

## Similarity and dissimilarity relationships based graphs for multimodal change detection

Yuli Sun <sup>a,b</sup>, Lin Lei <sup>c,\*</sup>, Zhang Li <sup>a,b</sup>, Gangyao Kuang <sup>c</sup>

<sup>a</sup> College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, The People's Republic of China

<sup>b</sup> Hunan Provincial Key Laboratory of Image Measurement and Vision Navigation, Changsha 410073, The People's Republic of China

<sup>c</sup> College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, The People's Republic of China



### ARTICLE INFO

#### Keywords:

Multimodal change detection  
Dissimilarity relationship  
k-farthest neighbors  
k-nearest neighbors  
Image regression

### ABSTRACT

Multimodal change detection (CD) is an increasingly interesting yet highly challenging subject in remote sensing. To facilitate the comparison of multimodal images, some image regression methods transform one image to the domain of the other image, allowing for images comparison in the same domain as in unimodal CD. In this paper, we begin by analyzing the limitations of previous image structure based regression models that only rely on similarity relationships. Then, we highlight the significance of incorporating dissimilarity relationships as a complementary approach to more comprehensively characterize and utilize the image structure. In light of this, we propose a novel method for multimodal CD called Similarity and Dissimilarity induced Image Regression (SDIR). Specifically, SDIR construct a similarity based k-nearest neighbors (KNN) graph and a dissimilarity based k-farthest neighbors (KFN) graph, where the former mainly characterizes the low-frequency information and the latter captures the high-frequency information in spectral domain. Therefore, the proposed SDIR that incorporates similarity (low-frequency) and dissimilarity (high-frequency) relationships enables to achieve better regression performance. After completing the image regression, we utilize a Markovian based fusion segmentation model to combine the change fusion and change extraction processes for improving the final CD accuracy. The proposed method's effectiveness is demonstrated through experiments on six real datasets and compared with eleven advanced and widely used methods, achieving 5.6% improvements in the average Kappa coefficient. The source code is accessible at <https://github.com/yulisun/SDIR>.

## 1. Introduction

### 1.1. Background

Remote sensing change detection (CD) is a technique used to extract information about changes on the Earth's surface by comparing multiple images taken at the same geographical location, but at different times (before and after an event). SINGH (1989). CD has been widely utilized in various fields such as disaster rescue, urban planning and environmental monitoring (Lv et al., 2021).

Based on the compared remote sensing images, CD can be classified into two categories: unimodal CD and multimodal CD (Lv et al., 2022b). Unimodal CD, also known as homogeneous CD, refers to the use of pre- and post-event images obtained by sensors with the same type and identical sensor parameters. Unimodal CD has been extensively investigated and is the main focus of prior CD research (Li et al., 2020). According to the sensor type, unimodal CD can also be subdivided into CD of

multipletspectral (Wu et al., 2023; Chen et al., 2023a), hyperspectral (Liu et al., 2022a) and synthetic aperture radar (SAR) images (Zhang et al., 2022, 2021). Due to the rapid advancements in remote sensing technology and a more open access mechanism to remote sensing data, a vast amount of data from different sensors with different imaging mechanisms can be used by CD to detect changes (Ferraris et al., 2020), forming the basis of multimodal CD, also known as heterogeneous CD (Lv et al., 2022b).

Compared to traditional unimodal CD, multimodal CD offers two distinct advantages (Sun et al., 2022b). Firstly, in the situations requiring emergency response such as earthquakes, landslides, explosions, etc., where unimodal images are unavailable because of the uncooperative imaging conditions (e.g., light, weather) or satellite revisit cycle limitations, multimodal CD can utilize any available pre- and post-event images for rapidly acquiring change information (Sun et al., 2021a). Secondly, in long-term dynamic monitoring processes, multimodal CD

\* Corresponding author.

E-mail address: [alaleilin@163.com](mailto:alaleilin@163.com) (L. Lei).

can incorporate images from various sources to enhance temporal resolution, especially being able to make full use of early remote sensing data (such as images acquired from outdated sensors).

## 1.2. Related work

Unimodal CD compares two homogeneous images directly to extract changes, whereas in multimodal CD, the representations of the same object in pre- and post-event images are different. Thus, applying mature paradigms developed for unimodal CD to accurately detect changes from heterogeneous images in multimodal CD is not feasible (Touati, 2019; Volpi et al., 2015). Consequently, the first task of multimodal CD is to construct a connection between the heterogeneous images and then transform them into a common domain, making them “comparable” (Chen et al., 2023b). To achieve this goal, researchers have proposed various approaches related to feature learning, data transformation and domain adaptation (Deng et al., 2021). The existing multimodal CD methods can be roughly categorized into five groups based on the transformation method and common domain.

(1) Fusion based methods that usually use image fusion or feature fusion to complement the modal discrepancies and establish comparability between the them. Ferraris et al. propose a method where two multimodal images are modeled as degraded versions of two latent images with same spatial and spectral resolutions (Ferraris et al., 2018). They transform the MCD task into a robust fusion task, assuming that the differences between the latent images are spatially sparse (Ferraris et al., 2017). Some other methods use the deep neural network (DNN) to complete the fusion, such as the cross-resolution difference learning method with coupled DNN (Zheng et al., 2022a), the deeply supervised image fusion network (Zhang et al., 2020), and deep homogeneous feature fusion method (Jiang et al., 2020). However, when considering the multimodal differences and the presence of noise in remote sensing images, particularly in SAR images with prominent speckle noise, the resulting fused image in these methods may contain outliers or anomalies. These outliers can potentially lead to false detections or pseudo-changes in the detection results (Lv et al., 2022b).

(2) Classification comparison-based methods, including post-classification comparison method (Camps-Valls et al., 2008), compound classification with multitemporal segmentation method (Wan et al., 2019), multidimensional evidential reasoning-based method (Liu et al., 2014), and hierarchical extreme learning machine-based post-classification method (Han et al., 2021), which transform the images to the same category domain. This type of method has two advantages: first, these methods are simple and intuitive, and existing image classification methods can be used to obtain certain change detection results; second, they can output a multivariate change map, which not only identifies change regions but also the change category. However, these methods also have certain limitations, including the following: (1) change detection results may be limited by the image classification algorithm; (2) there is a risk of error accumulation; (3) labeled samples are required to train the classifier.

(3) Similarity measure based methods that usually assume some connections between multimodal images regardless of the imaging conditions, and use these connections to construct a imaging modality invariant metric, such as Copula theory based local statistical metric (Mercier et al., 2008), manifold learning based similarity metric (Prendes et al., 2015), sorted histogram based distance (Wan et al., 2018), affinity matrix distance (Luppino et al., 2019), nonlocal patch distance (Sun et al., 2021c), and pixel pairwise energy based models (Touati and Mignotte, 2018; Touati et al., 2020). These methods have several advantages such as being intuitive, highly interpretable, and mostly unsupervised. However, the performance of such methods depends on two aspects: first, the reasonableness of the imaging modality invariant assumption, which requires that the assumptions used can be applied to different multimodal CD scenarios with universal applicability; second, the distinguishability of constructed similarity

metrics, which requires that the changed and unchanged regions can be well distinguished under these metrics.

(4) Feature leaning based methods that usually first transform two multimodal images into a common feature space and then compare the images in the transformed domain. This type of methods are mainly deep learning based methods that usually take the form of Siamese or pseudo-Siamese networks (Zhang et al., 2016), such as symmetric convolutional coupling network (Liu et al., 2018a), bipartite convolutional neural network (Liu et al., 2022b), common feature learning with commonality autoencoder (Wu et al., 2021), simple multiscale UNet (Lv et al., 2022a), self-supervised leaning based method (Chen and Bruzzone, 2022), structural relationship graph convolutional autoencoder (Chen et al., 2022). Deep feature comparison-based methods offer the advantage of establishing connections between multimodal images by using the powerful learning ability of neural networks, without the need for manually constructed metrics, which makes these methods adaptive and robust. However, these methods also often suffer from two limitations. First, deep network training is a time-consuming and computationally resource-intensive process. Second, their performance is strongly influenced by the training samples. Furthermore, due to the lack of large-scale datasets for multimodal CD, most of these methods have to use the pre- and post-event images themselves to train the deep network. This results in the need to retrain the network when the task scene, sensor type, imaging conditions are different, leading to weak transferability of the network.

(5) Image regression based methods aim to make multimodal images comparable by transforming one image to the domain of the other image and then comparing images as in the unimodal CD, such as homogeneous pixel transformation method (Liu et al., 2018b), fractal projection method (Mignotte, 2020), image structure regression based methods (Sun et al., 2021b), deep translation network (Li et al., 2021), deep homogeneous feature fusion method (Jiang et al., 2020), conditional generative adversarial network (CGAN) (Niu et al., 2019), and cycle-consistent adversarial network with K-means clustering and random forest classifier (Liu et al., 2022c). Compared to feature comparison-based methods, the benefit of image regression methods is that they could provide more visual information by obtaining regression image in addition to the difference image (DI) and the final change map (CM), which helps to identify change categories. However, these methods also often face two primary challenges: firstly, how to build an appropriate regression function in traditional regression methods, and secondly, how to train the regression network in deep regression methods. Additionally, deep regression methods are also limited by the availability of computational resources and time, as well as the transferability of the networks, which is particularly problematic due to the lack of large-scale datasets for multimodal CD.

## 1.3. Motivations

The objective of this study is to propose a traditional image regression method for multimodal CD. Let  $X$  and  $Y$  denote the pre- and post-event images that are co-registered, respectively. For the image regression based multimodal CD methods, they need to build two transform functions (or named mapping functions) of  $\mathcal{M}_1 : X \rightarrow Y'$  and  $\mathcal{M}_2 : Y \rightarrow X'$ , where  $X$  and  $X'$ ,  $Y$  and  $Y'$  have the similar statistical characteristics under the unchanged region, resulting in regression image  $X'$  (or  $Y'$ ) and original image  $X$  (or  $Y$ ) are in the same image domain. Obviously, the key to the image regression based multimodal CD is the design of the mapping functions ( $\mathcal{M}_1$  and  $\mathcal{M}_2$ ).

Recently, some regression methods based on image structure have been proposed for the multimodal CD (Jia et al., 2022; Zheng et al., 2022b). In the work of Luppino et al. Luppino et al. (2019), the heterogeneous images are divided into patches and two affinity matrices are built to capture the local structure of patches from different images, and then the affinity matrices distance (AMD) are calculated to measure the structure difference. AMD is further used to select training

samples, i.e., regions that are more likely to be unchanged, to train four different traditional image regression algorithms (Luppino et al., 2019) or the deep image translation networks, such as the adversarial cyclic encoder network (ACE-Net) (Luppino et al., 2022b) and the code-aligned autoencoder network (CAAE) (Luppino et al., 2022a). Mignotte has proposed a fractal projection and Markovian segmentation based method (FPMS) (Mignotte, 2020) for multimodal CD by exploiting the image self-similarity, which first fractally encodes the pre-event image to obtain its structure information, and maps it to the other domain by fractal projection, and then compares the mapped image and post-event image to detect changes. We have also proposed some graph-based regression methods for multimodal CD based on the structure consistency between images. These methods first construct a graph for the pre-event image to capture either local structure information (Sun et al., 2021b) or a combination of global and local structural information (Sun et al., 2022c), and then use the image decomposition model to translate the pre-event image by decomposing the post-event image into a change image and a regression image.

These structure-based regression methods have two advantages: first, they exploit the widespread structural consistency of images, so the methods are certain universal; second, they are image structure-based rather than pixel-based regression methods, so they are to some extent robust to image noise and imaging conditions. Despite their relatively good performance in multimodal CD, the existing image structure based regression methods still have two shortcomings as follows.

First, the structural information is not adequately expressed and used, i.e., not fully utilizing the structural information of the image, resulting in less accurate regression images and limited detection performance. They all focus solely on discovering and utilizing similarity relationships, while ignoring the dissimilarity relationship. For example, in FPMS (Mignotte, 2020), the structure of pre-event image is represented by the fractal encoding that identifies the most similar blocks of the target block in image; in the sparse-constrained adaptive structure consistency based method (SCASC) (Sun et al., 2021b) and the change smoothness based signal decomposition method (CSSD) (Zheng et al., 2022b), the structure of image is represent by the k-nearest neighbors (KNN) graph that connects each superpixel with its  $K$  most similar superpixels. Accordingly, they can only use this similarity relationship to obtain the regression image. For example, FPMS requires the regression image to be reconstructed by the fractal decoding, that is, using the blocks of post-event image from the same location to recreate the target block of the regression image; SCASC and CSSD require the similarity relationships of superpixels within regression image match those within pre-event image, that is, the superpixels within regression image connected by the KNN graph built on pre-event image are also very similar. In summary, these methods assume that the similarity relationship within regression image is identical to that of pre-event image, while ignoring dissimilarity relationships. In fact, the problem of neglecting dissimilarity also exists in many structure-based multimodal CD methods, such as the structure comparison based method (Sun et al., 2021c), however, this paper focuses only on structure-based regression methods.

In the structure regression based methods, if the image structure is more adequately characterized, the regression model will obtain more accurate regression images, and the corresponding extracted CD results will be better. In this study, we start by analyzing the limitations of regression models that rely solely on similarity relationships. Then, we highlight the significance of incorporating dissimilarity relationships as a complementary approach to more comprehensively characterize the image structure. Additionally, we analyze that the similarity relationships based KNN graph places similar superpixels closer to each other, so it mainly describes the attraction between the superpixels within the image and captures the low-frequency components of the image structure. To leverage the dissimilarity relationship, we construct a k-farthest neighbors (KFN) that connects dissimilar superpixels that represent different kinds of objects, which mainly describes the

repulsion between the superpixels within the image and captures the high-frequency components of the image structure. We then integrate a high-frequency constraint based on the dissimilarity relationship into the regression model, and propose an regression model that fuses both similarity (low-frequency) and dissimilarity (high-frequency) relationships. In this way, the regression model is able to fully utilize the structural information of the image and impose constraints on both similar and dissimilar superpixels, thereby obtaining better regression results.

Second, the fusion of change information from different domains is often overlooked. Intuitively, for the image regression based multimodal CD methods, one can obtain two change images with two transformation processes: (1) forward regression:  $\mathcal{M}_1 : X \rightarrow Y'$  and the corresponding forward change image calculated by comparing  $Y$  and  $Y'$ ; (2) backward regression:  $\mathcal{M}_2 : Y \rightarrow X'$  and the corresponding backward change image calculated by comparing  $X$  and  $X'$ . However, in the previous regression based methods, the fusion of change information from different domains has not been given sufficient attention. On the one hand, some regression based multimodal CD methods only complete a single regression process and neglect the fusion. For example, methods such as FPMS (Mignotte, 2020), CSSD (Zheng et al., 2022b), and hypergraph based image regression method (HGIR) (Sun et al., 2022c) only perform the forward regression that transforms the pre-event image to the domain of post-event image, CGAN (Niu et al., 2019) only translates the optical image to SAR image, while the deep homogeneous feature fusion (DHFF) (Jiang et al., 2020) and self-supervised image translation (SSIT) (Liu et al., 2022c) based multimodal CD methods only translate the SAR image to optical image. On the other hand, some methods directly fuse change information (difference images) obtained from different domains with a linear fusion approach, such as homogeneous pixel transformation based method (HPT) (Liu et al., 2018b), nonlocal patch graphs based method (NLPG) (Sun et al., 2021c), structural relationship graph convolutional autoencoder (SRGCAE) (Chen et al., 2022), and CAAE (Luppino et al., 2022a). However, such fusion strategies fail to fully utilize the change information in the DI calculated by different regression processes in different domains, resulting in limited gains in detection accuracy brought by fusion.

In this paper, to leverage the change information from different domains, we first perform forward regression and backward regression simultaneously to obtain forward and backward DI, which allows for capturing more change information compared to previous methods that only utilize a single-directional regression. Subsequently, to fully integrate the DI computed from different domains, we further use a Markov random field (MRF) model to combine DI fusion and DI segmentation. It can incorporate not only the spatial and change information of DI, but also the pairwise similarity and dissimilarity relationships presented in the original multitemporal images, which can avoid the loss of change information caused by the previous direct linear fusion.

#### 1.4. Contributions

The key contributions of this work are summarized below.

(1) We propose a similarity and dissimilarity induced image regression (SDIR) based multimodal CD method, which uses structure consistency constraints by incorporating both similarity (low-frequency) and dissimilarity (high-frequency) relationships of the image structure. To the best of our knowledge, this is the first method to introduce the dissimilarity relationships into CD method, which can overcome the shortcomings of previous methods that rely solely on similarity relationships.

(2) We construct two high-order graph models based on different neighborhood expansion principles to fully characterize the image structure, i.e., a similarity relationships based high-order KNN graph and a dissimilarity relationships based high-order KFN graph.

**Table 1**  
Notations.

Symbol	Description
$\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}$	Pre- and post-event images
$\tilde{\mathbf{X}}', \tilde{\mathbf{Y}}'$	Regression images of $\tilde{\mathbf{Y}}, \tilde{\mathbf{X}}$
$\tilde{\mathbf{X}}_i$	$i$ -th superpixel of $\tilde{\mathbf{X}}$
$\mathbf{X}, \mathbf{Y}$	Feature matrices of $\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}$
$\mathbf{X}', \mathbf{Y}'$	Regression feature matrices of $\mathbf{Y}, \mathbf{X}$
$\mathbf{x}_i$	$i$ -th feature vector of $\tilde{\mathbf{X}}$
$\Delta^x, \Delta^y$	Changed feature matrices
$G_n^{t1}, G_{h-n}^{t1}$	KNN graph and high-order KNN graph
$G_f^{t1}, G_{h-f}^{t1}$	KFN graph and high-order KFN graph
$\mathbf{L}^{t1}, \mathbf{L}^{t2}$	Laplacian matrices of the high-order KNN graphs
$\mathbf{I}_N$	An $N \times N$ identity matrix

(3) We address the issue of fusing change information from different domains in multimodal CD with an MRF based fusion segmentation model, which not only incorporates spatial information and similarity/dissimilarity information, but also takes into account the unbalanced change information.

The remainder of the paper is structured as follows. Section 2 analyzes the limitations of similarity relationship based KNN graph in the previous regression method. Section 3 describes the proposed SDIR in detail. Section 4 shows the experiments and gives some discussions. Finally, Section 5 provides conclusions.

## 2. Limitation of KNN graph in the structure based regression method

In this section, we first give the problem formulation of multimodal CD, then briefly review some previous regression method based on image structure, and then analyze the limitations of these methods.

### 2.1. Problem formulation

Given two co-registered images collected at times  $t_1$  (pre-event) and  $t_2$  (post-event) with different modalities, denoted as  $\tilde{\mathbf{X}} \in \mathbb{R}^{M \times N \times B_x}$  in  $\mathcal{X}$  domain and  $\tilde{\mathbf{Y}} \in \mathbb{R}^{M \times N \times B_y}$  in  $\mathcal{Y}$  domain, respectively. The image pixels of  $\tilde{\mathbf{X}}$  and  $\tilde{\mathbf{Y}}$  are denoted as  $\tilde{x}(m, n, b_x)$  and  $\tilde{y}(m, n, b_y)$ , respectively. Because the pixel values of  $\tilde{x}$  and  $\tilde{y}$  cannot be directly compared, the main purpose of multimodal CD is to enable the comparison of incomparable images to detect the changed areas represented by the CM.

We denote the feature matrices of pre-event and post-event images as  $\mathbf{X}$  and  $\mathbf{Y}$  respectively. In the decomposition based regression model, we decompose the feature matrix  $\mathbf{Y}$  into a regression feature matrix  $\mathbf{Y}'$  and a changed feature matrix  $\Delta^y$  as  $\mathbf{Y} = \mathbf{Y}' - \Delta^y$ , thereby obtaining the regression image  $\tilde{\mathbf{Y}}'$  and the matrix  $\Delta^y$  containing change information, defined as the forward regression process; it decomposes the feature matrix  $\mathbf{X}$  into a regression feature matrix  $\mathbf{X}'$  and a changed feature matrix  $\Delta^x$  as  $\mathbf{X} = \mathbf{X}' - \Delta^x$ , thereby obtaining the regression image  $\tilde{\mathbf{X}}'$  and the matrix  $\Delta^x$  containing change information, defined as the backward regression process. For easy reference, Table 1 lists some important notations.

### 2.2. Structure consistency

In the previous image structure based regression methods, such as SCASC (Sun et al., 2021b), FPMS (Mignotte, 2020), and CSSD (Zheng et al., 2022b), they transform  $\tilde{\mathbf{X}}$  to the domain  $\mathcal{Y}$  by using the structure consistency between multimodal images, which relies on the intrinsic self-similarity property of images. Suppose that images are divided into  $N_S$  small parts (e.g., square patches or superpixels), denoted as  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{Y}}_i$ ,  $i = 1, \dots, N_S$ . The self-similarity property implies that for the pre-event image, each small part  $\tilde{\mathbf{X}}_i$  can always find some similar parts within the same image, denoted as  $\tilde{\mathbf{X}}_j$ . Then, if both the  $i$ th and  $j$ th

parts are unchanged during the event, then the corresponding parts of  $\tilde{\mathbf{Y}}_i$  and  $\tilde{\mathbf{Y}}_j$  in post-event image are also very similar (Sun et al., 2021b).

We use a statistical model to analyze the structure consistency property between multimodal images, which draws on the statistical model proposed in Prendes et al. (2015). In the multimodal CD, the pre- and post-event images are often collected by different sensors, denoted as  $S_1$  and  $S_2$ , respectively. We have the following simple statistical model

$$\begin{aligned}\tilde{\mathbf{X}}_i | P_i^{t1} &= f_{S1} [T_{S1} (P_i^{t1}), n^{t1}], \\ \tilde{\mathbf{Y}}_i | P_i^{t2} &= f_{S2} [T_{S2} (P_i^{t2}), n^{t2}],\end{aligned}\quad (1)$$

where

- $P_i^{t1}$  and  $P_i^{t2}$  denote the ground objects represented by the  $i$ th parts of the pre- and post-event images, respectively;
- $T_{S1} (P_i^{t1})$  and  $T_{S2} (P_i^{t2})$  are the imaging functions that transforms  $P_i^{t1}$  and  $P_i^{t2}$  into pixel values by sensors  $S_1$  and  $S_2$  under ideal noise-free conditions, respectively;
- $n^{t1}$  and  $n^{t2}$  are the random variables that represent the various noises, e.g., sensor noise, atmospheric noise;
- $f_{S1} [\cdot, \cdot]$  and  $f_{S2} [\cdot, \cdot]$  are the noise influence functions that describe how the noise interact with the ideal sensor measurements of  $S_1$  and  $S_2$ , respectively.

For example, if the pre-event image is collected by the optical sensor, then the additive Gaussian white noise model is commonly used, i.e.,  $\tilde{\mathbf{X}}_i | P_i^{t1} = T_{S1} (P_i^{t1}) + n^{t1}$ , where  $n^{t1} \sim \mathcal{N}(0, \sigma^2)$  with  $\sigma^2$  denoting the variance; while for the SAR image, the multiplicative speckle noise model with gamma distribution is commonly used, i.e.,  $\tilde{\mathbf{X}}_i | P_i^{t1} = T_{S1} (P_i^{t1}) n^{t1}$ , where  $n^{t1} \sim \Gamma(L, L^{-1})$  with  $L$  denoting the equivalent number of looks.

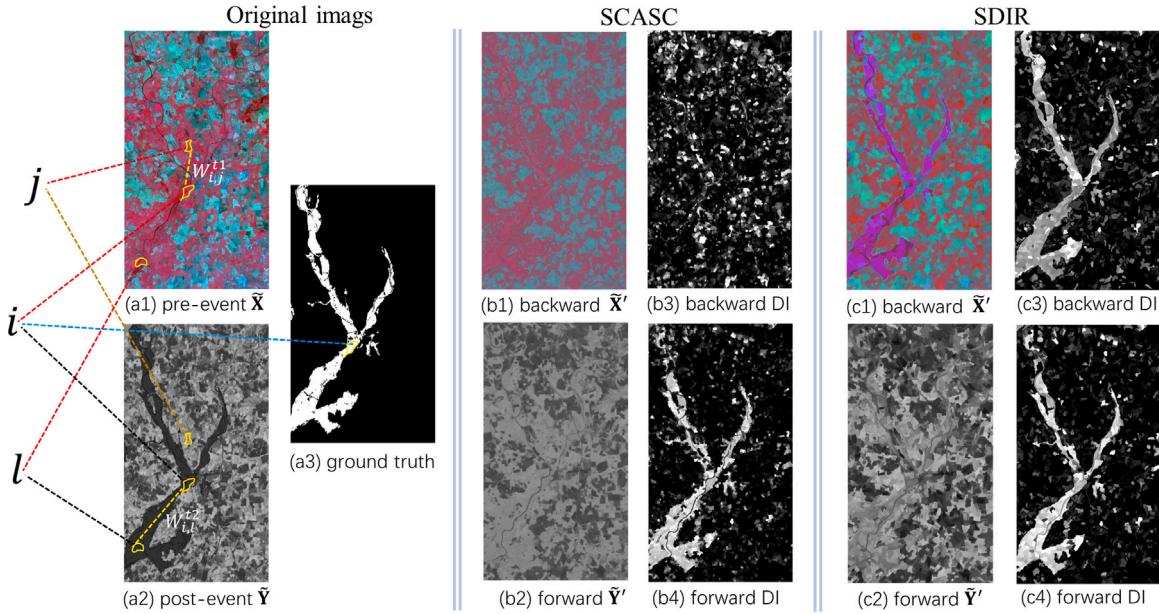
For one of the images, e.g., the pre-event image, if the  $i$ th and  $j$ th parts correspond to the same type of object, i.e.,  $P_i^{t1} = P_j^{t1}$ , then we have  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  obey the same statistical distribution and are very similar based on the model (1). If both the  $i$ th and  $j$ th region are unchanged during the event, i.e.,  $P_i^{t2} = P_j^{t2}$  and  $P_i^{t1} = P_j^{t1}$ , then the corresponding  $\tilde{\mathbf{Y}}_i$  and  $\tilde{\mathbf{Y}}_j$  in the other image obey the same statistical distribution as well (as  $P_i^{t2} = P_j^{t2}$ ) and are also very similar. On the contrary, if one of the regions is changed during the event, e.g., the  $i$ th region is unchanged but the  $j$ th region is changed during the event ( $P_i^{t2} = P_i^{t1}$ ,  $P_j^{t2} \neq P_j^{t1}$ ), then this similarity relationship cannot be conformed by the other image, i.e.,  $\tilde{\mathbf{Y}}_i$  and  $\tilde{\mathbf{Y}}_j$  is no longer similar as  $P_i^{t2} \neq P_j^{t2}$ . Because the nonlocal similarity relationship within the image itself can eliminate the discrepancy across different imaging modalities, the image structure characterized by the similarity relationship can be well preserved between unchanged multimodal images.

### 2.3. Structure based regression method

Based on the structure consistency, some regression methods have been proposed (Sun et al., 2021b; Zheng et al., 2022b; Sun et al., 2022b). They first construct a KNN graph  $G_n^{t1}$  for pre-event  $\tilde{\mathbf{X}}$ , which connects each part  $\tilde{\mathbf{X}}_i$  with its  $K$  nearest-neighbors (NNs) with the weight  $W_{i,j}^{t1} > 0$ . Obviously, the image parts of  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  connected by graph  $G_n^{t1}$  are very similar. Then, they transform  $\tilde{\mathbf{X}}$  to domain  $\mathcal{Y}$  to obtain the regression image  $\tilde{\mathbf{Y}}'$  with a signal decomposition model as

$$\begin{aligned}\min_{\mathbf{Y}', \Delta^y} \sum_{i,j=1}^{N_S} \left\| \mathbf{Y}'_i - \mathbf{Y}'_j \right\|_2^2 W_{i,j}^{t1} + \lambda \|\Delta^y\|_{2,1} \\ s.t. \quad \mathbf{Y} = \mathbf{Y}' - \Delta^y,\end{aligned}\quad (2)$$

where  $\lambda$  is the regularization parameter,  $\mathbf{W}^{t1}$  denotes the weighting matrix of  $G_n^{t1}$ ,  $\mathbf{Y}'_i$  is the  $i$ th column of regression feature matrix  $\mathbf{Y}'$ ,  $\Delta^y$  denotes the forward changed feature matrix that carries change information, and the  $\ell_{2,1}$ -norm of  $\|\Delta^y\|_{2,1} = \sum_{i=1}^{N_S} \|\Delta_i^y\|_2$  equals to the sum of the  $\ell_2$ -norm of each column vector of  $\Delta^y$ .



**Fig. 1.** An example of SCASC (Sun et al., 2021b) and the proposed SDIR. From (a1) to (a3) are: pre-event image  $\tilde{\mathbf{X}}$ , post-event image  $\tilde{\mathbf{Y}}$  and ground truth.  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  are connected by the KNN graph  $G_n^{t1}$  with  $W_{i,j}^{t1}$ .  $\tilde{\mathbf{Y}}_i$  and  $\tilde{\mathbf{Y}}_j$  are connected by the KNN graph  $G_n^{t2}$  with  $W_{i,j}^{t2}$ . From (b1) to (b4) are results generated by SCASC: backward regression  $\tilde{\mathbf{X}}'$ , forward regression  $\tilde{\mathbf{Y}}'$ , backward DI and forward DI. From (c1) to (c4) are results generated by SDIR: backward regression  $\tilde{\mathbf{X}}'$ , forward regression  $\tilde{\mathbf{Y}}'$ , backward DI and forward DI. During the forward regression process, both SCASC and SDIR can perform image transformation and change detection. However, during the backward regression process, only SDIR is able to detect changes while SCASC cannot, as SDIR utilizes the dissimilarity relationships.

In the model (2), the first term of  $\sum_{i,j=1}^{N_S} \|\mathbf{Y}'_i - \mathbf{Y}'_j\|_2^2 W_{i,j}^{t1}$  requires that regression image  $\tilde{\mathbf{Y}}'$  and pre-event image  $\tilde{\mathbf{X}}$  have the same structure, i.e., the images parts of  $\tilde{\mathbf{Y}}'_i$  and  $\tilde{\mathbf{Y}}'_j$  connected by the KNN graph  $G_n^{t1}$  are also very similar. The second term of  $\|\Delta^y\|_{2,1}$  is based on the sparse change prior in the practical CD problem. By solving the forward regression model (2), the changed region can be detected from the changed feature matrix  $\Delta^y$ .

#### 2.4. Limitation of similarity graph based regression

In Fig. 1, we give a simple example of the SCASC (Sun et al., 2021b) with the model (2). From Fig. 1, we can observe that the SCASC method is capable of performing the forward regression that transforms the pre-event  $\tilde{\mathbf{X}}$  to domain  $\mathcal{Y}$ , but fails to perform the backward regression that transforms the post-event  $\tilde{\mathbf{Y}}$  to domain  $\mathcal{X}$ , which is based on the model

$$\min_{\mathbf{X}', \Delta^x} \sum_{i,j=1}^{N_S} \|\mathbf{X}'_i - \mathbf{X}'_j\|_2^2 W_{i,j}^{t2} + \lambda \|\Delta^x\|_{2,1} \quad (3)$$

$$\text{s.t. } \mathbf{X} = \mathbf{X}' - \Delta^x,$$

where  $\lambda$  is the regularization parameter,  $\mathbf{W}^{t2}$  denotes the weighting matrix of  $G_n^{t2}$ ,  $\mathbf{X}'_i$  is the  $i$ th column of regression feature matrix  $\mathbf{X}'$ , and  $\Delta^x$  denotes the backward changed feature matrix. It is worth noting that although only the regression failure of SCASC is shown in Fig. 1, this similar failure also exists in other structure-based regression methods, such as FPMS (Mignotte, 2020), AMD-IR (Lupino et al., 2019), AGSCC (Sun et al., 2022b).

First, we analyze why SCASC behaves so differently in the forward and backward regression, from the perspective of the regression models. As shown in Fig. 1, the  $i$ th part that belongs to the changed area,  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  are connected by  $G_n^{t1}$ ,  $\tilde{\mathbf{Y}}_i$  and  $\tilde{\mathbf{Y}}_j$  are connected by  $G_n^{t2}$ . Then, SCASC uses the pairwise relationships between the  $i$ th part and  $j, l$ -th parts to help determine the label (changed or unchanged) of the  $i$ th part.

(1) In the forward regression model (2), because  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  are connected by  $G_n^{t1}$ ,  $\tilde{\mathbf{Y}}_i$  and  $\tilde{\mathbf{Y}}_j$  belong to different types of object and show very different features, the regularization of  $\|\mathbf{Y}'_i - \mathbf{Y}'_j\|_2^2 W_{i,j}^{t1}$  and

the sparse penalty of  $\|\Delta_i^y\|_2 + \|\Delta_j^y\|_2$  can prompt the solution to be  $\mathbf{Y}'_i = \mathbf{Y}_j$  and  $\mathbf{Y}'_j = \mathbf{Y}_i$ , resulting in the changed  $i$ th part with  $\Delta_i^y \neq 0$ , which can detect the changes in the forward regression model.

(2) However, in the backward regression model (3), because  $\tilde{\mathbf{Y}}_i$  and  $\tilde{\mathbf{Y}}_l$  are connected by  $G_n^{t2}$ ,  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_l$  belong to the same type of object and show very similar features as well, the regularization of  $\|\mathbf{X}'_i - \mathbf{X}'_l\|_2^2 W_{i,l}^{t2}$  and the sparse penalty of  $\|\Delta_i^x\|_2 + \|\Delta_l^x\|_2$  can prompt the solution to be  $\mathbf{X}'_i = \mathbf{X}_i$  and  $\mathbf{X}'_l = \mathbf{X}_l$ , resulting in the unchanged  $i$ th part with  $\Delta_i^x = 0$ , which cannot detect the changes in the backward regression model.

In fact, consider a special case where the pre-event image is a completely smooth image with  $\mathbf{X}_i = \mathbf{X}_j, \forall i, j \in \{1, \dots, N_S\}$ , i.e., the whole image represent the same type of object, then the optimal solution of backward regression model (3) is  $\mathbf{X}' = \mathbf{X}$  and  $\Delta_i^x = 0$ , which means that the model (3) cannot detect changes whatever the post-event image is. Similarly, if the post-event image is a completely smooth image with  $\mathbf{Y}_i = \mathbf{Y}_j, \forall i, j \in \{1, \dots, N_S\}$ , then the forward regression model (2) outputs the  $\mathbf{Y}' = \mathbf{Y}$  and  $\Delta_i^y = 0$ , which also fails to detect changes.

Next, we further analyze the reasons for the failure of SCASC in the forward or backward regression, from the perspective of the graph properties. In the similarity based graph constructed on the image (e.g., pre-event image), either KNN graph or fully connected graph, its edge weight  $W_{i,j}^{t1}$  are inversely proportional to the distance between image parts of  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$ , such as  $W_{i,j}^{t1} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\sigma^2}\right)$ . It describes the attraction between the internal parts of the image, in other words, it mainly characterizes the low-frequency information of the image structure, as analyzed later in Section 4.5. We can treat the feature matrix of the image as the graph signal defined on the graph, whose  $i$ th column vector represents the signal on the  $i$ th vertex. Then, it is well known that image  $\tilde{\mathbf{X}}$  is smooth on the similarity based graph (such as  $G_n^{t1}$ ) built on itself, that is, the difference between the signals connected by the graph edge is very small (Ortega et al., 2018; Shuman et al., 2013; Stanković et al., 2019). If, at this time, the other image  $\tilde{\mathbf{Y}}$  is also smooth on this graph  $G_n^{t1}$ , this graph cannot distinguish the structural differences between the two images, which leads to the failure of regression model (2). Furthermore, it can also be seen

from the structure constraint term of  $\sum_{i,j=1}^{N_S} \|\mathbf{Y}'_i - \mathbf{Y}'_j\|_2^2 W_{i,j}^{t1}$  in model (2), which focuses on those terms  $\|\mathbf{Y}'_i - \mathbf{Y}'_j\|_2^2 W_{i,j}^{t1}$  with large weights  $W_{i,j}^{t1}$ . However, for terms with small  $W_{i,j}^{t1}$  that the connected  $\mathbf{X}_i$  and  $\mathbf{X}_j$  are very different, it has little constraint on the corresponding  $\mathbf{Y}'_i$  and  $\mathbf{Y}'_j$ . This means that the similarity-based graph is incomplete in characterizing the image structure, i.e., it ignores the dissimilarity relationships (the high-frequency) within the image, which are also important in the regression model.

Recently, two papers have also observed the issue that the detection results obtained from forward and backward regression processes are significantly different (Florez-Ospina et al., 2023; Sun et al., 2023). In Florez-Ospina et al. (2023), Florez-Ospina et al. have raised a question in the conclusion section: “why is this happening in the first place, and how can we maintain performance while making the method invariant to such reversal”? In Sun et al. (2023), we have attributed this phenomenon to the structural asymmetry between multimodal images, suggesting that the varying complexity of image structures leads to differing difficulty levels in solving forward and backward regression models. Consequently, we have introduced a change alignment regularization term to enforce alignment between the DIs obtained by the two regression models. In contrast to Sun et al. (2023), this paper analyzes the performance inconsistency phenomenon from the perspective of regression model and graph representation capacity, which is both more accurate and more universally applicable. And then it presents an alternative solution without the interaction between the two regression processes, which can also address the question posed in Florez-Ospina et al. (2023).

## 2.5. Dissimilarity and fusion

Return to the simple example of Fig. 1, in the post-event image,  $\tilde{\mathbf{Y}}_i$  and  $\tilde{\mathbf{Y}}_j$  represent different types of objects (i.e.,  $P_i^{t2} \neq P_j^{t2}$ ). Based on the model (1), assuming that both the  $i$ th and  $j$ th regions are unchanged during the event, then the corresponding  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  also represent different types of objects. Therefore, we can add a penalty term in the regression model (3) that requires the regression  $\tilde{\mathbf{X}}'_i$  and  $\tilde{\mathbf{X}}'_j$  to represent different types of objects, showing that they are also very dissimilar, which naturally prevents the unchanged  $\tilde{\mathbf{X}}'_i$ .

On the other hand, we can find that combining the forward and backward regression processes is beneficial, which may improve the CD performance, as shown in Fig. 1. Therefore, in this paper we first propose a similarity and dissimilarity induced image regression method for multimodal CD, and then propose an MRF based fusion segmentation model that combines the forward and backward detection results.

## 3. Similarity and dissimilarity induced images regression method

In this section, we provide a detailed description of the proposed SDIR, which contains three main parts: (1) representing the image structure in terms of similarity and dissimilarity relationships, (2) utilizing the relationship-based structure to perform image regression, and (3) fusing the forward and backward detection results to enhance the accuracy of CD. Fig. 2 shows the framework of the proposed SDIR.

### 3.1. Preprocessing

As aforementioned in (1), we need to segment the multitemporal images into small parts with the same segmentation map. In addition, the pairwise relationship (similarity and dissimilarity) is used to represent the image structure in this paper. In light of this, we have opted to utilize superpixels as the basic analysis unit, which not only retains object edge and contextual information well but also significantly reduces algorithm complexity by decreasing the number of graph vertices.

We select the GMMSP (Ban et al., 2018) to segment the multitemporal images into homogeneous regions with the same segmentation map, which is known to produce superpixels that conform to object boundaries with linear complexity relative to the number of pixels (i.e.  $\mathcal{O}(MN)$ ). We construct a false RGB image: the first two channels are the normalized gray images of pre- and post-event images, which are obtained by summing the squared pixel values along the channels of the pre- and post-event images and then taking the square root; the third channel of the false RGB image is an all-zero channel. Then, GMMSP segments this false RGB image into  $N_S$  superpixels with the segmentation map  $\Lambda$ , and obtains the superpixels of  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{Y}}_i$ ,  $i \in \mathcal{I}$  denoted as

$$\begin{aligned}\tilde{\mathbf{X}}_i &= \{\tilde{x}(m, n, b_x) \mid (m, n) \in \Lambda_i, b_x = 1, \dots, B_x\}, \\ \tilde{\mathbf{Y}}_i &= \{\tilde{y}(m, n, b_y) \mid (m, n) \in \Lambda_i, b_y = 1, \dots, B_y\},\end{aligned}\quad (4)$$

where  $\mathcal{I} = \{1, \dots, N_S\}$  denotes the index set. Therefore,  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{Y}}_i$  represent the same geographical area and they have a high probability of being internally homogeneous, i.e., the pixels inside the superpixel of  $\tilde{\mathbf{X}}_i$  (and  $\tilde{\mathbf{Y}}_i$ ) have a high probability of representing the same type of object. However, since only the normalized gray image is used in constructing the false RGB image, it causes a certain loss of information to distinguish different objects, and the following situation may occur: spatially adjacent pixels representing different types of objects may be grouped within the same superpixel when they have similar pixel values in the normalized gray image. This is a shortcoming of the above superpixel segmentation method that needs to be improved in the future.

Once the co-segmentation superpixels are obtained, the mean and median values of each band are extracted as the superpixel feature for simplicity (it is not exclusive, other features are also available). Then, we obtain the feature matrices of  $\mathbf{X} \in \mathbb{R}^{2B_x \times N_S}$  and  $\mathbf{Y} \in \mathbb{R}^{2B_y \times N_S}$ , where each column represents the feature vector of a superpixel.

### 3.2. Structure representation

In the previous structure based multimodal CD methods, they only focus on the similarity relationships within the image (pixel-wise (Liu et al., 2018b; Touati and Mignotte, 2018), patch-wise (Mignotte, 2020; Sun et al., 2021c) or superpixel-wise (Zheng et al., 2022b; Sun et al., 2022c)) and construct the similarity induced graph to characterize the image structure, such as the KNN graph (Zheng et al., 2022b), full connected graph (Luppino et al., 2019), adaptive probabilistic graph (Sun et al., 2021b). In this paper, in addition to similarity relationships, we also explore the use of dissimilarity relationships to characterize the image structure, which are represented by two graphs: KNN graph and KFN graph, respectively.

#### 3.2.1. KNN graph

In order to characterize the similarity relationships between the objectives within the image, a high-order KNN graph is constructed. Take the pre-event image  $\tilde{\mathbf{X}}$  as an example, we first construct a KNN graph  $G_n^{t1} = \{V_n^{t1}, E_n^{t1}, w\}$  by setting each superpixel as a vertex and connecting each superpixel with other superpixels with the weight  $W_{i,j}^{t1}$ , which is similar as the adaptive probabilistic graph used in SCASC (Sun et al., 2021b) by using the following model

$$\begin{aligned}\min_{\mathbf{W}^{t1}} \sum_{i,j=1}^{N_S} D_{i,j}^{t1} W_{i,j}^{t1} + \alpha (W_{i,j}^{t1})^2 \\ \text{s.t. } 0 \leq W_{i,j}^{t1} \leq 1, \sum_{j=1}^{N_S} W_{i,j}^{t1} = 1,\end{aligned}\quad (5)$$

where  $D_{i,j}^{t1} = \|\mathbf{X}_i - \mathbf{X}_j\|_2^2$  denotes the feature distance between the  $i$ th and  $j$ th superpixels of pre-event image, and  $\alpha$  denotes the regularization parameter that controls the number of nearest neighbors of the

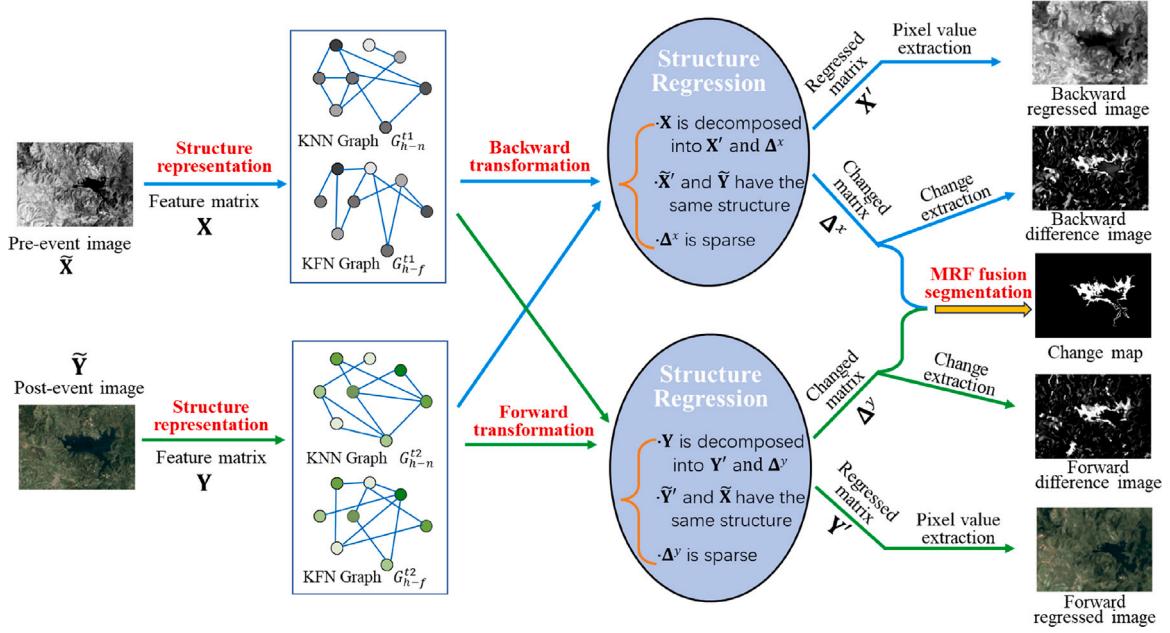


Fig. 2. Framework of the proposed similarity and dissimilarity induced images regression method.

superpixel. According to Sun et al. (2021b), the closed-form solution of  $W^{t1}$  is

$$W_{i,(j)}^{t1} = \begin{cases} \frac{D_{i,(k+1)}^{t1} - D_{i,(j)}^{t1}}{k D_{i,(k+1)}^{t1} - \sum_{h=1}^k D_{i,(h)}^{t1}}, & j \leq k \\ 0, & j > k \end{cases}, \quad (6)$$

where the distance vector  $D_i^{t1}$  is sorted in ascending order as  $D_{i,(1)}^{t1}, D_{i,(2)}^{t1}, \dots, D_{i,(N_S)}^{t1}$ , and then the symbol  $(j)$  in  $D_{i,(j)}^{t1}$  represents the index of the  $j$ th smallest value in  $D_i^{t1}$ . Then, it can be observed that  $G_n^{t1}$  is a KNN graph that connects each superpixel with its  $k$  NNs with weights  $W_i^{t1}$ .

In the KNN graph, the number  $k$  of NNs is a critical parameter. On the one hand, a very small value of  $k$  may not be robust enough for the graph, which cannot adequately characterize the similarity relationship of the images. On the other hand, a too large value of  $k$  is also inappropriate, which may cause over-connections that lead to confusion in the graph, i.e., dissimilar vertices are connected by the KNN graph that characterizes similar relationships. In this paper, we first construct the  $G_n^{t1}$  with a relatively small  $k$  as  $k = \sqrt{N_S}$ . Then, we construct a high-order KNN graph  $G_{h-n}^{t1} = \{V_{h-n}^{t1}, E_{h-n}^{t1}, w\}$  based on the  $G_n^{t1}$  by using the principle: “**a neighbor of the neighbor is also likely to be a neighbor**”, which can effectively expand each vertex’s similar neighbors, i.e., connecting the  $i$ th vertex within  $H$ -hop away from this vertex. The weighting matrix of the  $G_{h-n}^{t1}$  is computed by

$$W_{h-n}^{t1} = (\mathbf{D}_{h-n}^{t1})^{-1} \odot \mathbf{W}^H, \quad (7)$$

where  $\mathbf{W}^H = \sum_{h=1}^H (W_{h-n}^{t1})^h$  and  $\mathbf{D}_{h-n}$  denotes the diagonal matrix with the  $i$ th diagonal element being  $\sum_{j=1}^{N_S} W_{i,j}^H$ .

In this way, the  $G_{h-n}^{t1}$  can connect each vertex with more similar neighbors and capture the high-order information of the graph. For the post-event image, we can construct the high-order KNN graph  $G_{h-n}^{t2}$  in a similar way as  $G_{h-n}^{t1}$ , whose weighting matrix is denoted as  $W_{h-n}^{t2}$ .

### 3.2.2. KFN graph

In order to characterize the dissimilarity relationships between the objects within the image, a KFN graph is constructed. Take the pre-event image  $\tilde{X}$  as an example, we first construct a KFN graph  $G_f^{t1} = \{V_f^{t1}, E_f^{t1}, w\}$  by setting each superpixel as a vertex and connecting

each superpixel with the  $k$  most different superpixels, i.e.,  $(i, j) \in E_f^{t1}$  if and only if  $D_{i,j}^{t1}$  is among the  $k$ -largest elements of the distance vector  $\mathbf{D}_i^{t1}$ .

Different from the KNN graph, the KFN graph used to characterize the dissimilarity relationships has three aspects to be considered separately.

- First, the choice of the number  $k$  of farthest neighbors. It is well known that there are many object categories in remote sensing images, so there are more superpixel pairs representing different categories of objects than those representing the same object category. Therefore, the number  $k$  of farthest neighbors in KFN graph should be larger than the NNs number in the KNN graph  $G_n^{t1}$ . In this paper, we set  $k = 5\sqrt{N_S}$  for the KFN graph.

- Second, the high-order information of graph. In the high-order graph  $G_{h-n}^{t1}$ , we use the principle of “a neighbor of the neighbor is also likely to be a neighbor”. However, this principle is not applicable in the KFN graph  $G_f^{t1}$ . For example, suppose that  $\tilde{X}_i$  and  $\tilde{X}_j$  belong to the same object of “Grass”, and they are both connected by the  $\tilde{X}_i$  that represents the object of “Water” in the KFN graph  $G_f^{t1}$ , then if we use the neighborhood expansion principle in the KNN graph, the highly similar  $\tilde{X}_i$  and  $\tilde{X}_j$  will then be connected by KFN graph, which is inappropriate. Here, we use another two neighborhood expansion principles for the KFN graph instead: “**a nearest-neighbor of the farthest-neighbor is also likely to be a farthest-neighbor; and a farthest-neighbor of the nearest-neighbor is also likely to be a farthest-neighbor**”. Therefore, we can construct the high-order KFN graph  $G_{h-f}^{t1} = \{V_{h-f}^{t1}, E_{h-f}^{t1}, w\}$  by using the high-order KNN graph  $G_{h-n}^{t1}$  as follows

$$(i, j) \in E_{h-f}^{t1}; \quad \begin{aligned} &\text{if } (i, j) \in E_f^{t1}, \\ &\text{or } (i, t) \in E_f^{t1}, (t, j) \in E_{h-n}^{t1}, \\ &\text{or } (i, t) \in E_{h-n}^{t1}, (t, j) \in E_f^{t1}. \end{aligned} \quad (8)$$

- Third, the weight of the edges. In the KNN graph, the more similar the superpixels are, the larger the weight of the edges connecting them as in (6). However, in the KFN graph, the dissimilarity of the superpixels is hard to measure, that is, it is meaningless to measure the degree of difference between different categories. For example, we cannot say that the difference between “Water” and “Grass” is greater than that between “Water” and “Buildings”. Therefore, in the high-order

KFN graph, we assign the same weight to all edges connecting the vertices as

$$\mathbf{W}_{h-f}^{t1} = \left( \mathbf{D}_{h-f}^{t1} \right)^{-1} \odot \mathbf{A}_{h-f}^{t1}, \quad (9)$$

where  $\mathbf{W}_{h-f}^{t1}$ ,  $\mathbf{A}_{h-f}^{t1}$  and  $\mathbf{D}_{h-f}^{t1}$  denote the random walk matrix, adjacent matrix and degree matrix of the high-order KFN graph  $G_{h-f}^{t1}$ , respectively.

In this way, the  $G_{h-f}^{t1}$  can connect each vertex with more dissimilarity neighbors and capture the high-order information of the graph. Additionally, we can construct the high-order KFN graph  $G_{h-f}^{t2}$  for the post-event image in a similar way as  $G_{h-f}^{t1}$ , whose weighting matrix is denoted as  $\mathbf{W}_{h-f}^{t2}$ .

### 3.3. Structure based regression model

For the regression based multimodal CD methods, their intentions are to transform one image to the other image's domain for comparing images in the same domain. In the proposed SDIR, we choose superpixels as the fundamental analysis unit and extract their features to represent them, which requires finding the transformation functions between different feature matrices of the images, defined as  $\mathcal{M}_1 : \mathbf{X} \rightarrow \mathbf{Y}'$  and  $\mathcal{M}_2 : \mathbf{Y} \rightarrow \mathbf{X}'$  with  $\mathbf{X}'$  and  $\mathbf{Y}'$  denoting the transformed feature matrices. We define the superpixel feature extraction operator as  $\mathcal{F}$ , define the pixel value recovery operator from superpixel features as  $\mathcal{F}^{-1}$ , e.g., extracting the mean feature as the pixel value of each pixel inside the superpixel, define the  $\tilde{\mathbf{X}}'$  and  $\tilde{\mathbf{Y}}'$  as the translated images. Then we have the forward transformation as

$$\tilde{\mathbf{Y}}' = \mathcal{F}^{-1}(\mathbf{Y}') = \mathcal{F}^{-1}\mathcal{M}_1(\mathbf{X}) = \mathcal{F}^{-1}\mathcal{M}_1\mathcal{F}(\tilde{\mathbf{X}}), \quad (10)$$

and the backward transformation as

$$\tilde{\mathbf{X}}' = \mathcal{F}^{-1}(\mathbf{X}') = \mathcal{F}^{-1}\mathcal{M}_2(\mathbf{Y}) = \mathcal{F}^{-1}\mathcal{M}_2\mathcal{F}(\tilde{\mathbf{Y}}). \quad (11)$$

Next, we describe these two transformation processes in detail.

#### 3.3.1. Similarity relationships based regularization

As can be seen in (10), it is necessary to find the connection between the translated image  $\tilde{\mathbf{Y}}'$  and the original image  $\tilde{\mathbf{X}}$ . By using the statistical model (1), we have

$$\tilde{\mathbf{Y}}'_i | P_i^{t1} = f_{S2}[T_{S2}(P_i^{t1}), n^2]. \quad (12)$$

Then, we have that the similarity relationship between superpixels within the images of  $\tilde{\mathbf{X}}$  and  $\tilde{\mathbf{Y}}'$  are the same.

In the high-order KNN graph of  $G_{h-n}^{t1}$ , if the superpixels of  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  are connected by the edge  $E_{h-n}^{t1}$ , we have a high probability that  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  belong to the same type of object, i.e.  $P_i^{t1} = P_j^{t1}$ , then the corresponding  $\tilde{\mathbf{Y}}'_i$  and  $\tilde{\mathbf{Y}}'_j$  in the transformed image also are very similar by using (12), which means that the superpixels in the transformed image  $\tilde{\mathbf{Y}}'$  connected by the graph  $G_{h-n}^{t1}$  should be similar as well. Therefore, we need to construct a similarity relationship based regularization (SRR) that penalizes the dissimilarity of the connected  $\tilde{\mathbf{Y}}'_i$  and  $\tilde{\mathbf{Y}}'_j$ . With the same regularization term as in KNN graph learning model (2), we choose SRR as

$$\sum_{i,j=1}^{N_S} \|\mathbf{Y}'_i - \mathbf{Y}'_j\|_2^2 (W_{h-n}^{t1})_{i,j} = 2\text{Tr}(\mathbf{Y}' \mathbf{L}^{t1} \mathbf{Y}'^T), \quad (13)$$

where  $\mathbf{L}^{t1}$  denotes the Laplacian matrix of the high-order KNN graph  $G_{h-n}^{t1}$  and  $\text{Tr}(\cdot)$  denotes the trace of matrix.

#### 3.3.2. Dissimilarity relationships based regularization

In the high-order KFN graph of  $G_{h-f}^{t1}$ , if the superpixels of  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  are connected by the edge  $E_{h-f}^{t1}$ , we have a high probability that  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  belong to the different types of objects, i.e.  $P_i^{t1} \neq P_j^{t1}$ , then the corresponding  $\tilde{\mathbf{Y}}'_i$  and  $\tilde{\mathbf{Y}}'_j$  in the transformed image also belong to the different types of objects by using (12), which means that the

superpixels in the transformed image  $\tilde{\mathbf{Y}}'$  connected by the graph  $G_{h-f}^{t1}$  should be dissimilar as well. Here, we need to construct a dissimilarity relationship based regularization (DRR) that penalizes the similarity of the connected  $\tilde{\mathbf{Y}}'_i$  and  $\tilde{\mathbf{Y}}'_j$  as

$$\sum_{i,j=1}^{N_S} \phi\left(dist_{i,j}^{Y'}\right) (W_{h-f}^{t1})_{i,j}, \quad (14)$$

where the value of  $\phi\left(dist_{i,j}^{Y'}\right)$  is inversely proportional to the feature distance of  $dist_{i,j}^{Y'}$ .

Intuitively, the simplest and most straightforward penalty function is

$$\phi\left(dist_{i,j}^{Y'}\right) = -\|\mathbf{Y}'_i - \mathbf{Y}'_j\|_2^2. \quad (15)$$

However, we need to consider the specificity of dissimilarity based penalty, which has also been emphasized in the construction of KFN graph, i.e., the degree of difference between different categories is not comparable. Therefore, the purpose of DRR is to make the  $\tilde{\mathbf{Y}}'_i$  and  $\tilde{\mathbf{Y}}'_j$  connected by high-order KFN graph not belong to the same type of objects, not to make them increasingly different from each other in terms of features when they originally belong to different types of objects. It serves to impose penalties on very similar superpixels of  $\tilde{\mathbf{Y}}'_i$  and  $\tilde{\mathbf{Y}}'_j$ , while those less similar  $\tilde{\mathbf{Y}}'_i$  and  $\tilde{\mathbf{Y}}'_j$  should be ignored (because they already represent different types of objects), at this point it would lead to errors if  $\tilde{\mathbf{Y}}'_i$  and  $\tilde{\mathbf{Y}}'_j$  were separated further apart instead. So the value of  $\phi\left(dist_{i,j}^{Y'}\right)$  should be large around  $dist_{i,j}^{Y'} \approx 0$ , and relatively flat and converging to 0 in the region where  $dist_{i,j}^{Y'}$  is larger.

If we use a threshold  $th$  to determine whether  $\tilde{\mathbf{Y}}'_i$  and  $\tilde{\mathbf{Y}}'_j$  belong to the same type of object, the penalty function should be

$$\phi\left(dist_{i,j}^{Y'}\right) = \begin{cases} c, & \text{if } dist_{i,j}^{Y'} \leq th \\ 0, & \text{otherwise} \end{cases}, \quad (16)$$

where  $c$  is a constant. However, this threshold  $th$  is hard to determine. In this paper, we use the penalty function as follows

$$\phi\left(dist_{i,j}^{Y'}\right) = \frac{1}{\|\mathbf{Y}'_i - \mathbf{Y}'_j\|_2^2 + \varepsilon}, \quad (17)$$

where  $\varepsilon > 0$  is a parameter to adjust the steepness of the curve. Here, we set  $\varepsilon = \frac{1}{|E_n^{t2}|} \sum_{(i,j) \in E_n^{t2}} \|\mathbf{Y}_i - \mathbf{Y}_j\|_2^2$ , which is approximately equal to the average distance between superpixels representing the same type of objects in the post-event image.

#### 3.3.3. Prior sparsity based regularization

For the post-event feature matrix  $\mathbf{Y}$ , we decompose it into a transformed feature matrix  $\mathbf{Y}$  and a changed feature matrix  $\Delta^y$  as  $\mathbf{Y} = \mathbf{Y}' - \Delta^y$ , where the changed feature matrix  $\Delta^y$  contains the change information caused by the event. We have a prior sparsity based regularization (PSR) for the changed feature matrix, which is based on the fact that only a small part of objects are changed during the event in the practice. Intuitively, the  $\ell_{2,0}$ -norm regularization  $\|\Delta^y\|_{2,0}$  that equals to the number of non-zero column of  $\Delta^y$ , i.e., the number of changed superpixels, should be chosen as the PSR. However,  $\|\Delta^y\|_{2,0}$  is non-convex and imposing constraints on it in the model leads to difficulties in solving. Therefore, we use the  $\ell_{2,1}$ -norm based PSR as

$$\|\Delta^y\|_{2,1} = \sum_{i=1}^{N_S} \|\Delta_i^y\|_2, \quad (18)$$

which is a convex relaxation of  $\|\Delta^y\|_{2,0}$ .

### 3.3.4. Objective function

By combining the SRR (13), DRR (14) and PSR (18), we can obtain the similarity and dissimilarity induced forward regression model as

$$\min_{\mathbf{Y}', \Delta^y} 2\text{Tr}(\mathbf{Y}' \mathbf{L}^{t1} \mathbf{Y}'^T) + \beta \sum_{i,j=1}^{N_S} \frac{\left( W_{h-f}^{t1} \right)_{i,j}}{\|\mathbf{Y}'_i - \mathbf{Y}'_j\|_2^2 + \epsilon} + \lambda \|\Delta^y\|_{2,1} \quad (19)$$

$$\text{s.t. } \mathbf{Y}' = \mathbf{Y}' - \Delta^y,$$

where  $\beta, \lambda > 0$  are the balancing parameters.

From (19), we can find that the SRR tends to obtain the solution of smoothed  $\mathbf{Y}$ , such as  $\mathbf{Y} = 0$ , the DRR tends to obtain the non-smooth  $\mathbf{Y}$ , while the PSR tends to obtain the solution of  $\mathbf{Y}' = \mathbf{Y}$ , which means that these regularization terms have an adversarial balancing effect. In addition, by comparing the proposed SDIR (19) with the previous SCASC (2), we can find that with the help of DRR, even when the post-event image is a completely smooth image with  $\mathbf{Y}_i = \mathbf{Y}_j, \forall i, j \in \{1, \dots, N_S\}$ , the optimal solution of  $\mathbf{Y}'$  is not  $\mathbf{Y}' = \mathbf{Y}$  and  $\Delta_i^y = 0$  as in SCASC (2). Problem (19) can be efficiently optimized by using the alternating direction method of multipliers (ADMM), the details of which are shown in Appendix.

Similar to the forward regression model of (19), we can obtain the backward regression by using the following model

$$\min_{\mathbf{X}', \Delta^x} 2\text{Tr}(\mathbf{X}' \mathbf{L}^{t2} \mathbf{X}'^T) + \beta \sum_{i,j=1}^{N_S} \frac{\left( W_{h-f}^{t2} \right)_{i,j}}{\|\mathbf{X}'_i - \mathbf{X}'_j\|_2^2 + \epsilon} + \lambda \|\Delta^x\|_{2,1} \quad (20)$$

$$\text{s.t. } \mathbf{X}' = \mathbf{X}' - \Delta^x.$$

### 3.4. Change extraction

In case the transformed feature matrices of  $\mathbf{Y}$ ,  $\mathbf{X}$  and changed feature matrices of  $\Delta^y$ ,  $\Delta^x$  are calculated by the forward and backward regression models of (19) and (20), we can compute the regression images by using  $\tilde{\mathbf{Y}}' = \mathcal{F}^{-1}(\mathbf{Y}')$  and  $\tilde{\mathbf{X}}' = \mathcal{F}^{-1}(\mathbf{X}')$ , and obtain the DIs by using

$$\begin{aligned} \text{DI}^y(m, n) &= p_i^y; \quad (m, n) \in \Lambda_i, i \in \mathcal{I}, \\ \text{DI}^x(m, n) &= p_i^x; \quad (m, n) \in \Lambda_i, i \in \mathcal{I}, \end{aligned} \quad (21)$$

where  $p_i^y = \|\Delta_i^y\|_2$  and  $p_i^x = \|\Delta_i^x\|_2$  represent the forward and backward change levels of the  $i$ th superpixel, respectively.

From (21), we can find that the forward and backward DIs are computed from different domains. A linear fusion approach that sums and averages them may result in information loss. To fully fuse the change information in the two DIs to extract the final CM, we use a improved MRF fusion segmentation method based on IRG-McS (Sun et al., 2021a). In contrast to the uniform distribution assumption used in the segmentation model in IRG-McS, the proposed MRF fusion segmentation in this paper modifies the data energy term to take into account the class imbalance commonly observed in CD problems, that is, the proportion of changed and unchanged areas is uneven.

We define  $\mathbf{L} = \{L_i | i \in \mathcal{I}\}$  as the label set of superpixels with  $L_i \in \{\omega_u, \omega_c\}$ , where  $L_i = \omega_u$  represents that the region of  $\Lambda_i$  is unchanged and  $L_i = \omega_c$  represents that  $\Lambda_i$  is changed during the event. Then, the index set  $\mathcal{I}$  can be divided into unchanged set  $\Omega_u = \{i | L_i = \omega_u, i \in \mathcal{I}\}$  and changed set  $\Omega_c = \{i | L_i = \omega_c, i \in \mathcal{I}\}$ . The superpixel-labeling problem is equivalent to an energy minimization problem according to the MRF theory (Geman and Geman, 1984; Szeliski et al., 2008) given the change level vectors of  $\mathbf{p}^x$  and  $\mathbf{p}^y$

$$\mathbf{L}^* = \arg \min_{\mathbf{L} \in \{\omega_u, \omega_c\}^{\mathcal{I}}} \gamma \mathcal{J}_d(\mathbf{L}) + (1 - \gamma) \mathcal{J}_s(\mathbf{L}), \quad (22)$$

where  $\mathcal{J}_d(\mathbf{L})$  represents the data energy term based on the DIs,  $\mathcal{J}_s(\mathbf{L})$  represents the spatial energy term, and  $\gamma > 0$  is a weighting parameter.

In order to construct the data energy term  $\mathcal{J}_d(\mathbf{L})$ , we first normalize the change level vectors as  $\mathbf{p}'^x = \min\{\mathbf{p}^x / 2T^x, 1\}$ ,  $\mathbf{p}'^y = \min\{\mathbf{p}^y / 2T^y, 1\}$ ,

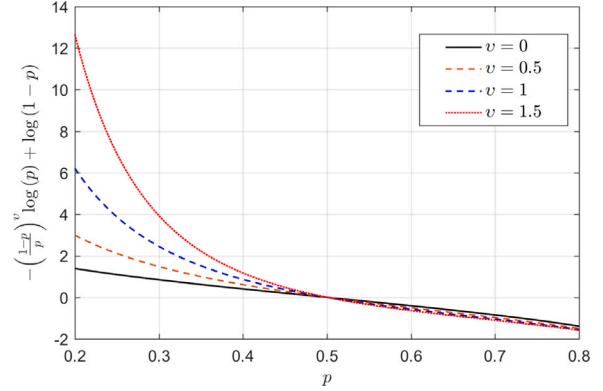


Fig. 3. Loss function with different  $v$ .

where  $T^x$  and  $T^y$  are the Otsu thresholding parameters (Otsu, 1979) on  $\mathbf{p}^x$  and  $\mathbf{p}^y$  respectively. Then, we define the  $\mathcal{J}_d(\mathbf{L})$  as

$$\mathcal{J}_d(\mathbf{L}) = \sum_{i=1}^{N_S} \varphi_c(L_i, p_i'^x) + \varphi_c(L_i, p_i'^y), \quad (23)$$

where  $\varphi_c$  is inspired by the focal loss function (Lin et al., 2017) defined as

$$\varphi_c(L_i, p) = \begin{cases} -\left(\frac{1-p}{p}\right)^v \log(p), & \text{if } L_i = \omega_c \\ -\log(1-p), & \text{if } L_i = \omega_u \end{cases}, \quad (24)$$

where  $v > 0$  is a tuning parameter.

From (23) and (24), we can find that when  $p_i^x > T^x$ ,  $\varphi_c(\omega_c, p_i'^x)$  is smaller than  $\varphi_c(\omega_u, p_i'^x)$ , which encourages the  $i$ th superpixel to be labeled as changed; when  $p_i^x < T^x$ ,  $\varphi_c(\omega_c, p_i'^x)$  is larger than  $\varphi_c(\omega_u, p_i'^x)$ , which encourages the  $i$ th superpixel to be labeled as unchanged. At the same time, the class imbalances, i.e. the very small percentage of changed classes and the very large percentage of unchanged classes in the binary CM, are taken into account by the energy function  $\mathcal{J}_d(\mathbf{L})$ . When  $v = 0$ , this type  $\mathcal{J}_d(\mathbf{L})$  can be regarded as derived from the assumption that  $\mathbf{p}'^x$  and  $\mathbf{p}'^y$  obey uniform distributions. Fig. 3 plots the loss function of  $\varphi_c(\omega_c, p) - \varphi_u(\omega_c, p)$  with different  $v$ , from which it can be seen that as  $v$  gets larger, the value of  $\varphi_c(\omega_c, p) - \varphi_u(\omega_c, p)$  increases faster as  $p$  decreases from 0.5 to 0.2. However, in the other direction as  $p$  increases from 0.5 to 0.8, the value of  $\varphi_c(\omega_c, p) - \varphi_u(\omega_c, p)$  changes very slowly. This means that the defined function (24) tends to obtain more unchanged regions, and the parameter  $v$  control the rate of the imbalance penalties.

For the spatial energy term  $\mathcal{J}_s(\mathbf{L})$ , we directly use the form in IRG-McS (Sun et al., 2021a) to penalize the spatial adjacent superpixels that are assigned with different labels, which can incorporate the contextual information of DI and pairwise similarity/dissimilarity relationships of original multitemporal images. Finally, the energy minimization problem (22) can be solved efficiently by the graph cuts algorithm (Boykov and Kolmogorov, 2004). After the  $\mathbf{L}^*$  is obtained, the binary CM can be computed as

$$\text{CM}(m, n) = L_i^*; \quad (m, n) \in \Lambda_i, i \in \mathcal{I}. \quad (25)$$

The overall framework of SDIR based multimodal CD is summarized in Algorithm 1.

**Table 2**  
Multimodal datasets.

Dataset	Date	Sensor (or modality)	Location	Image Size	Event (& Spatial resolution)
#1	Sept. 1995–July 1996	Landsat-5/Google Earth	Sardinia, Italy	300 × 412 × 1(3)	Lake expansion (30 m.)
#2	May 2012–July 2013	Pleiades/WorldView2	Toulouse, France	2000 × 2000 × 3(3)	Construction (0.52 m.)
#3	1999–2000	Spot/NDVI	Gloucester, England	990 × 554 × 3(1)	Flooding ( $\approx 25\text{m}$ )
#4	June 2008–Sept. 2012	Radarsat-2/Google Earth	Shuguang Village, China	593 × 921 × 1(3)	Building construction (8 m.)
#5	July 2006–July 2007	QuickBird 2/TerraSAR-X	Gloucester, England	4135 × 2325 × 3(1)	Flooding (0.65 m.)
#6	June 2008–Sept. 2010	Radarsat-2/Google Earth	Yellow River, China	343 × 291 × 1(3)	Embankment change (8 m.)

Algorithm 1: SDIR based multimodal change detection.

**Input:** Multimodal Images of  $\tilde{\mathbf{X}}$ ,  $\tilde{\mathbf{Y}}$ , parameters of  $N_S$ ,  $\beta$ ,  $\lambda$ .

**Structure representation:**

Segment images into superpixels by using GMMSP.

Extract features to obtain  $\mathbf{X}$  and  $\mathbf{Y}$ .

Construct the high-order KNN graphs of  $G_{h-n}^{(1)}$ ,  $G_{h-n}^{(2)}$ .

Construct the high-order KFN graphs of  $G_{h-f}^{(1)}$ ,  $G_{h-f}^{(2)}$ .

**Structural regression:**

Initialize: set  $\Delta^x$ ,  $\Delta^y = \mathbf{0}$ .

Repeat:

1: Update  $\mathbf{X}'$  and  $\mathbf{Y}'$  through (30).

2: Update  $\Delta^x$  and  $\Delta^y$  through (32).

3: Update the Lagrange multiplier through (33).

Until the stopping criterion is met.

**Change extraction:**

Calculate the DIs of  $\text{DI}^x$  and  $\text{DI}^y$ .

Compute the CM by using MRF fusion segmentation.

#### 4. Experimental results and discussions

This section shows the capability of the proposed SDIR in unsupervised multimodal CD problem. We consider six commonly used datasets listed in Table 2<sup>1</sup>. These datasets encompass a variety of sensors, image sizes, spatial resolutions, and change events, allowing for a comprehensive evaluation of the adaptability and robustness of the proposed algorithm. We compare the SDIR with some state-of-the-art methods (including five traditional methods of AMD-IR (Luppino et al., 2019), SCASC (Sun et al., 2021b), AGSCC (Sun et al., 2022b), FPMS (Mignotte, 2020), CICM (Touati, 2019) and six deep learning based methods of CGAN (Niu et al., 2019), SCCR (Liu et al., 2018a), CAAE (Luppino et al., 2022a), ACE-Net (Luppino et al., 2022b), X-Net (Luppino et al., 2022b), SRGCAE (Chen et al., 2022)) on regression images, DI and CM.

##### 4.1. Evaluation metric and implementation detail

To assess the performance of DI, we use the precision-recall (PR) curve and calculate the corresponding area under the curve, denoted as AUP. For evaluating the performance of CM, we use metrics such as true negatives (TN), true positives (TP), false negatives (FN), false positives (FP), miss alarms (MA) and false alarms (FA), overall accuracy (OA), F1-score (F1), and Kappa coefficient ( $\kappa$ ), where the first four indicators are marked with different colors in the qualitative results and the last five indicators are listed in the quantitative results.

For all the experiments of SDIR, we set the superpixel number  $N_S \approx 2500$ , and fix the balance parameter in the regression models of (19) and (20) as  $\lambda = 0.1$ , and adjust the  $\beta$  by varying  $\beta \in \{5, 10, 15, 20\}$ , and set the imbalance parameter in the MRF fusion segmentation as  $v = 1$ . The effects of parameters are discussed in Section 4.5.

##### 4.2. Regression images

To evaluate the regression performance of the proposed SDIR in multimodal CD, four image regression based methods are chosen for

comparison, including the image structure based methods of AMD-IR (Luppino et al., 2019), SCASC (Sun et al., 2021b), and AGSCC (Sun et al., 2022b), and the deep translation method of CGAN (Niu et al., 2019). In the case of methods that only have one-way regression, such as SCASC, AGSCC, and CGAN, we reversed the order of the input images to obtain the regression in the opposite direction.

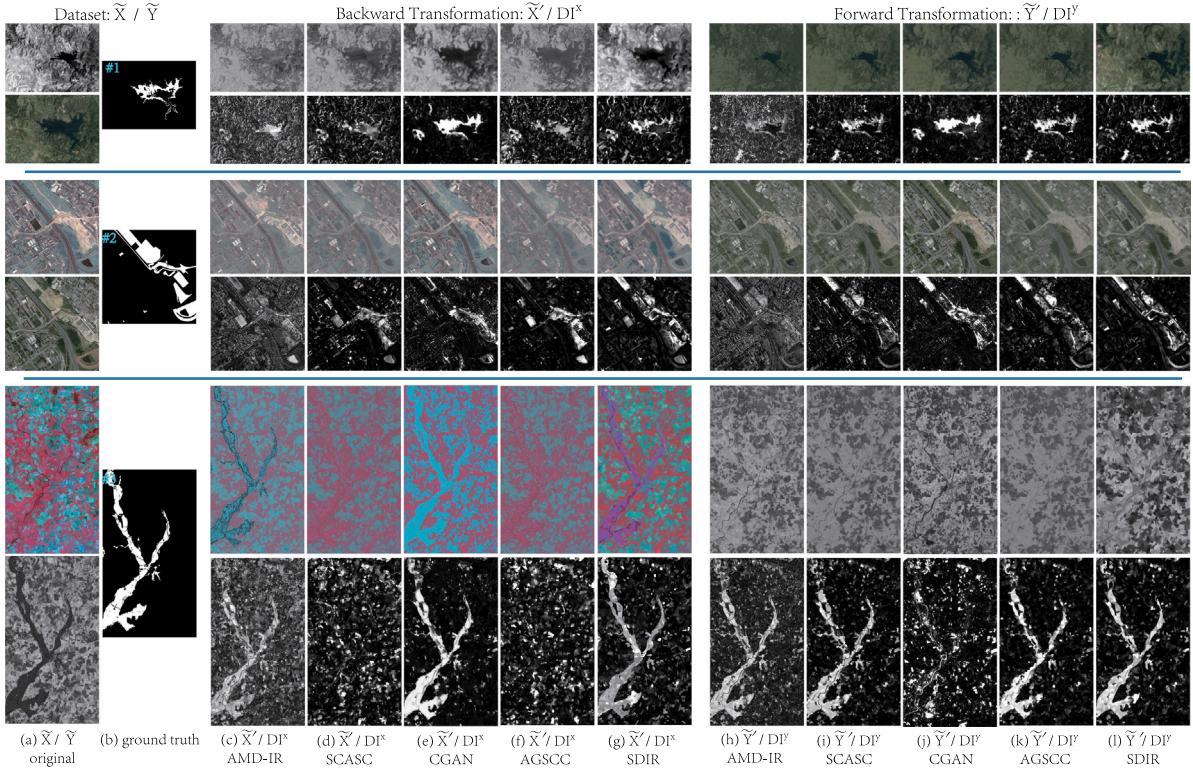
Figs. 4 and 5 show the forward and backward regression images in all the datasets. As seen in the regression images  $\tilde{\mathbf{X}}'$  and  $\tilde{\mathbf{Y}}'$  of Figs. 4 and 5, most of these methods are able to accomplish the style transformation of the images, i.e., the statistical properties of regression images and original images in the target domain are similar. However, when further comparing these regression images with the original images of the source domain carefully, it can be seen that the regression images obtained by some methods do not retain the structure of the original images and regression errors occur. For example, in the backward regression process of AMD-IR on Dataset #1, SCASC and AGSCC on Datasets #3 and #5, the structure of regressed  $\tilde{\mathbf{X}}'$  and original  $\tilde{\mathbf{Y}}$  are inconsistent. As analyzed in Section 2.4, this is due to the fact that the original image  $\tilde{\mathbf{Y}}$  of the target domain is smooth on the KNN graph  $G_n^{(1)}$  constructed over the image  $\tilde{\mathbf{X}}$  of source domain as well, which also validates the limitations of similarity-based graph in the structure based regression method. Similarly, the forward regression images of CGAN on Dataset #3 and AMD-IR on Dataset #5 are relatively different from original images in structure. Meanwhile, it can be seen that the regression method SDIR incorporating similarity and dissimilarity relationships proposed in this paper can not only avoid the failure of backward regression in Datasets #3 and #5, but also achieve better results in forward regression of other datasets, such as Datasets #2 and #6.

##### 4.3. Difference images

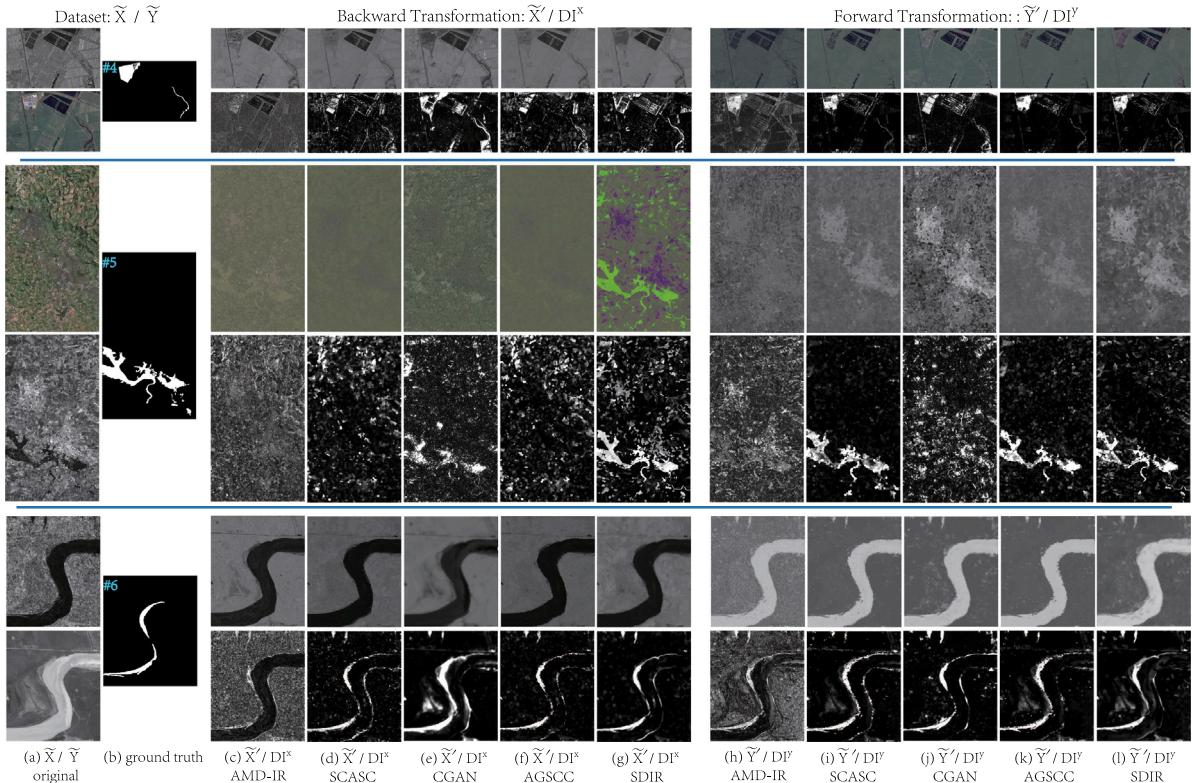
To assess the ability of the forward and backward regression processes to measure the changes, we show both forward and backward DIs obtained by different methods in Figs. 4 and 5. Three-fold can be observed: firstly, there are certain differences between the forward and backward DIs of each method, which are due to the fact that these DIs are calculated in different domains and also the detection ability of the forward and backward regression models are different, which again demonstrates the importance of fusing the forward and backward DIs. Secondly, the forward and backward DIs obtained by some methods are difficult to detect changes in some datasets due to the different performance ability of similarity-based graph (i.e., the KNN graph cannot distinguish between the pre- and post-event images), such as the backward DIs of AMD-IR, SCASC and AGSCC on Datasets #1 and #5, the forward DIs of CGAN on Dataset #3 and AMD-IR on Dataset #5. Thirdly, the proposed SDIR is able to overcome the shortcomings of the similarity-based KNN graph in terms of its insufficient structure-representation ability, and introduce a dissimilarity-based KFN graph, which enables the regression model to obtain better DIs in both forward and backward regression processes. In addition, it can also be found that the DI obtained by SDIR is sparse, which can be attributed to the use of prior sparsity-based regularization in the model.

Figs. 6 and 7 plot the PR curves of DIs generated by all the comparison methods except for CICM (which does not provide the DI), and Table 3 reports the corresponding AUP of different methods.

<sup>1</sup> Datasets #2, #3 and #5 are available at Professor Max Mignotte's homepage of <http://www-labs.iro.umontreal.ca/~mignotte>



**Fig. 4.** Multimodal datasets, forward and backward regression images and DIs on Datasets #1 to #3. From top to bottom, they correspond to Datasets #1 to #3, respectively. From left to right are: (a) pre-event image  $\tilde{X}$  and post-event image  $\tilde{Y}$ ; (b) the ground truth; (c)–(g) are the backward regression image of  $\tilde{X}'$  and backward DI of  $DI^x$  generated by (c) AMD-IR, (d) SCASC, (e) CGAN, (f) AGSCC, (g) the proposed SDIR; (h)–(l) are the forward regression image of  $\tilde{Y}'$  and forward DI of  $DI^y$  generated by (h) AMD-IR, (i) SCASC, (j) CGAN, (k) AGSCC, (l) the proposed SDIR.



**Fig. 5.** Multimodal datasets, forward and backward regression images and DIs on Datasets #4 to #6. From top to bottom, they correspond to Datasets #4 to #6, respectively. From left to right are: (a) pre-event image  $\tilde{X}$  and post-event image  $\tilde{Y}$ ; (b) the ground truth; (c)–(g) are the backward regression image of  $\tilde{X}'$  and backward DI of  $DI^x$  generated by (c) AMD-IR, (d) SCASC, (e) CGAN, (f) AGSCC, (g) the proposed SDIR; (h)–(l) are the forward regression image of  $\tilde{Y}'$  and forward DI of  $DI^y$  generated by (h) AMD-IR, (i) SCASC, (j) CGAN, (k) AGSCC, (l) the proposed SDIR.

**Table 3**

AUP of DIs on the multimodal datasets. The best and second best scores are marked in red and blue, respectively.

Methods	Dataset #1		Dataset #2		Dataset #3		Dataset #4		Dataset #5		Dataset #6		Average	
	Forward	Backward												
AMD-IR (Luppino et al., 2019)	0.155	0.060	0.264	0.237	0.741	0.537	0.564	0.090	0.103	0.057	0.216	0.137	0.341	0.186
SCASC (Sun et al., 2021b)	0.383	0.133	0.458	0.276	0.636	0.123	0.695	0.218	0.681	0.048	0.597	0.517	0.575	0.219
CGAN (Niu et al., 2019)	<b>0.550</b>	<b>0.590</b>	0.436	0.284	0.238	<b>0.887</b>	0.273	0.270	0.073	0.287	0.375	0.474	0.324	<b>0.465</b>
AGSCC (Sun et al., 2022b)	0.532	0.106	<b>0.534</b>	0.427	0.719	0.113	0.787	0.257	<b>0.797</b>	0.053	<b>0.664</b>	0.606	<b>0.672</b>	0.260
FPMS (Mignotte, 2020)	0.406	<b>0.590</b>	0.258	0.224	<b>0.774</b>	0.102	<b>0.904</b>	0.237	<b>0.836</b>	0.050	0.652	0.327	0.638	0.255
SCCN (Liu et al., 2018a)	0.449	0.048	0.158	0.259	0.080	<b>0.894</b>	0.096	0.185	0.054	<b>0.357</b>	0.049	0.078	0.148	0.303
CAAE (Luppino et al., 2022a)	0.266	0.406	0.421	0.318	0.095	0.184	0.458	<b>0.346</b>	0.078	0.067	0.065	0.214	0.231	0.256
ACE-Net (Luppino et al., 2022b)	0.368	0.417	0.412	0.307	0.735	0.766	0.420	0.204	0.083	0.318	0.105	0.351	0.354	0.394
X-Net (Luppino et al., 2022b)	0.363	0.474	0.381	0.373	0.707	0.707	0.448	0.218	0.068	0.284	0.092	0.445	0.343	0.417
SRGCAE (Chen et al., 2022)	0.240	0.149	0.469	<b>0.427</b>	<b>0.816</b>	0.157	0.720	0.117	0.669	0.058	<b>0.670</b>	<b>0.648</b>	0.598	0.259
Proposed SDIR	<b>0.585</b>	0.468	<b>0.554</b>	<b>0.543</b>	0.698	0.701	<b>0.791</b>	<b>0.274</b>	0.784	<b>0.618</b>	0.659	<b>0.747</b>	<b>0.678</b>	<b>0.559</b>

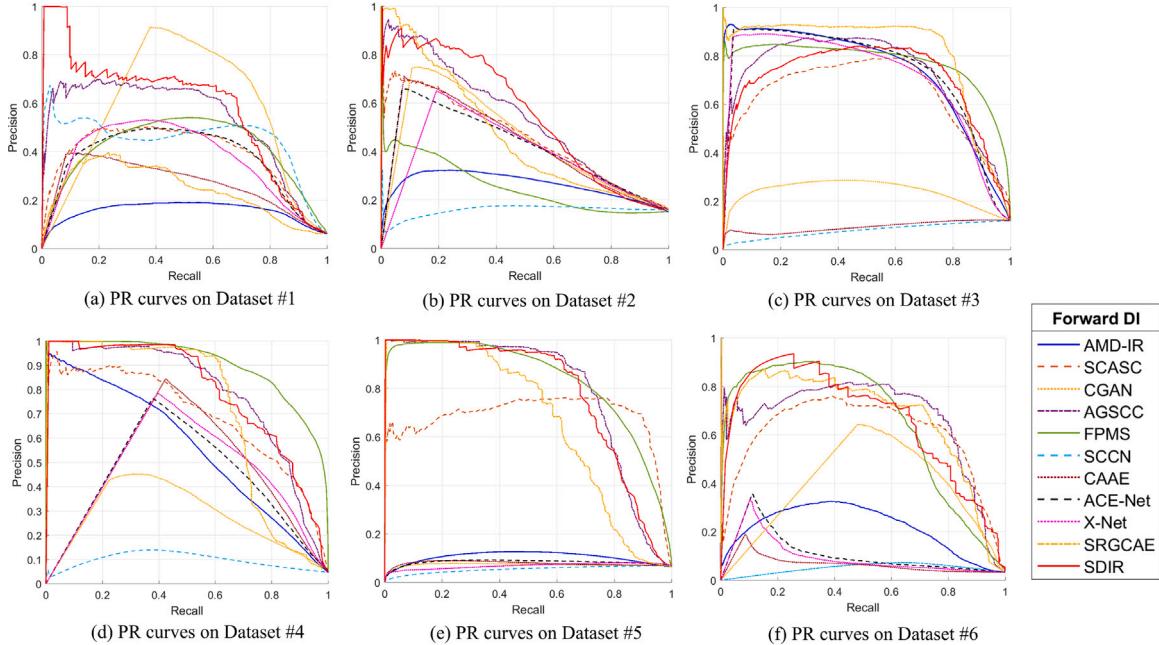


Fig. 6. PR curves of the forward DI generated by different methods. From (a) to (f) are the results on Datasets #1 to #6, respectively.

Unsurprisingly, the AUP values of backward DI obtained by SDIR are much higher than other methods on Datasets #2, #5 and #6. The average AUP values obtained for the forward and backward DIs of SDIR are 0.678 and 0.559 respectively, which are higher than the scores of all other methods. Specifically, the average AUP value of the backward DI is 9.4% higher than that of CGAN, which has the second highest score. From Fig. 6, Fig. 7 and Table 3, it can be inferred that the DI obtained by SDIR possesses a remarkable ability to detect changes, which enables the direct use of simple thresholding methods (Otsu, 1979) or clustering methods (Bezdek et al., 1984; Hartigan and Wong, 1979) for generating a relatively good CM.

#### 4.4. Change maps

To evaluate the detection performance of SDIR, we also show the CM generated by comparison methods, including not only the four methods compared in Figs. 4 and 5, but also the following seven methods, i.e., the fractal projection and Markovian segmentation based FPMS (Mignotte, 2020), the circular invariant convolution model (CICM) (Touati, 2019), the deep feature comparison based SCCN (Liu et al., 2018a), the code-aligned autoencoders (CAAE) (Luppino et al., 2022a), the AMD induced deep translation methods of ACE-Net (Luppino et al., 2022b) and X-Net (Luppino et al., 2022b), and the structural relationship graph convolutional autoencoder (SRGCAE) (Chen et al., 2022). To ensure a fair comparison, we selected the best results from

both the forward and backward processes of methods that only have one-way regression (or transformation) in their original papers, such as SCASC, CGAN, AGSCC, and FPMS.

Fig. 8 displays the CMs generated by various methods, where different colors are used to indicate TP, FP, TN, and FN. Tables 4 and 5 report the corresponding quantitative results. Since we selected the best results from the forward and backward CMs for the comparison methods, the CMs shown in Fig. 8 mostly detect changed areas. Nevertheless, some methods still lack robustness. For example, AMD-IR presents a large number of FP in Datasets #1, #2, #5, and #6; CICM and SCCN also have many false detections in Dataset #6, resulting in kappa coefficients of 0.024 and 0.183 respectively; CAAE have many miss detections in Datasets #3 and #5, resulting in very small kappa coefficients. As can be observed from Fig. 8 and Table 5, the proposed SDIR is able to get better results stably with small false alarms and miss detections, especially on the difficult Dataset #2 where SDIR performs far better than the other compared methods. The average F1 and  $\kappa$  of SDIR on the six datasets are 0.775 and 0.756 respectively, which are 5.3% and 5.6% higher than the second-ranked method. The results presented in Fig. 8, Tables 4 and 5 demonstrate the effectiveness and competitiveness of the proposed SDIR algorithm, even when compared to some deep learning-based methods, such as CGAN, SCCN, CAAE, ACE-Net, X-Net and SRGCAE.

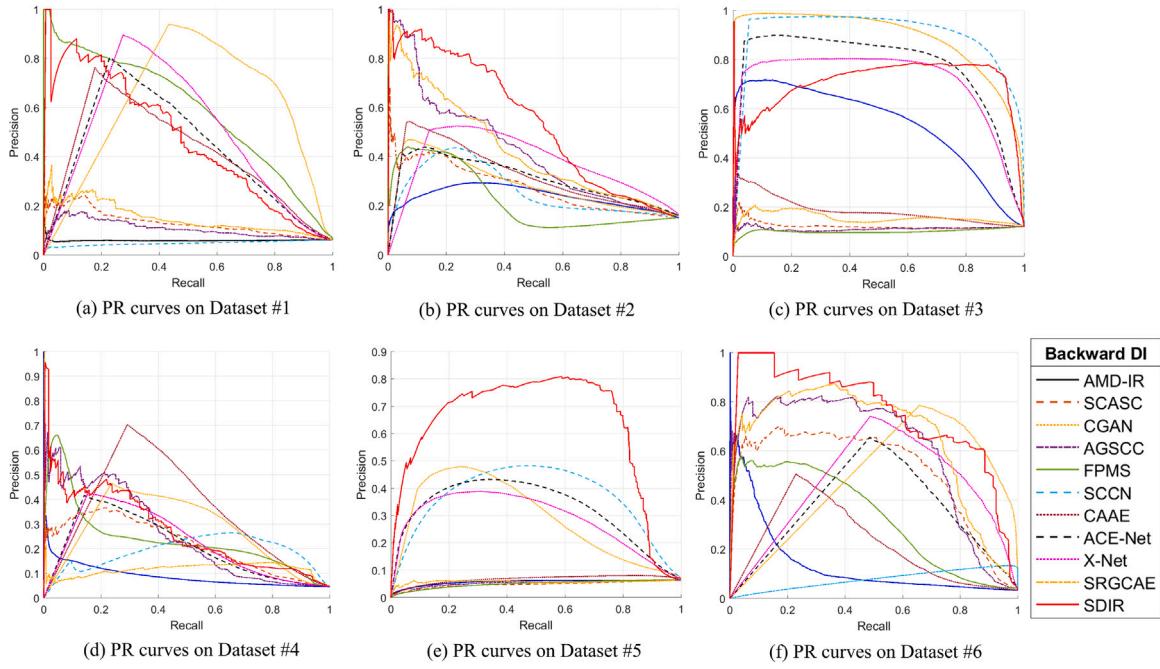


Fig. 7. PR curves of the backward DI generated by different methods. From (a) to (f) are the results on Datasets #1 to #6, respectively.

Table 4

MA and FA of CMs. The best and second best scores are marked in red and blue, respectively.

Methods	Dataset #1		Dataset #2		Dataset #3		Dataset #4		Dataset #5		Dataset #6		Average	
	MA	FA	MA	FA										
AMD-IR (Luppino et al., 2019)	0.206	0.200	0.366	0.259	0.153	0.095	0.198	0.043	0.421	0.204	0.047	0.323	0.232	0.187
SCASC (Sun et al., 2021b)	0.292	0.037	0.619	0.017	0.078	0.050	0.317	0.007	0.200	0.016	0.142	0.019	0.275	0.024
CGAN (Niu et al., 2019)	0.191	0.024	0.711	0.028	0.771	0.085	0.527	0.035	0.554	0.054	0.362	0.017	0.519	0.041
AGSCC (Sun et al., 2022b)	0.293	0.024	0.600	0.015	0.154	0.031	0.321	0.003	0.327	0.004	0.266	0.010	0.327	0.014
FPMS (Mignotte, 2020)	0.171	0.054	0.666	0.086	0.176	0.019	0.003	0.065	0.147	0.022	0.610	0.001	0.296	0.041
CICM (Touati, 2019)	0.525	0.023	0.525	0.023	0.525	0.023	0.525	0.023	0.525	0.023	0.525	0.023	0.525	0.023
SCCN (Liu et al., 2018a)	0.153	0.077	0.703	0.083	0.087	0.037	0.410	0.082	0.273	0.080	0.067	0.216	0.282	0.096
CAAE (Luppino et al., 2022a)	0.252	0.058	0.650	0.042	0.740	0.124	0.192	0.050	0.918	0.098	0.459	0.053	0.535	0.071
ACE-Net (Luppino et al., 2022b)	0.242	0.019	0.519	0.058	0.126	0.043	0.208	0.028	0.414	0.100	0.324	0.028	0.306	0.046
X-Net (Luppino et al., 2022b)	0.227	0.029	0.637	0.040	0.119	0.038	0.222	0.038	0.301	0.084	0.368	0.030	0.312	0.043
SRGCAE (Chen et al., 2022)	0.067	0.232	0.510	0.096	0.038	0.103	0.260	0.005	0.171	0.041	0.070	0.224	0.186	0.117
Proposed SDIR	0.352	0.015	0.424	0.020	0.053	0.041	0.133	0.012	0.187	0.006	0.216	0.009	0.227	0.017

Table 5

OA, F1 and  $\kappa$  of CMs. The best and second best scores are marked in red and blue, respectively.

Methods	Dataset #1			Dataset #2			Dataset #3			Dataset #4			Dataset #5			Dataset #6			Average		
	OA	F1	$\kappa$	OA	F1	$\kappa$															
AMD-IR (Luppino et al., 2019)	0.799	0.328	0.255	0.724	0.411	0.259	0.898	0.664	0.607	0.950	0.597	0.572	0.782	0.254	0.171	0.686	0.170	0.115	0.807	0.404	0.330
SCASC (Sun et al., 2021b)	0.947	0.621	0.593	0.892	0.516	0.464	0.946	0.804	0.773	0.979	0.751	0.741	0.973	0.788	0.774	0.977	0.711	0.700	0.952	0.699	0.674
CGAN (Niu et al., 2019)	0.965	0.742	0.724	0.863	0.402	0.338	0.833	0.248	0.155	0.943	0.432	0.402	0.911	0.412	0.364	0.971	0.599	0.584	0.914	0.473	0.428
AGSCC (Sun et al., 2022b)	0.959	0.680	0.658	0.897	0.540	0.490	0.955	0.817	0.791	0.983	0.782	0.773	0.976	0.779	0.766	0.982	0.733	0.724	0.959	0.722	0.700
FPMS (Mignotte, 2020)	0.938	0.625	0.593	0.827	0.368	0.269	0.962	0.837	0.816	0.938	0.597	0.569	0.970	0.786	0.770	0.979	0.553	0.544	0.936	0.628	0.594
CICM (Touati, 2019)	0.943	0.481	0.451	0.867	0.321	0.270	0.884	0.573	0.507	0.974	0.759	0.745	0.896	0.423	0.371	0.899	0.800	0.824	0.892	0.439	0.395
SCCN (Liu et al., 2018a)	0.919	0.562	0.522	0.818	0.342	0.240	0.957	0.835	0.810	0.903	0.359	0.315	0.907	0.521	0.474	0.789	0.232	0.183	0.882	0.475	0.424
CAAE (Luppino et al., 2022a)	0.930	0.569	0.534	0.861	0.446	0.373	0.803	0.239	0.126	0.943	0.568	0.540	0.844	0.069	0.014	0.933	0.355	0.324	0.886	0.374	0.314
ACE-Net (Luppino et al., 2022b)	0.967	0.740	0.723	0.868	0.538	0.463	0.947	0.798	0.768	0.964	0.670	0.651	0.878	0.401	0.341	0.962	0.546	0.526	0.931	0.616	0.579
X-Net (Luppino et al., 2022b)	0.958	0.696	0.674	0.864	0.461	0.389	0.952	0.816	0.789	0.954	0.609	0.586	0.901	0.497	0.447	0.959	0.510	0.490	0.931	0.598	0.563
SRGCAE (Chen et al., 2022)	0.778	0.341	0.268	0.838	0.491	0.395	0.905	0.706	0.654	0.983	0.800	0.791	0.951	0.682	0.656	0.781	0.222	0.173	0.873	0.541	0.490
Proposed SDIR	0.964	0.691	0.672	0.919	0.683	0.638	0.958	0.843	0.819	0.982	0.810	0.853	0.843	0.984	0.984	0.764	0.755	0.965	0.775	0.756	

#### 4.5. Discussions

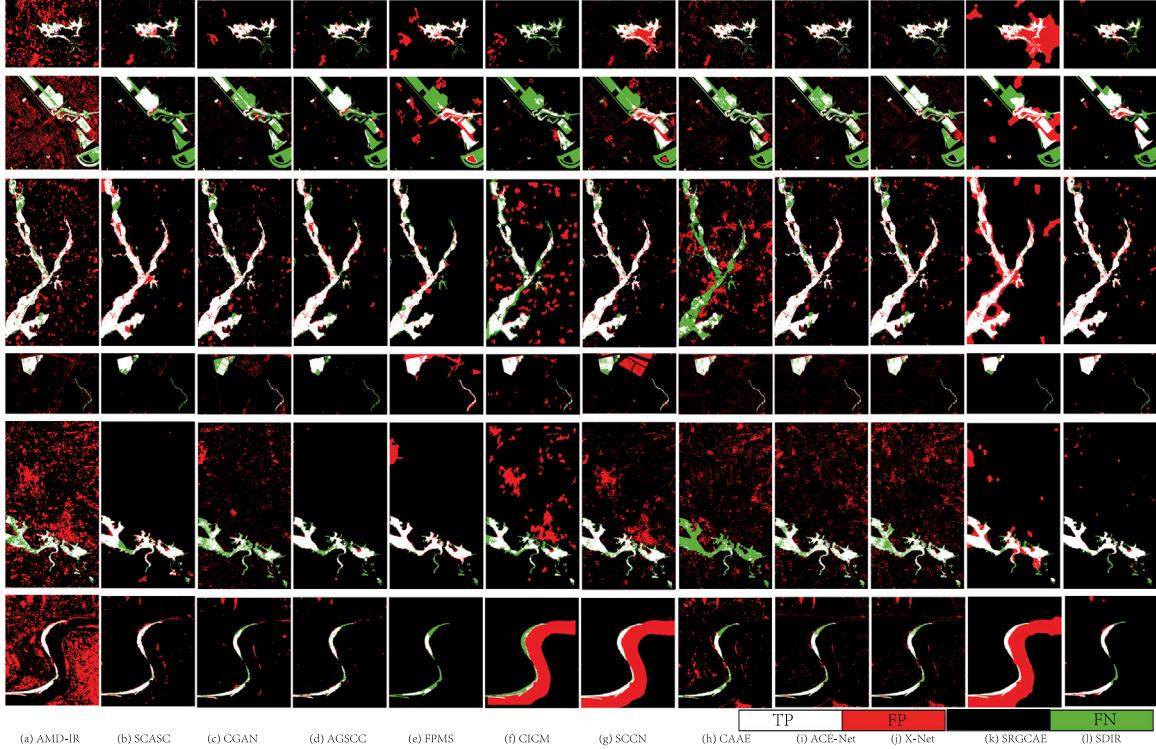
##### 4.5.1. Ablation study

Two main processes are included in the proposed SDIR: structure representation and structure regression, as shown in Fig. 2. The key techniques used in these two processes are analyzed separately below, i.e., higher-order neighborhood information (HNI) mining of KNN and KFN graphs in structure representation and the dissimilarity-based KFN graph induced DRR in structure regression. We construct a baseline model: structure representation without using HNI and structure regression without using DRR.

Table 6

Ablation study of SDIR measured by the average scores on all the evaluated datasets, where “B” represents the baseline model.

Settings	Forward DI		Backward DI		Final CM		
	AUP	AUP	AUP	AUP	OA	F1	$\kappa$
B	0.571	0.236	0.838	0.401	0.327		
B + HNI	0.651	0.283	0.851	0.446	0.383		
B + DRR	0.649	0.527	0.960	0.743	0.724		
B + HNI + DRR	0.678	0.559	0.965	0.775	0.756		



**Fig. 8.** CMs on multimodal datasets. From top to bottom, they correspond to Datasets #1 to #6, respectively. From left to right are the binary CMs generated by: (a) AMD-IR; (b) SCASC; (c) CGAN; (d) AGSCC; (e) FPMS; (f) CICM; (g) SCCN; (h) GAAE; (i) ACE-Net; (j) X-Net; (k) SRGCAE; (l) proposed SDIR. In the binary CM, White: true positives (TP); Red: false positives (FP); Black: true negatives (TN); Green: false negatives (FN). The original figure is available at <https://github.com/yulisun/SDIR>.

**Table 6** presents the average quantitative evaluation results of the DI and CM obtained from the model with and without the use of HNI and DRR on all the evaluated datasets. As evident from the results in **Table 6**, the performance of SDIR degrades a lot when HNI and DRR are not used. To be specific, by using the two neighborhood expansion principles proposed in Section 3.2 for constructing higher-order KNN graphs and higher-order KFN graphs, the higher-order structural information in the images can be preserved, which in turn makes the structure representation and structure regression more accurate. Specifically, adding HNI to the baseline model improves the AUP of forward DI and backward DI by 8.0% and 4.7% respectively; adding HNI to the baseline model with DRR improves the AUP of forward DI and backward DI by 2.9% and 3.2% respectively. By using the DRR of (14) that penalizes the similarity of superpixels connected by the KFN graphs, the structure regression model can make use of the dissimilarity relationships (high-frequency) of the image, and then obtain more accurate DI and CM, especially for the backward regression process. As presented in **Table 6**, when the dissimilarity base DRR term is added, the detection performance of SDIR is significantly improved, e.g., the average F1 is improved by nearly 0.342 compared to the baseline model. In addition, when using both HNI and DRR, the SDIR detection performance is improved even more significantly, e.g., its average F1 is improved by nearly 0.374 compared to the baseline model.

#### 4.5.2. Parameter analysis

The main parameters in SDIR are the superpixel number  $N_S$ , the balancing parameters of  $\lambda$  and  $\beta$  in the regression models of (19) and (20), and the imbalance parameter  $v$  of (24) in the MRF fusion segmentation.

In general, the choice of  $N_S$  should align with the image resolution and the granularity demands of the CD task. Opting for a larger  $N_S$  will result in smaller segmented superpixels, enhancing detection granularity. However, this also amplifies computational complexity, as discussed in the subsequent subsection of complexity analysis. In this paper, we simply set  $N_S \approx 2500$  as a compromise choice.

For the parameter  $\lambda$ , it is the weight of PSR, which is used to control the sparse level of changed feature matrix. Intuitively, it can be adjusted according to the percentage of the changed area. The value of  $\lambda$  should be small when the proportion of the changed area is known to be very low. **Fig. 9** shows the AUP values of forward and backward DIs obtained when  $\lambda$  takes different values (from  $2^{-8}$  to (1), from which it can be seen that SDIR is relatively robust to parameter  $\lambda$ . According to **Fig. 9**, we fix  $\lambda$  as 0.1 (i.e.  $\lambda \approx 2^{-3.3}$ ) for simplicity. In addition, in order to have a clearer view of the impact of  $\lambda$ , **Fig. 10** shows the DIs generated by SDIR with different  $\lambda$  on Dataset #6, from which it can be found that  $\lambda$  can control the sparsity of DI.

For the parameter  $\beta$ , it is the weight of DRR, which is used to adjust the weight of dissimilarity-based constraints in the regression model. As can be seen in **Fig. 11**, either too large or too small a  $\beta$  is inappropriate: first, too small a  $\beta$  would make the DRR weights little to play a corresponding role; second, too large a beta would make the model overly biased toward the DRR, destroying the balance among the constraint terms and tending to obtain the non-smooth DI. In addition, we can also find that the forward and backward regression processes have different requirements for the parameter  $\beta$ . Take Dataset #3 for example, it requires a larger  $\beta$  in the backward regression than in the forward regression. This is due to the fact that the post-event image in Dataset #3 also has low-pass properties on the KNN graph constructed on the pre-event image, so it is difficult to complete the backward regression using only the similarity-based constraints at this point. Therefore, it can be seen that the determination of  $\beta$  should be related to the spectral properties of pre- and post-event images on the graphs.

For the parameter  $v$ , it controls the rate of the imbalance penalties in the MRF fusion segmentation. In **Fig. 12**, we adjust  $v = 0, 0.5, 1, 1.5$  and plot the corresponding F1-scores on different datasets. From the results shown in **Fig. 12**, it can be observed that considering the class imbalance of change/unchanged categories in the MRF fusion segmentation model significantly improves the detection accuracy. The

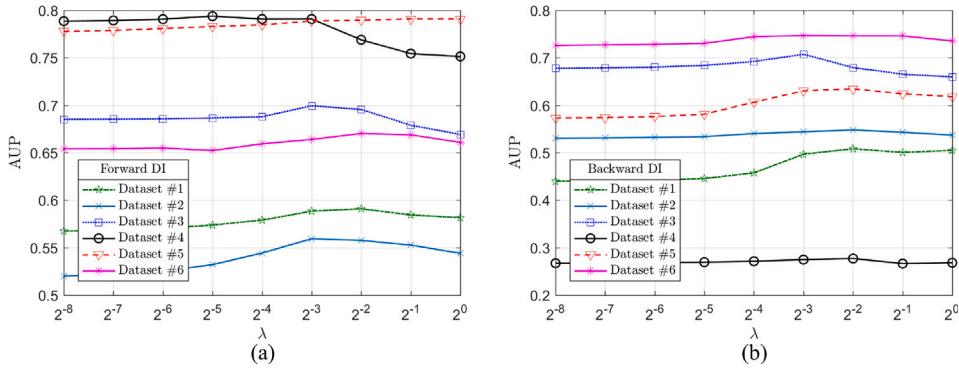


Fig. 9. Sensitivity analysis of parameter  $\lambda$ : (a) AUP of forward DI; (b) AUP of backward DI.

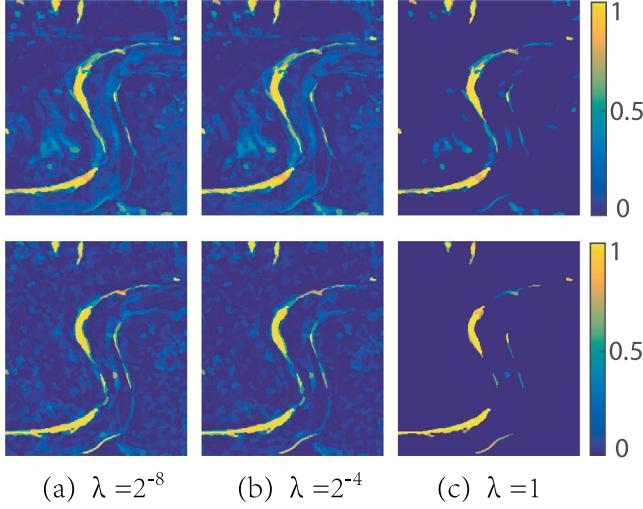


Fig. 10. Sensitivity analysis of parameter  $\lambda$  on Dataset #6. Top row is the forward DI, bottom is the backward DI, generated with different  $\lambda$ : (a)  $\lambda = 2^{-8}$ ; (b)  $\lambda = 2^{-4}$ ; (c)  $\lambda = 1$ .

F1 values are generally higher when  $v = 0.5, 1, 1.5$  compared to when  $v = 0$ . Additionally, for most datasets, the best results are achieved when the imbalance parameter  $v$  is set to 1. Therefore, we fix  $v = 1$  in this paper for simplicity.

#### 4.5.3. Computational complexity

The proposed SDIR consists of four processes: preprocessing, structure representation, structure regression fusion by solving the minimization models of (19) and (20), and the change extraction by using the MRF fusion segmentation.

- Pre-processing: the complexity of the GMMSP is linear in the number of pixels in the image  $\mathcal{O}(MN)$ , which is reported in Ban et al. (2018). The average number of pixels within each superpixel is  $MN/N_S$ , then the complexities of mean feature extraction and median feature extraction are around  $\mathcal{O}((B_x + B_y) MN)$  and  $\mathcal{O}((B_x + B_y) MN \log(MN/N_S))$ , respectively.

- Structure representation: first, calculating the distances between all the superpixels requires  $\mathcal{O}((B_x + B_y) N_S^2)$ . Second, sorting the distance matrix by column for finding the KNN and KFN requires  $\mathcal{O}(N_S^2 \log N_S)$  by using some accelerated sorting algorithms. Third, calculating the closed-form similarity vector  $W_i^{t1}$  by using (6) requires  $\mathcal{O}(N_S)$ . Fourth, constructing the high-order graphs requires  $\mathcal{O}(N_S^2)$ . Therefore, the structure representation by constructing high-order KNN graph and KFN graph requires  $\mathcal{O}(N_S^2 \log N_S)$ .

- Structural regression: first, updating  $\mathbf{X}'$  and  $\mathbf{Y}'$ . Taking  $\mathbf{Y}'$  update with (30) as an example, computing the matrix  $\Phi$  requires  $\mathcal{O}(B_y N_S^2)$ ,

the matrix multiplication requires  $\mathcal{O}(B_y N_S^2)$ . Second, updating  $\Delta^x$  and  $\Delta^y$  with (32). As  $\Delta^y$  can be solved column-wisely by using the closed-form operator, then it requires  $\mathcal{O}(B_y N_S)$ . Third, updating the Lagrange multipliers by using (33) requires  $\mathcal{O}(B_y N_S)$ . Therefore, the structural regression fusion requires  $\mathcal{O}((B_x + B_y) N_S^2)$ .

- Change extraction by using the MRF fusion segmentation method. As discussed in Sun et al. (2021a), the computation of energy terms requires  $\mathcal{O}(N_S)$  and  $\mathcal{O}(N_R)$  respectively, where  $N_R$  denotes the number of edges in the  $R$ -adjacency neighbor system. The investigation of using the min-cut/maxflow algorithm for energy minimization problem is presented in Boykov and Kolmogorov (2004). Specifically, the theoretical complexity in the worst-case scenario is  $\mathcal{O}(2N_R N_S^2)$ , while the empirical complexity tends to be relatively low for typical instances in vision problems.

Due to the inclusion of both inner and outer loops in the process of structural regression, and the fact that matrix  $\Phi$  needs to be updated in each iteration of the inner loop, it is actually the structural regression is the most time-consuming process in SDIR. The MATLAB running time of SDIR on the Dataset #1 is 24.21 s, with each of the four processes consuming 0.14, 1.46, 22.18, and 0.43 s respectively, in an environment of Windows Laptop with Intel Core i9-10980HK CPU and 64 GB of RAM.

#### 4.5.4. Spectral domain analysis of SDIR

Define the graph Laplacian matrix as  $\mathbf{L}$ , it can be decomposed into  $\mathbf{L} = \mathbf{U}\Lambda\mathbf{U}^{-1}$ , with  $\mathbf{U}$  denoting the orthonormal matrix of the eigenvectors  $\mathbf{u}_k$  in its column and  $\Lambda$  denoting the diagonal matrix of the corresponding eigenvalues  $\lambda_k$ . The graph Fourier transform (GFT) of a graph signal  $\mathbf{f}$  can be mathematically expressed as

$$\hat{\mathbf{f}} = \text{GFT}(\mathbf{f}) = \mathbf{U}^{-1}\mathbf{f}. \quad (26)$$

The inverse GFT can be expressed as

$$\mathbf{f} = \text{IGFT}(\hat{\mathbf{f}}) = \mathbf{U}\hat{\mathbf{f}}. \quad (27)$$

If we sort the eigenspectra of graph Laplacian matrix  $\mathbf{L}$  as  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{N_S}$ , then  $\lambda_{N_S}$  represents the highest frequency and  $\lambda_1$  represents the lowest frequency (Stankovic et al., 2019; Stanković et al., 2019).

In the previous work (Sun et al., 2022a), it treats each feature matrix as a graph signal on the constructed KNN graph, and converts the multimodal CD into a graph signal processing problem. Based on the defined graph and graph signals, it has analyzed the spectral properties of the image on its own constructed similarity based KNN graph, that is, both the regressed signal  $\mathbf{Y}'$  and original signal  $\mathbf{X}$  are approximate low-pass signals on the KNN graph of  $G_n^{t1}$ , and both the regressed signal  $\mathbf{X}'$  and original signal  $\mathbf{Y}$  are approximate low-pass signals on the KNN graph of  $G_n^{t2}$ . The high-frequency components of signals  $\mathbf{X}$  on  $G_n^{t1}$  and  $\mathbf{Y}$  on  $G_n^{t1}$  are introduced by the changes as illustrated by Fig. 13, where the spectra are defined as  $\hat{\mathbf{X}}_n^{t1} = (\mathbf{U}_n^{t1})^{-1} \mathbf{X}$ ,  $\hat{\mathbf{Y}}_n^{t1} = (\mathbf{U}_n^{t1})^{-1} \mathbf{Y}$ ,  $\hat{\mathbf{Y}}_n^{t2} = (\mathbf{U}_n^{t2})^{-1} \mathbf{Y}$ , and  $\hat{\mathbf{X}}_n^{t2} = (\mathbf{U}_n^{t2})^{-1} \mathbf{X}$ , and the orthonormal matrices of  $\mathbf{U}_n^{t1}$  and  $\mathbf{U}_n^{t2}$  are computed from the Laplacian matrices of  $G_n^{t1}$  and  $G_n^{t2}$  respectively.

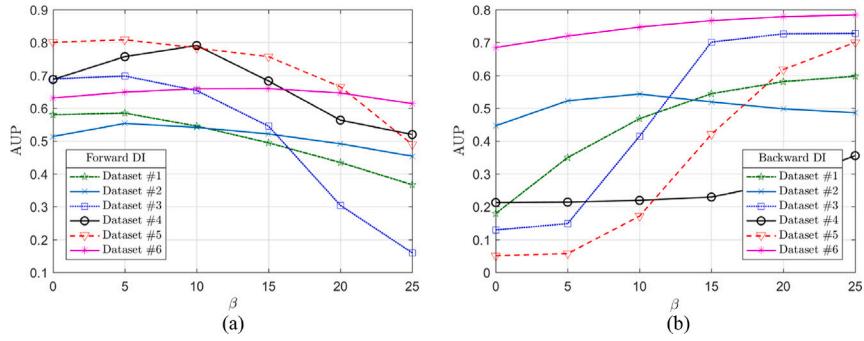


Fig. 11. Sensitivity analysis of parameter  $\beta$ : (a) AUP of forward DI; (b) AUP of backward DI.

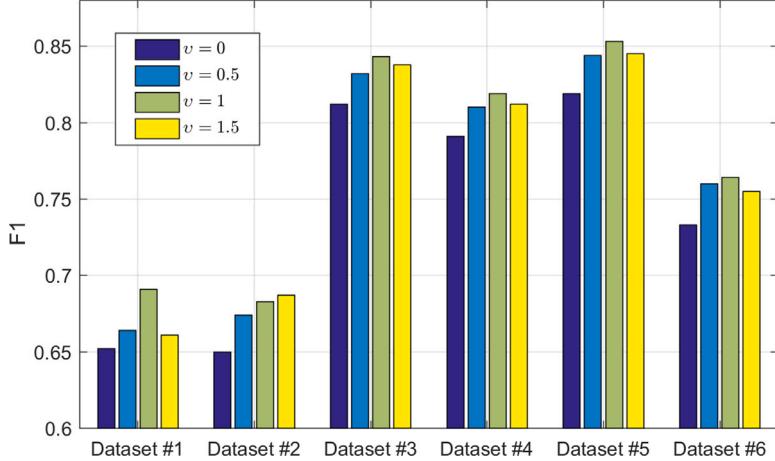


Fig. 12. Sensitivity analysis of parameter  $v$ .

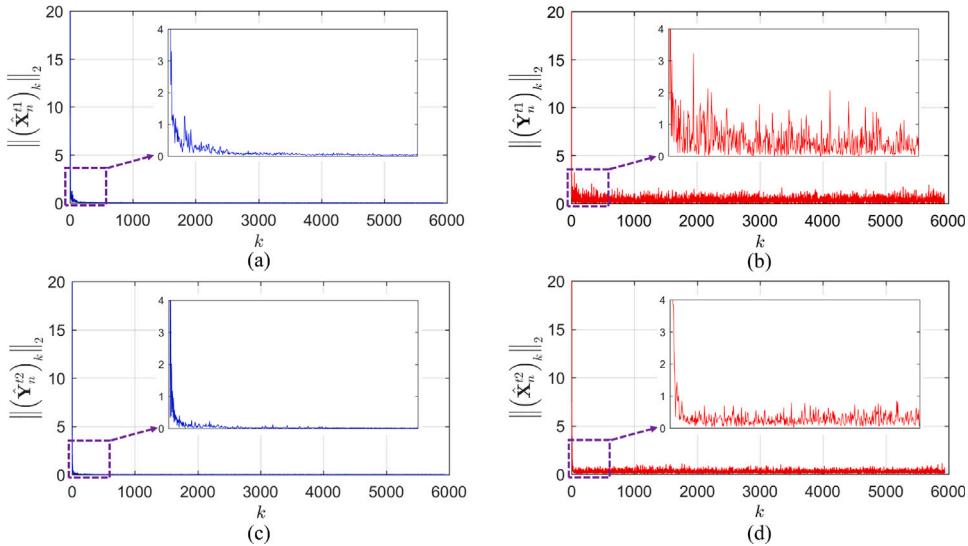
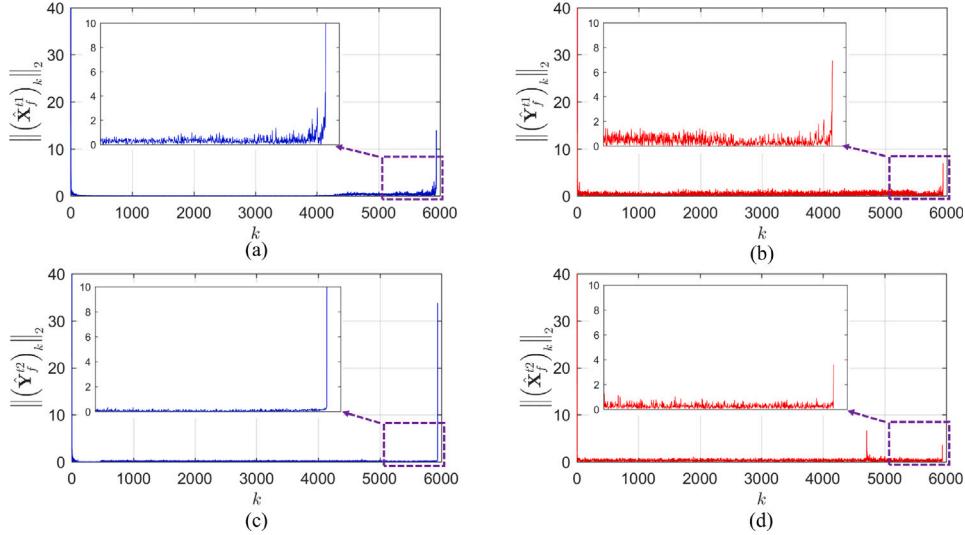


Fig. 13. Spectral properties of graph signals  $\mathbf{X}$  and  $\mathbf{Y}$  on the KNN graphs  $G_n^{t1}$  and  $G_n^{t2}$ . (a)  $\mathbf{X}$  on the  $G_n^{t1}$ ; (b)  $\mathbf{Y}$  on the  $G_n^{t1}$ ; (c)  $\mathbf{Y}$  on the  $G_n^{t2}$ ; (d)  $\mathbf{X}$  on the  $G_n^{t2}$ . The graph signals and KNN graphs are constructed from the multitemporal images of Dataset #1 in Fig. 4.

From Fig. 13 and the spectral analysis in Sun et al. (2022a), it can be seen that the KNN graph mainly captures low-frequency information of the image structure, while the SRR (13) in the regression model (19) constrains that the regression image and original image have the same low-frequency characteristics on the similarity based KNN graph.

In Fig. 14, we plots the spectral properties of graph signals  $\mathbf{X}$  and  $\mathbf{Y}$  on the KFN graphs of  $G_f^{t1}$  and  $G_f^{t2}$  respectively, where the spectra

are defined as  $\hat{\mathbf{X}}_f^{t1} = (\mathbf{U}_f^{t1})^{-1} \mathbf{X}$ ,  $\hat{\mathbf{Y}}_f^{t1} = (\mathbf{U}_f^{t1})^{-1} \mathbf{Y}$ ,  $\hat{\mathbf{Y}}_f^{t2} = (\mathbf{U}_f^{t2})^{-1} \mathbf{Y}$ , and  $\hat{\mathbf{X}}_f^{t2} = (\mathbf{U}_f^{t2})^{-1} \mathbf{X}$ , and the orthonormal matrices of  $\mathbf{U}_f^{t1}$  and  $\mathbf{U}_f^{t2}$  are computed from the Laplacian matrices of  $G_f^{t1}$  and  $G_f^{t2}$  respectively. From Fig. 14, we can find that the KFN graph can capture the high-frequency information of the image structure. This can also be illustrated by the construction process of the KFN, i.e., its edges are connected to superpixels representing different types of objects, so the variation level



**Fig. 14.** Spectral properties of graph signals on the KFN graphs  $G_f^{t1}$  and  $G_f^{t2}$ . (a)  $\mathbf{X}$  on the  $G_f^{t1}$ ; (b)  $\mathbf{Y}$  on the  $G_f^{t1}$ ; (c)  $\mathbf{Y}$  on the  $G_f^{t2}$ ; (d)  $\mathbf{X}$  on the  $G_f^{t2}$ . The graph signals and KFN graphs are constructed from the multitemporal images of Dataset #1 in Fig. 4.

between the graph signals connected by the edges in the KFN graph is very dramatic, i.e., the graph signals have a large high-frequency component on the  $G_f^{t1}$  and  $G_f^{t2}$ . The DRR (14) in the regression model (19) constrains that the regression image and the original image have the same high-frequency characteristics on the dissimilarity based KFN graph. Therefore, the proposed SDIR can incorporate both similarity (low-frequency) and dissimilarity (high-frequency) relationships of images by combining the high-order KNN graph and KFN graph.

## 5. Conclusion

This paper focuses on the problem of multimodal CD in remote sensing. We first analyze the previous structural regression-based methods and found the shortcomings of these methods using only similarity relationships. Then, we propose a similarity and dissimilarity induced image regression method for the multimodal CD. The proposed method constructs high-order KNN and KFN graphs to capture the similarity and dissimilarity relationships of the image, and then utilizes structure consistency constraints to decompose the target image into a regression image and a change image, which requires the regression image to share the same spectral properties as the original image on the KFN and KNN graphs. Once the forward and backward image regression are completed, an MRF based fusion segmentation model is employed to combine the change fusion and change extraction processes. The experimental results demonstrate the effectiveness of the proposed method under various scenarios.

As elaborated in Section 3.1, when employing the image superpixel co-segmentation method, there is a certain loss of information as only the grayscale information of the images is utilized in constructing the false RGB image, which may result in inaccuracies in superpixel segmentation. Therefore, future research involves the use of more advanced superpixel segmentation methods to ensure that segmented superpixels are internally homogeneous within their respective images.

The analysis of graphs has prompted us to focus on dissimilarity relationships and high-frequency components of graph spectra. In the future, we will explore the potential of incorporating graph wavelet techniques to better capture the structure of images and extend the dissimilarity relationships to graph neural network based multimodal CD (Florez-Ospina et al., 2023), thus improving the CD performance.

## Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No. 61971426. The authors would like to thank the editors and anonymous reviewers for their constructive suggestions that improved the presentation of this work.

## Appendix. Optimization of fused regression model

We use the alternating direction method of multipliers for solving the minimization problem of (20), the augmented Lagrangian function of (20) is

$$\begin{aligned} \Theta(\mathbf{Y}', \Delta^y, \mathbf{R}) = & 2\text{Tr}(\mathbf{Y}' \mathbf{L}'^T \mathbf{Y}'^T) + \beta \sum_{i,j=1}^{N_S} \frac{(W_{h-f}^{t1})_{i,j}}{\|\mathbf{Y}'_i - \mathbf{Y}'_j\|_2^2 + \epsilon} \\ & + \lambda \|\Delta^y\|_{2,1} + \text{Tr}(\mathbf{R}^T (\mathbf{Y}' - \mathbf{Y} - \Delta^y)) \\ & + \frac{\mu}{2} \|\mathbf{Y}' - \mathbf{Y} - \Delta^y\|_F^2, \end{aligned} \quad (28)$$

where  $\mathbf{R} \in \mathbb{R}^{2b_y \times N_S}$  is a Lagrange multiplier, and  $\mu > 0$  is a penalty parameter that affects the convergence performance of the ADMM (Boyd et al., 2011), which is fixed to 0.4 in the experiment. The optimization of (28) can be decomposed into three subproblems.

(1)  $\mathbf{Y}'$  subproblem. The optimization problem for minimizing (28) with respect to  $\mathbf{Y}'$  can be written as

$$\begin{aligned} \min_{\mathbf{Y}'} & 2\text{Tr}(\mathbf{Y}' \mathbf{L}'^T \mathbf{Y}'^T) + \beta \sum_{i,j=1}^{N_S} \frac{(W_{h-f}^{t1})_{i,j}}{\|\mathbf{Y}'_i - \mathbf{Y}'_j\|_2^2 + \epsilon} \\ & + \text{Tr}(\mathbf{R}^T (\mathbf{Y}' - \mathbf{Y} - \Delta^y)) + \frac{\mu}{2} \|\mathbf{Y}' - \mathbf{Y} - \Delta^y\|_F^2. \end{aligned} \quad (29)$$

The gradient descent method is used for this minimization problem. We set the iteration number of inner loop for  $\mathbf{Y}'$  subproblem as  $N_i$  and the

step size as  $\tau$ , then we have

$$\begin{aligned} g(\mathbf{Y}') &= \mathbf{Y}' \left( \mu \mathbf{I}_{N_S} + 4 (\mathbf{L}^{11} - \beta \Phi) \right) + \mathbf{R} - \mu (\mathbf{Y} + \Delta^y), \\ \mathbf{Y}' &\leftarrow \mathbf{Y}' - \tau g(\mathbf{Y}'), \end{aligned} \quad (30)$$

where  $\Phi = \Phi^1 - \Phi^2$  with  $\Phi_{i,j}^2 = \frac{(W_{h-f}^{11})_{i,j}}{\left( \|Y'_i - Y'_j\|_2^2 + \epsilon \right)^2}$  and  $\Phi^1 \in \mathbb{R}^{N_S \times N_S}$  being a diagonal matrix with  $\Phi_{i,i}^1 = \sum_{j=1}^{N_S} \frac{\Phi_{i,j}^2 + \Phi_{j,i}^2}{2}$ .

(2)  $\Delta^y$  subproblem. The optimization problem for minimizing (28) with respect to  $\Delta^y$  can be written as

$$\min_{\Delta^y} \lambda \|\Delta^y\|_{2,1} + \frac{\mu}{2} \left\| \Delta^y + \mathbf{Y} - \mathbf{Y}' - \frac{\mathbf{R}}{\mu} \right\|_F^2. \quad (31)$$

The closed-form solution of (31) can be obtained from (Yang et al., 2009) as

$$\Delta_i^y = \max \left\{ \|\mathbf{Q}_i\|_2 - \frac{\lambda}{\mu} \right\} \frac{\mathbf{Q}_i}{\|\mathbf{Q}_i\|_2}, \quad (32)$$

where  $\mathbf{Q} = \mathbf{Y}' - \mathbf{Y} + \frac{\mathbf{R}}{\mu}$  and we follow  $0 \cdot (0/0) = 0$ .

(3) Multiplier updating. Finally, we update the Lagrangian multiplier of  $\mathbf{R}$  by using

$$\mathbf{R} \leftarrow \mathbf{R} + \mu (\mathbf{Y}' - \mathbf{Y} - \Delta^y). \quad (33)$$

## References

- Ban, Z., Liu, J., Cao, L., 2018. Superpixel segmentation using Gaussian mixture model. *IEEE Trans. Image Process.* 27 (8), 4105–4117.
- Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* 10 (2–3), 191–203.
- Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al., 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* 3 (1), 1–122.
- Boykov, Y., Kolmogorov, V., 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (9), 1124–1137.
- Camps-Valls, G., Gomez-Chova, L., Munoz-Mari, J., Rojo-Alvarez, J.L., Martinez-Ramon, M., 2008. Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. *IEEE Trans. Geosci. Remote Sens.* 46 (6), 1822–1835.
- Chen, Y., Bruzzone, L., 2022. Self-supervised change detection in multiview remote sensing images. *IEEE Trans. Geosci. Remote Sens.* 60, 1–12.
- Chen, H., Song, J., Wu, C., Du, B., Yokoya, N., 2023a. Exchange means change: An unsupervised single-temporal change detection framework based on intra-and inter-image patch exchange. *ISPRS J. Photogramm. Remote Sens.* 206, 87–105.
- Chen, H., Yokoya, N., Chini, M., 2023b. Fourier domain structural relationship analysis for unsupervised multimodal change detection. *ISPRS J. Photogramm. Remote Sens.* 198, 99–114.
- Chen, H., Yokoya, N., Wu, C., Du, B., 2022. Unsupervised multimodal change detection based on structural relationship graph representation learning. *IEEE Trans. Geosci. Remote Sens.* 60, 1–18.
- Deng, W., Liao, Q., Zhao, L., Guo, D., Kuang, G., Hu, D., Liu, L., 2021. Joint clustering and discriminative feature alignment for unsupervised domain adaptation. *IEEE Trans. Image Process.* 30, 7842–7855.
- Ferraris, V., Dobigeon, N., Chabert, M., 2020. Robust fusion algorithms for unsupervised change detection between multi-band optical images—A comprehensive case study. *Inf. Fusion* 64, 293–317.
- Ferraris, V., Dobigeon, N., Wei, Q., Chabert, M., 2017. Robust fusion of multiband images with different spatial and spectral resolutions for change detection. *IEEE Trans. Comput. Imaging* 3 (2), 175–186.
- Ferraris, V., Dobigeon, N., Wei, Q., Chabert, M., 2018. Detecting changes between optical images of different spatial and spectral resolutions: A fusion-based approach. *IEEE Trans. Geosci. Remote Sens.* 56 (3), 1566–1578.
- Florez-Ospina, J.F., Jimenez-Sierra, D.A., Benitez-Restrepo, H.D., Arce, G.R., 2023. Exploiting variational inequalities for generalized change detection on graphs. *IEEE Trans. Geosci. Remote Sens.* 61, 1–16.
- Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* (6), 721–741.
- Han, T., Tang, Y., Yang, X., Lin, Z., Zou, B., Feng, H., 2021. Change detection for heterogeneous remote sensing images with improved training of hierarchical extreme learning machine (HELM). *Remote Sens.* 13 (23), 4918.
- Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A k-means clustering algorithm. *J. R. Stat. Soc. Ser. C Appl. Stat.* 28 (1), 100–108.
- Jia, M., Zhang, C., Lv, Z., Zhao, Z., Wang, L., 2022. Bipartite adversarial autoencoders with structural self-similarity for unsupervised heterogeneous remote sensing image change detection. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5.
- Jiang, X., Li, G., Liu, Y., Zhang, X.-P., He, Y., 2020. Change detection in heterogeneous optical and SAR remote sensing images via deep homogeneous feature fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 13, 1551–1566.
- Li, X., Du, Z., Huang, Y., Tan, Z., 2021. A deep translation (GAN) based change detection network for optical and SAR remote sensing images. *ISPRS J. Photogramm. Remote Sens.* 179, 14–34.
- Li, H.-C., Yang, G., Yang, W., Du, Q., Emery, W.J., 2020. Deep nonsmooth nonnegative matrix factorization network with semi-supervised learning for SAR image change detection. *ISPRS J. Photogramm. Remote Sens.* 160, 167–179.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P., 2017. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision. ICCV.
- Liu, J., Gong, M., Qin, K., Zhang, P., 2018a. A deep convolutional coupling network for change detection based on heterogeneous optical and radar images. *IEEE Trans. Neural Netw. Learn. Syst.* 29 (3), 545–559.
- Liu, L., Hong, D., Ni, L., Gao, L., 2022a. Multilayer cascade screening strategy for semi-supervised change detection in hyperspectral images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 15, 1926–1940.
- Liu, Z., Li, G., Mercier, G., He, Y., Pan, Q., 2018b. Change detection in heterogeneous remote sensing images via homogeneous pixel transformation. *IEEE Trans. Image Process.* 27 (4), 1822–1834.
- Liu, Z.-g., Mercier, G., Dezert, J., Pan, Q., 2014. Change detection in heterogeneous remote sensing images based on multidimensional evidential reasoning. *IEEE Geosci. Remote Sens. Lett.* 11 (1), 168–172.
- Liu, J., Zhang, W., Liu, F., Xiao, L., 2022b. A probabilistic model based on bipartite convolutional neural network for unsupervised change detection. *IEEE Trans. Geosci. Remote Sens.* 60, 1–14.
- Liu, Z., Zhang, Z., Pan, Q., Ning, L., 2022c. Unsupervised change detection from heterogeneous data based on image translation. *IEEE Trans. Geosci. Remote Sens.* 60, 1–13.
- Luppino, L.T., Bianchi, F.M., Moser, G., Anfinsen, S.N., 2019. Unsupervised image regression for heterogeneous change detection. *IEEE Trans. Geosci. Remote Sens.* 57 (12), 9960–9975.
- Luppino, L.T., Hansen, M.A., Kampffmeyer, M., Bianchi, F.M., Moser, G., Jenssen, R., Anfinsen, S.N., 2022a. Code-aligned autoencoders for unsupervised change detection in multimodal remote sensing images. *IEEE Trans. Neural Netw. Learn. Syst.* 1–13.
- Luppino, L.T., Kampffmeyer, M., Bianchi, F.M., Moser, G., Serpico, S.B., Jenssen, R., Anfinsen, S.N., 2022b. Deep image translation with an affinity-based change prior for unsupervised multimodal change detection. *IEEE Trans. Geosci. Remote Sens.* 60, 1–22.
- Lv, Z., Huang, H., Gao, L., Benediktsson, J.A., Zhao, M., Shi, C., 2022a. Simple multiscale UNet for change detection with heterogeneous remote sensing images. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5.
- Lv, Z., Huang, H., Li, X., Zhao, M., Benediktsson, J.A., Sun, W., Falco, N., 2022b. Land cover change detection with heterogeneous remote sensing images: Review, progress, and perspective. *Proc. IEEE* 110 (12), 1976–1991.
- Lv, Z., Liu, T., Benediktsson, J.A., Falco, N., 2021. Land cover change detection techniques: Very-high-resolution optical images: A review. *IEEE Geosci. Remote Sens. Mag.* 2–21.
- Mercier, G., Moser, G., Serpico, S.B., 2008. Conditional copulas for change detection in heterogeneous remote sensing images. *IEEE Trans. Geosci. Remote Sens.* 46 (5), 1428–1441.
- Mignotte, M., 2020. A fractal projection and Markovian segmentation-based approach for multimodal change detection. *IEEE Trans. Geosci. Remote Sens.* 58 (11), 8046–8058.
- Niu, X., Gong, M., Zhan, T., Yang, Y., 2019. A conditional adversarial network for change detection in heterogeneous images. *IEEE Geosci. Remote Sens. Lett.* 16 (1), 45–49.
- Ortega, A., Frossard, P., Kovačević, J., Moura, J.M., Vandergheynst, P., 2018. Graph signal processing: Overview, challenges, and applications. *Proc. IEEE* 106 (5), 808–828.
- Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* 9 (1), 62–66.
- Prendes, J., Chabert, M., Pascal, F., Giros, A., Tourneret, J.-Y., 2015. A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors. *IEEE Trans. Image Process.* 24 (3), 799–812.
- Shuman, D.I., Narang, S.K., Frossard, P., Ortega, A., Vandergheynst, P., 2013. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process. Mag.* 30 (3), 83–98.
- SINGH, A., 1989. Review article digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* 10 (6), 989–1003.
- Stanković, L., Daković, M., Sejdic, E., 2019. Introduction to graph signal processing. In: Vertex-Frequency Analysis of Graph Signals. Springer, pp. 3–108.
- Stankovic, L., Mandic, D.P., Dakovic, M., Kisil, I., Sejdic, E., Constantinides, A.G., 2019. Understanding the basis of graph signal processing via an intuitive example-driven approach [Lecture Notes]. *IEEE Signal Process. Mag.* 36 (6), 133–145.

- Sun, Y., Lei, L., Guan, D., Kuang, G., 2021a. Iterative robust graph for unsupervised change detection of heterogeneous remote sensing images. *IEEE Trans. Image Process.* 30, 6277–6291.
- Sun, Y., Lei, L., Guan, D., Kuang, G., Liu, L., 2022a. Graph signal processing for heterogeneous change detection. *IEEE Trans. Geosci. Remote Sens.* 60, 1–23.
- Sun, Y., Lei, L., Guan, D., Li, M., Kuang, G., 2021b. Sparse-constrained adaptive structure consistency-based unsupervised image regression for heterogeneous remote sensing change detection. *IEEE Trans. Geosci. Remote Sens.* 1–14.
- Sun, Y., Lei, L., Guan, D., Wu, J., Kuang, G., Liu, L., 2022b. Image regression with structure cycle consistency for heterogeneous change detection. *IEEE Trans. Neural Netw. Learn. Syst.* 1–15.
- Sun, Y., Lei, L., Li, X., Sun, H., Kuang, G., 2021c. Nonlocal patch similarity based heterogeneous remote sensing change detection. *Pattern Recognit.* 109, 107598.
- Sun, Y., Lei, L., Liu, L., Kuang, G., 2023. Structural regression fusion for unsupervised multimodal change detection. *IEEE Trans. Geosci. Remote Sens.* 61, 1–18.
- Sun, Y., Lei, L., Tan, X., Guan, D., Wu, J., Kuang, G., 2022c. Structured graph based image regression for unsupervised multimodal change detection. *ISPRS J. Photogramm. Remote Sens.* 185, 16–31.
- Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., Rother, C., 2008. A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *IEEE Trans. Pattern Anal. Mach. Intell.* 30 (6), 1068–1080.
- Touati, R., 2019. Détection de changement en imagerie satellitaire multimodale (Ph.D. thesis) Université de Montréal.
- Touati, R., Mignotte, M., 2018. An energy-based model encoding nonlocal pairwise pixel interactions for multisensor change detection. *IEEE Trans. Geosci. Remote Sens.* 56 (2), 1046–1058.
- Touati, R., Mignotte, M., Dahmane, M., 2020. Multimodal change detection in remote sensing images using an unsupervised pixel pairwise-based Markov random field model. *IEEE Trans. Image Process.* 29, 757–767.
- Volpi, M., Camps-Valls, G., Tuia, D., 2015. Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis. *ISPRS J. Photogramm. Remote Sens.* 107, 50–63, Multitemporal remote sensing data analysis.
- Wan, L., Xiang, Y., You, H., 2019. An object-based hierarchical compound classification method for change detection in heterogeneous optical and SAR images. *IEEE Trans. Geosci. Remote Sens.* 57 (12), 9941–9959.
- Wan, L., Zhang, T., You, H., 2018. Multi-sensor remote sensing image change detection based on sorted histograms. *Int. J. Remote Sens.* 39 (11), 3753–3775.
- Wu, C., Du, B., Zhang, L., 2023. Fully convolutional change detection framework with generative adversarial network for unsupervised, weakly supervised and regional supervised change detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 1–15.
- Wu, Y., Li, J., Yuan, Y., Qin, A.K., Miao, Q.-G., Gong, M.-G., 2021. Commonality autoencoder: Learning common features for change detection from heterogeneous images. *IEEE Trans. Neural Netw. Learn. Syst.* 1–14.
- Yang, J., Yin, W., Zhang, Y., Wang, Y., 2009. A fast algorithm for edge-preserving variational multichannel image restoration. *SIAM J. Imaging Sci.* 2 (2), 569–592.
- Zhang, P., Gong, M., Su, L., Liu, J., Li, Z., 2016. Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images. *ISPRS J. Photogramm. Remote Sens.* 116, 24–41.
- Zhang, W., Jiao, L., Liu, F., Yang, S., Song, W., Liu, J., 2022. Sparse feature clustering network for unsupervised SAR image change detection. *IEEE Trans. Geosci. Remote Sens.* 60, 1–13.
- Zhang, X., Su, H., Zhang, C., Gu, X., Tan, X., Atkinson, P.M., 2021. Robust unsupervised small area change detection from SAR imagery using deep learning. *ISPRS J. Photogramm. Remote Sens.* 173, 79–94.
- Zhang, C., Yue, P., Tapete, D., Jiang, L., Shangguan, B., Huang, L., Liu, G., 2020. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. *ISPRS J. Photogramm. Remote Sens.* 166, 183–200, Available: <https://www.sciencedirect.com/science/article/pii/S0924271620301532>.
- Zheng, X., Chen, X., Lu, X., Sun, B., 2022a. Unsupervised change detection by cross-resolution difference learning. *IEEE Trans. Geosci. Remote Sens.* 60, 1–16.
- Zheng, X., Guan, D., Li, B., Chen, Z., Li, X., 2022b. Change smoothness-based signal decomposition method for multimodal change detection. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5.