D^{2}MDM^2: A Brain-Inspired Deep Network Based on DDM Decision-Making Mechanism for Remote Sensing Change Detection

Kang Zhao , **Ye Zhang** , **Bin Wang** * , **Jianchao Zeng**
School of Computer Science and Technology, North University of China, Taiyuan 030051, China
kang_zhao_nuc@163.com, ye_zhang_nuc@163.com, binwang_nuc@163.com, zjc@nuc.edu.cn

Abstract

Remote sensing change detection (RSCD) aims to identify changed regions in bitemporal images. However, conventional one-step modeling suffers from performance degradation caused by imaging temporal differences (e.g., illumination disturbances, seasonal variations). To address this issue, we formulate the problem as an "adversarial attack and defense" paradigm. Then, instead of traditional adversarial training, we propose a brain-inspired deep network based on the **drift diffusion model (DDM)** decision-making mechanism for RSCD, which is rooted in cognitive neuroscience. By simulating brain decision processes, the network enhances model robustness significantly. Specifically, we design the multi-stage reasoning unit to simulate DDM's **iterative evidence generation**, stochastic representation unit with reparameterization sampling to model **stochastic diffusion** in decision-making, decision aggregation unit to replicate DDM's **evidence accumulation**, and loss-guided unit with multi-loss joint constraints to simulate **directional drift**. Experimental results verify the model's superior performance on multiple datasets and their perturbed versions.

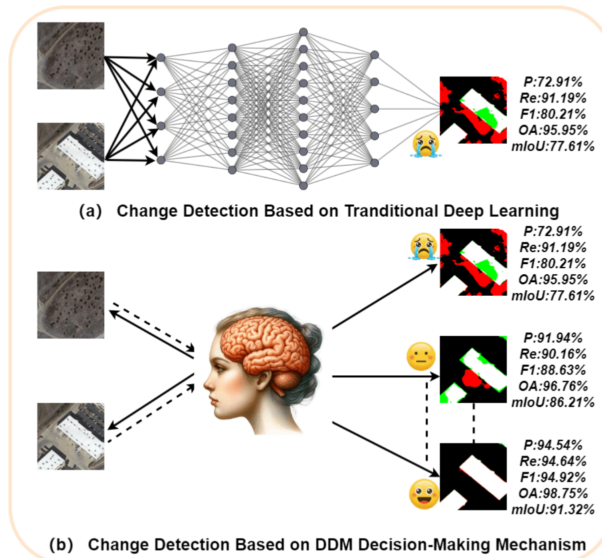

Figure 1: Different frameworkss for RSCD. (a) Traditional deep learning network: perform one-step modeling on bitemporal images to complete change detection. (b) DDM decision-making deep learning network: complete the change detection on bitemporal image based on the brain-inspired mechanism in a multi-stage manner.

1 Introduction

RSCD has become a research hotspot with advancing imaging technologies. By comparing bitemporal images of same region [Chen *et al.*, 2023], RSCD identifies surface transformations to support urban planning and environmental protection [Gu *et al.*, 2025; Wang *et al.*, 2025b; Zhao *et al.*, 2025]. However, precisely controlling imaging conditions (e.g., illumination, climate) remains challenging, as cross-acquisition variations often introduce substantial detection noise.

Existing RSCD methodologies are primarily categorized into traditional [Xu *et al.*, 2025] and deep learning [Wang *et al.*, 2025a] approaches. Traditional methods rely on manually designed features and expert-led hyperparameter tuning. They suffer from restricted representation capabilities and weak generalization. In contrast, deep learning has become the dominant paradigm [Wang *et al.*, 2023]. CNN-based frameworks [Zheng *et al.*, 2021; Zhang *et al.*, 2022] (e.g., FCN, U-Net) are widely used but are often limited by narrow receptive fields; Transformers have been introduced to capture global contextual information [Lin *et al.*, 2025; Bernhard *et al.*, 2023; Hills *et al.*, 2024]; Mamba-based models achieve superior spatiotemporal modeling efficiency through unique selective scanning mechanisms [Feng *et al.*, 2025; Zhang *et al.*, 2026; He *et al.*, 2025]; and DDPMs (Denoising Diffusion Probabilistic Models) offer a novel technical path by modeling complex data distributions through iterative diffusion and denoising processes [Dai *et al.*, 2025; Vinholi *et al.*, 2024; Li *et al.*, 2024]. However, deep learning models often struggle with environmental disturbances like illumination fluctuations, which shift data distributions and degrade performance. These unpredictable variations can be

*Corresponding Author

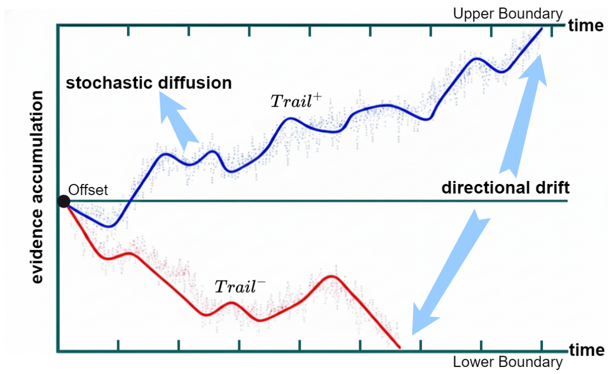

Figure 2: Schematica diagram of the drift-diffusion model (DDM)

framed as adversarial attacks on the CD. Consequently, bolstering robustness through an adversarial defense perspective offers a promising solution [Hung-Quang *et al.*, 2024].

Existing adversarial training [Cho *et al.*, 2023] often lack adaptability and thus fail to meet these practical. To address this critical limitation, we draw inspiration from the human brain, which is the only intelligent system endowed with innate robustness. As illustrated in Figure 1, traditional methods typically adopt one-step modeling to directly generate outputs. In contrast, the brain-inspired decision mechanism facilitates a coarse-to-fine detection process, a pattern consistently corroborated by cognitive neuroscience [Pla, 2010].

Inspired by this, this paper proposes a brain-inspired deep network based on DDM decision-making mechanism for RSCD. Firstly, U-Net performs multi-scale decomposition of input difference features, with the results fed into multi-stage reasoning unit; iterative updating of decision prior features enhances change region localization in a multi-stage progressive manner. Secondly, posterior probability is learned from bitemporal images and labels to train a prior distribution generator; random sampling from the prior distribution and fusion of decision features lay the foundation for modeling random diffusion characteristics. Thirdly, multi-stage generated decision maps are sequentially input into decision aggregation unit to deeply mine temporal correlations and achieve brain-inspired evidence accumulation. Finally, multi-loss directional guidance from the loss-guided unit drives the model to accurately focus on real change regions. The main contributions of this paper are as follows:

- Aiming at the impact of dynamic and uncertain illumination interference on RSCD, we treat it as adversarial attacks; adopting the brain-inspired mechanism of DDM significantly enhances model robustness.
- A multi-stage reasoning unit is designed to simulate the iterative evidence generation of DDM, and a stochastic representation unit to introduce randomness via reparameterization sampling, which jointly model brain-inspired stochastic diffusion.
- Leveraging DDM-based brain-inspired evidence accumulation, decision aggregation unit mines temporal correlations of multi-stage decision maps, with loss-guided unit incorporated to model DDM directional drift.

2 Relation Work

2.1 Drift Diffusion Model (DDM)

In neural learning, the DDM characterizes the brain's response patterns under noisy inputs. Perceptual decision-making research [Yang and Shadlen, 2007] has leveraged random dot motion experiments to reveal these neural mechanisms, while subsequent studies [Ratcliff and McKoon, 2008] have used the DDM to quantify correlations between accuracy, reaction time, evidence accumulation rate, and decision thresholds. Beyond the prefrontal-parietal network, similar accumulation patterns identified in brain regions like the basal ganglia [Wang *et al.*, 2024] confirm that evidence accumulation is a universal cerebral decision mechanism. As illustrated in Figure 2, the accumulation trajectory is governed by two core components: directional drift, representing a systematic bias throughout the process, and random diffusion, denoting instantaneous, directionless noise perturbations. A decision is triggered when the accumulated evidence crosses a predefined threshold, with the duration from stimulus onset to threshold crossing defining the response time.

2.2 Distributed Sampling Learning

Distributed sampling learning refers to extracting features from a given or learned data distribution for model training. Through rational sampling, sample features that are representative of the overall data distribution can be obtained, facilitating model to learn the intrinsic patterns of the data. Among these techniques, reparameterization sampling represents the sampling process as a combination of deterministic function and noise variable, enabling sampling from complex probability distributions and efficient gradient computation simultaneously [Wu *et al.*, 2024]. During deep learning training, sampling from the learned data distribution for model optimization can significantly reduce reliance on feature extraction operators such as convolution and self-attention, as well as cut down on computational resource requirements. Meanwhile, the randomness introduced by sampling can enrich feature diversity, thereby effectively enhancing the generalization and stability of the model [Yu *et al.*, 2025].

3 Method

Figure 3 illustrates the overall architecture of our proposed network, i.e., D^2MDM^2. Notably, the DDM encompasses three core characteristics, namely stochastic diffusion, evidence accumulation and directional drift. To simulate brain-like decision-making process, several interconnected modules are constructed in the subsequent sections.

3.1 Multi-stage Reasoning Unit (MRU)

In DDM, evidence is generated iteratively and updated dynamically over multiple time steps. Inspired by the human visual system and DDM, this paper proposes MRU. This module aims to continuously revise decision prior features via a series of small-scale guidance factors across multiple stages, simulating the evidence iterative generation process of DDM.

We employ a U-Net backbone to extract multi-scale bitemporal features. Specifically, the second-largest scale output

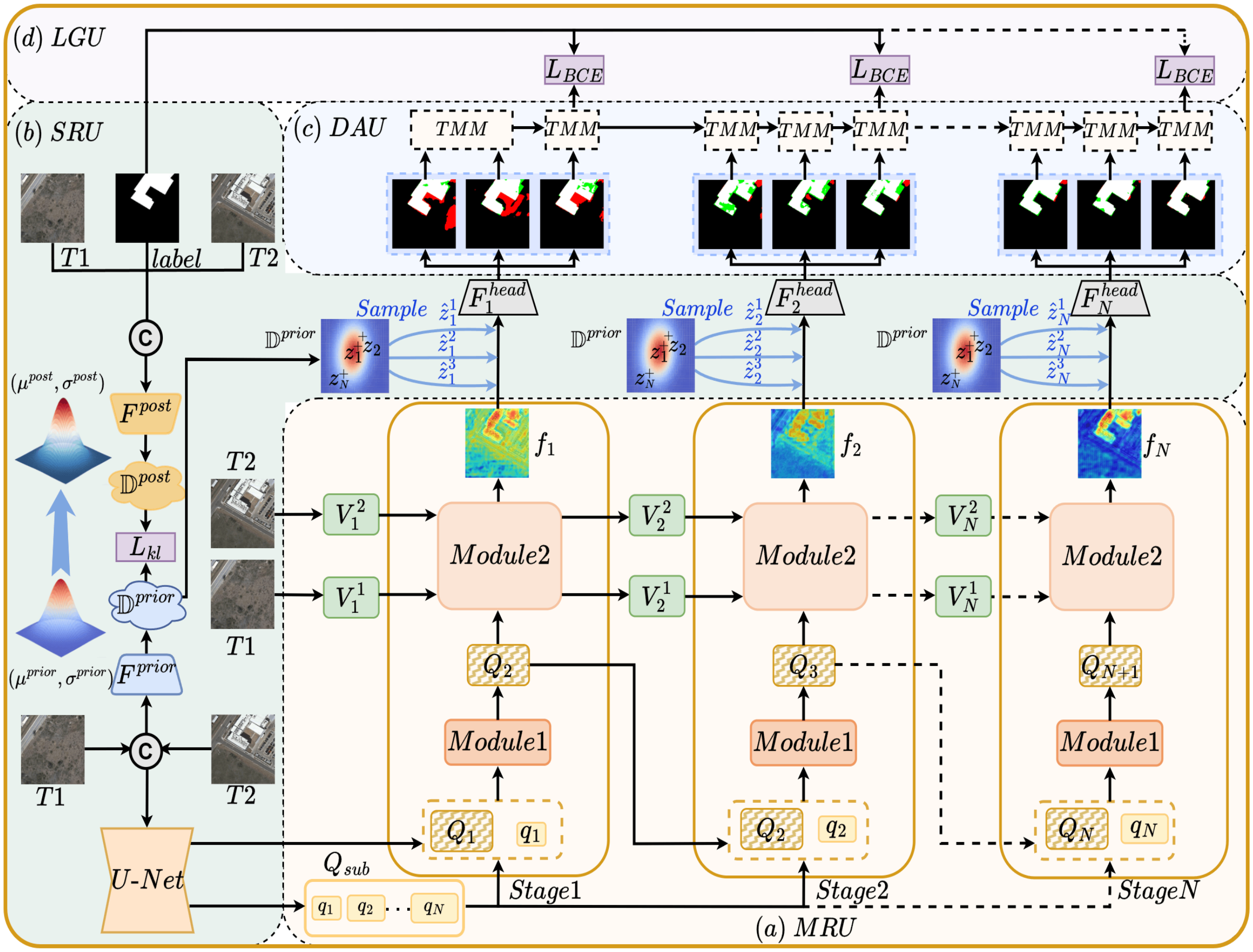

Figure 3: A Brain-Inspired Deep Network based on DDM Decision-Making Mechanism for RSCD. (a)Multi-stage reasoning unit (MRU): mimicking the iterative evidence generation of DDM; (b)Stochastic representation unit (SRU): simulating stochastic diffusion; (c)Decision aggregation unit (DAU): replicating evidence accumulation; (d)Loss-guided unit (LGU): emulating directional drift of DDM.

is designated as the initial decision prior $Q_1 \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times C}$, while the remaining smaller-scale features constitute a guidance factor set $Q_{sub} = \{q_1, \ldots, q_N\}$. These guidance factors provide scale-specific discriminative cues to iteratively update and refine the decision prior. MRU operates through N stages. Each stage comprises guidance update and decision feature generation module, as illustrated in Figure 3.

Guidance Update Module. This module refines decision prior Q_i by leveraging guidance factor q_i, as illustrated in Figure 4. Through an attention mechanism, the module integrates multi-scale discriminative features to produce updated decision prior Q_{i+1}. The operation is formulated as:

$$A_i^{q \to Q} = SoftMax\left(\left(Q_i W_1^Q\right) \otimes \left(q_i W_2^q\right) / \sqrt{C}\right) \tag{1}$$

$$Q_{i+1} = MLP\left(A_i^{q \to Q} \otimes \left(Q_i W_3^Q\right)\right) + Q_i \tag{2}$$

Where $A_i^{q \to Q}$ denotes the attention weight, W_1^Q, W_2^q, and W_3^Q are learnable projection layers, and $\bigotimes$ denotes the matrix multiplication.

Decision Feature Generation Module. This module incorporates the updated decision prior Q_{i+1} into the bitemporal features, as illustrated in Figure 4. Leveraging a guided attention mechanism, it augments feature representations to explicitly distill and characterize change-discriminative information, ultimately yielding stage-specific decision feature f_i.

First, we propose a reverse softmax (re-softmax) that can suppress the weights of similar regions and amplify the responses of differential regions. The formula is shown as:

$$re\text{-}SoftMax\,(x) = SoftMax\,(-x) \tag{3}$$

Second, this paper adopts a bidirectional interaction mechanism to balance feature transfer and current decision-making demands. The decision prior feature Q_{i+1} is interacted with the stage-specific bitemporal features $V_i^k (k \in \{1, 2\})$, generating the state update features V_{i+1}^k for transfer and the projection features $V_i^{\hat{k}}$ for decision reasoning, respectively. The detailed process is as follows:

$$A_{i+1}^k = SoftMax\left(\left(Q_{i+1} W_1^{\hat{Q}}\right) \otimes \left(V_i^k W_2^{\hat{V}}\right) / \sqrt{C}\right) \tag{4}$$

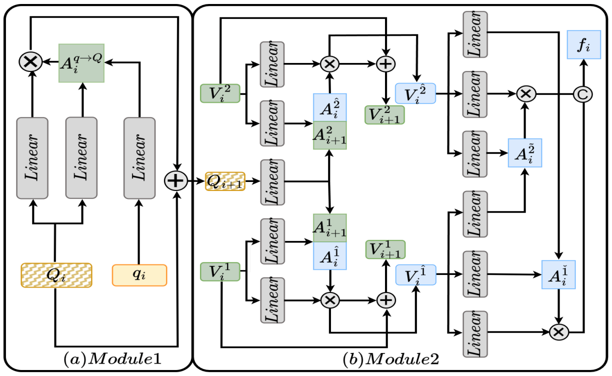

Figure 4: (a) Module 1: Guidance update module. (b) Module 2: Decision feature generation module.

$$V_{i+1}^{k} = MLP\left(A_{i+1}^{k} \otimes \left(V_i^k W_3^{\hat{V}}\right)\right) + V_i^k \tag{5}$$

$$A_i^{\hat{k}} = re\text{-}SoftMax\left(\left(Q_{i+1} W_1^{\hat{Q}}\right) \otimes \left(V_i^k W_2^{\hat{V}}\right)/\sqrt{C}\right) \tag{6}$$

$$V_i^{\hat{k}} = MLP\left(A_i^{\hat{k}} \otimes \left(V_i^k W_3^{\hat{V}}\right)\right) \tag{7}$$

Notably, to enhance parameter efficiency, the forward and reverse attention components share a unified set of weight matrices $W_1^{\hat{Q}}$, $W_2^{\hat{V}}$ and $W_3^{\hat{V}}$.

Finally, a cross reverse attention mechanism is introduced to capture the differential information between the stage-specific bitemporal features $V_i^{\hat{1}}$ and $V_i^{\hat{2}}$ and generate the final decision feature f_i. The formula is as follows:

$$\begin{aligned} f_i = {} & re\text{-}SoftMax\left(\left(V_i^{\hat{1}} W_1^{\tilde{V}}\right) \otimes \left(V_i^{\hat{2}} W_2^{\tilde{V}}\right)/\sqrt{C}\right) \otimes \left(V_i^{\hat{2}} W_3^{\tilde{V}}\right) \\ & + re\text{-}SoftMax\left(\left(V_i^{\hat{2}} W_1^{\tilde{V}}\right) \otimes \left(V_i^{\hat{1}} W_2^{\tilde{V}}\right)/\sqrt{C}\right) \otimes \left(V_i^{\hat{1}} W_3^{\tilde{V}}\right) \end{aligned} \tag{8}$$

3.2 Stochastic Representation Unit (SRU)

In DDM, stochastic diffusion denotes the continuous noise injected during evidence accumulation, resulting in stochastic walk trajectories that drive decision variability. To simulate this mechanism, we propose SRU. Its integration with MRU enables the model to capture the inherent uncertainty and fluctuations during the decision-making process.

First, the concatenation map X derived from the dual-temporal images $T_1, T_2 \in \mathbb{R}^{H\times W\times 3}$ is fed into a prior network parameterized by ω to generate a prior distribution $\mathbb{D}^{prior}(X)$. Moreover, this prior distribution is modeled as an axis-aligned multivariate Gaussian distribution, denoted as $\mathcal{N}\left(\mu^{prior}, \sigma^{prior}\right)$, where the mean and variance are obtained as follows:

$$\mu^{prior}, \log\sigma^{prior} = F^{prior}(X, \omega) \tag{9}$$

Where F^{prior} denotes a learnable generator for the prior distribution, as illustrated in Figure 3.

Additionally, to generate posterior distribution, we feed original images T_1, T_2 and label Y into posterior network parameterized by weights υ, thereby learning posterior distribution $\mathbb{D}^{post}(T_1, T_2, Y)$, denoted as $\mathcal{N}\left(\mu^{post}, \sigma^{post}\right)$. The mean and variance are obtained as follows:

$$\mu^{post}, \log\sigma^{post} = F^{post}(T_1, T_2, Y, \upsilon) \tag{10}$$

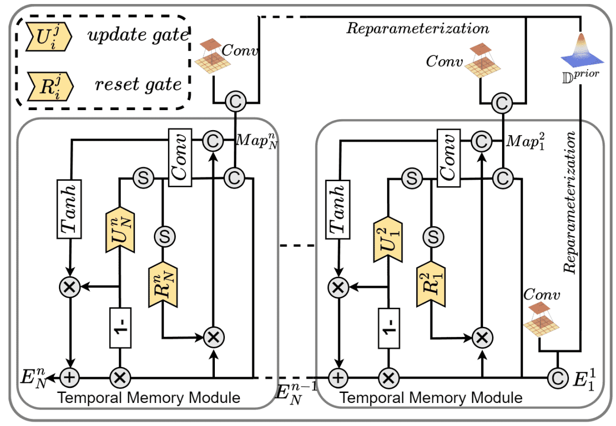

Figure 5: Decision aggregation unit.

Where F^{post} denotes learnable generator for posterior distribution, as illustrated in Figure 3.

To train prior network via posterior network, we adopt Kullback-Leibler (KL) divergence loss $KL(Q \parallel P) = \mathbb{E}_{z\sim Q}\left[\log\frac{Q(z)}{P(z)}\right]$ to penalize the discrepancy between posterior and prior distribution:

$$\mathcal{L}_{kl} = KL\left(\mathbb{D}_{prior}(X), \mathbb{D}^{post}(T_1, T_2, Y)\right) \tag{11}$$

Then, reparameterization sampling is leveraged to perform multiple samplings from the prior distribution, yielding sampled features $(\hat{z}_1^1, \hat{z}_i^2 ..., \hat{z}_N^n)$ (where n denotes the number of samplings and N represents the number of stages). This not only ensures the differentiability of the sampling process but also introduces randomness into the model.

Finally, to simulate the stochastic diffusion characteristics, the sampled features $(\hat{z}_1^1, \hat{z}_i^2 ..., \hat{z}_N^n)$ and the decision feature f_i are fed into the change detection head F_i^{head} separately to obtain the decision results. The process is as follows:

$$Map_i^j = F_i^{head}\left(\hat{z}_i^j \copyright f_i\right) \tag{12}$$

Where © denotes the feature concatenation operation, and the decision map sequence $\{Map_i^j\}$ ($i \in [1, N]$, $j \in [1, n]$) lays the foundation for subsequent evidence accumulation.

3.3 Decision Aggregation Unit (DAU)

In DDM, evidence accumulation is a dynamic process that continuously integrates task-related information until a decision threshold is reached and a choice is output. In this paper, the sequences of multiple decision maps generated at each stage exhibit significant temporal dependence, prompting us to introduce DAU to explore the evolution rules among multi-stage decision features. This design is highly consistent with evidence accumulation in terms of physical mechanism.

As illustrated in Figure 5, the previous hidden state E_i^{j-1} and the current time-step state Map_i^j are fed into the temporal memory module (TMM), yielding the activation values of the reset gate and update gate. The reset gate filters out redundant information irrelevant to the CD task, while the update gate

Method	LEVIR-CD					SECOND					SYSUCD				
	P	Re	F1	OA	mIoU	P	Re	F1	OA	mIoU	P	Re	F1	OA	mIoU
DTCDSCN	72.86	<u>89.23</u>	80.22	98.56	82.75	62.03	**70.92**	66.17	87.23	67.43	<u>82.91</u>	70.47	76.19	89.57	74.51
BIT	92.16	70.68	80.01	98.20	82.40	69.42	64.37	<u>66.80</u>	**88.72**	<u>68.71</u>	80.52	72.55	76.16	89.33	74.32
Change Former	91.17	78.57	84.40	98.52	85.74	<u>74.61</u>	50.58	66.79	<u>87.95</u>	68.21	80.83	73.68	77.09	89.67	75.11
ICIFNet	**95.70**	70.12	80.94	98.32	83.12	**78.21**	53.96	63.86	87.79	66.61	**84.60**	70.70	77.03	90.06	75.36
USSF-Net	90.44	84.02	<u>87.11</u>	<u>98.73</u>	<u>87.92</u>	73.15	54.43	62.48	86.90	65.37	80.66	75.62	78.06	89.97	75.90
DMINet	<u>94.38</u>	74.08	83.01	98.46	84.68	74.18	58.50	65.41	87.63	67.30	79.17	**80.66**	<u>79.91</u>	<u>90.43</u>	<u>77.36</u>
Change Mamba	94.16	75.81	84.01	98.53	85.44	69.29	63.19	66.10	87.04	67.26	79.55	73.91	76.63	89.37	74.62
Ours	89.08	**91.00**	**90.03**	**98.97**	**90.40**	71.66	<u>64.91</u>	**68.13**	87.85	**68.84**	80.44	<u>79.90</u>	**80.17**	**90.68**	**77.71**

*All values are reported in percentage (%). Color convention: **best**, <u>2-nd best</u>.

Table 1: Performance comparison of diffenert methods across various datasets

determines the key valid information that should be retained and passed to subsequent stages. As shown in the formulas:

$$R_i^j, U_i^j = \sigma\left(Split\left(Conv\left(Map_i^j \copyright E_i^{j-1}\right)\right)\right) \tag{13}$$

Where R_i^j denotes reset gate, U_i^j denotes update gate, and $\sigma(\cdot)$ represents Sigmoid activation function. Based on the filtering results of the reset gate, the weight of the hidden state E_i^{j-1} from the previous stage is adjusted; then, it is concatenated with the current time-step state Map_i^j, and the reset feature $\hat{E}_i^j$ is generated via convolution and activation functions. This feature focuses on the fusion of key decision information in current stage and valid historical evidence:

$$\hat{E}_i^j = Tanh\left(Conv\left(\left(R_i^j \odot E_i^{j-1}\right) \copyright Map_i^j\right)\right) \tag{14}$$

Through weight modulation of the update gate, the evidence accumulated in the previous stage and the currently generated reset feature are dynamically fused to obtain the hidden state E_i^j of the current stage, thus completing one round of evidence accumulation:

$$E_i^j = (1 - U_i^j) \odot E_i^{j-1} + U_i^j \odot \hat{E}_i^j \tag{15}$$

Here, we set $E_1^1 = Map_1^1$. Through the above temporal iterative computation, the model can gradually integrate the valid information of multi-stage decision maps and complete the brain-inspired evidence accumulation process.

3.4 Loss-Guided Unit (LGU)

The core of directional drift in DDM lies in the characteristic that decision-making process is driven by goal-oriented signals and gradually shifts toward the correct decision direction. To simulate this characteristic, this paper constructs LGU. Through phased supervision and multi-loss collaboration, it balances the stability of model training and final detection accuracy, guides model to gradually focus on real change regions, and achieves brain-like directional drift.

We adopt the Binary Cross-Entropy (BCE) loss to supervise the prediction maps. Considering the multi-stage progressive localization characteristic of our model, to fully utilize middle-layer features and avoid gradient vanishing or overfitting during training, we impose constraints on the final hidden state h_i^n of each stage, which is defined as the prediction map P^i.

The BCE loss at the $i-th$ stage is defined as follows:

$$\mathcal{L}_{BCE}^i = \frac{1}{H \times W} \sum_{h=1}^{H} \sum_{w=1}^{W} l_{bce}\left(P_{hw}^i, Y_{hw}\right) \tag{16}$$

Where Y_{hw} represents the pixel value of label at position (h, w), and P_{hw}^i denotes the probability that model predicts the pixel at position (h, w) as a changed pixel in the i-th stage.

To integrate constraint effects of multiple losses, we construct a joint loss function $\mathcal{L}_T$, which is defined as follows:

$$\mathcal{L}_T = \lambda_1 * \sum_{i=1}^{N} \frac{1}{2^{i-1}} \mathcal{L}_{BCE}^i + \lambda_2 * \mathcal{L}_{kl} \tag{17}$$

This loss function ensures the training stability of each stage through multiple constraints while clarifying the model's optimization direction, which is highly consistent with the directed drift characteristic of DDM. The output of the final stage h_N^n is ultimately taken as the final CD result.

4 Experiment Setting

We evaluate our proposed D^2MDM^2 change detector on three public change detection datasets, namely LEVIR-CD, SECOND, and SYSUCD. LEVIR-CD is a large-scale U.S. building change detection dataset with 637 ultra-high-resolution image pairs split into 256×256 sub-images for training/validation/testing. SECOND is a Chinese land cover change dataset with 4,662 aerial images cropped into sub-images and split into three sets at a 7:2:1 ratio. SYSUCD is a 3-band Hong Kong aerial dataset with 20,000 image pairs capturing diverse changes and split 6:2:2 for experiments.

The model is implemented in PyTorch and trained on a single NVIDIA GeForce RTX 4090 GPU with batch size of 16. We utilize the AdamW optimizer with an initial learning rate of 0.001 and a weight decay of 5×10^{-4}. To facilitate convergence, the learning rate follows a step decay strategy, reducing to 70% of its previous value every 20 epochs across a total of 150 training epochs. For the multi-task objective function, the loss weight hyperparameters are empirically set as $\lambda_1 = 1$ and $\lambda_2 = 1$.

For objective evaluation, this paper adopts five evaluation metrics—Precision (P), Recall (Re), F1-score (F1), Overall Accuracy (OA), and mean Intersection over Union (mIoU)—to comprehensively assess the experimental results.

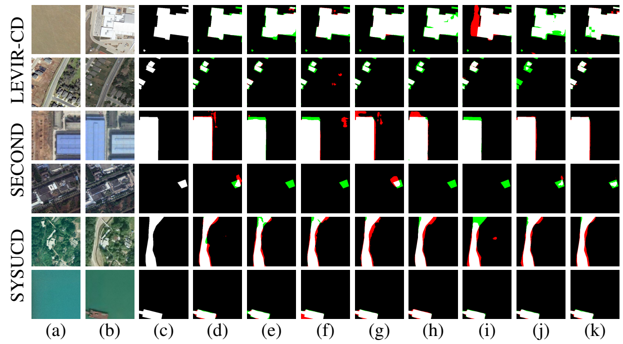

Figure 6: Qualitative results of different CD methods on LEVIR-CD, SECOND and SYSUCD datasets. (a) T1 Image, (b) T2 Image, (c) Ground Truth, (d) DTCDSCN, (e) BIT, (f) Change Former, (g) ICIFNet, (h) DMINet, (i) USSFC-Net, (j) ChangeMamba, (k) Ours.

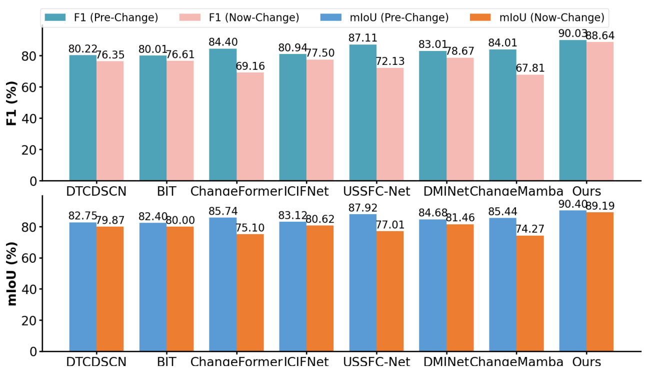

Figure 7: Indicator Changes of Local Perturbation on LEVIR-CD

5 Experiment Result

5.1 Comparison for Regular Data

To verify the effectiveness of D^{2}MDM^2 on RSCD, we compare it with several CD methods, including DTCDSCN [Liu *et al.*, 2021], BIT [Chen *et al.*, 2022], Change Former [Bandara and Patel, 2022], ICIFNet [Feng *et al.*, 2022], USSFC-Net [Lei *et al.*, 2023], DMINet [Wu *et al.*, 2023] and Change Mamba [Chen *et al.*, 2024]. For fairness, all comparative methods are implemented using the codes released by their authors, and their respective parameters are set in accordance with the original literatures.

Objective Indicator. Table 1 presents the quantitative metrics of different models on three clean datasets without attacks. D^{2}MDM^2 achieves excellent performance in comprehensive metrics such as F1-score and mIoU. For instance, its F1-score reaches 90.03% on LEVIR-CD dataset. This superior performance may be attributed to the DDM-inspired brain-like decision mechanism, which is highly consistent with the human cognitive process and gradually realizes the accurate localization of changed regions.

Visual Comparison. As illustrated in Figure 6, existing methods like DMINet and ChangeMamba suffer from missed detections of small targets, while DTCDSCN and BIT exhibit edge-related false alarms. Additionally, USSFC-Net yields significant false positives in large regions. In contrast, MRU effectively refines decision priors via guidance factors, achieving superior qualitative results across various scales.

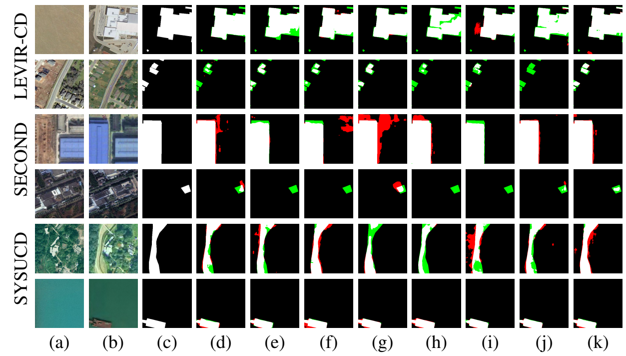

Figure 8: Qualitative results on Detection results of different CD methods on LEVIR-CD, SECOND and SYSUCD dataset with local perturbation. (a) T1 Image, (b) T2 Image, (c) Ground Truth, (d) DTCDSCN, (e) BIT, (f) Change Former, (g) ICIFNet, (h) DMINet, (i) USSFC-Net, (j) ChangeMamba, (k) Ours.

5.2 Comparison for Local Perturbation Data

To evaluate robustness, we constructed an illumination disturbance set by simulating real HSV color space variations. For LEVIR-CD, SECOND, and SYSUCD, the perturbation ranges were determined as $[-5.25, 12.57]$, $[4.65, 30.82]$, and $[-28.81, 29.95]$. By applying random pixel-wise adjustments within these bounds, we synthesized realistic samples that reflect unpredictable environmental dynamics.

Objective Indicator. Figure 7 presents the quantitative metrics of different models for LEVIR-CD under illumination perturbation. It can be seen that the F1 and mIoU of our model remain leading across all datasets. This is because SRU effectively simulates random perturbations caused by noise, thus greatly improving the model's anti-interference ability.

Visual Comparison. Figure 8 evaluates performance under illumination perturbations, which significantly degrade the accuracy of baseline methods (e.g., BIT, ICIFNet, DMINet, and ChangeMamba), leading to increased false alarms and misses. In contrast, our model maintains superior target integrity and robustness. This resilience is attributed to the loss-guided unit, which simulates the directed drift process of DDM. By focusing on change-related information and suppressing irrelevant redundant information through multi-loss constraints, it further improves the accuracy of changed region localization and the reliability of results.

5.3 Ablation Experiments

DSF	P.	Re.	F1	OA	mIoU
2	73.01	<u>88.35</u>	79.95	97.74	82.12
3	89.08	**91.00**	**90.03**	**98.97**	**90.40**
4	**94.71**	78.13	<u>85.63</u>	<u>98.66</u>	<u>86.74</u>
5	<u>93.26</u>	77.80	84.83	98.58	86.09

Table 2: Abaltion of distribution samples frequency

SRU	DAU	P.	Re.	F1	OA	mIoU
✓		92.97	79.16	85.51	98.63	86.63
	✓	**94.80**	76.92	84.93	98.61	86.18
✓	✓	89.08	**91.00**	**90.03**	**98.97**	**90.40**

Table 3: Ablation of module

N	P.	Re.	F1	OA	mIoU
2	88.35	89.71	89.27	98.87	89.52
3	89.08	**91.00**	**90.03**	**98.97**	**90.40**
4	**92.73**	80.89	86.40	98.70	87.35

Table 4: Ablation of stage number

Ablation of Distribution Sampling Frequency (DSF). To introduce controllable randomness, multiple independent samples are fused with decision features. As shown in Table 2, model performance peaks at three samples. While moderate sampling enhances adaptability to complex interferences, excessive repetitions introduce redundant information and noise, leading to overfitting and increased overhead.

Ablation of Module. We conducted ablation studies to verify the effectiveness of the proposed modules (i.e., SRU and DAU). As shown in Table 3, SRU effectively simulates the stochastic diffusion mechanism via random sampling, while DAU implements the evidence accumulation function. Together, these elements constitute the core of the DDM framework. The results clearly demonstrate that their integration significantly enhances the model's performance, validating the proposed brain-inspired decision-making logic.

Ablation of Stage Number. In Table 4, we analyze the impact of the number of iterative stages N. Since the effectiveness of DAU has been verified in previous experiments, N in this experiment starts from 2. It is not difficult to observe that the performance improves when N increases from 2 to 3, but decreases significantly when N exceeds 3. This may be because as the number of stages increases, extremely small-scale features emerge, making it harder for the model to learn more accurate changed regions. More experiment are provided in supplementary materials [1].

5.4 Visualization

Visualization for Uncouple Space. To intuitively assess the CD performance, we utilize T-SNE to project changed (yellow) and unchanged (blue) pixel features onto a 2D plane. As illustrated in Figure 9, our model exhibits a significantly clearer semantic boundary and tighter feature clustering compared to existing methods. This superior feature separation aligns with our quantitative results, confirming the efficacy of the proposed brain-inspired mechanism.

Visualization for Feature Heatmap. To intuitively analyze D^{2}MDM^2 mechanism, Figure 10 visualizes key features. As the stages progress [Figure 10(b-d)], decision prior features exhibit an evident transition from coarse localization to precise differentiation, while bitemporal features increasingly concentrate on change regions. Furthermore, final change

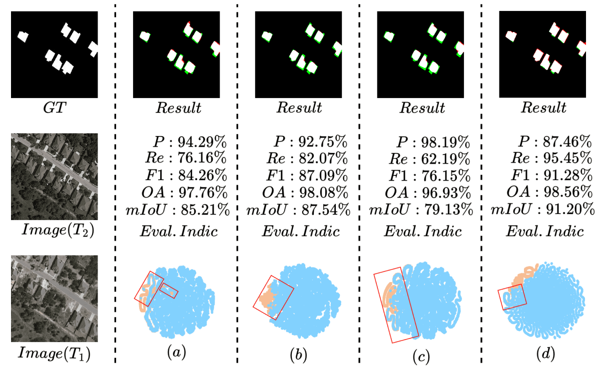

Figure 9: Some visualization examples produced by typical deep models. (a) Results of the BIT. (b) Results of the Change Former. (c) Results of the ICIFNet. (d) Results of our D^{2}MDM^2.

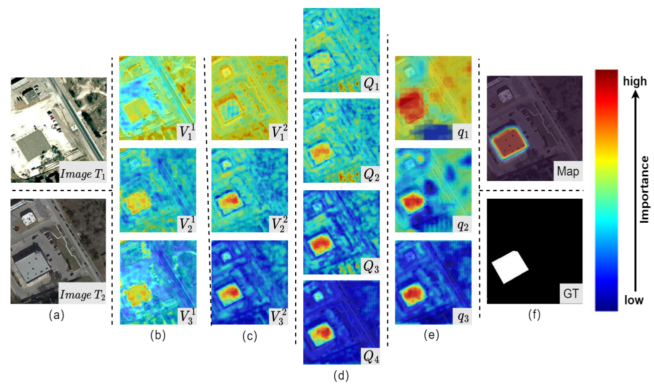

Figure 10: Visualization of key features from D^{2}MDM^2. (a) Dual-phase images T_1 and T_2; (b) Stage-wise features V_i^1 of image T_1; (c) Stage-wise features V_i^2 of image T_2; (d) Initial decision prior features Q and their stage-wise updated results Q_i; (e) Guidance factors q_i; (f) Heatmaps of the change map and ground truth (GT).

heatmap [Figure 10(f)] demonstrates sharp boundaries and accurate target localization. This evolutionary process validates that brain-inspired DDM mechanism effectively refines change cues through multi-stage evidence accumulation, thereby enhancing the discriminative power of model.

6 Conclusion

In this paper, we break through the constraints of traditional one-step CD paradigms by framing illumination-induced performance degradation as an adversarial attack-and-defense problem. We propose D^{2}MDM^2, a novel brain-inspired deep network that integrates DDM with cognitive neuroscience principles. By implementing MRU and the SRU, our model effectively captures the inherent uncertainty in RS environments while mimicking the human "coarse-to-fine" cognitive process. The inclusion of DAU and LGU ensure that model maintains focus on authentic change regions through directional drift modeling and evidence accumulation. Extensive evaluations on multiple datasets with complex illumination perturbations demonstrate that D^{2}MDM^2 significantly outperforms several methods in both accuracy and robustness.

[1]https://github.com/iceking111/D2MDM2

Acknowledgments

This work was supported in part by the Teaching Reform Innovation Program for Higher Education Institutions in Shanxi Province under Grant J20240856; Graduate Education Innovation Program in Shanxi Province under Grant 2025AL12; in part by the Basic Applied Research Projects of Shanxi Province in China under Grants 20210302124165 and TZLH20230818007; in part by the Ministry of Education industry-university cooperative collaborative education project under Grant 221002722143739; in part by the Natural Science Foundation of Hainan Province under Grant422QN350; in part by the National Natural Science Foundation of Youth Science Foundation Project under Grant42001360; in part by the Shanxi Province Graduate Education Innovation Plan under Grants 2025XS444 and 2025SJ326; and in part by the Research Project Supported by the Shanxi Scholarship Council of China under Grants 2024-118.

References

[Bandara and Patel, 2022] 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.

[Bernhard *et al.*, 2023] Maximilian Bernhard, Niklas Strauß, and Matthias Schubert. Mapformer: Boosting change detection by using pre-change information. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 16791–16800, 2023.

[Chen *et al.*, 2022] 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.

[Chen *et al.*, 2023] Chao Chen, Jun-Wei Hsieh, Ping-Yang Chen, Yi-Kuan Hsieh, and Bo Wang. Saras-net: Scale and relation aware siamese network for change detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, abs/2212.01287, 2023.

[Chen *et al.*, 2024] Hongruixuan Chen, Jian Song, Chengxi Han, Junshi Xia, and Naoto Yokoya. Changemamba: Remote sensing change detection with spatiotemporal state space model. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–20, 2024.

[Cho *et al.*, 2023] Seungju Cho, Hongsin Lee, and Changick Kim. Enhancing robustness in incremental learning with adversarial training. In *AAAI Conference on Artificial Intelligence*, 2023.

[Dai *et al.*, 2025] Tao Dai, Song Wang, Hang Guo, Jianping Wang, and Zexuan Zhu. Diin: Diffusion iterative implicit network for arbitrary-scale super-resolution. In *Proceedings of International Joint Conference on Artificia Intelligence (IJCAI-25)*, 2025.

[Feng *et al.*, 2022] Yuchao Feng, Honghui Xu, Jiawei Jiang, Hao Liu, and Jianwei Zheng. Icif-net: Intra-scale cross-interaction and inter-scale feature fusion network for bitemporal remote sensing images change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022.

[Feng *et al.*, 2025] Yuxin Feng, Li Zhuo, Hui Zhang, and Jiafeng Li. Hybrid-mambacd: Hybrid mamba-cnn network for remote sensing image change detection with region-channel attention mechanism and iterative global-local feature fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–12, 2025.

[Gu *et al.*, 2025] Binhao Gu, Lei Song, Youyong Kong, and Binjie Gu. M2f2net: Multi-stage mixed feature fusion network for remote sensing change detection. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2025.

[He *et al.*, 2025] Xin He, Yili Wang, Wenqi Fan, Xu Shen, Xin Juan, Rui Miao, and Xin Wang. Mamba-based graph convolutional networks: tackling over-smoothing with selective state space. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, IJCAI '25, 2025.

[Hills *et al.*, 2024] Anthony Hills, Talia Tseriotou, Xenia Miscouridou, Adam Tsakalidis, and Maria Liakata. Exciting mood changes: A time-aware hierarchical transformer for change detection modelling. In *Annual Meeting of the Association for Computational Linguistics*, 2024.

[Hung-Quang *et al.*, 2024] Nguyen Hung-Quang, Yingjie Lao, Tung Pham, Kok-Seng Wong, and Khoa D Doan. Understanding the robustness of randomized feature defense against query-based adversarial attacks. In *The Twelfth International Conference on Learning Representations*, 2024.

[Lei *et al.*, 2023] 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.

[Li *et al.*, 2024] Xiaoli Li, Hongzhong Li, Luyi Sun, Pan Chen, Longlong Zhao, and Jinsong Chen. Object-oriented sar image change detection based on speckle reducing anisotropic diffusion. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 10080–10083, 2024.

[Lin *et al.*, 2025] Shaoxiong Lin, Tao Lei, Tongfei Liu, Shuxin Zhang, Chongdan Min, and Asoke K. Nandi. Dynamic sparse encoding and cross-temporal attention for remote sensing image change detection. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2025.

[Liu *et al.*, 2021] Yi Liu, Chao Pang, Zongqian Zhan, Xiaomeng Zhang, and Xue Yang. Building change detection for remote sensing images using a dual-task constrained deep siamese convolutional network model. *IEEE Geoscience and Remote Sensing Letters*, 18(5):811–815, 2021.

[Pla, 2010] *Cognitive Load Theory*. Cambridge University Press, 2010.

[Ratcliff and McKoon, 2008] Roger Ratcliff and Gail McKoon. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20(4):873–922, 2008.

[Vinholi *et al.*, 2024] João Gabriel Vinholi, Marco Chini, Anis Amziane, Patrick Matgen, and Renato Machado. Optical image translation using diffusion models in support of heterogeneous change detection. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 7231–7234, 2024.

[Wang *et al.*, 2023] Supeng Wang, Yuxi Li, Ming Xie, Mingmin Chi, Yabiao Wang, Chengjie Wang, and Wenbing Zhu. Align, perturb and decouple: toward better leverage of difference information for rsi change detection. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, IJCAI '23, 2023.

[Wang *et al.*, 2024] Song Wang, Zhenyu Lei, Zhen Tan, Jiaqi Ding, Xinyu Zhao, Yushun Dong, Guorong Wu, Tianlong Chen, Chen Chen, Aiying Zhang, and Jundong Li. Brainmap: Learning multiple activation pathways in brain networks. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2024.

[Wang *et al.*, 2025a] Bin Wang, Xiaohu Jiang, Pinle Qin, and Jianchao Zeng. Dsfdcd: Joint distribution sampling and feature decoupling deep network for remote sensing change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–12, 2025.

[Wang *et al.*, 2025b] Bin Wang, Kang Zhao, Tong Xiao, Pinle Qin, and Jianchao Zeng. Rahff-net: Recall-adjustable hierarchical feature fusion network for remote sensing image change detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18:176–190, 2025.

[Wu *et al.*, 2023] Xiao Wu, Yu Chen, Qiuyang Yan, Yuhang Zhao, Jilu Zhao, Xiaoqing Zhang, Risa Higashita, and Jiang Liu. Dminet: A lightweight dual-mixed channel-independent network for cataract recognition. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2023.

[Wu *et al.*, 2024] Yicheng Wu, Xiangde Luo, Zhe Xu, Xiaoqing Guo, Lie Ju, Zongyuan Ge, Wenjun Liao, and Jianfei Cai. Diversified and personalized multi-rater medical image segmentation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11470–11479, 2024.

[Xu *et al.*, 2025] Shuyi Xu, He Sun, Xu Sun, Li Ni, and Lianru Gao. A hyperspectral change detection method for small vehicles. *IEEE Transactions on Image Processing*, 34:7874–7888, 2025.

[Yang and Shadlen, 2007] Tianming Yang and Michael N. Shadlen. Probabilistic reasoning by neurons. *Nature*, 447(7148):1075–1080, 2007.

[Yu *et al.*, 2025] Xiaofei Yu, Yitong Li, Jie Ma, Chang Li, and Hanlin Wu. Diffusion-rscc: Diffusion probabilistic model for change captioning in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–13, 2025.

[Zhang *et al.*, 2022] Jinsong Zhang, Mengdao Xing, Guang-Cai Sun, and Xin Shi. Vehicle trace detection in two-pass sar coherent change detection images with spatial feature enhanced unet and adaptive augmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022.

[Zhang *et al.*, 2026] Mingwei Zhang, Yuan Jiang, and Qi Wang. Enhancing vmamba for change detection via lightweight feature interaction and selection. *Pattern Recognition*, 172:112648, 2026.

[Zhao *et al.*, 2025] Jingyu Zhao, Jiahui Qu, and Wenqian Dong. Bi-diffcd: bidirectional diffusion guided collaborative change detection for arbitrary-modal remote sensing images. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, IJCAI '25, 2025.

[Zheng *et al.*, 2021] Zhuo Zheng, Ailong Ma, Liangpei Zhang, and Yanfei Zhong. Change is everywhere: Single-temporal supervised object change detection in remote sensing imagery. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15173–15182, 2021.