

Structural Regression Fusion for Unsupervised Multimodal Change Detection

Yuli Sun[✉], Lin Lei[✉], Li Liu[✉], Senior Member, IEEE, and Gangyao Kuang, Senior Member, IEEE

Abstract—Multimodal change detection (MCD) is an increasingly interesting but very challenging topic in remote sensing, which is due to the unavailability of detecting changes by directly comparing multimodal images from different domains. In this article, we first analyze the structural asymmetry between multitemporal images and show their negative impact on the previous MCD methods using image structures. Specifically, when there is a structural asymmetry, previous structure-based methods can only complete a structure comparison or image regression in one direction and fail in the other direction; that is, they cannot transform or convert from complex structural images (with more categories) to simple structural images (with fewer categories). To reduce the influence of structural asymmetry, we propose a structural regression fusion (SRF)-based method that simultaneously transforms the pre-event and post-event images into the image domain of each other, calculating the forward and backward changed images, respectively. Noteworthy, different from previous late fusion methods that fuse the forward and backward changed images in the postprocessing stage, SRF incorporates fusion into the regression process, which can fully explore the connection between changed images and, thus, improve image transformation performance and obtain better changed images. Specifically, SRF yields three types of constraints to perform the fused image transformation: structure consistency-based regression term, change smoothness and alignment-based fusion term, and prior sparsity-based penalty term. Finally, the changes can be extracted by comparing the transformed and original images. The proposed SRF is verified on six real datasets by comparing with some state-of-the-art (SOTA) methods. Source code of the proposed method will be made available at <https://github.com/yulisun/SRF>.

Index Terms—Change detection (CD), fusion, image regression, multimodal data, structural asymmetry.

I. INTRODUCTION

A. Background

CHANGE detection (CD) of remote sensing images aims at identifying changes on the Earth's surface by jointly analyzing two or more co-registered images acquired over the same geographical area but at different epochs [1]. CD plays an important role in Earth observation applications, such

Manuscript received 13 March 2023; revised 31 May 2023; accepted 21 June 2023. Date of publication 13 July 2023; date of current version 27 July 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFB3100800 and in part by the National Natural Science Foundation of China under Grant 61872379 and Grant 61971426. (*Corresponding author: Gangyao Kuang.*)

The authors are with the College of Electronic Science, National University of Defense Technology, Changsha 410073, China (e-mail: kuangyeats@hotmail.com).

Digital Object Identifier 10.1109/TGRS.2023.3294884

as damage assessment [2], [3], urban development [4], and environment monitoring [5].

Currently, most research in CD is focused on homogeneous CD (or named monomodal CD); that is, it is relying on homogeneous data acquired from the same sensor under similar sensing conditions and parameters, such as the homogeneous CD of optical images [6], synthetic aperture radar (SAR) images [7], and hyperspectral images [8]. With the rapid development of remote sensing-related techniques (e.g., new sensor systems and new data processing methods) and a more open access mechanism to remote sensing data, current Earth observation sensors (including satellite-based, airborne-based, and unmanned aerial vehicle-based) can provide a vast amount of data from different types of sensors with different imaging mechanisms, e.g., multispectral images, SAR images, and hyperspectral images. This places new demands on CD technology, that is, the development of multimodal CD (or named heterogeneous CD) [9], [10], [11].

Multimodal CD (MCD) identifies the changes based on multimodal data, i.e., comparing images acquired by different sensors, which mainly contains two categories [12]: multi-source images acquired by different types of sensors, such as a pair of SAR and multispectral images (e.g., images obtained from Sentinel-1 and Sentinel-2, respectively), and multisensor (or named cross sensor) images acquired by different sensors with the same type, such as two multispectral images obtained from Sentinel-2 and Landsat-5, or two SAR images obtained from TerraSAR-X and COSMO-SkyMed, respectively. MCD can be regarded as a promotion of conventional homogeneous CD, which can relax the restriction of the homogeneous data that has some limitations on many real-world applications, especially in some emergency situations, such as earthquakes and floods [13], [14], [15].

B. Related Work

Despite its great practical value, MCD encounters a greater challenge, because it must be capable of comparing remote sensing images of arbitrary different modalities to identify changes [16]. Let X and Y be the compared multimodal images acquired at time t_1 (pre-event) and t_2 (post-event), respectively, and let x and y be the data samples drawn from X and Y at the same position, respectively, which can be pixels, square patches, or superpixels according to the basic analysis unit. Since different sensors have different imaging characteristics/parameters, the acquired multimodal images

exhibit quite different statistical properties (such as SAR and optical images), resulting in the inability to detect changes by directly comparing x and y as in traditional homogeneous CD. Therefore, MCD aims to make the “incomparable” multimodal images “comparable,” which is related to the topic of image transformation [17], [18].

The transformation-based MCD methods first transform the multimodal images to the common domain and then compare them in the transformed domain to obtain the difference image (DI), which can be generalized with the paradigms formulated as $\text{DI} = \mathcal{M}_1(X) \ominus \mathcal{M}_2(Y)$ [19], with \mathcal{M}_1 and \mathcal{M}_2 denoting the transform functions and \ominus denoting the difference operator.

Based on the transformed domain, the MCD methods can be divided into classification comparison-based, feature transformation-based, and image regression-based. The classification comparison-based methods first transform the images into a common category space by using classifiers of \mathcal{M}_1 and \mathcal{M}_2 trained by samples, and then compare the classification results [20], [21]. The feature transformation-based methods first transform the images into a common constructed feature space as $\mathcal{M}_1 : X \rightarrow Z$ or $\mathcal{M}_2 : Y \rightarrow Z'$ by using traditional method [22], [23], [24], or latent learned feature space by using deep neural networks [25], [26], [27], [28], and then compare the features (i.e., Z and Z') to detect changes. The image regression-based methods first transform one image to the domain of the other image as $\mathcal{M}_1 : X \rightarrow Y'$ or $\mathcal{M}_2 : Y \rightarrow X'$, and then compare the translated images Y' and X' and the target images Y and X , such as the traditional image translation methods [17], [29], [30] and deep translation methods [31], [32], [33], [34].

C. Motivations

There are three important unique issues of transformation-based MCD compared with homogeneous CD.

First and foremost, how to obtain the transform functions of \mathcal{M}_1 and \mathcal{M}_2 in the MCD, while the images can be directly compared without transformation in the homogeneous CD. Training \mathcal{M}_1 and \mathcal{M}_2 relies on the large training sets under the supervised mode, such as homogeneous pixel transformation (HPT) method [29], graph convolutional network (GCN)-based method [35], semisupervised Siamese network (S3N) [36], deep translation with generative adversarial network (GAN)-based method [31], domain adaptation-based neural network [37], [38], [39], and multiscale UNet-based method [27]. However, labeling samples is labor-intensive in practice, particularly for MCD problems, which also require great expert knowledge. Alternatively, for methods under the unsupervised mode, they tend to rely on the change prior, such as the affinity matrix difference (AMD)-guided traditional image regression [17], convolutional neural networks-based X-Net and adversarial cyclic encoder network (ACE-Net) [18], and code-aligned autoencoders (CAEs) [33], [40]. However, how to obtain an accurate change prior is a challenge itself.

Recently, some unsupervised MCD methods using graph models have been proposed, which exploit the structure consistency of the multitemporal images in the unchanged regions

to construct the transform functions of \mathcal{M}_1 and \mathcal{M}_2 . These methods can be broadly classified into two categories: first, the structure transformation and comparison-based methods, which construct a graph for each image to capture the structure information and then compare the structures by graph fusion [41], [42] or graph mapping [12], [24], [43], [44]; second, the structural regression-based methods, which require the structures of the original image X (or Y) and translated image Y' (or X') that are similar [30], [45], [46]. The proposed method in this article also fits into this subcategory.

However, for these image structure-based transformation methods, they suffer a great drawback: ignoring the structural asymmetry, that is, the structural complexity of the pre- and post-event images is not the same, especially when a new category of objects appears or when a category of objects disappears completely between the two images. In these cases, they can only complete a mapping comparison or image regression in one direction and will fail in the other direction, i.e., cannot map or convert from complex structural images (with more categories) to simple structural images (with fewer categories). As illustrated by Fig. 3, SCASC [30] can perform the forward transformation that translates the pre-event image to the domain of post-event image, but fails the backward transformation. In this article, we first analyze the structural asymmetry of compared images and show the reasons why the previous structure-based methods fail (such as IST-CRF [12], SCASC [30], FPMS [45], and AGSCC [46]), and then introduce a structural regression fusion (SRF)-based method to address this drawback.

Second, how to suppress the influence of unknown changed samples in the transformation, including training the transformation functions of \mathcal{M}_1 and \mathcal{M}_2 and completing the transformation processes of $X \xrightarrow{\mathcal{M}_1} Y'$ and $Y \xrightarrow{\mathcal{M}_2} X'$ when they are not the individual-pixel-based mappings. In particular, this issue is rarely mentioned by other studies, probably for two reasons: first, it is unique to MCD but not homogeneous CD; second, previous works usually adopt a step-by-step strategy, i.e., transformations followed by comparisons and then segmentation, such as HPT [29], X-Net [18], CAE [33], FPMS [45], INLPG [24], and GBF [41], which hides a risk that the transformation will be affected by the changed samples. Even though a few studies have focused on this challenge, they still need to construct a complex, coarse-to-refine iterative framework to reduce the changes influence [43], [47]. For example, in IRG-McS, Sun et al. [43] show that the changed samples will make the transformation unstable and the DI confusing, and then, they propagate the detection results of previous segmentation process back to the structure transformation process of next round, to gradually alleviate the influence of changes and correct the final CD results.

Different from the previous complex, redundant iterative framework, we decompose the original image into a translated image and changed image in the structural regression model similar to [30], which can bring two benefits: first, it can reduce the negative impact of changed samples on image transformation and directly output the DI by combining the transformations and comparisons in one model [46]; second,

it fuses the changes in the regression model with some constraints to make the output changes more accurate, thus mutually improving both regression performance and detection performance.

Third, how to fuse the detection results calculated from different transformed domains to improve MCD performance, especially for the image regression-based methods. Intuitively, for the regression-based MCD, we can compute two changed images to measure the change level after obtaining the mapping functions of $\mathcal{M}_1 : X \rightarrow Y'$ or $\mathcal{M}_2 : Y \rightarrow X'$. However, since the changed images are calculated in different domains (i.e., $X - X'$ and $Y - Y'$), we cannot directly add them as the final result. For the feature transformation-based MCD, when there are two transformations, such as the forward and backward transformations in NLPG [22] and INLPG [24], and the local and nonlocal structural transformations in SRGRL [44], two corresponding DIs will also be generated at this time. More importantly, sometimes fusing the changed images is particularly important when the mapping function is not accurate or robust enough due to the complex scenarios, training sample limitations, and noise influence.

In the previous studies, some MCD methods only perform one transformation and neglect the fusion. For example, the SCASC [30] and FPMS [45] only translate the pre-event image to the domain of post-event image, the CDRL [48] only translates the post-event image to the domain of pre-event image, the CGAN [34] only translates the optical image to the domain of SAR image, and the USCD [32] and DHFF [49] based on image style transfer only translate the SAR image to the domain of optical image. On the other hand, some methods fuse the change results after the transformation process similar to the late fusion; that is, the fusion and transformation are two separate processes, for example, the direct linear fusion in NLPG [22], HPT [29], AMD-IR [17], X-Net [18], CAE [33], and SRGRL [44]; the discrete wavelet transform-based fusion in INLPG [24]; and the Markov random field (MRF)-based fusion in the segmentation process of IRG-MCs [43] and IST-CRF [12]. However, none of the fusion results of these methods can be directly used to improve the transformation performance (since they perform transformation first and then fusion), which, in turn, limits the performance of image transformation, change fusion, and change extraction.

In this article, we propose an SRF model, which fuses the changes during the image transformation process. Specifically, based on the inherent physical constraints between changed images calculated from different domains (they are both characterizing the same change event), we combine the forward and backward transformations in a single optimization model and add constraints between the transformations to fuse the changes and improve the performance of forward and backward transformations. This can bring two benefits: first, it combines the two processes of fusion and transformation, which avoids completing forward transformation, backward transformation, and change fusion separately. Alternatively, we only need to solve one regression fusion model. Second, more importantly, by fusing the changes in the transformation, we can further reduce the impact of change on the

TABLE I
LIST OF IMPORTANT NOTATIONS

Symbol	Description
$\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}$	pre-event and post-even images
$\tilde{\mathbf{X}}', \tilde{\mathbf{Y}}'$	regression images of $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{X}}$
$\tilde{\mathbf{X}}_i$	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} and \mathbf{X}
\mathbf{X}_i	i -th feature vector of $\tilde{\mathbf{X}}$
Δ^x, Δ^y	changed feature matrices
$\mathcal{G}^{t1} = \{V^{t1}, E^{t1}, w\}$	hypergraph with vertex set V^{t1} , hyperedge E^{t1} and hyperedge weight w
$\mathbf{L}^{t1}, \mathbf{L}^{t2}, \mathbf{L}^f$	graph Laplacian matrices
\mathbf{I}_N	an $N \times N$ identity matrix

transformation and can overcome the one-way transformation failure caused by structure asymmetry. That is, by putting the two mutually reinforcing processes of fusion and transformation in one model, we can simultaneously make the transformation better and the change extraction more accurate.

D. Contributions

The main contributions of this article are summarized as follows.

- 1) We analyze the structural asymmetry of compared images in the MCD problem and explain why the previous structure-based methods perform differently in the forward and backward transformations, especially for the structural regression-based MCD methods.
- 2) We propose an SRF-based method for unsupervised MCD that incorporates the fusion into the combined regression model of forward and backward transformations, which can overcome the drawback of the structural regression caused by the structural asymmetry.
- 3) We use the HG instead of the normal graph to capture the high-order information of the image, which can improve the structure representation.
- 4) Extensive experiments demonstrate the effectiveness of the proposed method by comparing with some state-of-the-art (SOTA) methods.

E. Outline and Notation

The rest of this article is structured as follows. Section II describes the related structure-based method and analyzes the structure asymmetry of the images. Section III describes the proposed SRF in detail. Section IV gives the experimental results. Finally, we conclude this article in Section V. For convenience, Table I lists some important notations used in the rest of this article.

II. STRUCTURE-BASED MCD AND STRUCTURE ASYMMETRY

Given a pair of co-registered images of different modalities collected at times t_1 (pre-event) and t_2 (post-event), denoted as $\tilde{\mathbf{X}} \in \mathbb{R}^{M \times N \times b_x}$ in domain \mathcal{X} and $\tilde{\mathbf{Y}} \in \mathbb{R}^{M \times N \times b_y}$ in domain \mathcal{Y} , respectively, with pixels defined as $\tilde{x}(m, n, b)$ and $\tilde{y}(m, n, b)$,

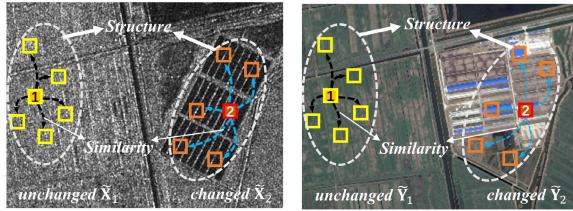


Fig. 1. Illustration of structure consistency in multimodal images. (Left) SAR image. (Right) Optical image. The similarity between image parts is reflected by the thickness of connecting lines. The structures of the unchanged parts of $\tilde{\mathbf{X}}_1$ and $\tilde{\mathbf{Y}}_1$ are the same, while the structures of the changed parts of $\tilde{\mathbf{X}}_2$ and $\tilde{\mathbf{Y}}_2$ are different.

respectively, the purpose of MCD is to detect the changed areas represented by the change map (CM).

Since the proposed method is to use the structure consistency to complete the image regression fusion, we first briefly review the transformation-based MCD methods using structure consistency, which contain the structure comparison-based and structure regression-based, and then analyze the structure asymmetry in MCD problem.

A. Structure Consistency

As mentioned in Section I, because the multimodal images show quite different characteristics for the same object, we cannot directly compare the pixel values of \tilde{x} and \tilde{y} to detect the changes. Therefore, we need to find the connection between multimodal images and make them comparable.

The structure consistency [24] is based on the inherent self-similarity property of images, where the “structure” is characterized by the similarity relationships within the image. As illustrated in Fig. 1, each image is divided into N_S small parts (such as square patches [22] or superpixels [43]) with the same segmentation form. Then, for one of the images (e.g., pre-event $\tilde{\mathbf{X}}$), if the parts of $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{X}}_j$ represent the same kind of objects, showing that they are very similar, and neither of them has changed, then the corresponding pair of $\tilde{\mathbf{Y}}_i$ and $\tilde{\mathbf{Y}}_j$ in the other image is also very similar, as they also represent the same kind of objects. On the contrary, if $\tilde{\mathbf{X}}_i$ has changed during the event, this similarity relationship is no longer conformed by $\tilde{\mathbf{Y}}_i$. Since this nonlocal similarity relationship within the image itself can eliminate the discrepancy across different imaging modalities, the structure can be well preserved between multimodal images and can be used to establish the connection between multimodal images to make them comparable, using structure comparison or structure regression.

B. MCD Methods Using Structure Consistency

1) *Structure Comparison*: For the structure comparison-based methods, such as INLPG [24], IRG-McS [43], and IST-CRF [12], they first construct K -nearest neighbor (KNN) graphs of G^{t1} and G^{t2} for images of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ to capture the structure information, respectively, and then project one graph to the domain of the other image to compute the DIs

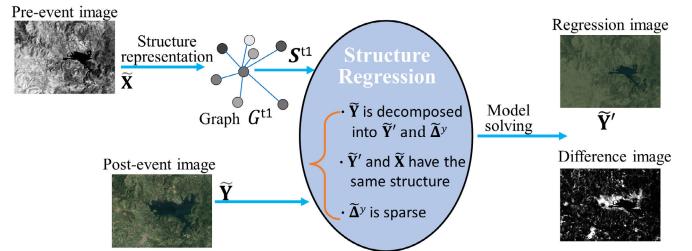


Fig. 2. Illustration of structure regression-based SCASC [30].

similar to

$$\begin{aligned} f_i^{\text{fw}} &= \frac{1}{K} \left| \sum_{j=1}^{N_S} (A_{i,j}^{t1} - A_{i,j}^{t2}) |D_{i,j}^{t2}| \right| \\ f_i^{\text{bw}} &= \frac{1}{K} \left| \sum_{j=1}^{N_S} (A_{i,j}^{t1} - A_{i,j}^{t2}) |D_{i,j}^{t1}| \right| \end{aligned} \quad (1)$$

where f_i^{fw} and f_i^{bw} represent the forward and backward change level of the i th vertex (i.e., superpixel or square patch), respectively; \mathbf{A}^{t1} and \mathbf{A}^{t2} represent the adjacent matrices of G^{t1} and G^{t2} , respectively; and $D_{i,j}^{t1}$ and $D_{i,j}^{t2}$ represent the feature difference vectors (e.g., IST-CRF [12]) or feature distance (e.g., INLPG [24] and IRG-McS [43]) between $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{X}}_j$, and $\tilde{\mathbf{Y}}_i$ and $\tilde{\mathbf{Y}}_j$, respectively. Intuitively, the forward f_i^{fw} and backward f_i^{bw} calculate the structure difference by measuring how different the adjacent matrices of \mathbf{A}^{t1} and \mathbf{A}^{t2} are in the image differential domain.

2) *Structure Regression*: For the structure regression-based methods, such as SCASC [30] and AGSCC [46] in Fig. 2, they first construct a KNN graph G^{t1} for the pre-event image $\tilde{\mathbf{X}}$, and then translate $\tilde{\mathbf{X}}$ to the domain \mathcal{Y} of post-event image by a structural regression-based image decomposition model as follows:

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

where \mathbf{Y}' represents the feature of regression image $\tilde{\mathbf{Y}}'$, Δ^y represents changed feature matrix, S^{t1} represents the weighing matrix of G^{t1} , and $\lambda > 0$ is a regularization parameter. The first regularization term of $\sum_{i,j=1}^{N_S} \| \mathbf{Y}'_i - \mathbf{Y}'_j \|_2^2 S_{i,j}^{t1}$ is used to constrain the regression image \mathbf{Y}' and the original image $\tilde{\mathbf{X}}$ to have the same structure, and the second regularization term of $\| \Delta^y \|_{2,1}$ is based on the sparse change prior in the practical CD problem.

In the previous structure regression-based methods of [30], [46], and [45], they all only consider the forward regression that translates the pre-event image to the domain of post-event image. Similar to (2), we can complete the backward regression by translating the post-event image $\tilde{\mathbf{Y}}$ to the domain \mathcal{X} of pre-event image with the model as follows:

$$\begin{aligned} \min_{\mathbf{X}', \Delta^x} \sum_{i,j=1}^{N_S} & \| \mathbf{X}'_i - \mathbf{X}'_j \|_2^2 S_{i,j}^{t2} + \lambda \| \Delta^x \|_{2,1} \\ \text{s.t. } \mathbf{X} &= \mathbf{X}' - \Delta^x \end{aligned} \quad (3)$$

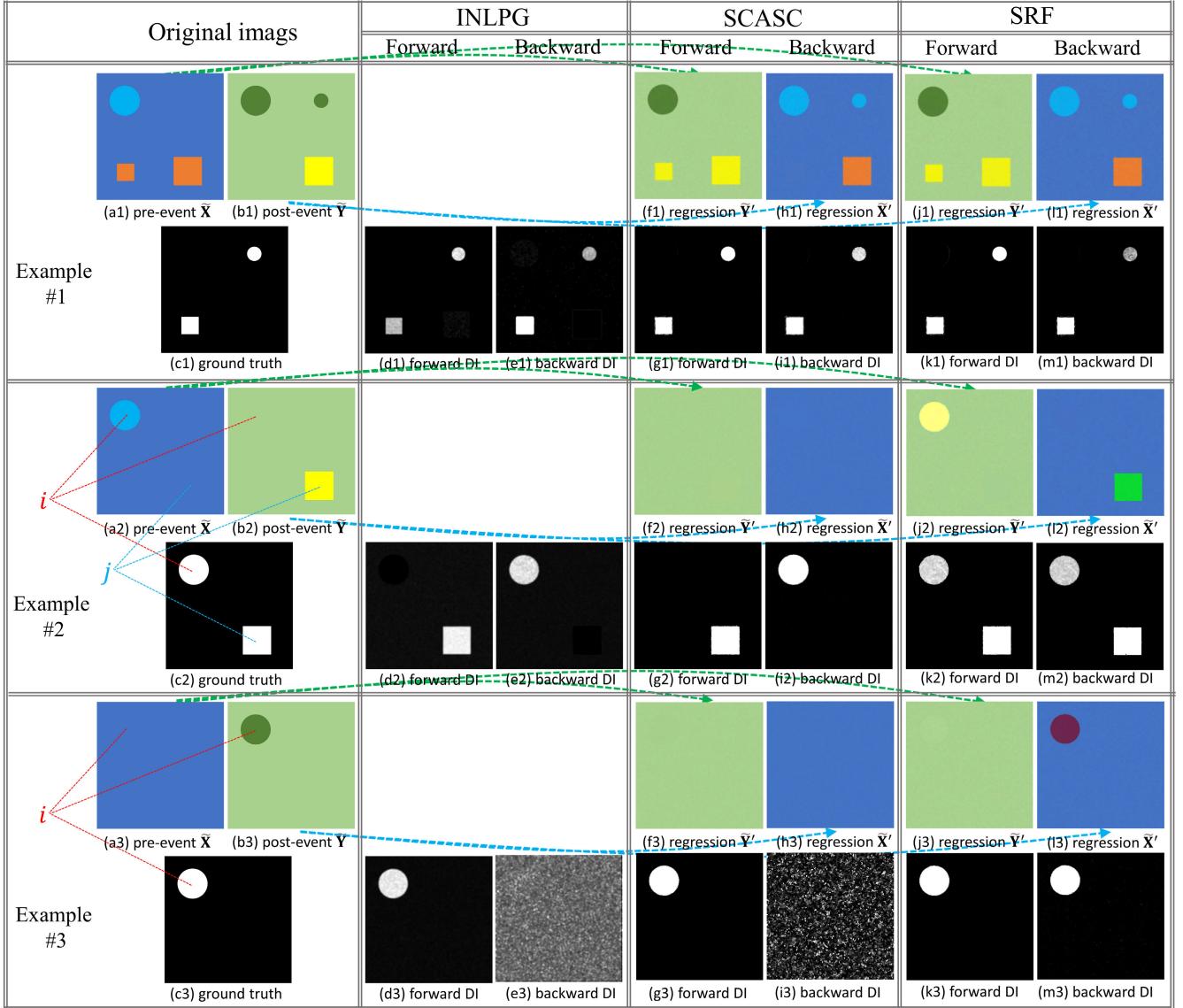


Fig. 3. Three examples of INLPG [24], SCASC [30], and the proposed SRF. From top to bottom, they correspond to results on the first, second, and third examples, respectively. From left to right are (a) pre-event image \bar{X} , (b) post-event image \bar{Y} , (c) ground truth, (d) forward DI of INLPG, (e) backward DI of INLPG, (f) forward regression \bar{Y}' of SCASC, (g) forward DI of SCASC, (h) backward regression \bar{X}' of SCASC, (i) backward DI of SCASC, (j) forward regression \bar{Y}' of SRF, (k) forward DI of SRF, (l) backward regression \bar{X}' of SRF, and (m) backward DI of SRF. In Example 1, all three methods of INLPG, SCASC, and SRF can detect changes, as there is no structural asymmetry in either pre-event image or post-event image. In Example 2, the forward and backward comparisons of INLPG or regressions of SCASC can only detect partial changes because of the structural asymmetry between the pre-event and post-event images. For the marked i in (a2)–(c2), there is $C_i^x \subseteq C_i^y$, so the change of i cannot be detected in the forward DIs of INLPG and SCASC; for the marked j in (a2)–(c2), there is $C_j^y \subseteq C_j^x$, so the change of j cannot be detected in the backward DIs of INLPG and SCASC. In Example 3, because $C_i^y \subseteq C_i^x$ holds for any i , the backward comparison of INLPG or regression of SCASC cannot detect changes at all. Meanwhile, in Examples 1–3, INLPG and SCASC are able to detect changes for regions where there is no structural asymmetry. However, the proposed SRF can detect all changed regions in both forward regression and backward regression, regardless of the presence of structural asymmetry.

where \mathbf{X}' represents the feature of regression image $\tilde{\mathbf{X}}'$, Δ^x represents the changed feature matrix, and \mathbf{S}^{t2} represents the weighing matrix of KNN graph G^{t2} constructed for image $\tilde{\mathbf{Y}}$.

3) *Simple MCD Example:* In Fig. 3, we show three different MCD results of structure comparison-based method of INLPG [24] and structure regression-based method of SCASC [30]. In the first experiment, we can find that both INLPG and SCASC can detect the changes well, both in forward and backward comparisons or regressions. However, in the second experiment, the forward and backward comparisons of INLPG or regressions of SCASC can only detect

partial changes, and in the third experiment, the backward comparison of INLPG or regression of SCASC cannot detect changes at all. The three examples in Fig. 3 also demonstrate the importance of fusion (such as the transformation fusion, regression fusion, and the DI fusion). Next, we will analyze the reasons why these structure-based MCD approaches fail in certain scenarios.

C. Structural Asymmetry

Taking a deeper analysis of the second and third experiments in Fig. 3 again, we can find that there is a common factor

where the structure comparison or regression fails: for the undetected changed region, there is a structural asymmetry between the compared multitemporal images.

Definition of Congeneric Sets: For $\tilde{\mathbf{X}}_i, \tilde{\mathbf{Y}}_i, i \in \mathcal{I}$ with $\mathcal{I} = \{1, \dots, N_S\}$, we define the congeneric sets for $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{Y}}_i$ as the index sets of the following:

$$\begin{aligned}\mathcal{C}_i^x &= \{j | j \in \mathcal{I}, \tilde{\mathbf{X}}_i \text{ and } \tilde{\mathbf{X}}_j \text{ represent the same type of object}\} \\ \mathcal{C}_i^y &= \{j | j \in \mathcal{I}, \tilde{\mathbf{Y}}_i \text{ and } \tilde{\mathbf{Y}}_j \text{ represent the same type of object}\}.\end{aligned}\quad (4)$$

Easily, we have $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{X}}_j, j \in \mathcal{C}_i^x$, that are very similar; and $\tilde{\mathbf{Y}}_i$ and $\tilde{\mathbf{Y}}_j, j \in \mathcal{C}_i^y$, that are also very similar.

Definition of Structural Asymmetry: For the i th part of images, if $\mathcal{C}_i^x \subseteq \mathcal{C}_i^y$, we say that the structure of $\tilde{\mathbf{X}}_i$ is contained in the structure of $\tilde{\mathbf{Y}}_i$; on the contrary, if $\mathcal{C}_i^y \subseteq \mathcal{C}_i^x$, we say that the structure of $\tilde{\mathbf{Y}}_i$ is contained in the structure of $\tilde{\mathbf{X}}_i$. We define that there is a structural asymmetry between two images if this structure containment phenomenon exists in them, i.e., there exists $i \in \mathcal{I}$, such that $\mathcal{C}_i^x \subseteq \mathcal{C}_i^y$ or $\mathcal{C}_i^y \subseteq \mathcal{C}_i^x$.

For the i th part of image, as \mathbf{A}^{t1} is the adjacent matrix of the KNN graph G^{t1} , then we have $\{h | h \in \mathcal{I}, A_{i,h}^{t1} \neq 0\} \subseteq \mathcal{C}_i^x$. When $\mathcal{C}_i^x \subseteq \mathcal{C}_i^y$, then we have $\tilde{\mathbf{Y}}_i$ and $\tilde{\mathbf{Y}}_j, j \in \{h | h \in \mathcal{I}, A_{i,h}^{t1} \neq 0\}$ or $j \in \{h | h \in \mathcal{I}, S_{i,h}^{t1} \neq 0\}$, which also represent the same type of object.

Then, we can find that if $\mathcal{C}_i^x \subseteq \mathcal{C}_i^y$ and the i th part of image is changed, then the forward structure comparison and regression will fail.

First, for the forward comparison of (1), if $\mathcal{C}_i^x \subseteq \mathcal{C}_i^y$, then $A_{i,j}^{t1}|D_{i,j}^{t2}|$ is very small, because $\tilde{\mathbf{Y}}_i$ and $\tilde{\mathbf{Y}}_j, j \in \{h | h \in \mathcal{I}, A_{i,h}^{t1} \neq 0\}$, represent the same type of object; then, the change level of the changed i th vertex calculated by the structure difference (1) is very small (because $A_{i,j}^{t2}|D_{i,j}^{t2}|$ is very small due to the nature of KNN graph). Therefore, the forward comparison cannot detect the changes when the structure of this changed $\tilde{\mathbf{X}}_i$ is contained in the structure of $\tilde{\mathbf{Y}}_i$.

Second, for the forward regression of (2), if $\mathcal{C}_i^x \subseteq \mathcal{C}_i^y$, then $\sum_{j=1}^{N_S} \|\mathbf{Y}_i - \mathbf{Y}_j\|_2^2 S_{i,j}^{t1}$ is very small, because $\tilde{\mathbf{Y}}_i$ and $\tilde{\mathbf{Y}}_j, j \in \{h | h \in \mathcal{I}, S_{i,h}^{t1} \neq 0\}$, represent the same type of object. With the change penalty of $\|\Delta^y\|_{2,1}$, we have $\mathbf{Y}'_i = \mathbf{Y}_i$ and $\Delta_i^y = 0$ for the changed i . Therefore, the forward regression fails when the structure of this changed $\tilde{\mathbf{X}}_i$ is contained in the structure of $\tilde{\mathbf{Y}}_i$.

Similarly, we can also find that if the structure of changed $\tilde{\mathbf{Y}}_i$ is contained in the structure of $\tilde{\mathbf{X}}_i$, then the backward comparison and regression will also fail. At the same time, as shown in Fig. 3, when the structure $\tilde{\mathbf{Y}}_i$ is not contained in the structure of $\tilde{\mathbf{X}}_i$, the backward comparison and regression can detect the changes well.

D. Fusion of the Regressions

From the above analysis and experiments shown in Fig. 3, we can find that there is an urgent need to fuse the forward and backward transformations to eliminate the influence of structural asymmetric in the structure consistency-based methods. In the previous methods, the forward and backward DIs are fused after the transformation process, such as NLPG [22], HPT [29], AMD-IR [17], X-Net [18], CAE [33], INLPG [24],

IRG-McS [43], and IST-CRF [12], which is a late fusion process and may carry the risk of worse outcomes after the fusion. In addition, based on the structural asymmetry, we can also provide an alternative strategy for this late fusion of DIs. For example, we can first find out these regions where structural asymmetry exist, and then assign different weights to the forward and backward DIs in the fusion according to the asymmetry. However, regarding the late DIs fusion in these methods is not the focus of this study. In this article, we propose a change fusion-based regression method that fuses the forward and backward regressions in a single regression model, which means that the fusion and regression are combined.

III. STRUCTURAL REGRESSION FUSION

In this section, we describe the proposed SRF in detail, which contains two main parts: how to represent the image structure and how to complete the fused structural regression to obtain the DI. Fig. 4 shows the framework of the proposed SRF.

A. Preprocessing

As aforementioned, we use the pairwise similarity relationships to represent the structure of image. Here, we choose the superpixel that internally belongs to the same object as the analysis-based unit, which has two advantages compared with the individual pixel or square patch: first, the superpixel can capture the contextual information and preserve the edge of object; second, the computational complexity can be greatly reduced by reducing the graph size.

The Gaussian mixture model-based superpixel segmentation method (GMMSp) [50] is selected to generate the superpixels, which can efficiently produce superpixels that adhere to object boundaries in linear complexity with respect to the number of pixels. In order to obtain the co-segmentation map, we first construct a false RGB image, where the channels are the gray pre-event image, gray post-event image, and a zero channel. Then, the GMMSp is employed to segment the false RGB image into N_S regions with the co-segmentation map Λ . Then, we can obtain the superpixels of $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{Y}}_i$, denoted as follows:

$$\begin{aligned}\tilde{\mathbf{X}}_i &= \{\tilde{x}(m, n, b) | (m, n) \in \Lambda_i, b = 1, \dots, b_x\} \\ \tilde{\mathbf{Y}}_i &= \{\tilde{y}(m, n, b) | (m, n) \in \Lambda_i, b = 1, \dots, b_y\}.\end{aligned}\quad (5)$$

Then, $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{Y}}_i$ represent the same region, and they are both internally homogeneous in the multitemporal images of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$, respectively.

Once the superpixels are obtained, different features can be extracted to represent different information of the superpixels, such as spectral (intensity), spatial, and textual information. In this article, we simply extract the mean and median values of each band as the feature of superpixel (other discriminative features can also be added). Then, we can obtain the feature matrices of the multitemporal images, denoted as $\mathbf{X} \in \mathbb{R}^{2b_x \times N_S}$ and $\mathbf{Y} \in \mathbb{R}^{2b_y \times N_S}$, where each column (\mathbf{X}_i and \mathbf{Y}_i) represents the feature vector of the superpixels ($\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{Y}}_i$).

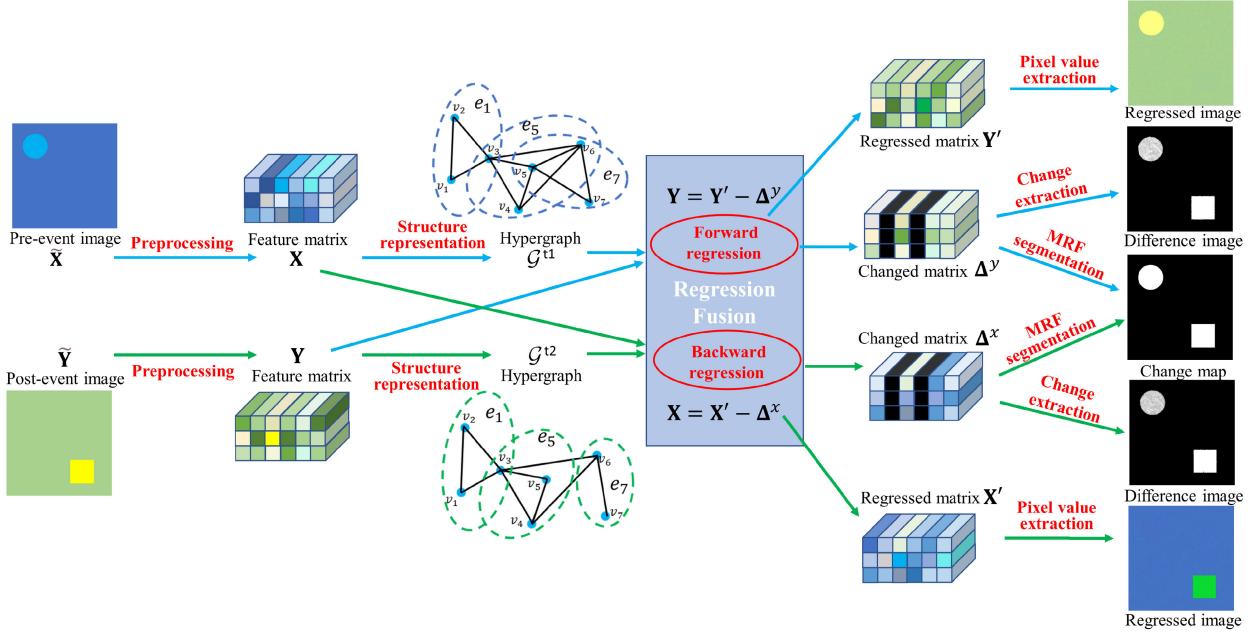


Fig. 4. Framework of the proposed SRF-based MCD method.

B. Structure Representation

In order to capture the structure information of images, the KNN-type graphs are constructed. For example, in SCASC [30], an adaptive KNN probabilistic graph G^{t1} is constructed for pre-event image with the weighting matrix \mathbf{S}^{t1} , which connects each superpixel $\tilde{\mathbf{X}}_i$ to its k_i NNs with weights $S_{i,j}^{t1}$, $j = 1, \dots, N_S$. The \mathbf{S}^{t1} is calculated by using the following minimization model:

$$\min_{\mathbf{S}^{t1}} \sum_{j=1}^{N_S} D_{i,j}^{t1} S_{i,j}^{t1} + \alpha_i (S_{i,j}^{t1})^2 \quad \text{s.t. } 0 \leq S_{i,j}^{t1} \leq 1, \quad \sum_{j=1}^{N_S} S_{i,j}^{t1} = 1 \quad (6)$$

where $D_{i,j}^{t1} = \|\mathbf{X}_i - \mathbf{X}_j\|_2^2$ denotes the distance between two superpixels of $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{X}}_j$, and $\alpha_i > 0$ is a balance parameter that controls the NNs number k_i of $\tilde{\mathbf{X}}_i$. According to the SCASC, the closed-form solution of \mathbf{S}^{t1} can be calculated by

$$S_{i,(j)}^{t1} = \begin{cases} \frac{D_{i,(k_i+1)}^{t1} - D_{i,(j)}^{t1}}{k_i D_{i,(k_i+1)}^{t1} - \sum_{h=1}^{k_i} D_{i,(h)}^{t1}}, & j \leq k_i \\ 0, & j > k_i \end{cases} \quad (7)$$

where we sort \mathbf{D}_i^{t1} in ascending order as $D_{i,(1)}^{t1}, D_{i,(2)}^{t1}, \dots, D_{i,(N_S)}^{t1}$, and then, (j) of $D_{i,(j)}^{t1}$ denotes the position of the j th smallest value in \mathbf{D}_i^{t1} . The k_i is adaptively determined by an in-degree-based strategy proposed in SCASC [30].

In this article, we construct HGs of \mathcal{G}^{t1} and \mathcal{G}^{t2} for images of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ on the basis of the adaptive KNN probabilistic graphs of G^{t1} of (7) and G^{t2} (it is constructed similar to G^{t1}), respectively. In contrast to the pairwise graph, an HG can link more than two vertices, which can preserve the high-order neighborhood relationships between superpixels and then capture more comprehensively structure information [51], as shown in Fig. 4.

Denote V^{t1} and E^{t1} as the vertex set and hyperedge set of the HG $\mathcal{G}^{t1} = \{V^{t1}, E^{t1}, w\}$, respectively; we construct the \mathcal{G}^{t1} by setting each superpixel as a vertex $v \in V^{t1}$ and link each superpixel with its neighbors as a hyperedge $e_i^{t1} \in E^{t1}$ with the associate hyperedge weight $w(e_i^{t1})$, which is a subset of the vertex set as $e_i^{t1} \subseteq V^{t1}$. Specifically, with the \mathbf{S}^{t1} of (7), the HG \mathcal{G}^{t1} is constructed as follows:

$$V^{t1} = \mathcal{I}; \quad e_i^{t1} = \{j | S_{j,i}^{t1} \neq 0; j \in \mathcal{I}\}. \quad (8)$$

We use a weighted incident matrix \mathbf{H}^{t1} as $\mathbf{H}^{t1} = \mathbf{S}^{t1}$, which gives different weights for superpixels $\tilde{\mathbf{X}}_j$ connected to the center superpixel $\tilde{\mathbf{X}}_i$. We set the hyperedge weight $w(e^{t1})$ to be the mean of heat kernel similarity of the intraclass superpixels as follows:

$$w(e_i^{t1}) = \frac{1}{k_i(k_i-1)} \sum_{j,l \in e_i^{t1}} \exp(-\|\mathbf{X}_j - \mathbf{X}_l\|_2^2). \quad (9)$$

With this hyperedge weight $w(e^{t1})$, the constructed HG with a larger average intraclass similarity is assigned a higher weight. Based on \mathbf{H}^{t1} and $w(e^{t1})$, the vertex degree for each vertex is $d(v) = \sum_{i=1}^{N_S} w(e_i^{t1}) H_{v,i}^{t1}$, and the hyperedge degree for each hyperedge is $\psi(e_i^{t1}) = \sum_{v=1}^{N_S} H_{v,i}^{t1}$.

Let \mathbf{d}^{t1} , $\boldsymbol{\psi}^{t1}$, and \mathbf{W}^{t1} denote the diagonal matrices of the vertex degree, hyperedge degree, and hyperedge weight, respectively. The unnormalized HG Laplacian matrix is defined as follows [52], [53]:

$$\mathbf{L}^{t1} = \mathbf{d}^{t1} - \mathbf{H}^{t1} \mathbf{W}^{t1} (\boldsymbol{\psi}^{t1})^{-1} (\mathbf{H}^{t1})^T. \quad (10)$$

Similar to \mathcal{G}^{t1} of pre-event image $\tilde{\mathbf{X}}$, we also construct $\mathcal{G}^{t2} = \{V^{t2}, E^{t2}, w\}$ for post-event image $\tilde{\mathbf{Y}}$ and obtain the \mathbf{d}^{t2} , $\boldsymbol{\psi}^{t2}$, and \mathbf{L}^{t2} .

C. Fused Regression

For the image regression-based MCD, the main goal is to transform one image to the domain of the other image and obtain the DI between the transformed image and the target image. In the structure consistency-based image regression method, we require the transformed image and the original image to have the same structure. Since we use the superpixel as the vertex (basic unit) and extract the features to represent the superpixel, then we need to find the regression functions between the two feature matrices. Define the transformation functions between the domains of feature matrices as \mathcal{M}_1 and \mathcal{M}_2 , define the feature extraction operator as \mathcal{F} , and define the pixel value extraction operator as \mathcal{F}^{-1} , e.g., extracting the mean feature from the feature matrix as the pixel value of each pixel inside the superpixel. Denote the \mathbf{X}' and \mathbf{Y}' as the transformed feature matrices, and define $\tilde{\mathbf{X}}'$ and $\tilde{\mathbf{Y}}'$ as the translated image. Then, we have the forward transformation as follows:

$$\tilde{\mathbf{Y}}' = \mathcal{F}^{-1}(\mathbf{Y}') = \mathcal{F}^{-1}\mathcal{M}_1(\mathbf{X}) = \mathcal{F}^{-1}\mathcal{M}_1\mathcal{F}(\tilde{\mathbf{X}}) \quad (11)$$

and have the backward transformation as follows:

$$\tilde{\mathbf{X}}' = \mathcal{F}^{-1}(\mathbf{X}') = \mathcal{F}^{-1}\mathcal{M}_2(\mathbf{Y}) = \mathcal{F}^{-1}\mathcal{M}_2\mathcal{F}(\tilde{\mathbf{Y}}). \quad (12)$$

Next, we describe these two regression functions of \mathcal{M}_1 and \mathcal{M}_2 and show how to fuse the two regression processes into one model.

1) *Forward Transformation*: In (11), it needs to find the connection between $\tilde{\mathbf{Y}}'$ and $\tilde{\mathbf{X}}$. Based on the structure consistency between multimodal images, we have that the similarity relationships between superpixels within the images are the same. That is, if two superpixels ($\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{X}}_j$) in the pre-event image represent the same kind of object (showing very similar), then the transformed superpixels $\tilde{\mathbf{Y}}'_i$ and $\tilde{\mathbf{Y}}'_j$ should represent the same kind of object (with very small difference/distance). Since the structure information of the pre-event image is characterized by the HG \mathcal{G}'^1 , then we require that the transformed superpixels $\tilde{\mathbf{Y}}'_i$ and $\tilde{\mathbf{Y}}'_j$ corresponding to the $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{X}}_j$ within the same hyperedge e_l^{t1} of \mathcal{G}'^1 , i.e., $i, j \in e_l^{t1}$, should be similar to each other. Therefore, we can obtain the HG Laplacian-based regularization (HGLR) as follows:

$$\sum_{l=1}^{N_S} \sum_{i,j \in e_l^{t1}} \frac{w(e_l^{t1}) H_{i,l}^{t1} H_{j,l}^{t1}}{\psi(e_l^{t1})} \|\mathbf{Y}'_i - \mathbf{Y}'_j\|_2^2 = 2\text{Tr}(\mathbf{Y}' \mathbf{L}'^{t1} \mathbf{Y}'^T). \quad (13)$$

The detailed similar derivation of (13) can be found in the Supplementary Material of [54]. The HGLR constrains that the similarity relationships between superpixels of $\tilde{\mathbf{X}}$ inside the same hyperedge of \mathcal{G}'^1 are preserved by the regression image $\tilde{\mathbf{Y}}'$.

For the post-event image $\tilde{\mathbf{Y}}$, we can decompose it into a regression image $\tilde{\mathbf{Y}}'$ and a changed image Δ^y ; then, we have $\mathbf{Y} = \mathbf{Y}' - \Delta^y$, where $\Delta^y \in \mathbb{R}^{2b_y \times N_S}$ represents the changed feature matrix. For the changed image, we have a prior sparsity-based regularization (PSR), which is based on the fact that only a small part of objects are changed in the practice [30], [55]. In this article, we choose the $\ell_{2,1}$ -norm-based

sparsity regularization defined as $\|\Delta^y\|_{2,1} = \sum_{i=1}^{N_S} \|\Delta_i^y\|_2$, which is a convex relaxation of the original $\ell_{2,0}$ -norm $\|\Delta^y\|_{2,0}$ that equals to the number of nonzero column of Δ^y (the number of changed superpixels).

By combining the HGLR and the PSR, we have the forward transformation model

$$\min_{\mathbf{Y}', \Delta^y} 2\text{Tr}(\mathbf{Y}' \mathbf{L}'^{t1} \mathbf{Y}'^T) + \lambda \|\Delta^y\|_{2,1} \text{ s.t. } \mathbf{Y} = \mathbf{Y}' - \Delta^y. \quad (14)$$

2) *Backward Transformation*: Similar to the forward transformation, we require that the similarity relationships within the post-event image $\tilde{\mathbf{Y}}$ characterized by the HG \mathcal{G}'^2 should be preserved by the regression image $\tilde{\mathbf{X}}'$; that is, the transformed superpixels $\tilde{\mathbf{X}}'_i$ and $\tilde{\mathbf{X}}'_j$ corresponding to the $\tilde{\mathbf{Y}}_i$ and $\tilde{\mathbf{Y}}_j$ within the same hyperedge e_l^{t2} of \mathcal{G}'^2 , i.e., $i, j \in e_l^{t2}$, should be similar to each other. Then, we have the backward transformation model

$$\min_{\mathbf{X}', \Delta^x} 2\text{Tr}(\mathbf{X}' \mathbf{L}'^{t2} \mathbf{X}'^T) + \lambda \|\Delta^x\|_{2,1} \text{ s.t. } \mathbf{X} = \mathbf{X}' - \Delta^x. \quad (15)$$

3) *Connection Between Changes*: From the models of (14) and (15), the backward and forward changed images $\tilde{\Delta}_i^x = \mathcal{F}^{-1}(\Delta^x)$ and $\tilde{\Delta}_i^y = \mathcal{F}^{-1}(\Delta^y)$ are from different domains, and we cannot directly fuse them to obtain the detection results; that is, directly fusing these changed images by taking the mean value may not improve detection accuracy. Instead of the late fusing of the changed images as in previous methods, we fuse the forward and backward transformations in the regression process to improve the regression performance and, thus, obtain a more accurate changed image. Then, we need to find the connections between the changed images.

First, the changed images are simultaneously smooth on the fused HG \mathcal{G}^f . In [41], a signal smoothness representation has been proposed to obtain a smoother DI, where the fused graph is constructed by simply taking the minimum of the two edge weights. In this article, we define the fused HG $\mathcal{G}^f = \{V^f, E^f, w\}$ as follows: $V^f = \mathcal{I}$; $e_i^f = e_i^{t1} \cap e_i^{t2}$; $H_{v,i}^f = 1$ if $v \in e_i^f$, and $H_{v,i}^f = 0$, otherwise. We set the fused hyperedge weight as follows:

$$w(e_i^f) = \frac{\sum_{j,l \in e_i^f} \exp(-\|\mathbf{X}_j - \mathbf{X}_l\|_2^2 - \|\mathbf{Y}_j - \mathbf{Y}_l\|_2^2)}{\left(\text{card}(e_i^f)\right)^2} \quad (16)$$

where $\text{card}(\cdot)$ denotes the cardinality of a set.

Because $e_i^f \subseteq e_i^{t1}$, then $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{X}}_j$, $j \in e_i^f$, represent the same kind of object (from the construction of \mathcal{G}'^1), and $\tilde{\mathbf{Y}}'_i$ and $\tilde{\mathbf{Y}}'_j$, $j \in e_i^f$, also represent the same kind of object (from the constraint of structure consistency of forward transformation). Similarly, because $e_i^f \subseteq e_i^{t2}$, then $\tilde{\mathbf{Y}}_i$ and $\tilde{\mathbf{Y}}_j$, $j \in e_i^f$, also represent the same kind of object (from the construction of \mathcal{G}'^2), and $\tilde{\mathbf{X}}'_i$ and $\tilde{\mathbf{X}}'_j$, $j \in e_i^f$, also represent the same kind of object (from the constraint of structure consistency of backward transformation). Then, we have that $\tilde{\Delta}_i^x$ and $\tilde{\Delta}_j^x$, and $\tilde{\Delta}_i^y$ and $\tilde{\Delta}_j^y$, $j \in e_i^f$, also represent the same kind of changes. A more intuitive mathematical description is as follows: because $\|\mathbf{X}_i - \mathbf{X}_j\|_2$ and $\|\mathbf{Y}_i - \mathbf{Y}_j\|_2$ are very small due to $e_i^f \subseteq e_i^{t1}$ and $e_i^f \subseteq e_i^{t2}$, respectively, and $\|\mathbf{X}'_i - \mathbf{X}'_j\|_2$ and $\|\mathbf{Y}'_i - \mathbf{Y}'_j\|_2$ are also very small due to the structure consistency-based HGLR, then we have $\|\Delta_i^x - \Delta_j^x\|_2 = \|(\mathbf{X}'_i - \mathbf{X}'_j) - (\tilde{\mathbf{X}}'_i - \tilde{\mathbf{X}}'_j)\|_2$

$(\mathbf{X}_i - \mathbf{X}_j)_2$ and $\|\Delta_i^x - \Delta_j^x\|_2 = \|(\mathbf{Y}'_i - \mathbf{Y}'_j) - (\mathbf{Y}_i - \mathbf{Y}_j)\|_2$ that are also very small. Then, we can obtain the fused HG-based smoothness regularization (FHSR) as follows:

$$\sum_{l=1}^{N_s} \sum_{i,j \in e_l^f} \frac{w(e_l^f) H_{i,l}^f H_{j,l}^f}{\psi(e_l^f)} \left(\|\Delta_i^x - \Delta_j^x\|_2^2 + \|\Delta_i^y - \Delta_j^y\|_2^2 \right) \\ = 2\text{Tr}(\Delta^x \mathbf{L}^f \Delta^{xT}) + 2\text{Tr}(\Delta^y \mathbf{L}^f \Delta^{yT}) \quad (17)$$

where \mathbf{L}^f is the unnormalized HG Laplacian matrix of the \mathcal{G}^f .

Second, the changed images are aligned on the support sets. Because the two changed images are describing the same change event, i.e., the changed areas in the Δ^x and Δ^y are the same, then we have the following constraint:

$$\{i | i \in \mathcal{I}, \|\Delta_i^x\|_2 \neq 0\} = \{i | i \in \mathcal{I}, \|\Delta_i^y\|_2 \neq 0\}. \quad (18)$$

However, directly using this constraint in the regression model is very difficult. Alternatively, we use the following change alignment-based regularization (CAR) of $\sum_{i=1}^{N_s} \phi(\|\Delta_i^x\|_2, \|\Delta_i^y\|_2)$ with the function ϕ being defined as follows:

$$\phi^1(\|\Delta_i^x\|_2, \|\Delta_i^y\|_2) = -\|\Delta_i^x\|_2 \|\Delta_i^y\|_2 \quad (19a)$$

$$\phi^2(\|\Delta_i^x\|_2, \|\Delta_i^y\|_2) = \exp(-\|\Delta_i^x\|_2 \|\Delta_i^y\|_2). \quad (19b)$$

In the CAR, we use $\|\Delta_i^x\|_2$ and $\|\Delta_i^y\|_2$ to represent the change probability of the i th superpixel in $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$, respectively. If the region represented by the i th superpixel ($\tilde{\mathbf{X}}_i$ or $\tilde{\mathbf{Y}}_i$) is changed during the event, then both the $\|\Delta_i^x\|_2$ and $\|\Delta_i^y\|_2$ should be large, and then, we have a small CAR. On the contrary, if this region is unchanged during the event, then both the $\|\Delta_i^x\|_2$ and $\|\Delta_i^y\|_2$ should be close to zero; then, we have a larger CAR. The minimization of CAR requires that whenever either of the change levels of $\|\Delta_i^x\|_2$ or $\|\Delta_i^y\|_2$ is large, the other one should also be large, which means that $\|\Delta_i^x\|_2$ and $\|\Delta_i^y\|_2$ are aligned.

4) *Fused Regression Model*: By combining the forward transformation (14), backward transformation (15), FHSR (17), and the CAR (19a) or (19b), we have the fused regression model as follows:

$$\begin{aligned} & \min_{\mathbf{X}', \mathbf{Y}', \Delta^x, \Delta^y} 2\text{Tr}(\mathbf{X}' \mathbf{L}^{t2} \mathbf{X}'^T) + 2\text{Tr}(\mathbf{Y}' \mathbf{L}^{t1} \mathbf{Y}'^T) \\ & + 2\beta \text{Tr}(\Delta^x \mathbf{L}^f \Delta^{xT}) + 2\beta \text{Tr}(\Delta^y \mathbf{L}^f \Delta^{yT}) \\ & + \lambda \|\Delta^x\|_{2,1} + \lambda \|\Delta^y\|_{2,1} + \eta \sum_{i=1}^{N_s} \phi(\|\Delta_i^x\|_2, \|\Delta_i^y\|_2) \end{aligned}$$

s.t. $\mathbf{X} = \mathbf{X}' - \Delta^x$, $\mathbf{Y} = \mathbf{Y}' - \Delta^y \quad (20)$

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

Problem (20) can be efficiently solved by using the alternating direction method of multipliers (ADMM) in the Appendix. From (20), we can find that the HGLR tends to obtain the solution of $\mathbf{X}' = \mathbf{Y}' = \mathbf{0}$, and both the PSR and FHSR tend to obtain the solution of $\Delta^x = \Delta^y = \mathbf{0}$, and the CAR tends to obtain the nonzero solution of Δ^x and Δ^y , which means that these regularization terms have an adversarial balancing effect. From (20), it can also be found that the \mathbf{X} and \mathbf{X}' are in the same domain based on two reasons: first, \mathbf{X}' is

Algorithm 1 SRF-Based MCD

Input: Images of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$, parameters of N_s , β , λ , and η .

Structure representation:

Segment images into superpixels by GMMSP.

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

Construct the graphs of G^{t1} and G^{t2} with (7).

Construct the hypergraphs of \mathcal{G}^{t1} , \mathcal{G}^{t2} and \mathcal{G}^f .

Compute the hypergraph Laplacian matrices of \mathbf{L}^{t1} , \mathbf{L}^{t2} and \mathbf{L}^f .

Structural regression fusion:

Initialize: set Δ^x , Δ^y , \mathbf{R}_1 , \mathbf{R}_2 , \mathbf{R}_3 , and $\mathbf{R}_4 = \mathbf{0}$.

Repeat:

1: Update \mathbf{X}' and \mathbf{Y}' through (24) and (25), respectively.

2: Update Δ^x and Δ^y through (27) based on different ϕ .

3: Update \mathbf{P}_1 and \mathbf{P}_2 through (30) and (31), respectively.

4: Update \mathbf{R}_1 , \mathbf{R}_2 , \mathbf{R}_3 , and \mathbf{R}_4 through (32).

Until the stopping criterion is met.

Change extraction:

Compute the difference images of DI^x and DI^y .

Segment the DI^x and DI^y to obtain the changed map.

decomposed from \mathbf{X} , and only a few columns between them are different (Δ^x is column sparse), which means that $\mathbf{X}'_i = \mathbf{X}_i$ ideally holds for most $i \in \mathcal{I}$; second, for the change \mathbf{X}_j , the regression model constrains it to be similar to its neighbors of by the HGLR, which prevents anomalous \mathbf{X}'_j . Similarly, \mathbf{Y} and \mathbf{Y}' are also in the same domain.

From the fused regression model (20), it can be found that forward and backward transformations are fused in one model by using the CAR. Turning back to Example 2 in Section II, the regression image and changed image obtained by the SRF are shown in Fig. 3. We can find that due to the structural asymmetry, there is always one failure and one success in forward or backward regression of SCASC [30]. By introducing the CAR, the two changed images are aligned in SRF, which can also be regarded as a change fusion. Furthermore, with this change fusion in SRF, one regression (e.g., forward) can use the supervised information (changed information) from another regression (e.g., backward). Therefore, this allows forward regression to overcome the influence of structural asymmetry, thus improving the performance of fused regression and obtaining more accurate changed images, as illustrated by Fig. 3.

D. Change Extraction

Once the regressed feature matrices of \mathbf{X}' and \mathbf{Y}' and the changed feature matrices of Δ^x and Δ^y are output from (20), the regression images can be calculated by $\tilde{\mathbf{X}}' = \mathcal{F}^{-1}(\mathbf{X}')$ and $\tilde{\mathbf{Y}}' = \mathcal{F}^{-1}(\mathbf{Y}')$, the change level vectors of \mathbf{p}^x and \mathbf{p}^y can be computed by $p_i^x = \|\Delta_i^x\|_2$ and $p_i^y = \|\Delta_i^y\|_2$, $i = 1, \dots, N_s$, and the corresponding DIs can be calculated by

$$\text{DI}^x(m, n) = p_i^x; \quad (m, n) \in \Lambda_i, \quad i = 1, \dots, N_s$$

$$\text{DI}^y(m, n) = p_i^y; \quad (m, n) \in \Lambda_i, \quad i = 1, \dots, N_s. \quad (21)$$

Then, we can obtain the binary CM by segmenting the DIs into two classes of changed and unchanged, which can

TABLE II
DESCRIPTION OF THE SIX MULTIMODAL DATASETS

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

be obtained by employing thresholding methods, such as Otsu threshold [56], or clustering methods, such as the K -means [57], or the random field-based methods, such as MRF [30], [43]. In this article, we directly use the MRF co-segmentation method proposed in IRG-McS [43], where the only hyperparameter is the balancing parameter.

The overall framework of the proposed SRF-based MCD method is summarized in Algorithm 1. The stopping criteria of the inner loop of updating Δ^x and Δ^y and the outer loop of solving minimization of (20) in SRF are as follows: the maximum numbers of iteration N_i and N_o are reached, or the relative difference between two iteration results is less than the tolerance.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The performance of the proposed SRF is verified on six pairs of real multimodal datasets by comparing with some SOTA methods, including six traditional methods (such as SCASC [30], AMD-IR [17], AGSCC [46], FPMS [45], M3CD [10], and IRG-McS [43]) and four deep-learning-based methods (such as CGAN [34], SCCN [19], X-Net [18], and ACE-Net [18]).

A. Experimental Setting

1) *Multimodal Datasets*: Six real datasets are employed in the experiments, as shown in Figs. 5 and 6, which includes two different multimodal cases: 1) the cross-sensor case, that is, the multitemporal images are acquired by different sensors but with the same sensor type, e.g., the images in Datasets 1–3 and 2) the multisource case, that is, the multitemporal images are acquired by different types of sensors, e.g., the images in Datasets 4–6. As listed in Table II, these datasets also cover different spatial resolutions (from 0.52 to 30 m), image sizes (from 300 to 4135 pixels in width or length), and types of change events (such as flooding, construction, and lake overflow), which can adequately evaluate the performance of comparison methods.

2) *Metric*: Two types of evaluation metrics are selected: 1) to evaluate the DI, the empirical receiver operating characteristics (ROCs) curve and the precision-recall (PR) curve are employed, and the corresponding areas under ROC curve (AUR) and PR curve (AUP) are also used and 2) to evaluate the binary CM, seven widely used metrics are adopted for assessment, including true positives (TPs), false positives (FPs), true negatives (TNs), false negatives (FNs), overall accuracy (OA), Kappa coefficient (κ), and $F1$ score ($F1$), where the TP, FP,

TN, and FN are marked in different colors in the qualitative results, and OA, κ , and $F1$ are listed in quantitative results.

3) *Implementation Detail*: For all the experiments of SRF, we adjust the scale parameter of GMMSP¹ to make the number of segmented superpixels $N_S \approx 5000$; and choose the CAR with function ϕ^1 (19a), and fix the balancing parameters in the fused regression model (20) as $\beta = 1$ and $\lambda = 0.1$ for all the datasets, and adjust the η by varying $\eta \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and selecting the best one as the result. The impact of these parameters will be analyzed in detail in Section IV-C.

B. Experimental Results

1) *Regression Images*: We compare the proposed SRF with four image regression-based MCD methods to evaluate the forward and backward regression images.

SCASC [30]: A sparse constrained adaptive structure Consistency-based regression method based on the image decomposition model (2), where only one-way (forward) regression is considered in the original paper.

AMD-IR [17]: An AMD-based image regression method that uses the AMD to construct the pseudo-training set, where both the forward and backward regressions are performed, but the change results are simply averaged for fusion.

CGAN [34]: A deep translation method that uses a conditional GAN, where only one-way image translation is performed in this article.

AGSCC [46]: An adaptive graph and structure cycle consistency-based regression method, which also only performs a one-way regression.

To show the difference between forward and backward regressions and to demonstrate the effect of structural asymmetry on the image regression, we present the forward regression image and backward regression image of each method separately, as shown in Figs. 5(c)–(g) and 6(c)–(g). For the one-way regression-based SCASC, CGAN, and AGSCC, we swap the order of the input images (pre- and post-event images) to get another regression image.

From the regression images of $\tilde{\mathbf{X}}'$ and $\tilde{\mathbf{Y}}'$ in Figs. 5 and 6, we can find that most methods can achieve image translation with a certain degree, i.e., the translated image is similar in style to the target image, such as the $\tilde{\mathbf{X}}'$ and $\tilde{\mathbf{X}}$, and $\tilde{\mathbf{Y}}'$ and $\tilde{\mathbf{Y}}$. However, when we carefully compare the translated image with the original image, we find that some of the methods fail to preserve the structure of the original image, although

¹ Available at <https://github.com/ahban/GMSP-superpixel>.

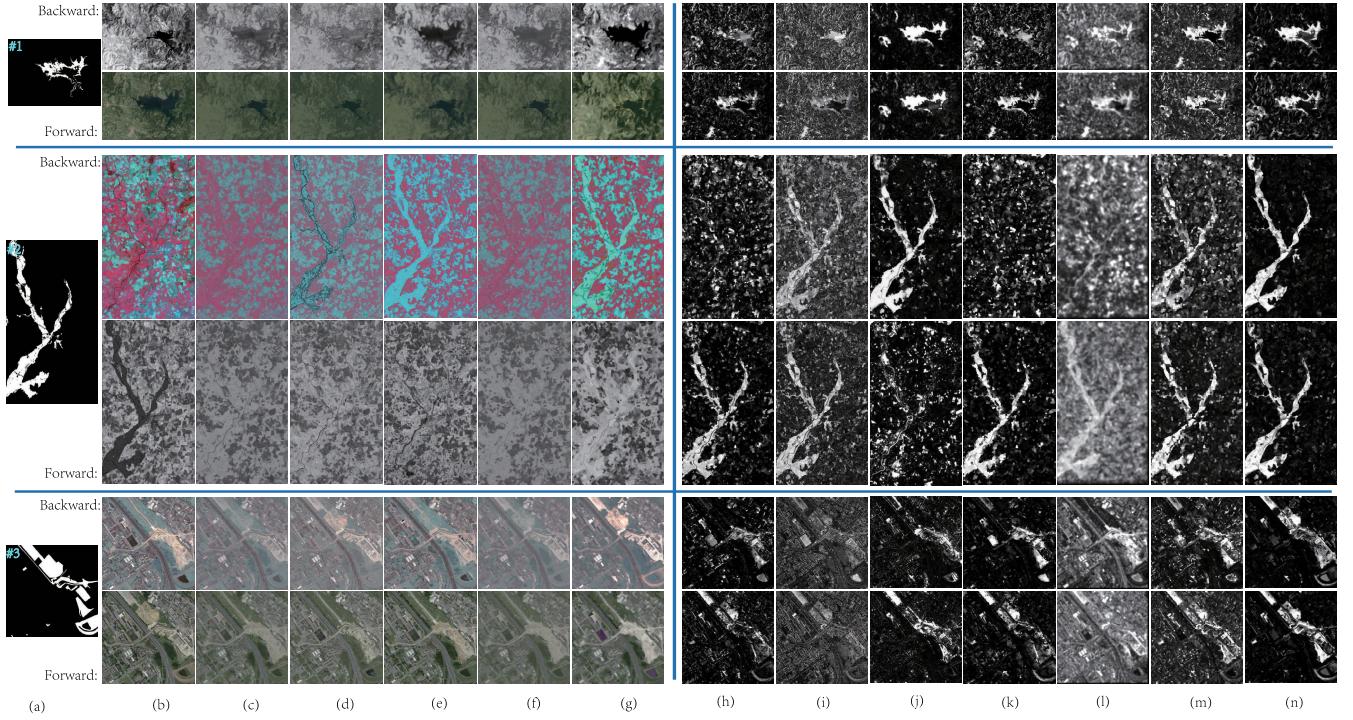


Fig. 5. Multimodal datasets, forward and backward regression images, and DIs on Datasets 1–3. From top to bottom, they correspond to Datasets 1–3, respectively. From left to right are (a) ground truth; (b) pre-event image \tilde{X} and post-event image \tilde{Y} ; (c)–(g) backward and forward regression images of \tilde{X}' and \tilde{Y}' generated by (c) SCASC, (d) AMD-IR, (e) CGAN, (f) AGSCC, and (g) proposed SRF; and (h)–(n) backward and forward DIs of DI^x and DI^y generated by (h) SCASC, (i) AMD-IR, (j) CGAN, (k) AGSCC, (l) FPMS, (m) IRG-McS, and (n) proposed SRF.

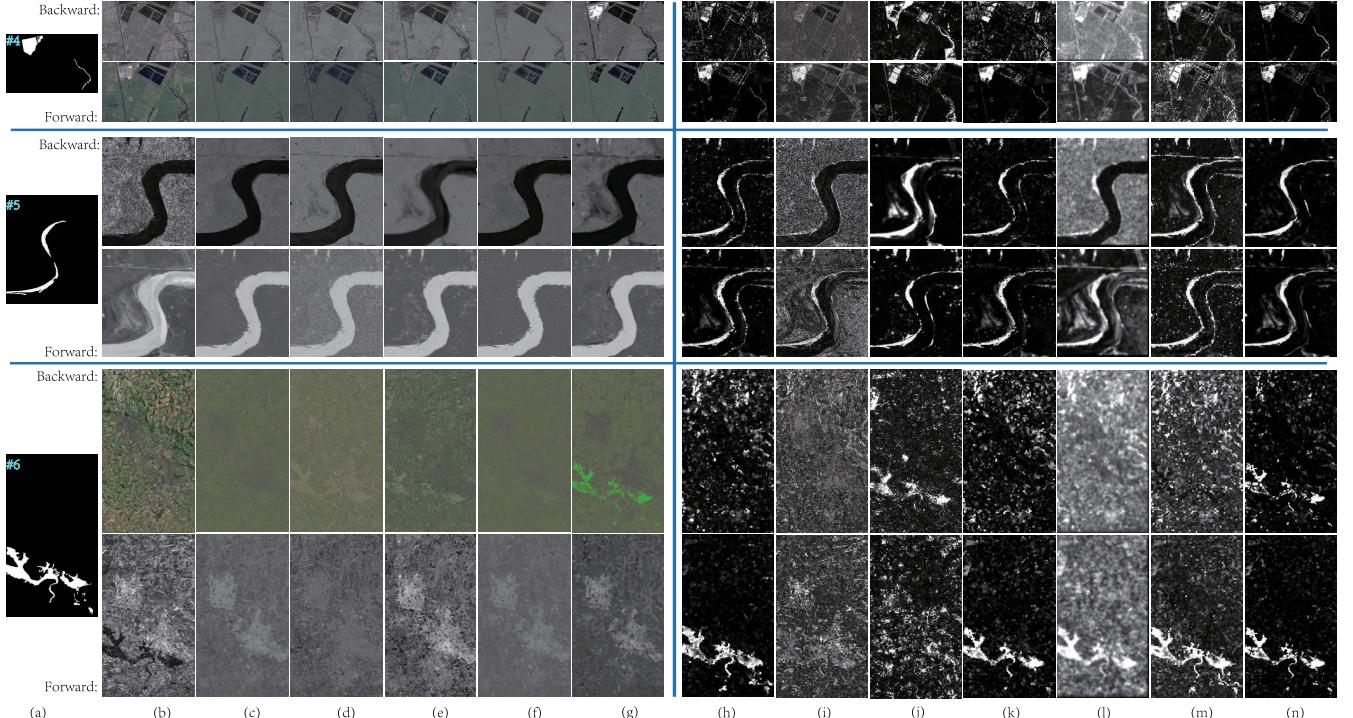


Fig. 6. Multimodal datasets, forward and backward regression images, and DIs on Datasets 4–6. From top to bottom, they correspond to Datasets 4–6, respectively. From left to right are (a) ground truth; (b) pre-event image \tilde{X} and post-event image \tilde{Y} ; (c)–(g) backward and forward regression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

the image style is transferred, for example, the backward \tilde{X}' of AMD-IR in Dataset 1, AGSCC in Dataset 2, CGAN in Dataset 3, and SCASC in Dataset 6 as shown in Figs. 5 and 6.

In particular, this regression error is most pronounced for Datasets 2 and 6, where the structural asymmetry is prominent. At the same time, it can be seen that the proposed SRF can

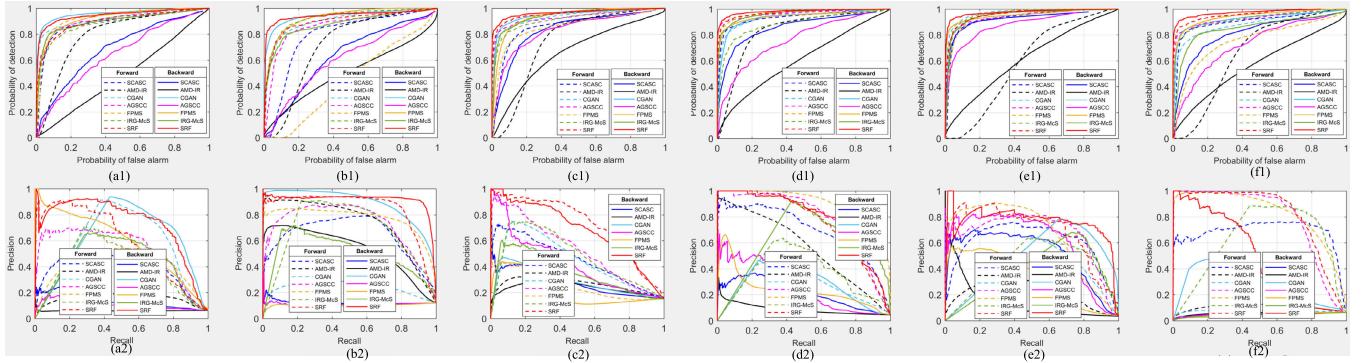


Fig. 7. ROC and PR curves of DIs generated by different methods. (Top row) ROC curves. (Bottom row) PR curves. (a)–(f) Results on Datasets 1–6, respectively.

TABLE III

QUANTITATIVE MEASURES OF DIs ON THE MULTIMODAL DATASETS. THE HIGHEST SCORES ARE HIGHLIGHTED IN BOLD

Methods	Dataset #1				Dataset #2				Dataset #3				Dataset #4				Dataset #5				Dataset #6				Average			
	Forward		Backward																									
	AUR	AUP																										
SCASC [30]	0.885	0.383	0.688	0.133	0.936	0.636	0.491	0.123	0.793	0.458	0.652	0.276	0.968	0.695	0.806	0.218	0.969	0.597	0.927	0.517	0.971	0.681	0.388	0.048	0.920	0.575	0.659	0.219
AMD-IR [17]	0.795	0.155	0.482	0.060	0.927	0.741	0.867	0.537	0.708	0.264	0.676	0.237	0.922	0.564	0.631	0.090	0.859	0.216	0.697	0.137	0.691	0.103	0.475	0.057	0.817	0.341	0.638	0.186
CGAN [34]	0.922	0.550	0.957	0.590	0.746	0.238	0.979	0.887	0.773	0.436	0.667	0.284	0.890	0.273	0.853	0.270	0.950	0.375	0.990	0.474	0.521	0.073	0.812	0.287	0.800	0.324	0.876	0.465
AGSCC [46]	0.905	0.522	0.646	0.106	0.940	0.719	0.470	0.113	0.802	0.534	0.735	0.427	0.959	0.787	0.746	0.257	0.955	0.664	0.917	0.606	0.928	0.797	0.443	0.053	0.915	0.672	0.660	0.260
FPMS [45]	0.925	0.406	0.922	0.590	0.969	0.774	0.421	0.102	0.597	0.258	0.467	0.224	0.994	0.904	0.845	0.237	0.952	0.652	0.848	0.327	0.977	0.836	0.421	0.050	0.902	0.638	0.654	0.255
IRG-McS [43]	0.881	0.393	0.897	0.424	0.941	0.658	0.900	0.499	0.813	0.469	0.717	0.352	0.934	0.359	0.984	0.582	0.972	0.414	0.976	0.372	0.964	0.605	0.234	0.039	0.918	0.483	0.785	0.378
proposed SRF	0.900	0.591	0.945	0.734	0.958	0.845	0.988	0.914	0.898	0.727	0.886	0.683	0.962	0.760	0.963	0.782	0.975	0.694	0.979	0.761	0.946	0.823	0.792	0.486	0.940	0.740	0.926	0.727

not only overcome this structural asymmetry on these two datasets (2 and 6), but also achieve better regression results on other datasets by combining the forward and backward transformation processes, e.g., the regression images of SRF in Datasets 1 and 4. Also, we can find that the superpixel-based regression methods have a block smoothing effect, such as the regression images of SCASC, AGSCC, and SRF on Datasets 1, 4, and 6.

2) *Difference Images*: In order to assess the ability of forward regression (or transformation) and backward regression (or transformation) to measure change, we show the forward DI and backward DI generated by different methods, which not only contain the four methods compared above (in Figs. 5 and 6), but also two additional ones as follows.

FPMS [45]: A fractal projection and Markovian segmentation-based method that translates the image by using a spatial fractal decomposition and a contractive projection, which only performs a forward translation and neglects the fusion.

IRG-McS [43]: An iterative robust graph and MRF co-segmentation-based MCD method, which calculates the forward and backward DIs (1) by transforming images into the same image differential domain and fuses the DIs in the segmentation process.

Figs. 5(h)–(n) and 6(h)–(n) show the forward and backward DIs generated by different methods. Four facets can be observed: first, there is a difference between the forward and backward results of each method, which is because these DIs are computed in different domains and the performances of the forward and backward regressions (or transformations) are also different. This once again confirms the importance of fusing forward and backward regressions. Second, the backward

regression of some methods cannot obtain a useful DI because of the structural asymmetry between the multimodal images, such as the backward DI^x of SCASC, AMD-IR, and AGSCC in Dataset 1; DI^x of SCASC, AGSCC, and FPMS in Dataset 2; and DI^x of SCASC, AMD-IR, AGSCC, FPMS, and IRG-McS in Dataset 6. Third, the proposed SRF can obtain both forward and backward DIs with good performance that overcomes the effect of structural asymmetry. Moreover, the forward and backward DIs of SRF are aligned, i.e., the regions they highlighted are mostly identical, which is attributed to the CAR regularization term of (19a). Fourth, the DIs of SRF are sparse by using the sparsity regularization of $\|\Delta^x\|_{2,1}$ and $\|\Delta^y\|_{2,1}$ in the model (20), which means that satisfactory segmentation maps can be directly obtained even by the simple thresholding methods (such as Otsu threshold [56]) or clustering methods (such as K-means [57]).

Fig. 7 plots the ROC and PR curves of DIs generated by the comparison methods, and Table III lists the corresponding AUR and AUP. Although SRF is not optimal in terms of metrics on some datasets, such as the AUR and AUP of forward DIs on Datasets 4 and 6, SRF is able to consistently obtain high-quality DIs across different datasets, both in the forward and backward transformations. The average AUR of the forward and backward DIs generated by SRF is 0.940 and 0.926, respectively, and the corresponding average AUP is 0.740 and 0.727, respectively, which are all higher than that of other methods. For example, the average forward AUP is 6.8% higher than the second-ranked AGSCC, and the average backward AUP is 26.2% higher than the second-ranked CGAN. This once again demonstrates the benefits of structure regression fusion: first, the ability to overcome the influence of structural asymmetry (e.g., the performance

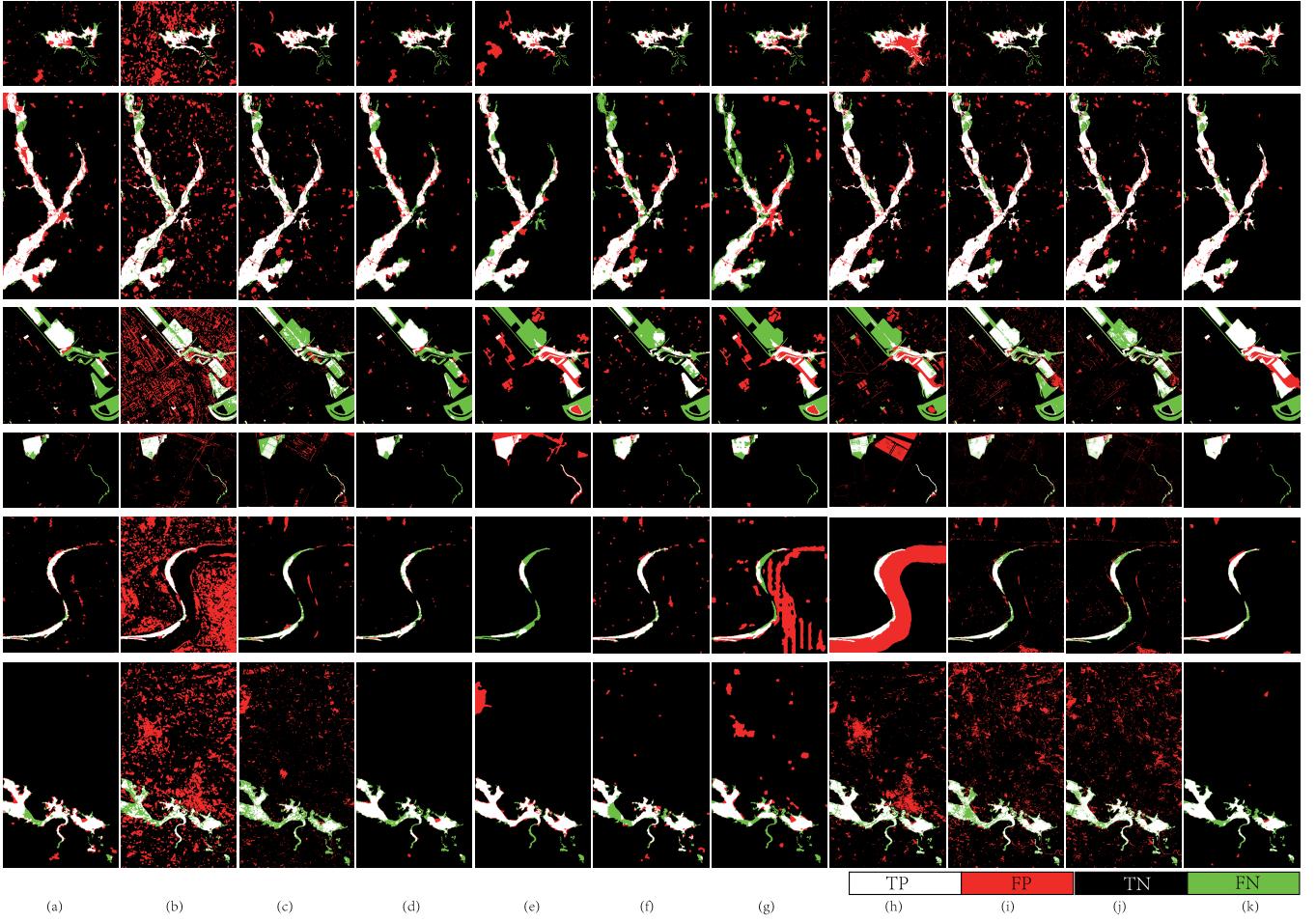


Fig. 8. Binary CMs of different methods on multimodal datasets. From top to bottom, they correspond to Datasets 1–6, respectively. From left to right are binary CMs generated by (a) SCASC, (b) AMD-IR, (c) CGAN, (d) AGSCC, (e) FPMS, (f) IRG-McS, (g) M3CD, (h) SCCN, (i) X-Net, (j) ACE-Net, and (k) proposed SRF. In the binary CM, white: TPs, red: FPs, black: TNs, and green: FNs.

improvement of backward DI), and second, the ability to gain better regression images and more accurate changed images by fusion (e.g., the performance improvement of forward DI).

3) Change Maps: In the third experiment, we compare the final CMs generated by different methods, including the six methods of SCASC, AMD-IR, CGAN, AGSCC, IRG-McS, and FPMS compared above (in Figs. 5–7) and another four methods as follows.

M3CD [10]: An MRF-based method that builds up the observation field from a pixel pairwise modeling and solves the energy minimization problem by using iterative conditional estimation.

SCCN [19]: A deep-learning-based method that uses a symmetric convolutional coupling network to transform the multimodal images into a common feature space.

X-Net [18]: A weighted translation network that consists of two fully convolutional networks, each dedicated to mapping the data from one domain to the other.

ACE-Net [18]: A deep image translation method that uses an ACE-Net consisting of two autoencoders. Both the X-Net and ACE-Net use the affinity-based change prior to train the network.

For the sake of fair comparison, for these methods with only one-way regression or transformation (such as SCASC,

CGAN, AGSCC, M3CD, and FPMS) in their original papers, we choose the better of the forward and backward results for the comparison.

Fig. 8 shows the final CMs generated by different methods, and Table IV lists the quantitative measures of these CMs. Because we choose the better results of the forward and backward CMs of some comparison methods, most of the CMs in Fig. 8 can roughly reflect the change information, i.e., detect the changed regions. However, some methods do not perform robustly enough. For example, the AMD-IR introduces a large number of FPs in Datasets 3, 5, and 6; the M3CD and SCCN gain a lot of FPs in Dataset 5, resulting in $\kappa = 0.158$ and $\kappa = 0.183$, respectively. In addition, we can see that most methods do not perform well on Dataset 3, mainly for the following reasons: the resolution of the multitemporal images of Dataset 3 is very high (0.52 m); it contains more types of objects in the images than other datasets (such as buildings, grass, trees, roads, and pitches), and the proportion of these ground objects is quite uneven; the changed region in Dataset 3 is also relatively larger than other datasets, as shown in Figs. 5(a) and (b) and 6(a) and (b), which poses difficulties for accurate detection of changes. Nevertheless, the proposed SRF can still detect most of the changes for Dataset 3 ($F1 = 0.703$). The average κ and $F1$ of the proposed SRF on these

TABLE IV
QUANTITATIVE MEASURES OF CMs ON THE MULTIMODAL DATASETS. THE HIGHEST SCORES ARE HIGHLIGHTED IN BOLD

Methods	Dataset #1			Dataset #2			Dataset #3			Dataset #4			Dataset #5			Dataset #6			Average		
	OA	κ	F1																		
SCASC [30]	0.947	0.593	0.621	0.946	0.773	0.804	0.892	0.464	0.516	0.979	0.741	0.751	0.977	0.700	0.711	0.973	0.774	0.788	0.952	0.674	0.699
AMD-IR [17]	0.799	0.255	0.328	0.898	0.607	0.664	0.724	0.259	0.411	0.950	0.572	0.597	0.686	0.115	0.170	0.782	0.171	0.254	0.807	0.330	0.404
CGAN [34]	0.965	0.724	0.742	0.833	0.155	0.248	0.863	0.338	0.402	0.943	0.402	0.432	0.971	0.584	0.599	0.911	0.364	0.412	0.914	0.428	0.473
AGSCC [46]	0.959	0.658	0.680	0.955	0.791	0.817	0.897	0.490	0.540	0.983	0.773	0.782	0.982	0.724	0.733	0.976	0.766	0.779	0.959	0.700	0.722
FPMS [45]	0.938	0.593	0.625	0.962	0.816	0.837	0.827	0.269	0.368	0.938	0.569	0.597	0.979	0.544	0.553	0.970	0.770	0.786	0.936	0.594	0.628
IRG-McS [43]	0.971	0.739	0.754	0.939	0.714	0.749	0.882	0.420	0.478	0.983	0.794	0.804	0.976	0.690	0.702	0.971	0.740	0.755	0.954	0.683	0.707
M3CD [10]	0.958	0.650	0.672	0.944	0.756	0.788	0.827	0.269	0.368	0.976	0.684	0.696	0.856	0.158	0.204	0.952	0.618	0.643	0.919	0.523	0.562
SCCN [19]	0.919	0.522	0.562	0.957	0.810	0.835	0.818	0.240	0.342	0.903	0.315	0.359	0.789	0.183	0.232	0.907	0.474	0.521	0.882	0.424	0.475
X-Net [18]	0.958	0.674	0.696	0.952	0.789	0.816	0.864	0.389	0.461	0.954	0.586	0.609	0.959	0.490	0.510	0.901	0.447	0.497	0.931	0.563	0.598
ACE-Net [18]	0.967	0.723	0.740	0.947	0.768	0.798	0.868	0.463	0.538	0.964	0.651	0.670	0.962	0.526	0.546	0.878	0.341	0.401	0.931	0.579	0.616
proposed SRF	0.971	0.755	0.771	0.968	0.851	0.869	0.914	0.652	0.703	0.987	0.838	0.845	0.982	0.752	0.761	0.978	0.788	0.800	0.967	0.773	0.792

TABLE V
ABLATION STUDY OF SRF MEASURED THE AVERAGE SCORES

Methods	Components			Forward DI		Backward DI		Final CM		
	HGs	FHSR	CAR	AUR	AUP	AUR	AUP	OA	κ	F1
Baseline	x	x	x	0.918	0.574	0.674	0.230	0.951	0.677	0.701
Settings	✓	✗	✗	0.919	0.595	0.689	0.244	0.952	0.688	0.712
	✗	✓	✗	0.923	0.622	0.697	0.265	0.956	0.695	0.719
	✓	✓	✗	0.925	0.630	0.711	0.281	0.956	0.700	0.726
	✗	✗	✓	0.933	0.699	0.911	0.687	0.960	0.741	0.763
	✓	✗	✓	0.936	0.713	0.918	0.706	0.962	0.750	0.769
SRF	✓	✓	✓	0.937	0.729	0.920	0.715	0.964	0.761	0.778

six datasets are about 0.773 and 0.792, respectively, which are 7.3% and 7.0% higher than that of the second-ranked AGSCC, respectively. This shows that the proposed SRF is very effective and competitive, even compared with some deep-learning-based methods.

C. Discussions

1) *Ablation Study*: The proposed SRF mainly contains two processes: structure representation and SRF. Then, we investigate the contributions of different components in these two processes for the MCD problem, i.e., the HGs of \mathcal{G}^{t1} and \mathcal{G}^{t2} in the structure representation, and the regularization terms of FHSR and CAR in the fused regression model (20). We construct a baseline by replacing the HG with the graph G^{t1} learned by model (6) and deleting the FHSR and CAR in the regression model (20). The detection performances of the SRF with and without the HGs, FHSR, and CAR are reported in Table V.

According to Table V, it is clear that the performance of SRF degrades without HGs, FHSR, and CAR. To be specific, by using the HG that links more than two vertices, the high-order neighborhood relationship can be preserved by \mathcal{G}^{t1} , \mathcal{G}^{t2} , and \mathcal{G}^f , which helps the graphs capture more comprehensively structure information of images, thus improving the performance of structural regression. Specifically, when adding HGs to the baseline model, the average AUP of forward DI increases by about 2.1%. By applying the FHSR of (17) that constrains the smoothness of changed images on the fused HG, the scores of average AUP of forward DI and κ of CM are about 4.8% and 1.8% higher than that without the FHSR of baseline. By introducing the CAR of (19a) to the baseline model that requires the forward and backward changed images that are aligned on the support sets, the SRF-based model (20)

can obtain much better regression images and more accurate DIs, especially for the backward regression process (e.g., the AUR and AUP of the backward DI). As can be seen from Table V, a significant improvement in detection performance is achieved when CAR is employed, e.g., the average κ is improved by 6.4% compared with that without CAR. In addition, when two combinations of HGs, FHSR, and CAR are used, the detection performance improvement is even more pronounced. Finally, the performance improvement is greatest when all modules are introduced, e.g., a 9.6% improvement on the average κ compared with the baseline.

2) *Parameter Analysis*: The main parameters in the proposed SRF are as follows: the number of superpixels N_S ; and the balancing parameters of β , λ , and η that control the weights of FHSR, PSR, and CAR in the fused regression model (20), respectively.

In general, the number of superpixels N_S should be selected according to the resolution of dataset and taking into account the computational efficiency requirements of CD task. A large N_S can improve the detection granularity while increasing the computational burden of the algorithm as analyzed in Section IV-C3. The computational time of each process of SRF with different values of N_S on Datasets 2 and 6 is listed in Table VI.

The parameters of β , λ , and η are used to balance the regularization terms of the fused regression model (20). For the parameter β , it controls the weight of FHSR that constrains the smoothness of changed images ($\tilde{\Delta}^x$ and $\tilde{\Delta}^y$) on the fused HG (\mathcal{G}^f). From its formulation of (17), we can see that FHSR is very similar to the HGLR that constrains the smoothness of regression images ($\tilde{\mathbf{Y}}'$ and $\tilde{\mathbf{X}}'$) on the HGs (\mathcal{G}^{t1} and \mathcal{G}^{t2}) as in (13). Since FHSR and HGLR are similar in form and have a similar magnitude of values, we set $\beta = 1$ for simplicity.

To measure the impact of λ and η on SRF, we adjust the λ from 2^{-8} to 1 with the ratio of 2 and adjust the η from 0.1 to 0.9 with the step of 0.1, and plot the OA generated by SRF with different values of λ and η in Fig. 9. The λ is used to control the sparsity level of changed images in the model (20), which should be adjusted according to the percentage of the changed region. Based on Fig. 9(a), we fix $\lambda = 0.1$ (i.e., $\lambda \approx 2^{-3.3}$) in our experiments for simplicity.

For the parameter η , it is used to control the CAR that constrains the change alignment in the model (20). From Fig. 9(b), we can find that too small and too large η are not suitable: first, if η is very small, then CAR does not

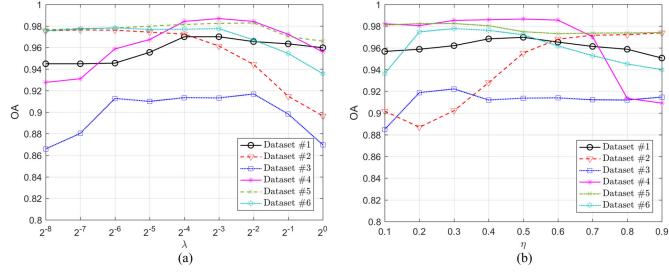
Fig. 9. Sensitivity analysis of parameters in SRF. (a) λ . (b) η .

TABLE VI

COMPUTATIONAL TIME (S) OF SRF, WHERE t_{SR} , t_{RF} , t_{CE} , AND t_{TOTAL} DENOTE THE COMPUTATIONAL TIME OF STRUCTURAL REPRESENTATION, REGRESSION FUSION, CHANGE EXTRACTION, AND THE WHOLE SRF, RESPECTIVELY

Data sets	N_S	t_{sr}	t_{rf}	t_{ce}	t_{total}
Dataset #2 900 × 554 × 3(1)	2500	1.93	1.51	1.90	5.41
	5000	2.61	7.98	5.13	15.88
	10000	6.39	49.95	16.66	74.91
Dataset #6 4135 × 2325 × 3(1)	2500	3.42	1.69	2.58	7.93
	5000	4.19	7.99	5.89	18.67
	10000	9.27	55.14	18.02	83.50

play a noticeable effect in the structural regression, which will limit the regression and detection performance of the model; second, if η is very large, the CAR tends to generate changed superpixels; i.e., it will break the balance between different regularization terms in the model (20). It can be seen from Fig. 9(b) that SRF is a little sensitive to the parameter η , and how to reduce this sensitivity, i.e., how to make SRF robust to η needs to be further investigated.

3) *Complexity Analysis*: The main computational complexity of the proposed SRF is concentrating on the structure representation, regression fusion by solving the minimization model (20), and the change extraction by using the MRF co-segmentation method.

In the structure representation, calculating distance matrices of \mathbf{D}^{t1} and \mathbf{D}^{t2} requires $\mathcal{O}((b_x + b_y)N_S^2)$, sorting the distance matrices by column requires $\mathcal{O}(N_S^2 \log N_S)$, calculating the adaptive k by an in-degree-based strategy [30] requires $\mathcal{O}(N_S^2 \log N_S)$, calculating the closed-form solution of \mathbf{S}^{t1} and \mathbf{S}^{t2} requires $\mathcal{O}(N_S^2)$, and constructing the HG requires $\mathcal{O}((b_x + b_y)N_S^2)$. Therefore, the structure representation requires $\mathcal{O}(N_S^2 \log N_S)$. In the regression fusion, updating \mathbf{X}' and \mathbf{Y}' through (24) and (25) requires $\mathcal{O}(N_S^3)$, updating Δ^x and Δ^y through (27) requires $\mathcal{O}((b_x + b_y)N_S)$, updating \mathbf{P}_1 and \mathbf{P}_2 through (30) and (31) requires $\mathcal{O}(N_S^3)$, and updating the Lagrange multipliers through (32) requires $\mathcal{O}((b_x + b_y)N_S)$. Therefore, the regression fusion requires $\mathcal{O}(N_S^3)$. In the change extraction, the MRF co-segmentation requires $\mathcal{O}(2N_R N_S^2)$ for the worst case, where N_R is the number of edges in the R -adjacency neighbor system of the MRF co-segmentation model proposed in [43].

Although the SRF requires $\mathcal{O}(N_S^3)$ for each iteration in the solving model (20), two acceleration strategies are available

to improve the efficiency of updating \mathbf{X}' , \mathbf{Y}' , \mathbf{P}_1 , and \mathbf{P}_2 as introduced in [30]. Taking \mathbf{X}' update as an example, we can compute the matrix inversion of $(\mu_1 \mathbf{I}_{N_S} + 4\mathbf{L}^{t2})^{-1}$ offline in advance or solving the \mathbf{X}' subproblem (23) with preconditioned conjugate gradient (PCG) method.

The computational time of each process of SRF with different values of N_S on Datasets 2 and 6 is listed in Table VI, where the SRF is performed in MATLAB 2016a running with Intel Core i9-10980HK CPU. As can be seen in Table VI, the number of superpixels is the main factor affecting the running time rather than the image size, and the regression fusion is the most time-consuming process in SRF, which is in accordance with the theoretical analysis.

V. CONCLUSION

In this article, we focus on the CD of multimodal remote sensing images. We first investigate the previous structure-based MCD approaches and show the negative effects of structural asymmetry on these approaches. Then, to address the challenge posed by structural asymmetry, we propose an SRF method to fuse the forward and backward transformation processes. Specifically, the proposed method first segments the images into superpixels and constructs HGs to represent the structure information, which can capture the high-order information. Then, based on the inherent connection between forward and backward transformations, the proposed method uses three types of constraints to perform the SRF: structural regression term, change fusion term, and sparse penalty term. By putting the two mutually reinforcing processes of transformation and fusion in one model, it can obtain better regression images and changed images and, thus, improve the accuracy of CD. The experimental results on six real datasets confirm the effectiveness of the proposed method by comparing with other related methods.

The analysis of structural asymmetry has motivated us to propose the SRF in this article. However, the structural asymmetry can also be used for the fusion of structure comparison and the late fusion of DIs generated by some other structure-based MCD methods, especially for the graph neural networks-based methods, thus improving the performance of these methods. In addition, this article analyzes the impact of structural asymmetry from the perspective of structure comparison and structure regression models, but it does not provide an insight into structural asymmetry from the perspective of graph representation capability or the spectral domain properties of graphs. Besides, the proposed method is a little sensitive to the CAR. Our future work is to further analyze the structural asymmetry for the MCD problem and consider a more robust regularization or adopt some parameter adaptive strategies to improve the robustness of the algorithm.

APPENDIX

We use the ADMM for solving the minimization problem of (20). By introducing the auxiliary constraints of $\mathbf{P}_1 = \Delta^x$

and $\mathbf{P}_2 = \Delta^y$, the augmented Lagrangian function of (20) is

$$\begin{aligned} \Theta(\mathbf{X}', \mathbf{Y}', \Delta^x, \Delta^y) &= 2\text{Tr}(\mathbf{X}'\mathbf{L}^{t2}\mathbf{X}'^T) + 2\text{Tr}(\mathbf{Y}'\mathbf{L}^{t1}\mathbf{Y}'^T) \\ &\quad + 2\beta\text{Tr}(\mathbf{P}_1\mathbf{L}^f\mathbf{P}_1^T) + 2\beta\text{Tr}(\mathbf{P}_2\mathbf{L}^f\mathbf{P}_2^T) \\ &\quad + \lambda\|\Delta^x\|_{2,1} + \lambda\|\Delta^y\|_{2,1} + \eta \sum_{i=1}^{N_S} \phi(\|\Delta_i^x\|_2, \|\Delta_i^y\|_2) \\ &\quad + \text{Tr}(\mathbf{R}_1^T(\mathbf{X}' - \mathbf{X} - \Delta^x)) + \text{Tr}(\mathbf{R}_2^T(\Delta^x - \mathbf{P}_1)) \\ &\quad + \text{Tr}(\mathbf{R}_3^T(\mathbf{Y}' - \mathbf{Y} - \Delta^y)) + \text{Tr}(\mathbf{R}_4^T(\Delta^y - \mathbf{P}_2)) \\ &\quad + \frac{\mu_1}{2}\|\mathbf{X}' - \mathbf{X} - \Delta^x\|_F^2 + \frac{\mu_2}{2}\|\Delta^x - \mathbf{P}_1\|_F^2 \\ &\quad + \frac{\mu_3}{2}\|\mathbf{Y}' - \mathbf{Y} - \Delta^y\|_F^2 + \frac{\mu_4}{2}\|\Delta^y - \mathbf{P}_2\|_F^2 \end{aligned} \quad (22)$$

where $\mathbf{R}_1, \mathbf{R}_2 \in \mathbb{R}^{2b_x \times N_S}$ and $\mathbf{R}_3, \mathbf{R}_4 \in \mathbb{R}^{2b_y \times N_S}$ are Lagrangian multipliers, and $\mu_1, \mu_2, \mu_3, \mu_4 > 0$ are penalty parameters. The minimization of (22) can be divided into the following subproblems.

1) \mathbf{X}' and \mathbf{Y}' Subproblems: The minimization of (22) with respect to \mathbf{X}' can be written as follows:

$$\min_{\mathbf{X}'} 2\text{Tr}(\mathbf{X}'\mathbf{L}^{t2}\mathbf{X}'^T) + \frac{\mu_1}{2}\left\|\mathbf{X}' - \mathbf{X} - \Delta^x + \frac{\mathbf{R}_1}{\mu_1}\right\|_F^2 \quad (23)$$

which can be solved by taking the first-order derivative of the objective function to zero. Then, we can update \mathbf{X}' as follows:

$$\mathbf{X}' = (\mu_1\mathbf{X} + \mu_1\Delta^x - \mathbf{R}_1) \times (\mu_1\mathbf{I}_{N_S} + 4\mathbf{L}^{t2})^{-1} \quad (24)$$

where \mathbf{I}_{N_S} denotes an $N_S \times N_S$ identity matrix. Similarly, for the regressed \mathbf{Y}' , we can update it as follows:

$$\mathbf{Y}' = (\mu_3\mathbf{Y} + \mu_3\Delta^y - \mathbf{R}_3) \times (\mu_3\mathbf{I}_{N_S} + 4\mathbf{L}^{t1})^{-1}. \quad (25)$$

2) Δ^x and Δ^y Subproblems: The minimization of (22) with respect to Δ^x can be written as follows:

$$\begin{aligned} \min_{\Delta^x} \lambda\|\Delta^x\|_{2,1} + \eta \sum_{i=1}^{N_S} \phi(\|\Delta_i^x\|_2, \|\Delta_i^y\|_2) \\ + \frac{\mu_1}{2}\left\|\mathbf{X}' - \mathbf{X} - \Delta^x + \frac{\mathbf{R}_1}{\mu_1}\right\|_F^2 + \frac{\mu_2}{2}\left\|\Delta^x - \mathbf{P}_1 + \frac{\mathbf{R}_2}{\mu_2}\right\|_F^2. \end{aligned} \quad (26)$$

The gradient descent method is used for this minimization problem. We set the iteration number of inner loop for Δ^x as N_i and the step size as τ ; then, we have

$$\begin{aligned} g(\Delta^x) &= \varrho \odot \Delta^x + \mu_1\mathbf{X} - \mu_1\mathbf{X}' - \mathbf{R}_1 - \mu_2\mathbf{P}_1 + \mathbf{R}_2 \\ \Delta^x &\leftarrow \Delta^x - \tau g(\Delta^x) \end{aligned} \quad (27)$$

where \odot denotes the Hadamard product and $\varrho \in \mathbb{R}^{2b_x \times N_S}$ is calculated by

$$\varrho_{i,j}^1 = \frac{\lambda - \eta\|\Delta_i^y\|_2}{\|\Delta_i^x\|_2} + \mu_1 + \mu_2 \quad (28a)$$

$$\varrho_{i,j}^2 = \frac{\lambda - \eta\|\Delta_i^y\|_2 \exp(-\|\Delta_i^x\|_2\|\Delta_i^y\|_2)}{\|\Delta_i^x\|_2} + \mu_1 + \mu_2 \quad (28b)$$

according to the choice of ϕ^1 (19a) or ϕ^2 (19b). Similarly, we can update Δ^y in the same way as updating Δ^x .

3) \mathbf{P}_1 and \mathbf{P}_2 Subproblems: The minimization of (22) with respect to \mathbf{P}_1 can be written as follows:

$$\min_{\mathbf{P}_1} 2\beta\text{Tr}(\mathbf{P}_1\mathbf{L}^f\mathbf{P}_1^T) + \frac{\mu_2}{2}\left\|\Delta^x - \mathbf{P}_1 + \frac{\mathbf{R}_2}{\mu_2}\right\|_F^2 \quad (29)$$

which can be solved as follows:

$$\mathbf{P}_1 = (\mu_2\Delta^x + \mathbf{R}_2) \times (\mu_2\mathbf{I}_{N_S} + 4\beta\mathbf{L}^f)^{-1}. \quad (30)$$

Similarly, for the \mathbf{P}_2 , we can update it as follows:

$$\mathbf{P}_2 = (\mu_4\Delta^y + \mathbf{R}_4) \times (\mu_4\mathbf{I}_{N_S} + 4\beta\mathbf{L}^f)^{-1}. \quad (31)$$

Finally, the Lagrangian multipliers of $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$, and \mathbf{R}_4 can be updated as follows:

$$\begin{aligned} \mathbf{R}_1 &\leftarrow \mathbf{R}_1 + \mu_1(\mathbf{X}' - \mathbf{X} - \Delta^x) \\ \mathbf{R}_2 &\leftarrow \mathbf{R}_2 + \mu_2(\Delta^x - \mathbf{P}_1) \\ \mathbf{R}_3 &\leftarrow \mathbf{R}_3 + \mu_3(\mathbf{Y}' - \mathbf{Y} - \Delta^y) \\ \mathbf{R}_4 &\leftarrow \mathbf{R}_4 + \mu_4(\Delta^y - \mathbf{P}_2). \end{aligned} \quad (32)$$

ACKNOWLEDGMENT

The authors would like to thank the researchers for their friendly sharing of multimodal change detection codes and datasets. They would also like to thank the editors and anonymous reviewers for their constructive suggestions that improved the presentation of this work.

REFERENCES

- [1] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, Jun. 1989.
- [2] Z. Y. Lv, W. Shi, X. Zhang, and J. A. Benediktsson, "Landslide inventory mapping from bitemporal high-resolution remote sensing images using change detection and multiscale segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1520–1532, May 2018.
- [3] H. Chen, E. Nemni, S. Vallecorsa, X. Li, C. Wu, and L. Bromley, "Dual-tasks Siamese transformer framework for building damage assessment," 2022, *arXiv:2201.10953*.
- [4] Z. Lv, T. Liu, J. A. Benediktsson, and N. Falco, "Land cover change detection techniques: Very-high-resolution optical images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 44–63, Mar. 2022.
- [5] D. Wen et al., "Change detection from very-high-spatial-resolution optical remote sensing images: Methods, applications, and future directions," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 68–101, Dec. 2021.
- [6] Z. Lv, F. Wang, G. Cui, J. A. Benediktsson, T. Lei, and W. Sun, "Spatial-spectral attention network guided with change magnitude image for land cover change detection using remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4412712.
- [7] H.-C. Li, G. Yang, W. Yang, Q. Du, and W. J. Emery, "Deep nonsmooth nonnegative matrix factorization network with semi-supervised learning for SAR image change detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 160, pp. 167–179, Feb. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271619302898>
- [8] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2019.
- [9] G. Mercier, G. Moser, and S. B. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1428–1441, May 2008.
- [10] R. Touati, M. Mignotte, and M. Dahmane, "Multimodal change detection in remote sensing images using an unsupervised pixel pairwise-based Markov random field model," *IEEE Trans. Image Process.*, vol. 29, pp. 757–767, 2020.
- [11] R. Touati, "Détection de changement en imagerie satellitaire multimodale," Ph.D. dissertation, Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, Montreal, QC, Canada, 2019.

- [12] Y. Sun, L. Lei, D. Guan, J. Wu, and G. Kuang, "Iterative structure transformation and conditional random field based method for unsupervised multimodal change detection," *Pattern Recognit.*, vol. 131, Nov. 2022, Art. no. 108845. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031320322003260>
- [13] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using VHR optical and SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2403–2420, May 2010.
- [14] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 799–812, Mar. 2015.
- [15] R. Touati and M. Mignotte, "An energy-based model encoding nonlocal pairwise pixel interactions for multisensor change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 1046–1058, Feb. 2018.
- [16] Z. Lv et al., "Land cover change detection with heterogeneous remote sensing images: Review, progress, and perspective," *Proc. IEEE*, vol. 110, no. 12, pp. 1976–1991, Dec. 2022.
- [17] L. T. Luppino, F. M. Bianchi, G. Moser, and S. N. Anfinsen, "Unsupervised image regression for heterogeneous change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9960–9975, Dec. 2019.
- [18] L. T. Luppino et al., "Deep image translation with an affinity-based change prior for unsupervised multimodal change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4700422.
- [19] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 545–559, Mar. 2018.
- [20] L. Wan, Y. Xiang, and H. You, "An object-based hierarchical compound classification method for change detection in heterogeneous optical and SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9941–9959, Dec. 2019.
- [21] Y. Wu, Z. Bai, Q. Miao, W. Ma, Y. Yang, and M. Gong, "A classified adversarial network for multi-spectral remote sensing image change detection," *Remote Sens.*, vol. 12, no. 13, p. 2098, Jun. 2020.
- [22] Y. Sun, L. Lei, X. Li, H. Sun, and G. Kuang, "Nonlocal patch similarity based heterogeneous remote sensing change detection," *Pattern Recognit.*, vol. 109, Jan. 2021, Art. no. 107598.
- [23] M. Volpi, G. Camps-Valls, and D. Tuia, "Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis," *ISPRS J. Photogramm. Remote Sens.*, vol. 107, pp. 50–63, Sep. 2015.
- [24] Y. Sun, L. Lei, X. Li, X. Tan, and G. Kuang, "Structure consistency-based graph for unsupervised change detection with homogeneous and heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4700221.
- [25] J. Liu, W. Zhang, F. Liu, and L. Xiao, "A probabilistic model based on bipartite convolutional neural network for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4701514.
- [26] Y. Wu, J. Li, Y. Yuan, A. K. Qin, Q.-G. Miao, and M.-G. Gong, "Commonality autoencoder: Learning common features for change detection from heterogeneous images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4257–4270, Sep. 2022.
- [27] Z. Lv, H. Huang, L. Gao, J. A. Benediktsson, M. Zhao, and C. Shi, "Simple multiscale UNet for change detection with heterogeneous remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [28] P. Zhang, M. Gong, L. Su, J. Liu, and Z. Li, "Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 24–41, Jun. 2016.
- [29] Z. Liu, G. Li, G. Mercier, Y. He, and Q. Pan, "Change detection in heterogenous remote sensing images via homogeneous pixel transformation," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1822–1834, Apr. 2018.
- [30] Y. Sun, L. Lei, D. Guan, M. Li, and G. Kuang, "Sparse-constrained adaptive structure consistency-based unsupervised image regression for heterogeneous remote-sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4405814.
- [31] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 179, pp. 14–34, Sep. 2021.
- [32] Z.-G. Liu, Z.-W. Zhang, Q. Pan, and L.-B. Ning, "Unsupervised change detection from heterogeneous data based on image translation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4403413.
- [33] L. T. Luppino et al., "Code-aligned autoencoders for unsupervised change detection in multimodal remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, May 12, 2022, doi: [10.1109/TNNLS.2022.3172183](https://doi.org/10.1109/TNNLS.2022.3172183).
- [34] X. Niu, M. Gong, T. Zhan, and Y. Yang, "A conditional adversarial network for change detection in heterogeneous images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 45–49, Jan. 2019.
- [35] J. Wu et al., "A multiscale graph convolutional network for change detection in homogeneous and heterogeneous remote sensing images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, Dec. 2021, Art. no. 102615. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0303243421003226>
- [36] X. Jiang, G. Li, X.-P. Zhang, and Y. He, "A semisupervised Siamese network for efficient change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4700718.
- [37] C. Zhang et al., "A domain adaptation neural network for change detection with heterogeneous optical and SAR remote sensing images," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 109, May 2022, Art. no. 102769. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0303243422000952>
- [38] J. Liu, W. Xuan, Y. Gan, Y. Zhan, J. Liu, and B. Du, "An end-to-end supervised domain adaptation framework for cross-domain change detection," *Pattern Recognit.*, vol. 132, Dec. 2022, Art. no. 108960.
- [39] P. J. S. Vega et al., "An unsupervised domain adaptation approach for change detection and its application to deforestation mapping in tropical biomes," *ISPRS J. Photogramm. Remote Sens.*, vol. 181, pp. 113–128, Nov. 2021.
- [40] J. A. Agersborg, L. T. Luppino, S. N. Anfinsen, and J. U. Jepsen, "Towards targeted change detection with heterogeneous remote sensing images for forest mortality mapping," 2022, *arXiv:2203.00049*.
- [41] D. A. Jimenez-Sierra, D. A. Quintero-Olaya, J. C. Alvear-Muñoz, H. D. Benítez-Restrepo, J. F. Florez-Ospina, and J. Chanussot, "Graph learning based on signal smoothness representation for homogeneous and heterogeneous change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4410416.
- [42] D. A. Jimenez-Sierra, H. D. Benítez-Restrepo, H. D. Vargas-Cardona, and J. Chanussot, "Graph-based data fusion applied to: Change detection and biomass estimation in Rice crops," *Remote Sens.*, vol. 12, no. 17, p. 2683, Aug. 2020. [Online]. Available: <https://www.mdpi.com/2072-4292/12/17/2683>
- [43] Y. Sun, L. Lei, D. Guan, and G. Kuang, "Iterative robust graph for unsupervised change detection of heterogeneous remote sensing images," *IEEE Trans. Image Process.*, vol. 30, pp. 6277–6291, 2021.
- [44] H. Chen, N. Yokoya, C. Wu, and B. Du, "Unsupervised multimodal change detection based on structural relationship graph representation learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5635318.
- [45] M. Mignotte, "A fractal projection and Markovian segmentation-based approach for multimodal change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8046–8058, Nov. 2020.
- [46] Y. Sun, L. Lei, D. Guan, J. Wu, G. Kuang, and L. Liu, "Image regression with structure cycle consistency for heterogeneous change detection," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 29, 2022, doi: [10.1109/TNNLS.2022.3184414](https://doi.org/10.1109/TNNLS.2022.3184414).
- [47] H. Chen, F. He, and J. Liu, "Heterogeneous images change detection based on iterative joint global-local translation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9680–9698, 2022.
- [48] H. Noh, J. Ju, M. Seo, J. Park, and D.-G. Choi, "Unsupervised change detection based on image reconstruction loss," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 1351–1360.
- [49] X. Jiang, G. Li, Y. Liu, X.-P. Zhang, and Y. He, "Change detection in heterogeneous optical and SAR remote sensing images via deep homogeneous feature fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1551–1566, 2020.
- [50] Z. Ban, J. Liu, and L. Cao, "Superpixel segmentation using Gaussian mixture model," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4105–4117, Aug. 2018.

- [51] J. Wang, X. Yang, X. Yang, L. Jia, and S. Fang, "Unsupervised change detection between SAR images based on hypergraphs," *ISPRS J. Photogramm. Remote Sens.*, vol. 164, pp. 61–72, Jun. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271620301040>
- [52] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 19, 2007, pp. 1601–1608.
- [53] S. Agarwal, K. Branson, and S. Belongie, "Higher order learning with graphs," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 17–24.
- [54] S. Gao, I. W. Tsang, and L.-T. Chia, "Laplacian sparse coding, hypergraph Laplacian sparse coding, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 92–104, Jan. 2013.
- [55] X. Zhang, H. Su, C. Zhang, X. Gu, X. Tan, and P. M. Atkinson, "Robust unsupervised small area change detection from SAR imagery using deep learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 79–94, Mar. 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271621000046>
- [56] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
- [57] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Appl. Statist.*, vol. 28, no. 1, pp. 100–108, 1979.



Li Liu (Senior Member, IEEE) received the Ph.D. degree in information and communication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2012.

From 2015 to 2016, she spent ten months visiting the Multimedia Laboratory, The Chinese University of Hong Kong, Hong Kong. From 2016 to 2018, she worked as a Senior Researcher with the Machine Vision Group, University of Oulu, Oulu, Finland. She is currently a Full Professor with the College of System Engineering, NUDT. Her research interests include computer vision, pattern recognition, and machine learning.

Dr. Liu was the Cochair of nine International Workshops at CVPR, ICCV, and ECCV. She serves as the Area Chair of ICME'20, ICME'21, ICME'22, ACCV'20, and ACCV'22. She served as a Leading Guest Editor for special issues in the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (IEEE TPAMI) and the *International Journal of Computer Vision*. She is serving as a Leading Guest Editor for the IEEE TPAMI special issue on "Learning With Fewer Labels in Computer Vision." She currently serves as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, *Pattern Recognition*, and *Pattern Recognition Letters*.



Yuli Sun received the B.S. degree from Xiangtan University, Xiangtan, China, in 2009, and the M.S. degree from the University of Science and Technology of China, Hefei, China, in 2014. He is currently pursuing the Ph.D. degree in information and communication engineering with the College of Electronic Science, National University of Defense Technology.

He has published papers in the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, ISPRS P&RS, and *Pattern Recognition*. His research interests include remote sensing image processing and multitemporal image analysis.



Lin Lei received the Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2008.

She is currently a Professor with the School of Electronic Science, National University of Defense Technology. Her research interests include computer vision, remote sensing image interpretation, and data fusion.



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

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