DGKAN: Dual-branch Graph Kolmogorov-Arnold Network for Unsupervised Multimodal Change Detection

Tongfei Liu[1], Jianjian Xu[1], Tao Lei[1]*, Yingbo Wang[1], Xiaogang Du[1], Zhiyong Lv[2]

[1]Shaanxi Joint Laboratory of Artificial Intelligence and School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology

[2]School of Computer Science and Engineering, Xi'an University of Technology

liutongfei_home@hotmail.com, {xujianjian_leo@, leitao@, wangyingbo@, duxiaogang@}sust.edu.cn, lvzhiyong_fly@hotmail.com

Abstract

Multimodal change detection (MCD) has important applications in disaster assessment, but the nonlinear distortion of features and spatial misalignment caused by sensor imaging differences make it difficult to obtain changes through direct comparison. To overcome the above problems, this study aims to realize MCD by capturing the modality-independent structural commonality features between Multimodal Remote Sensing Images (MRSIs). To achieve this, we devise a basic Graph Kolmogorov-Arnold Network (GKAN) to excavate spatial structural relationships and cross-modal nonlinear mappings simultaneously. Based on this, we propose a Dual-branch GKAN (DGKAN) for unsupervised MCD, which can capture spatial-spectral structural commonality features, and compare them directly to detect changes. Concretely, the GKAN is used within the DGKAN to build two autoencoders consisting of a Siamese encoder and two independent decoders to learn spatial-spectral structural commonality features through feature reconstruction. Besides, we introduce a Covariance Structural Commonality Loss (CSCL), which guides the network in extracting spatial-spectral structural commonality features between MRSIs by unsupervised constraints on the distributional consistency of cross-modal features. Experiments on several MCD datasets show that the proposed DGKAN can achieve convincing results, and ablation studies verify the effectiveness of the GKAN and CSCL.

Code — https://github.com/TongfeiLiu/DGKAN-for-MCD

Introduction

Change Detection (CD) plays an important role in applications such as land use change and disaster assessment (Noh et al. 2022; Wu, Du, and Zhang 2023; Zheng et al. 2025). It is worth noting that with the improvement of multi-source remote sensing data acquisition capabilities, Multimodal CD (MCD) is gradually becoming mainstream (Lv et al. 2022). However, spatial differences in Multimodal Remote Sensing Images (MRSIs) features and mismatch in structural representations are the main challenges constraining its performance improvement (Ye et al. 2017; Sommervold, Gazzea, and Arghandeh 2023).

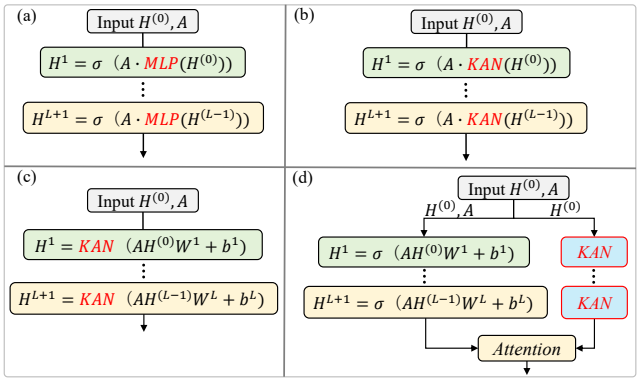

Figure 1: Motivation of our proposed GKAN: (a) GCNs with MLP, (b) GCNs with KAN instead of MLP, (c) GCNs with KAN instead of activation function, (d) Our GKAN adopts a parallel fusion strategy of GCNs and KAN. Note: $H^{(0)}$ and A denote node feature matrix and adjacency matrix, respectively; σ refers to activation function; W and b denote weight matrices and biases, respectively.

Multimodal learning based on deep learning (DL) provides many effective solutions for multimodal remote sensing image interpretation (Guo et al. 2024; Ding et al. 2025). In particular, MCD has received extensive attention and achieved remarkable success (Lv et al. 2022; Peng et al. 2025; Xu et al. 2025). But the fixed kernel size and static receptive field of CNNs still limit its ability to capture contextual information of MRSIs (Lu et al. 2022). Transformer methods can alleviate this problem through the attention mechanism but tend to ignore local structural information (Han et al. 2022; Liu et al. 2024a). Subsequently, people began to use Graph Convolutional Networks (GCNs) to model structural relationships in MRSIs (Jia et al. 2022; Han et al. 2024) and achieved good results. Nevertheless, traditional GCNs (e.g., (Kipf and Welling 2016)) mainly rely on linear graph convolutional layers, which makes it difficult to model nonlinear features of MRSIs. Recently, some studies have suggested that the combination of GCN and Kolmogorov-Arnold Network (KAN) can help enhance nonlinear modeling capabilities (Kiamari, Kiamari, and Krishnamachari

*Corresponding Author.

2024; De Carlo, Mastropietro, and Anagnostopoulos 2024).

Based on this, our motivation comes from two aspects: On the one hand, existing GCNs-based methods mainly rely on multilayer stacking or simple nonlinear activation functions (Wu et al. 2020), which are difficult to adequately capture complex nonlinear relationships between MRSIs in the MCD task. On the other hand, most current graph structure modeling approaches mainly use object-level graphs, which reduce computational complexity through feature aggregation but lose the detailed features of pixel-level spatial-spectral structure (Wu et al. 2021a). A few pixel-level graph construction methods do not adequately incorporate the constraint mechanism of spatial neighborhood structure on attribute consistency (Wu et al. 2020), limiting further improvement in MCD accuracy. Recently, the emergent KAN (Liu et al. 2024b) has achieved good results in CD tasks by virtue of its nonlinear modeling capabilities and interpretability advantages (Wang et al. 2024; Liu et al. 2025a). This suggests that combining KAN (Kiamari, Kiamari, and Krishnamachari 2024; De Carlo, Mastropietro, and Anagnostopoulos 2024) with GCNs may be able to make up for the lack of nonlinear modeling capability of GCNs and obtain more powerful modeling capability. Therefore, we aim to explore the synergy of GCN and KAN to achieve better spatial-spectral structural feature modeling between MRSIs in the MCD task. Our contributions can be summarized in the following:

- We propose a GKAN block that combines KAN with GCNs to enhance the spatial-spectral structural relationship representation between MRSIs by augmenting the GCNs paradigm through joint optimization of spatial structure modeling and higher-order nonlinear decoupling.
- We devise a novel DGKAN framework based on GKAN block for unsupervised MCD, which can capture spatial-spectral structural commonality features between MRSIs, and compare them directly to detect changes.
- We introduce the covariance structure commonality loss (CSCL) to guide the network to acquire the spatial-spectral structural commonality features in MRSIs, overcoming the limitation of direct feature incomparability in MCD.

Related Work

Traditional MCD Methods

Earlier traditional MCD methods mostly used human-designed features such as histogram trend similarity and chunked similarity matrix for modeling (Chen et al. 2024), but the quality of the features limits the generalization ability. Therefore, people try to use statistical methods (multidimensional scaling technology (Touati, Mignotte, and Dahmane 2018), fractal coding (Mignotte 2020)) to reduce human intervention through shallow feature modeling, but ignore the deep semantic associations. Additionally, structure optimization-based methods enhance discriminability through local constraints (e.g., Sparse Constrained Adaptive Structure Consistency (SCASC) (Sun et al. 2021b), Adaptive Graph with Structural Cycle Consistency (AGSCC) (Sun et al. 2022b), and Local neighborhood Preservation Energy Model (LPEM) (Sun et al. 2024)), but may not adequately model the global situation. Recently, some progress has been made in global structure representation based on graph structure theory (Jimenez-Sierra et al. 2020) and frequency domain analysis methods (Chen, Yokoya, and Chini 2023), but these methods are still limited to modeling shallow feature correlations and cannot fully characterize the deep features of MRSIs.

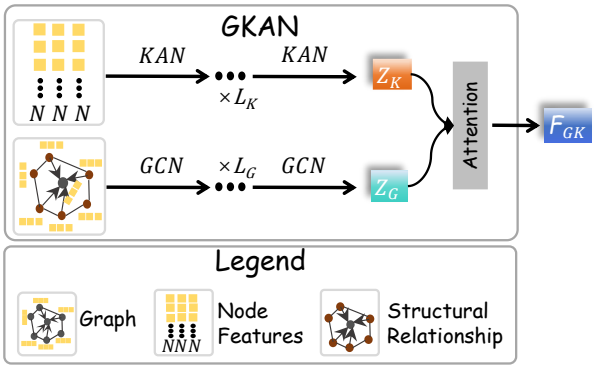

Figure 2: Structure of the proposed GKAN.

DL-based MCD Methods

Current DL-based MCD methods mainly include supervised, semi-supervised and unsupervised (Lv et al. 2022). The former, such as (Liu et al. 2024a) use of label-supervised training to achieve high-precision detection, has the significant limitation that obtaining labeled samples is usually costly. Semi-supervised methods such as PRBCD (Hu et al. 2023) adopt pseudo-label generation strategies to alleviate the need for labels, but there is still the problem of pseudo-label noise propagation. Unsupervised methods eliminate label dependency entirely through autoencoders (Han et al. 2024; Liu et al. 2025b) or adversarial architectures (e.g., CANet (Niu et al. 2018)), focusing on purely data-driven representation learning.

Despite the above methods differing in how they utilize labeled data, they all aim to address the challenge of incomparability between MRSIs. The core detection strategies can be categorized into three approaches: 1) Feature fusion via simple concatenation or attention mechanisms (Feng et al. 2022; Luo et al. 2023); 2) Cross-modal alignment through spatial projection or contrastive learning to bridge heterogeneous features (Luppino et al. 2022; Xing et al. 2023); 3) Image translation using encoder-decoders (e.g., X-Net and ACE-Net (Luppino et al. 2021)) to generate pseudo-isomorphic representations. Therefore, it is meaningful to further study the MCD approach.

Methodology

Proposed GKAN

In this paper, the proposed GKAN achieves the joint optimization of structural modeling and nonlinear feature decoupling by constructing a parallel path architecture. As shown

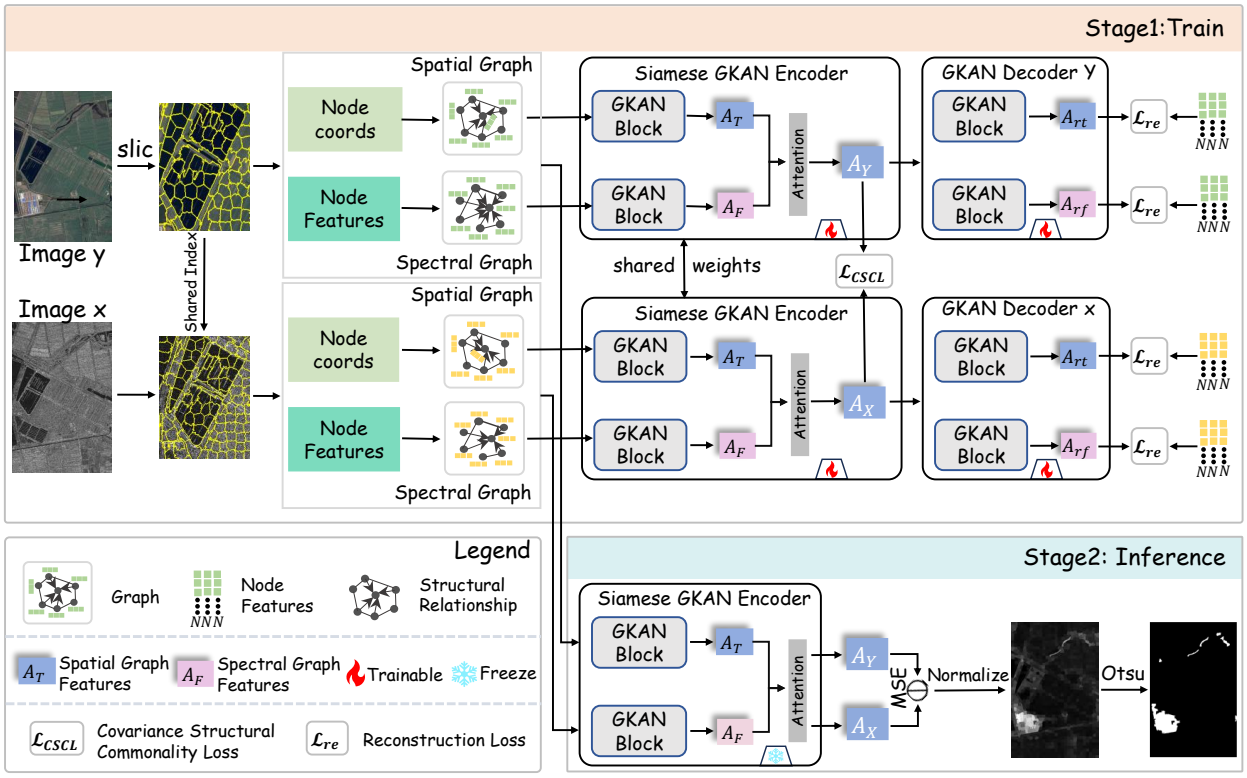

Figure 3: Framework of the proposed DGKAN. Training: Under the guidance of CSCL, the Siamese GKAN encoder is prompted to capture the spatial-spectral structural commonality features between MRSIs. Inference: Using the trained Siamese GKAN encoder to extract the structural commonality features, and generates the DI by the Mean Square Error (MSE). Subsequently, the DI was segmented into a BCM using the Otsu (Otsu et al. 1975) algorithm.

in Fig. 2, the nonlinear feature mapping path uses the KAN (Liu et al. 2024b) adaptive spline basis function of the L_k layer to perform nonlinear feature mapping. For the input feature x of the L-th layer, the transformation formula is:

$$K^{(L)}(x) = \sigma\left(W_{\text{base}}^{(L)} x + \sum_{j=1}^{d_{\text{in}}} \sum_{k=1}^{n} w_{jk}^{(L)} \psi_{jk}^{(L)}(x_j)\right), \quad (1)$$

where $W_{\text{base}}^{(L)}$ is a learnable linear weight matrix, d_{in} is the input dimension, $\psi_{jk}^{(L)}$ is a differentiable spline basis function defined on a dynamic learnable grid, and $w_{jk}^{(L)}$ is a sparse combination weight that satisfies $\|w_{jk}^{(L)}\|_0 \ll n \cdot d_{\text{in}}$, $\sigma(\cdot)$ is an element-by-element nonlinear activation function. The structural modeling path parallel to KAN completes feature update through the GCN of the L_G layer. For the L-th layer, the neighborhood relationship modeling process can be expressed as:

$$H^{(L+1)} = \sigma\left(A \cdot H^{(L)} W_{\text{struct}}^{(L)}\right), \quad (2)$$

where A represents the normalized adjacency matrix, $W_{\text{struct}}^{(L)}$ is the learnable parameter matrix of the GCN structural path of the L-th layer, and $\sigma(\cdot)$ is a nonlinear activation function.

Then, the output Z_G of the L-th layer of the GCN path and the output Z_K of the L-th layer of the feature mapping path are stacked along the sequence dimension to form the input feature X^{seq}. Subsequently, X^{seq} is linearly projected to obtain the query Q, key K and value V. The final F_{GK} output is calculated as follows:

$$F_{GK} = \sum_{t}^{2} \left(\text{softmax}\left(\frac{QK^{\top}}{\sqrt{d}}\right) V\right)_t. \quad (3)$$

Finally, F_{GK} is used as a fusion feature for downstream tasks.

Proposed DGKAN for Unsupervised MCD

Overview of the Proposed DGKAN In this study, a DGKAN based on the GKAN is proposed, aiming to solve the incomparability problem of MRSIs by capturing modality-independent structural commonality features in them. Its core component, the GKAN, achieves the decoupling of spatial structural relationship modeling and cross-modal nonlinear mapping through the combination of graph convolution and KAN. As shown in Fig. 3, MRSIs are first segmented into superpixel primitives using SLIC (Achanta et al. 2012), and pixel-level spectral similarity graphs along with spatial neighborhood graphs are concurrently constructed as inputs to the Siamese encoder. Then, the encoder is constrained through the decoder feature reconstruction

method and CSCL to capture the commonality characteristics of the spatial-spectral structure among MRSIs. In the inference stage, the commonality features of the spatial spectral structure extracted by the encoder are used to generate a Difference Image (DI), and the Otsu threshold (Otsu et al. 1975) is used to generate a Binary Change Map (BCM).

Graph Construction In the proposed DGAKN, a dual graph construction strategy, i.e., a spectral similarity graph and a spatial neighborhood graph are bulit as inputs to improve represenstation ability of spatial-spectral structural commonality features. Here, the detailed dual graph strategy is presents as follows:

Data Preprocessing: This study uses MRSIs data, including the pre-event image $X \in \mathbb{R}^{H\times W\times C_X}$ and post-event image $Y \in \mathbb{R}^{H\times W\times C_Y}$, where $H \times W$ denotes the spatial dimensions, C_X and C_Y represent the number of channels of images of different modalities, respectively.

We perform firstly superpixel segmentation based on the post-event image Y to unify the spatial units. We use SLIC to generate the superpixel mask $S \in \{1,\ldots,N\}^{H\times W}$, where N is the number of superpixels. Apply the generated mask S to the pre-event image X to ensure spatial alignment. In addition, it is also very important to normalize the data after segmentation, which not only helps to eliminate the numerical differences between different modes but also improves the numerical stability of the data. Specifically, for optical images from different sources, we used maximum-minimum normalization, while for images from different sources and of different types (opt and SAR), we used mean-variance normalization. However, before normalizing the SAR images, pixel values with zero absolute values first need to be replaced with non-zero minima and the images need to be log-transformed to reduce the effect of noise and enhance the usability of the images. Through this series of steps, subsequent processing can be carried out on the basis of the normalized data.

Finally, based on the superpixel mask S, the pre-event image X and post-event image Y are partitioned into superpixel sets: $\mathcal{I}_X = \{X_1,\ldots,X_N\}$ and $\mathcal{I}_Y = \{Y_1,\ldots,Y_N\}$, where N represents the number of superpixels. This step aggregates pixel-level data into superpixel units as input for subsequent processing.

Dual Graph Construction: Given a pre-event image superpixel region $X_i \subseteq \mathcal{I}_X$, which contains n_i pixel nodes, the pixel value of each node p forms the feature matrix $V_X \in \mathbb{R}^{n_i\times d_f}$. For modality X, the $n_i \times n_i$ similarity matrix is generated by calculating the euclidean distance between node features, which is used to define the edge set E_f^X of the spectral similarity graph $G^X_{\text{spectral}}(V_X, E_f^X)$. Meanwhile, based on the spatial coordinates of the nodes, their 2D planar Euclidean distances are computed, and the set of neighborhood edges E_t^X is generated by k-nearest neighbor filtering to form a spatial neighborhood graph $G^X_{\text{spatial}}(V_X, E_t^X)$ that constrains the local spatial continuity.

In the above manner, the superpixel segmentation region $X_i \subseteq \mathcal{I}_X$ of the pre-event image X can be constructed as a dual-graph structure data:

$$\mathbf{G}_X = \left\{G^X_{\text{spectral}}(V_X, E_f^X), G^X_{\text{spatial}}(V_X, E_t^X)\right\}. \quad (4)$$

Similarly, the superpixel segmentation region $Y_i \subseteq \mathcal{I}_Y$ of the image Y after the event can be constructed as a dual-graph structure data, expressed as:

$$\mathbf{G}_Y = \left\{G^Y_{\text{spectral}}(V_Y, E_f^Y), G^Y_{\text{spatial}}(V_Y, E_t^Y)\right\}. \quad (5)$$

Architecture of the Proposed DGKAN The overall architecture of the proposed DGKAN is shown in Fig. 3, and its core lies in applying the proposed GKAN. DGKAN uses an autoencoder consisting of a Siamese encoder and two independent decoders to extract the common features of the spatial-spectral structure among MRSIs. The specific process of feature extraction task is as follows:

Here, we define the GKAN for spectral similarity graph modeling as $\mathcal{E}_F(\cdot)$, the GKAN for spatial neighborhood graph modeling as $\mathcal{E}_T(\cdot)$, and the dual-graph feature fusion process as $\mathcal{A}_{\text{fu}}(\cdot,\cdot)$. Therefore, the encoding process of image X can be expressed in the following form:

$$\mathbf{A}_F^X = \mathcal{E}_F(V_X, E_f^X), \mathbf{A}_T^X = \mathcal{E}_T(V_X, E_t^X), \quad (6)$$

where (V_X, E_f^X) and (V_X, E_t^X) are the initial spectral similarity graph and spatial neighborhood graph constructed based on image X respectively. Next, through the feature fusion function $\mathcal{A}_{\text{fu}}(\cdot,\cdot)$, the encoder jointly expresses the extracted features:

$$\mathbf{A}_X = \mathcal{A}_{\text{fu}}(\mathbf{A}_F^X, \mathbf{A}_T^X), \quad (7)$$

where $\mathbf{A}_X$ is the final output of the encoder, representing the fused joint feature representation, and serves as the input of the subsequent CD task.

Similarly, the encoding process of the image Y after the event can be expressed as:

$$\mathbf{A}_F^Y = \mathcal{E}_F(V_Y, E_f^Y), \mathbf{A}_T^Y = \mathcal{E}_T(V_Y, E_t^Y), \quad (8)$$

$$\mathbf{A}_Y = \mathcal{A}_{\text{fu}}(\mathbf{A}_F^Y, \mathbf{A}_T^Y), \quad (9)$$

where (V_Y, E_f^Y) and (V_Y, E_t^Y) are the initial spectral similarity graph and spatial neighborhood graph constructed based on image Y respectively; $\mathbf{A}_Y$ is the encoder output of image Y, representing the fused joint feature representation.

The purpose of the decoder is to reconstruct the original node feature matrix based on the output of the encoder to ensure the effectiveness of the learned features. Specifically, the spectral similarity graph reconstruction task is implemented by the independent GKAN $\mathcal{D}_F(\cdot)$, and the spatial neighborhood graph reconstruction task is completed by GKAN $\mathcal{D}_T(\cdot)$. Based on this decoding architecture, the process of decoding the embedded features generated by the encoder for the input image $\mathbf{A}_X$ can be formally expressed as:

$$\mathbf{A}_{rf}^X = \mathcal{D}_F(A_X, E_f^X), \mathbf{A}_{rt}^X = \mathcal{D}_T(A_X, E_t^X), \quad (10)$$

where (A_X, E_f^X) and (A_X, E_t^X) represent the intermediate spectral similarity graph and spatial neighborhood graph constructed based on X, respectively.

Similarly, for the embedded feature $\mathbf{A}_Y$ of image Y, the decoding process can be expressed as:

$$\mathbf{A}_{rf}^Y = \mathcal{D}_F((A_Y, E_f^Y)), \mathbf{A}_{rt}^Y = \mathcal{D}_T((A_Y, E_t^Y)), \quad (11)$$

where (A_Y, E_f^Y) and (A_Y, E_t^Y) represent the intermediate spectral similarity graph and spatial neighborhood graph constructed based on Y, respectively. In this way, the proposed DGKAN achieves MCD by deeply mining the spatial-spectral structural commonality features between MRSIs.

Loss Function In the proposed DGKAN, we proposes a dual loss function consists of a basic reconstruction loss ($\mathcal{L}_{re}$) and a CSCL ($\mathcal{L}_{CSCL}$). The ($\mathcal{L}_{re}$) ensures feature validity by minimizing the MSE between the original node feature matrix and the reconstructed node feature matrix. The mathematical construction process of $\mathcal{L}re$ can be expressed as:

$$\mathcal{L}_{\text{re}} = \sum_{j \in \{X,Y\}} \left(\|\mathbf{V}_j - \mathbf{A}_{\text{rf}}^j\|_{\text{MSE}} + \|\mathbf{V}_j - \mathbf{A}_{\text{rt}}^j\|_{\text{MSE}} \right), \tag{12}$$

where $\mathbf{V}_j$ represents the original node feature matrix of image j. $\mathbf{A}_{\text{rf}}^j$ and $\mathbf{A}_{\text{rt}}^j$ are the reconstructed node feature matrices, respectively. Every term $||\cdot||_{\text{MSE}}$ denotes the mean square error between the computed feature matrices, which is used as a measure of the similarity between the decoder output features and the original features.

The other $\mathcal{L}_{CSCL}$ aims to guide the network in extracting spatial spectral structural commonality features among multimodal remote sensing images by minimizing the covariance difference between multimodal features in the shared feature space. The detailed process can be described as: We mean-center and L2 normalization the features A_X and A_Y extracted by the encoder:

$$\begin{aligned} \mathbf{A}'_X &= \mathbf{A}_X - \text{mean}(\mathbf{A}_X, \dim = 0) \\ \mathbf{A}'_Y &= \mathbf{A}_Y - \text{mean}(\mathbf{A}_Y, \dim = 0). \end{aligned} \tag{13}$$

After mean centering, we further perform L2 normalization on the features to eliminate feature scale differences (see formula 14).

$$\mathbf{A}'_X = \frac{\mathbf{A}'_X}{|\mathbf{A}'_X|_2}, \quad \mathbf{A}'_Y = \frac{\mathbf{A}'_Y}{|\mathbf{A}'_Y|_2}, \tag{14}$$

where $|\mathbf{A}'_X|_2$ and $|\mathbf{A}'_Y|_2$ represent the L2 normalizations of A'_X and A'_Y along specific dimensions, respectively. Next we calculate the covariance matrix of the cross-modal features A'_X and A'_Y as:

$$\mathbf{C}_X = \mathbf{A}'_X(\mathbf{A}'_X)^T, \quad \mathbf{C}_Y = \mathbf{A}'_Y(\mathbf{A}'_Y)^T. \tag{15}$$

Finally, the difference between their distributions is measured by calculating the MSE between the covariance matrices $\mathbf{C}_X$ and $\mathbf{C}_Y$, that is:

$$\mathcal{L}_{CSCL} = \frac{1}{N} \sum_{m,n} \left(C_X^{(m,n)} - C_Y^{(m,n)} \right)^2, \tag{16}$$

where N is the number of elements in the matrix, m and n are the indices in the covariance matrix. Based on the above description, we can express the total loss of optimizing network parameters as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{re}} + \mathcal{L}_{\text{CSCL}}. \tag{17}$$

Based on the above loss function $\mathcal{L}_{\text{total}}$, our DGKAN will be constrained to learn modality-independent spatial-spectral structural commonality features between MRSIs.

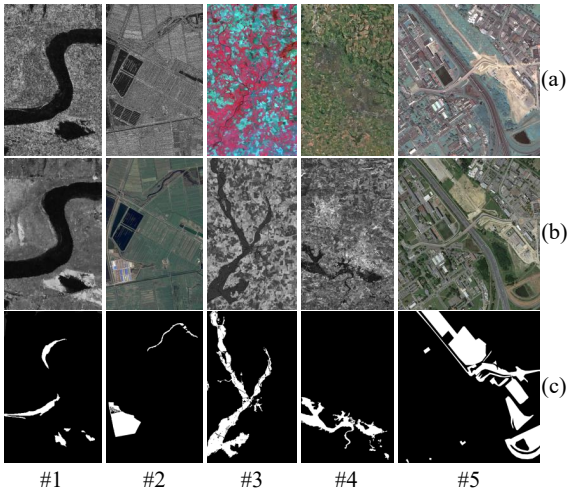

Figure 4: Dataset #1-#5: (a) pre-event images, (b) post-event images, (c) ground reference images.

Measuring DI and BCM In the inference stage, the Siamese GKAN encoder within the DGKAN can efficiently extract the spatial-spectral structural commonality features between MRSIs, and compare them to acquire changes. To achieve this, the DI can be generated by calculating the MSE distances between these structural commonality features, thereby reflecting the change magnitude between MRSIs effectively. Herein, DI can be computed as:

$$\mathcal{DI} = \left\{ \|A_X^i - A_Y^i\|_{\text{MSE}}, \quad i \in \{1, 2, \ldots, N\} \right\}, \tag{18}$$

where A_X^i and A_Y^i represent the commonality structural features between the i-th superpixel regions of the MRSIs. After obtaining the $\mathcal{DI}$, the Otsu threshold method is adopted to divide the DI into a BCM.

Experiments and Results

Data Descriptions and Evaluation Metrics

As shown in Fig. 4, we selected five public MCD datasets in the experiments. We numbered datasets #1 to #5, which include changes caused by urban land use changes and natural disasters. Tab. 1 lists the details of the dataset. Besides, we used three widely used evaluation metrics (Liu et al. 2025a), including Overall Accuracy (OA), Kappa coefficient (Ka), and F1-Score (F1), to measure the accuracy of MCD.

Experimental Setup

In the proposed DGKAN, we experimentally set L_K = L_G = 2 within the GAKN. It is worth noting that expansion with arbitrary path depth is allowed in the GKAN by adjusting L_K and L_G. During the training process, the encoder-decoder uses the Adam optimizer, the learning rate is $lr = 0.0001$, and the weight decay is $weight\_decay = 0.0001$ to ensure stable convergence of the network. Set $K = 10$ when constructing the spatial graph in order to preserve local spatial information in the topology. During training, we adopt the strategy of updating the network weights after processing every ten superpixels. In addition to the fixed parameter

Datasets	Date	Modality	Size (used band)	Spatial resolution (m)	Change Event	Location
#1	June 2008/Sept. 2010	Radarsat-2/Google Earth	$444 \times 291 \times 1/1$ (all)	8	Riverbank erosion	Yellow River, China
#2	June 2008/Sept. 2012	Radarsat-2/Google Earth	$593 \times 921 \times 1/3$ (all)	8	Constructions	Shuguang, China
#3	1999/2000	Spot/NDVI	$990 \times 554 \times 3/1$ (all)	≈ 25	Flooding	Gloucester, England
#4	July 2006/July 2007	QuickBird-2/TerraSAR-X	$4135 \times 2325 \times 3/1$ (all)	≈ 0.65	Flooding	Gloucester, England
#5	May 2012/July 2013	Pleiades/WorldView2	$2000 \times 2000 \times 3/3$ (all)	0.52	Constructions	Toulouse, France

Table 1: Detailed descriptions of the five MCD datasets in the experiments.

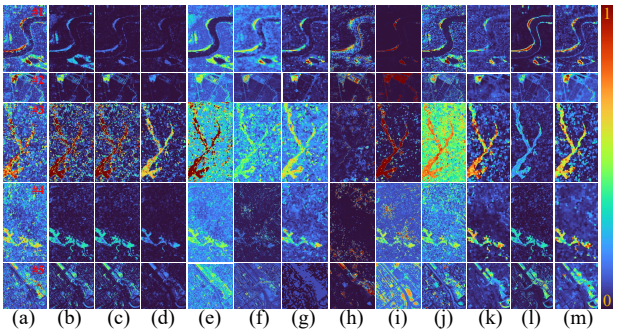

Figure 5: Visualization of DIs comparison of different methods: (a) IRG-McS, (b) GIR-MRF, (c) SCASC, (d) AGSCC, (e) IST-CRF, (f) CACD, (g) SR-GCAE, (h) BASNet, (i) PRBCD-Net, (j) SDC-GAE, (k) AEKAN, (l) CFRL and (m) Ours.

settings described above, there are two key adjustable parameters for SLIC superpixel segmentation of event images: n_segments (P_n) and compactness (P_c), which control the number of superpixel and the degree of regularity of the superpixel region, respectively.

Experimental Comparisons on DIs

As shown in Fig. 5, we visualize DIs generated by different methods as heatmaps to intuitively present the change areas and their intensity distribution. Judging from the heatmap results, the heatmap of DIs generated by our DGKAN has clear levels of change intensity, which is consistent with the actual change degree gradient, and its visual discernibility is significantly better than other algorithms. For example, for datasets #1, #3, and #4, compared with the heatmaps of other methods, our method generates heatmaps with continuous color change and natural gradient, sharp edges that closely fit the river boundary, and stable bank areas without obvious scatter noise. This is a strong indication that our proposed DGKAN can learn the spatial-spectral structural commonality features between MRSIs and achieve effective MCDs through direct comparison.

Experimental Comparisons on BCMs

For the BCMs, quantitative results (Table 2) show that DGKAN achieves SOTA performance on most datasets. This indicates that the detection results of our DGKAN have fewer false detections and misses compared to other methods and can recognize detailed changes that are difficult to detect by other methods. Furthermore, visual comparison

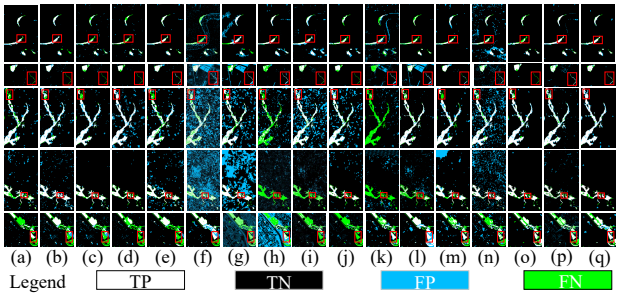

Figure 6: BCMs of different methods on MCD datasets #1-#5: (a) IRG-McS, (b) GIR-MRF, (c) SCASC, (d) AGSCC, (e) IST-CRF, (f) GBF-CD, (g) GLSS, (h) CANet, (i) CACD, (j) SR-GCAE, (k) BGAAE, (l) PRBCD, (m) LPEM, (n) SDC-GAE, (o) AEKAN, (p) CFRL and (q) DGKAN (Ours).

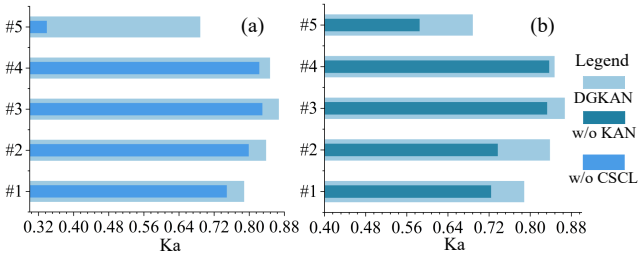

Figure 7: Ablation study results on datasets #1-#5: (a) CSCL, (b) GKAN.

can also reach the same conclusion as the quantitative comparison, as shown in Fig. 6. This advantage stems from the proposed GKAN, which enhances the extraction of complex features by decoupling joint structural modeling from nonlinear features. In addition, under the guidance of CSCL, our proposed DGKAN can capture the commonality features of spatial-spectral structure among MRSIs more accurately. In addition, we evaluated the computational cost. Taking dataset #5 as an example, GCN has 58B parameters, 21.76M FLOPs, and an inference time of 1.59s; GKAN has 334B parameters, 65.28M FLOPs, and an inference time of 4.19s. Despite the increased computational cost, the performance improvement more than compensates for the difference.

Discussion

Ablation Study

Ablation Study for CSCL To verify the effectiveness of CSCL, we remove CSCL to test whether it can effectively

Methods	Dataset #1			Dataset #2			Dataset #3			Dataset #4			Dataset #5		
	OA	Ka	F1	OA	Ka	F1	OA	Ka	F1	OA	Ka	F1	OA	Ka	F1
IRG-McS (Sun et al. 2021a)	**0.986**	**0.788**	**0.795**	0.983	0.794	0.803	0.936	0.704	0.740	0.950	0.606	0.633	0.875	0.349	0.403
GIR-MRF (Sun et al. 2022c)	**0.986**	**0.788**	**0.795**	0.981	0.798	0.807	0.937	0.734	0.770	0.957	0.711	0.734	0.873	0.447	0.519
SCASC (Sun et al. 2021b)	0.976	0.623	0.636	0.979	0.741	0.751	0.950	0.776	0.804	0.960	0.828	0.838	0.872	0.327	0.381
AGSCC (Sun et al. 2022b)	0.976	0.591	0.603	0.983	0.772	0.781	0.955	0.791	0.817	0.960	0.809	0.819	0.895	0.489	0.542
IST-CRF (Sun et al. 2022a)	**0.986**	0.783	0.790	0.981	0.801	0.811	0.920	0.649	0.695	0.935	0.585	0.618	0.899	0.614	0.674
GBF-CD (Jimenez-Sierra et al. 2020)	0.924	0.297	0.330	0.720	0.161	0.229	0.720	0.197	0.345	0.586	0.072	0.176	0.690	0.114	0.289
GLSS (Jimenez-Sierra et al. 2022)	0.949	0.456	0.480	0.854	0.200	0.256	0.850	0.490	0.571	0.642	0.141	0.236	0.459	-0.1	0.155
LPEM (Sun et al. 2024)	0.978	0.679	0.690	0.982	0.807	0.817	0.947	0.761	0.791	0.928	0.538	0.575	0.909	0.687	**0.741**
CANet (Niu et al. 2018)	0.980	0.727	0.737	0.954	0.414	0.437	0.844	0.225	0.313	0.854	-0.026	0.511	0.838	0.199	0.277
X-Net (Luppino et al. 2021)	0.918	0.232	0.268	0.954	0.595	0.618	0.909	0.637	0.688	/	/	/	0.830	0.229	0.322
ACE-Net (Luppino et al. 2021)	0.928	0.297	0.329	0.956	0.613	0.635	0.928	0.659	0.701	/	/	/	0.842	0.294	0.382
CACD (Wu et al. 2021b)	0.967	0.614	0.630	0.978	0.738	0.750	0.798	0.417	0.516	0.906	0.190	0.240	0.829	0.029	0.085
SR-GCAE (Chen et al. 2022)	0.981	0.694	0.704	0.982	0.761	0.770	0.884	0.586	0.649	0.953	0.626	0.651	0.874	0.460	0.531
BGAAE (Jia et al. 2022)	0.945	0.398	0.424	0.899	0.142	0.192	0.863	0.016	0.585	0.891	-0.030	0.264	0.840	0.173	0.247
PRBCD-Net (Hu et al. 2023)	0.983	0.662	0.671	0.865	0.315	0.365	0.968	0.850	0.868	0.924	0.286	0.326	0.783	-0.095	0.178
SDC-GAE (Han et al. 2024)	0.916	0.345	0.378	0.935	0.490	0.521	0.841	0.489	0.572	0.852	0.363	0.424	0.834	0.266	0.358
AEKAN (Liu et al. 2025a)	0.982	0.681	0.690	**0.987**	0.831	0.838	0.964	0.837	0.858	0.983	0.839	0.847	0.921	0.658	0.702
CFRL (Liu et al. 2025b)	0.985	0.787	**0.795**	0.970	0.693	0.709	0.960	0.822	0.845	0.963	0.703	0.723	0.885	0.518	0.584
DGKAN (Ours)	**0.986**	**0.788**	**0.795**	0.986	**0.838**	**0.845**	**0.970**	**0.867**	**0.884**	**0.984**	**0.847**	**0.856**	**0.923**	**0.688**	0.734

Table 2: Comparisons between our DGKAN and other SOTA methods for datasets #1-#5. The best is presented in bold.

guide the DGKAN to better learn the spatial-spectral structural commonality features between MRSIs. As shown in Figure Ablation(a), on all datasets, the complete model including $\mathcal{L}_{CSCL}$ outperformed the ablation model relying solely on reconstruction loss. There are different improvements for several other datasets. Therefore, the validity of $\mathcal{L}_{CSCL}$ is verified.

Ablation Study for KAN To test the effect of the GKAN, the KAN path is removed from the complete architecture in this ablation experiment, leaving the remaining structure and hyperparameters unchanged. On this basis, we conducted the experiments and obtained the corresponding Ka results. As shown in Fig. 7(b), after removing the KAN path, the Ka performance indicator on each dataset shows a systematic decay. Furthermore, experiments replacing the KAN branch with MLP show that GKAN improves the Ka metrics by an average of 8.1% across all datasets, further demonstrating its advantages in jointly modeling graph structural features and decoupling nonlinear features, effectively alleviating the problem of incomparability between MRSIs in the MCD scenario.

Parameter Analysis for the P_c And P_n

By setting the SLIC parameter P_n in the range of 1000 to 2600 with an interval of 200 for all data sets, and setting P_c between 0.2 and 0.4 and between 10 and 30 with intervals of 0.02 and 2 for data #1 and #2 to #5, respectively, their impact on the experimental results is analyzed. As shown in Fig. 8, it can be observed that changes in P_c and P_n have some effect on the ka assessment indicator values for all datasets. As for P_c, it affects the regularity of superpixels on image boundaries, thereby affecting the detection details and causing fluctuations in detection results. While the parameter P_n controls the number of superpixels and modifies the amount of information in a single superpixel, thereby affecting the extraction and evaluation results of structural commonality

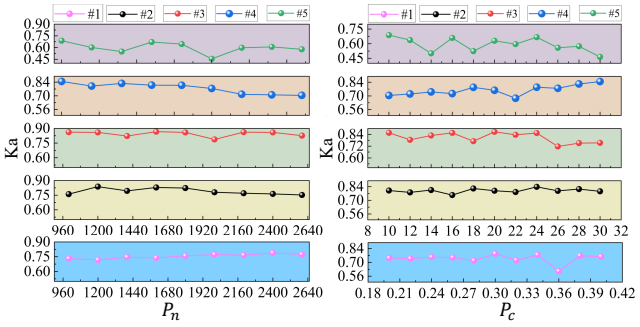

Figure 8: Parameter analysis results of the parameter P_c and P_n in the proposed DGKAN on datasets #1-#5.

features. Therefore, the appropriate P_c and P_n parameters must be chosen to obtain satisfactory detection results for MCD.

Conclusion

In this paper, a novel GKAN was proposed that can improve the spatial-spectral structural features of MRSIs by modeling spatial structural relationships and cross-modal nonlinear mapping simultaneously. And our proposed DGKAN also achieved effective unsupervised MCD by extracting the modality-independent spatial-spectral structural commonality features between MRSIs. Extensive experimental comparisons on five MCD datasets demonstrate that the proposed DGKAN outperforms SOTA MCD methods. In addition, ablation studies verify the effectiveness of the proposed GKAN and CSCL in MCD. In the future, we will further test the effectiveness of the proposed DGKAN on other more complex MCD datasets and the applicability of GKAN in different fields.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Program No. 62271296, 62201334), in part by The Innovation Capability Support Plan Project in Shaanxi Province (Program No. 2025RS-CXTD-012), in part by the Natural Science Basic Research Program of Shaanxi (Program No. 2025JC-YBQN-800, 2024JC-YBQN-0037) and the Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No. 24JK0350, 23JP014, 23JP022), in part by the Key Research and Development Program of Shaanxi (Program No. 2024GX-YBXM-121), in part by the Young Science and Technology Innovation Leading Talents Program of Xi'an, China (Program No. 25ZQRC00019).

References

Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; and Süsstrunk, S. 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(11): 2274–2282.

Chen, H.; Yokoya, N.; and Chini, M. 2023. Fourier domain structural relationship analysis for unsupervised multimodal change detection. *ISPRS J. Photogramm. Remote Sens.*, 198: 99–114.

Chen, H.; Yokoya, N.; Wu, C.; and Du, B. 2022. Unsupervised multimodal change detection based on structural relationship graph representation learning. *IEEE Trans. Geosci. Remote Sens.*, 60: 1–18.

Chen, J.; Hou, D.; He, C.; Liu, Y.; Guo, Y.; and Yang, B. 2024. Change detection with cross-domain remote sensing images: A systematic review. *Int. J. Appl. Earth Obs. Geoinf.*

De Carlo, G.; Mastropietro, A.; and Anagnostopoulos, A. 2024. Kolmogorov-arnold graph neural networks. *arXiv preprint arXiv:2406.18354*.

Ding, L.; Zuo, X.; Hong, D.; Guo, H.; Lu, J.; Gong, Z.; and Bruzzone, L. 2025. S2C: Learning Noise-Resistant Differences for Unsupervised Change Detection in Multimodal Remote Sensing Images. *arXiv preprint arXiv:2502.12604*.

Feng, Y.; Xu, H.; Jiang, J.; Liu, H.; and Zheng, J. 2022. ICIF-Net: Intra-scale cross-interaction and inter-scale feature fusion network for bitemporal remote sensing images change detection. *IEEE Trans. Geosci. Remote Sens.*, 60: 1–13.

Guo, X.; Lao, J.; Dang, B.; Zhang, Y.; Yu, L.; Ru, L.; Zhong, L.; Huang, Z.; Wu, K.; Hu, D.; et al. 2024. Skysense: A multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery. In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 27672–27683.

Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. 2022. A survey on vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(1): 87–110.

Han, T.; Tang, Y.; Chen, Y.; Yang, X.; Guo, Y.; and Jiang, S. 2024. SDC-GAE: Structural Difference Compensation Graph Autoencoder for Unsupervised Multimodal Change Detection. *IEEE Trans. Geosci. Remote Sens.*

Hu, L.; Liu, Q.; Liu, J.; and Xiao, L. 2023. Prbcd-net: Predict-refining-involved bidirectional contrastive difference network for unsupervised change detection. *IEEE Trans. Geosci. Remote Sens.*

Jia, M.; Zhang, C.; Zhao, Z.; and Wang, L. 2022. Bipartite graph attention autoencoders for unsupervised change detection using vhr remote sensing images. *IEEE Trans. Geosci. Remote Sens.*, 60: 1–15.

Jimenez-Sierra, D. A.; Benítez-Restrepo, H. D.; Vargas-Cardona, H. D.; and Chanussot, J. 2020. Graph-based data fusion applied to: Change detection and biomass estimation in rice crops. *Remote Sens.*, 12(17): 2683.

Jimenez-Sierra, D. A.; Quintero-Olaya, D. A.; Alvear-Munoz, J. C.; Benitez-Restrepo, H. D.; Florez-Ospina, J. F.; and Chanussot, J. 2022. Graph learning based on signal smoothness representation for homogeneous and heterogeneous change detection. *IEEE Trans. Geosci. Remote Sens.*, 60: 1–16.

Kiamari, M.; Kiamari, M.; and Krishnamachari, B. 2024. Gkan: Graph kolmogorov-arnold networks. *arXiv preprint arXiv:2406.06470*.

Kipf, T. N.; and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Liu, B.; Chen, H.; Li, K.; and Yang, M. Y. 2024a. Transformer-based multimodal change detection with multitask consistency constraints. *Inf. Fusion*, 108: 102358.

Liu, T.; Xu, J.; Lei, T.; Wang, Y.; Du, X.; Zhang, W.; Lv, Z.; and Gong, M. 2025a. AEKAN: Exploring Superpixel-based AutoEncoder Kolmogorov-Arnold Network for Unsupervised Multimodal Change Detection. *IEEE Trans. Geosci. Remote Sens.*

Liu, T.; Zhang, M.; Gong, M.; Zhang, Q.; Jiang, F.; Zheng, H.; and Lu, D. 2025b. Commonality Feature Representation Learning for Unsupervised Multimodal Change Detection. *IEEE Trans. Image Process.*, 34: 1219–1233.

Liu, Z.; Wang, Y.; Vaidya, S.; Ruehle, F.; Halverson, J.; Soljačić, M.; Hou, T. Y.; and Tegmark, M. 2024b. Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756*.

Lu, C.; Xia, M.; Qian, M.; and Chen, B. 2022. Dual-branch network for cloud and cloud shadow segmentation. *IEEE Trans. Geosci. Remote Sens.*, 60: 1–12.

Luo, F.; Zhou, T.; Liu, J.; Guo, T.; Gong, X.; and Ren, J. 2023. Multiscale diff-changed feature fusion network for hyperspectral image change detection. *IEEE Trans. Geosci. Remote Sens.*, 61: 1–13.

Luppino, L. T.; Hansen, M. A.; Kampffmeyer, M.; Bianchi, F. M.; Moser, G.; Jenssen, R.; and Anfinsen, S. N. 2022. Code-aligned autoencoders for unsupervised change detection in multimodal remote sensing images. *IEEE Trans. Neural Networks Learn. Syst.*, 35(1): 60–72.

Luppino, L. T.; Kampffmeyer, M.; Bianchi, F. M.; Moser, G.; Serpico, S. B.; Jenssen, R.; and Anfinsen, S. N. 2021. Deep image translation with an affinity-based change prior for unsupervised multimodal change detection. *IEEE Trans. Geosci. Remote Sens.*, 60: 1–22.

Lv, Z.; Huang, H.; Li, X.; Zhao, M.; Benediktsson, J. A.; Sun, W.; and Falco, N. 2022. Land cover change detection with heterogeneous remote sensing images: Review, progress, and perspective. *Proc. of the IEEE*, 110(12): 1976–1991.

Mignotte, M. 2020. A fractal projection and Markovian segmentation-based approach for multimodal change detection. *IEEE Trans. Geosci. Remote Sens.*, 58(11): 8046–8058.

Niu, X.; Gong, M.; Zhan, T.; and Yang, Y. 2018. A conditional adversarial network for change detection in heterogeneous images. *IEEE Geosci. Remote Sens. Lett.*, 16(1): 45–49.

Noh, H.; Ju, J.; Seo, M.; Park, J.; and Choi, D.-G. 2022. Unsupervised change detection based on image reconstruction loss. In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 1352–1361.

Otsu, N.; et al. 1975. A threshold selection method from gray-level histograms. *Autom.*, 11(285-296): 23–27.

Peng, D.; Liu, X.; Zhang, Y.; Guan, H.; Li, Y.; and Bruzzone, L. 2025. Deep learning change detection techniques for optical remote sensing imagery: Status, perspectives and challenges. *Int. J. Appl. Earth Obs. Geoinf.*, 136: 104282.

Sommervold, O.; Gazzea, M.; and Arghandeh, R. 2023. A survey on SAR and optical satellite image registration. *Remote Sens.*, 15(3): 850.

Sun, Y.; Lei, L.; Guan, D.; and Kuang, G. 2021a. Iterative robust graph for unsupervised change detection of heterogeneous remote sensing images. *IEEE Trans. Image Process.*, 30: 6277–6291.

Sun, Y.; Lei, L.; Guan, D.; Kuang, G.; Li, Z.; and Liu, L. 2024. Locality Preservation for Unsupervised Multimodal Change Detection in Remote Sensing Imagery. *IEEE Trans. Neural Networks Learn. Syst.*

Sun, Y.; Lei, L.; Guan, D.; Li, M.; and Kuang, G. 2021b. Sparse-constrained adaptive structure consistency-based unsupervised image regression for heterogeneous remote-sensing change detection. *IEEE Trans. Geosci. Remote Sens.*, 60: 1–14.

Sun, Y.; Lei, L.; Guan, D.; Wu, J.; and Kuang, G. 2022a. Iterative structure transformation and conditional random field based method for unsupervised multimodal change detection. *Pattern Recog.*, 131: 108845.

Sun, Y.; Lei, L.; Guan, D.; Wu, J.; Kuang, G.; and Liu, L. 2022b. Image regression with structure cycle consistency for heterogeneous change detection. *IEEE Trans. Neural Networks Learn. Syst.*, 35(2): 1613–1627.

Sun, Y.; Lei, L.; Tan, X.; Guan, D.; Wu, J.; and Kuang, G. 2022c. Structured graph based image regression for unsupervised multimodal change detection. *ISPRS J. Photogramm. Remote Sens.*, 185: 16–31.

Touati, R.; Mignotte, M.; and Dahmane, M. 2018. Change detection in heterogeneous remote sensing images based on an imaging modality-invariant MDS representation. In *IEEE Int. Conf. Image Process. (ICIP)*, 3998–4002. IEEE.

Wang, Y.; Yu, X.; Gao, Y.; Sha, J.; Wang, J.; Gao, L.; Zhang, Y.; and Rong, X. 2024. Spectralkan: Kolmogorov-arnold network for hyperspectral images change detection. *arXiv preprint arXiv:2407.00949*.

Wu, C.; Du, B.; and Zhang, L. 2023. Fully convolutional change detection framework with generative adversarial network for unsupervised, weakly supervised and regional supervised change detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(8): 9774–9788.

Wu, J.; Li, B.; Qin, Y.; Ni, W.; Zhang, H.; Fu, R.; and Sun, Y. 2021a. A multiscale graph convolutional network for change detection in homogeneous and heterogeneous remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.*, 105: 102615.

Wu, Y.; Li, J.; Yuan, Y.; Qin, A. K.; Miao, Q.-G.; and Gong, M.-G. 2021b. Commonality autoencoder: Learning common features for change detection from heterogeneous images. *IEEE Trans. Neural Networks Learn. Syst.*, 33(9): 4257–4270.

Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; and Philip, S. Y. 2020. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Networks Learn. Syst.*, 32(1): 4–24.

Xing, Y.; Zhang, Q.; Ran, L.; Zhang, X.; Yin, H.; and Zhang, Y. 2023. Progressive modality-alignment for unsupervised heterogeneous change detection. *IEEE Trans. Geosci. Remote Sens.*

Xu, J.; Liu, T.; Lei, T.; Chen, H.; Yokoya, N.; Lv, Z.; and Gong, M. 2025. CGSL: Commonality graph structure learning for unsupervised multimodal change detection. *ISPRS J. Photogramm. Remote Sens.*, 229: 92–106.

Ye, Y.; Shan, J.; Bruzzone, L.; and Shen, L. 2017. Robust registration of multimodal remote sensing images based on structural similarity. *IEEE Trans. Geosci. Remote Sens.*, 55(5): 2941–2958.

Zheng, Z.; Ermon, S.; Kim, D.; Zhang, L.; and Zhong, Y. 2025. Changen2: Multi-temporal remote sensing generative change foundation model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 47(2): 725–741.