SDC-GAE: Structural Difference Compensation Graph Autoencoder for Unsupervised Multimodal Change Detection

Te Han, Yuqi Tang, Yuzeng Chen, Xin Yang, Yuqiang Guo, Shujing Jiang

***Abstract*—Multimodal change detection (MCD) is a crucial technology for applications in natural resource monitoring, disaster assessment, and urban planning. To address the reliance on labeled data and enhance the robustness of structural features in existing methods, we propose a structure difference compensation graph autoencoder (SDC-GAE) for unsupervised MCD. It is recognized that the registered multimodal images exhibit consistency in structural features in unchanged areas, while the structural features in changed areas are distinct. SDC-GAE utilizes a graph convolutional network (GCN) to extract deep structural features from multimodal images. It uses the structural features of one time-phase image to reconstruct its spectral features in the spectral feature space of the target image. Through structural difference compensation, SDC-GAE learns the structural disparities between different images, with the compensation value directly reflecting the intensity of the changes. The SDC-GAE loss function consists of three components: image reconstruction loss, which evaluates the spectral feature discrepancy between the reconstructed and target images, guiding the model to reduce these differences via structural difference compensation; sparse constraint loss, which accounts for the fact that changes are typically confined to a few areas, ensuring the sparsity of the detected changes; and structural consistency loss, which aligns the structural features of the reconstructed image closely with those of the target image. The efficacy of our method is validated through experiments on eight multimodal datasets, where it is compared with state-of-the-art methods.**

***Index Terms*—Multimodal change detection, multi-source data, structural feature, structured graph, graph convolutional network, structural difference, compensation.**

I. Introduction

A. Background

Remote sensing image change detection (CD) is a technique that has garnered attention in fields such as natural resource monitoring [1], disaster assessment [2] and urban planning [3], [4]. This technique enables the detection and analysis of changes on Earth's surface by comparing remote sensing images of the same geographical area captured at different times [5]. The advent of various remote sensing satellites, including multispectral, hyperspectral and synthetic aperture radar (SAR) ones, has increased the variety and availability of remote sensing images, thereby advancing the development of CD. Based on the attributes of the remote sensing data, CD can be categorized into unimodal CD (UCD) and multimodal CD (MCD).

UCD primarily relies on data from a single type of satellite sensor. However, this approach often encounters challenges due to low data quality and potential data loss, influenced by factors such as satellite performance and environmental conditions. For instance, optical satellite images are prone to cloud cover and solar illumination issues.

In contrast, MCD offers several advantages over UCD:

1) It leverages the observational strengths of various sensors, allowing for a combination of data types. For instance, optical remote sensing satellites provide high-resolution surface information, while SAR satellites can observe under any weather or lighting conditions.
2) It adapts to complex spatial-temporal characteristics, as surface changes are influenced by numerous natural and human factors, exhibiting intricate spatial-temporal patterns. MCD can better accommodate these complexities by integrating information from different remote sensing data sources, enhancing the accuracy and robustness of CD.
3) It improves the temporal frequency and coverage of CD, as different satellites have distinct revisit cycles and coverage capabilities. By utilizing a diverse range of satellites, more frequent and extensive remote sensing data can be collected for continuous monitoring of surface changes.

However, the varying spatial, spectral, radiometric, and temporal resolutions of different remote sensing data pose challenges for applying traditional UCD methods. Consequently, there is an urgent need to develop specialized methods for MCD.

B. Related work

Multimodal imagery presents challenges due to imaging differences that cause the same surface feature to exhibit varying

Manuscript received **** This work was Supported by the National Natural Science Foundation of China (Grant 42271411); the Scientific Research Innovation Project for Graduate Students in Hunan Province (No. CX20220169); Research Project on Monitoring and Early Warning Technologies for Implementation of Land Use Planning in Guangzhou City (2020B0101130009); Collaborative Innovation Center for Natural Resources Planning and Marine Technology of Guangzhou (No. 2023B04J0326) *(Corresponding author: Yuqi Tang).*

Te Han, Yuqi Tang, Xin Yang and Yuqiang Guo are with the School of Geosciences and Info-Physics, Central South University, Changsha, 410083, China (e-mail: tehanrs@163.com; yqtang@csu.edu.cn; yang_x@csu.edu.cn; 235012186@csu.edu.cn)

Yuzeng Chen is with the School of Geodesy and Geomatics, Wuhan University, Wuhan, 430072, China. (yuzeng_chen@whu.edu.cn)

Shujing Jiang is with the Guangzhou Urban Planning & Design Survey Research Institute and Collaborative Innovation Center for Natural Resources Planning and Marine Technology of Guangzhou, Guangzhou, 510060, China. (jiangshujing128@126.com)

characteristics across different images, complicating direct comparison. To address this, researchers have developed methods to construct comparable features, enabling a uniform analysis of images from various sensors. These efforts have led to the development of four main types of MCD methods: post-classification comparison, similarity measurement, feature space mapping, and image space transformation.

1) Post-classification comparison-based methods. These methods involve selecting appropriate classification algorithms for individual processing and then comparing the results to detect changes. Representative methods include the kernel-based framework (KBF) [6], multitemporal segmentation and compound classification (MS-CC) [7], cooperative multitemporal segmentation and hierarchical compound classification (CMS-HCC) [8], and hierarchical extreme learning machine classification (HELMC) [9]. These methods are straightforward and can be tailored to specific applications. They also facilitate the understanding and interpretation of detection results by categorizing image data. Despite the advantages of post-classification comparison methods, several challenges remain: Firstly, classification errors inherent in these algorithms tend to accumulate during the comparison process, potentially diminishing the accuracy of CD. Secondly, these methods often necessitate extensive training datasets to effectively learn the characteristics of different feature types or changes. However, acquiring high-quality labeled data for MCD is particularly arduous, especially for intricate tasks. Thirdly, multimodal images present diverse data features. It is crucial to select classification algorithms or feature extraction methods that are compatible with these varying data types. However, ensuring the consistency of classification criteria across different methods is a laborious and complex endeavor that can undermine the methods' efficiency.

2) Similarity measurement-based methods. These methods posit a pattern correlation between multimodal images and leverage this correlation to construct invariant operators for measuring image similarity. For instance, the multidimensional statistical model (MSM) [10] employs statistical methods to model multimodal images, evaluating changes by comparing pixel-level statistical features. The multimodal change detection Markov model (M3CD) [11] identifies changed regions by establishing a Markov model to describe pixel relationships across different modalities. Other methods, such as energy-based model (EBM) [12], use energy distribution or difference metrics to assess similarity or changes. These approaches determine changes by comparing pixel value distributions or statistical information, bypassing the need for complex training processes. However, these methods may underutilize spatial information and are vulnerable to image noise. To utilize spatial information from imagery, sorted histogram (SH) [13] assesses pixel similarity by sorting and comparing image histograms, and [14] proposes a novel Bayesian statistical approach for MCD, involving a two-stage process that begins with preliminary estimation of spatially adaptive class conditional likelihoods specific to the imaging modality pair, followed by segmentation based on these likelihoods for each pixel and modality. The advantage of these methods is their ability to perform unsupervised CD without specific data type dependencies, offering flexibility across various change scenarios. Nonetheless, the reliance on hand-designed mode operators, which depend on prior knowledge and expert insights, limits their ability to capture the complex dependencies between multimodal data. This challenge persists in creating an invariant operator that accurately reflects the correlation between multimodal images.

3) Feature space mapping-based methods. These methods project multimodal remote sensing images into a shared feature space, ensuring that similar objects are represented similarly within this space. Techniques such as symmetric convolutional coupling network (SCCN) [15], approximately symmetrical deep neural network (ASDNN) [16], two-stage joint feature learning (TSJFL) [17], multicue contrastive self-supervised learning (MC-CSSL) [18], deep sparse residual model (DSRM) [19] , commonality autoencoder (CAE) [20] and log-based transformation feature learning (LTFL) [21]. For multiscale feature learning, methods like deep pyramid feature learning networks (DPFL-Nets) [22], deep homogeneous feature fusion (DHFF) [23], iterative joint global-local translation (IJGLT) [24]and Topological structure coupling network (TSCNet) [25], structural relationship graph convolutional autoencoder (SR-GCAE) [26] are proposed. These approaches enhance consistency across different modalities by mapping multimodal images into a unified feature space, making them adaptable to various types of remote sensing data. However, the varying noise levels and imaging features of multimodal images can lead to differences in the relationships between similar object features within the shared feature space.

4) Image space transformation-based methods. These methods establish image transformation models between multimodal images, enabling the transformation of multimodal images from their original image space to another image space. This means that the transformed images are closer to the original images in terms of imaging features, reducing the impact of modality differences on CD. For example, homogeneous pixel transformation (HPT) [27] and deep translation based change detection network (DTCDN) [28] construct spatial transformation relationships between multimodal images using label data. To enhance the autonomy of the algorithm, methods such as unsupervised image regression (UIR) [29] and coupled dictionary learning (CDL) [30] have been proposed. To utilize the structural information of image space, some graph-based methods have been introduced, such as patch similarity graph matrix (PSGM) [31], sparse-constrained adaptive structure consistency (SCASC) [32] and graph based image regression and Markov random field (GIR-MRF) [33]. Additionally, some scholars have used deep learning methods to achieve spatial transformation of multimodal images, such as generative adversarial networks under cutmix transformations (GANCT) [34], unsupervised change detection (USCD) [35], conditional adversarial network (CAN) [36], image translation network and post-processing (ITNPP) [37], hierarchical extreme learning machine image transformation (HELMIT) [38], code-aligned autoencoders (CAA) [39], and adversarial cyclic encoder network (ACE-Net) [40]. Through image transformation, the feature representation of one temporal image in the image space of another temporal image can be obtained, enhancing the diversity and richness of the image data before and after the change event, and providing more comprehensive change information. To further enhance the performance of such algorithms, it is

necessary to consider how to establish accurate multimodal image space transformation models.

C. Motivations and contributions

1) Advancements of unsupervised MCD Methods. Recent advancements in MCD have seen the introduction of several supervised learning methods that achieve remarkable results by training models with annotated ground truth changes. Notably, multitask change detection network (MTCDN) [41] and deep translation based change detection network (DTCDN) [28] have developed end-to-end image conversion frameworks utilizing UNet++ and generative adversarial networks (GAN), respectively. Hierarchical attention feature fusion (HAFF) [42] has enhanced CD capabilities through a hierarchical attention mechanism and feature fusion. Furthermore, domain adaptive cross reconstruction (DACR) [43] has facilitated domain adaptation between heterogeneous remote sensing images through feedback guidance. Leveraging spatial structural information, the dual neighborhood hypergraph neural network (DHGNN) [44] has introduced a novel network structure for high-resolution CD using a dual-neighbor hypergraph neural network. In this paper, we aim to propose an unsupervised MCD method to reduce reliance on label data. Unsupervised methods offer several advantages over supervised methods: a) They eliminate the need for training data, reducing the labor and subjectivity associated with manual data label, which is especially beneficial for large-scale datasets or when label data are scarce; b) They are versatile, capable of handling image data from various sensors, bands, or time points, making them adaptable to a wide range of CD tasks; c) Their lack of dependence on predefined change patterns or prior knowledge grants them robustness against unknown or complex changes.

2) Structural feature consistency in multimodal images. Despite the substantial differences in imaging features among multimodal images, they share consistent structural features in regions that have not changed [45], [46]. Detecting changes in multimodal images involves encoding these structural features into structural graphs and assessing the differences between them. Fractal projection and Markovian segmentation (FPMS) [47] leverages the spatial self-similarity of images. It projects patterns from one time phase to another using fractal encoding and employs pixel-level difference map binarization and Markov segmentation strategies within an unsupervised Bayesian framework to detect changes between multimodal images. Convolution model-based mapping (CMM) [48] captures local structural information through convolutional operations. Improved nonlocal patch-based graph (INLPG) [49] focuses on generating non-local structural features by considering the relationships between distant image patches, which are then quantified by mapping them into a common image domain for change measurement. On the other hand, Graph based fusion (GBF) [50] treats multimodal images as graph data, leveraging their intrinsic similarities to detect changes by fusing graph data and minimizing graph similarity. Graph learning based on signal smoothness representation (GLSSR) [51] integrates signal smoothness using graph structures to enhance detection accuracy. To further refine structural features, the iterative structure transformation and conditional random field (IST-CRF) [52] combines iterative optimization of structural transformations with conditional random field models for unsupervised CD. To reduce the influence of changed areas on image structural features, methods like enhanced graph structure representation (EGSR) [53], iterative robust graph and Markov co-segmentation (IRG-McS) [45] and adaptive optimization of structured graph (AOSG) [46] improve detection accuracy by iteratively optimizing graph structure. Structure graph-based methods offer several advantages over pixel, image patch, or superpixel-based methods in CD: a) They can overcome imaging differences in multimodal images by mining and comparing consistent structural features in unchanged regions, enhancing detection precision and robustness; b) The vertices in the structure graph represent image objects, and the edges between them reflect the similarity and correlation between these objects, providing valuable contextual information. This comprehensive consideration of vertex and edge attributes allows for more accurate differentiation between changed and unchanged areas; c) Encoding structural features into structure graphs can mitigate the effects of image noise on CD. However, these methods rely on traditional K-nearest neighbors (KNN) graphs to establish structural relationships, which typically consider only the direct proximity between pixels or objects, failing to capture more complex spatial structures and contextual information.

3) The potential of graph convolutional network (GCN) in extracting structural features of multimodal images. GCN [54] are capable of uncovering deeper structural features in images, which is particularly advantageous in MCD. Traditional methods, relying on pixel or superpixel analysis, often struggle to capture global structural information due to imaging disparities and a focus on local features. GCN address this limitation by acting as a robust tool for graph structure learning. They perform convolutional operations on graph structures, integrating not only the local information of nodes but also learning the intricate relationships between them. This process unveils the images' global structural features, aiding models in comprehending complex changes. Furthermore, GCN's high-dimensional feature representation enriches the contextual information of images, enabling a clearer distinction between actual changes and those falsely induced by imaging differences, thereby improving the reliability of detection.

Therefore, this paper proposes a structure difference compensation graph autoencoder (SDC-GAE) as an unsupervised method for MCD. The rationale behind this method is the assumption that, in the absence of changes, the structural features of registered multimodal images $\mathbf{X}$ and $\mathbf{Y}$ should align perfectly. By leveraging the structural features of image $\mathbf{X}$, we can reconstruct an image $\mathbf{Y}'$ in the domain of $\mathbf{Y}$, ensuring that the spectral features of $\mathbf{Y}$ and $\mathbf{Y}'$ are identical. However, changes in the imagery introduce structural discrepancies, resulting in spectral feature differences between $\mathbf{Y}$ and $\mathbf{Y}'$. To reconcile these differences, spectral difference compensation is employed, reflecting the intensity of the changes in the multimodal images. SDC-GAE constructs a graph model with superpixels as vertices and employs a GCN to extract deep structural features from the multimodal images. This process also involves learning the structural differences between images through structural difference compensation. The encoder component of SDC-GAE maps image $\mathbf{X}$ into a latent space,

utilizing its structural features and the spectral features of image $\mathbf{Y}$. The decoder then reconstructs image $\mathbf{Y}'$ from this latent space, striving for structural consistency with image $\mathbf{Y}$. To address spectral feature discrepancies due to changes, SDC-GAE incorporates a structural difference compensation to align the spectral features of $\mathbf{Y}'$ with those of $\mathbf{Y}$. The loss function of SDC-GAE is composed of three components: image reconstruction loss, which quantifies the spectral feature differences between $\mathbf{Y}'$ and $\mathbf{Y}$, guides the model to minimize these through structural difference compensation; sparse constraint loss, which is designed based on the fact that changes in images are typically confined to a few areas; and structural consistency loss, which ensures that $\mathbf{Y}'$ closely mirrors the structural features of $\mathbf{Y}$. The contributions of this paper are:

1） The proposed SDC-GAE for MCD has been developed to eliminate the need for additional supervision signals. SDC-GAE uses the structural features of imagery from one time phase to reconstruct the spectral spatial features in another image, establishing a spectral mapping relationship between the same objects in multimodal imagery, thus obtaining imagery of different modalities at the same time.

2） Unlike traditional methods that rely on shallow features, SDC-GAE extracts deep structural features from multimodal imagery, taking into account the complex spatial contextual relationships within the imagery.

3） SDC-GAE introduces a structural difference compensation mechanism, which optimizes the compensation value to make the reconstructed imagery structurally closer to the target imagery. The loss function design of SDC-GAE considers the characteristics of MCD, employing three types of loss functions to achieve the reconstruction of multimodal imagery, enabling the accurate identification of changed areas through structural difference compensation.

4） Validation of the proposed method's effectiveness through experiments on 8 datasets and comparisons with state-of-the-art methods.

II. Methodology

Given a pair of registered multimodal images $\mathbf{X} \in \mathbb{R}^{M\times N\times B_{\mathbf{X}}}$ and $\mathbf{Y} \in \mathbb{R}^{M\times N\times B_{\mathbf{Y}}}$, where M, N, and $B_{\mathbf{X}}\left(B_{\mathbf{Y}}\right)$ denote the length, width, and number of bands of image $\mathbf{X}\left(\mathbf{Y}\right)$, respectively, and the pixels are represented as $x\left(m,n,b_{\mathbf{X}}\right)$ and $y\left(m,n,b_{\mathbf{Y}}\right)$. Despite the significant imaging differences, the structural features in regions that have not changed remain consistent. This consistency enables the detection of changed areas by measuring the differences in the structural features between the multimodal images. As depicted in Fig. 1, squares and circles symbolize image objects, with the line thickness indicating the degree of similarity between them. Image $\mathbf{Y}'$ represents the spectral expression of image $\mathbf{X}$ within the image domain $\mathcal{Y}$. The structural features of multimodal images are manifested through the similarity relationships among these image objects. If images $\mathbf{X}$ and $\mathbf{Y}$ have not any changes, the structural features of the corresponding regions should be identical, meaning that the spectral features of image $\mathbf{Y}'$ can be represented by those of the same objects in image $\mathbf{Y}$. However, if changes are present in the multimodal images, their structural features will differ (as illustrated in Fig. 1 by the changed similarity relationships between objects in images $\mathbf{X}$ and $\mathbf{Y}$), resulting in the spectral features of image $\mathbf{Y}'$ in the changed areas being unable to be represented by the spectral features of the corresponding objects in image $\mathbf{Y}$. To address this, structural difference compensation can be employed to rectify structural discrepancies in areas experiencing changes. This approach refines the spectral properties of the reconstructed image $\mathbf{Y}'$ to correspond with those of image $\mathbf{Y}$. The compensation process adeptly detects changes in image intensity across different modalities, thereby enhancing the detection of changes.

The method proposed in this paper consists of three primary components (Fig. 2): structural graph construction, SDC-GAE learning, and change map (CM) generation. The following sections will detail these steps.

A. Structural Graph Construction

Structural graphs intuitively represent the structural features of images by using vertices and edges to illustrate the connections between them. In this study, we utilize structural graphs to depict the inherent structural features of images, with superpixels acting as graph vertices. Superpixels are pixel clusters within an image that share similar color and texture, and each superpixel is represented by a vertex in the graph. These vertices not only capture the local image features but also indicate the similarity between superpixels through their connections. This approach is more efficient than traditional methods that rely on image patches or individual pixels, as it better preserves the image's structural information, accurately captures regional boundaries, and minimizes fragmentation and noise issues associated with small processing units. Moreover, since superpixels consist of multiple pixels, the overall data processing volume is reduced.

To segment superpixels in images, we apply the simple linear iterative clustering (SLIC) [55], which delineates superpixel boundaries based on local color similarity and spatial continuity, yielding well-defined superpixels. To ensure that multimodal images have consistent superpixel boundaries for comparison, we overlay images $\mathbf{X}$ and $\mathbf{Y}$, and segment the combined image using SLIC to create a superpixel map $\mathrm{P}=\left\{\mathrm{P}_i \mid i=1,2,...,N_p\right\}$. This map is then mapped back to the original images $\mathbf{X}$ and $\mathbf{Y}$, yielding superpixel sets $\mathbf{X}=\left\{\mathbf{X}_i \mid i=1,2,...,N_p\right\}$ and $\mathbf{Y}=\left\{\mathbf{Y}_i \mid i=1,2,...,N_p\right\}$, with N_p indicating the total number of superpixels. To characterize each superpixel, we calculate the mean and median of its constituent pixels, providing insights into the color distribution. The mean indicates the average color trend, while the median is less affected by outliers and better represents the central tendency of the color spectrum. This process results in superpixel feature matrices $\tilde{\mathbf{X}} \in \mathbb{R}^{N_p\times 3B_{\mathbf{X}}}$ and $\tilde{\mathbf{Y}} \in \mathbb{R}^{N_p\times 3B_{\mathbf{Y}}}$ for images $\mathbf{X}$ and $\mathbf{Y}$, respectively.

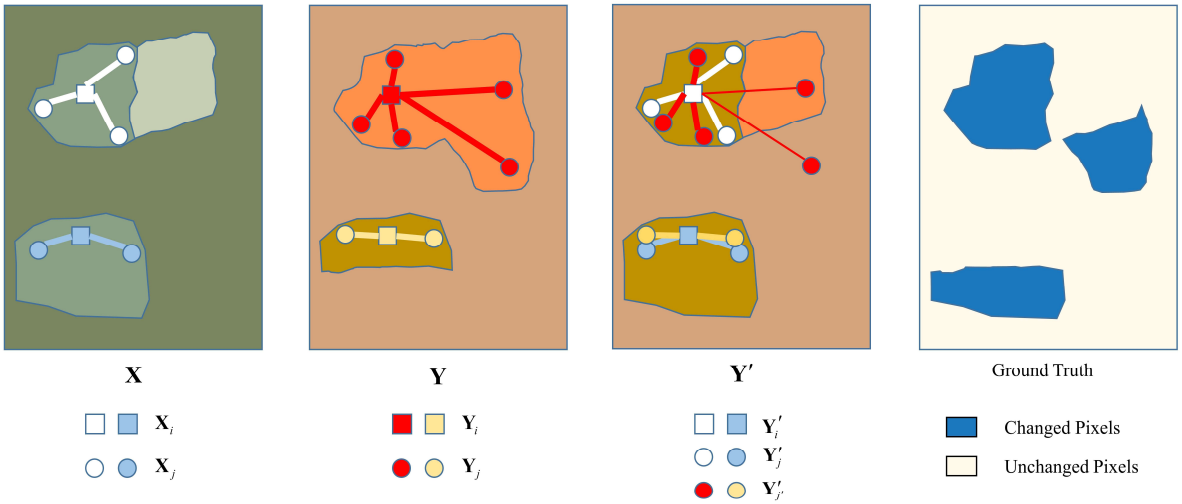


Fig. 1. Schematic diagram of structural difference compensation in multimodal images. Squares and circles represent image objects, with the thickness of the connecting lines indicating the strength of their similarity. This similarity reflects the structural characteristics of the images. Image $\mathbf{Y}'$ represents the spectral expression of image $\mathbf{X}$ in the image domain $\mathcal{Y}$. In the unchanged areas, the consistent structural features of image $\mathbf{X}$ and $\mathbf{Y}$ in that region allow the objects in image $\mathbf{Y}$ to be characterized by the spectral features of the same objects in image $\mathbf{X}$. However, in the changed areas, the structural features of images $\mathbf{X}$ and $\mathbf{Y}$ will differ, preventing image $\mathbf{Y}$ from being characterized by the spectral features of the corresponding objects in image $\mathbf{X}$. Therefore, structural difference compensation can be applied to the changed areas, making the reconstructed image $\mathbf{Y}'$ have the same spectral features as the original image $\mathbf{Y}$. This compensation value reflects the intensity of the changes between multimodal images.

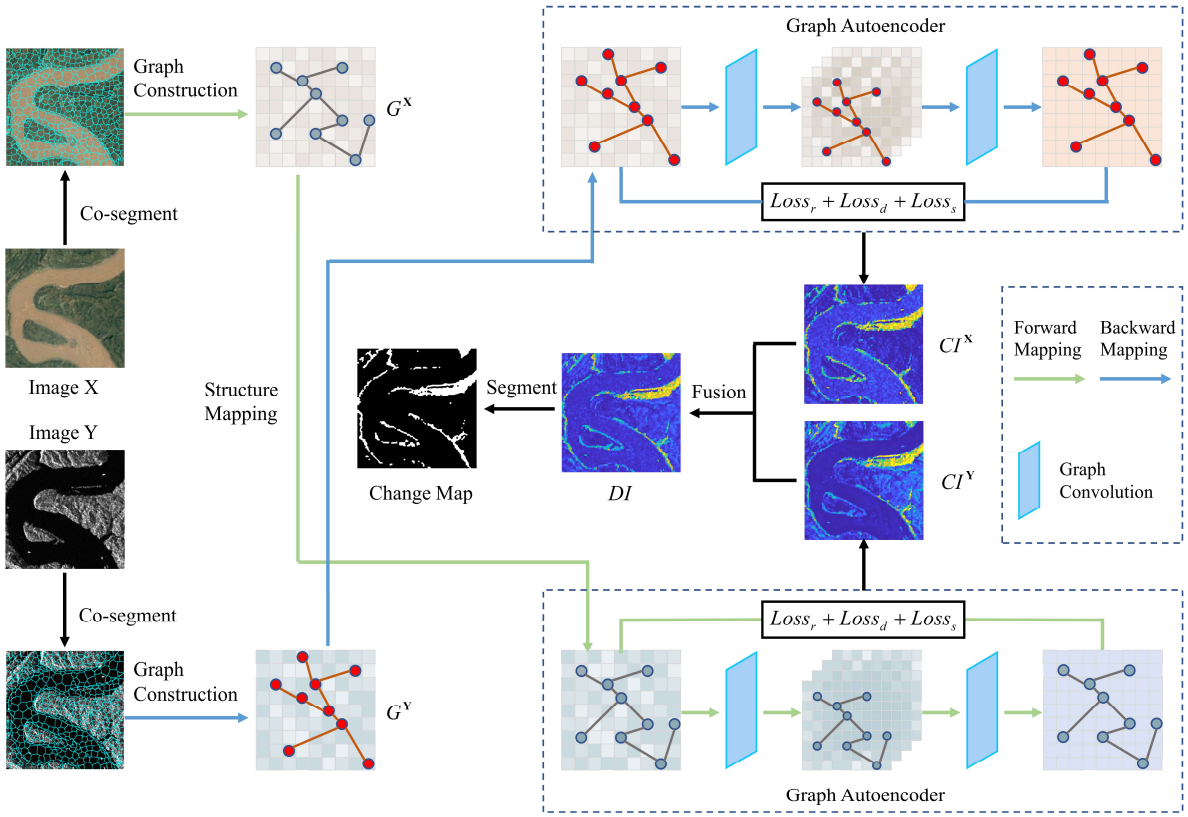


Fig. 2. Framework of SDC-GAE based MCD method. Following the superpixel segmentation of images $\mathbf{X}$ and $\mathbf{Y}$, an initial structural graph is constructed, where superpixels serve as vertices. Utilizing the structural features of image $\mathbf{X}(\mathbf{Y})$, the SDC-GAE models the spectral feature expressions of image $\mathbf{X}(\mathbf{Y})$ within image $\mathbf{Y}(\mathbf{X})$. The process yields a measure of the intensity of changes within the images through structural difference compensation.

Based on the obtained superpixels, we can construct a structural graph $G_{\mathbf{X}}$ for image $\mathbf{X}$ to represent its structural features. We define $G_{\mathbf{X}} = \{V_{\mathbf{X}}, E_{\mathbf{X}}\}$ as follows:

$$V_{\mathbf{X}} = \left\{ \mathbf{X}_i \mid i = 1,2,...,N_p \right\}$$
$$E_{\mathbf{X}} = \left\{ \left(\mathbf{X}_i, \mathbf{X}_j \right) \mid i = 1,2,...,N_p, j \in \Omega_{\mathbf{X}_i}, j \neq i \right\} \quad (1)$$

where, $V_{\mathbf{X}}$ and $E_{\mathbf{X}}$ represent the vertices and edges of the graph $G_{\mathbf{X}}$, respectively, while $\Omega_{\mathbf{X}_i}$ denotes the set of indices of vertices $\mathbf{X}_j$ connected to vertex $\mathbf{X}_i$. Similarly, we can construct the structural graph $G_{\mathbf{Y}}$ for image $\mathbf{Y}$.

To more conveniently describe the connection relationships between vertices in graph $G_{\mathbf{X}}$, we introduce the adjacency matrix $\mathbf{A}^{\mathbf{X}} \in \mathbb{R}^{N_p \times N_p}$ of graph $G_{\mathbf{X}}$:

$$\mathbf{A}_{i,j}^{\mathbf{X}} = \begin{cases} 1, & \left(\mathbf{X}_i, \mathbf{X}_j \right) \in E_{\mathbf{X}} \\ 0, & \left(\mathbf{X}_i, \mathbf{X}_j \right) \notin E_{\mathbf{X}} \end{cases} \quad (2)$$

This paper introduces the degree matrix $\mathbf{D}^{\mathbf{X}}$ for graph $G_{\mathbf{X}}$ to analyze the connection strength between vertices and to uncover the graph's clustering features and the significance of individual vertices. The degree matrix, a diagonal matrix where $\mathbf{D}_{i,i}^{\mathbf{X}} = \sum_j^{N_S} \mathbf{A}_{i,j}^{\mathbf{X}} \in \mathbb{R}^{N_S \times N_S}$ represents the frequency of vertex connections, serves as a direct mapping of each vertex's connectivity. In parallel, a structural graph $G_{\mathbf{Y}}$ is constructed for image $\mathbf{Y}$ to apply similar analysis.

The selection of an appropriate number of associated vertices K_i for vertex $\mathbf{X}_i$ is pivotal for accurately capturing the graph's structural features. A suboptimal K_i, either too small or too large, can lead to an inaccurate representation of the graph's structure. To determine K_i, we adopt a method similar to that described in [32], which involves setting a range for the number of neighbors with $k_{\max} = \lfloor K_{ratio} \times N_p \rfloor$ as the upper limit and $k_{\min} = \lfloor k_{\max} / 10 \rfloor$ as the lower limit, where $\lfloor \bullet \rfloor$ denotes the floor function and K_{ratio} is neighbor ratio. We calculate the feature distance $dist_{i,j}^{\mathbf{X}} = \left\| \mathbf{X}_i - \mathbf{X}_j \right\|_2^2$ between vertex $\mathbf{X}_i$ and other vertices to identify the $k_{\max}$ closest neighbors. The degree value $\mathbf{D}^{\mathbf{X}}$ of vertex $\mathbf{X}_i$ is then computed. Subsequently, K_i is set to the minimum of $\max\left\{ \mathbf{D}_{i,i}^{\mathbf{X}}, k_{\min} \right\}$ and $k_{\max}$ to ensure a balanced representation of the graph's structure.

B. SDC-GAE Learning

1) SDC-GAE: GCNs are specialized network models designed for processing graph data, adept at capturing and learning the structural characteristics inherent in graphs. Their strong interpretability and expressive capabilities have made them a staple in fields ranging from computer vision to recommendation systems. To precisely determine spectral difference compensation, which measure the changes intensity in multimodal images, the proposed SDC-GAE is composed of an encoder and a decoder. Both components are equipped with graph convolutional layers to facilitate efficient feature extraction and reconstruction, akin to traditional autoencoders.

The SDC-GAE reconstructs the original image to obtain its spectral feature representation in a different temporal image domain, while also calculating the structural difference compensation value dif. As established earlier, if multimodal images remain unchanged, their structural features should be consistent. Leveraging this consistency, we preserve the structural features of image $\mathbf{X}$ and reconstruct it within the image domain $\mathcal{Y}$, guided by the spectral features of image $\mathbf{Y}$. In the reconstructed image $\mathbf{Y}'$, the feature representation $h_{\mathbf{Y}_i'}^{(l+1)}$ of superpixel $\mathbf{Y}_i'$ at layer l+1 is refined through a weighted aggregation of the features of its l-layer neighboring vertices $h_{\mathbf{Y}_j'}^{(l)}$:

$$h_{\mathbf{Y}_i'}^{(l+1)} = \sigma\left(\sum\nolimits_{j \in \Omega_{\mathbf{X}_i}} \alpha_{i,j}^{\mathbf{X}(l)} W^{(l)} h_{\mathbf{Y}_j'}^{(l)} \right) \quad (3)$$

where $\sigma(\bullet)$ is the activation function, and $W^{(l)}$ represents the learnable weight matrix. $\alpha_{i,j}^{\mathbf{X}(l)}$ denotes the dynamically allocated contribution degree for each vertex, which reflects the importance of neighboring vertex in updating the current vertex. $j \in \Omega_{\mathbf{X}_i}$ denotes that in the superpixel $\mathbf{Y}_i'$ update process, the spatial index utilized for neighboring vertices corresponds to that of the adjacent vertices in graph $G_{\mathbf{X}}$. Consequently, this implies that the feature learning for reconstructing image $\mathbf{Y}'$ is performed using the structural features of image $\mathbf{X}$. It is worth noting that the input layer features of vertex $\tilde{\mathbf{Y}}_j'$ are $h_{\mathbf{Y}_j'}^{(0)} = \tilde{\mathbf{Y}}_j$.

To enhance the accuracy of establishing connection relationships between graph vertices and to improve the graph neural network's ability to model image structural features, this paper introduces the graph attention mechanism (GAM) [56], [57]. This mechanism is used to learn the contribution degree $\alpha_{i,j}^{\mathbf{X}(l)}$ for each vertex. Specifically, for each vertex $\mathbf{X}_i$ in graph $G_{\mathbf{X}} = \{V_{\mathbf{X}}, E_{\mathbf{X}}\}$, the attention weight $\alpha_{i,j}^{\mathbf{X}(l)}$ with its neighboring vertices can be calculated as follows:

$$\alpha_{i,j}^{\mathbf{X}(l)} = \frac{\exp\left(\mathrm{LeakyReLU}\left(a^{(l)^T} \left[W^{(l)} h_{\mathbf{Y}_i}^{(l)} \,\|\, W^{(l)} h_{\mathbf{Y}_j}^{(l)} \right] \right) \right)}{\sum_{k \in \Omega_{\mathbf{X}_i}} \exp\left(\mathrm{LeakyReLU}\left(a^{(l)^T} \left[W^{(l)} h_{\mathbf{Y}_i}^{(l)} \,\|\, W^{(l)} h_{\mathbf{Y}_k}^{(l)} \right] \right) \right)} \quad (4)$$

where $a^{(l)}$ represents a parameterized vector of a learnable attention mechanism, $\|$ denotes the concatenation of feature vectors, and $\mathrm{LeakyReLU}(\bullet)$ is the activation function.

Additionally, to achieve a more stable update process, we introduce multi-head attention. Assuming there are R attention heads for vertex feature updates, (3) can be rewritten as:

$$h_{\mathbf{Y}_i'}^{(l+1)} = \sigma\left(\frac{1}{R} \sum_{r=1}^{R} \sum_{j \in \Omega_{\mathbf{X}_i}} \alpha_{i,j}^{\mathbf{X}(l)^r} W^{(l)^r} h_{\mathbf{Y}_j'}^{(l)} \right) \quad (5)$$

where $\alpha_{i,j}^{\mathbf{X}(l)^r}$ represents the weight of the r-th ($r = 1,2,...,R$) attention mechanism a^r, while $W^{(l)^r}$ is the corresponding

learnable weight matrix.

SDC-GAE employs graph convolutional layers to progressively learn and reconstruct the image $\mathbf{Y}'$ of $\mathbf{X}$ within the image domain $\mathcal{Y}$. By leveraging the structural features of image $\mathbf{Y}$, we can similarly construct the reconstructed image $\mathbf{X}'$ within the image domain $\mathcal{X}$.

2) Loss Function: Influenced by changed areas, images $\mathbf{Y}'(\mathbf{X}')$ and $\mathbf{Y}(\mathbf{X})$ may exhibit differences, necessitating compensation with values $dif^{\mathbf{X}}$ and $dif^{\mathbf{Y}}$. These values, along with other SDC-GAE parameters, are optimized through the backpropagation algorithm [58] and gradient descent. To ensure that the reconstructed images $\mathbf{Y}'$ and $\mathbf{X}'$ closely resemble images $\mathbf{Y}$ and $\mathbf{X}$ in spectral features, respectively, we introduce a reconstruction loss function to achieve this objective.

$$loss_r = \sum_{i=1}^{N_p} \left\| \tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}'_i + dif_i^{\mathbf{X}} \right\|_2^2 + \sum_{i=1}^{N_p} \left\| \tilde{\mathbf{Y}}_i - \tilde{\mathbf{Y}}'_i + dif_i^{\mathbf{Y}} \right\|_2^2 \quad (6)$$

where $\tilde{\mathbf{Y}}' \in \mathbb{R}^{N_p \times 3B_{\mathbf{Y}}}$ represents the feature matrix of the reconstructed image $\mathbf{Y}'$.

The sparse constraint loss is engineered to minimize the structural difference compensation values $dif^{\mathbf{X}}$ and $dif^{\mathbf{Y}}$, acknowledging that changed regions are generally smaller compared to unchanged regions. Consequently, the majority of pixels are categorized as unchanged (assigned values of zero or close to zero), whereas a small subset of pixels is designated as changed (assigned a larger value values):

$$loss_d = \left\| dif^{\mathbf{X}} \right\|_2^2 + \left\| dif^{\mathbf{Y}} \right\|_2^2 \quad (7)$$

As shown in Fig. 1, the image $\mathbf{X}$ and its reconstructed image $\mathbf{Y}'$ share consistent structural features. Specifically, the interconnections and interactions between $\mathbf{Y}'_i$ and $\mathbf{Y}'_j$ should reflect the same pattern and intensity as those between $\mathbf{X}_i$ and $\mathbf{X}_j$, which means:

$$\min \sum_{i,j=1}^{N_p} \left\| \tilde{\mathbf{Y}}'_i - \tilde{\mathbf{Y}}'_j \right\|_2^2 \mathbf{A}_{i,j}^{\mathbf{X}} = 2Tr\left(\tilde{\mathbf{Y}}'^T \mathbf{L}^{\mathbf{X}} \tilde{\mathbf{Y}}' \right) \quad (8)$$

where $\mathbf{L}^{\mathbf{X}} = \mathbf{D}^{\mathbf{X}} - \mathbf{A}^{\mathbf{X}}$ is the Laplacian matrix of the graph $G_{\mathbf{X}}$.

Additionally, we have performed normalization on the Laplacian matrix, which offers three significant benefits. First, normalization stabilizes the degree values of vertices, essential for maintaining nodal force equilibrium during analysis. This step also minimizes computational errors, boosting the reliability and accuracy of numerical calculations. Moreover, the normalized matrix adapts to different network structures, making the algorithm versatile for a variety of scenarios. Thus, we can obtain the structural consistency loss:

$$loss_s^{\mathbf{X}} = 2Tr\left(\tilde{\mathbf{Y}}'^T \tilde{\mathbf{L}}^{\mathbf{X}} \tilde{\mathbf{Y}}' \right) \quad (9)$$

where $\tilde{\mathbf{L}}^{\mathbf{X}} = \mathbf{I} - \left(\mathbf{D}^{\mathbf{X}}\right)^{-1/2} \mathbf{A}^{\mathbf{X}} \left(\mathbf{D}^{\mathbf{X}}\right)^{-1/2}$ represents the normalized Laplacian matrix, while $\mathbf{I} \in \mathbb{R}^{N_p \times N_p}$ represents the identity matrix. Similarly, for the reconstructed image $\mathbf{X}'$, we can obtain:

$$loss_s^{\mathbf{Y}} = 2Tr\left(\tilde{\mathbf{X}}'^T \tilde{\mathbf{L}}^{\mathbf{Y}} \tilde{\mathbf{X}}' \right) \quad (10)$$

where $\tilde{\mathbf{X}}' \in \mathbb{R}^{N_p \times 3B_{\mathbf{X}}}$ represents the feature matrix of the reconstructed image $\mathbf{X}'$, and $\tilde{\mathbf{L}}^{\mathbf{Y}} = \mathbf{I} - \left(\mathbf{D}^{\mathbf{Y}}\right)^{-1/2} \mathbf{A}^{\mathbf{Y}} \left(\mathbf{D}^{\mathbf{Y}}\right)^{-1/2}$ represents the normalized Laplacian matrix of graph $G_{\mathbf{Y}}$. The overall structural consistency loss function is then given by:

$$loss_s = loss_s^{\mathbf{X}} + loss_s^{\mathbf{Y}} \quad (11)$$

Then the overall loss function for SDC-GAE is:

$$loss = loss_r + loss_d + loss_s \quad (12)$$

C. CM generation

Structural difference compensation $dif^{\mathbf{X}}$ and $dif^{\mathbf{Y}}$ reflect the change intensity of image, thus the final change intensity is a combination of both:

$$dif^{final} = dif^{\mathbf{X}} / mean\left(dif^{\mathbf{X}}\right) + dif^{\mathbf{Y}} / mean\left(dif^{\mathbf{Y}}\right) \quad (13)$$

The change intensity map (CIM) is obtained as follows:

$$CI_{(m,n)} = dif_i^{final} \;;\; (m,n) \in P_i \quad (14)$$

The generation of the final CM can be visualized as a binary segmentation problem, and we obtain the CM using Otsu's threshold segmentation [59]. The summary of the proposed SDC-GAE is presented in Table I.

TABLE I.
FRAMEWORK OF SDC-GAE

SDC-GAE
Input: images $\mathbf{X}$ and $\mathbf{Y}$, parameters of SDC-GAE **1. Structure Graph Construction:** Image superpixel segmentation and superpixel feature extraction; Calculation of similarity between graph vertices; Determination of the number of neighboring vertices for a graph vertex; **2. SDC-GAE Learning:** Construct the structure of SDC-GAE; Construct the loss function; Update vertex features according to (3); **3. CM Generation:** Compute the binary CM with (13), (14) and Otsu's threshold segmentation.

III. EXPERIMENTS

In this section, we will provide a detailed introduction to the experimental setup, including comparative methods, evaluation metrics, experimental parameters, etc., following with the experimental datasets and the results of the experiments.

A. Comparative Methods and Evaluation Metrics

To validate the effectiveness of the methods presented in this paper, we employ several state-of-the-art methods as comparative methods:

1) LTFL[21] utilizes a stacked denoising autoencoder to extract deep features from multimodal images and applies difference metrics to these features. High-confidence samples are selected for training a classifier, which performs binary classification on feature difference maps to generate CM.
2) INLPG [49] constructs non-local structural features for multimodal images by considering non-local correlations between image patches. These features are mapped into a

TABLE II
DESCRIPTION OF THE DATASETS.

Dataset	Sensor	Size(pixels)	Date	Location	Event (& Spatial resolution)
#1	Google Earth/Sentinel-1	600×600×3(1)	Dec. 1999-Nov. 2017	Chongqing, China	River expansion (10 m)
#2	Sentinel-2/ Sentinel-1	444×571×3(1)	Apr. 2017 - Oct. 2020	Lake Poyang, China	Lake expansion (10 m)
#3	Pleiades/WorldView2	2000 × 2000 × 3(3)	May 2012 – July 2013	Toulouse, France	Urban construction (0.52m)
#4	QuickBird/LiDAR	700×700×4(1)	Nov. 2007 – June 2011	San Francisco, USA	Urban construction (0.5m)
#5	QuickBird-2/TerraSAR-X	4135×2325×3(1)	July 2006 – July 2007	Gloucester, England	Flooding (≈0.65m)
#6	Spot/NDVI	990 × 554 × 3(1)	1999 - 2000	Gloucester, England	Flooding (≈25m)
#7	Landsat-8/Sentinel-1	3500×2000×11(3)	Jan. 2017 – Feb. 2017	Sutter County, USA	Flooding (≈15m)
#8	Landsat-5/Google Earth	300 × 412 × 1(3)	Sept. 1995 - July 1996	Sardinia, Italy	Lake expansion (30m.)

consistent image domain for comparison, enabling the quantification of changes within multimodal images.

3) GBF [50] treats multimodal images as graph data, capturing their intrinsic similarities. It fuses multi-temporal graph data and minimizes graph similarity to detect changes in multimodal images.
4) IRG-McS [45] develops an adaptive non-local structural graph based on superpixels to represent image structure features. It refines the graph structure of unchanged regions using structural difference metrics and Markov co-segmentation, enhancing the accuracy of CD.
5) SCASC [32] retains the structural features of the source image and applies sparse constraints to transform these features into the target image domain. Change information is extracted through a comparison between the original and transformed images.
6) GIR-MRF [33] employs an unsupervised image regression method that integrates global and local graph structure learning to capture image features and relationships. It uses a Markov segmentation model to segment difference maps, resulting in CM.
7) SRGCAE [26] leverages a GCN to learn graph structure relationships in multimodal images, expressing their structural features. It extracts change information by comparing these structural features.
8) AOSG [46] constructs an adaptive structured graph by combining the spatial scale of image objects with the feature distance between image patches. It optimizes the graph by considering the change attributes of neighboring regions, thereby improving the precision of structural feature representation.

In order to evaluate the effectiveness of various methods quantitatively, we have employed a suite of evaluation metrics, including overall accuracy (OA), kappa coefficient (KC), and F1-measure (F1). These metrics provide a comprehensive assessment of model performance. Below are detailed descriptions of these evaluation metrics:

The OA serves as a key indicator when assessing the efficacy of a classification model. It measures the model's ability to correctly classify samples, expressed as a ratio of the correctly classified instances to the total sample count N. The OA is calculated as follows:

$$\mathrm{OA}=\left(\mathrm{TP}+\mathrm{TN}\right)/\left(\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}\right) \tag{15}$$

where, TP, FP, TN, FN denote true positives, false positives, true negatives, and false negatives respectively.

The KC is used to measure the degree of agreement among data, and it is calculated as follows:

$$\mathrm{KC}=\left(\mathrm{OA}-\mathrm{PRE}\right)/\left(1-\mathrm{PRE}\right) \tag{16}$$

where,

$$\mathrm{PRE}=\left(\left(\mathrm{TP}+\mathrm{FN}\right)\left(\mathrm{TP}+\mathrm{FP}\right)+\left(\mathrm{TN}+\mathrm{FP}\right)\left(\mathrm{TN}+\mathrm{FN}\right)\right)/\mathrm{N}^2 \tag{17}$$

The F1 Score is the harmonic mean of Precision and Recall, used to measure the comprehensive performance of a classification model on a specific category, especially in cases of class imbalance. It is calculated as follows:

$$\mathrm{F1}=\frac{2\times Precision\times Recall}{Precision+Recall} \tag{18}$$

where,

$$Precision=\mathrm{TP}/\left(\mathrm{TP}+\mathrm{FP}\right) \tag{19}$$

$$Recall=\mathrm{TP}/\left(\mathrm{TP}+\mathrm{FN}\right) \tag{20}$$

In this paper, the proposed SDC-GAE employs the Adam optimizer [60] with a learning rate set to 0.01, weight decay to 0.0001, and the number of attention heads R is set to 4. Both layers of the graph convolutional layers use convolutional kernels of 16. The epochs are set to 300. In addition, the main parameters of SDC-GAE, the neighbor ratio k_{ratio} and the number of superpixels N_p, are set to 0.1 and 5000, respectively, and these parameters will be analyzed in section IV. All experiments of SDC-GAE are conducted in the following environment: Pytorch, CPU is AMD Ryzen 7 3800X 8-Core Processor, 3.89 GHz, Windows 11, 32GB RAM, and an NVIDIA GeForce GTX 2070 SUPER.

B. Experimental Datasets

This paper validates the effectiveness of the proposed multimodal image CD methods using eight diverse remote sensing datasets, as depicted in Fig. 3(a) and (b). The datasets include a range of sensor images from optical satellites like Sentinel-1, QuickBird, and Landsat-8, as well as SAR satellites, capturing both short-term and long-term changes. This variety ensures a comprehensive assessment of the methods' performance across different sensors and time frames. Spanning various geographical regions, the datasets feature changes at multiple scales, such as urban development and river expansion, which are essential for testing the algorithms' applicability and generalization capabilities. The reference images, derived from expert knowledge and high-resolution, temporally and spatially proximate images, provide a reliable benchmark for validation. The selection of these datasets aims to rigorously evaluate the

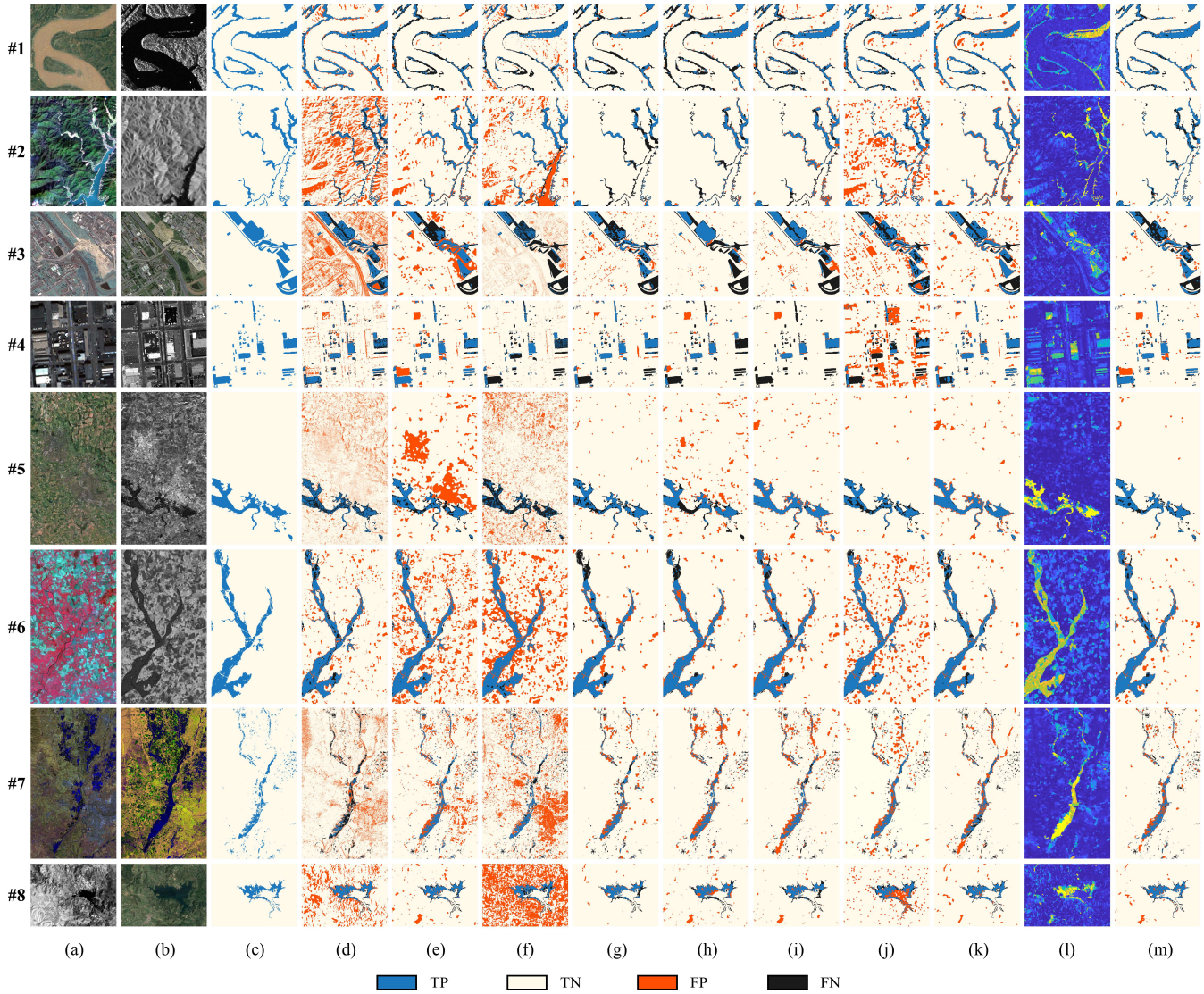


Fig. 3. CMs of different methods and the CIMs for SDC-GAE. From top to bottom, they correspond to datasets #1 to #8. From left to right, they are as follows: (a) image **X** ; (b) image **Y** ; (c) reference image; (d) CMs of LTFL; (e) CMs of INLPG; (f) CMs of GBF; (g) CMs of IRG-McS; (h) CMs of SCASC; (i) CMs of GIR-MRF; (j) CMs of SRGCAE; (k) CMs of AOSG; (l) CIMs of SDC-GAE; (m) CMs of SDC-GAE.

proposed methods' effectiveness in diverse environmental contexts. All datasets underwent preprocessing such as radiometric correction, atmospheric correction, and geometric correction, with each dataset's images resampled to the same spatial resolution. For further details, refer to Table II.

C. Experimental Results

Fig. 3 illustrates the CMs for different methods applied to datasets #1 through #8, alongside the CIMs of SDC-GAE. Datasets #1 and #2, which depict river changes, present a challenge due to “pseudo-changes” resulting from varying shadow distributions over land. Visual inspection reveals that all methods successfully identified the primary change areas in dataset #1. However, GBF and SRGCAE exhibited notable FPs, while INLPG, IRG-McS, SCASC, and GIR-MRF, though having fewer FPs, missed several detections. AOSG managed to detect more comprehensive change areas but encountered isolated FPs in the image's center. In dataset #2, LTFL, INLPG, GBF, and SRGCAE showed significant FPs, whereas IRG-McS, SCASC, GIR-MRF, and AOSG had fewer FPs but more missed detections. Dataset #3, with its complex change scenario involving bare land, grassland, buildings, and roads, saw most methods, except SCASC, producing many false positives. SCASC, however, failed to identify the change area in the lower right corner of the image. Dataset #4, which reflects changes in buildings and vehicles, saw LTFL and SRGCAE generating more FPs than other methods. Meanwhile, INLPG, IRG-McS, SCASC, and GIR-MRF missed three distinct building change areas. AOSG achieved a balance between FPs and FNs but still overlooked a building change in the lower left corner. Datasets #5 to #7, all showing river changes, faced challenges due to extensive image coverage and complex object textures. In dataset #5, GBF's performance was compromised by numerous false and FNs. LTFL and INLPG had many FPs, yet other methods

TABLE III
ACCURACY EVALUATIONS ON DATASETS #1- #4. THE MAXIMUM AND SECOND MAXIMUM VALUES ARE REPRESENTED BY BOLD AND UNDERLINED TEXT, RESPECTIVELY.

Methods	#1			#2			#3			#4		
	OA	KC	F1	OA	KC	F1	OA	KC	F1	OA	KC	F1
LTFL	0.9204	0.7119	0.7579	0.7016	0.1490	0.2600	0.6800	0.2181	0.3862	0.9109	0.5367	0.5862
INLPG	0.9083	0.5635	0.6092	0.9148	0.5426	0.5879	0.8171	0.3395	0.4482	0.9279	0.6445	0.6843
GBF	0.8989	0.5530	0.6109	0.8300	0.3458	0.4231	0.8261	0.2155	0.3105	0.9202	0.4167	0.4544
IRG-McS	0.9128	0.6022	0.6483	0.9450	0.4948	0.5182	0.8685	0.4239	0.4973	0.9387	0.6220	0.6552
SCASC	0.8955	0.5069	0.5599	0.9537	0.6443	0.6684	0.8918	0.4711	0.5247	0.9148	0.3929	0.4345
SRGCAE	0.9223	0.6812	0.7259	0.8354	0.3585	0.4339	0.8231	0.3817	0.4867	0.7708	0.0821	0.2005
GIR-MRF	0.9037	0.5913	0.6457	0.9481	0.6398	0.6678	**0.8960**	0.4840	0.5350	0.9300	0.4957	0.5294
AOSG	0.9270	0.7291	0.7725	0.9465	0.7061	**0.7347**	0.8745	**0.5131**	**0.5871**	0.9445	0.6870	0.7177
SDC-GAE	**0.9429**	**0.7590**	**0.7914**	**0.9627**	**0.7067**	0.7259	0.8860	0.4932	0.5561	**0.9615**	**0.6978**	**0.7178**

TABLE IV
ACCURACY EVALUATIONS ON DATASETS #5- #8. THE MAXIMUM AND SECOND MAXIMUM VALUES ARE REPRESENTED BY BOLD AND UNDERLINED TEXT, RESPECTIVELY.

Methods	#5			#6			#7			#8		
	OA	KC	F1	OA	KC	F1	OA	KC	F1	OA	KC	F1
LTFL	0.8658	0.2879	0.3509	0.9195	0.6689	0.7145	0.8504	0.0631	0.1240	0.8078	0.2849	0.3549
INLPG	0.8316	0.2615	0.3321	0.7829	0.4134	0.5162	0.9059	0.3730	0.4128	0.9503	0.6128	0.6392
GBF	0.8284	0.1032	0.1832	0.8272	0.4827	0.5695	0.7915	0.1093	0.1737	0.4667	0.0344	0.1418
IRG-McS	0.9672	0.7029	0.7201	0.9358	0.7136	0.7502	0.9469	0.4703	0.4975	**0.9700**	**0.7242**	**0.7401**
SCASC	0.9295	0.5010	0.5381	0.9503	0.7762	0.8046	0.9381	0.4585	0.4888	0.9472	0.5958	0.6238
SRGCAE	0.9707	0.7184	0.7335	0.8728	0.5724	0.6400	0.9376	0.4246	0.4557	0.9129	0.4737	0.5164
GIR-MRF	0.9639	0.7524	0.7713	0.9332	0.7249	0.7627	0.9446	0.4674	0.4954	0.9587	0.6332	0.6551
AOSG	0.9644	0.7587	0.7773	**0.9553**	0.7788	0.8040	0.9450	0.4259	0.4544	0.9605	0.6733	0.6944
SDC-GAE	**0.9774**	**0.7945**	**0.8063**	0.9530	**0.7902**	**0.8169**	**0.9525**	**0.5134**	**0.5378**	0.9645	0.7185	0.7374

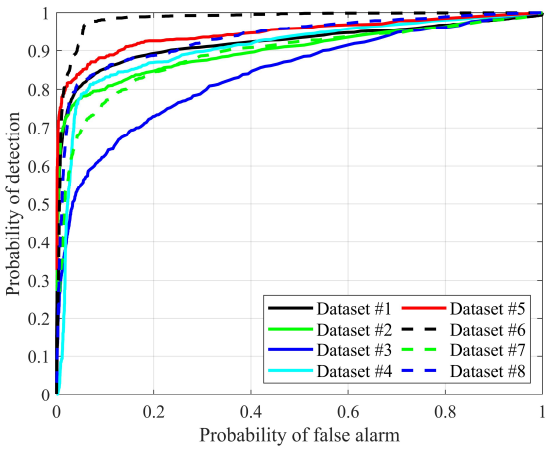


Fig. 4. ROC curves of CIMs generated by SDC-GAE across all datasets.

managed to detect change areas more thoroughly, albeit with some FPs. Dataset #6 saw INLPG, GBF, and SRGCAE with significant FPs, while other methods detected change areas more completely with minimal FPs. In dataset #7, LTFL, INLPG, and GBF had more FPs, but other methods detected the main change areas with relatively fewer FPs. Finally, dataset #8, which captured subtle lake changes, found LTFL and GBF with many FPs in land areas, and SRGCAE incorrectly identified the lake's interior as a change area. In contrast, other methods detected a more accurate change area. In summary, the proposed SDC-GAE method not only detected more complete change areas but also had the fewest FPs. This is attributed to its capability to extract high-level image features via a graph encoder and accurately pinpoint change areas through structural difference compensation.

Fig. 3(l) showcases the CIMs produced by the proposed SDC-GAE across various datasets. To evaluate the performance of these CIMs, we utilize the area under the receiver operating characteristics (ROC) curve (AUC) as a metric. As depicted in Fig. 4, the ROC curves for SDC-GAE on datasets #1 through #8 yield AUC values of 0.9207, 0.9018, 0.8383, 0.9070, 0.9460, 0.9835, 0.8899, and 0.9331, respectively. These results indicate that SDC-GAE is adept at generating high-quality CIMs. Furthermore, these maps can be efficiently converted into CMs using straightforward threshold segmentation techniques.

The accuracy assessment results for various methods on datasets #1-#4 and #5-#8 are detailed in Tables III and IV, respectively. In these Tables, the boldface denotes the maximum values, while underlines highlight the second maximum values. With the exception of dataset #3, the proposed SDC-GAE consistently ranks within the top two for OA across the remaining datasets. Moreover, SDC-GAE secures a position in the top two for both KC and F1 scores in all datasets, achieving the highest KC in six out of the eight datasets. These results underscore the

TABLE V
THE EFFECTIVENESS of GAM

	AT	#1	#2	#3	#4	#5	#6	#7	#8
OA	×	0.9379	0.9378	0.8786	0.9392	0.9397	0.9627	0.8722	0.9591
	√	0.9429	0.9627	0.8860	0.9615	0.9774	0.9530	0.9525	0.9645
KC	×	0.7263	0.6703	0.4748	0.6719	0.7404	0.7067	0.4513	0.6661
	√	0.7590	0.7067	0.4932	0.6978	0.7945	0.7902	0.5134	0.7185
F1	×	0.7604	0.7049	0.5432	0.7057	0.7741	0.7259	0.5236	0.6879
	√	0.7914	0.7259	0.5561	0.7178	0.8063	0.8169	0.5378	0.7374

effectiveness and robustness of the SDC-GAE.

IV. DISCUSSIONS

A. Parameter Analysis

1) Neighbor ratio K_{ratio}. This section examines the sensitivity of the proposed SDC-GAE to the parameter K_{ratio}. To do this, we maintained N_P at 5000 and varied K from 0.02 to 0.20 in increments of 0.02. The resulting accuracy changes for SDC-GAE are depicted in Fig. 5(a). Analyzing Fig. 5(a) reveals that SDC-GAE's accuracy generally ascends with K_{ratio} values from 0.02 to 0.08, peaks and stabilizes when K_{ratio} is between 0.08 and 0.12, and subsequently declines for K_{ratio} values from 0.12 to 0.20. This trend underscores the impact of the number of neighbors on the graph's structural representation quality. The rationale is twofold: Firstly, a vertex with an excessive number of neighbors may be inundated with noisy information, compromising the graph's accuracy. On the other hand, a vertex with insufficient neighbors might lack the context necessary for precise feature representation. Secondly, graph neural networks are prone to oversmoothing, where an abundance of neighbors can lead to each vertex being overly influenced by similar ones during information aggregation, thus diminishing the distinctiveness of vertex features. Given these considerations, we suggest an optimal K_{ratio} value of 0.1 in this paper.

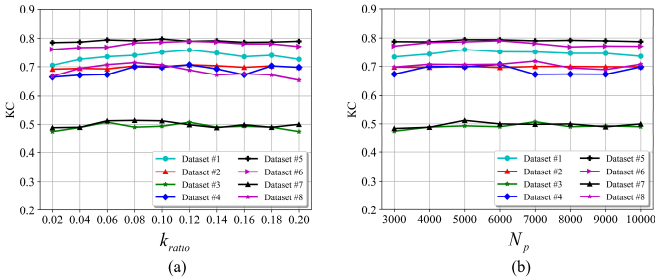


Fig. 5. Sensitivity analysis of parameters in SDC-GAE: (a) k_{ratio} -KC curves. (b) N_P -KC curves.

2) Number of Superpixels N_P. To investigate the effect of the N_P on the proposed SDC-GAE, we varied N_P from 3,000 to 10,000 in increments of 1,000, with the neighbor ratio K_{ratio} held constant at 0.1. Fig. 5(b) illustrates the accuracy of SDC-GAE across this range of N_P values. The KC for SDC-GAE initially increases as N_P rises from 3,000 to 4,000, stabilizes between 4,000 and 6,000, and then exhibits a slight decline for N_P values from 6,000 to 10,000. This is because the number of superpixels determines the spatial resolution of the segmented image. Too few superpixels may fail to capture subtle changes, while too many can lead to over-segmentation, making it difficult to distinguish real changes from noise in CD. However, overall, the impact on the proposed SDC-GAE is not significant. Moreover, an excessive number of superpixels will increase the algorithm's running time. Therefore, considering both the accuracy and efficiency of the algorithm, we recommend setting N_P to 5000.

B. Similarity Between the Reconstructed Image and the Target Image.

Fig. 6 displays the reconstructed images generated by SDC-GAE across datasets #1 to #8. It can be observed from Fig. 6 that the reconstructed image $\mathbf{X}'(\mathbf{Y}')$ shares similar spectral characteristics with image $\mathbf{X}(\mathbf{Y})$ and structural features with image $\mathbf{Y}(\mathbf{X})$. To further illustrate the differences in spectral characteristics before and after image reconstruction, we selected unaltered regions within each dataset (indicated by the red boxes) and calculated the histogram distribution of each band of the images (red, green, and blue bands are labeled with numbers (1), (2), and (3), respectively, in Fig. 7). It is evident that there are significant differences in the shape, peak position, and distribution range of the probability density function (PDF) curves between the original images $\mathbf{X}$ and $\mathbf{Y}$. In contrast, the PDF curves of the reconstructed image $\mathbf{X}'(\mathbf{Y}')$ and image $\mathbf{X}(\mathbf{Y})$ are much closer. This indicates that SDC-GAE can encode image $\mathbf{X}(\mathbf{Y})$ into the spectral space of image $\mathbf{Y}(\mathbf{X})$ while preserving the structural features of image $\mathbf{X}(\mathbf{Y})$.

C. The Effectiveness of GAM

The GAM allows SDC-GAE to dynamically adjust the weights of graph vertices during processing, prioritizing those vertices deemed more important. This adaptive approach facilitates the learning of vertex representations. To assess the impact of GAM, we compared the accuracy of SDC-GAE both with and without GAM (Table V). The results, as presented in Table V, show that SDC-GAE achieved an average enhancement of 2.16% in OA, 7.30% in KC, and 4.71% in F1 when GAM was employed. These improvements underscore the

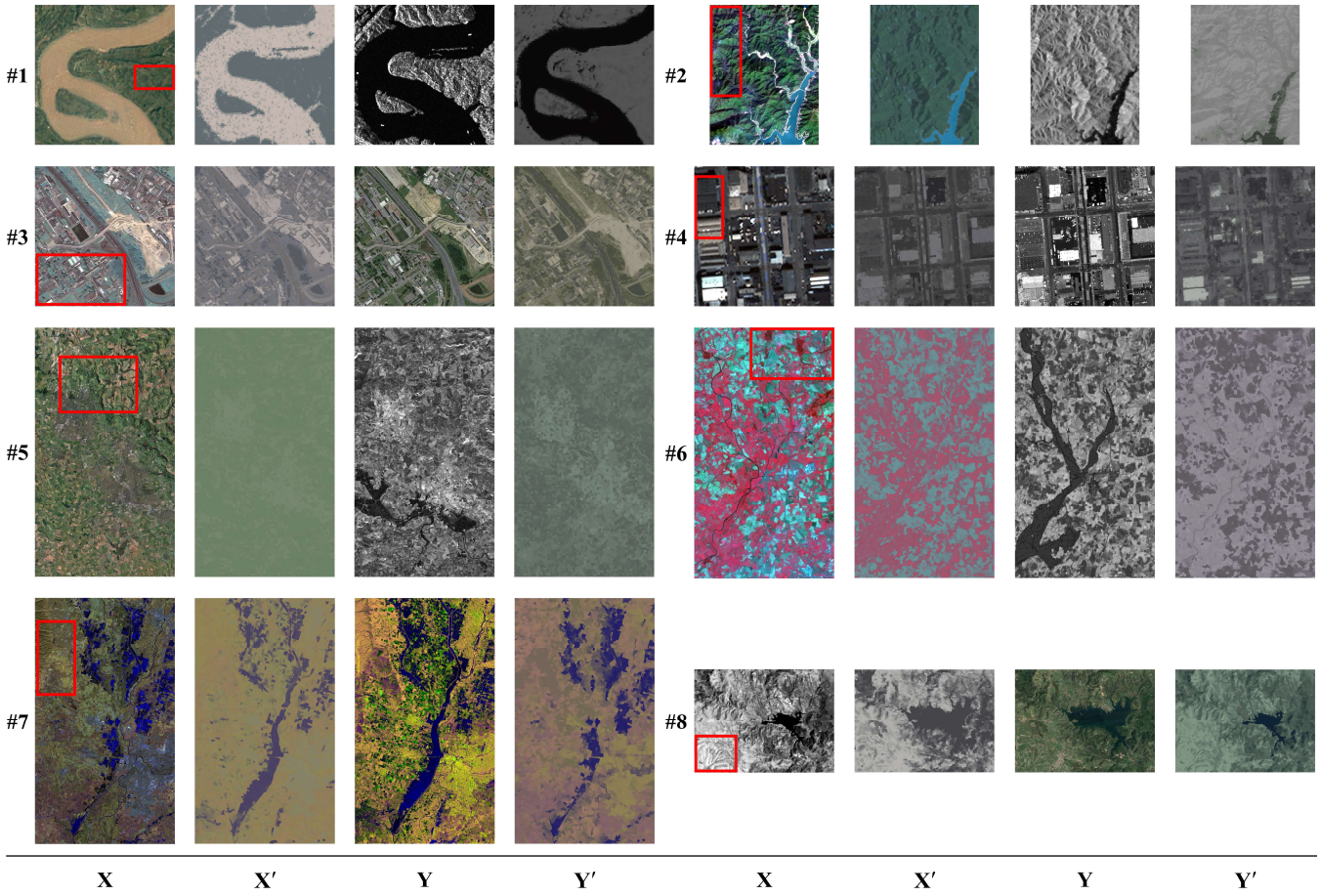


Fig. 6. The reconstructed images of SDC-GAE on datasets #1 to #8 are presented. Within each dataset, from left to right, they are as follows: image $\mathbf{X}$, the reconstructed image $\mathbf{X}'$ of image $\mathbf{Y}$ in the domain of $\mathcal{X}$, image $\mathbf{Y}$, and the reconstructed image $\mathbf{Y}'$ of image $\mathbf{X}$ in the domain of $\mathcal{Y}$. The red boxes indicate the selected changed regions.

GAM's ability to direct SDC-GAE's focus towards salient vertex features, which in turn optimizes the structural difference compensation values and enhances the accuracy of CD.

D. Ablation Study of the Loss Function

In the structural difference compensation learning process of SDC-GAE, three loss functions collaboratively guide the model: image reconstruction loss $loss_r$, sparsity constraint loss $loss_d$, and structural consistency loss $loss_s$. $loss_r$ includes the loss term of the reconstructed image and the structural difference compensation value, and thus serves as the fundamental loss function of the proposed SDC-GAE. Therefore, this paper focuses on the ablation of $loss_d$ and $loss_s$ to evaluate their individual impacts on algorithm performance, as detailed in Table VI. Fig. 6 demonstrates that the accuracy of SDC-GAE, when relying solely on $loss_r$, is markedly inferior to that of the comprehensive model incorporating all loss functions. This disparity highlights the insufficiency of $loss_r$ for precise CD and underscores the model's reliance on $loss_d$ and $loss_s$ for enhanced performance. By incorporating $loss_d$ and $loss_s$ into $loss_r$, SDC-GAE achieves an average improvement of 4.7% and 8.70% in OA, 55.28% and 67.10% in KC, and 39.26% and 46.28% in F1, respectively. Conversely, when $loss_d$ and $loss_s$ are omitted from the total loss function, the model with the complete loss function still shows an average enhancement of 3.53% and 1.03% in OA, 14.16% and 5.55% in KC, and 10.08% and 4.41% in F1 Score, respectively. These findings underscore the critical roles of $loss_d$ and $loss_s$ in refining the model's detection of change areas and preserving structural consistency. $loss_d$, by enforcing sparsity in change areas, aids in the precise localization of changes and serves as a regularization term to prevent overfitting, particularly in scenarios with subtle or minimal changes. $loss_s$, on the other hand, ensures the spatial structural consistency between the reconstructed and original images, which is crucial for identifying genuine changes and mitigating false positives due to noise, shadows, or other non-structural elements.

E. Computational Time

This paper presents an analysis of the computational time for the proposed SDC-GAE model, focusing on the smallest (Dataset #8) and largest (Dataset #5) datasets by size, as detailed in Table VII. The data reveals a direct correlation between the

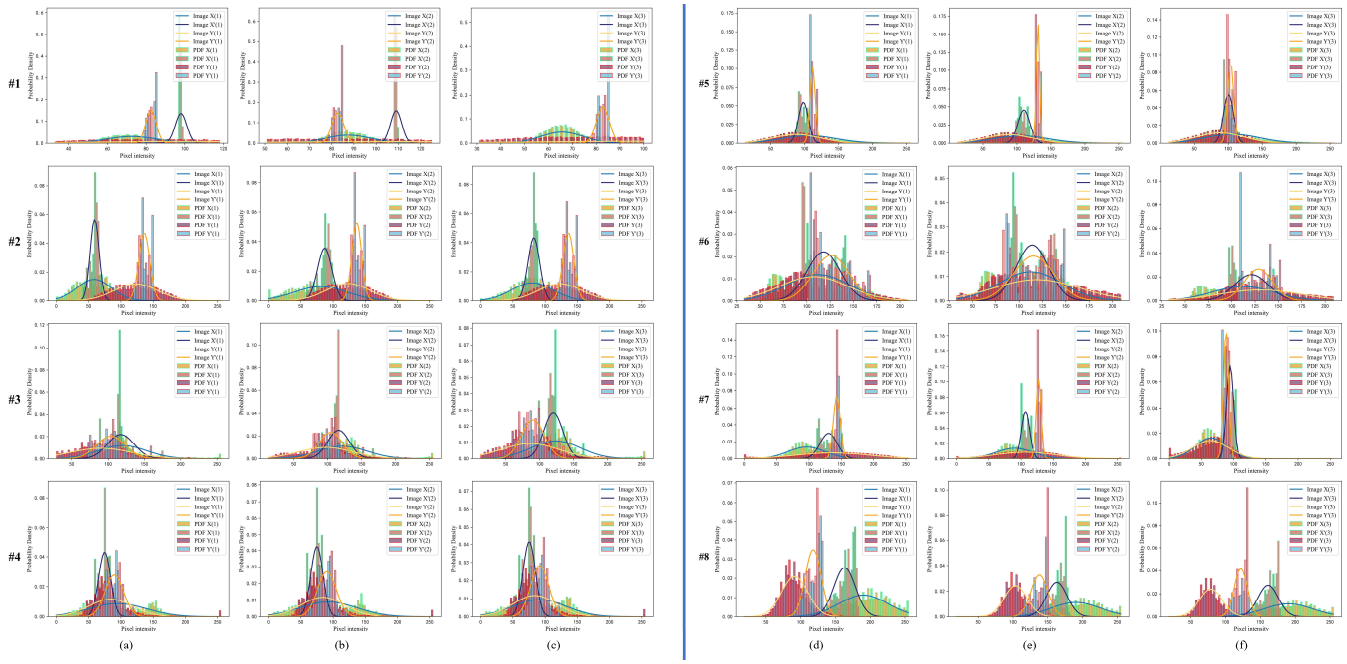


Fig. 7. Comparative analysis of histograms for the reconstructed image $\mathbf{X}'(\mathbf{Y}')$ and the original image $\mathbf{X}(\mathbf{Y})$ across eight datasets. The datasets are divided into two groups: datasets #1 to #4 are displayed in the left column, and datasets #5 to #8 are displayed in the right column. For each dataset, the histograms of the respective bands for both images are juxtaposed from left to right, allowing for a side-by-side comparison.

TABLE VI
THE EFFECTIVENESS OF SPARSITY CONSTRAINT LOSS $loss_d$ AND STRUCTURAL CONSISTENCY LOSS $loss_s$

	$loss_r$	$loss_d$	$loss_s$	#1	#2	#3	#4	#5	#6	#7	#8
OA	√	×	×	0.9202	0.8309	0.8354	0.8447	0.9531	0.9197	0.8400	0.7979
	√	×	√	0.9133	0.9393	0.8361	0.9371	0.9629	0.9251	0.8722	0.9625
	√	√	×	0.9279	0.9591	0.8818	0.9346	0.9749	0.9468	0.9495	0.9484
	√	√	√	0.9429	0.9627	0.8860	0.9615	0.9774	0.9530	0.9525	0.9645
KC	√	×	×	0.7122	0.3158	0.2342	0.3516	0.6675	0.6842	0.2579	0.2910
	√	×	√	0.6952	0.6715	0.3701	0.6486	0.7193	0.7002	0.4513	0.5949
	√	√	×	0.7232	0.6661	0.4738	0.6659	0.7862	0.7757	0.4983	0.6061
	√	√	√	0.7590	0.7067	0.4932	0.6978	0.7945	0.7902	0.5134	0.7185
F1	√	×	×	0.7598	0.4070	0.3212	0.4318	0.6922	0.7292	0.3105	0.3615
	√	×	√	0.7470	0.7053	0.4670	0.6486	0.7391	0.7424	0.5236	0.6136
	√	√	×	0.7658	0.6879	0.5390	0.7022	0.7996	0.8060	0.5241	0.6333
	√	√	√	0.7914	0.7259	0.5561	0.7178	0.8063	0.8169	0.5378	0.7374

TABLE VII
COMPUTATIONAL TIME (SECONDS) OF SDC-GAE

		N_P =3000	N_P =5000	N_P =10000
Dataset #8 300 × 412 × 1(3)	K_{ratio} =0.02	9.12	12.18	22.76
	K_{ratio} =0.10	17.91	32.33	623.45
Dataset #5 4135×2325×3(1)	K_{ratio} =0.02	120.02	163.68	289.29
	K_{ratio} =0.10	125.55	200.06	810.66

number of superpixels N_P and the neighbor ratio K_{ratio} with the computational time required for SDC-GAE. Notably, when the N_P is set to a higher value, such as 10,000, an increase in K_{ratio} significantly extends the computational time. This increase is attributed to the fact that a larger N_P necessitates the processing of more neighbor vertices for each vertex in SDC-GAE, complicating the aggregation operation. Furthermore, managing a higher volume of vertices and their relationships demands additional memory for storing vertex features, weight matrices, and intermediate computation results. Consequently, the augmented memory requirements can potentially restrict the model's efficiency when implemented on hardware.

V. CONCLUSION

This research tackles the challenge of imaging feature discrepancies in MCD by introducing an innovative SDC-GAE model. Utilizing an unsupervised learning framework, SDC-GAE delves into the extraction of deep structural features from images. This model, augmented with a structural difference compensation mechanism, enables precise detection of change

areas. The loss function of SDC-GAE is meticulously designed with three key components: image reconstruction loss, sparsity constraint loss, and structural consistency loss. These components work in tandem to guide the model in minimizing spectral differences between reconstructed and target images, focusing on change areas, and ensuring the structural features of both images remain consistent. Comparative experiments on eight multimodal datasets with state-of-the-art methods have validated the effectiveness and superiority of SDC-GAE in MCD tasks.

SDC-GAE demonstrates its potential in processing multimodal remote sensing data, providing valuable technical support for CD in domains such as natural resource monitoring, disaster assessment, and urban planning. Future research will focus on the continued refinement of the SDC-GAE's performance. The planned enhancements aim to address the following key areas: 1) Advanced graph representation learning: We suggest employing cutting-edge GCN frameworks to more effectively extract and encode image structural features. Integrating multi-scale graph learning will facilitate the simultaneous capture of both local and global structural elements, thereby increasing the model's responsiveness to changes across different scales; 2) Refined attention mechanisms: we advocate for the development of more nuanced attention mechanisms that prioritize change detection areas and minimize overfitting to stable regions. An adaptive attention distribution should be implemented, which adjusts focus based on image content, enabling more precise identification of change features; 3) Multi-task learning: we recommend integrating change detection with complementary tasks, such as classification and segmentation, to benefit from mutual information and enhance the precision of structural difference compensation; 4) Loss function optimization: the design of novel loss functions or the enhancement of existing ones should be pursued to better address structural discrepancies. This could involve the integration of gradient-based loss components to diminish reconstruction inaccuracies; 5) Data augmentation and preprocessing: The use of data augmentation strategies, including image rotation, scaling, and brightness adjustments, during training can improve the model's adaptability to diverse change scenarios. Additionally, preprocessing steps, such as noise reduction and contrast enhancement, should be applied to refine input data quality and, consequently, the accuracy of structural difference compensation.

In subsequent research, efforts also will be directed towards improving the model's real-time detection capabilities. This can be achieved through the following approaches: 1) Employing techniques such as model pruning and knowledge distillation to reduce the complexity of the model while ensuring its performance is maintained; 2) Improving the computation of the loss function to minimize redundant calculations during the model's gradient descent; 3) Further increasing computational efficiency by utilizing parallel computing frameworks and specialized hardware accelerators.

Moreover, we intend to investigate the application of SDC-GAE to a broader range of remote sensing image analysis tasks, with the goal of offering more extensive technical support to related fields.

REFERENCES

[1] M. Hansen and T. Loveland, "A review of large area monitoring of land cover change using Landsat data," *REMOTE SENSING OF ENVIRONMENT*, vol. 122, pp. 66–74, Jul. 2012, doi: 10.1016/j.rse.2011.08.024.

[2] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake Damage Assessment of Buildings Using VHR Optical and SAR Imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 5, pp. 2403–2420, May 2010, doi: 10.1109/TGRS.2009.2038274.

[3] Y. Tang, L. Zhang, and X. Huang, "Object-oriented change detection based on the Kolmogorov–Smirnov test using high-resolution multispectral imagery," *International Journal of Remote Sensing*, vol. 32, no. 20, pp. 5719–5740, Oct. 2011, doi: 10.1080/01431161.2010.507263.

[4] Y. Tang and L. Zhang, "Urban Change Analysis with Multi-Sensor Multispectral Imagery," *Remote Sensing*, vol. 9, no. 3, p. 252, Mar. 2017, doi: 10.3390/rs9030252.

[5] T. Bai *et al.*, "Deep learning for change detection in remote sensing: a review," *Geo-spatial Information Science*, vol. 26, no. 3, pp. 262–288, Jul. 2023, doi: 10.1080/10095020.2022.2085633.

[6] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. L. Rojo-Alvarez, and M. Martinez-Ramon, "Kernel-Based Framework for Multitemporal and Multisource Remote Sensing Data Classification and Change Detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 46, no. 6, pp. 1822–1835, Jun. 2008, doi: 10.1109/TGRS.2008.916201.

[7] L. Wan, Y. Xiang, and H. You, "A Post-Classification Comparison Method for SAR and Optical Images Change Detection," *IEEE Geosci. Remote Sensing Lett.*, vol. 16, no. 7, pp. 1026–1030, Jul. 2019, doi: 10.1109/LGRS.2019.2892432.

[8] 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 Sensing*, vol. 57, no. 12, pp. 9941–9959, Dec. 2019, doi: 10.1109/TGRS.2019.2930322.

[9] T. Han, Y. Tang, X. Yang, Z. Lin, B. Zou, and H. Feng, "Change Detection for Heterogeneous Remote Sensing Images with Improved Training of Hierarchical Extreme Learning Machine (HELM)," *Remote Sensing*, vol. 13, no. 23, p. 4918, Dec. 2021, doi: 10.3390/rs13234918.

[10] 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. on Image Process.*, vol. 24, no. 3, pp. 799–812, Mar. 2015, doi: 10.1109/TIP.2014.2387013.

[11] 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. on Image Process.*, vol. 29, pp. 757–767, 2020, doi: 10.1109/TIP.2019.2933747.

[12] R. Touati and M. Mignotte, "An Energy-Based Model Encoding Nonlocal Pairwise Pixel Interactions for Multisensor Change Detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 56, no. 2, pp. 1046–1058, Feb. 2018, doi: 10.1109/TGRS.2017.2758359.

[13] L. Wan, T. Zhang, and H. J. You, "Multi-sensor remote sensing image change detection based on sorted histograms," *International Journal of Remote Sensing*, vol. 39, no. 11, pp. 3753–3775, Jun. 2018, doi: 10.1080/01431161.2018.1448481.

[14] M. Mignotte, "MRF Models Based on a Neighborhood Adaptive Class Conditional Likelihood For Multimodal Change Detection," *ACRT*, vol. 2022, pp. 1–20, Mar. 2022, doi: 10.5772/acrt.02.

[15] 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. Learning Syst.*, vol. 29, no. 3, pp. 545–559, Mar. 2018, doi: 10.1109/TNNLS.2016.2636227.

[16] W. Zhao, Z. Wang, M. Gong, and J. Liu, "Discriminative Feature Learning for Unsupervised Change Detection in Heterogeneous Images Based on a Coupled Neural Network," *IEEE Trans. Geosci. Remote Sensing*, vol. 55, no. 12, pp. 7066–7080, Dec. 2017, doi: 10.1109/TGRS.2017.2739800.

[17] T. Han, Y. Tang, and Y. Chen, "Heterogeneous Image Change Detection Based on Two-Stage Joint Feature Learning," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 3215–3218. doi: 10.1109/IGARSS46834.2022.9883323.

[18] M. Yang *et al.*, "Multicue Contrastive Self-Supervised Learning for Change Detection in Remote Sensing," *IEEE Trans. Geosci. Remote Sensing*, vol. 61, pp. 1–14, 2023, doi: 10.1109/TGRS.2023.3330494.

[19] R. Touati, M. Mignotte, and M. Dahmane, "Anomaly Feature Learning for Unsupervised Change Detection in Heterogeneous Images: A Deep Sparse Residual Model," *IEEE J. Sel. Top. Appl. Earth Observations*

Remote Sensing, vol. 13, pp. 588–600, 2020, doi: 10.1109/JSTARS.2020.2964409.

[20] 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. Learning Syst.*, vol. 33, no. 9, pp. 4257–4270, Sep. 2022, doi: 10.1109/TNNLS.2021.3056238.

[21] T. Zhan, M. Gong, X. Jiang, and S. Li, "Log-Based Transformation Feature Learning for Change Detection in Heterogeneous Images," *IEEE Geosci. Remote Sensing Lett.*, vol. 15, no. 9, pp. 1352–1356, Sep. 2018, doi: 10.1109/LGRS.2018.2843385.

[22] M. Yang, F. Liu, and M. Jian, "DPFL-Nets: Deep Pyramid Feature Learning Networks for Multiscale Change Detection," *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, p. 15.

[23] 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. Top. Appl. Earth Observations Remote Sensing*, vol. 13, pp. 1551–1566, 2020, doi: 10.1109/JSTARS.2020.2983993.

[24] H. Chen, F. He, and J. Liu, "Heterogeneous Images Change Detection Based on Iterative Joint Global-Local Translation," *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, pp. 1–19, 2022, doi: 10.1109/JSTARS.2022.3192251.

[25] X. Wang, W. Cheng, Y. Feng, and R. Song, "TSCNet: Topological Structure Coupling Network for Change Detection of Heterogeneous Remote Sensing Images," *Remote Sensing*, vol. 15, no. 3, p. 621, Jan. 2023, doi: 10.3390/rs15030621.

[26] H. Chen, N. Yokoya, C. Wu, and B. Du, "Unsupervised Multimodal Change Detection Based on Structural Relationship Graph Representation Learning," *IEEE Trans. Geosci. Remote Sensing*, vol. 60, pp. 1–18, 2022, doi: 10.1109/TGRS.2022.3229027.

[27] Z. Liu, G. Li, G. Mercier, Y. He, and Q. Pan, "Change Detection in Heterogenous Remote Sensing Images via Homogeneous Pixel Transformation," *IEEE Trans. on Image Process.*, vol. 27, no. 4, pp. 1822–1834, Apr. 2018, doi: 10.1109/TIP.2017.2784560.

[28] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 179, pp. 14–34, Sep. 2021, doi: 10.1016/j.isprsjprs.2021.07.007.

[29] L. T. Luppino, F. M. Bianchi, G. Moser, and S. N. Anfinsen, "Unsupervised Image Regression for Heterogeneous Change Detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 57, no. 12, pp. 9960–9975, Dec. 2019, doi: 10.1109/TGRS.2019.2930348.

[30] M. Gong, P. Zhang, L. Su, and J. Liu, "Coupled Dictionary Learning for Change Detection From Multisource Data," *IEEE Trans. Geosci. Remote Sensing*, vol. 54, no. 12, pp. 7077–7091, Dec. 2016, doi: 10.1109/TGRS.2016.2594952.

[31] Y. Sun, L. Lei, X. Li, X. Tan, and G. Kuang, "Patch Similarity Graph Matrix-Based Unsupervised Remote Sensing Change Detection With Homogeneous and Heterogeneous Sensors," *IEEE Trans. Geosci. Remote Sensing*, vol. 59, no. 6, pp. 4841–4861, Jun. 2021, doi: 10.1109/TGRS.2020.3013673.

[32] 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 Sensing*, vol. 60, pp. 1–14, 2022, doi: 10.1109/TGRS.2021.3110998.

[33] Y. Sun, L. Lei, X. Tan, D. Guan, J. Wu, and G. Kuang, "Structured graph based image regression for unsupervised multimodal change detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 185, pp. 16–31, Mar. 2022, doi: 10.1016/j.isprsjprs.2022.01.004.

[34] A. Radoi, "Generative Adversarial Networks Under CutMix Transformations for Multimodal Change Detection," *IEEE Geosci. Remote Sensing Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2022.3201003.

[35] 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 Sensing*, vol. 60, pp. 1–13, 2022, doi: 10.1109/TGRS.2021.3097717.

[36] X. Niu, M. Gong, T. Zhan, and Y. Yang, "A Conditional Adversarial Network for Change Detection in Heterogeneous Images," *IEEE Geosci. Remote Sensing Lett.*, vol. 16, no. 1, pp. 45–49, Jan. 2019, doi: 10.1109/LGRS.2018.2868704.

[37] D. Wang, F. Zhao, H. Yi, Y. Li, and X. Chen, "An unsupervised heterogeneous change detection method based on image translation network and post-processing algorithm," *International Journal of Digital Earth*, vol. 15, no. 1, pp. 1056–1080, Dec. 2022, doi: 10.1080/17538947.2022.2092658.

[38] T. Han, Y. Tang, B. Zou, H. Feng, and F. Zhang, "Heterogeneous Images Change Detection Method based on Hierarchical Extreme Learning Machine Image Transformation," *Journal of Geo-Information Science*, vol. 24, no. 11, pp. 2212–2224, 2022, doi: 10.12082/dqxxkx.2022.220089.

[39] L. T. Luppino *et al.*, "Code-Aligned Autoencoders for Unsupervised Change Detection in Multimodal Remote Sensing Images," *IEEE Trans. Neural Netw. Learning Syst.*, pp. 1–13, 2022, doi: 10.1109/TNNLS.2022.3172183.

[40] L. T. Luppino *et al.*, "Deep Image Translation With an Affinity-Based Change Prior for Unsupervised Multimodal Change Detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 60, pp. 1–22, 2022, doi: 10.1109/TGRS.2021.3056196.

[41] Z. Du, X. Li, J. Miao, Y. Huang, H. Shen, and L. Zhang, "Concatenated Deep-Learning Framework for Multitask Change Detection of Optical and SAR Images," *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, vol. 17, pp. 719–731, 2024, doi: 10.1109/JSTARS.2023.3333959.

[42] Z. Lv, J. Liu, W. Sun, T. Lei, J. A. Benediktsson, and X. Jia, "Hierarchical Attention Feature Fusion-Based Network for Land Cover Change Detection With Homogeneous and Heterogeneous Remote Sensing Images," *IEEE Trans. Geosci. Remote Sensing*, vol. 61, pp. 1–15, 2023, doi: 10.1109/TGRS.2023.3334521.

[43] Q. Liu, K. Ren, X. Meng, and F. Shao, "Domain Adaptive Cross Reconstruction for Change Detection of Heterogeneous Remote Sensing Images via a Feedback Guidance Mechanism," *IEEE Trans. Geosci. Remote Sensing*, vol. 61, pp. 1–16, 2023, doi: 10.1109/TGRS.2023.3320805.

[44] J. Wu *et al.*, "A Dual Neighborhood Hypergraph Neural Network for Change Detection in VHR Remote Sensing Images," *Remote Sensing*, vol. 15, no. 3, p. 694, Jan. 2023, doi: 10.3390/rs15030694.

[45] Y. Sun, L. Lei, D. Guan, and G. Kuang, "Iterative Robust Graph for Unsupervised Change Detection of Heterogeneous Remote Sensing Images," *IEEE Trans. on Image Process.*, vol. 30, pp. 6277–6291, 2021, doi: 10.1109/TIP.2021.3093766.

[46] T. Han, Y. Tang, B. Zou, and H. Feng, "Unsupervised multimodal change detection based on adaptive optimization of structured graph," *International Journal of Applied Earth Observation and Geoinformation*, vol. 126, p. 103630, Feb. 2024, doi: 10.1016/j.jag.2023.103630.

[47] M. Mignotte, "A Fractal Projection and Markovian Segmentation-Based Approach for Multimodal Change Detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 58, no. 11, pp. 8046–8058, Nov. 2020, doi: 10.1109/TGRS.2020.2986239.

[48] R. Touati, M. Mignotte, and M. Dahmane, "Multimodal Change Detection Using a Convolution Model-Based Mapping," in *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Istanbul, Turkey: IEEE, Nov. 2019, pp. 1–6. doi: 10.1109/IPTA.2019.8936127.

[49] 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 Sensing*, vol. 60, pp. 1–21, 2022, doi: 10.1109/TGRS.2021.3053571.

[50] 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 Sensing*, vol. 12, no. 17, p. 2683, Aug. 2020, doi: 10.3390/rs12172683.

[51] D. A. Jimenez-Sierra, D. A. Quintero-Olaya, J. C. Alvear-Munoz, H. D. Benitez-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 Sensing*, vol. 60, pp. 1–16, 2022, doi: 10.1109/TGRS.2022.3168126.

[52] 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 Recognition*, vol. 131, p. 108845, Nov. 2022, doi: 10.1016/j.patcog.2022.108845.

[53] Y. Tang, X. Yang, T. Han, F. Zhang, B. Zou, and H. Feng, "Enhanced Graph Structure Representation for Unsupervised Heterogeneous Change Detection," *Remote Sensing*, vol. 16, no. 4, p. 721, Feb. 2024, doi: 10.3390/rs16040721.

[54] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks." arXiv, Feb. 22, 2017. Accessed: Feb. 25, 2024. [Online]. Available: http://arxiv.org/abs/1609.02907

[55] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012, doi: 10.1109/TPAMI.2012.120.

[56] A. Vaswani *et al.*, "Attention Is All You Need." arXiv, Aug. 01, 2023. Accessed: Feb. 24, 2024. [Online]. Available: http://arxiv.org/abs/1706.03762

[57] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks." arXiv, Feb. 04, 2018. Accessed: Feb. 24, 2024. [Online]. Available: http://arxiv.org/abs/1710.10903

[58] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986, doi: 10.1038/323533a0.

[59] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.

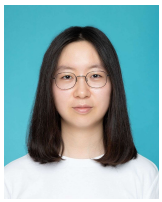
[60] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization." arXiv, Jan. 29, 2017. Accessed: Feb. 24, 2024. [Online]. Available: http://arxiv.org/abs/1412.6980

Te Han received the Bachelor's and Master's degrees in Surveying and Mapping Science and Technology from School of Geosciences and Info-Physics at Central South University in 2017 and 2020, respectively. He is currently pursuing his Ph.D. at the same institution. His research interests include machine learning and deep learning, as well as their application in remote sensing image change detection.

Yuqi Tang received the Master's degree from Wuhan University of Technology in 2008, and a Ph.D. in Engineering from Wuhan University in 2013.

Since 2013, she has been working at the School of Geosciences and Info-Physics at Central South University, where she holds the position of Associate Professor and is a doctoral supervisor. She has published over 30 papers in SCI-indexed journals. Her research has long been focused on the intelligent interpretation of multispectral/hyperspectral remote sensing data and the analysis and monitoring of natural resources. Her research includes multimodal remote sensing image change detection, satellite remote sensing video motion target detection, and hyperspectral remote sensing for water resource monitoring, among other applications in natural resource monitoring.

Yuzeng Chen received the B.S. degree in geographic information science from Southwest University of Science and Technology, Mianyang, China, in 2020 and the M.S. degree from Central South University, in 2023, Changsha, China. He is pursuing the Ph.D. degree with the School of Geodesy and Geomatics, Wuhan University, Wuhan. His research interests include remote sensing/hyperspectral video object detection and tracking, change detection. More details can be found at https://yzcu.github.io. e-mail: yuzeng_chen@whu.edu.cn



Xin Yang received the B.S. degree in Geographic Information Systems (GIS) from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2021. She is currently pursuing her M.Eng. in Surveying Engineering within the Department of Surveying and Remote Sensing at the School of Geosciences and Info-Physics, Central South University, China. Her research interest lies in change detection using multimodal remote sensing imagery.

Yuqiang Guo received the B.S. degree in Geographic Information Science from Zhengzhou University, Zhengzhou, China, in 2023. He is studying for the M.Eng. degree in Surveying Engineering with the Department of Surveying and Remote Sensing, School of Geosciences and Info-Physics, Central South University, China. His research interest is land-cover/-use change detection with homogeneous/heterogeneous remote sensing images.

Shujing Jiang graduated from Wuhan University School in 2012 with a Master's degree. She is currently working as a Senior Engineer at the Guangzhou Urban Planning & Design Survey Research Institute, China. Her research interests include surveying natural resources, monitoring and evaluation, early warning systems for land spatial planning, and the processing and application of real estate data.