# Change-Prior-Guided Unsupervised Change Detection of Heterogeneous Remote Sensing Images

Yuli Sun, Lin Lei, Gangyao Kuang

***Abstract*—Heterogeneous change detection (HeCD) enables the identification of land-cover changes using remote sensing imagery obtained from different sensors. Most existing methods overly emphasize modality transformation and shared feature extraction to bridge the gap between heterogeneous images. While these strategies facilitate comparable representations, they tend to neglect the intrinsic characteristics of the changes themselves, which limits their effectiveness in complex scenarios. To overcome this limitation, we propose a change prior-guided image transformation model (CPIT) for unsupervised HeCD. Specifically, starting from the definition of change detection, we analyze the connections among pairwise object relationships, change labels, and change semantics, and then derive change semantic consistency and inconsistency rules solely from the inherent nature of the change detection problem, without relying on data-specific assumptions. These rules are subsequently encoded as change semantic consistency and inconsistency constraints, which, from the perspective of graph signal processing, correspond to low-pass and high-pass spectral properties of the change signals. Finally, by integrating these semantic constraints with sparsity priors and image transformation constraints, we formulate a more precise transformation model for HeCD. Solving this model produces change detection results that conform to the change priors, thereby improving the detection performance. The derivation, formulation, and utilization of change priors in this work offer valuable insights for broader change detection research. Extensive experiments on five datasets validate the effectiveness of CPIT. The code will be released at https://github.com/yulisun/CPIT.**

***Index Terms*—Change prior, unsupervised change detection, heterogeneous images, image transformation.**

## I. Introduction

### A. Background

Heterogeneous change detection (HeCD) refers to the process of comparing remote sensing imagery acquired by different sensors over the same area to identify land-cover changes [1], [2]. By eliminating the requirement for same-sensor image acquisition, HeCD surpasses the constraints of traditional homogeneous change detection (HoCD) methods and has emerged as a topic of growing interest in remote sensing research [3], [4].

*Manuscript received XXX. (Corresponding author: Y. Sun. Email: sunyuli@mail.ustc.edu.cn)*

Y. Sun, L. Lei and G. Kuang are with the College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073.

This work was supported by the National Natural Science Foundation of China under Grant No. 62401577, the Natural Science Foundation of Hunan Province under Grant No. 2024JJ6466, and the China Postdoctoral Science Foundation under Grant No. 2025T181178.

This paper is accompanied by supplementary material, which provides additional details on spectral analysis, optimization, parameter analysis, spatial-prior integration, and complexity analysis.

HeCD has many significant real-world applications. For example, during emergency events such as earthquakes, floods, and landslides, unfavorable imaging conditions (e.g., poor illumination and adverse weather associated with disasters) and the availability of image resources often restrict the use of homogeneous images [5], [6]. Consequently, change detection typically relies on pre-event optical and post-event SAR images [7]. Additionally, in long-term monitoring, multisensor remote sensing can mitigate the limitations of satellite revisit cycles, thereby improving the temporal resolution of change detection [8], [9].

Despite the considerable practical value of HeCD, it also introduces additional challenges compared to HoCD. In HeCD, multi-temporal images are captured from different sensors (such as optical and SAR sensors), and the variations in their imaging mechanisms and conditions lead to distinct representations of the same ground objects in heterogeneous images [10]. As a result, changes cannot be detected through direct image comparison as in HoCD [11], [12].

### B. Related Work

A key step in HeCD is bridging the modality gap between images by mapping them into a comparable latent space. Therefore, existing methods have primarily focused on achieving image transformation and comparison [13], [14], which can be broadly categorized into four main approaches.

**1) Image classification-based** approaches construct classifiers to independently categorize multi-temporal images, thereby mapping heterogeneous images into a common land-cover category space for comparison. Representative approaches include the Bayesian soft fusion method [15], post-classification inference using multidimensional evidence theory [16], and compound classification based on Markov random field (MRF) [17]. **2) Energy model-based** approaches formulate the correlations between heterogeneous images as energy functions and employ random field-based models to detect changes. Representative approaches include the MRF model based on nonlocal pairwise pixel relationships [18], [19], the circular invariant convolution model [12], the adaptive class-conditional likelihood MRF model [20], and the locality preservation-based energy model [21]. **3) Feature comparison-based** approaches map heterogeneous images into a shared feature space by either manually designing similarity metrics or training feature extraction networks for comparison [22], [23]. Representative approaches include the manifold space metric based on multivariate statistical models [24], the structural comparison method [25], [26], and the commonality feature representation network [3]. **4) Image transformation-based** approaches utilize pixel mapping or style migration to convert images from one domain to another, thereby enabling direct comparison within a unified image domain [27], [28]. Representative approaches include homogeneous pixel translation [29], fractal encoding-decoding [30], structured graph regression [31], code-aligned autoencoders [32], diffusion-based translation networks [33], and fusion-based generative adversarial networks [34]. In addition, some recent methods combine image translation with feature-level modeling to further improve HeCD performance. For example, Li et al. proposed a copula mixture and CycleGAN-based method [35], which first translates the pre-event image into the modality of the post-event image using CycleGAN, and then performs change detection by modeling the joint distribution of image features with a copula mixture.

In summary, existing studies primarily focus on transforming heterogeneous images into a common domain to facilitate difference measurement, thereby tackling the issue of incomparability in HeCD [36], [37]. In addition, during the transformation and comparison processes, some change priors are also used as constraints, mainly including spatial change priors: the spatial sparseness of changes, the spatial smoothness of changes, and the spatial consistency of changes.

**1) Spatial sparseness** of change is based on the observation that, in real-world scenarios, only a small portion of the area undergoes change while the majority remains unchanged [21], [38]. Some image decomposition-based approaches integrate this prior into HeCD by modeling the post-event image as a combination of an unchanged translated image and a distinct changed image, while enforcing sparsity constraints on the changed image [31], [39], [40]. Change sparsity is a foundational concept for many unsupervised HeCD methods, as it treats change detection as an anomaly detection problem.

**2) Spatial smoothness** of change posits that neighboring pixels are spatially correlated and tend to share the same change status. This characteristic is frequently used in post-processing steps for difference image segmentation, where it helps to eliminate isolated changed/unchanged regions, effectively reducing salt-and-pepper noise in the change map [25], [41], [42]. This property is beneficial in both HoCD and HeCD.

**3) Spatial consistency** of change is applicable when multiple detection results are available, such as in cases involving different modalities or temporal variations. It is based on the fact that multiple detection results describing the same change event should exhibit consistent change regions. This consistency is often leveraged for fusing detection results [6], [39], [43], [44], ensuring that the detected change regions are stable and reliable across diverse sources.

### *C. Motivations and Contributions*

Although these aforementioned spatial change priors have significantly influenced HeCD research, they tend to overlook a deeper investigation into the intrinsic nature of change itself. Among the aforementioned change priors, spatial smoothness and spatial consistency are typically applied as post-processing techniques for difference image (DI) fusion and segmentation. As a result, their benefits do not directly contribute to improving the image transformation process or the DI computation. In contrast, only spatial sparsity has been explicitly incorporated into transformation models to constrain the change process, thereby enhancing both image transformation and DI computation [31], [44]. However, relying solely on change sparsity is insufficient for the transformation process of HeCD, particularly in unsupervised transformation-based methods, where no labeled training samples are available. This limitation arises primarily for two reasons.

• **First**, unsupervised HeCD is inherently more challenging due to the absence of labeled data, making it difficult to ensure stable and accurate image transformation. Over-reliance on sparsity as the sole change constraint may lead to model failure, increased sensitivity to parameter settings, and insufficient robustness.

• **Second**, incorporating additional change priors into the transformation model can improve its ability to capture change patterns, leading to more reliable image transformation and DI computation. By leveraging multiple change priors in a unified framework, the model can better characterize the change process and enhance the overall performance of HeCD.

In this paper, we delve into the characteristics of change in HeCD tasks and incorporate them into unsupervised image transformation models. First, starting from the definition of the HeCD problem, we analyze the inherent connections between pairwise objects, binary change labels (change/unchanged) and change semantics, which forms the basis of existing structure regression-based methods. Furthermore, building on these pairwise connections and the intrinsic properties of the HeCD problem, we derive change rules, including semantic consistency and semantic inconsistency, thereby introducing new prior knowledge about change in HeCD, i.e., semantic change priors. Notably, these change priors not only encompass binary change labels but also describe change semantics between different categories. Importantly, these change semantic rules are derived directly from the fundamental characteristics of HeCD without additional assumptions, making them theoretically general and broadly transferable across diverse HeCD task scenarios.

Based on these principles, we establish the change semantic consistency and inconsistency constraints. The former enforces that connected pairwise regions (represented by superpixels) in the changed image share the same change semantics, while the latter requires connected pairwise regions to exhibit different change semantics. From the perspective of graph signal processing, these two constraints correspond to low-pass and high-pass signals on the graph models, respectively. Building upon these constraints, we propose a change prior-guided image transformation model (CPIT), which decomposes one image into a transformed image and a changed image. CPIT ensures that the transformed image preserves the structural relationships based on the connections between pairwise objects and binary change labels, and constrains the changed image using both change semantic priors and sparsity priors, leading to a more accurate image transformation model and DI. Finally, an MRF-based fusion segmentation model

TABLE I
NOTATIONS.

| Symbol | Description |
|---|---|
| $\tilde{\mathbf{X}}$, $\tilde{\mathbf{Y}}$ | Pre- and post-event images |
| $P^{t1}$, $P^{t2}$ | Ground objects represented by images |
| $\tilde{\mathbf{X}}_i$, $\tilde{\mathbf{Y}}_i$ | $i$-th superpixels of $\tilde{\mathbf{X}}$, $\tilde{\mathbf{Y}}$ |
| $\mathbf{X}$, $\mathbf{Y}$ | Feature matrices of $\tilde{\mathbf{X}}$, $\tilde{\mathbf{Y}}$ |
| $\mathbf{D}^x$, $\mathbf{D}^y$ | Feature distance matrices |
| $\mathbf{X}'$, $\mathbf{Y}'$ | Transformed feature matrices of $\mathbf{Y}$, $\mathbf{X}$ |
| $\boldsymbol{\Delta}^x$, $\boldsymbol{\Delta}^y$ | Changed feature matrices |
| $G_{hn}^{t1}$, $G_{hf}^{t1}$ | KNN and KFN graphs |
| $E_{hn}^{t1}$, $E_{hf}^{t1}$ | Edge sets of KNN and KFN graphs |
| $\mathbf{L}^{t1}$, $\mathbf{L}^{t2}$, $\mathbf{L}^{sc}$ | Graph Laplacian matrices |

refines the DI using spatial consistency and spatial smoothness priors, producing the final change map (CM). The derivation, formulation, and utilization of change priors presented in this paper provide a valuable insight for both HoCD and HeCD tasks. This work makes the following contributions.

1) Starting from the definition of the HeCD problem, we analyze the connections among object relationships, change labels, and change semantics in pairwise regions, and derive theoretically general change rules. This analysis provides richer change prior knowledge for HeCD.

2) Based on change semantic priors, we construct change semantic consistency and inconsistency constraints, and integrate them with change sparsity and image structure constraints to develop a more precise change prior-guided image transformation model for HeCD.

3) Using graph signal processing tools, we analyze the spectral characteristics of changes on the semantic consistency and inconsistency induced graphs. The proposed CPIT is evaluated on five benchmark heterogeneous remote sensing image datasets, demonstrating superior performance.

## II. Connections between Pairwise Object Relationships, Change Labels and Change Semantics

Given a pair of co-registered images acquired at time points $t_1$ (before the event) and $t_2$ (after the event) using different sensors, we denote them as $\tilde{\mathbf{X}} \in \mathbb{R}^{M\times N\times B_x}$ and $\tilde{\mathbf{Y}} \in \mathbb{R}^{M\times N\times B_y}$, respectively, where the pixels are represented as $\tilde{x}(m,n,b_x)$ and $\tilde{y}(m,n,b_y)$, respectively. The goal of HeCD is to generate a binary change map that indicates the changed/unchanged pixels. Since the imaging sensors and conditions differ between $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$, directly comparing the multi-temporal images is infeasible for accurate change detection. Table I lists some important notations for easy reference.

### A. Connections between Object Relationships and Change Labels

First, we investigate the connections between pairwise object relationships and their change labels. We define $P^{t1}$ and $P^{t2}$ as the ground objects represented by the multi-temporal images respectively, and define $\boldsymbol{L}$ as the change label of images, with $L_i = 0$ representing "unchanged" and $L_i = 1$ representing "changed". For the pairwise regions represented by the $i$-th and $j$-th pixels, based on the definition of change detection problem, we can obtain 4 combinations of pairwise object relationships and their associated change labels as follows

$$\begin{aligned} P_i^{t1} = P_i^{t2},\ P_j^{t1} = P_j^{t2} \quad &\Leftrightarrow \quad L_i = 0,\ L_j = 0;\\ P_i^{t1} = P_i^{t2},\ P_j^{t1} \neq P_j^{t2} \quad &\Leftrightarrow \quad L_i = 0,\ L_j = 1;\\ P_i^{t1} \neq P_i^{t2},\ P_j^{t1} = P_j^{t2} \quad &\Leftrightarrow \quad L_i = 1,\ L_j = 0;\\ P_i^{t1} \neq P_i^{t2},\ P_j^{t1} \neq P_j^{t2} \quad &\Leftrightarrow \quad L_i = 1,\ L_j = 1. \end{aligned} \tag{1}$$

Then, we can build the connections between pairwise objects and change labels as

$$\begin{aligned} &P_i^{t1} = P_j^{t1}, P_i^{t2} = P_j^{t2} \Rightarrow (L_i, L_j) \in \{(0,0), (1,1)\};\\ &P_i^{t1} \neq P_j^{t1}, P_i^{t2} = P_j^{t2} \Rightarrow (L_i, L_j) \in \{(0,1), (1,0), (1,1)\};\\ &P_i^{t1} = P_j^{t1}, P_i^{t2} \neq P_j^{t2} \Rightarrow (L_i, L_j) \in \{(0,1), (1,0), (1,1)\};\\ &P_i^{t1} \neq P_j^{t1}, P_i^{t2} \neq P_j^{t2} \Rightarrow (L_i, L_j) \in \{(0,0), (0,1), (1,0), (1,1)\}. \end{aligned} \tag{2}$$

By combining (1) and (2), we can obtain

$$\text{If } L_i = 0, L_j = 0, \quad \text{then} \begin{cases} P_i^{t1} = P_j^{t1} \Leftrightarrow P_i^{t2} = P_j^{t2};\\ P_i^{t1} \neq P_j^{t1} \Leftrightarrow P_i^{t2} \neq P_j^{t2}; \end{cases} \tag{3}$$

This principle directly motivates the structure-consistency regularization terms in the image transformation model, whose detailed implementation is presented in Section III-D.

TABLE II
THE PROCESS OF RULE DERIVATION FROM THE CONNECTIONS BETWEEN PAIRWISE OBJECT RELATIONSHIPS AND CHANGE SEMANTICS.

| Rules | Restrictions | Criteria | Examples $\begin{pmatrix}\tilde{\mathbf{X}}_i \rightarrow \tilde{\mathbf{Y}}_i\\ \tilde{\mathbf{X}}_j \rightarrow \tilde{\mathbf{Y}}_j\end{pmatrix}$ | | Conclusions |
|---|---|---|---|---|---|
| Rule 1 | $P_i^{t1} = P_j^{t1}$ | $P_i^{t2} = P_j^{t2}$ | (🌲→🌲 / 🌲→🌲) (🌲→💧 / 🌲→💧) | ⇨ | $C_i^{t1\rightarrow t2} = C_j^{t1\rightarrow t2}$ |
| Rule 3 | $P_i^{t1} \neq P_j^{t1}$ | $P_i^{t2} = P_j^{t2}$ | (🌲→🌲 / 💧→🌲) (🌲→💧 / 💧→💧) (🌲→🏠 / 💧→🏠) | ⇨ | $C_i^{t1\rightarrow t2} \neq C_j^{t1\rightarrow t2}$ |
| Rule 5 | $P_i^{t1} = P_j^{t1}$ | $P_i^{t2} \neq P_j^{t2}$ | (🌲→🌲 / 🌲→💧) (🌲→💧 / 🌲→🌲) (🌲→🏠 / 🌲→💧) | ⇨ | $C_i^{t1\rightarrow t2} \neq C_j^{t1\rightarrow t2}$ |

The symbols 🌲 💧 🏠 represent different kinds of objects, e.g., 🌲 Forest, 💧 Water, 🏠 Building.

### B. Connections between Object Relationships and Change Semantics

Second, we investigate the connections between pairwise object relationships and their change semantics. We use $C_i^{t1\rightarrow t2} = P_i^{t1} \ominus P_i^{t2}$ to denote the "from to" change state (i.e., change from the object category $P_i^{t1}$ to the object category $P_i^{t2}$), defined as the "change semantic" in this paper. Substituting $C_i^{t1\rightarrow t2}$ into (2) and applying the definition of change detection, we can obtain

$$P_i^{t1} = P_j^{t1}, P_i^{t2} = P_j^{t2} \;\Rightarrow\; C_i^{t1\rightarrow t2} = C_j^{t1\rightarrow t2}; \tag{4a}$$

$$P_i^{t1} \neq P_j^{t1}, P_i^{t2} = P_j^{t2} \;\Rightarrow\; C_i^{t1\rightarrow t2} \neq C_j^{t1\rightarrow t2}; \tag{4b}$$

$$P_i^{t1} = P_j^{t1}, P_i^{t2} \neq P_j^{t2} \;\Rightarrow\; C_i^{t1\rightarrow t2} \neq C_j^{t1\rightarrow t2}; \tag{4c}$$

$$P_i^{t1} \neq P_j^{t1}, P_i^{t2} \neq P_j^{t2} \;\Rightarrow\; \begin{cases} C_i^{t1\rightarrow t2} = C_j^{t1\rightarrow t2};\\ C_i^{t1\rightarrow t2} \neq C_j^{t1\rightarrow t2}; \end{cases} \tag{4d}$$

From (4), it can be seen that the change behavior is not inferred from the internal relationships of a single image, but rather from the combined pairwise relationships of the two images. For practical implementation, we further decompose these relationships into a restriction term and a criterion term. Table II presents illustrative symbolic examples to intuitively show the connections between pairwise object relationships and their corresponding change semantics. These examples serve to clarify the logical foundation underlying the formulation of change semantics determination rules. Based on (4) and Table II, we derive the rules of change semantics in the context of the change detection problem.

**Change Semantic Consistency Rules**

**Rule 1**: when restricting $P_i^{t1} = P_j^{t1}$, then if $P_i^{t2} = P_j^{t2}$ holds, we have $C_i^{t1\rightarrow t2} = C_j^{t1\rightarrow t2}$.

**Rule 2**: when restricting $P_i^{t2} = P_j^{t2}$, then if $P_i^{t1} = P_j^{t1}$ holds, we have $C_i^{t1\rightarrow t2} = C_j^{t1\rightarrow t2}$.

**Change Semantic Inconsistency Rules**

**Rule 3**: when restricting $P_i^{t1} \neq P_j^{t1}$, then if $P_i^{t2} = P_j^{t2}$ holds, we have $C_i^{t1\rightarrow t2} \neq C_j^{t1\rightarrow t2}$.

**Rule 4**: when restricting $P_i^{t2} \neq P_j^{t2}$, then if $P_i^{t1} = P_j^{t1}$ holds, we have $C_i^{t1\rightarrow t2} \neq C_j^{t1\rightarrow t2}$.

**Rule 5**: when restricting $P_i^{t1} = P_j^{t1}$, then if $P_i^{t2} \neq P_j^{t2}$ holds, we have $C_i^{t1\rightarrow t2} \neq C_j^{t1\rightarrow t2}$.

**Rule 6**: when restricting $P_i^{t2} = P_j^{t2}$, then if $P_i^{t1} \neq P_j^{t1}$ holds, we have $C_i^{t1\rightarrow t2} \neq C_j^{t1\rightarrow t2}$.

Therefore, we derived a semantic change prior consisting of six rules from pairwise connections of (4): two rules for semantic consistency determination and four rules for semantic inconsistency determination. This semantic prior has two highly appealing characteristics.

- **First**, this change prior is derived solely from the inherent nature of the change detection problem without relying on any additional assumptions, making it theoretically general and broadly transferable to different change detection tasks, including both homogeneous and heterogeneous change detection.
- **Second**, this prior provides a superior solution for HeCD. It extracts change information solely based on the internal object relationships within each image (i.e., comparing $P_i^{t1}$ and $P_j^{t1}$, $P_i^{t2}$ and $P_j^{t2}$), rather than performing cross-image comparisons (e.g., comparing $P_i^{t1}$ and $P_i^{t2}$). This significantly reduces the complexity of solving the HeCD problem, as it avoids direct comparisons between heterogeneous images and enables change detection without requiring labeled samples.

In addition, comparing (2) and (4) provides several insights: **(i)** Analyzing pairwise object relationships can extract change-related information, including change labels and change semantics. **(ii)** Connections to change labels are more numerous and complex (12 mappings in (2)), while connections to change semantics are fewer (5 mappings in (4), with 3 one-to-one), making them easier to use for constructing constraints. **(iii)** These mappings suit different purposes: change-label connections are suitable for energy-based models or post-processing of segmentation results (constraining the final change map with $\boldsymbol{L}$), whereas change-semantic connections are better for image transformation, as they can be directly embedded into both the transformed and changed images. For example, in $\tilde{\mathbf{Y}} = \tilde{\mathbf{Y}}' \ominus \tilde{\boldsymbol{\Delta}}^y$, where $\tilde{\boldsymbol{\Delta}}^y$ denotes changed image and $\tilde{\mathbf{Y}}'$ denotes transformed image from $\tilde{\mathbf{X}}$, change semantics can be naturally incorporated into the transformation process.

Nevertheless, the practical effectiveness of semantic change rules depends on their implementation in a specific HeCD model. Therefore, a key challenge lies in effectively leveraging these semantic change priors for HeCD—specifically, how to impose their restrictions and mathematically implement their inferences into a practical HeCD framework. To address this, we construct change semantic consistency/inconsistency constraints for the image transformation model.

## III. Change Prior-Guided Image Transformation Model

This section presents the proposed change prior-guided image transformation model (CPIT) for HeCD in detail, comprising three main parts: (i) constructing graphs to characterize pairwise object relationships and formulating constraints of semantic change priors; (ii) building and solving the CPIT model; (iii) fusing and segmenting the changed image to extract change information, as illustrated in Fig. 1.

### A. Preprocessing

Modeling pairwise object relationships at the pixel level results in a quadratic computational complexity of $\mathcal{O}(M^2N^2)$. To alleviate this issue, we adopt superpixels rather than individual pixels as the basic analysis units, which substantially reduces the computational burden while exploiting contextual information and preserving object boundaries [45], [46].

With superpixels as the base unit, the pixels within a superpixel are assumed to share the same change label. This requires each superpixel to be homogeneous within each image, meaning that all pixels inside a superpixel in both multi-temporal images should belong to the same land-cover category, respectively. We adopt the preprocessing method to generate the co-segmented superpixels similar to that in [31]. It begins by normalizing multi-temporal images—applying linear normalization for optical images and logarithmic normalization for SAR images [47], and then stacks multi-temporal images along the channel dimension. Then, superpixel segmentation [48] is performed on the stacked image to generate a segmentation map $\Lambda$ with $N_S$ superpixels, which efficiently preserves object boundaries with linear complexity in the number of pixels. This segmentation map is subsequently used to partition both multi-temporal images into corresponding superpixels, ensuring the same edge contours, defined as $\tilde{\mathbf{X}}_i$, $\tilde{\mathbf{Y}}_i$, $i = 1, \cdots, N_S$. After segmentation, we extract the three quartiles and mean values as superpixel features, forming feature matrices $\mathbf{X} \in \mathbb{R}^{4B_x \times N_S}$ and $\mathbf{Y} \in \mathbb{R}^{4B_y \times N_S}$. Although more discriminative features, such as deep features, may provide stronger representational and discriminative abilities, simple statistical features are adopted in this paper for simplicity and computational efficiency. We define the feature distance matrices as $\mathbf{D}^x$ and $\mathbf{D}^y$ respectively, where the elements of $D_{i,j}^x = \|\mathbf{X}_i - \mathbf{X}_j\|_2^2$ and $D_{i,j}^y = \|\mathbf{Y}_i - \mathbf{Y}_j\|_2^2$ denote feature distances between superpixels within the pre- and post-event images, respectively.

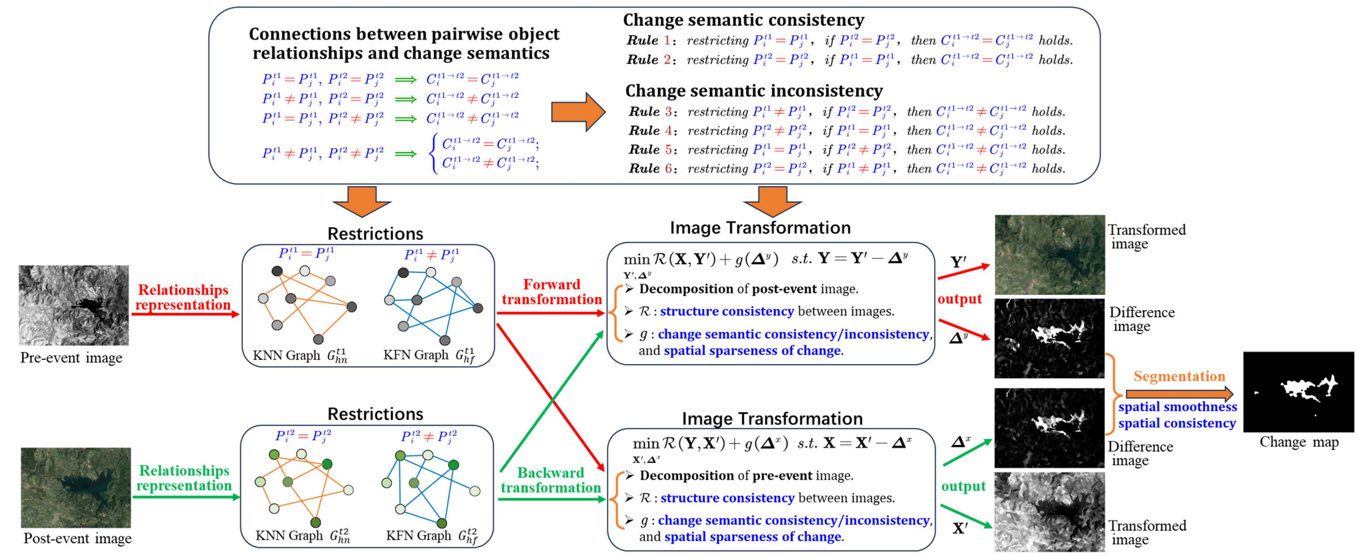



Fig. 1. Illustration of the proposed change prior-guided image transformation model.

## B. Image Transformation Model

Image transformation-based HeCD approaches aim to map one image into the domain of the other. Given that our method uses superpixels as fundamental units and represents them via extracted features, it is necessary to build transformation functions between the corresponding features. Let $\mathbf{\Delta}^x$ and $\mathbf{\Delta}^y$ represent the changed feature matrices that capture the change information. Then, the decomposition-based image transformation models can be formulated as

$$\begin{aligned} \text{Forward}: \ & \min_{\mathbf{Y}',\mathbf{\Delta}^y} \mathcal{R}\left(\mathbf{X},\mathbf{Y}'\right)+g\left(\mathbf{\Delta}^y\right) \ \ s.t. \ \mathbf{Y}=\mathbf{Y}'-\mathbf{\Delta}^y \\ \text{Backward}: \ & \min_{\mathbf{X}',\mathbf{\Delta}^x} \mathcal{R}\left(\mathbf{Y},\mathbf{X}'\right)+g\left(\mathbf{\Delta}^x\right) \ \ s.t. \ \mathbf{X}=\mathbf{X}'-\mathbf{\Delta}^x \end{aligned} \tag{5}$$

Taking the forward transformation as an example, the regularization term $\mathcal{R}\left(\mathbf{X},\mathbf{Y}'\right)$ serves as a constraint on the transformed matrix $\mathbf{Y}'$, capturing the relationship between the transformed and the original images, such as the structure consistency proposed in [49]. Meanwhile, the regularization term $g\left(\mathbf{\Delta}^y\right)$ serves as a constraint on the changed features, based on the change prior.

## C. From Change Prior to Change Constraints

*1) Change Semantic Consistency:* In Rule 1, the restriction is $P_i^{t1}=P_j^{t1}$, i.e., $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{X}}_j$ represent the same land-cover category. To approximate this restriction, we construct a K-nearest-neighbor (KNN) graph $G_n^{t1}=\left\{V_n^{t1},E_n^{t1},\mathbf{A}_n^{t1}\right\}$ for the pre-event image $\tilde{\mathbf{X}}$, which is used to capture feature-induced similar superpixel relationships within the image:

$$\begin{aligned} V_n^{t1} &= \left\{ i | i\in\mathcal{I}\right\}; \\ E_n^{t1} &= \left\{ (i,j) \,|j\in\mathcal{N}_i^x, i\in\mathcal{I}\right\} \end{aligned} \tag{6}$$

where $\mathcal{I}=\left\{1,2,\cdots,N_S\right\}$ is the index set, $\mathbf{A}_n^{t1}$ denotes the affinity matrix, $V_n^{t1}$ and $E_n^{t1}$ denote the vertex set and edge set, respectively. $\mathcal{N}_i^x$ is the index set of the KNN of the $i$-th superpixel $\tilde{\mathbf{X}}_i$ within the image by sorting the feature distance vector $\mathbf{D}_i^x$.

The choice of $k$ in the KNN graph is critical. A large $k$ may connect superpixels from different objects, while a small $k$ leads to too few edges in $G_n^{t1}$, making it hard to capture object relationships and reducing the effect of Rule 1. In this paper, we start by constructing an initial KNN graph using a small value of $k=\sqrt{N_S}$. Then, we apply a neighbor expansion strategy inspired by balance theory from social psychology [50], [51]. Following the principle of "*the friend of my friend is my friend*" in balance theory, we build high-order KNN graphs by treating the "neighbor" as "friend."[1] For instance, a $h$-order ($h\geq 2$) graph $G_{hn}^{t1}=\left\{V_{hn}^{t1},E_{hn}^{t1},\mathbf{A}_{hn}^{t1}\right\}$ can be defined with $V_{hn}^{t1}=\mathcal{I}$ and

$$E_{hn}^{t1}=E_{(h-1)n}^{t1}\cup\left\{(i,j)|(i,t)\in E_n^{t1},(t,j)\in E_{(h-1)n}^{t1}\right\} \tag{7}$$

Similarly, a high-order KNN graph $G_{hn}^{t2}=\left\{V_{hn}^{t2},E_{hn}^{t2},\mathbf{A}_{hn}^{t2}\right\}$ can be constructed for the post-event image.

Superpixel pairs connected by $G_{hn}^{t1}$ are more likely to belong to the same land-cover category in the pre-event image. According to Rule 1, a higher probability that $P_i^{t2}=P_j^{t2}$ implies a higher likelihood that $C_i^{t1\to t2}=C_j^{t1\to t2}$. This, in turn, leads to smaller values of $\left\|\mathbf{\Delta}_i^x-\mathbf{\Delta}_j^x\right\|$ and $\left\|\mathbf{\Delta}_i^y-\mathbf{\Delta}_j^y\right\|$. Based on this, we construct the following penalty function for Rule 1, taking $\mathbf{\Delta}^y$ in the forward transformation (5) as an example:

$$g_1\left(\mathbf{\Delta}^y\right)=\sum_{(i,j)\in E_{hn}^{t1}}\left\|\mathbf{\Delta}_i^y-\mathbf{\Delta}_j^y\right\|_2^2 W_{i,j}^y \tag{8}$$

with

$$W_{i,j}^y=\left\{\begin{array}{cc} \exp\left(-D_{i,j}^y\right), & \text{if} \quad D_{i,j}^y\leq\sigma_i^y \\ 0, & \text{otherwise} \end{array}\right. \tag{9}$$

[1]Note that the principle "the enemy of my enemy is my friend" in social psychology [50], [51] is not adopted in constructing the high-order KNN graph, since remote sensing images usually contain multiple land-cover classes. Therefore, two vertices that are both dissimilar to a third one do not necessarily belong to the same class.

where the condition $(i,j) \in E_{hn}^{t1}$ in (8) is used to approximately satisfy the restriction of $P_i^{t1} = P_j^{t1}$ required by Rule 1, and the threshold $\sigma_i^y = \max_j \left\{ D_{i,j}^y | (i,j) \in E_{hn}^{t2} \right\}$ is used to exclude superpixel pairs that likely do not satisfy $P_i^{t2} = P_j^{t2}$ from participating in the penalty. In (8), if the feature distance $D_{i,j}^y$ is very small, which means that $\tilde{\mathbf{Y}}_i$ and $\tilde{\mathbf{Y}}_j$ most likely belong to the same object class (i.e., $P_i^{t2} = P_j^{t2}$), then minimizing the penalty term $g_1$ requires a small value of $\left\| \mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y \right\|_2$, thereby meeting the requirement of $C_i^{t1 \to t2} = C_j^{t1 \to t2}$.

In the same manner, we can construct the following penalty function $g_2$ by using the $G_{hn}^{t2}$ for Rule 2:

$$g_2 \left( \mathbf{\Delta}^y \right) = \sum_{(i,j) \in E_{hn}^{t2}} \left\| \mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y \right\|_2^2 W_{i,j}^x \tag{10}$$

with

$$W_{i,j}^x = \begin{cases} \exp\left( -D_{i,j}^x \right), & \text{if} \quad D_{i,j}^x \le \sigma_i^x \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

We define a matrix $\mathbf{W}^{sc} \in \mathbb{R}^{N_S \times N_S}$ as

$$W_{i,j}^{sc} = W_{i,j}^y \delta\left( (i,j) \in E_{hn}^{t1} \right) + W_{i,j}^x \delta\left( (i,j) \in E_{hn}^{t2} \right) \tag{12}$$

where $\delta(\cdot)$ is 1 when the condition in parentheses holds, and 0 otherwise. Define the diagonal matrix $\mathbf{D}^{sc} \in \mathbb{R}^{N_S \times N_S}$, whose $i$-th diagonal element is $D_{i,i}^{sc} = \sum_{j=1}^{N_S} \left( W_{i,j}^{sc} + W_{j,i}^{sc} \right)/2$, and define the Laplacian matrix as $\mathbf{L}^{sc} = \mathbf{D}^{sc} - \frac{\mathbf{W}^{sc} + (\mathbf{W}^{sc})^T}{2}$.

By combining (8) and (10), and substituting (12), we have the change semantic consistency regularization (CSSR) as

$$\begin{aligned} g_{sc}\left( \mathbf{\Delta}^y \right) &= g_1\left( \mathbf{\Delta}^y \right) + g_2\left( \mathbf{\Delta}^y \right) \\ &= 2\mathrm{Tr}\left( \mathbf{\Delta}^y \mathbf{L}^{sc} \left( \mathbf{\Delta}^y \right)^T \right) \end{aligned} \tag{13}$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix. This type of CSSR can be regarded as a graph smoothness constraint or low-pass constraint on the $\mathbf{W}^{sc}$ induced graph, as discussed in the supplementary material.

*2) Change Semantic Inconsistency:* In Rule 3, the restriction is $P_i^{t1} \neq P_j^{t1}$, meaning that $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{X}}_j$ belong to different land-cover categories. To approximate this restriction, we construct a K-farthest-neighbor (KFN) graph $G_f^{t1} = \left\{ V_f^{t1}, E_f^{t1}, \mathbf{A}_f^{t1} \right\}$ for the pre-event image $\tilde{\mathbf{X}}$, which characterizes the feature-induced dissimilar superpixel relationships within the image as follows:

$$\begin{aligned} V_f^{t1} &= \{ i | i \in \mathcal{I} \}; \\ E_f^{t1} &= \{ (i,j) | j \in \mathcal{F}_i^x, i \in \mathcal{I} \} \end{aligned} \tag{14}$$

where $\mathcal{F}_i^x$ denotes the index set of the farthest-neighbors. Intuitively, farthest-neighbors can be obtained by selecting the largest $k$ elements in $\mathbf{D}_i^x$. However, this often links a vertex to superpixels from only a few or even the same class, reducing the KFN graph's ability to capture diverse dissimilarity relationships required by Rule 3.

In this paper, the construction of $\mathcal{F}_i^x$ proceeds as follows: first, sort $\mathbf{D}_i^x$ in descending order, then sample the top $k = \sqrt{N_S}$ elements at a fixed interval (set to 5 in our method). Second, we also apply a neighbor expansion strategy inspired by balance theory in social psychology [50], [51]. Following the principles of "*the enemy of my friend is my enemy*" and "*the friend of my enemy is my enemy*" in balance theory, we construct a high-order KFN graph by treating the "nearest-neighbor" as "friend" and "farthest-neighbor" as "enemy." Accordingly, we obtain the high-order KFN graph $G_{hf}^{t1} = \left\{ V_{hf}^{t1}, E_{hf}^{t1}, \mathbf{A}_{hf}^{t1} \right\}$ with $V_{hf}^{t1} = \mathcal{I}$ and

$$\begin{aligned} (i,j) \in E_{hf}^{t1}; \quad & \text{if} \quad (i,j) \in E_f^{t1} \\ & \text{or} \quad (i,t) \in E_{hn}^{t1}, (t,j) \in E_f^{t1} \\ & \text{or} \quad (i,t) \in E_f^{t1}, (t,j) \in E_{hn}^{t1} \end{aligned} \tag{15}$$

Similarly, a high-order KFN graph $G_{hf}^{t2} = \left\{ V_{hf}^{t2}, E_{hf}^{t2}, \mathbf{A}_{hf}^{t2} \right\}$ can be constructed for the post-event image.

Superpixel pairs connected by $G_{hf}^{t1}$ are more likely to belong to different land-cover categories in the pre-event image. According to Rule 3, a higher probability that $P_i^{t2} = P_j^{t2}$ implies a higher probability that $C_i^{t1 \to t2} \neq C_j^{t1 \to t2}$, leading to larger values of $\left\| \mathbf{\Delta}_i^x - \mathbf{\Delta}_j^x \right\|$ and $\left\| \mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y \right\|$. Based on this, we can construct the following penalty function for Rule 3, taking $\mathbf{\Delta}^y$ in the forward transformation (5) as an example:

$$g_3\left( \mathbf{\Delta}^y \right) = \sum_{(i,j) \in E_{hf}^{t1}} \varphi\left( \left\| \mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y \right\| \right) W_{i,j}^y \tag{16}$$

where $(i,j) \in E_{hf}^{t1}$ is used to approximately satisfy the condition $P_i^{t1} \neq P_j^{t1}$ as required by Rule 3, and $W_{i,j}^y$ is defined in (9). The key idea of (16) is to prevent the connected $\tilde{\mathbf{\Delta}}_i^y$ and $\tilde{\mathbf{\Delta}}_j^y$ from being assigned the same change semantics, rather than artificially enlarging their feature distance when they are already different. Accordingly, the penalty is applied only to overly similar pairs: $\varphi$ takes large values when $\left\| \mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y \right\| \approx 0$ and gradually vanishes as the difference increases. In this way, the penalty does not enforce unnecessary separation, but instead stabilizes and reinforces meaningful contrasts.

In the proposed method, we choose $\varphi$ as

$$\varphi\left( \left\| \mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y \right\| \right) = \frac{1}{\left\| \mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y \right\|_2^2 + \varepsilon} \tag{17}$$

where $\varepsilon$ is the parameter that controls the steepness of the curve. We set $\varepsilon$ as the average feature distance between connected superpixels in $E_n^{t2}$, which roughly represents the intra-class noise level.

In the same manner, we can construct the penalty function $g_4$ by using the $G_{hf}^{t2}$ for Rule 4:

$$g_4\left( \mathbf{\Delta}^y \right) = \sum_{(i,j) \in E_{hf}^{t2}} \varphi\left( \left\| \mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y \right\| \right) W_{i,j}^x \tag{18}$$

where $(i,j) \in E_{hf}^{t2}$ is used to approximately meet the condition of $P_i^{t2} \neq P_j^{t2}$ as required by Rule 4, and $W_{i,j}^x$ is defined in (11).

In Rule 5, the restriction $P_i^{t1} = P_j^{t1}$ is approximately implemented via the high-order KNN graph $G_{hn}^{t1}$. For the superpixel pairs connected by $E_{hn}^{t1}$, Rule 5 indicates that if $P_i^{t2} \neq P_j^{t2}$, then $C_i^{t1 \to t2} \neq C_j^{t1 \to t2}$ is more likely to hold, leading to larger $\left\| \mathbf{\Delta}_i^x - \mathbf{\Delta}_j^x \right\|$ and $\left\| \mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y \right\|$. However, for Rule 5, once superpixel pairs satisfy $P_i^{t2} \neq P_j^{t2}$, they already meet the same rule criterion, and they should not be further distinguished according to the exact magnitude of the feature distance. For example, even if the feature distance between a "Grasslands–Buildings" pair is larger than that between a "Grasslands–Water" pair in the post-event image, it is still inappropriate to assign them different penalties solely based on this difference. Under Rule 5, both cases correspond to $P_i^{t2} \neq P_j^{t2}$, and therefore should be treated in the same manner. Therefore, we construct the following penalty function

$$g_5\left(\mathbf{\Delta}^y\right) = \sum_{(i,j)\in E_{hn}^{t1}} \varphi\left(\left\|\mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y\right\|\right)\delta\left(D_{i,j}^y \geq \rho_i^y\right) \tag{19}$$

where the threshold $\rho_i^y = \min_j \left\{D_{i,j}^y \middle| (i,j) \in E_{hf}^{t2}\right\}$ is used to exclude superpixel pairs that likely do not satisfy $P_i^{t2} \neq P_j^{t2}$ from participating in the penalty, and $\delta(\cdot)$ assigns the same weight for eligible superpixel pairs.

In the same manner, we can construct the penalty function $g_6$ by using the $G_{hn}^{t2}$ for Rule 6:

$$g_6\left(\mathbf{\Delta}^y\right) = \sum_{(i,j)\in E_{hn}^{t2}} \varphi\left(\left\|\mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y\right\|\right)\delta\left(D_{i,j}^x \geq \rho_i^x\right) \tag{20}$$

Define a matrix $\mathbf{W}^{si} \in \mathbb{R}^{N_S \times N_S}$ as

$$\begin{aligned} W_{i,j}^{si} =& W_{i,j}^y \delta\left((i,j)\in E_{hf}^{t1}\right) + W_{i,j}^x \delta\left((i,j)\in E_{hf}^{t2}\right) \\ &+ \delta\left((i,j)\in E_{hn}^{t1}\right)\delta\left(D_{i,j}^y \geq \rho_i^y\right) \\ &+ \delta\left((i,j)\in E_{hn}^{t2}\right)\delta\left(D_{i,j}^x \geq \rho_i^x\right) \end{aligned} \tag{21}$$

By combining (16), (18), (19), and (20), and substituting (21), we have the change semantic inconsistency regularization (CSIR) as

$$\begin{aligned} g_{si}\left(\mathbf{\Delta}^y\right) &= g_3\left(\mathbf{\Delta}^y\right) + g_4\left(\mathbf{\Delta}^y\right) + g_5\left(\mathbf{\Delta}^y\right) + g_6\left(\mathbf{\Delta}^y\right) \\ &= \sum\nolimits_{i,j\in\mathcal{I}} \varphi\left(\left\|\mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y\right\|\right) W_{i,j}^{si} \end{aligned} \tag{22}$$

This type of CSIR can be regarded as a graph non-smoothness constraint or a high-pass constraint on the $\mathbf{W}^{si}$ induced graph, as discussed in the supplementary material.

*3) Change Sparsity:* The change sparsity assumption is based on the fact that, in real-world scenarios, most geographic surfaces remain stable over time, while only a small portion of the area typically undergoes change due to human activities or natural events. To enforce sparsity in the detected changes, we introduce a $\ell_{2,1}$-norm based change sparsity regularization (CSR) that ensures only a limited number of superpixels exhibit nonzero values, expressed as

$$\left\|\mathbf{\Delta}^y\right\|_{2,1} = \sum\nolimits_{i\in\mathcal{I}} \left\|\mathbf{\Delta}_i^y\right\|_2 \tag{23}$$

*4) Change Constraint:* By combining the semantic prior based (13), (22) and sparsity prior based (23), we can obtain the $g\left(\mathbf{\Delta}^y\right)$ for model (5) as

$$\begin{aligned} g\left(\mathbf{\Delta}^y\right) =& \alpha \sum\nolimits_{i,j\in\mathcal{I}} \varphi\left(\left\|\mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y\right\|\right) W_{i,j}^{si} \\ &+ 2\mathrm{Tr}\left(\mathbf{\Delta}^y \mathbf{L}^{sc} \left(\mathbf{\Delta}^y\right)^T\right) + \lambda \left\|\mathbf{\Delta}^y\right\|_{2,1} \end{aligned} \tag{24}$$

where $\alpha, \lambda > 0$ are the weighting parameters.

### D. CPIT Model

*1) Image Structural Consistency based $\mathcal{R}$:* For the transformed image $\tilde{\mathbf{Y}}'$, obtained by mapping the pre-event image into the post-event domain, no changes exist relative to $\tilde{\mathbf{X}}$ by definition. Consequently, for the image pair $\left(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}'\right)$, we naturally have $L_i = L_j = 0$. Based on the connections between pairwise object relationships and change labels in (3), this constraint yields

$$\begin{aligned} P_i^{t1} = P_j^{t1} \;&\Leftrightarrow\; P_i^{t2} = P_j^{t2} \\ P_i^{t1} \neq P_j^{t1} \;&\Leftrightarrow\; P_i^{t2} \neq P_j^{t2} \end{aligned} \tag{25}$$

Based on (25), superpixel pairs connected by $G_{hn}^{t1}$ are expected to represent the same object type and remain similar after transformation, while those connected by $G_{hf}^{t1}$ are expected to belong to different object types and remain dissimilar. Accordingly, the following penalty function $\mathcal{R}$ can be constructed for (25):

$$\mathcal{R}\left(\mathbf{X}, \mathbf{Y}'\right) = 2\mathrm{Tr}\left(\mathbf{Y}'\mathbf{L}^{t1}\mathbf{Y}'^T\right) + \beta \sum_{(i,j)\in E_{hf}^{t1}} \varphi\left(\left\|\mathbf{Y}_i' - \mathbf{Y}_j'\right\|\right) \tag{26}$$

where $\mathbf{L}^{t1}$ is the Laplacian matrix of graph $G_{hn}^{t1}$ with the edge weight $w_{i,j} = \exp\left(-D_{i,j}^x\right)$ and $\beta > 0$ is the weighting parameter.

From (26), we can find that the regularization term $\mathcal{R}\left(\mathbf{X}, \mathbf{Y}'\right)$ requires the transformed image $\tilde{\mathbf{Y}}'$ to preserve the similarity relationship and dissimilarity relationships of the original image $\tilde{\mathbf{X}}$ on the KNN graph $G_{hn}^{t1}$ and KFN graph $G_{hf}^{t1}$, respectively. If we define the relationships within the image as topological structures, then $\mathcal{R}\left(\mathbf{X}, \mathbf{Y}'\right)$ can be regarded as an image structural consistency regularization, which has been employed in previous works, especially for structural consistency based on similarity relationships [40], [49], [52].

*2) Objective Function of CPIT:* By substituting the $g\left(\mathbf{\Delta}^y\right)$ of (24) and $\mathcal{R}\left(\mathbf{X}, \mathbf{Y}'\right)$ of (26) into the forward transformation model (5), we can obtain the CPIT model based on pairwise connections derived from (3) and (4) as

$$\begin{aligned} \min_{\mathbf{Y}', \mathbf{\Delta}^y} \;& 2\mathrm{Tr}\left(\mathbf{Y}'\mathbf{L}^{t1}\mathbf{Y}'^T\right) + 2\mathrm{Tr}\left(\mathbf{\Delta}^y \mathbf{L}^{sc}\left(\mathbf{\Delta}^y\right)^T\right) + \lambda\left\|\mathbf{\Delta}^y\right\|_{2,1} \\ &+ \beta \sum_{(i,j)\in E_{hf}^{t1}} \varphi\left(\left\|\mathbf{Y}_i' - \mathbf{Y}_j'\right\|\right) + \alpha \sum_{i,j\in\mathcal{I}} \varphi\left(\left\|\mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y\right\|\right) W_{i,j}^{si} \\ & s.t. \quad \mathbf{Y} = \mathbf{Y}' - \mathbf{\Delta}^y \end{aligned} \tag{27}$$

From (27), it is evident that the different regularization terms impose conflicting constraints on $\mathbf{Y}'$ and $\mathbf{\Delta}^y$. The first term, enforcing similarity regularization, encourages $\mathbf{Y}'$ to be smooth over the graph $G_{hn}^{t1}$, potentially leading to a trivial solution like $\mathbf{Y}' = 0$. The second term, derived from the change semantic consistency, prefers a smooth $\mathbf{\Delta}^y$, such as $\mathbf{\Delta}^y = 0$. The third term, change sparsity regularization, pushes $\mathbf{Y}'$ to deviate minimally from the original $\mathbf{Y}$, preserving unchanged regions as $\mathbf{\Delta}^y = 0$. The fourth term, enforcing dissimilarity regularization, promotes non-smooth variations in $\mathbf{Y}'$ over the graph $G_{hf}^{t1}$, driving differences between connected superpixels of $\mathbf{Y}'$. Finally, the last term, derived from the change semantic inconsistency, enforces differences between connected superpixels of $\mathbf{\Delta}^y$. Hence, these competing terms create an inherent tension, pulling $\mathbf{Y}'$ and $\boldsymbol{\Delta}^y$ in different directions.

Similarly, we can obtain the change prior-guided backward transformation model based on pairwise connections derived from (3) and (4) as

$$\begin{aligned}\min_{\mathbf{X}',\boldsymbol{\Delta}^x} & 2\mathrm{Tr}\left(\mathbf{X}'\mathbf{L}^{t2}\mathbf{X}'^T\right) + 2\mathrm{Tr}\left(\boldsymbol{\Delta}^x\mathbf{L}^{sc}\left(\boldsymbol{\Delta}^x\right)^T\right) + \lambda\|\boldsymbol{\Delta}^x\|_{2,1} \\ & + \beta \sum_{(i,j)\in E_{hf}^{t2}} \varphi\left(\left\|\mathbf{X}'_i - \mathbf{X}'_j\right\|\right) + \alpha \sum_{i,j\in\mathcal{I}} \varphi\left(\left\|\boldsymbol{\Delta}^x_i - \boldsymbol{\Delta}^x_j\right\|\right) W^{si}_{i,j} \\ & s.t. \quad \mathbf{X} = \mathbf{X}' - \boldsymbol{\Delta}^x\end{aligned} \tag{28}$$

The optimization problems of (27) and (28) can be approximately tackled using the efficient alternating direction method of multipliers (ADMM), as provided in the supplementary material.

### E. Spatial Consistency and Spatial Smoothness based Change Extraction

After obtaining the changed feature matrices from the transformation models in (27) and (28), we calculate the corresponding forward and backward DIs as follows

$$\left.\begin{aligned}\mathrm{DI}^y_{m,n} &= \|\boldsymbol{\Delta}^y_i\|_2 \\ \mathrm{DI}^x_{m,n} &= \|\boldsymbol{\Delta}^x_i\|_2\end{aligned}\right\} (m,n) \in \Lambda_i,\ i \in \mathcal{I} \tag{29}$$

where $\|\boldsymbol{\Delta}^y_i\|_2$ and $\|\boldsymbol{\Delta}^x_i\|_2$ represent the magnitude of change for the $i$-th superpixel of forward and backward DIs, respectively.

Based on the spatial consistency of change, i.e., both the forward and backward image transformation models describe the same change event, the change regions on the two DIs should be consistent. According to Tobler's First Law of Geography, spatially adjacent objects tend to be similar and correlated. Thus, neighboring regions usually share similar physical properties and temporal evolution, making it reasonable to assume correlated change states among adjacent superpixels, i.e., the spatial smoothness of change.

In order to fuse and segment the forward and backward DIs for computing the change map (CM), we adopt the MRF co-segmentation approach proposed in [25], which incorporates the change spatial consistency and change spatial smoothness in the unified fusion and segmentation framework, ultimately leading to a more precise binary CM. The overall CPIT framework is summarized in Algorithm 1.

Algorithm 1: Change prior-guided image transformation for HeCD.

**Input:** Heterogeneous images and parameters of $N_S$, $\alpha$, $\beta$, $\lambda$.
**Relationships representation:**
Divide heterogeneous images into superpixels.
Compute the feature distance matrices $\mathbf{D}^x$ and $\mathbf{D}^y$.
Construct KNN graphs $G^{t1}_{hn}$ and $G^{t2}_{hn}$.
Construct KFN graphs $G^{t1}_{hf}$ and $G^{t2}_{hf}$.
**Image transformation:**
Compute the matrix $\mathbf{L}^{sc}$ based on change semantic consistency.
Compute the matrix $\mathbf{W}^{si}$ based on change semantic inconsistency.
Compute the matrices $\mathbf{L}^{t1}$, $\mathbf{L}^{t2}$ based on structural consistency.
Solve the forward transformation model (27) to obtain $\mathbf{Y}'$ and $\boldsymbol{\Delta}^y$.
Solve the backward transformation model (28) to obtain $\mathbf{X}'$ and $\boldsymbol{\Delta}^x$.
**Change extraction:**
Compute the DIs from $\boldsymbol{\Delta}^x$ and $\boldsymbol{\Delta}^y$ by using (29).
Calculate the CM with MRF co-segmentation.

## IV. Experiments and Discussions

### A. Datasets and Setting

*1) Datasets:* To assess the capability of CPIT, we conduct comparative experiments on five real-world datasets, as detailed in Table III and Fig. 2. These datasets include: (i) image pairs acquired from different optical sensors (Datasets #1 and #2), and (ii) image pairs acquired from SAR and optical sensors (Datasets #3-#5). These datasets span a wide range of spatial resolutions (from 30 m to 0.52 m), image dimensions (ranging from $343 \times 291$ to $4135 \times 2325$), and various change scenarios. This diversity enables a comprehensive evaluation of each algorithm's generalization ability and robustness under different conditions.

*2) Experimental Setting:* In CPIT, we set $N_S = 2500$ for superpixel co-segmentation, fix $k = \sqrt{N_S}$ for KNN and KFN graphs, and set the weighting parameters of transformation models as $\lambda = 0.1$, $\alpha = 2^{-2}$ for all datasets. For the forward transformation (27), we fix $\beta = 5$; for the backward transformation (28), we set $\beta = 5$ for Datasets #2 and #3, and $\beta = 30$ for Datasets #1, #4 and #5. A detailed discussion of these parameters is provided in Section IV-C.

*3) Evaluation Metric:* To evaluate the DI and CM generated by different HeCD methods, we adopt two groups of evaluation metrics: (i) For DI assessment, we utilize the precision-recall (PR) curves, where the area under the PR curve (AUP) serves as the quantitative indicator; (ii) For CM evaluation, we visualize classification results using different colors to denote true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Meanwhile, quantitative evaluation of CM performance is conducted using overall accuracy (OA), F1-score, and the Kappa coefficient ($\kappa$).

### B. Experimental Results

In our experiments, we compare CPIT against 11 representative HeCD methods, including the image translation-based methods such as FPMS [30], CICM [12], SCASC [49], ACE-Net [36], X-Net [36], SDIR [31], and CAAE [32], as well as the feature transformation-based methods such as IRG-McS [25], AOSG [13], GSGM [14], and USSD [37]. All results were obtained using publicly available codes with default recommended parameters. Among these, FPMS, SCASC, ACE-Net, X-Net, SDIR, CAAE, AOSG, and GSGM generate both forward and backward DIs, similar to CPIT, and are thus considered in the DI performance evaluation. All methods participate in the final CM performance comparison.

*1) Experimental Comparison on DIs:* Qualitative Comparison: Figure 2 shows the forward and backward DIs across five datasets. Overall, a noticeable discrepancy can be observed between the forward and backward DIs produced by the same method, particularly for FPMS, SCASC, AOSG, and GSGM. This phenomenon primarily arises from the inherent asymmetry in the complexity of forward and backward transformations. For instance, transforming an optical image into the SAR domain is not equivalent in complexity to the reverse transformation; similarly, mapping from a structurally simpler image to a more complex one is typically more challenging

TABLE III
HETEROGENEOUS DATASETS.

| Dataset | Sensor (or modality) | Image size (pixels) | Date | Location | Spatial resolution | Change event |
|---|---|---|---|---|---|---|
| #1 | Landsat-5<br>Google Earth | $300 \times 412 \times 1$<br>$300 \times 412 \times 3$ | Sept. 1995<br>July 1996 | Sardinia, Italy | 30 m | Lake expansion |
| #2 | Pleiades<br>WorldView2 | $2000 \times 2000 \times 3$<br>$2000 \times 2000 \times 3$ | May 2012<br>July 2013 | Toulouse, France | 0.52 m | Construction |
| #3 | Radarsat-2<br>Google Earth | $593 \times 921 \times 1$<br>$593 \times 921 \times 3$ | June 2008<br>Sept. 2012 | Shuguang Village, China | 8 m | Construction |
| #4 | Landsat-8<br>Sentinel-1A | $3500 \times 2000 \times 11$<br>$3500 \times 2000 \times 3$ | Jan. 2017<br>Feb. 2017 | Sutter County, USA | $\approx 15$ m | Flooding |
| #5 | QuickBird-2<br>TerraSAR-X | $4135 \times 2325 \times 3$<br>$4135 \times 2325 \times 1$ | July 2006<br>July 2007 | Gloucester, England | 0.65 m | Flooding |

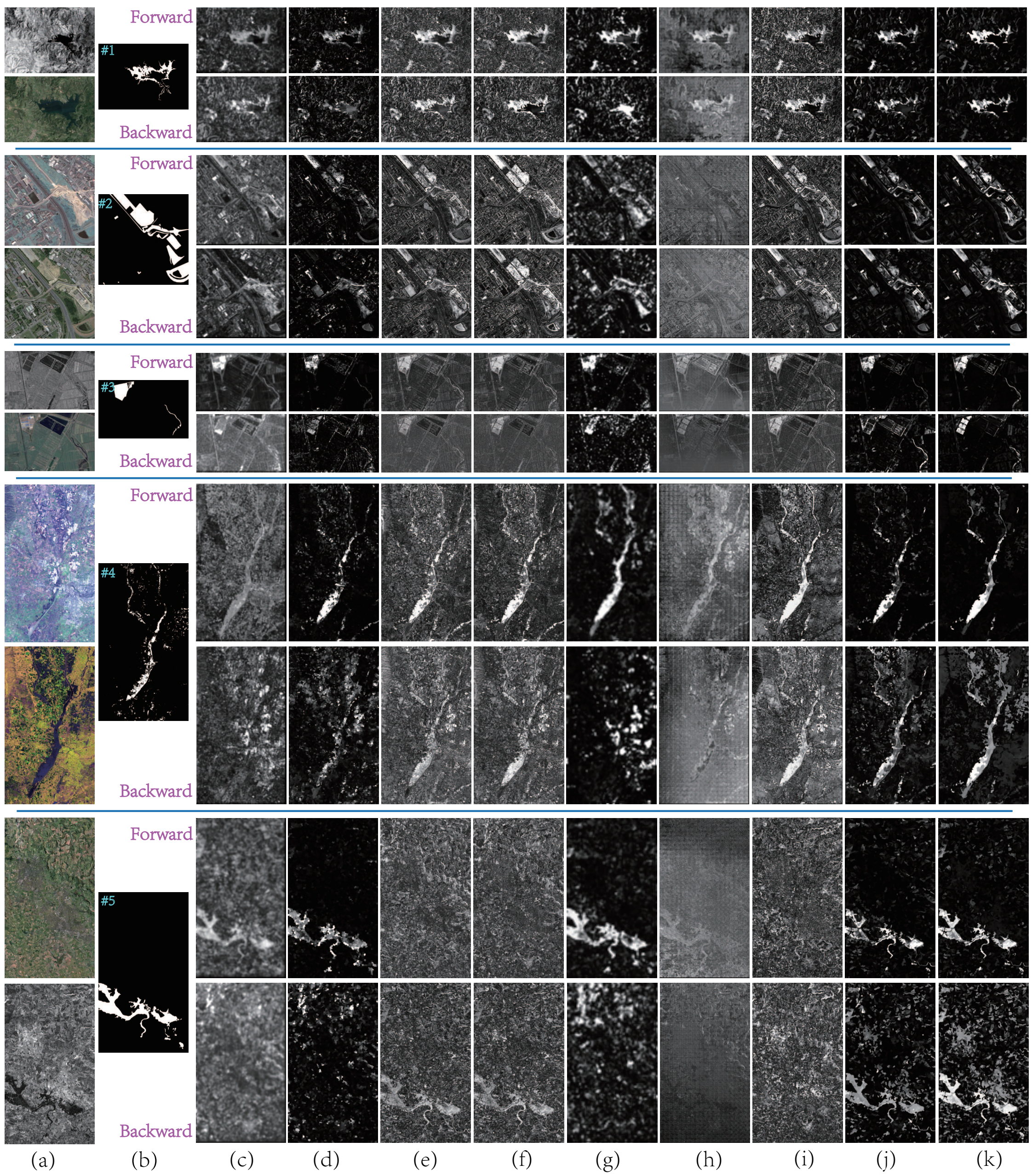
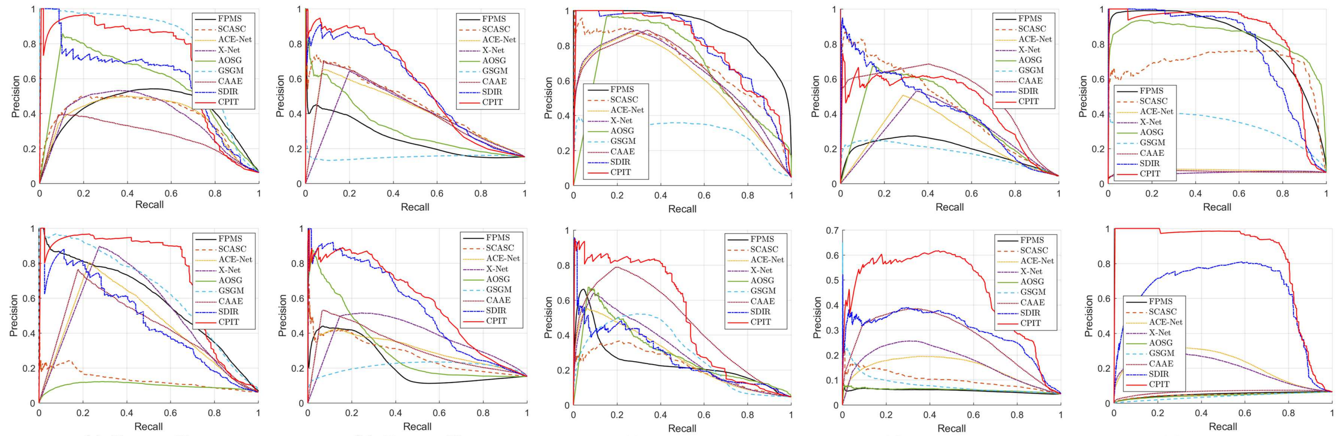



Fig. 2. Forward and backward difference images generated by different methods. From top to bottom, they correspond to Datasets #1 to #5, respectively. (a) pre- and post-event images; (b) ground truth; from (c) to (k) are the difference images generated by (c) FPMS, (d) SCASC, (e) ACE-Net, (f) X-Net, (g) AOSG, (h) GSGM, (i) CAAE, (j) SDIR, and (k) the proposed CPIT.




(a) Dataset #1 (b) Dataset #2 (c) Dataset #3 (d) Dataset #4 (e) Dataset #5

Fig. 3. PR curves of forward (top) and backward (bottom) DIs. From (a) to (e) are the curves obtained from Datasets #1 to #5, respectively.

TABLE IV
THE AUP OF DIS. SCORES WITH THE **BEST** AND **SECOND-BEST** PERFORMANCE ARE SHOWN IN **RED** AND **BLUE**, RESPECTIVELY.

| Methods | Dataset #1 | | Dataset #2 | | Dataset #3 | | Dataset #4 | | Dataset #5 | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Forward | Backward | Forward | Backward | Forward | Backward | Forward | Backward | Forward | Backward | Forward | Backward |
| FPMS [30] | 0.406 | 0.590 | 0.258 | 0.224 | **0.904** | 0.237 | 0.191 | 0.056 | 0.836 | 0.050 | 0.519 | 0.231 |
| SCASC [49] | 0.383 | 0.133 | 0.458 | 0.276 | 0.695 | 0.218 | 0.447 | 0.102 | 0.681 | 0.048 | 0.533 | 0.155 |
| ACE-Net [36] | 0.368 | 0.417 | 0.408 | 0.297 | 0.593 | 0.271 | 0.255 | 0.143 | 0.074 | 0.218 | 0.340 | 0.269 |
| X-Net [36] | 0.363 | 0.474 | 0.376 | 0.366 | 0.630 | 0.302 | 0.292 | 0.168 | 0.062 | 0.198 | 0.345 | 0.301 |
| AOSG [13] | 0.544 | 0.097 | 0.323 | 0.310 | 0.646 | 0.286 | 0.386 | 0.060 | **0.839** | 0.044 | 0.548 | 0.159 |
| GSGM [14] | **0.786** | **0.642** | 0.151 | 0.206 | 0.296 | 0.290 | 0.175 | 0.082 | 0.331 | 0.037 | 0.348 | 0.252 |
| CAAE [32] | 0.266 | 0.406 | 0.416 | 0.310 | 0.617 | **0.458** | **0.512** | 0.265 | 0.068 | 0.064 | 0.376 | 0.300 |
| SDIR [31] | 0.585 | 0.468 | **0.553** | **0.543** | 0.791 | 0.274 | 0.421 | **0.296** | 0.784 | **0.618** | **0.627** | **0.440** |
| proposed CPIT | **0.687** | **0.717** | **0.568** | **0.582** | **0.794** | **0.500** | **0.458** | **0.451** | **0.846** | **0.833** | **0.670** | **0.616** |

than the opposite direction. Such asymmetric behavior has been analyzed in detail in [31], [44] from the perspectives of image structural asymmetry and graph signal processing, respectively. In contrast, the proposed CPIT method effectively mitigates the DI inconsistency problem and yields high-quality DIs in both forward and backward transformations.

Specifically, on Dataset #1, although most methods can generate satisfactory forward DIs, there are substantial differences in the quality of backward DIs, e.g., the backward DIs produced by SCASC and AOSG fail to distinguish the change regions. In Dataset #2, the presence of a greater variety of land-cover types (e.g., roads, buildings, grass, sports fields, construction sites) with highly imbalanced proportions and higher image resolution (0.52 m) imposes significant challenges for HeCD. As a result, some methods, such as FPMS and GSGM, produce DIs in which changed and unchanged regions are severely confused. On Dataset #3, most methods achieve relatively good DI performance; however, some exhibit loss of fine change details, as seen in the backward DIs of ACE-Net, X-Net, and AOSG. Dataset #4 presents challenges due to the large number of land-cover categories and rich change details. On this dataset, CAAE, SDIR, and CPIT demonstrate better performance in both forward and backward DIs. On Dataset #5, only SDIR and CPIT maintain consistency between forward and backward DIs, while the rest show significant divergence, indicating a lack of robustness in those models. Furthermore, it is worth noting that the DIs of CPIT not only consistently highlight the changed regions but also exhibit a degree of sparsity. This sparsity can be attributed to the change-prior constraint incorporated into the transformation model, which encourages the detection of only meaningful changes.

Quantitative Comparison: Figure 3 presents the PR curves of generated DIs, while the corresponding AUP values are summarized in Table IV. Among all the methods, AOSG and SCASC exhibit the most pronounced discrepancies between forward and backward DI performance. For instance, their forward AUPs reach 0.548 and 0.533, whereas their backward AUPs drop sharply to only 0.159 and 0.155, respectively. In contrast, CPIT demonstrates both superior and more consistent performance across forward and backward directions, underscoring its robustness in HeCD tasks. Specifically, CPIT achieves the highest average AUP values in both forward and backward directions with scores of 0.670 and 0.616, which surpass those of the second-best method, SDIR, by 0.043 and 0.176, respectively. This indicates that CPIT can deliver satisfactory change detection results even when coupled with simple thresholding or clustering techniques.

*2) Experimental Comparison on CMs:* Figure 4 shows the binary CMs generated by CPIT and the other 11 comparison methods across all datasets. In these CMs, white, black, red, and blue pixels denote TP, TN, FP, and FN, respectively. The corresponding quantitative evaluation results (OA, $\kappa$ and F1) are summarized in Table V.

According to the visual results displayed in Fig. 4, several approaches exhibit limited robustness on certain datasets. For

TABLE V
QUANTITATIVE MEASURES OF CMS. SCORES WITH THE **BEST** AND **SECOND-BEST** PERFORMANCE ARE SHOWN IN **RED** AND **BLUE**, RESPECTIVELY.

| Methods | Dataset #1 | | | Dataset #2 | | | Dataset #3 | | | Dataset #4 | | | Dataset #5 | | | Average | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | OA | $\kappa$ | F1 | OA | $\kappa$ | F1 | OA | $\kappa$ | F1 | OA | $\kappa$ | F1 | OA | $\kappa$ | F1 | OA | $\kappa$ | F1 |
| FPMS [30] | 0.925 | 0.552 | 0.588 | 0.838 | 0.215 | 0.296 | 0.938 | 0.569 | 0.597 | 0.947 | 0.329 | 0.356 | 0.970 | 0.770 | 0.786 | 0.924 | 0.487 | 0.525 |
| SCASC [49] | 0.947 | 0.593 | 0.621 | 0.892 | 0.464 | 0.516 | 0.979 | 0.741 | 0.751 | 0.960 | 0.479 | 0.499 | 0.973 | 0.774 | 0.788 | 0.950 | 0.610 | 0.635 |
| ACE-Net [36] | 0.958 | 0.674 | 0.696 | 0.870 | 0.397 | 0.466 | 0.960 | 0.610 | 0.631 | 0.911 | 0.367 | 0.406 | 0.895 | 0.365 | 0.417 | 0.919 | 0.483 | 0.523 |
| X-Net [36] | 0.967 | 0.723 | 0.740 | 0.874 | 0.471 | 0.544 | 0.968 | 0.672 | 0.689 | 0.931 | 0.428 | 0.461 | 0.876 | 0.270 | 0.331 | 0.923 | 0.513 | 0.553 |
| AOSG [13] | 0.874 | 0.397 | 0.451 | 0.783 | 0.215 | 0.344 | 0.919 | 0.452 | 0.487 | 0.898 | 0.329 | 0.372 | 0.911 | 0.533 | 0.574 | 0.877 | 0.385 | 0.446 |
| GSGM [14] | 0.958 | 0.677 | 0.699 | 0.528 | 0.078 | 0.294 | 0.928 | 0.396 | 0.432 | 0.858 | 0.233 | 0.286 | 0.853 | 0.166 | 0.237 | 0.825 | 0.310 | 0.390 |
| CAAE [32] | 0.930 | 0.534 | 0.569 | 0.867 | 0.380 | 0.451 | 0.950 | 0.563 | 0.588 | 0.931 | 0.460 | 0.492 | 0.858 | 0.016 | 0.059 | 0.907 | 0.384 | 0.432 |
| SDIR [31] | 0.964 | 0.672 | 0.691 | **0.919** | **0.638** | **0.683** | **0.982** | **0.810** | **0.819** | 0.958 | 0.485 | 0.507 | **0.982** | **0.843** | **0.853** | **0.961** | **0.690** | **0.711** |
| CICM [12] | 0.943 | 0.451 | 0.481 | 0.867 | 0.270 | 0.321 | 0.975 | 0.749 | 0.763 | 0.899 | 0.081 | 0.131 | 0.898 | 0.372 | 0.423 | 0.916 | 0.385 | 0.424 |
| IRG-McS [25] | **0.971** | **0.739** | **0.754** | 0.882 | 0.418 | 0.475 | 0.983 | 0.794 | 0.804 | **0.959** | **0.490** | **0.512** | 0.971 | 0.740 | 0.755 | 0.953 | 0.636 | 0.660 |
| USSD [37] | 0.965 | 0.726 | 0.745 | 0.694 | 0.214 | 0.379 | 0.942 | 0.376 | 0.406 | 0.896 | 0.384 | 0.425 | 0.578 | 0.084 | 0.187 | 0.815 | 0.357 | 0.428 |
| proposed CPIT | **0.972** | **0.744** | **0.759** | **0.919** | **0.630** | **0.674** | **0.984** | **0.816** | **0.824** | **0.956** | **0.533** | **0.556** | **0.982** | **0.850** | **0.860** | **0.963** | **0.715** | **0.734** |

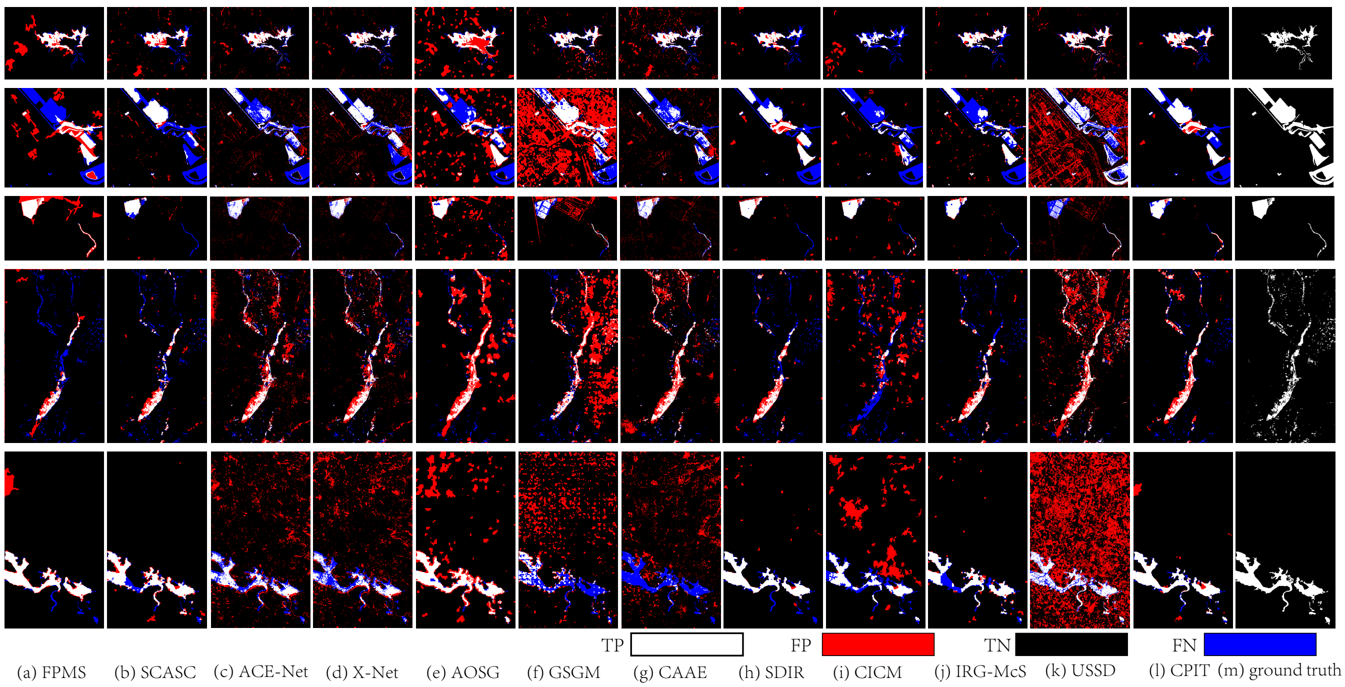



Fig. 4. Change maps generated by different methods. From top to bottom, they correspond to Datasets #1 to #5, respectively. From (a) to (l) are the change maps generated by (a) FPMS, (b) SCASC, (c) ACE-Net, (d) X-Net, (e) AOSG, (f) GSGM, (g) CAAE, (h) SDIR, (i) CICM, (j) IRG-McS, (k) USSD, and (l) the proposed CPIT. In the change map, white, **red**, black, and **blue** mark true positives (TP), **false positives (FP)**, true negatives (TN), and **false negatives (FN)**, respectively.

instance, AOSG and CAAE produce a high number of false alarms on Dataset #1. GSGM and USSD suffer from numerous false alarms on Dataset #2, while FPMS, SCASC, and CICM exhibit a large number of missed detections. In Dataset #3, most methods achieve satisfactory performance, though some fail to preserve fine-grained change details such as GSGM and SCASC. For Dataset #4, false alarms and missed detections are common across most methods, yet CPIT yields noticeably fewer errors. On Dataset #5, USSD shows extensive false alarms, while CAAE experiences a significant number of missed detections. According to the evaluation metrics listed in Table V, CPIT achieves the best results on Datasets #1, #3, #4, and #5, and ranks second on Dataset #2. These findings further validate the robustness and effectiveness of CPIT in handling diverse and challenging change detection scenarios.

## C. Discussions

*1) Ablation Study:* The proposed CPIT mainly consists of four components: semantic change priors (SeP), structural consistency priors (StP), sparsity prior (SpP), and MRF fusion. To investigate the contribution of each component, we construct four variants by removing one component at a time from the full model. The average quantitative results, including the AUP of the DIs and the OA, $\kappa$, and F1 of the final CMs, are reported in Table VI.

It can be observed that removing either the semantic or structural component leads to a clear performance drop, indicating that both priors play important roles in CPIT. Specifically, compared with the full CPIT, removing the semantic change priors decreases the average F1 from 0.734 to 0.653, while removing the structural consistency priors reduces it to 0.662. These results suggest that the semantic and structural priors are complementary, and that their joint use yields

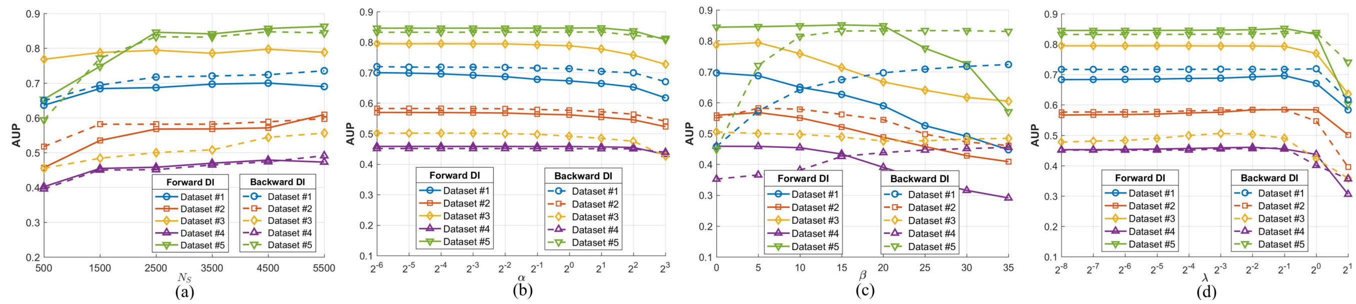



Fig. 5. Sensitivity analysis of parameters (a) $N_S$, (b) $\alpha$, (c) $\beta$, and (d) $\lambda$.

TABLE VI
ABLATION STUDY OF CPIT MEASURED BY AVERAGE QUANTITATIVE SCORES. SEP, STP, AND SPP DENOTE SEMANTIC CHANGE PRIORS, STRUCTURAL CONSISTENCY PRIORS, AND SPARSITY PRIOR, RESPECTIVELY.

| Settings | AUP of DIs | | Final CMs | | |
|---|---|---|---|---|---|
| | Forward | Backward | OA | $\kappa$ | F1 |
| CPIT w/o SeP | 0.626 | 0.431 | 0.953 | 0.628 | 0.653 |
| CPIT w/o StP | 0.630 | 0.446 | 0.955 | 0.637 | 0.662 |
| CPIT w/o SpP | 0.648 | 0.592 | 0.961 | 0.694 | 0.715 |
| CPIT w/o MRF | 0.670 | 0.616 | 0.962 | 0.705 | 0.723 |
| Full CPIT | 0.670 | 0.616 | 0.963 | 0.715 | 0.734 |

more reliable change inference than either of them alone. In addition, removing the change sparsity prior or the MRF fusion also degrades the results, although to a smaller extent. Since MRF fusion is only used as a post-processing step, it does not affect the image transformation process or the generated DIs, but still reduces the accuracy of the final CMs. Overall, the full CPIT achieves the best performance, which verifies the effectiveness of combining all four components.

*2) Parameter Analysis:* The main hyperparameters in CPIT include the superpixel number $N_S$, along with the weighting coefficients $\alpha$, $\beta$, and $\lambda$ of the transformation models in (27) and (28).

The number of superpixels $N_S$ should be selected by considering both image spatial resolution and real-time requirements of the HeCD task. Increasing $N_S$ leads to finer superpixel segmentation, which enhances the detail and precision of HeCD. However, this also introduces greater computational burden, as will be elaborated later in this subsection. On the other hand, a smaller $N_S$ reduces computational demands but risks overly coarse segmentation. In such cases, a single superpixel may span multiple land-cover types, compromising internal consistency and potentially reducing detection accuracy. As shown in Fig. 5(a), the AUP increases rapidly as $N_S$ grows from 500 to 2500, and then gradually stabilizes. This robustness arises because CPIT emphasizes pairwise relationship-based structural modeling rather than individual superpixel features, together with neighborhood expansion strategies in high-order graph construction, which enhance the representational capacity of the graph. To strike a balance between computational efficiency and segmentation quality, this paper sets $N_S \approx 2500$.

The parameters $\alpha$, $\beta$, and $\lambda$ serve as weighting coefficients in the transformation models, controlling the relative influence of the change semantic inconsistency term, the structural consistency term, and the sparsity regularization term, respectively. Figure 5 shows the sensitivity performance of weighting parameters, where the AUP values of DIs are employed as the as evaluation indicators. In Fig. 5(b), we vary the parameter $\alpha$ from $2^{-6}$ to $2^3$ to evaluate the sensitivity of CPIT to the semantic inconsistency constraint weight. The results indicate that the proposed CPIT is relatively robust to this parameter, and achieves consistently good performance across datasets when $\alpha = 2^{-2}$. Therefore, for simplicity, we set $\alpha = 2^{-2}$ in CPIT. As for parameter $\beta$, which controls the structural consistency of dissimilarity relationships, Fig. 5(c) shows that the forward and backward transformation models exhibit different sensitivities to $\beta$. According to Fig. 5(c), we set $\beta = 5$ for the forward transformation on all datasets, as well as for the backward transformation on Datasets #2 and #3, while $\beta = 30$ is used for the backward transformation on Datasets #1, #4, and #5. This discrepancy arises from the structural differences between the pre-event and post-event images in different datasets. When the target image already exhibits low-pass characteristics on the similarity-based KNN graph constructed from the source image, the KNN graph model alone becomes insufficient to distinguish the source and target images. In this case, a larger $\beta$ is needed to strengthen the dissimilarity constraint imposed by the KFN graph model. Therefore, different strengths of structural constraints may be required when transforming from the pre-event image domain to the post-event image domain and vice versa, as discussed in [31].

The parameter $\lambda$ serves as the weighting factor for the change sparsity regularization term. Conceptually, $\lambda$ should be adjusted in accordance with the expected proportion of changed regions—smaller values are preferable when the change area is known to be minimal. As illustrated in Fig. 5(d), the AUP values for both forward and backward DIs remain relatively stable across a wide range of $\lambda$ values, from $2^{-8}$ to $2^1$. Based on these results, we select $\lambda = 0.1$ (approximately $2^{-3.3}$) for all experiments to ensure a balance between performance and simplicity.

To further evaluate the joint sensitivity of $\alpha$, $\beta$, and $\lambda$, we additionally evaluate 12 representative parameter combinations, and the corresponding AUP results are reported in Table VII. The results show that CPIT remains relatively stable under these joint parameter variations. Moreover, the

TABLE VII

AUP OF THE DIs GENERATED BY CPIT ON DIFFERENT DATASETS UNDER DIFFERENT JOINT SETTINGS OF $\{\alpha, \beta, \lambda\}$. THE RECOMMENDED SETTINGS ARE HIGHLIGHTED IN BOLD AND UNDERLINED.

| $\{\alpha, \beta, \lambda\}$ | Forward AUP on different datasets | | | | | Backward AUP on different datasets | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | #1 | #2 | #3 | #4 | #5 | #1 | #2 | #3 | #4 | #5 |
| $\{0.25, 5, 0.01\}$ | 0.684 | 0.568 | 0.795 | 0.453 | 0.845 | 0.571 | 0.577 | 0.483 | 0.366 | 0.721 |
| $\{0.25, 5, 0.1\}$ | **<u>0.687</u>** | **<u>0.568</u>** | **<u>0.794</u>** | **<u>0.458</u>** | **<u>0.846</u>** | 0.573 | **<u>0.582</u>** | **<u>0.500</u>** | 0.365 | 0.720 |
| $\{1, 5, 0.01\}$ | 0.676 | 0.562 | 0.787 | 0.458 | 0.845 | 0.576 | 0.575 | 0.490 | 0.364 | 0.721 |
| $\{1, 5, 0.1\}$ | 0.673 | 0.562 | 0.788 | 0.458 | 0.846 | 0.577 | 0.577 | 0.492 | 0.364 | 0.723 |
| $\{0.25, 15, 0.01\}$ | 0.625 | 0.520 | 0.714 | 0.429 | 0.850 | 0.672 | 0.556 | 0.480 | 0.424 | 0.831 |
| $\{0.25, 15, 0.1\}$ | 0.626 | 0.521 | 0.713 | 0.433 | 0.851 | 0.674 | 0.562 | 0.489 | 0.425 | 0.833 |
| $\{1, 15, 0.01\}$ | 0.620 | 0.522 | 0.702 | 0.441 | 0.851 | 0.676 | 0.551 | 0.488 | 0.426 | 0.833 |
| $\{1, 15, 0.1\}$ | 0.619 | 0.521 | 0.701 | 0.438 | 0.850 | 0.671 | 0.550 | 0.487 | 0.425 | 0.835 |
| $\{0.25, 30, 0.01\}$ | 0.487 | 0.426 | 0.617 | 0.315 | 0.725 | 0.717 | 0.469 | 0.472 | 0.451 | 0.832 |
| $\{0.25, 30, 0.1\}$ | 0.491 | 0.428 | 0.617 | 0.316 | 0.726 | **<u>0.717</u>** | 0.472 | 0.482 | **<u>0.451</u>** | **<u>0.833</u>** |
| $\{1, 30, 0.01\}$ | 0.490 | 0.423 | 0.598 | 0.319 | 0.731 | 0.714 | 0.464 | 0.479 | 0.452 | 0.832 |
| $\{1, 30, 0.1\}$ | 0.491 | 0.424 | 0.596 | 0.319 | 0.727 | 0.713 | 0.465 | 0.480 | 0.451 | 0.833 |

recommended settings are either optimal or near-optimal on most datasets, indicating that the practical tuning burden is manageable. In practice, $\alpha$, $\beta$, and $\lambda$ are tuned by keeping the corresponding regularization terms numerically balanced, so that no single term dominates the optimization excessively. Based on our experiments, $\alpha$ and $\lambda$ are relatively stable across different datasets and can therefore be fixed to default values once a reasonable range is identified. By contrast, $\beta$ is the main tunable parameter, especially in the backward transformation model, which can be selected from a small candidate set such as $\{5, 15, 30\}$.

*3) Convergence Behavior:* Because the overall objective of CPIT is nonconvex, among the ADMM subproblems, only the $\mathbf{\Delta}^x, \mathbf{\Delta}^y$-subproblems and the $\mathbf{R}_1, \mathbf{R}_2$-subproblems are convex, while the remaining two subproblems are nonconvex (see the supplementary materials for details). Therefore, strict theoretical convergence guarantees or global optimality cannot be established for CPIT. In practice, however, the required number of ADMM iterations is small and stable, mainly due to the following two strategies.

Initialization strategy. In CPIT, the initialization is not random, but derived from the structural consistency prior of the transformation model. From the graph signal processing perspective, taking the forward transformation as an example, the transformed image $\tilde{\mathbf{Y}}'$ is expected to behave as a low-pass graph signal on the KNN graph $G_{hn}^{t1}$ constructed from $\tilde{\mathbf{X}}$. After column-normalizing the graph weight matrix $\mathbf{W}_{hn}^{t1}$, we have $\mathbf{X} \approx \mathbf{X}\mathbf{W}_{hn}^{t1}$, which indicates that each superpixel can be approximately reconstructed as a weighted average of its graph neighbors. Based on the structural consistency of (25), $\mathbf{Y}'$ is expected to satisfy a similar property, i.e., $\mathbf{Y}' \approx \mathbf{Y}'\mathbf{W}_{hn}^{t1}$. Considering that $\mathbf{\Delta}^y = \mathbf{Y}' - \mathbf{Y}$ and $\mathbf{\Delta}^y$ is sparse, we adopt

$$\mathbf{Y}'^{(0)} = \mathbf{Y}\mathbf{W}_{hn}^{t1}, \quad \mathbf{\Delta}^{y(0)} = \mathbf{Y}\mathbf{W}_{hn}^{t1} - \mathbf{Y} \tag{30}$$

as the initialization. This provides a meaningful starting point and improves the practical stability of the nonconvex optimization.

Early-stopping strategy. In our experiments, according to the empirical convergence behavior shown in Fig. 6, an early-stopping criterion based on the relative change $\mathrm{RelDiff}\,(i) < 10^{-2}$ is adopted, and the maximum number of ADMM iterations is set to 20, where $\mathrm{RelDiff}\,(i) = \frac{\|\mathbf{\Delta} - \mathbf{\Delta}_{\text{old}}\|_F}{\|\mathbf{\Delta}\|_F}$. This strategy effectively prevents unnecessary iterations and ensures computational efficiency in practice.

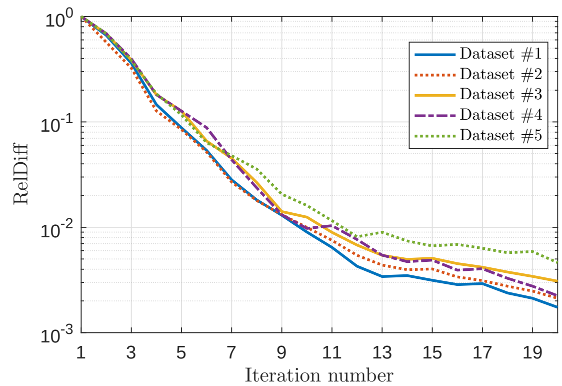



Fig. 6. Convergence performance of CPIT.

Overall, although CPIT is optimized under a nonconvex setting, the proposed initialization and early-stopping strategies make the optimization empirically stable and efficient, as also evidenced by the convergence curves in Fig. 6.

*4) Computational Complexity:* The CPIT framework consists of three stages: preprocessing, relationship representation, and image transformation via optimization of (27) and (28). As detailed in the supplementary material, (i) superpixel segmentation [48] and feature extraction require $\mathcal{O}(MN)$ and $\mathcal{O}(MNB\log(MN/N_S))$ respectively, with $B = B_x + B_y$; (ii) relationship representation requires $\mathcal{O}(N_S^2 \log N_S)$; (iii) image transformation requires $\mathcal{O}(N_S^2 B)$.

The execution times (in seconds) of conventional comparison methods on Dataset #2 (image size: $2000 \times 2000 \times 3$) are summarized as follows: FPMS: 92.4, SCASC: 26.7, AOSG: 206.9, GSGM: 788.8, SDIR: 49.1, CICM: 113.0, IRG-McS: 20.5, USSD: 795.2, and CPIT: 52.3. Specifically, FPMS and CICM were implemented in C++ and executed on a Linux workstation with an Intel Xeon Silver 4110 processor, whereas SCASC, AOSG, GSGM, SDIR, IRG-McS, USSD, and CPIT were conducted in MATLAB on a Windows desktop configured with an Intel Core i7-8700K CPU. These results indicate that CPIT offers a reasonable balance between computational cost and performance. In addition, the forward and backward image transformations take 41.8 seconds, making them the most time-consuming component in CPIT (additional details are provided in the supplementary material). To further improve efficiency, future work will explore more advanced optimization strategies for solving the underlying image transformation models. Moreover, replacing the full-graph models (KNN and KFN graphs) with anchor-graph models could further reduce graph complexity. For large-scale images, region-wise parallel processing is also a promising strategy to further improve scalability.

## V. Conclusion

This work focuses on the exploration and utilization of change priors in HeCD. Beyond conventional spatial change priors (including sparsity, smoothness, and consistency constraints), we delve into semantic change priors. Starting from the formal definition of the HeCD problem, we establish the intrinsic connections between pairwise object relationships, change labels, and change semantics. Based on these connections, we derive semantic consistency and inconsistency rules, forming the basis of our semantic change priors. We then formulate corresponding constraints based on these semantic priors, which exhibit low-pass and high-pass spectral characteristics on graph models, respectively. By integrating change semantic constraints with spatial sparsity and image transformation constraints, we develop a change-guided HeCD model, which yields more accurate change detection results. Nevertheless, since the current implementation constructs fixed KNN/KFN graphs using simple superpixel features, inaccurate graph connections may occur in complex scenarios. Incorporating more discriminative feature representations, learning-based distance metrics, or graph neural networks could further improve the robustness and adaptivity of the proposed change-prior-guided framework. As a possible extension, these semantic change priors may also be incorporated into other HeCD frameworks, such as feature-comparison-based or statistical-modeling-based methods, to further enhance performance and interpretability. More generally, because they are derived directly from the intrinsic nature of change detection, they are not restricted to HeCD and may also be applicable to homogeneous and semantic change detection.

## References


[1] Z. Lv, H. Huang, X. Li, M. Zhao, J. A. Benediktsson, W. Sun, and N. Falco, "Land cover change detection with heterogeneous remote sensing images: Review, progress, and perspective," *Proceedings of the IEEE*, vol. 110, no. 12, pp. 1976–1991, 2022. 1

[2] Y. Sun, L. Lei, D. Guan, G. Kuang, and L. Liu, "Graph signal processing for heterogeneous change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–23, 2022. 1

[3] T. Liu, M. Zhang, M. Gong, Q. Zhang, F. Jiang, H. Zheng, and D. Lu, "Commonality feature representation learning for unsupervised multimodal change detection," *IEEE Transactions on Image Processing*, vol. 34, pp. 1219–1233, 2025. 1, 2

[4] T. Liu, J. Xu, T. Lei, Y. Wang, X. Du, W. Zhang, Z. Lv, and M. Gong, "Aekan: Exploring superpixel-based autoencoder kolmogorov-arnold network for unsupervised multimodal change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–14, 2025. 1

[5] H. Chen, J. Song, O. Dietrich, C. Broni-Bediako, W. Xuan, J. Wang, X. Shao, Y. Wei, J. Xia, C. Lan *et al.*, "Bright: A globally distributed multimodal building damage assessment dataset with very-high-resolution for all-weather disaster response," *arXiv preprint arXiv:2501.06019*, 2025. 1

[6] H. Chen, N. Yokoya, and M. Chini, "Fourier domain structural relationship analysis for unsupervised multimodal change detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 198, pp. 99–114, 2023. 1, 2

[7] H. Zhang, G. Ma, D. Wang, and Y. Zhang, "M3icnet: A cross-modal resolution preserving building damage detection method with optical and sar remote sensing imagery and two heterogeneous image disaster datasets," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 221, pp. 224–250, 2025. 1

[8] B. Zhao, H. Sui, J. Liu, W. Shi, W. Wang, C. Xu, and J. Wang, "Flood inundation monitoring using multi-source satellite imagery: a knowledge transfer strategy for heterogeneous image change detection," *Remote Sensing of Environment*, vol. 314, p. 114373, 2024. 1

[9] Z. Wang, X. Wang, W. Wu, and G. Li, "Continuous change detection of flood extents with multisource heterogeneous satellite image time series," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–18, 2023. 1

[10] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 179, pp. 14–34, 2021. 1

[11] T. Liu, Y. Pu, T. Lei, J. Xu, M. Gong, L. He, and A. K. Nandi, "Hierarchical feature alignment-based progressive addition network for multimodal change detection," *Pattern Recognition*, vol. 162, p. 111355, 2025. 1

[12] R. Touati, "Détection de changement en imagerie satellitaire multimodale," Ph.D. dissertation, Université de Montréal, 2019. 1, 8, 11

[13] T. Han, Y. Tang, B. Zou, and H. Feng, "Unsupervised multimodal change detection based on adaptive optimization of structured graph," *International Journal of Applied Earth Observation and Geoinformation*, vol. 126, p. 103630, 2024. 1, 8, 10, 11

[14] T. Han, Y. Tang, Y. Chen, B. Zou, and H. Feng, "Global structure graph mapping for multimodal change detection," *International Journal of Digital Earth*, vol. 17, no. 1, p. 2347457, 2024. 1, 8, 10, 11

[15] C. Wu, B. Du, X. Cui, and L. Zhang, "A post-classification change detection method based on iterative slow feature analysis and bayesian soft fusion," *Remote Sensing of Environment*, vol. 199, pp. 241–255, 2017. 1

[16] Z.-g. Liu, G. Mercier, J. Dezert, and Q. Pan, "Change detection in heterogeneous remote sensing images based on multidimensional evidential reasoning," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 1, pp. 168–172, 2014. 1

[17] L. Wan, Y. Xiang, and H. You, "An object-based hierarchical compound classification method for change detection in heterogeneous optical and SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9941–9959, 2019. 1

[18] R. Touati and M. Mignotte, "An energy-based model encoding nonlocal pairwise pixel interactions for multisensor change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 1046–1058, 2018. 1

[19] R. Touati, M. Mignotte, and M. Dahmane, "Multimodal change detection in remote sensing images using an unsupervised pixel pairwise-based markov random field model," *IEEE Transactions on Image Processing*, vol. 29, pp. 757–767, 2020. 1

[20] M. Mignotte, "Mrf models based on a neighborhood adaptive class conditional likelihood for multimodal change detection," *AI, Computer Science and Robotics Technology*, Mar 2022. 1

[21] Y. Sun, L. Lei, D. Guan, G. Kuang, Z. Li, and L. Liu, "Locality preservation for unsupervised multimodal change detection in remote sensing imagery," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2025. 1, 2

[22] W. Jing, H. Bai, B. Song, W. Ni, J. Wu, and Q. Wang, "Hetecd: Feature consistency alignment and difference mining for heterogeneous remote sensing image change detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 223, pp. 317–327, 2025. 1

[23] Z. Yang, X. Wang, H. Lin, M. Li, and M. L. and, "Cross-modal feature interaction network for heterogeneous change detection," *Geo-spatial Information Science*, vol. 0, no. 0, pp. 1–22, 2025. 1

[24] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 799–812, 2015. 1

[25] Y. Sun, L. Lei, D. Guan, and G. Kuang, "Iterative robust graph for unsupervised change detection of heterogeneous remote sensing images," *IEEE Transactions on Image Processing*, vol. 30, pp. 6277–6291, 2021. 1, 2, 8, 11

[26] Y. Sun, L. Lei, X. Li, X. Tan, and G. Kuang, "Structure consistency-based graph for unsupervised change detection with homogeneous and heterogeneous remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–21, 2022. 1

[27] Z. Zhang, C. Liu, F. Hao, and Z. Liu, "Style transfer-based unsupervised change detection from heterogeneous images," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 1–15, 2025. 2

[28] J. Li, M. Wu, L. Lin, Q. Yuan, and H. Shen, "Glcd-da: Change detection from optical and sar imagery using a global-local network with diversified attention," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 226, pp. 396–414, 2025. 2

[29] Z. Liu, G. Li, G. Mercier, Y. He, and Q. Pan, "Change detection in heterogenous remote sensing images via homogeneous pixel transformation," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1822–1834, 2018. 2

[30] M. Mignotte, "A fractal projection and markovian segmentation-based approach for multimodal change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 8046–8058, 2020. 2, 8, 10, 11

[31] Y. Sun, L. Lei, Z. Li, and G. Kuang, "Similarity and dissimilarity relationships based graphs for multimodal change detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 208, pp. 70–88, 2024. 2, 4, 8, 10, 11, 12

[32] L. T. Luppino, M. A. Hansen, M. Kampffmeyer, F. M. Bianchi, G. Moser, R. Jenssen, and S. N. Anfinsen, "Code-aligned autoencoders for unsupervised change detection in multimodal remote sensing images," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 1, pp. 60–72, 2024. 2, 8, 10, 11

[33] J. G. Vinholi, M. Chini, A. Amziane, P. Matgen, and R. Machado, "Optical image translation using diffusion models in support of heterogeneous change detection," in *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2024, pp. 7231–7234. 2

[34] J.-J. Wang, N. Dobigeon, M. Chabert, D.-C. Wang, T.-Z. Huang, and J. Huang, "Cd-gan: A robust fusion-based generative adversarial network for unsupervised remote sensing change detection with heterogeneous sensors," *Information Fusion*, p. 102313, 2024. 2

[35] C. Li, G. Li, Z. Wang, X. Wang, and P. K. Varshney, "Comic: An unsupervised change detection method for heterogeneous remote sensing images based on copula mixtures and cycle-consistent adversarial networks," *Information Fusion*, p. 102240, 2024. 2

[36] L. T. Luppino, M. Kampffmeyer, F. M. Bianchi, G. Moser, S. B. Serpico, R. Jenssen, and S. N. Anfinsen, "Deep image translation with an affinity-based change prior for unsupervised multimodal change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–22, 2022. 2, 8, 10, 11

[37] L. Zhu, W. Sun, D. Fan, H. Xing, and X. Liu, "Unsupervised spatial self-similarity difference-based change detection method for multi-source heterogeneous images," *Pattern Recognition*, vol. 149, p. 110237, 2024. 2, 8, 11

[38] J. F. Florez-Ospina, H. D. Benitez-Restrepo, and D. A. Jimenez-Sierra, "Change detection on graphs: Exploiting graph structure from bi-temporal satellite imagery," in *Signal and Image Processing for Remote Sensing*. CRC Press, 2024, pp. 369–389. 2

[39] T. Han, Y. Tang, Y. Chen, X. Yang, Y. Guo, and S. Jiang, "Sdc-gae: Structural difference compensation graph autoencoder for unsupervised multimodal change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–16, 2024. 2

[40] J. F. Florez-Ospina, D. A. Jimenez-Sierra, H. D. Benitez-Restrepo, and G. R. Arce, "Exploiting variational inequalities for generalized change detection on graphs," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023. 2, 7

[41] J. Wang, L. Yan, J. Yang, H. Xie, Q. Yuan, P. Wei, Z. Gao, C. Zhang, and P. M. Atkinson, "Macon: A generic self-supervised framework for unsupervised multimodal change detection," *IEEE Transactions on Image Processing*, vol. 34, pp. 1485–1500, 2025. 2

[42] E. Zhang, H. Zong, X. Li, M. Feng, and J. Ren, "Icsf: Integrating inter-modal and cross-modal learning framework for self-supervised heterogeneous change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–16, 2025. 2

[43] H. Chen, N. Yokoya, C. Wu, and B. Du, "Unsupervised multimodal change detection based on structural relationship graph representation learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2022. 2

[44] Y. Sun, L. Lei, L. Liu, and G. Kuang, "Structural regression fusion for unsupervised multimodal change detection," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–1, 2023. 2, 10

[45] W. Li, X. Wang, G. Li, B. Geng, and P. K. Varshney, "A copula-guided in-model interpretable neural network for change detection in heterogeneous remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–1, 2025. 4

[46] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS Journal of photogrammetry and remote sensing*, vol. 80, pp. 91–106, 2013. 4

[47] C.-A. Deledalle, L. Denis, and F. Tupin, "How to compare noisy patches? patch similarity beyond gaussian noise," *International journal of computer vision*, vol. 99, pp. 86–102, 2012. 4

[48] Z. Ban, J. Liu, and L. Cao, "Superpixel segmentation using gaussian mixture model," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4105–4117, 2018. 4, 13

[49] Y. Sun, L. Lei, D. Guan, M. Li, and G. Kuang, "Sparse-constrained adaptive structure consistency-based unsupervised image regression for heterogeneous remote-sensing change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022. 5, 7, 8, 10, 11

[50] F. Heider, "Attitudes and cognitive organization," *The Journal of psychology*, vol. 21, no. 1, pp. 107–112, 1946. 5, 6

[51] D. Cartwright and F. Harary, "Structural balance: a generalization of heider's theory." *Psychological review*, vol. 63, no. 5, p. 277, 1956. 5, 6

[52] Y. Zheng, S. Liu, and L. Bruzzone, "Scribble-guided structural regression fusion for multimodal remote sensing change detection," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–1, 2025. 7



**Yuli Sun** received the Ph.D. degree in Information and Communication Engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2023. He is currently a lecturer with the College of Electronic Science and Technology in NUDT. His research interests cover remote sensing image processing and multitemporal image analysis. He has published over 40 journal papers in IEEE TIP, IEEE TNNLS, IEEE TGRS, IJCV, ISPRS P&RS, Pattern Recognition, etc.



**Lin Lei** received the Ph.D. degree in Information and Communication Engineering from the National University of Defense Technology, Changsha, China, in 2008. She is currently a Professor with the College of Electronic Science and Technology, National University of Defense Technology. Her research interests include computer vision, remote sensing image interpretation and data fusion.

**Gangyao Kuang** (Senior Member, IEEE) received B.S. and M.S. degrees in geophysics from the Central South University of Technology, Changsha, China, in 1988 and 1991, and the Ph.D. degree in communication and information from the National University of Defense Technology, Changsha, China, in 1995.

He is currently a Professor at the College of Electronic Science and Technology, National University of Defense Technology. His research interests include remote sensing, SAR image processing, change detection, SAR ground moving target indication, and classification with polarimetric SAR images.