UCDNET:UNCERTAINTY-GUIDED CAUSAL DISENTANGLED NETWORK FOR REMOTE SENSING CHANGE DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Existing remote sensing change detection (RSCD) methods typically rely on intertwined feature representation learning, jointly encoding semantic backgrounds, real changes, and environmental perturbations. This entangled representation couples change indicators with invariant background semantics, rendering models highly susceptible to spurious correlations induced by variations in illumination, seasonal phenology, and sensor discrepancies. Furthermore, existing paradigms lack the capacity to model pixel-level prediction reliability. To address these limitations, this paper proposes an Uncertainty-Guided Causal Disentangled Network for RSCD. We first formalize the RSCD process via a Structural Causal Model (SCM), establishing a causal generative graph governing invariant backgrounds, authentic semantic changes, and confounding noise, and theoretically demonstrating that explicit feature disentanglement effectively blocks the backdoor path through which confounders bias change inference. Building upon this formulation, we design a Pearson correlation-guided feature disentanglement that distills shared invariant representations via cross-temporal correlations and explicitly segregates change components from background factors. To enhance the discriminative capacity of the disentangled representations, we further introduce a maximum-entropy clustering-based disentangled prototypical contrastive learning strategy. Unlike conventional prototype learning that forces all features to rigidly converge toward a single prototype, this strategy guides initial prototypes toward core prototypes, thereby better preserving the multimodal topology and diversity of real-world change manifolds. Finally, to bolster feature confidence, a progressive evidence-accumulating uncertainty estimation module is incorporated. By parameterizing a Dirichlet distribution over the disentangled representations, this module not only quantifies predictive uncertainty but also optimizes epistemic reliability via an evidence-based loss. Extensive experiments demonstrate that the proposed framework achieves substantial improvements across both qualitative and quantitative benchmarks.

1 INTRODUCTION

Remote sensing change detection identifies and delineates surface changes across bitemporal images, widely applied in urban expansion, disaster assessment, military surveillance, and vegetation monitoring.(Xu et al., 2025; Zhang et al., 2026; Zhao et al., 2025; Lv et al., 2025).Illumination shifts pose critical challenges due to uncontrolled multi-temporal and seasonal imaging conditions. Existing solutions primarily rely on feature extraction or domain adaptation. However, feature-based networks generalize poorly across domains, while domain adaptation degrades markedly under severe domain shifts.(Wang et al., 2025; Qu et al., 2025).

As in Fig1(a), traditional networks extract and fuse entangled features, which confuses superficial variations with authentic semantic changes and blurs decision boundaries. Conversely, our approach disentangles images into independent attribute subsets first Fig1(b), establishing clear causal relations with change predictions.

To eliminate false alarms caused by noise and background spurious correlations, we propose the Uncertainty-Guided Causal Disentangled Network. From a causal perspective, the invariant back-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

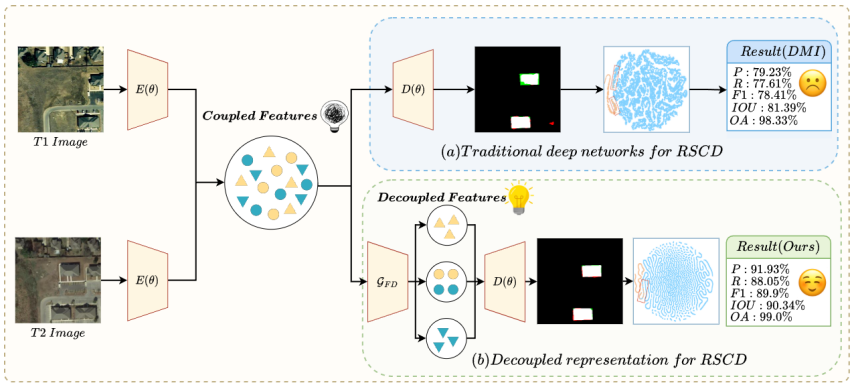

Figure 1: Comparison of change detection paradigms. (a) **Traditional (Coupled)**: Entangled features confuse environmental shifts with genuine changes, causing severe manifold overlap and false alarms. (b) **Ours (Decoupled)**: The decoupler $\mathcal{G}_{FD}$ separates shared background from private changes, eliminating confounding bias and achieving clear manifold separation.

ground acts as an open backdoor confounder between bitemporal observations and change predictions. By designing a principled Structural Causal Model, UCDNet blocks spurious paths via Pearson correlation-guided feature disentanglement feature decoupling, expands feature margins via maximum-entropy prototypical contrast, and quantifies predictive confidence through progressive Dirichlet evidence accumulation.

The main contributions of this paper are summarized as follows:

(1)The causal disentanglement framework for RSCD is proposed from a structural causal perspective. Based on remote sensing physical imaging mechanisms, a SCM is constructed using invariant background, genuine semantic change, and environmental perturbations as its main variables. Furthermore, causal analysis theoretically proves that explicitly isolating shared invariant features and private change features severs spurious backdoor paths induced by background context bias, thereby eliminating shortcut learning and improving model causal interpretability.

(2)Based on the structural causal model, the dedicated Pearson channel-covariance feature decoupler and maximum-entropy prototypical contrastive learning scheme are designed to achieve decoupled representation of background and change. Specifically, cross-temporal Pearson channel-covariance matrices are utilized to adaptively isolate and strip the shared public background substrate via subspace projection. Furthermore, maximum-entropy regularization is introduced to relax the rigid single-center prototype constraint over multimodal change distributions, which preserves fine-grained change diversity while significantly expanding the geometric margin between positive and negative samples on the decoupled manifold.

(3)Progressive evidential uncertainty quantification framework is introduced by incorporating evidential deep learning into decoupled change reasoning. Decoupled change representations are directly mapped into the parameter space of a Dirichlet distribution, concurrently generating pixel-level change predictions and closed-form epistemic uncertainty in a single forward pass. Furthermore, a multi-stage progressive evidence accumulation module is designed to aggregate hierarchical spatial textures and semantic evidence across scales.Additionally, evidence-calibrated loss functions suppress overconfident predictions in ambiguous boundary and transition zones, which strengthens model prediction reliability and trustworthiness for RSCD.

2 Related Work

2.1 Feature Disentanglement

Feature disentanglement maps mixed representations into independent subspaces encoding specific factors, improving model interpretability(Zhang et al., 2025). In RSCD, methods are broadly classified into direct and indirect decoupling. Indirect methods implicitly extract attributes (e.g., style and content) via encoder-decoders(Chen et al., 2024b; Lu et al., 2024) and infer changes from content representations. Dai et al. (2024) and Feng et al. (2025) proposed an adaptive fractional-order domain disentanglement network for hyperspectral CD.

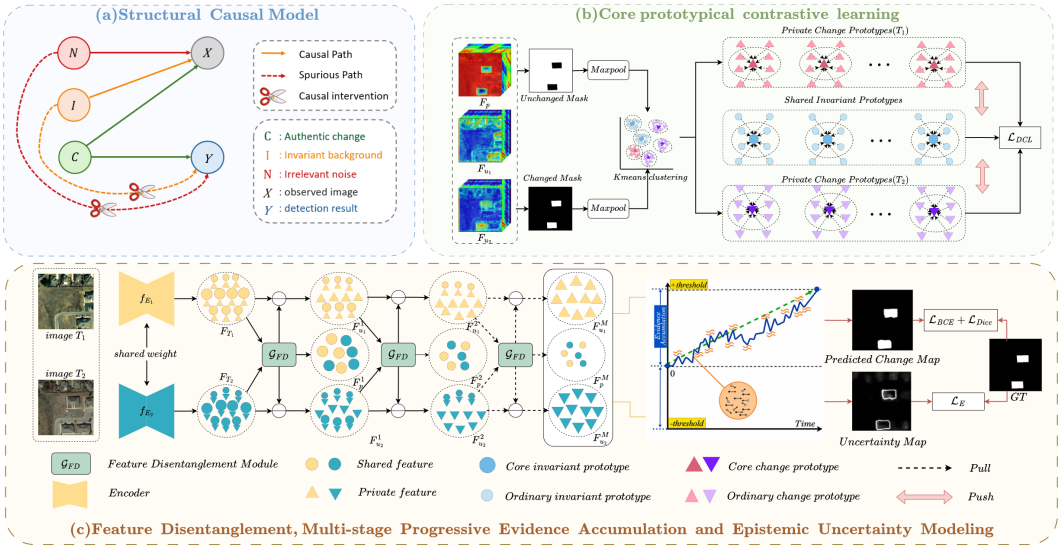

Figure 2: Architecture of the proposed UCDNet. (a) SCM establishes causal dependencies among latent variables and executes causal intervention to sever confounding backdoor paths ($\boldsymbol{I} \rightarrow \boldsymbol{Y}, \boldsymbol{N} \rightarrow \boldsymbol{Y}$). (b) Core prototypical contrastive learning clusters strip prototypes and anchors them to entropy-maximized core prototypes, pulling intra-class samples and pushing inter-class prototypes to retain change diversity while repelling noise. (c) Main pipeline progressively separates shared and private features across multi-stage Feature Disentanglement modules, accumulating evidential Dirichlet strength to generate calibrated change predictions and overcome boundary uncertainty.

2.2 Prototypical and Contrastive Learning

Prototype learning extracts representative prototypes to model class-level semantics for classification and inference(Liu et al., 2025; Ma et al., 2025). For RSCD, inputs are projected into the prototype space, and distance metrics (e.g., Euclidean or cosine similarity) determine final decisions(Chen et al., 2025) .Contrastive learning pulls positive pairs closer and pushes negative pairs apart, learning discriminative representations from unlabeled data (Li et al., 2025; Tian et al., 2025; Yu et al., 2025). In RSCD, Lv et al. (2026)modeled graph dependencies,Ding et al. (2025) used cross-view contrastive loss for noise invariance, and triplet schemes constrained anchor-positive-negative pairs to boost discriminability(Liu et al., 2021; Yan et al., 2024; Zheng et al., 2023).

3 Method

3.1 SCM and Causal Intervention

As formulated in the DAG Fig2(a), the observed bitemporal pair $\boldsymbol{X} = \{\boldsymbol{T_1}, \boldsymbol{T_2}\}$ is generated by three latent causal variables: invariant background $\boldsymbol{I}$, authentic change $\boldsymbol{C}$, and nuisance variations $\boldsymbol{N}$ (illumination, sensor noise). The causal inference path is $\boldsymbol{C} \rightarrow \boldsymbol{Y}$, whereas $\boldsymbol{I} \rightarrow \boldsymbol{Y}$ and $\boldsymbol{N} \rightarrow \boldsymbol{Y}$ constitute confounding backdoor paths.

Initial features $\boldsymbol{F_{T_i}} \in \mathbb{R}^{H \times W \times C}$ extracted by encoder f_{E_i} from $\boldsymbol{T_i} \in \mathbb{R}^{4H \times 4W \times 3}$ are nonlinear mixtures: $\boldsymbol{F_{T_i}} = f_{E_i}(\boldsymbol{I}, \boldsymbol{C_i}, \boldsymbol{N_i})$. A first-order Taylor expansion around $(\boldsymbol{I}, \boldsymbol{0}, \boldsymbol{0})$ yields the additive approximation:

$$\boldsymbol{F_{T_i}} \approx \Phi(\boldsymbol{I}) + \Delta\Phi(\boldsymbol{C_i}) + \Delta\Phi(\boldsymbol{N_i}) \tag{1}$$

where $\boldsymbol{\Phi_I} = \boldsymbol{f_{E_i}}(\boldsymbol{I}, \boldsymbol{0}, \boldsymbol{0})$ denotes the pure invariant background embedding, while $\boldsymbol{\Delta\Phi_{C_i}}$ and $\boldsymbol{\Delta\Phi_{N_i}}$ represent local displacements induced by authentic changes and nuisance variations, respectively. Conditioned on $\boldsymbol{I}$, change and noise displacements satisfy conditional independence: $\boldsymbol{C_1} \perp\!\!\!\perp \boldsymbol{C_2} \mid \boldsymbol{I}$ and $\boldsymbol{N_1} \perp\!\!\!\perp \boldsymbol{N_2} \mid \boldsymbol{I}$. Hence, cross-terms vanish in expectation:

$$\mathbb{E}\left[\boldsymbol{\Delta\Phi}(\boldsymbol{C_1})^{\top} \boldsymbol{\Delta\Phi}(\boldsymbol{C_2})\right] = \boldsymbol{0}, \quad \mathbb{E}\left[\boldsymbol{\Delta\Phi}(\boldsymbol{N_1})^{\top} \boldsymbol{\Delta\Phi}(\boldsymbol{N_2})\right] = \boldsymbol{0} \tag{2}$$

The cross-temporal correlation expectation simplifies to $\mathbb{E}[\boldsymbol{F_{T_1}^{\top}} \boldsymbol{F_{T_2}}] \approx \mathbb{E}[\boldsymbol{\Phi_I^{\top}} \boldsymbol{\Phi_I}]$, proving that cross-temporal statistics suppress uncorrelated change vectors, forcing shared feature $\boldsymbol{F_p} \rightarrow \boldsymbol{\Phi_I}$.

Subtracting $\boldsymbol{F_p}$ explicitly severs the backdoor path $\boldsymbol{I} \rightarrow \boldsymbol{Y}$:

$$\boldsymbol{F_{u_i}} = \boldsymbol{F_{T_i}} - \boldsymbol{F_p} \approx (\boldsymbol{\Phi}(\boldsymbol{I}) + \boldsymbol{\Delta\Phi}(\boldsymbol{C_i}) + \boldsymbol{\Delta\Phi}(\boldsymbol{N_i})) - \boldsymbol{\Phi}(\boldsymbol{I}) = \boldsymbol{\Delta\Phi}(\boldsymbol{C_i}) + \boldsymbol{\Delta\Phi}(\boldsymbol{N_i}) \quad (3)$$

Following the backdoor adjustment formula, the true causal effect is obtained by marginalizing over $\boldsymbol{I}$:

$$\boldsymbol{P}(\boldsymbol{Y} \mid do(\boldsymbol{X})) = \int P(\boldsymbol{Y} \mid \boldsymbol{\Delta\Phi_C}, \boldsymbol{\Delta\Phi_N}, \boldsymbol{\Phi_I}) P(\boldsymbol{\Phi_I}) \, \mathrm{d}\boldsymbol{\Phi_I} \approx P(\boldsymbol{Y} \mid \boldsymbol{\Delta\Phi_C}, \boldsymbol{\Delta\Phi_N}) \quad (4)$$

Explicit disentanglement neutralizes the confounding effect of $\boldsymbol{\Phi_I}$, and prototypical contrastive learning subsequently filters out noise offsets $\boldsymbol{\Delta\Phi_N}$, compelling predictions to rely strictly on true change vectors $\boldsymbol{\Delta\Phi_C}$.

3.2 Pearson correlation-guided feature disentanglement

To extract $\boldsymbol{F_p}$, $\boldsymbol{F_{T_i}}$ is projected to $\boldsymbol{Q}, \boldsymbol{K} \in \mathbb{R}^{N_s \times C}$ ($N_s = HW$).The channel Pearson correlation matrix $\boldsymbol{r} \in \mathbb{R}^{C \times C}$ is computed as:

$$r_{u,v} = \frac{\sum_{s=1}^{N_s} \left(\boldsymbol{Q}_{s,u} - \bar{\boldsymbol{Q}}_u\right)\left(\boldsymbol{K}_{s,v} - \bar{\boldsymbol{K}}_v\right)}{\sqrt{\sum_{s=1}^{N_s} \left(\boldsymbol{Q}_{s,u} - \bar{\boldsymbol{Q}}_u\right)^2} \sqrt{\sum_{s=1}^{N_s} \left(\boldsymbol{K}_{s,v} - \bar{\boldsymbol{K}}_v\right)^2}} \quad (5)$$

The matrix $\boldsymbol{r}$ quantifies the linear statistical dependency between any channel pair (u, v) across time, attenuating time-specific channels while emphasizing consistent cross-temporal responses. Modulating spatial cross-attention with $\boldsymbol{r}$ aggregates the shared invariant feature $\boldsymbol{F_p}$:

$$\boldsymbol{F_p} = \mathrm{Softmax}\left(\frac{\boldsymbol{Q r K}^\top}{\sqrt{C}}\right)\left(\frac{\boldsymbol{Q} + \boldsymbol{K}}{2}\right) \quad (6)$$

Following the additive decomposition of our causal formulation, subtracting the shared background baseline $\boldsymbol{F_p}$ isolates private feature representations $\boldsymbol{F_{u_1}}$ and $\boldsymbol{F_{u_2}}$, which contain only authentic land-cover changes and temporal perturbations:

$$\boldsymbol{F_{u_1}} = \boldsymbol{F_{T_1}} - \boldsymbol{F_p}, \boldsymbol{F_{u_2}} = \boldsymbol{F_{T_2}} - \boldsymbol{F_p} \quad (7)$$

To prevent feature degradation, a multi-stage progressive architecture with symmetric deconvolutional decoders is adopted to reconstruct images in a self-supervised manner, jointly optimized with an L_1 reconstruction loss $\mathcal{L}_{\mathrm{RE}}$.

$$\mathcal{L}_{RE} = \sum_{l=1}^{L} \left(\left| \widehat{\boldsymbol{T}}_{\boldsymbol{1}}^{\boldsymbol{l}} - \boldsymbol{T_1} \right|_{\boldsymbol{1}} + \left| \widehat{\boldsymbol{T}}_{\boldsymbol{2}}^{\boldsymbol{l}} - \boldsymbol{T_2} \right|_{\boldsymbol{1}} \right) \quad (8)$$

3.3 Core Prototypical Contrastive Learning

Conventional prototype contrast forces representations toward rigid unimodal centroids, which smooths out authentic multimodal change distributions and admits nuisance noise backdoors ($\boldsymbol{N} \rightarrow \boldsymbol{Y}$). To overcome this, we propose core prototypical contrastive learning. Rather than rigid pixel-to-centroid alignment, contrastive learning guides regional strip prototypes toward entropy-maximized core anchors, preserving the intrinsic spectral diversity of genuine changes $\boldsymbol{\Delta\Phi_C}$ while expelling spatially fragmented noise $\boldsymbol{\Delta\Phi_N}$.

For class $\boldsymbol{\psi} \in \{pri, pub\}$ with binary mask $\boldsymbol{M_\psi}$ ($\boldsymbol{M_{pri}} = \mathbf{GT}, \boldsymbol{M_{pub}} = \mathbf{1} - \mathbf{GT}$), horizontal pooling along the vertical dimension extracts strip prototypes $\boldsymbol{p_\psi} = \mathrm{MaxPool}_H(\boldsymbol{F_\psi} \odot \boldsymbol{M_\psi}) \in \mathbb{R}^{W \times C}$.

To capture diverse sub-semantics, K-means partitions $\boldsymbol{p_\psi}$ into K_ψ sub-clusters. Denoting $\boldsymbol{C_\psi^j}$ as the $\boldsymbol{j}$-th cluster centroid, the subset Ω_ψ^k for sub-cluster $\boldsymbol{k}$ is:

$$\Omega_\psi^k = \left\{ \boldsymbol{p} \in \boldsymbol{p_\psi} \mid k = \arg\max_j \left\langle \boldsymbol{p}, \boldsymbol{C_\psi^j} \right\rangle \right\}, \quad j, k \in \{1, 2, \ldots, K_\psi\} \quad (9)$$

To bypass noisy boundary outliers, we calculate the similarity distribution $\pi_\psi^{i,n}$and information entropy $\mathcal{H}(\boldsymbol{p_\psi^i})$ for candidate prototypes $\boldsymbol{p_\psi^i} \in \Omega_\psi^k$ against feature tensor $\boldsymbol{F_\psi}$:

$$\pi_\psi^{i,n} = \frac{\exp\left(\left\langle \boldsymbol{p_\psi^i}, \boldsymbol{F_\psi^n} \right\rangle / \tau\right)}{\sum_{m=1}^{N} \exp\left(\left\langle \boldsymbol{p_\psi^i}, \boldsymbol{F_\psi^m} \right\rangle / \tau\right)} \quad (10)$$

The candidate with maximal information entropy serves as the core anchor prototype $\boldsymbol{p}^{\boldsymbol{k}}_{\boldsymbol{core},\boldsymbol{\psi}}$:

$$\boldsymbol{p}^{\boldsymbol{k}}_{\boldsymbol{core},\boldsymbol{\psi}} = \underset{\boldsymbol{p}^{\boldsymbol{i}}_{\boldsymbol{\psi}} \in \boldsymbol{\Omega}^{\boldsymbol{k}}_{\boldsymbol{\psi}}}{\arg\max}\, \mathcal{H}(\boldsymbol{p}^{\boldsymbol{i}}_{\boldsymbol{\psi}}) = \underset{\boldsymbol{p}^{\boldsymbol{i}}_{\boldsymbol{\psi}} \in \boldsymbol{\Omega}^{\boldsymbol{k}}_{\boldsymbol{\psi}}}{\arg\max} \left(-\sum_{n=1}^{N} \boldsymbol{\pi}^{\boldsymbol{i},\boldsymbol{n}}_{\boldsymbol{\psi}} \log \left(\boldsymbol{\pi}^{\boldsymbol{i},\boldsymbol{n}}_{\boldsymbol{\psi}} \right) \right) \tag{11}$$

Taking $\boldsymbol{p}^{\boldsymbol{k}}_{\boldsymbol{core},\boldsymbol{\psi}}$ as the anchor, remaining intra-cluster prototypes form the positive set $\boldsymbol{p}^{\boldsymbol{k}}_{\boldsymbol{pos},\boldsymbol{\psi}} = \Omega^{k}_{\psi} \setminus \{\boldsymbol{p}^{\boldsymbol{k}}_{\boldsymbol{core},\boldsymbol{\psi}}\}$, while prototypes from the opposing class constitute the negative set $\boldsymbol{p}^{\boldsymbol{k}}_{\boldsymbol{neg},\boldsymbol{\psi}}$ (i.e., when $\boldsymbol{\psi} = pri$, negatives are drawn from $\boldsymbol{p}_{\boldsymbol{pub}}$; when $\boldsymbol{\psi} = pub$, negatives are drawn from $\boldsymbol{p}_{\boldsymbol{pri}}$). The multi-positive contrastive loss for cluster k of class $\boldsymbol{\psi}$ is formulated as:

$$\mathcal{L}^{k}_{sub,\psi} = -\frac{1}{\left|\boldsymbol{p}^{+}_{\boldsymbol{\psi},\boldsymbol{k}}\right|} \sum_{\boldsymbol{p}^{+} \in \boldsymbol{p}^{k}_{pos,\psi}} \log \frac{\exp\left(\left\langle \boldsymbol{p}^{+}, \boldsymbol{p}^{k}_{\boldsymbol{core},\boldsymbol{\psi}} \right\rangle / \tau\right)}{\exp\left(\left\langle \boldsymbol{p}^{+}, \boldsymbol{p}^{k}_{\boldsymbol{core},\boldsymbol{\psi}} \right\rangle / \tau\right) + \sum\limits_{\boldsymbol{p}^{-} \in \boldsymbol{p}^{k}_{\boldsymbol{neg},\boldsymbol{\psi}}} \exp\left(\left\langle \boldsymbol{p}^{-}, \boldsymbol{p}^{k}_{\boldsymbol{core},\boldsymbol{\psi}} \right\rangle / \tau\right)} \tag{12}$$

Aggregating over all clusters across both private and public feature categories, the overall disentangled prototypical contrastive learning objective is defined as:

$$\mathcal{L}_{\boldsymbol{DCL}} = \sum_{\boldsymbol{\psi} \in \{pri,pub\}} \sum_{k=1}^{K_{\psi}} \mathcal{L}^{k}_{sub,\psi} \tag{13}$$

This contrastive formulation maximizes the geometric margin between change and invariant manifolds, effectively severing spurious non-causal paths.

3.4 Progressive Evidential Uncertainty Modeling

To mitigate overconfident classifications in ambiguous transition zones, we integrate Evidential Deep Learning (EDL). At stage K, non-negative evidence vectors for change features $\boldsymbol{F}^{\boldsymbol{k}}_{\boldsymbol{u_i}}$ and background features $\boldsymbol{F}^{\boldsymbol{k}}_{\boldsymbol{p}}$ are extracted via bounded activations:

$$\boldsymbol{e}^{\boldsymbol{k}}_{\boldsymbol{u_i}} = \exp\left(\tanh\left(\boldsymbol{F}^{\boldsymbol{k}}_{\boldsymbol{u_i}}\right)\right), \quad \boldsymbol{e}^{\boldsymbol{k}}_{\boldsymbol{p}} = \exp\left(\tanh\left(\boldsymbol{F}^{\boldsymbol{k}}_{\boldsymbol{p}}\right)\right) \tag{14}$$

We accumulate evidence across all K stages to obtain aggregated representations:

$$\boldsymbol{E_c} = \sum_{\boldsymbol{k=1}}^{\boldsymbol{K}} \boldsymbol{e}^{\boldsymbol{k}}_{\boldsymbol{u_i}}, \quad \boldsymbol{E_{nc}} = \sum_{\boldsymbol{k=1}}^{\boldsymbol{K}} \boldsymbol{e}^{\boldsymbol{k}}_{\boldsymbol{p}} \tag{15}$$

Under Dempster-Shafer(Bao et al., 2021) Theory and Subjective Logic, class belief masses $\boldsymbol{B_c}, \boldsymbol{B_{nc}}$ and epistemic uncertainty $\boldsymbol{U}$ satisfy:

$$\boldsymbol{B_c} = \frac{\boldsymbol{E_c}}{\boldsymbol{S}}, \quad \boldsymbol{B_{nc}} = \frac{\boldsymbol{E_{nc}}}{\boldsymbol{S}}, \quad \boldsymbol{U} = \frac{\boldsymbol{2}}{\boldsymbol{S}}, \quad \text{where } \boldsymbol{S} = \boldsymbol{E_c} + \boldsymbol{E_{nc}} + \boldsymbol{2} \tag{16}$$

$$\boldsymbol{U} + \boldsymbol{B_c} + \boldsymbol{B_{nc}} = \boldsymbol{1}_{H \times W} \tag{17}$$

where $\boldsymbol{S}$ denotes total Dirichlet strength. When $\boldsymbol{E_c} = \boldsymbol{E_{nc}} = \boldsymbol{0}$, uncertainty reaches its maximum ($\boldsymbol{U} = \boldsymbol{1}$). Expected class probabilities $\boldsymbol{P} = [\boldsymbol{P_c}, \boldsymbol{P_{nc}}]$ derive from concentration parameters $\boldsymbol{\alpha} = [\boldsymbol{\alpha_c}, \boldsymbol{\alpha_{nc}}] = [\boldsymbol{E_c} + \boldsymbol{1}, \boldsymbol{E_{nc}} + \boldsymbol{1}]$:

$$[\boldsymbol{P_c}, \boldsymbol{P_{nc}}] = \frac{[\boldsymbol{\alpha_c}, \boldsymbol{\alpha_{nc}}]}{\boldsymbol{S}} \tag{18}$$

Parameter $\boldsymbol{\alpha}$ defines a Dirichlet distribution over the class probability simplex $\boldsymbol{\mathcal{S}_K}$ ($K = 2$):

$$\mathcal{D}\left(\boldsymbol{P}|\boldsymbol{\alpha}\right) = \begin{cases} \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^{K} P_k^{\alpha_k - 1}, & \text{for } \boldsymbol{P} \in \boldsymbol{\mathcal{S}_K} \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

We minimize the expected cross-entropy loss under the Dirichlet distribution:

$$\mathcal{L}_E = \int \left[\sum_{k=1}^{K} \boldsymbol{y}(\boldsymbol{k}) \log\left(\boldsymbol{S_K}\right) \right] \mathcal{D}\left(\boldsymbol{S}|\boldsymbol{\alpha}\right) d\boldsymbol{S} = \sum_{k=1}^{K} \boldsymbol{y}(\boldsymbol{k}) \left(\boldsymbol{\psi}\left(\boldsymbol{S_K}\right) - \boldsymbol{\psi}\left(\boldsymbol{\alpha}\right) \right) \tag{20}$$

where $\boldsymbol{\psi}(\cdot)$ denotes the Digamma function. Standard BCE and Dice losses are jointly optimized across stages to supervise segmentation: The overall composite training objective is:

$$\mathcal{L}_{total} = \lambda_1 \cdot \mathcal{L}_{BCE} + \lambda_2 \cdot \mathcal{L}_{Dice} + \lambda_3 \cdot \mathcal{L}_{DCL} + \lambda_4 \cdot \mathcal{L}_{RE} + \lambda_5 \cdot \mathcal{L}_{E} \tag{21}$$

4 EXPERIMENTAL SETUP

UCDNet is implemented in PyTorch on a single NVIDIA RTX 4090 GPU. We optimize the model using AdamW (momentum 0.9, weight decay 3×10^{-4}) with an initial learning rate of 0.001, decayed by a factor of 0.9 every 20 epochs.In addition,total number of training epochs is set to 150 for all datasets with batch size of 16 and validation is performed after each training epoch,while model with optimal performance on validation set is finally used for test set evaluation.

5 EXPERIMENTAL RESULTS

5.1 PERFORMANCE COMPARISON

5.1.1 OBJECTIVE INDICATOR

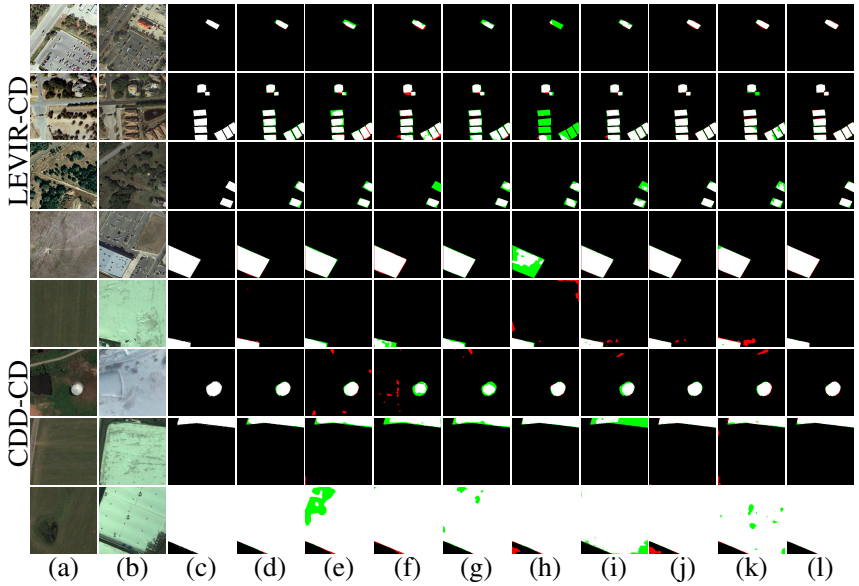

Figure 3: Detection results across different methods: (a) T_1 image; (b) T_2 image; (c) Ground truth; (d) ChangeFormer; (e) ICIFNet; (f) USSFC-Net; (g) DMINet; (h) ChangeMamba; (i) ConvFormer-CD; (j) FCdDNet; (k) UAFD-cd; (l) Ours.

To validate our model, we benchmark it against state-of-the-art CD methods, including ChangeFormer(Bandara & Patel, 2022), ICIFNet (Feng et al., 2022), USSFC-Net (Lei et al., 2023), DMINet (Feng et al., 2023), ChangeMamba (Chen et al., 2024a), ConvFormer-CD (Yang et al., 2025), FCdDNet (Wang et al., 2026b), and UAFD-cd(Wang et al., 2026a). For fair comparison, all baselines are executed using official open-source codebases with hyperparameters configured as reported in their original publications.

Table 1: Performance comparison of different methods on various datasets.

Methods	LEVIR-CD					SYSU-CD					SECOND					CDD-CD				
	P.	Re.	F1	OA	mIoU	P.	Re.	F1	OA	mIoU	P.	Re.	F1	OA	mIoU	P.	Re.	F1	OA	mIoU
ChangeFormer	91.17	78.57	84.40	98.52	85.74	80.83	73.68	77.09	89.67	75.11	74.61	50.58	66.79	87.95	68.21	95.68	90.55	93.05	98.33	92.56
ICIFNet	76.65	81.25	78.88	98.30	81.69	84.60	70.70	77.03	90.06	75.36	**78.21**	53.96	63.86	87.79	66.61	94.71	88.56	91.53	98.07	91.11
USSFC-Net	**94.85**	73.98	83.13	98.47	84.77	81.66	68.24	74.35	88.90	72.97	73.15	54.43	62.48	86.90	65.37	93.11	83.42	88.00	97.19	87.72
DMINet	79.23	77.61	78.41	98.33	81.39	86.07	71.13	77.89	90.48	76.17	74.18	58.50	65.41	87.63	67.30	95.22	90.41	92.75	98.33	92.31
ChangeMamba	94.16	75.81	84.01	98.53	85.44	79.55	73.91	76.63	89.37	74.62	69.29	63.19	66.10	87.04	67.26	96.20	90.38	93.20	98.37	92.71
ConvFormer-CD	76.98	85.99	81.24	98.45	83.40	**86.22**	73.49	79.35	**90.98**	**77.43**	69.03	63.28	66.03	87.01	67.20	94.81	93.14	93.96	98.59	93.52
FCdDNet	92.05	87.13	89.52	98.96	89.97	74.96	77.66	76.28	88.61	73.86	68.42	**69.01**	68.72	87.33	68.88	**96.91**	90.24	93.46	98.44	92.98
UAFD-cd	90.69	**88.73**	89.70	98.96	90.12	75.16	76.70	75.97	88.54	73.63	67.47	68.45	67.96	87.09	68.25	94.17	90.60	92.35	98.15	91.85
Ours	91.93	88.05	**89.95**	**99.00**	**90.34**	77.35	**82.53**	**79.85**	90.18	77.13	71.47	67.24	**69.29**	**88.08**	**69.62**	95.24	**93.66**	**94.44**	**98.64**	**93.96**

* All values are reported in percentage (%). Color convention: **best**, 2nd best.

Table 1 reports quantitative results on LEVIR-CD, SYSU-CD, SECOND, and CDD. Our model achieves superior performance across comprehensive metrics, attaining an 89.95 F1-score and 90.34 mIoU on LEVIR-CD.

5.1.2 VISUAL COMPARISON

Visual comparisons across datasets are depicted in Fig.3 and 4. In small-target scenarios, baselines such as USSFC-Net and DMINet suffer from severe omissions, whereas models like ICIFNet incur frequent false alarms, with error rates escalating under complex scenes. Subtle changes are often missed entirely, while extensive continuous regions frequently suffer from concurrent boundary false alarms and missed detections. By severing noise backdoor paths via prototypical contrast and quantifying boundary ambiguity through evidential learning, our model demonstrates high resilience to complex terrains, delivering sharp, regularized boundaries and effectively suppressing pseudo-changes.

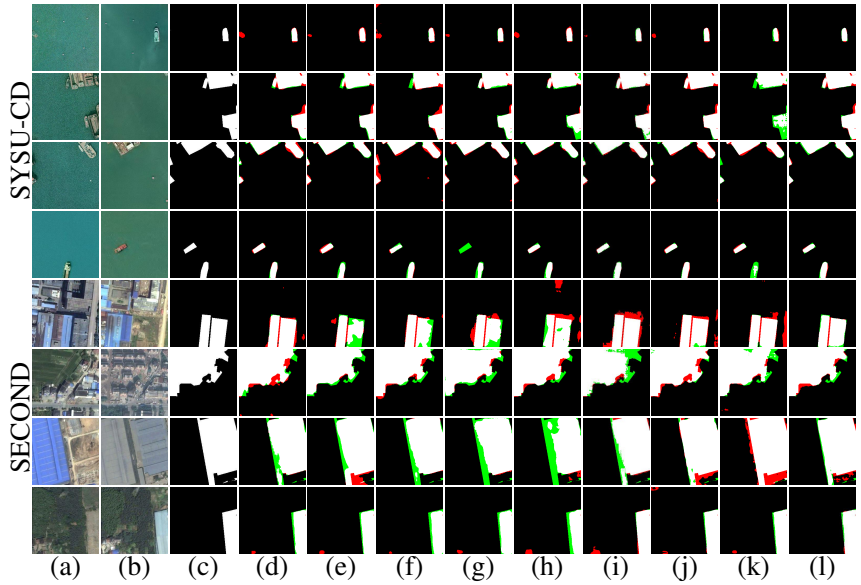

Figure 4: Detection results across different methods: (a) T_1 image; (b) T_2 image; (c) Ground truth; (d) ChangeFormer; (e) ICIFNet; (f) USSFC-Net; (g) DMINet; (h) ChangeMamba; (i) ConvFormer-CD; (j) FCdDNet; (k) UAFD-cd; (l) Ours.

5.2 ABLATION EXPERIMENTS

5.2.1 ABLATION EXPERIMENTS OF MODULES

We conduct ablation experiments to isolate the contribution of each key component (feature disentanglement $\mathcal{G}_{FD}$, $\mathcal{L}_{DCL}$,$\mathcal{L}_{E}$, and$\mathcal{L}_{RE}$). As shown in Table 2 (where "✓" indicates module inclusion), each component consistently contributes to the overall performance gains across benchmark datasets.

Table 2: Comparative experimental results

$\mathcal{G}_{FD}$	$\mathcal{L}_{DCL}$	$\mathcal{L}_{E}$	$\mathcal{L}_{RE}^{T_i}$	P.	Re.	F1	OA	mIoU
	✓	✓	✓	89.10	**89.19**	89.15	98.89	89.63
✓		✓	✓	**99.59**	86.66	89.20	98.90	89.70
✓	✓		✓	90.44	86.06	88.20	98.83	88.83
✓	✓	✓		89.07	88.51	88.79	98.86	89.32
✓	✓	✓	✓	91.93	88.05	**89.95**	**99.00**	**90.34**

Table 3: Evaluation of cluster numbers

Number of clusters		P	Re	F1	OA	mIoU
private	public					
3	4	91.93	**88.05**	**89.95**	**99.00**	**90.34**
4	4	91.14	85.41	88.18	98.83	88.82
2	4	**92.10**	87.06	89.51	98.96	89.96
3	5	91.01	86.58	88.74	98.88	89.29
3	3	89.92	87.92	88.91	98.88	89.43

5.2.2 ABLATION OF REGULARIZATION COEFFICIENTS

To evaluate sensitivity to regularization coefficients and identify an optimal trade-off, we tune the loss weights on the SYSU dataset. As listed in Table 4, the configuration $\lambda = [5, 5, 1.5, 0.5, 0.5]$ achieves the best overall performance. Applying this setting directly to the remaining datasets yields consistently strong results, demonstrating the robustness and generalizability of the selected parameters.

Table 4: Influence of regularization on SYSU

λ_1	λ_2	λ_3	λ_4	λ_5	P.	Re.	F1	OA	mIoU
5	5	1.5	0.5	0.5	77.35	82.53	**79.85**	90.18	**77.13**
10	5	1.5	0.5	0.5	72.30	81.58	76.66	88.29	73.83
2.5	5	1.5	0.5	0.5	66.41	**86.81**	75.25	86.54	71.70
5	7.5	1.5	0.5	0.5	74.64	82.23	78.25	89.22	75.45
5	2.5	1.5	0.5	0.5	79.12	77.07	78.08	89.80	75.79
5	5	3	0.5	0.5	81.09	75.26	78.07	90.03	75.95
5	5	0.75	0.5	0.5	**86.12**	70.67	77.64	**90.40**	75.96
5	5	1.5	1	0.5	74.73	81.11	77.79	89.08	75.07
5	5	1.5	0.25	0.5	85.53	69.23	76.52	89.98	75.00
5	5	1.5	0.5	1	76.43	78.13	79.17	89.81	76.44
5	5	1.5	0.5	0.25	82.81	72.11	77.09	89.89	75.27

Table 5: Cross-Domain generalization test

Methods	CDD		SECOND		SYSU	
	mF1	mIoU	mF1	mIoU	mF1	mIoU
Change Former	54.36	44.33	52.42	44.31	43.98	38.57
ICIFNet	51.51	45.49	55.57	46.47	44.46	38.83
USSFC-Net	53.84	44.67	51.89	43.90	43.79	38.47
DMINet	56.13	46.89	57.66	48.17	43.74	38.44
ChangeMamba	55.72	46.42	49.48	42.63	43.46	38.29
ConvFormer-CD	47.74	44.54	50.51	44.2	44.49	38.86
UAFD-cd	56.40	45.94	49.99	43.08	44.09	38.61
FCdDNet	54.45	44.80	50.80	43.55	43.83	38.48
Ours	**56.89**	**47.35**	**54.09**	**45.41**	**45.35**	**39.29**

5.2.3 The impact of cluster values on the LEVIR dataset

Cluster Count Ablation: During prototype construction, cluster numbers determine representative anchor selection. In Table 3,we ablate prototype cluster counts on LEVIR-CD. The results indicate that an excessive number of clusters dilutes prototype representativeness, whereas too few clusters make prototypes susceptible to noise, both degrading change detection accuracy.

5.2.4 Cross-Domain Generalization and Model Complexity

We evaluate the generalization of models trained on LEVIR across several unseen datasets Table 5. UCDNet achieves the best cross-domain performance on CDD and SYSU and remains competitive on SECOND, although DMINet performs better on the latter., all metrics drop significantly: domain shifts from background, seasonal, and imaging differences exceed the capacity of our causal decoupling, which limits generalization and underscores it as a key future direction.Table 6 summarizes the model complexity. Our model is not the lightest in parameters or computation, but the cost is acceptable, and improving efficiency for lightweight deployment is a primary focus of future work.

Table 6: FLOPs and params comparison of different methods

Methods	ChangeFormer	ICIFNet	USSFC-Net	DMINet	ChangeMamba	ConvFormer-CD	FCdDNet	UAFD-cd	**Ours**
FLOPs(G)	21.18	25.36	**4.86**	14.55	45.74	5.14	9.59	12.08	9.5
Params(M)	29.75	23.82	**1.52**	6.24	17.13	37.72	30.99	8.32	26.84

5.3 Visualization Analysis

5.3.1 Feature Separation Visualization

To intuitively assess feature discriminability, we present t-SNE visualizations in Fig.5, where orange and blue dots denote changed and unchanged pixel embeddings, and red boxes highlight confusion regions. Compared to baselines, our model exhibits clearer feature separation and minimal distribution overlap, aligning with its superior single-image F1-score, OA, and mIoU metrics. This advantage stems from the SCM and prototypical contrast, which explicitly decouple feature manifolds and sever spurious noise backdoors.

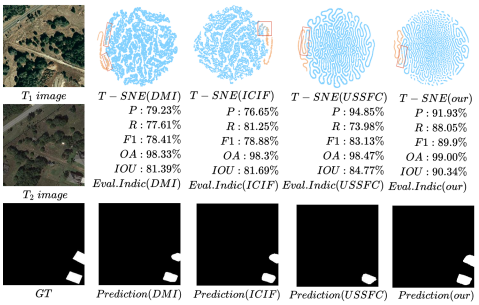

Figure 5: Feature-space separability comparison on LEVIR-CD. t-SNE embeddings of changed and unchanged pixels obtained by DMINet, ICIFNet, USSFC-Net, and UCDNet.

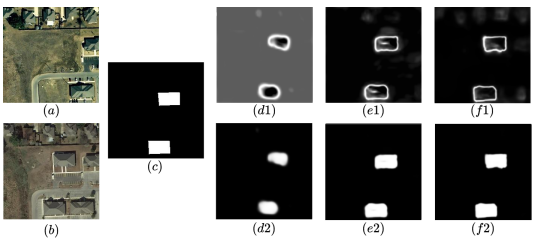

Figure 6: Uncertainty evolution across training epochs: (a) T_1; (b) T_2; (c) GT; (d1)–(f1) Uncertainty maps at epochs 10, 30, and 90; (d2)–(f2) Corresponding predicted change maps.

5.3.2 VISUALIZATION OF THE IMPACT OF UNCERTAINTY

Uncertainty Visualization: As illustrated in Fig.6, high-uncertainty regions strongly correlate with predictive errors across training epochs. In converged predictions, elevated uncertainty sharply localizes along ambiguous change boundaries, which aligns with human perceptual ambiguity in delineating transition zones.

5.3.3 FEATURE EXTRACTION VISUALIZATION

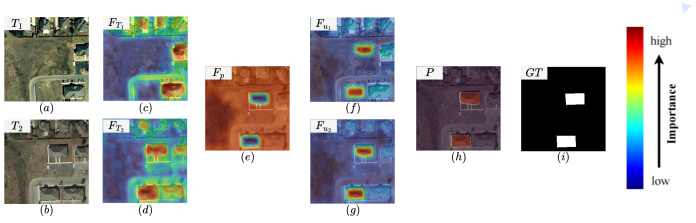

Figure 7: Visualization of key modules in UCDNet: (a) Pre-temporal image T1; (b) Post-temporal image T2; (c) (d) The features F_{T_i} of T_i; (e) Decoupled public feature F_p; (f) (g) Decoupled unique feature F_{u_i}; (h) Final change probability map; (i) Ground Truth

To examine stage-wise feature learning, representative LEVIR-CD samples are visualized in Fig.7. Key observations include: Both initial and disentangled features consistently focus attention on target regions (buildings); Shared feature $\boldsymbol{F_p}$ captures invariant backgrounds, while private features focus on phase-specific changes, particularly$\boldsymbol{F_{u_2}}$; In $\boldsymbol{T_2}$, shared feature $\boldsymbol{F_p}$ and initial feature $\boldsymbol{F_{T_2}}$exhibit similar responses on unchanged structures, allowing difference decomposition to isolate newly added buildings; in $\boldsymbol{T_1}$, initial feature $\boldsymbol{F_{T_1}}$ retains stronger responses than $\boldsymbol{F_p}$, enabling disentangled $\boldsymbol{F_{u_1}}$ to preserve pre-change cues. In summary, our framework achieves clear feature separation while maintaining sharp, target-centric attention. We evaluate feature separability via the

Table 7: Objective evaluation about decoupled feature discriminability

Methods	ICIFNet	DMINet	USSFC-Net	**Ours**
SC	0.6983	0.7052	0.7398	**0.9009**
FDR	2.4892	2.4301	3.6641	**4.7372**

Silhouette Coefficient (SC) and Fisher Discriminant Ratio (FDR). Table 7 shows our method attains the best values on both, confirming compact intra-class and well-separated inter-class features, in line with the qualitative results.

6 CONCLUSION

This paper proposes an Uncertainty-Guided Causal Disentangled Network for remote sensing change detection. Addressing the entangled representations, implicit change modeling, and uncalibrated risks in conventional methods, we first formalize an SCM to design a Pearson correlation-guided feature disentanglement, theoretically severing background confounding backdoors. Furthermore, a maximum-entropy prototypical contrastive learning framework is introduced to capture multimodal land-cover sub-distributions and sharpen decision boundaries. Finally, integrating Subjective Logic with multi-stage Dirichlet evidence accumulation enables joint pixel-level change detection and reliable epistemic uncertainty quantification. Extensive experiments demonstrate that our framework achieves state-of-the-art performance in detection accuracy, noise/illumination robustness, and interpretability.

AI USE STATEMENT

The authors did not use generative AI tools in the preparation of this manuscript.

REFERENCES

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, pp. 207–210, 2022. doi: 10.1109/IGARSS46834.2022.9883686.

Wentao Bao, Qi Yu, and Yu Kong. Evidential deep learning for open set action recognition. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13329–13338, 2021. doi: 10.1109/ICCV48922.2021.01310.

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*, 62:1–20, 2024a.

Jie Chen, Jingru Zhu, Peien He, Ya Guo, Liang Hong, Yin Yang, Min Deng, and Geng Sun. Unsupervised domain adaptation for building extraction of high-resolution remote sensing imagery based on decoupling style and semantic features. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–17, 2024b. doi: 10.1109/TGRS.2024.3387491.

Yongqi Chen, Shou Feng, Chunhui Zhao, Yingjie Tang, Wei Li, and Ran Tao. A prototype-aware learning and dual-view regularization network for weakly supervised change detection in vhr remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–15, 2025. doi: 10.1109/TGRS.2025.3608817.

Anjin Dai, Jianyu Yang, Tingting Zhang, Bingbo Gao, Kaixuan Tang, and Xinyue Chen. Dadr-hcd: A deep domain adaptation and disentangled representation network for unsupervised heterogeneous change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024. doi: 10.1109/TGRS.2024.3403727.

Yao Ding, Zhili Zhang, Aitao Yang, Yaoming Cai, Xiongwu Xiao, Danfeng Hong, and Junsong Yuan. Slcgc: A lightweight self-supervised low-pass contrastive graph clustering network for hyperspectral images. *IEEE Transactions on Multimedia*, 27:8251–8262, 2025. doi: 10.1109/TMM.2025.3604954.

Shou Feng, Tianyu Lan, Yuanze Fan, Mengmeng Zhang, Chunhui Zhao, Wei Li, and Ran Tao. An adaptive weighted metric learning network based on fractional domain decoupling for hyperspectral change detection. *IEEE Transactions on Neural Networks and Learning Systems*, 36(10): 17856–17870, 2025. doi: 10.1109/TNNLS.2025.3586714.

Y. Feng, H. Xu, J. Jiang, H. Liu, and J. 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.

Y. Feng, J. Jiang, H. Xu, and J. Zheng. Change detection on remote sensing images using dual-branch multilevel intertemporal network. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2023.

T. Lei, X. Geng, H. Ning, Z. Lv, M. Gong, Y. Jin, and A. 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.

Kaiwen Li, Hangzhou He, Shuang Zeng, Xinliang Zhang, Yuanwei Li, Lei Zhu, and Yanye Lu. Points-supervised fundus vessel segmentation via shape priors and contrastive learning. *IEEE Transactions on Medical Imaging*, 44(12):5038–5050, 2025. doi: 10.1109/TMI.2025.3586692.

Yang Liu, Xinshuo Wang, Xinbo Gao, Jungong Han, and Ling Shao. Multi-level contextual prototype modulation for compositional zero-shot learning. *IEEE Transactions on Image Processing*, 34:4856–4868, 2025. doi: 10.1109/TIP.2025.3592560.

Yunze Liu, Qingnan Fan, Shanghang Zhang, Hao Dong, Thomas Funkhouser, and Li Yi. Contrastive multimodal fusion with tupleinfonce. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 734–743, 2021. doi: 10.1109/ICCV48922.2021.00079.

Wei Lu, Si-Bao Chen, Qing-Ling Shu, Jin Tang, and Bin Luo. Decouplenet: A lightweight backbone network with efficient feature decoupling for remote sensing visual tasks. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–13, 2024. doi: 10.1109/TGRS.2024.3465496.

Zhiyong Lv, Ming Zhang, Weiwei Sun, Tao Lei, Jón Atli Benediktsson, and Tongfei Liu. Land cover change detection with hyperspectral remote sensing images: A survey. *Information Fusion*, 123:103257, 2025. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2025.103257. URL https://www.sciencedirect.com/science/article/pii/S1566253525003306.

Zhiyong Lv, Sizhe Cheng, Linfu Xie, Junhuai Li, and Minghua Zhao. A graph contrastive learning network for change detection with heterogeneous remote sensing images. *Pattern Recognition*, 172:112394, 2026. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2025.112394. URL https://www.sciencedirect.com/science/article/pii/S0031320325010556.

Shijie Ma, Fei Zhu, Xu-Yao Zhang, and Cheng-Lin Liu. Protogcd: Unified and unbiased prototype learning for generalized category discovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(7):6022–6038, 2025. doi: 10.1109/TPAMI.2025.3557502.

J. Qu, W. Dong, Y. Yang, T. Zhang, Y. Li, and Q. Du. Cycle-refined multidecision joint alignment network for unsupervised domain adaptive hyperspectral change detection. *IEEE Transactions on Neural Networks and Learning Systems*, 36(2):2634–2647, 2025.

Pengyu Tian, Hao Zhu, Biao Hou, Lu Chen, Pute Guo, Kefan Chen, and Licheng Jiao. Sample-level improved cross-source contrastive learning for pan and ms joint classification. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–13, 2025. doi: 10.1109/TGRS.2025.3574030.

B. Wang, X. Jiang, P. Qin, and J. 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.

Bin Wang, Sichao Qiao, Ziyv Zhao, Pinle Qin, and Jianchao Zeng. Uafdcd: Uncertainty-aware feature decoupling network for remote sensing change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 64:5625815–5625815, 2026a. doi: 10.1109/TGRS.2026.3701174.

Bin Wang, Bo Wang, Pinle Qin, and Jianchao Zeng. Fcddnet: Feature cross-domain decoupling network for remote sensing change detection. *Pattern Recognition*, 177:113259, 2026b. doi: 10.1016/j.patcog.2026.113259.

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. doi: 10.1109/TIP.2025.3635479.

Caixia Yan, Xiaojun Chang, Minnan Luo, Huan Liu, Xiaoqin Zhang, and Qinghua Zheng. Semantics-guided contrastive network for zero-shot object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(3):1530–1544, 2024. doi: 10.1109/TPAMI.2021.3140070.

Feng Yang, Mengtao Li, Wenqiang Shu, Anyong Qin, Tiecheng Song, Chenqiang Gao, and Gui-Song Xia. Convformer-cd: Hybrid cnn-transformer with temporal attention for detecting changes in remote sensing imagery. *IEEE Trans. Geosci. Remote. Sens.*, 63:1–15, 2025. doi: 10.1109/TGRS.2025.3544651. URL https://doi.org/10.1109/TGRS.2025.3544651.

Renping Yu, Chao Peng, Jixiang Zhu, Mingming Chen, and Rui Zhang. Weighted multi-modal contrastive learning based hybrid network for alzheimer's disease diagnosis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 33:1135–1144, 2025. doi: 10.1109/TNSRE.2025.3549730.

Mingwei Zhang, Yuan Jiang, and Qi Wang. Enhancing vmamba for change detection via lightweight feature interaction and selection. *Pattern Recognition*, 172:112648, 2026. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2025.112648. URL https://www.sciencedirect.com/science/article/pii/S0031320325013111.

Xiangyue Zhang, Yifan Jia, Jiaxu Zhang, Yijie Yang, and Zhigang Tu. Robust 2d skeleton action recognition via decoupling and distilling 3d latent features. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(10):10410–10422, 2025. doi: 10.1109/TCSVT.2025.3562877.

Jingyu Zhao, Jiahui Qu, and Wenqian Dong. Bi-diffcd: Bidirectional diffusion guided collaborative change detection for arbitrary-modal remote sensing images. In James Kwok (ed.), *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pp. 2449–2457. International Joint Conferences on Artificial Intelligence Organization, 8 2025. doi: 10.24963/ijcai.2025/273. URL https://doi.org/10.24963/ijcai.2025/273. Main Track.

Wenzhao Zheng, Jiwen Lu, and Jie Zhou. Deep metric learning with adaptively composite dynamic constraints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7):8265–8283, 2023. doi: 10.1109/TPAMI.2023.3234536.