322)	Flood-induced Water Mapping

Experimental Settings

Evaluation Measures

TP: True Positive TN: True Negative

FP: False Positive FN: False Negative

- $Precision = \frac{TP}{TP+FP}$
- $Recall = \frac{TP}{TP+FN}$
- $F1\ Score = \frac{2(Precision * Recall)}{Precision+Recall}$
- Pe(Expected Agreement)= $\frac{((TP+FN)\times(TP+FP)+(TN+FP)\times(TN+FN))}{(TP+FP+TN+FN)^2}$
- $Kappa(Cohen's\ Kappa) = \frac{OA-Pe}{1-Pe}$
- OA(Overall Accuracy)= $\frac{TP+TN}{TP+TN+FP+FN}$
- IoU(Intersection over Union)= $\frac{TP}{TP+FP+FN}$

Cross Validation

- Number of Iteration: 100K
- Three Predefined Independent Splits
- Data Split Ratio (Train:Val:Test) = 6:2:2

Hyperparameter Settings

- Batch size : 16
- Learning Rate Scheduler : Poly
- Initial Learning Rate : 1e-4
- Optimizer : Adam
- Weight Decay : 5e-4

Experimental Results

Comparison on Inference Time

- The framework achieves an outstanding 4.13ms inference time and 242.38 FPS on LEVIR-CD+, operating up to 73 times faster than computationally heavy architectures like CDMamba.
- The model realizes an 88.2% reduction in FLOPs while maintaining competitive F1 performance on LEVIR-CD+, indicating that strong efficiency can be achieved without a substantial loss of localization quality.

[Table 4] RS-CD performance comparison on the LEVIR-CD+ dataset

Model	Inference Time	F1	Frame Per Second	Flops
Proposed	**4.13 ± 0.20 ms**	**0.8912 ± 0.0021**	**242.4 ± 10.7**	**6.05 G**
Change3D	19.50 ± 2.44 ms	0.9074 ± 0.0025	52.1 ± 0.7	15.59 G
Changer	19.66 ± 1.59 ms	0.8371 ± 0.0063	51.5 ± 1.3	7.88 G
ChangeCLIP	47.60 ± 2.85 ms	0.9109 ± 0.0021	21.4 ± 0.2	14.15 G
ChangeMamba	57.48 ± 5.67 ms	0.9003 ± 0.0026	17.6 ± 0.2	28.70 G
CDMamba	305.08 ± 1.32 ms	0.9185 ± 0.0070	3.3 ± 0.1	51.17 G

Experimental Results

Comparison on Accuracy

[Table 5] RS-CD performance comparison on the all dataset (Averaged over 3 runs)

Evaluation Measure	Dataset	**Proposed**	Change3D	Changer	ChangeCLIP	ChangeMamba	CDMamba
F1	LEVIR-CD+	**0.8912 ± 0.0021**	0.9074 ± 0.0025	0.8371 ± 0.0063	0.9109 ± 0.0021	0.9003 ± 0.0026	0.9185 ± 0.0070
	WHU-CD	**0.9153 ± 0.0025**	0.9404 ± 0.0034	0.9307 ± 0.0159	0.9307 ± 0.0112	0.9354 ± 0.0024	0.9579 ± 0.0060
	CLCD	**0.7799 ± 0.0094**	0.8022 ± 0.0104	0.7570 ± 0.0146	0.7652 ± 0.0116	0.8269 ± 0.0070	0.8709 ± 0.0354
	S2Looking	**0.5145 ± 0.0298**	0.5920 ± 0.0069	0.6504 ± 0.0299	0.5647 ± 0.0263	0.6076 ± 0.0153	0.5360 ± 0.0062
	CaBuAr	**0.8178 ± 0.0192**	0.8598 ± 0.0389	0.6866 ± 0.0953	0.7231 ± 0.0600	0.8294 ± 0.0260	0.8809 ± 0.0617
	SEN1Floods11	**0.7307 ± 0.0045**	0.7480 ± 0.0117	0.8048 ± 0.0316	0.7286 ± 0.0266	0.7503 ± 0.0051	0.8112 ± 0.0045
Precision	LEVIR-CD+	**0.9048 ± 0.0048**	0.9097 ± 0.0067	0.8234 ± 0.0155	0.9243 ± 0.0061	0.9113 ± 0.0033	0.9290 ± 0.0068
	WHU-CD	**0.9358 ± 0.0095**	0.9510 ± 0.0095	0.9414 ± 0.0224	0.9504 ± 0.0240	0.9567 ± 0.0160	0.9601 ± 0.0095
	CLCD	**0.8084 ± 0.0372**	0.8339 ± 0.0179	0.8079 ± 0.0219	0.8448 ± 0.0188	0.8600 ± 0.0119	0.8906 ± 0.0490
	S2Looking	**0.5790 ± 0.0148**	0.6581 ± 0.0110	0.7431 ± 0.0593	0.6794 ± 0.0457	0.6982 ± 0.0613	0.6003 ± 0.0466
	CaBuAr	**0.8143 ± 0.0311**	0.8278 ± 0.0636	0.8070 ± 0.1483	0.7040 ± 0.1174	0.7834 ± 0.0527	0.9197 ± 0.0121
	SEN1Floods11	**0.7191 ± 0.0253**	0.7357 ± 0.0047	0.8801 ± 0.0168	0.7566 ± 0.0376	0.7676 ± 0.0155	0.8250 ± 0.0076
Recall	LEVIR-CD+	**0.8779 ± 0.0008**	0.9051 ± 0.0020	0.8515 ± 0.0083	0.8980 ± 0.0046	0.8895 ± 0.0024	0.9081 ± 0.0089
	WHU-CD	**0.8958 ± 0.0076**	0.9303 ± 0.0119	0.9204 ± 0.0128	0.9122 ± 0.0178	0.9153 ± 0.0107	0.9558 ± 0.0026
	CLCD	**0.7549 ± 0.0168**	0.7735 ± 0.0251	0.7128 ± 0.0276	0.6994 ± 0.0134	0.7966 ± 0.0211	0.8524 ± 0.0231
	S2Looking	**0.4637 ± 0.0394**	0.5379 ± 0.0085	0.4824 ± 0.0328	0.4886 ± 0.0650	0.5398 ± 0.0143	0.4883 ± 0.0449
	CaBuAr	**0.8219 ± 0.0125**	0.8959 ± 0.0198	0.6312 ± 0.1714	0.7522 ± 0.0294	0.8841 ± 0.0361	0.8517 ± 0.1185
	SEN1Floods11	**0.7442 ± 0.0229**	0.7612 ± 0.0268	0.7438 ± 0.0639	0.7059 ± 0.0597	0.7342 ± 0.0146	0.7980 ± 0.0052

Experimental Results

Comparison on Accuracy

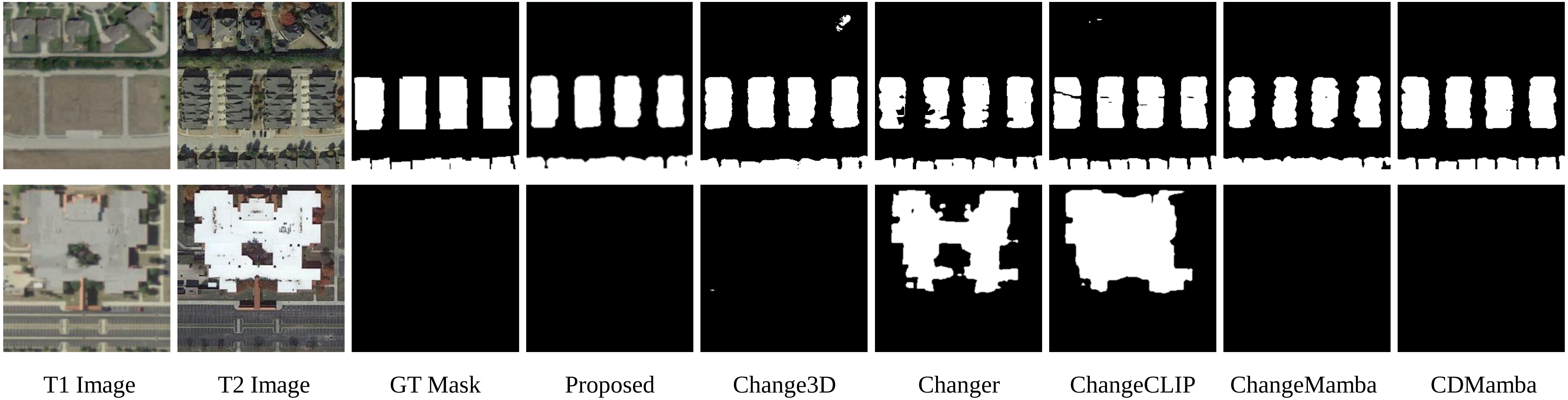
[Table 5] RS-CD performance comparison on the all dataset (Averaged over 3 runs)

Evaluation Measure	Dataset	**Proposed**	Change3D	Changer	ChangeCLIP	ChangeMamba	CDMamba
IoU	LEVIR-CD+	**0.8037 ± 0.0035**	0.8305 ± 0.0042	0.7199 ± 0.0093	0.8364 ± 0.0036	0.8186 ± 0.0043	0.8492 ± 0.0119
	WHU-CD	**0.8438 ± 0.0043**	0.8876 ± 0.0060	0.8707 ± 0.0278	0.8705 ± 0.0196	0.8787 ± 0.0042	0.9193 ± 0.0111
	CLCD	**0.6393 ± 0.0127**	0.6698 ± 0.0146	0.6092 ± 0.0188	0.6197 ± 0.0153	0.7049 ± 0.0101	0.7725 ± 0.0546
	S2Looking	**0.3469 ± 0.0274**	0.4204 ± 0.0069	0.4824 ± 0.0328	0.3938 ± 0.0256	0.4365 ± 0.0157	0.3661 ± 0.0059
	CaBuAr	**0.6922 ± 0.0272**	0.7554 ± 0.0592	0.5282 ± 0.1123	0.5686 ± 0.0719	0.7091 ± 0.0378	0.7907 ± 0.0961
	SEN1Floods11	**0.5756 ± 0.0057**	0.5976 ± 0.0150	0.6741 ± 0.0438	0.5735 ± 0.0328	0.6005 ± 0.0065	0.6824 ± 0.0064
Kappa	LEVIR-CD+	**0.8862 ± 0.0021**	0.9031 ± 0.0025	0.8300 ± 0.0061	0.9069 ± 0.0020	0.8957 ± 0.0024	0.9150 ± 0.0073
	WHU-CD	**0.9116 ± 0.0025**	0.9378 ± 0.0036	0.9282 ± 0.0167	0.9280 ± 0.0114	0.9326 ± 0.0025	0.9563 ± 0.0062
	CLCD	**0.7629 ± 0.0096**	0.7870 ± 0.0099	0.7401 ± 0.0172	0.7483 ± 0.0111	0.8136 ± 0.0064	0.8607 ± 0.0384
	S2Looking	**0.5088 ± 0.0301**	0.5870 ± 0.0071	0.6467 ± 0.0300	0.5599 ± 0.0262	0.6030 ± 0.0159	0.5309 ± 0.0059
	CaBuAr	**0.8034 ± 0.0190**	0.8484 ± 0.0406	0.6619 ± 0.0889	0.7007 ± 0.0630	0.8152 ± 0.0263	0.8695 ± 0.0665
	SEN1Floods11	**0.7071 ± 0.0060**	0.7260 ± 0.0128	0.7847 ± 0.0350	0.7062 ± 0.0276	0.7294 ± 0.0060	0.7890 ± 0.0051
Overall Accuracy	LEVIR-CD+	**0.9905 ± 0.0004**	0.9887 ± 0.0004	0.9865 ± 0.0008	0.9922 ± 0.0002	0.9912 ± 0.0003	0.9934 ± 0.0006
	WHU-CD	**0.9929 ± 0.0000**	0.9983 ± 0.0001	0.9951 ± 0.0015	0.9948 ± 0.0005	0.9946 ± 0.0004	0.9970 ± 0.0005
	CLCD	**0.9685 ± 0.0006**	0.9692 ± 0.0004	0.9683 ± 0.0056	0.9682 ± 0.0007	0.9753 ± 0.0013	0.9812 ± 0.0055
	S2Looking	**0.9885 ± 0.0006**	0.9898 ± 0.0003	0.9926 ± 0.0009	0.9901 ± 0.0003	0.9908 ± 0.0010	0.9898 ± 0.0007
	CaBuAr	**0.9734 ± 0.0006**	0.9835 ± 0.0005	0.9530 ± 0.0073	0.9736 ± 0.0009	0.9736 ± 0.0009	0.9791 ± 0.0091
	SEN1Floods11	**0.9566 ± 0.0033**	0.9595 ± 0.0025	0.9634 ± 0.0093	0.9585 ± 0.0041	0.9614 ± 0.0025	0.9603 ± 0.0010

Experimental Results

Qualitative Analysis

- The proposed framework delivers competitive semantic change detection performance, effectively capturing intricate boundaries while maintaining substantially higher efficiency than heavier architectures.
- Spectral filtering suppresses image-wide stylistic pseudo-changes that more easily affect conventional 2D models such as Changer and ChangeCLIP, leading to more stable change localization with substantially higher efficiency than heavier architectures.

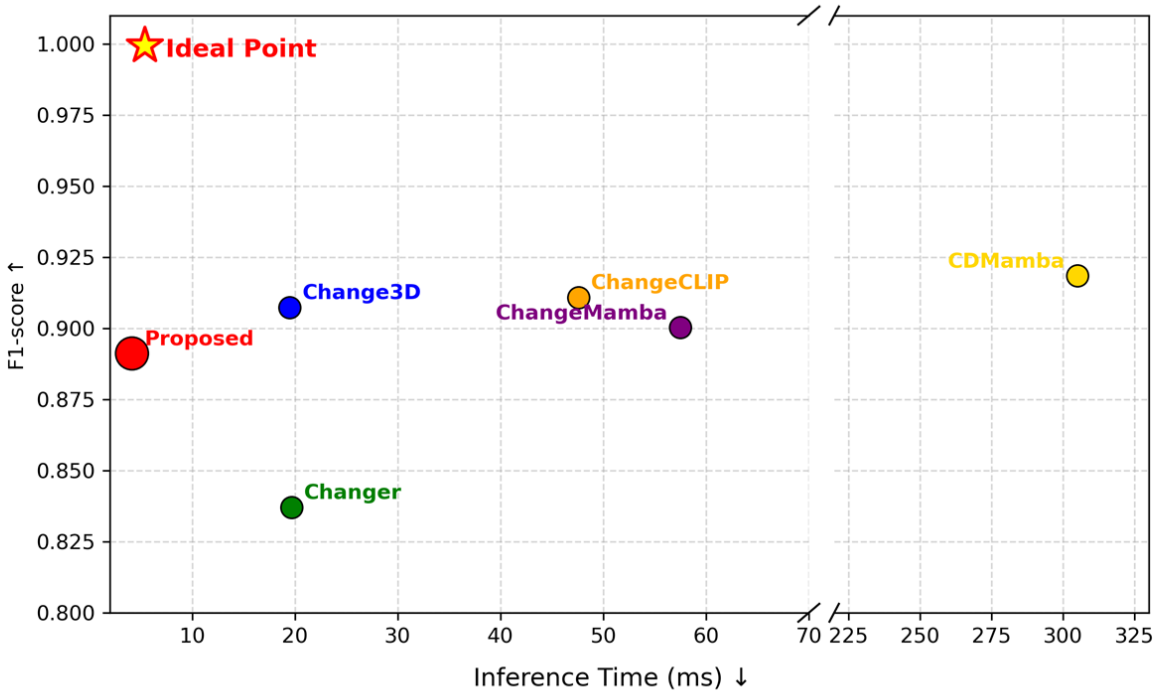


[Figure 21] Results Comparison on Data

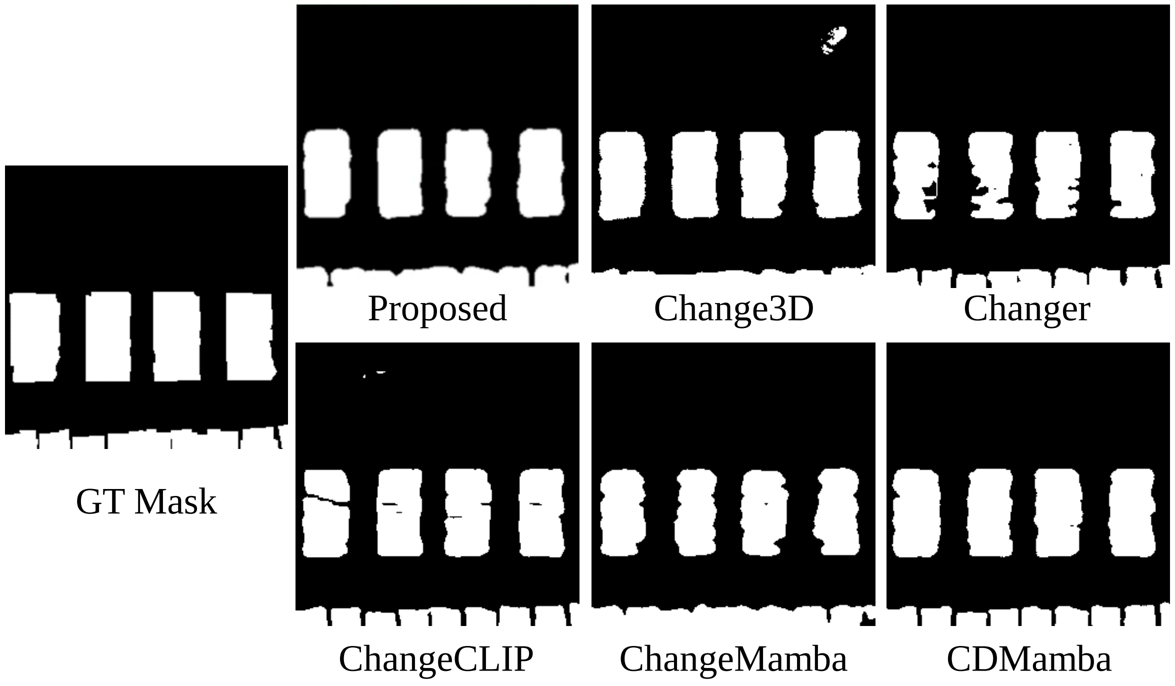
Experimental Results

Performance Comparison

- This trade-off plot on LEVIR-CD+ shows that the proposed framework remains the lightest model among the compared methods while still maintaining a favorable accuracy-cost trade-off under the same unified inference setting.
- On LEVIR-CD+, the model maintains competitive performance and even outperforms some compared structures, showing that meaningful localization quality can still be preserved while operating at a much lower inference cost.



[Figure 21] Accuracy–Efficiency Trade-off on LEVIR-CD+



[Figure 23] Results Comparison on LEVIR-CD+

Conclusion and Contribution

- This study introduces an Adaptive Frequency Masking (AFM) module that enhances efficient backbones by filtering global stylistic noise in the frequency domain at the earliest stage of the network.
- By employing learnable spectral masks, the proposed framework effectively suppresses widespread environmental pseudo-changes, allowing the backbone to focus on genuine semantic features without the need for heavy spatial operations.
- The framework achieves a 4.13 ms inference time, runs 4.7 to 73 times faster than existing methods, and reduces FLOPs by 88.2%, offering a highly efficient solution for resource-constrained RS-CD tasks.

Achievement

- Projects
  - The Development of Automatic Search Technology for Hyperparameter Optimization for Vision AI Model, ***Hyundai***, 2025
  - Participation in a Project at the International Trade and Logistics, ***Chung-Ang University***, 2024-2026
- In Domestic Conferences
  - A Survey on Multi-Label Chest X-Ray Classification, ***IEIE***, 2025

Thank you

Reference

[1] G. Cheng et al., "Change Detection Methods for Remote Sensing in the Last Decade: A Comprehensive Review," Remote Sensing, vol. 16, p. 2355, 2024.

[2] M. Hussain, D. Chen, A. Cheng, H. Wei and D. Stanley, "Change Detection from Remotely Sensed Images: From Pixel-Based to Object-Based Approaches," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 80, pp. 91-106, 2013.

[3] D. Hong et al., "Deep Learning for Change Detection in Remote Sensing: A Comprehensive Survey," IEEE Geoscience and Remote Sensing Magazine, vol. 13, no. 3, pp. 164-189, 2025.

[4] Z. Zhu, "Change Detection Using Landsat Time Series: A Review of Frequencies, Preprocessing, Algorithms, and Applications," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 130, pp. 370-384, 2017.

[5] Yang et al., "Edge-guided Hierarchical Network for Building Change Detection in Remote Sensing Images," Applied Sciences, vol. 14, no. 13, p. 5415, 2024.

[6] I. Caballero et al., "Use of the Sentinel-2 and Landsat-8 Satellites for Water Quality Monitoring: An Early Warning Tool in the Mar Menor Coastal Lagoon," Remote Sensing, vol. 14, p. 2744, 2022.

[7] H. Chen et al., "ChangeMamba: Remote Sensing Change Detection with Spatio-Temporal State Space Model," IEEE Transactions on Geoscience and Remote Sensing, 2024.

[8] H. Zhang et al., "CDMamba: Incorporating Local Clues Into Mamba for Remote Sensing Image Binary Change Detection," IEEE Transactions on Geoscience and Remote Sensing, vol. 63, 2025.

[9] D. Zhu et al., "Change3D: Revisiting Change Detection and Captioning from A Video Modeling Perspective," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 24011-24022, 2025.

[10] S. Fang, K. Li and Z. Li, "Changer: Feature Interaction Is What You Need for Change Detection," IEEE Transactions on Geoscience and Remote Sensing, vol. 61, 2023.

[11] S. Dong, L. Wang, B. Du and X. Meng, "ChangeCLIP: Remote Sensing Change Detection with Multimodal Vision-Language Representation Learning," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 208, pp. 53-69, 2024.