

# Causal Decoupling Domain Generalization for Remote Sensing Change Detection

Jiaqi Zhao<sup>1,2</sup>, Jianpeng Xie<sup>1,2</sup>, Yong Zhou<sup>1,2,\*</sup>, Wen-Liang Du<sup>1,2</sup>, Hancheng Zhu<sup>1,2</sup>, Rui Yao<sup>1,2</sup>

<sup>1</sup>School of Computer Science and Technology / School of Artificial Intelligence, China University of Mining and Technology, Xuzhou 221116, China

<sup>2</sup>Mine Digitization Engineering Research Center of the Ministry of Education, Xuzhou 221116, China  
{jiaqizhao, jianpengxie, yzhou, wldu, zhuhancheng, ruiyao}@cumt.edu.cn

## Abstract

While current state-of-the-art Remote Sensing Change Detection (RSCD) methods can achieve impressive results on individual datasets, they become unreliable in unseen environments and imaging conditions, with performance metrics declining by as much as 60% to 80%. Simultaneously, variable environments and complex imaging conditions are the main characteristics of remote sensing data, calling for generalizable RSCD methods. To address this issue, we propose a novel RSCD method capable of domain generalization—CDDGNet. This method is based on causal decoupling theory, which progressively decouples invariant change features from variable domain features to extract generalizable characteristics. This enables a network trained on a single domain to accurately identify change regions in other domains. Specifically, firstly, the Causal Feature Adaptation Module is proposed to preliminarily decouple and simplify feature information during the encoding process by using wavelet transformation and feature energy spectralization methods. Secondly, the Causal Feature Fusion Module is presented to fully decouple features and aggregate significant change features during the decoding process through frequency domain processing and feature re-attention mechanisms. Thirdly, the Decoupling Effect Loss Function is proposed to optimize the process by evaluating the effectiveness of causal decoupling. Extensive experiments have shown that our model significantly outperforms existing methods across multiple groups of generalization tasks with varying levels of difficulty.

**Code** — <https://github.com/M2WindRunner/CDDGNet>

## Introduction

The process of identifying changes in the condition of objects or phenomena of interest over time using remote sensing techniques is known as Remote Sensing Change Detection (RSCD) (Tewkesbury et al. 2015). This technique plays a significant role in various scenarios, including disaster assessment (Bazila and Ankush 2024), agricultural management (Liu et al. 2022), urban planning (Yu et al. 2024), and environmental monitoring (Ren et al. 2024). In remote sensing data across various application scenarios, and even within the same scenario, there exist complex and

\*Corresponding author.

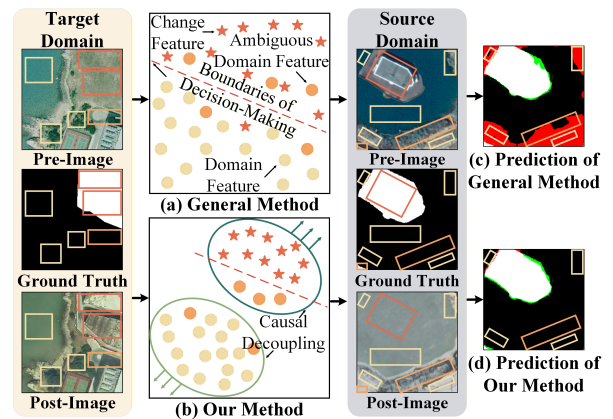


Figure 1: In (a) and (b), the general methods determine the decision boundary (red) using only the distribution of features in the source domain. By contrast, our method effectively decouples change features from domain features. In the source and target domain illustrations, different colored boxes denote the feature types as in (a) and (b). In (c) and (d), white indicates true positives, black true negatives, red false positives, and green false negatives.

unpredictable environmental conditions, as well as diverse and non-standardizable imaging conditions. These objectively existing domain-related characteristics significantly contribute to confusion in RSCD networks.

Recent methods have made significant progress in single-scene applications by improving model architecture and optimizing feature processing techniques. However, in practical applications, the diversity and complexity of scenarios, combined with the high cost of annotating the RSCD dataset, make it impractical to encompass all possible scenarios in the training sets. Consequently, RSCD networks must exhibit robust generalization capabilities to handle variations in any real-world scenario. Unfortunately, current RSCD methods demonstrate inadequate generalization performance in the target domain with different scenes after being trained on the source domain. As shown in Fig. 1 (a) and (c), general methods that directly extract features from original images are heavily reliant on the complexity of the train-

ing dataset for their generalization ability. When faced with unseen scenes, they misinterpret certain domain features as change features, causing a sharp drop in performance.

To address this issue, we propose a new RSCD method called the Causal Decoupling Domain Generalization Network (CDDGNet). Unlike existing domain generalization methods that rely on pre-trained models such as CLIP (Radford et al. 2021) or SAM (Kirillov et al. 2023), CDDGNet learns to directly decouple change features associated with change objects from domain features related to the scene, as shown in Fig. 1 (b) and (d). In RSCD tasks, although scenes are complex and diverse, the changed objects tend to remain consistent across different scenes. Therefore, a network that effectively decouples change features from domain features for the same changed object in the source domain can be directly transferred to the target domain.

Specifically, the CDDGNet network architecture is designed based on the theory of causal decoupling. During the encoding stage, the preliminary features extracted by the backbone are input into a plug-and-play Causal Feature Adaptation Module (CFAM). CFAM uses wavelet transforms to decompose features into high-frequency components, which primarily describe the edges and textures of changed objects, and low-frequency components, which mainly convey scene information and global context. Subsequently, within the same frequency domain, it combines energy spectrum features and spatial features for feature refinement and extraction, achieving simplification of feature information and preliminary feature decoupling. In the decoding stage, the Causal Feature Fusion Module (CFFM) is proposed to fully decouple change features from domain features while fusing change features at various scales. CFFM further decouples features using wavelet transforms and performs feature re-attention operations on low-frequency components, utilizing boundary information from changed objects in high-frequency components, thereby diminishing the influence of domain features on detection results. Additionally, the Decoupling Effect Loss Function (DELFF) measures CDDGNet’s ability to distinguish between ambiguous domain features, thereby enhancing the overall efficiency of the causal decoupling process. Experiments conducted on six sets of domain-generalized RSCD tasks of varying difficulty prove that our method achieves state-of-the-art performance. Our contributions can be summarized as follows:

- We propose a new RSCD domain generalization paradigm called CDDGNet. As the first method developed for this task, it is based on causal decoupling theory and designs a new network architecture that greatly improves the method’s generalization capabilities.
- A plug-and-play Causal Feature Adaptation Module is proposed, which employs wavelet transforms and feature energy spectralization techniques to simplify feature representations while initially decoupling change features from domain features.
- The Causal Feature Fusion Module is designed to fully decouple change features from domain features while effectively aggregating multi-scale features by using frequency domain-specific feature processing methods and

the feature re-attention mechanism.

- We designed six sets of RSCD domain generalization experiments using seven public datasets with varying levels of difficulty, and our method achieved the best results in each set of experiments.

## Related Works

### CD Methods using Deep Learning

FCN (Cayé Daudt, Le Saux, and Boulch 2018) was the first to introduce deep learning to RSCD. Since then, researchers have aimed to improve feature extraction by leveraging contextual relationships (Fang et al. 2022). BIT (Chen, Qi, and Shi 2022) and MambaCD (Chen et al. 2024a) applied Transformer and Mamba models to RSCD, respectively, to explore additional contextual modeling approaches. Moreover, FIMP (Chen et al. 2024c) and WS-Net (Xiong et al. 2024) employ frequency-domain techniques—specifically Fourier and wavelet transforms—to mitigate pseudo-change features. ST-Mamba (Zhao et al. 2025b) proposes a novel state-space model that unifies background information from multi-temporal images within a single domain. ChangeCLIP (Dong et al. 2024) introduces the CLIP framework into RSCD, fusing multi-modal information to detect change regions. Although current RSCD methods have advanced in integrating contextual information and improving accuracy, they often struggle with target-domain data that differ substantially from source-domain scenes. In such cases, existing methods frequently misclassify unseen domain features as changes, leading to degraded detection performance.

### Methods using Causal Inference

To enhance robustness and generalization, methods that combine causal learning with deep learning have emerged (Zhao et al. 2025a). DQN (Zhu, Yu, and Zhang 2023) reweights and resamples data in real time to improve performance. ConvNeXt (Li et al. 2023) analyzes non-causal relationships by intervening in clutter and evaluating parameter comparability, thereby clarifying factors contributing to overfitting. CausalCD (Li et al. 2024) constructs a structural causal model to formalize self-supervised SAR change detection, linking data to labels. However, current causal-correlated RSCD methods are limited to single-domain SAR images and do not meet the cross-scenario demands of domain-generalization tasks.

## Methodology

### Overview

The architecture of CDDGNet is shown in Fig. 2 (a). In the encoding process, ResNet50 has been selected as the backbone network. For each layer of extracted features, the Causal Feature Adaptation Module (CFAM) is introduced to initially decouple the change features from the domain features. Subsequently, during the decoding process, the Causal Feature Fusion Module (CFFM) is designed to further decouple the features while fusing multi-scale features in the frequency domain. Finally, we develop the Decoupling Effect Loss Function (DELFF) to evaluate the effectiveness of

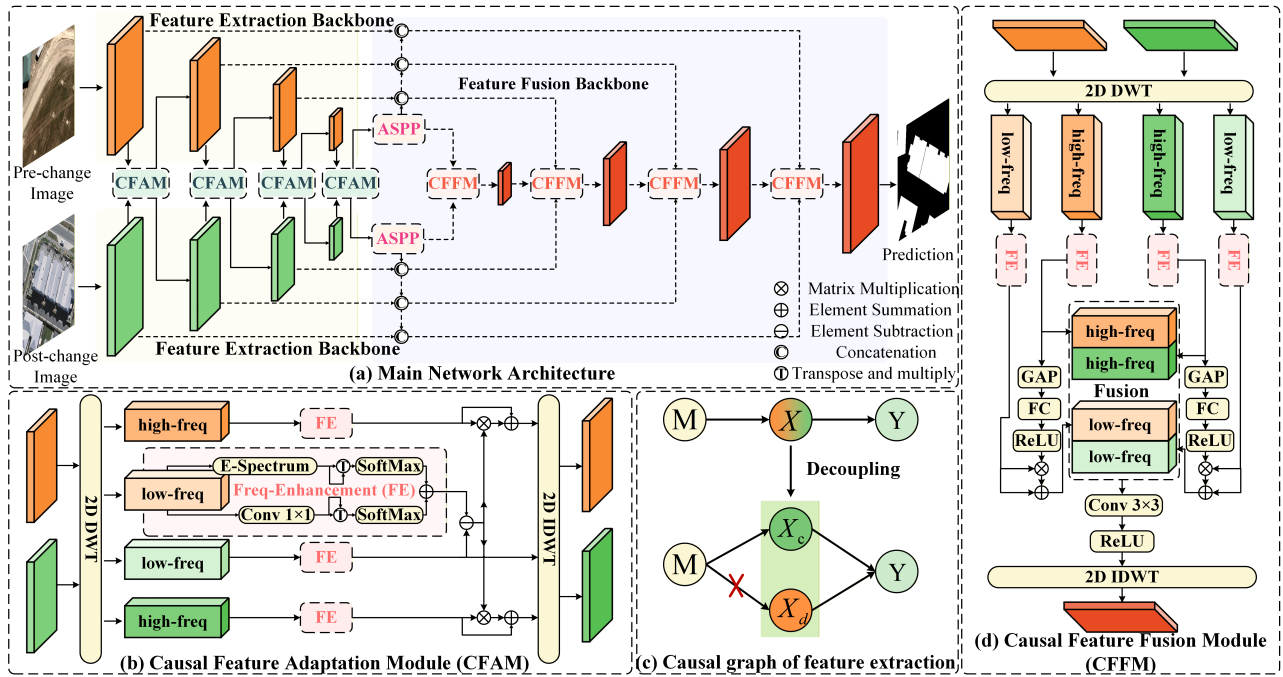


Figure 2: Illustration of our method. (a) presents an overview of the CDDGNet framework. Given pre- and post-change images, feature extraction is conducted using ResNet50 as the backbone. Initial feature decoupling and change feature enhancement are performed by CFAM. Then the features are fused by CFFM in different frequency components, allowing the change features to be gradually decoupled from the domain features in a stepwise manner. (b) illustrates the structural diagram of CFAM. (c) depicts the causal graph of feature extraction. (d) presents the structural graph of CFFM.

the decoupling process, as well as to adjust and enhance subsequent feature decoupling operations.

In the following, we will analyze the reasonableness of the causal decoupling method presented in this paper based on the theory of causal inference. Current RSCD methods tend to extract image features  $X$  directly from a multi-temporal image  $M$  to obtain the predicted label  $Y$ . Their causal map  $M \rightarrow X \rightarrow Y$  is illustrated in Fig. 2 (c). In a single scene, this causality is robust because the feature distribution remains essentially unchanged. However, there are significant differences between the target and source domains regarding environmental and imaging conditions. The change features learned directly from the source domain are overly reliant on domain features for accurate judgment. Therefore, this results in the RSCD network’s inability to accurately predict the target domain label  $Y$  based on the image features  $X$  learned from the source domain image  $M$  in domain generalization tasks. To address this limitation, we decouple the image feature  $X$  into the change feature  $X_c$  and the domain feature  $X_d$ . The decoupling process is illustrated as follows:

$$\begin{aligned}
 P(Y = y \mid do(M = m)) &= P(Y = y \mid M = m) \\
 &= \sum_{x \in x_c, x_d} P(Y = y \mid M = m, X = x)P(X = x \mid M = m) \\
 &= \sum_{x \in x_c, x_d} P(Y = y \mid M = m, X = x)P(X = x),
 \end{aligned} \tag{1}$$

where  $x_c$  denotes the change feature, while  $x_d$  denotes the domain feature. When these features are fully decoupled, the causal graph becomes  $M \rightarrow X_c \rightarrow Y$  and  $M \rightarrow X_d \rightarrow Y$ . This process is equivalent to training the network to distinguish between  $x_c$  and  $x_d$ , as shown in Fig. 2 (c). Upon achieving complete decoupling, the influence of domain features on the predicted label  $Y$  can be eliminated, resulting in a causal graph for the feature extraction process of  $M \rightarrow X_c \rightarrow Y$ . This approach enables accurate detection of change regions in the target domain without being affected by scene changes.

### Causal Feature Adaptation Module

To initially decouple invariant change features from variable and complex domain features, we designed a plug-and-play Causal Feature Adaptation Module (CFAM). This module uses discrete wavelet transform (DWT) to decouple features in the frequency domain and simplify the the associated features, as shown in Fig. 2 (b).

In CFAM, to ensure the continuity of texture features in changing objects, the four frequency domain components generated by DWT are first merged into two components: high-frequency and low-frequency. The high-frequency component primarily includes the contour texture and positional distribution of the changing objects, while the low-frequency component mainly reflects the background environment and the distribution of global objects. Therefore, the initial decoupling process requires separate pro-

cessing of these frequency components. The Frequency Enhancement (FE) method is designed to simplify feature information without losing data. FE refines features by analyzing the energy spectrum and positional correlation of frequency domain features, thereby enhancing the ability to capture both global and local features.

As shown in Fig. 2 (b), the input frequency domain feature maps  $F_{fc}^i \in \mathbb{R}^{C \times H \times W}$ , where  $i = 1, 2$  denotes the pre-change and post-change feature maps, respectively. The variable  $fc \in \{\text{high}, \text{low}\}$  represents the high-frequency and low-frequency components, respectively. Here,  $C$ ,  $H$ , and  $W$  denote the number of channels, height, and width, respectively. The energy spectrum matrix  $E(c, h, w)$  can be described as follows:

$$E(c, h, w) = |F_{fc}^i(c, h, w)|^2, \quad (2)$$

where  $E(c, h, w)$  represents the energy value at channel  $c$  and position  $(h, w)$ , the normalized energy spectrum matrix  $M_{es} \in \mathbb{R}^{C \times H \times W}$  can be constructed as follows:

$$M_{es}(c, h, w) = \frac{E(c, h, w)}{\sum_{c=1}^C \sum_{h=1}^H \sum_{w=1}^W E(c, h, w)}. \quad (3)$$

Subsequently, the correlation of the positional energy spectrum is calculated as follows:

$$R_{es}(h, w) = \text{Softmax} \left( \sum_{c=1}^C M_{es}(c, h) \cdot M_{es}(c, w) \right), \quad (4)$$

where  $R_{es} \in \mathbb{R}^{HW \times HW}$  is the location energy spectrum correlation matrix. Similarly, the bottom branch computes the self-attention weights of each location feature  $R_{fp} \in \mathbb{R}^{HW \times HW}$ . By performing a weighted summation followed by fusion, we can obtain the output of the FE as follows:

$$R_{fe} = \alpha R_{es} + (1 - \alpha) R_{fp}, \quad (5)$$

where  $\alpha$  is the weighting factor (set to 0.5 in the experiment). FE can highlight details and edges in changing objects by analyzing energy spectrum features, and can capture the global distribution of features by computing spatial self-attention weights.

However, the edge contour of the changed object in the enhanced high-frequency component is not entirely continuous, indicating that some relevant features of the changed object remain embedded in the low-frequency component. Therefore, we extract the global change contour from the low-frequency component by calculating the absolute value difference, which integrates the positional information of the changed object, to enhance the change contour in the high-frequency component, as shown below:

$$\theta = |FE(\{F_{low}^1\}^{in}) - FE(\{F_{low}^2\}^{in})|, \quad (6)$$

$$\{F_{high}^i\}^{out} = FE(\{F_{high}^i\}^{in}) + FE(\{F_{high}^i\}^{in}) \times \theta, \quad (7)$$

where  $FE(\cdot)$  represents the result after FE processing. Finally, all frequency components are passed through an inverse discrete wavelet transform (IDWT) and then fed into the subsequent feature extraction layer.

## Causal Feature Fusion Module

Since CFAM must ensure the complete extraction of change features, it is impossible to fully decouple these features during the encoding stage. Therefore, the Causal Feature Fusion Module (CFFM) is proposed to compose the decoder, enabling it to fully decouple change features from domain features while integrating multi-scale features.

Firstly, we upsample the smaller feature maps from the subsequent encoding layers and superimpose them onto the input of each decoding layer to expand the receptive field of the CFFM. For example, the input to the last decoder layer is illustrated below:

$$F_1^i = \text{Concat}(F_1^i, F_2^i, F_3^i, F_4^i), \quad (8)$$

where  $F_j^i$  denotes the feature map output from the  $j$ -th encoding layer, and  $\text{Concat}(\cdot)$  represents the concatenation operation on the feature map. The ASPP module (Chen et al. 2017) is used to enhance the first-layer decoder's ability to capture global information.

Secondly, we use DWT and FE to further decouple change features from domain features in CFFM. Unlike CFAM, which utilizes low-frequency components to enhance the completeness of change features in high-frequency components, we leverage change features in high-frequency components to guide low-frequency components to focus more on changed objects rather than domain-related features, as shown in Fig. 2 (d). The specific steps are as follows:

$$\beta = \text{ReLU}(\text{FC}(\text{GAP}(\text{FE}(F_{high}^i))))), \quad (9)$$

$$F_{low}^i = FE(F_{low}^i) + \beta \times FE(F_{low}^i), \quad (10)$$

where  $FC(\cdot)$  denotes the fully connected layer, and  $\text{GAP}(\cdot)$  denotes Global Average Pooling. This feature re-attention operation, which targets low-frequency components, can effectively reduce the influence of domain features on the extraction of change features. Subsequently, the different frequency domain components are aggregated separately to complete the final feature decoupling. Finally, the resulting feature map is passed through a  $3 \times 3$  convolutional layer followed by a ReLU activation function. After applying the IDWT, the output of the CFFM is obtained.

## Decoupling Effect Loss Function

The current loss function is calculated based on the difference between the predicted labels and the ground truth (GT) labels. However, this detection result only provides a broad indication of the network's performance and does not directly measure the specific effect of causal decoupling. To facilitate more effective iteration of the causal decoupling process, the Decoupling Effect Loss Function (DELFF) has been designed. DELFF evaluates the decoupling capability of CDDGNet by assessing its ability to handle ambiguous domain features that are challenging to decouple.

Specifically, because ambiguous domain features resemble change features and differ significantly from other domain features, they present a key challenge in the decoupling process. Additionally, since GT can identify the re-

Methods	L to W	L+ to W	S to C	C to S	C to Y	C to G
	OA/F1/IoU	OA/F1/IoU	OA/F1/IoU	OA/F1/IoU	OA/F1/IoU	OA/F1/IoU
SNUNet	97.02/71.13/55.19	97.69/73.34/57.90	59.82/31.79/18.89	77.86/35.52/21.60	86.77/29.55/17.33	71.27/17.43/9.55
BIT	97.51/71.25/55.34	97.24/71.98/56.23	59.31/30.75/18.16	77.12/36.46/22.29	<b>86.92/39.12/24.31</b>	<b>81.78/15.22/8.23</b>
P2V	96.47/64.51/47.61	97.12/67.62/51.08	57.91/30.70/18.14	78.73/39.26/24.43	72.66/35.33/21.45	68.94/ <b>21.80/12.23</b>
SGSLN	97.60/72.95/57.42	97.78/74.15/58.92	62.48/32.73/19.57	64.79/43.69/27.95	85.45/33.10/19.83	68.33/20.58/11.47
MambaCD	97.08/69.14/52.84	97.38/70.72/54.70	62.37/30.41/17.93	73.15/43.46/27.76	85.32/29.74/17.47	76.14/17.03/9.31
RS-Mamba	97.32/70.15/54.03	97.62/71.42/55.54	61.96/31.83/18.92	<u>79.24/48.20/31.76</u>	68.64/29.65/17.41	70.65/17.12/9.37
FIMP	97.56/72.56/56.93	97.79/74.76/59.70	<b>64.53/35.58/21.64</b>	<b>79.44/36.04/21.98</b>	80.20/11.82/6.28	<u>76.23/17.57/9.63</u>
WS-Net	97.42/73.64/58.28	97.62/74.70/59.62	71.81/34.50/20.84	78.39/42.38/26.89	81.11/28.91/16.90	74.24/17.39/9.53
ST-Mamba	<u>97.72/73.63/58.27</u>	<u>97.87/77.19/62.85</u>	62.30/32.36/19.31	79.12/47.70/31.33	73.37/31.14/18.44	72.05/ <u>20.82/11.62</u>
ChangeClip	<b>97.95/75.15/60.19</b>	<b>97.98/76.10/61.42</b>	62.77/32.11/19.12	79.18/28.56/16.66	78.43/5.36/2.75	69.02/5.30/2.70
CDDGNet	<b>98.21/79.47/65.94</b>	<b>98.52/82.99/70.93</b>	<b>89.36/60.77/43.65</b>	<b>84.75/68.40/51.98</b>	<b>89.93/65.21/48.38</b>	<b>90.26/46.35/30.17</b>

Table 1: Comparison of CDDGNet with respect to SOTA RSCD methods. The best result is both bolded and underlined, the second-best result is bolded, and the third-best result is underlined.

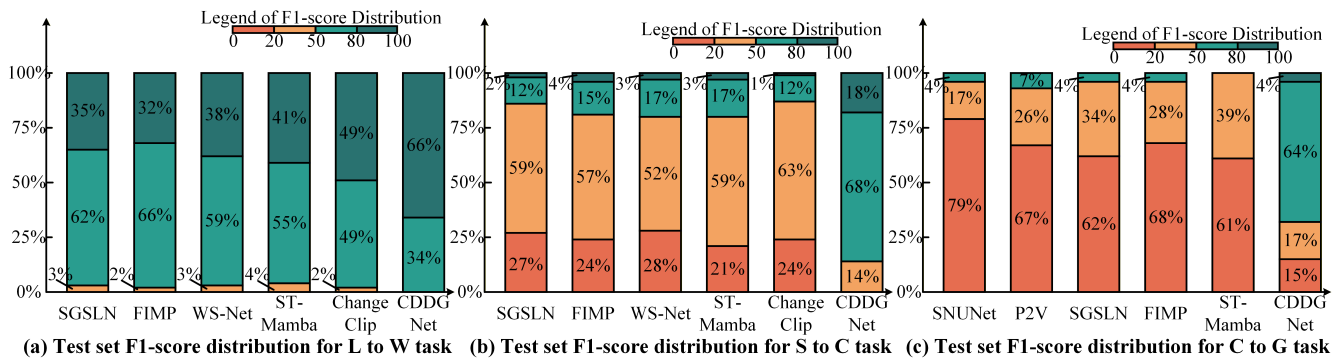


Figure 3: This figure shows the proportion of detection results within each F1 score segment relative to the total number of results on the test set for each method. The figure presents one group from each of the three difficulty-level tasks and displays the proportions for the top six methods based on F1 score for each task. Each number is rounded to the nearest whole number.

gions where change features are located, we combine clustering algorithms to classify the features of misclassified regions, the surrounding change regions (change features), and the invariant regions (domain features) along specific dimensions, and measure the  $L_2$  distance between them.

If the domain features are confirmed to be ambiguous, the DELF result is obtained by calculating the average difference in  $P_b$  among the surrounding domain features:

$$\mathcal{L}_{de} = \frac{1}{N} \sum \exp\left(\frac{\|P_b - P_p\|_2}{\tau}\right), \quad (11)$$

where  $N$  represents the number of pixels in ambiguous domain features, and  $\tau$  is the temperature parameter, which we set to 10. When combined with DELF, the total loss function of our method is defined as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{BCE} + \mathcal{L}_{Dice} + \lambda \mathcal{L}_{de}, \quad (12)$$

where  $\mathcal{L}_{BCE}$  represents the classification loss,  $\mathcal{L}_{Dice}$  addresses the sample imbalance between changed and unchanged regions. In our experiments, we set  $\lambda$  to 0.015 to balance each loss function.

## Experiments

### Experimental Settings

**Datasets.** To evaluate the model’s performance in RSCD domain generalization, we chose five high-resolution RSCD datasets: CDD (Lebedev et al. 2018), SYSU (Shi et al. 2021), LEVIR-CD (Chen and Shi 2020), WHU (Ji, Wei, and Lu 2018), and LEVIR-CD+, along with two radar RSCD datasets: Yellow and GBF-CD. We created six RSCD domain generalization tasks: L to W, L+ to W, S to C, C to S, C to Y, and C to G. The initial letters of the dataset are used as abbreviations; for example, L+ to W indicates training on LEVIR-CD+ and validation and testing on WHU. These six experiments are categorized into three difficulty levels: L to W and L+ to W involve a single type of changed object, with minor differences in environment and imaging conditions; S to C and C to S are full-class RSCD tasks with significant environmental differences and moderate imaging condition differences; C to Y and C to G are full-class RSCD tasks with substantial differences in both environment and imaging conditions. The method for dividing the data set follows the procedure described in (Zhao et al. 2023).

**Experimental Setup.** We use the AdamW optimizer

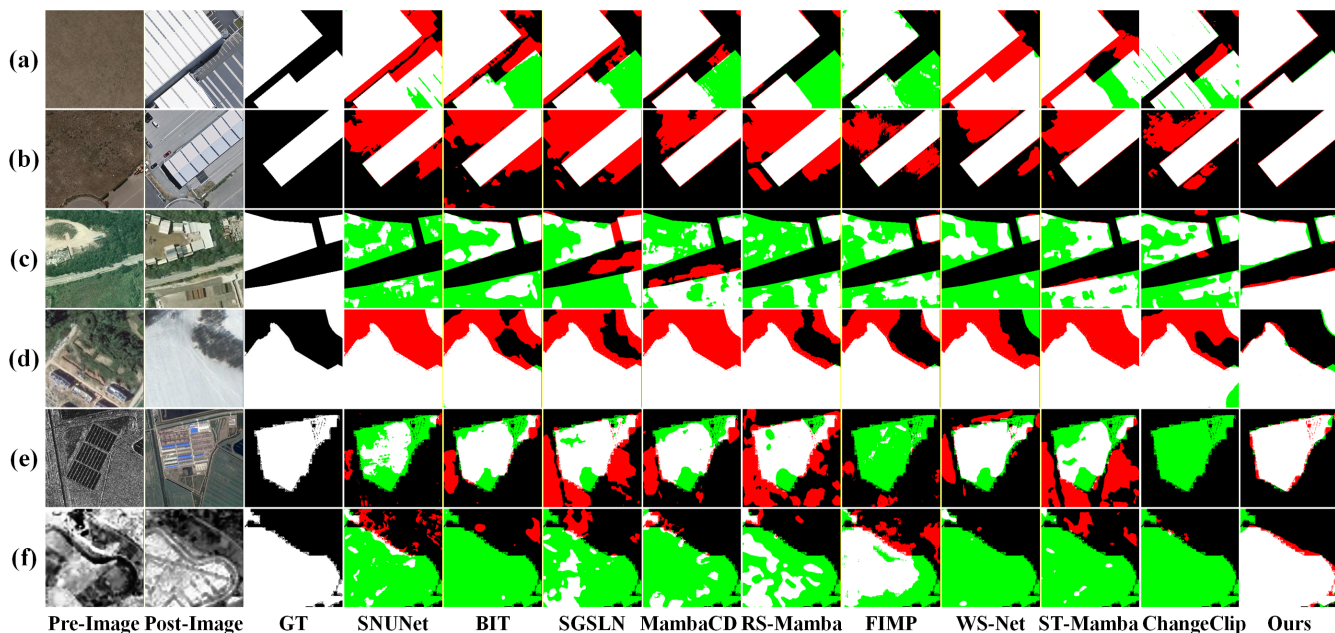


Figure 4: Qualitative results from six sets of experiments. In the result plots for each method, white indicates true positives, black indicates true negatives, green indicates false positives, and red indicates false negatives. (a) From L to W. (b) From L+ to W. (c) From C to S. (d) From S to C. (e) From C to Y. (f) From C to G.

(Loshchilov 2017) with initial learning rate  $1e^{-3}$  and weight decay  $5e^{-3}$ . The batch size during training is kept at 32 and the model is trained with 120 epochs on each dataset. We use PyTorch (Paszke et al. 2019) to implement the CDDGNet and train it on an RTX 3090 GPU (24 GB). Following (Chaminda Bandara, Nair, and Patel 2025), we use three metrics for evaluation: F1 score, Intersection-Over-Union (IoU), and Overall Accuracy (OA).

**Comparison Methods.** We selected ten classic or advanced methods for comparison. SNUNet (Fang et al. 2022) and BIT (Chen, Qi, and Shi 2022) are classic methods. P2V (Lin, Yang, and Zhang 2023) treats multi-temporal images as pseudo-video frames for feature extraction. SGSLN (Zhao et al. 2023) proposes a novel and lightweight RSCD architecture. MambaCD (Chen et al. 2024a), RS-Mamba (Chen et al. 2024b), and ST-Mamba (Zhao et al. 2025b) explore the use of Mamba architecture in RSCD. FIMP (Chen et al. 2024c) and WS-Net (Xiong et al. 2024) explore the use of frequency-domain correlation theory in RSCD. ChangeClip (Dong et al. 2024) is the first to apply CLIP to RSCD.

### Comparisons with Prior Works

**Quantitative results.** Tab. 1 shows the overall performance of all methods across the six sets of RSCD domain generalization experiments we designed. Clearly, our method demonstrates the best performance in each experiment set, and the advantages of CDDGNet become increasingly apparent as the difficulty of the task increases.

In the relatively simple L to W and L+ to W tasks, compared to the second-best method, CDDGNet improved the F1 score by 4.23% and 5.80%, and increased the IoU

score by 5.77% and 8.08%, respectively. Even in these relatively simple domain generalization tasks, our method significantly outperforms others, as clearly shown in Fig. 3 (a). CDDGNet achieved a high-quality detection rate exceeding 66% on the test set without any poor results. By contrast, other methods produced high-quality results below 50%, with some falling below 30%, and all produced a certain number of poor outcomes. In the moderately difficult S to C and C to S tasks, compared to the second-best method, CDDGNet improved the F1 score by 25.19% and 20.20%, and the IoU score by 22.01% and 20.22%, respectively. As shown in Fig. 3 (b), the proportion of qualified detection results produced by other methods on the test set was below 20%, whereas CDDGNet achieved a high rate of 86%. This clearly demonstrates that CDDGNet can accurately identify change regions even when confronted with numerous unknown scenarios. In the most challenging C to Y and C to G tasks, compared to the second-best method, CDDGNet improved the F1 score by 26.09% and 24.55%, and the IoU score by 24.07% and 17.94%, respectively. In this domain generalization task, where the model has virtually no prior knowledge of the environment or imaging conditions, other methods completely lose their ability to detect change regions, with a qualified result rate below 7%, nearly reaching zero. In contrast, CDDGNet achieves a rate exceeding 68%, remaining capable of identifying approximate change regions, as shown in Fig. 3 (c). This comparison highlights the exceptional generalization capability of CDDGNet.

**Qualitative results.** To more intuitively demonstrate the effectiveness of our proposed method, visualization results for each task group are presented in Fig. 4. Specifically, in the

Methods	Params (M)	FLOPs (G)
SNUNet	12.16	27.84
BIT	3.62	10.55
P2V	5.57	34.16
SGSLN	6.29	11.73
MambaCD	85.53	44.83
RS-Mamba	51.35	28.17
FIMP	10.34	20.28
WS-Net	8.87	41.84
ST-Mamba	77.86	40.81
ChangeClip	25.69	42.59
CDDGNet	9.35	35.24

Table 2: This compares the computational costs of different RSCD models, where Params indicates the number of parameters and FLOPs denotes the floating-point operations.

L to W and L+ to W tasks, other methods failed to distinguish between the extracted change features and the easily confused unseen domain features, resulting in large areas of false positives and false negatives. In contrast, CDDGNet is largely unaffected by such simple scene changes and accurately identifies the change regions. In the S to C and C to S tasks, the increased number of unknown scenes causes other methods to misdetect entire changed objects. However, CDDGNet, with its ability to decouple change features from domain features, maintains precise detection of each changed object. In the C to Y and C to G tasks, the situation is similar to the previous set of tasks, with other methods nearly losing their ability to detect change regions. However, CDDGNet, benefiting from its strong generalization capability, effectively delineates the change regions.

**Model efficiency.** To validate the proposed model’s efficiency, Tab. 2 lists the parameters and GFLOPs of each method. CDDGNet and lightweight models (BIT, P2V, SNUNet) use comparable computational resources but outperform ST-Mamba, MambaCD, and RS-Mamba, which have substantially more parameters. This indicates that CDDGNet and these lightweight models strike an effective balance between performance and efficiency.

### Ablation Studies

To validate the effectiveness of each innovative module in CDDGNet, we conducted seven ablation experiments, as shown in Tab. 3. Each experiment used a backbone model only with ResNet50 and the ASPP module. Since DELF is a loss function that assesses and guides feature decoupling, it was not included in the baseline without decoupling. It can be observed that the addition of either CFAM or CFFM alone results in only a limited performance improvement. However, when CFAM or CFFM is combined with DELF, there is a more significant increase in performance. This suggests that DELF effectively evaluates and enhances the decoupling ability of the network. When CFAM and CFFM work together, there is an even more substantial improvement in model performance. This indicates that CFAM and CFFM are well designed, as their combined use can more effectively decouple change features from domain features

CFAM	CFFM	DELF	L to W	S to C	C to G
			69.23	32.76	19.88
✓			73.85	42.12	27.85
	✓		74.29	38.55	25.29
✓		✓	75.61	48.34	35.41
	✓	✓	76.73	43.91	28.36
✓	✓		<u>77.60</u>	<u>55.47</u>	<u>40.94</u>
✓	✓	✓	<b>79.47</b>	<b>60.77</b>	<b>46.35</b>

Table 3: Ablation study of different components on L to W, S to C and C to G task. The best results are highlighted in bold, while the second-best results are underlined.

Methods	L to W	S to C	C to G
SNUNet	74.61(↑ <b>3.48</b> )	36.51(↑ <b>4.72</b> )	21.60(↑ <b>4.17</b> )
BIT	73.85(↑ <b>2.60</b> )	32.70(↑ <b>1.95</b> )	17.98(↑ <b>2.76</b> )
MambaCD	73.17(↑ <b>4.03</b> )	37.92(↑ <b>7.51</b> )	26.73(↑ <b>9.70</b> )
RS-Mamba	74.33(↑ <b>4.18</b> )	37.69(↑ <b>5.86</b> )	27.46(↑ <b>10.34</b> )
WS-Net	76.46(↑ <b>2.82</b> )	40.48(↑ <b>5.98</b> )	20.62(↑ <b>3.23</b> )
ChangeClip	77.08(↑ <b>1.93</b> )	34.37(↑ <b>2.26</b> )	16.11(↑ <b>10.81</b> )

Table 4: Ablation experiments applying CFAM to other methods on L to W, S to C and C to G task.

in a phased manner. In addition, to verify the plug-and-play performance of CFAM, we applied it to the comparison methods, as shown in Tab. 4. Methods not listed were excluded because other modules already occupied the positions where CFAM needed to be placed. It is evident that across the three domain generalization tasks of varying difficulty, adding CFAM improved the methods to different extents, and as the difficulty increased, the F1 scores also gradually improved. This clearly demonstrates the outstanding performance of CFAM as a plug-and-play module.

### Conclusion

To address the challenges posed by environmental changes and varying imaging conditions in real-world scenarios for Remote Sensing Change Detection (RSCD), we propose a novel domain generalization method called CDDGNet. This approach leverages causal inference theory to design a novel network architecture that decouples invariant change features across scenarios from highly variable and diverse domain features, thereby achieving robust generalization capabilities with limited source domain data. CDDGNet comprises three components: the Causal Feature Adaptation Module and the Causal Feature Fusion Module, which work together to decouple change features from domain features while accurately detecting the change regions; and the Decoupling Effect Loss Function, which optimizes model performance by assessing the impact of causal decoupling. We conducted six sets of RSCD domain generalization experiments with varying levels of difficulty using seven publicly available cross-domain datasets. Compared to existing state-of-the-art methods, CDDGNet achieves superior performance across all difficulty levels and demonstrates reliable results even in completely unknown scenarios.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62272461, Grant 62172417, and Grant 62277046; in part by the Double First-Class Project of China University of Mining and Technology for Independent Innovation and Social Service under Grant 2022ZZCX06; in part by the Six Talent Peaks Project in Jiangsu Province under Grant 2015-DZXX-010 and Grant 2018-XYDXX-044.

## References

- Bazila, F.; and Ankush, M. 2024. Satellite-based change detection in multi-objective scenarios: A comprehensive review. *Remote Sensing Applications: Society and Environment*, 34: 101168.
- Caye Daudt, R.; Le Saux, B.; and Boulch, A. 2018. Fully Convolutional Siamese Networks for Change Detection. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, 4063–4067.
- Chaminda Bandara, W. G.; Nair, N. G.; and Patel, V. M. 2025. DDPM-CD: Denoising Diffusion Probabilistic Models as Feature Extractors for Remote Sensing Change Detection. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 5250–5262.
- Chen, H.; Qi, Z.; and Shi, Z. 2022. Remote Sensing Image Change Detection With Transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–14.
- Chen, H.; and Shi, Z. 2020. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sensing*, 12(10): 1662.
- Chen, H.; Song, J.; Han, C.; Xia, J.; and Yokoya, N. 2024a. ChangeMamba: Remote Sensing Change Detection With Spatiotemporal State Space Model. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–20.
- Chen, K.; Chen, B.; Liu, C.; Li, W.; Zou, Z.; and Shi, Z. 2024b. RSMamba: Remote Sensing Image Classification With State Space Model. *IEEE Geoscience and Remote Sensing Letters*, 21: 1–5.
- Chen, L.-C.; Papandreou, G.; Schroff, F.; and Adam, H. 2017. Rethinking Atrous Convolution for Semantic Image Segmentation. [arXiv:1706.05587](https://arxiv.org/abs/1706.05587).
- Chen, Y.; Feng, S.; Zhao, C.; Su, N.; Li, W.; Tao, R.; and Ren, J. 2024c. High-Resolution Remote Sensing Image Change Detection Based on Fourier Feature Interaction and Multiscale Perception. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–15.
- Dong, S.; Wang, L.; Du, B.; and Meng, X. 2024. Change-CLIP: Remote sensing change detection with multimodal vision-language representation learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 208: 53–69.
- Fang, S.; Li, K.; Shao, J.; and Li, Z. 2022. SNUNet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5.
- Ji, S.; Wei, S.; and Lu, M. 2018. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on geoscience and remote sensing*, 57(1): 574–586.
- Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A. C.; Lo, W.-Y.; Dollar, P.; and Girshick, R. 2023. Segment Anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 4015–4026.
- Lebedev, M.; Vizilter, Y. V.; Vygolov, O.; Knyaz, V. A.; and Rubis, A. Y. 2018. Change detection in remote sensing images using conditional adversarial networks. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42: 565–571.
- Li, H.; Zou, B.; Zhang, L.; and Qin, J. 2024. CausalCD: A Causal Graph Contrastive Learning Framework for Self-Supervised SAR Image Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–16.
- Li, W.; Yang, W.; Liu, L.; Zhang, W.; and Liu, Y. 2023. Discovering and Explaining the Noncausality of Deep Learning in SAR ATR. *IEEE Geoscience and Remote Sensing Letters*, 20: 1–5.
- Lin, M.; Yang, G.; and Zhang, H. 2023. Transition Is a Process: Pair-to-Video Change Detection Networks for Very High Resolution Remote Sensing Images. *IEEE Transactions on Image Processing*, 32: 57–71.
- Liu, M.; Chai, Z.; Deng, H.; and Liu, R. 2022. A CNN-Transformer Network With Multiscale Context Aggregation for Fine-Grained Cropland Change Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 4297–4306.
- Loshchilov, I. 2017. Decoupled weight decay regularization. [arXiv preprint arXiv:1711.05101](https://arxiv.org/abs/1711.05101).
- Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.
- Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; Krueger, G.; and Sutskever, I. 2021. Learning Transferable Visual Models From Natural Language Supervision. In Meila, M.; and Zhang, T., eds., *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, 8748–8763. PMLR.
- Ren, H.; Xia, M.; Weng, L.; Hu, K.; and Lin, H. 2024. Dual-Attention-Guided Multiscale Feature Aggregation Network for Remote Sensing Image Change Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17(42075130): 4899–4916.
- Shi, Q.; Liu, M.; Li, S.; Liu, X.; Wang, F.; and Zhang, L. 2021. A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection. *IEEE transactions on geoscience and remote sensing*, 60: 1–16.
- Tewkesbury, A. P.; Comber, A. J.; Tate, N. J.; Lamb, A.; and Fisher, P. F. 2015. A critical synthesis of remotely sensed

optical image change detection techniques. *Remote Sensing of Environment*, 160: 1–14.

Xiong, F.; Li, T.; Yang, Y.; Zhou, J.; Lu, J.; and Qian, Y. 2024. Wavelet Siamese Network With Semi-Supervised Domain Adaptation for Remote Sensing Image Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–13.

Yu, S.; Li, J.; Chen, Z.; Ren, L.; and Hua, Z. 2024. Multiscale Attention Fusion Graph Network for Remote Sensing Building Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–18.

Zhao, J.; Li, Y.; Zhou, Y.; Du, W.-L.; Li, X.; Yao, R.; and Saddik, A. E. 2025a. DDCI: Unsupervised Domain Adaptation for Remote Sensing Images Based on Diffusion Causal Distillation. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1–12.

Zhao, J.; Xie, J.; Zhou, Y.; Du, W.-L.; Yao, R.; and Saddik, A. E. 2025b. ST-Mamba: Spatio-Temporal Synergistic Model for Remote Sensing Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1–13.

Zhao, S.; Zhang, X.; Xiao, P.; and He, G. 2023. Exchanging Dual-Encoder–Decoder: A New Strategy for Change Detection With Semantic Guidance and Spatial Localization. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–16.

Zhu, W.; Yu, C.; and Zhang, Q. 2023. Causal deep reinforcement learning using observational data. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, IJCAI '23. ISBN 978-1-956792-03-4.