Nonlocal Feature Learning Based on a Variational Graph Auto-Encoder Network for Small Area Change Detection using SAR Imagery

Hang Su[a,#], Xinzheng Zhang[a,b,*], Yuqing Luo [a,#], Ce Zhang[c,d,*], Xichuan Zhou[a], Peter M. Atkinson[c,e,f]

[a] *School of Microelectronics and Communication Engineering, Chongqing University, Chongqing, 400044, China*

[b] *Chongqing Key Laboratory of Space Information Network and Intelligent Information Fusion, Chongqing, 400044, China*

[c] *Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, United Kingdom*

[d] *UK Centre for Ecology & Hydrology, Library Avenue, Lancaster, LA1 4AP, United Kingdom*

[e] *Geography and Environmental Science, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom*

[f] *Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 11A Datun Road, Beijing 100101, China*

# These authors contributed equally to this work and should be considered co-first authors.

Abstract

Synthetic aperture radar (SAR) image change detection is a challenging task due to inherent speckle noise, imbalanced class occurrence and the requirement for discriminative feature learning. The traditional handcrafted feature extraction and current convolution-based deep learning techniques have some advantages, but suffer from being limited to neighborhood-based spatial information. The nonlocally observable imbalance phenomenon that exists naturally in small area change detection has presented a huge challenge to methods that focus on local features only. In this paper, an unsupervised method based on a variational graph auto-encoder (VGAE) network was developed for object-based small area change detection using SAR images, with the advantages of alleviating the negative impact of class imbalance and suppressing speckle noise. The main steps include: 1) Three types of difference image (DI) are combined to establish a three-channel fused DI (TCFDI), which lays the data-level foundation for subsequent analysis. 2) Simple linear iterative clustering (SLIC) is used to divide the TCFDI into superpixels regarded as nodes. Two functions are proposed and developed to measure the similarity between nodes to build a weighted undirected graph. 3) A VGAE network is designed and trained using the graph and nodes, and high-level nonlocal feature representations of each node are extracted. The network, with a Gaussian Radial Basis Function constrained by geospatial distances, establishes the connection among nonlocal, but similar superpixels in the process of feature learning, which leads to speckle noise suppression and distinguishable features learned in latent space. The nodes are then identified as changed or unchanged classes via *k*-means clustering. Five real SAR datasets were used in comparative experiments. Up to 99.72% accuracy was achieved, which is superior to state-of-the-art methods that pay attention only to local information, thus, demonstrating the effectiveness and robustness of the proposed approach.

Keywords:

Synthetic aperture radar, Change detection, Difference image, Graph auto-encoder network, Deep learning.

1. Introduction

Change detection using bi-temporal remotely sensed imagery is a common goal in a wide range of applications including environmental protection, land-cover monitoring and forest resource management (Muster et al., 2015; Pantze et al., 2013; Zhang et al., 2016; Lu et al., 2011; Jia et al., 2016). Synthetic aperture radar (SAR) images, compared with optical remote sensing images, have significant advantages including their relative insensitivity to atmospheric and sunlight conditions (Gong et al., 2017; Zhang et al., 2021; Li et al., 2019). However, they are usually contaminated by speckle noise, which brings interference and loss of signal to some extent. Furthermore, the changed area is commonly far smaller than the unchanged area in large scenes observed by SAR, presenting a significant imbalance and bringing great challenges for automatic change detection methods.

From the perspective of the basic unit of classification, change detection methods can be divided into pixel-based and object-based methods (Zhuang et al., 2020; Hussain et al., 2013). Compared with pixel-based methods, object-based approaches exhibit higher accuracy and efficiency due to utilizing homogeneous pixel groups as the identification unit. Change detection methods can also be classified into supervised and unsupervised methods. Unsupervised methods have been studied extensively and attracted much attention, because ground reference data containing the pixel labels are commonly unavailable or insufficient. The main steps of unsupervised approaches usually include: 1) preprocessing (e.g., geometric registration, denoising); 2) generating a difference image (DI); 3) analyzing the DI and identifying changed or unchanged pixels. This article focuses on an unsupervised, object-based method.

The step of generating the DI aims to provide valuable guidance for later procedures, in which subtraction and ratio operators are two classic methods for discriminating changed from unchanged pixels. The logarithmic ratio is popular for SAR images since it transforms multiplicative speckle into additive noise. Local spatial information can be exploited to suppress speckle noise (Zhang et al., 2013). For example, the mean ratio and neighborhood-based ratio can increase the signal-to-noise ratio by averaging and, thus, enhance the discriminative ability between changed and unchanged classes (Gong et al., 2012). The spatial-temporal adaptive neighborhood-based ratio (Zhuang et al., 2018) and adaptive generalized likelihood ratio test (Zhuang et al., 2020) were developed to select the optimal window size for generating the DI, to avoid image geometric degradation and texture loss caused by using neighborhood information from a fixed regular window. In (Zhang et al., 2021) and (Wang et al., 2020), irregular local homogeneous information was considered and used to increase texture and edge details. The high-quality DI provides more reliable guiding information for subsequent image interpretation and analysis.

In the process of DI analysis and pixel classification, threshold-based and clustering-based methods are prevalent (Gong et al., 2014). The former is limited due to using only pixel intensity information (Bazi et al., 2005). Clustering-based methods, such as k-means and fuzzy c-means (FCM), have attracted much attention because they exploit more information in the DI (e.g., multi-dimensional features). Research was undertaken to explore feature representations to improve clustering. Li et al. (2015) developed the Gabor wavelet representation to extract multi-dimensional information from the DI, which demonstrated outstanding noise robustness. Celik. (2009) applied principal component analysis (PCA) to extract key spatial features. Recently, deep learning-based techniques have received great attention and been applied widely in the field of remote sensing image processing. Deep learning algorithms can extract high-level semantic features automatically, and build more discriminative feature representations than hand-crafted features (Tajbakhsh et al., 2016; Wang et al., 2019; Cheng et al., 2018). A convolutional neural network (CNN) (Li et al., 2019) and a convolutional wavelet neural network (CWNN) (Gao et al., 2019) were introduced for local feature learning and change detection, achieving state-of-the art performance. Jaswanth et al. (2022) investigated the curvelet transform, which was used in the pre-classification of change detection, to assist a CNN in building more discriminative feature representations. Based on a neural network framework for change detection, Zhang et al. (2022) introduced a multi-objective sparse feature learning (MO-SFL) model where the sparsity of representation was adaptively learned, increasing the algorithm robustness to speckle noise. Dong et al. (2022) integrated a CNN with clustering to learn clustering-friendly feature representations, which showed advantages in preserving details of changed areas and suppressing speckle noise. Those studies indicate that deep learning models can transform visual features into a high-level semantic feature space and eliminate the deleterious effects of speckle noise effortlessly, effectively boosting SAR image change detection accuracy.

Auto-encoder (AE) networks play an important role in unsupervised deep learning. A classic AE contains an encoder and a decoder to remove redundant information by minimizing reconstruction errors. AEs have been studied extensively and adopted for SAR image change detection due to their predominant denoising and feature learning abilities. *Gong et al* (Gong et al., 2017) reshaped the image patches as spatial feature vectors, and developed a sparse AE to learn the relationship among neighboring pixels to establish robust high-level representations. In (Lv et al., 2018), simple linear iterative clustering (SLIC) was used for superpixel object segmentation on a DI to obtain homogeneous local regions, and a stacked AE (SAE) was introduced for denoising and deep feature extraction. *Liu et al* incorporated the Fisher discriminant criterion into SAE to further strengthen the discriminative ability (Liu et al., 2019). However, using only the local pixels and their neighboring information is insufficient for feature representation. In addition, both image patches and superpixels are isolated during the learning process of the aforementioned AEs and their variants, which makes it hard to capture deep discriminative features in change detection, especially in the situation of severe imbalances between the changed and unchanged pixels. Inspired by the fact that human understanding is not only based on local observations, but also on nonlocal or long-range observations, we explore the possibility to establish relations among nonlocal samples to obtain more robust high-level feature representations.

Recently, the graph neural network (GNN) was introduced with the capability to learn nonlocal features by harnessing the graph structure of samples. Kipf et al. (2016) used a GNN as an encoder to develop a framework for unsupervised learning on graph-structured data, and applied it to several challenging tasks, such as link prediction (Cai et al., 2021; Grover et al., 2019) and node clustering (Yang et al., 2019; Salha et al., 2019; Wang et al., 2017). In this research, a novel unsupervised change detection method based on GNN was proposed for bi-temporal SAR images. It is inappropriate to apply the GNN directly to images, which are non-graph structured data. Therefore, we obtain superpixels from DIs as nodes which are the basic units of classification. Then, the similarity measure function is developed to evaluate the relationship among nodes (superpixels) to build a weighted undirected graph. Here, three different types of DI are used to build graphs to integrate fully the capability of these DIs. Then, a Variational Graph Auto-encoder (VGAE) is employed to learn nonlocal features, the learning process of which can be understood as the collaborative representation of homogeneous nodes on the entire DI. Because VGAE is suitable for solving unbalanced classification tasks with graph-structured data, we adopt VGAE to extract features and improve the representation and increase the discrimination ability of the acquired features. The contributions of this article are, thus:

1) A novel unsupervised method based on VGAE was developed for small area change detection with bi-temporal SAR images, which can effectively suppress speckle noise and obtain powerful high-level representations in latent feature space.
2) A novel similarity metric for nodes was proposed to build the graph, which integrates the similarity in the visual intensity space and the geospatial distance between the nodes. The obtained reliable graph supported VGAE to capture the core semantic features and remove redundant information in noisy environments.
3) A three-channel fusion DI (TCFDI) was developed to provide a wealth of change information for nodes, conducive to learning more generalized features.

The remainder of this article is organized as follows. Section II and Section III describe the existing relevant knowledge and the proposed methodology, respectively. Section IV provides the experimental results and the analysis. Finally, the conclusions are drawn in Section V.

2. Existing relevant knowledge

Recently, the success of deep learning, including through CNNs, has promoted research in the field of pattern recognition and computer vision. Image analysis tasks have been completely changed by various deep learning paradigms, such as object detection (Redmon et al., 2016; Ren et al., 2017), semantic segmentation (Han et al., 2021; Li et al., 2021), and image enhancement (Liu et al., 2021; Dai et al., 2021). In an image, pixels are attributed to a regular rigid grid in Euclidean space, while CNNs are able to exploit the shift-invariance and local connectivity of the image to extract meaningful local features for further analysis and recognition (Wu et al., 2021). Although CNNs can effectively capture the hidden characteristics of the image local space most tasks, in reality, exhibit a unique, non-Euclidean data structure. Hence, the much-anticipated GNN was created aiming to be the analysis method of the deep learning model in the graph domain.

Examining CNNs and graphs, it can be found that the keys to the convolutional layers in the CNN can be summarized as local connection and shared weights, which can be generalized to the graph domain and developed into graph convolution. Thus, graph convolutional networks (GCN) emerged, exploiting the connectivity and dependencies between nodes to construct an associated graph structure. GCN can learn the global information of an image and allow information to flow over the entire association graph to learn discriminative global feature representations. A graph $G(V,E)$ can be defined by the relationship between nodes, where V represents the node set and E is the edge set. $G(V,E)$ can be specifically represented by a *weighted adjacency matrix* $\mathbf{A} \in \mathbf{R}^{N \times N}$. The degree matrix $\mathbf{D} \in \mathbf{R}^{N \times N}$ can be obtained by the adjacency matrix $\mathbf{A}$, whose element d_{ii} can be calculated by Eq.1:

$$d_{ii} = \sum_{j=1}^{j=N} a_{ij},\ a_{ij} \in \mathbf{A} \tag{1}$$

The symmetrically normalized Laplacian matrix is:

$$\mathbf{L}^{\mathbf{sym}} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \tag{2}$$

where $\mathbf{I} \in \mathbf{R}^{N \times N}$ is the identity matrix. Given the graph data $\mathbf{s} \in \mathbf{R}^{N}$, which denotes the feature vector of all nodes of a graph where $\mathbf{s}_i$ is the value of the i^{th} node (Wu et al., 2021). A filter $\mathbf{g}_{\boldsymbol{\theta}} = diag(\boldsymbol{\theta})$ parameterized by $\boldsymbol{\theta}$, the graph convolution is defined as:

$$\mathbf{g}_{\boldsymbol{\theta}} * \mathbf{s} = \mathbf{U} \mathbf{g}_{\boldsymbol{\theta}} \mathbf{U}^{\mathbf{T}} \mathbf{s} \tag{3}$$

where $\mathbf{U}$ is the matrix of eigenvectors of $\mathbf{L}^{\mathbf{sym}}$.

The eigendecomposition of the Laplacian matrix imposes an extremely high computational cost, and any perturbation to the graph results in a change of eigenbasis (Wu et al., 2021). ChebNet (Defferrard et al., 2016) utilizes the Chebyshev polynomial of the Eigenvalue diagonal matrix to approximate the filter operator $\mathbf{g}_{\boldsymbol{\theta}}$ to achieve K-order local convolution on the graph:

$$\mathbf{g}_{\theta} * \mathbf{s} \approx \sum_{k=0}^{k=K} \theta_k^{'} \mathbf{T}_k \tilde{\mathbf{L}} \mathbf{s} \tag{4}$$

where $\tilde{\mathbf{L}} = (2\mathbf{L}/\upsilon_{max}) - \mathbf{I}$ and the υ_{max} denotes the largest eigenvalue of $\mathbf{L}$. $\theta_k^{'}$ denotes the learnable parameters of *K*-local convolution. The Chebyshev polynomial is defined recursively by $\mathbf{T}_k(s) = 2s\mathbf{T}_{k-1}(s) - \mathbf{T}_{k-2}(s)$, which is the weight parameter matrix of *K*-local convolution.

The *K*-order Chebyshev polynomial is restricted to $K = 1$ to alleviate the over-fitting problem of the graph with a wide distribution of node degrees on the local neighborhood structure. So the GCN (Kipf et al., 2016) is further defined as:

$$\mathbf{g}_\theta * \mathbf{s} \approx \boldsymbol{\theta}_0\mathbf{s} + \boldsymbol{\theta}_1(\mathbf{L} - \mathbf{I}_N)\mathbf{s} \tag{5}$$

Thus, Eq. 5 can be rewritten as:

$$\mathbf{g}_\theta * \mathbf{s} \approx \boldsymbol{\theta}(\mathbf{I} + \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}})\mathbf{s} \tag{6}$$

Then, the single-layer GCN can be formulated as:

$$\begin{aligned}\mathbf{X}^{(l+1)} &= GCN(\mathbf{X}^{(l)}, \mathbf{A}) \\ &= (\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{X}^{(l)}\mathbf{W}^{(l)})\end{aligned} \tag{7}$$

where $\mathbf{X}^{(l+1)}$ and $\mathbf{X}^{(l)}$ are the output and input, respectively, and $\mathbf{W}^{(l)}$ represents the network parameters. The $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$ denote the adjacency matrix, which is self-connected, and $\tilde{\mathbf{D}}$ is the degree matrix of $\tilde{\mathbf{A}}$.

The Graph Auto-Encoder (GAE) is proposed to map nodes to embedding space to establish a low-dimensional representation through unsupervised training. It employs a multi-layer GCN to encode the nodes into embedded representations, uses a dot product decoder to reconstruct the adjacency matrix, and finally minimizes the reconstruction error between the original adjacency matrix $\mathbf{A}$ and the reconstructed adjacency matrix $\mathbf{B}$. The encoder and decoder can be denoted as Eq. 8 and Eq. 9:

$$\mathbf{Z} = GCN(\mathbf{X}, \mathbf{A}) \tag{8}$$

$$\mathbf{B} = \mathrm{Dot}(\mathbf{Z}\mathbf{Z}^{\mathbf{T}}) \tag{9}$$

where $\mathbf{Z}$ is the learned embedding low-dimensional vector, and $\mathrm{Dot}(\cdot)$ is the inner product function.

3. Methodology

The methodology of the proposed approach is exhibited in Fig. 1, which includes four parts: 1) Three kinds of DI are generated to form a TCFDI, and TCFDI is implemented with superpixel segmentation. 2) The superpixels are treated as nodes, and a graph structure is built. It should be noted that the three established graphs maintain a unified structure, but the node features in the three graphs are different. 3) A nonlocal feature representation is learned using VGAE. 4) *k*-means clustering is employed for node classification.

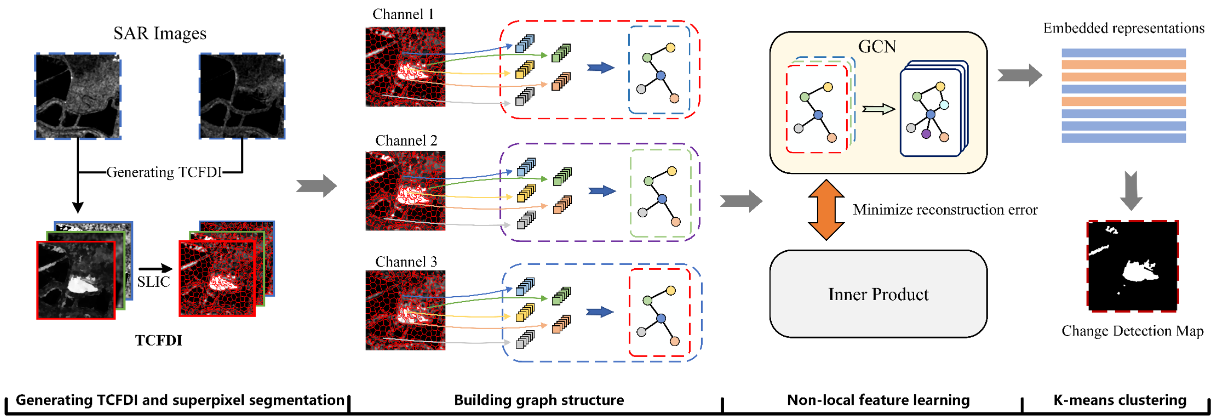


Fig. 1. Methodology of the proposed change detection approach.

3.1 Difference image generation

A TCFDI is developed to discover more abundant guidance information for subsequent analysis. Three types of DI, specifically the log-ratio DI (LRDI) (Li et al., 2019), the Combined DI (CDI) (Zheng et al., 2014) and the DI based on multi-scale superpixel reconstruction (MSRDI) (Zhang et al., 2021), are produced as the ingredients of subsequent analysis. Among them, LRDI provides strong robustness to the multiplicative speckle noise inherent in SAR images. Compared to the original version, we replaced the log-ratio operator of the CDI with the ratio operator, preventing the inhibition of weakly changed pixels. MSRDI suppresses speckle noise by exploiting homogeneous information in the local neighborhood, while retaining rich detailed edges. The TCFDI can be expressed as:

$$\mathbf{I}_{\mathbf{FDI}}^{1} = \mathrm{MSRDI}(\mathbf{I}_{\mathbf{SAR}}^{1}, \mathbf{I}_{\mathbf{SAR}}^{2}) \tag{10}$$

$$\mathbf{I}_{\mathbf{FDI}}^{2} = \log(\left|\mathbf{I}_{\mathbf{SAR}}^{1} / \mathbf{I}_{\mathbf{SAR}}^{2}\right|) \tag{11}$$

$$\mathbf{I}_{\mathbf{FDI}}^{3} = mean(\left|\mathbf{I}_{\mathbf{SAR}}^{1} - \mathbf{I}_{\mathbf{SAR}}^{2}\right|) + median(\max(\left|\mathbf{I}_{\mathbf{SAR}}^{1} / \mathbf{I}_{\mathbf{SAR}}^{2}\right|, \left|\mathbf{I}_{\mathbf{SAR}}^{2} / \mathbf{I}_{\mathbf{SAR}}^{1}\right|)) \tag{12}$$

where $\mathbf{I}_{\mathbf{SAR}}^{1}$ and $\mathbf{I}_{\mathbf{SAR}}^{2}$ are bi-temporal SAR images, $\mathbf{I}_{\mathbf{FDI}}^{c}$ is the c-th channel of the TCFDI and $c = 1,2,3$ denotes the channel index. $mean(\cdot)$ and $median(\cdot)$ are the mean and median filter operators, respectively.

The main motivations for developing the TCFDI are as follows: (1) the rich fused information in the TCFDI can ensure that the subsequent superpixel segmentation has better edge adhesion; (2) the three DIs focus on different types of information: MSRDI has good edge discriminating ability, CDI can capture weak intensity changes, and LRDI combined with filter operators can effectively suppress speckle noise. Information in the TCFDI gathered from the three DIs facilitates VGAEN to learn the most salient, generalized knowledge relating to the changed and unchanged classes. (3) the pixel features of the three DIs are combined to provide more valuable guidance for establishing reliable graph structures in the follow-up system.

3.2 Building the graph structure

Simple Linear Iterative Clustering (SLIC) is used to segment the TCFDI $\mathbf{I}_{\mathbf{FDI}}^{\mathbf{c}}$ to obtain superpixels. The set of N superpixels is expressed as $\{\mathbf{O}_n^1, \mathbf{O}_n^2, \mathbf{O}_n^3\}_{n=1}^{n=N}$, where 1, 2, 3 refer to the channel index. Then, each superpixel in $\{\mathbf{O}_n^1, \mathbf{O}_n^2, \mathbf{O}_n^3\}_{n=1}^{n=N}$ is reshaped into a M-dimensional feature vector, where M is the maximum number of pixels in all superpixels. When the number of pixels inside a superpixel is smaller than M, the median value of the current superpixel is used to fill the corresponding vector. All reshaped superpixel vectors are represented as $\{\mathbf{X}^1, \mathbf{X}^2, \mathbf{X}^3\}$, where $\mathbf{X}^c = [\mathbf{X}_1^c, \mathbf{X}_2^c, \ldots, \mathbf{X}_n^c, \ldots, \mathbf{X}_N^c] \in \mathbf{R}^{N \times M}$. That is, the task of classifying each pixel is transformed into that of identifying the reshaped superpixel vectors. For the purpose of establishing connections between analogous samples during training, two methods were developed to build the graph structure, respectively.

Gaussian Radial Basis Function:

The first method is that the graph is constructed by measuring similarities between vertices in intensity feature space. Here, the Gaussian radial basis function (GRBF) was introduced to calculate the similarity between nodes, as in Eq. 13:

$$S_{ij} = \exp(-\lambda(\sum_{c=1}^{c=3} \alpha_c \left\| \mathbf{X}_i^c - \mathbf{X}_j^c \right\|_F)^2) \tag{13}$$

where λ is the control parameter in intensity feature space, α_c are the weight parameters controlling the contribution of the three types of DI to the composition, and $\| \cdot \|_F$ is the Frobenius norm.

GRBF Constrained by Geospatial Distance:

Furthermore, the spatial position information of the superpixels (nodes) in the visual space can enhance the descriptiveness of the graph for representing global knowledge. Thus, a novel similarity metric function that combines geospatial position information and intensity features is proposed to construct a more reliable graph structure, as in Eq. 14:

$$S_{ij} = \exp(-\lambda(\sum_{c=1}^{c=3} \alpha_c \left\| \mathbf{X}_i^c - \mathbf{X}_j^c \right\|_F)^2 - \eta(\left\| \mathbf{P}_i^c - \mathbf{P}_j^c \right\|_F)^2) \tag{14}$$

where η is control parameter in visual space, and $\mathbf{P}_i^c$ is a vector that records the centroid position of the superpixel on the TCFDI. The adjacency matrix can be built as:

$$\mathbf{A} = \begin{pmatrix} S_{11} & S_{12} & \ldots & S_{1N} \\ S_{21} & S_{22} & \ldots & S_{2N} \\ \vdots & \vdots & \ldots & \vdots \\ S_{N1} & S_{N2} & \ldots & S_{NN} \end{pmatrix} \tag{15}$$

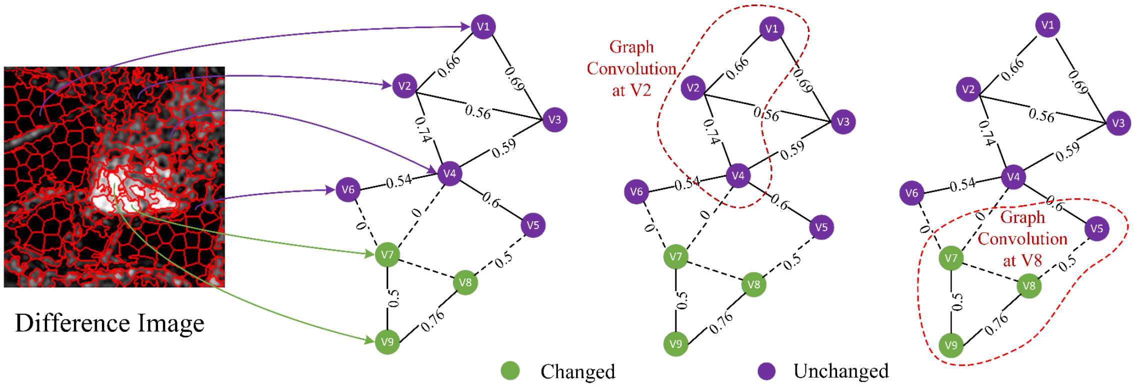


Fig. 2 Principle of graph convolution based on similarity graph structure.

Fig. 2 demonstrates the graph convolution to explain deeply the motivation and purpose of the proposed method. The graph convolution process is regarded as the fusion of local spatial information of nodes on the graph structure. The above methods are designed to establish the edges between nodes, which can convert the arrangement of superpixels from the DI in the visual space into the similarity space. Therefore, it can be regarded as the gradual creation of a joint representation of the interested node and its similar nodes in the learning process of GAE. The information flows on the graph structure of the similarity space, so as to traverse the DI for a long distance and realize non-local learning. The important point is that the proposed method based on the graph structure effectively describes and models the imbalanced phenomenon that the changed nodes (superpixels) are far less than the unchanged nodes. Such a graph will always maintain the constraint of imbalance in the subsequent feature learning process, so that the nodes belonging to the minority class will not be regarded as noise and removed.

3.3 Variational graph autoencoder

VGAE is a probabilistic model, which takes the adjacency matrix $\mathbf{A} \in \mathbf{R}^{N \times N}$ and feature matrix $\mathbf{X}^c \in \mathbf{R}^{N \times M}$ $c = 1,2,3$ as inputs, and aims at embedding the $\mathbf{X}^c$ into the latent subspace as the stochastic latent variables $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_N] \in \mathbf{R}^{N \times F}$, where $M > F$. The model (encoder) is defined as:

$$\begin{aligned} q(\mathbf{Z} \mid \mathbf{X}^c, \mathbf{A}) &= \prod\nolimits_{i=1}^{N} q(\mathbf{z}_i \mid \mathbf{X}^c, \mathbf{A}) \\ q(\mathbf{z}_i \mid \mathbf{X}^c, \mathbf{A}) &= N(\mathbf{z}_i \mid \boldsymbol{\mu}_i, diag(\boldsymbol{\sigma}_i^2)) \end{aligned} \tag{16}$$

where $N(\cdot)$ is the Gaussian Normal distribution, and the matrix $\boldsymbol{\mu}$ of means $\boldsymbol{\mu}_i$ and the matrix $\boldsymbol{\sigma}$ of variances $\boldsymbol{\sigma}_i$ are parameterized by GCN. That is, GCN learns the mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}$ of low-dimensional vector representations of nodes. The final output of the encoder is $\mathbf{Z}$, and the latent vectors $\mathbf{z}_i$ are realizations drawn from $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ distributions. The encoder is designed as a two-layer GCN:

$$\mathbf{Z} = \Gamma(GCN_{\boldsymbol{\mu}}(\mathbf{X}^c, \mathbf{A}) \,\&\, GCN_{\boldsymbol{\sigma}}(\mathbf{X}^c, \mathbf{A})) \tag{17}$$

Where $\Gamma(\cdot)$ is sampling function. According to the encoder designed above, the distributional inference model can be parameterized as $\boldsymbol{\mu} = GCN_{\boldsymbol{\mu}}(\mathbf{X}^c, \mathbf{A})$ and $\log \boldsymbol{\sigma} = GCN_{\boldsymbol{\sigma}}(\mathbf{X}^c, \mathbf{A})$, the two GCN models shared parameters in the first layer, and which are defined as:

$$\begin{aligned} GCN_{\boldsymbol{\mu}}(\mathbf{X}^c, \mathbf{A}) &= \mathbf{H}\mathrm{ReLU}(\mathbf{H}\mathbf{X}^c\mathbf{W}^{(0)})\mathbf{W}_{\boldsymbol{\mu}}^{(1)} \\ GCN_{\boldsymbol{\sigma}}(\mathbf{X}^c, \mathbf{A}) &= \mathbf{H}\mathrm{ReLU}(\mathbf{H}\mathbf{X}^c\mathbf{W}^{(0)})\mathbf{W}_{\boldsymbol{\sigma}}^{(1)} \end{aligned} \tag{18}$$

where $\mathbf{H} = \tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}$ is the symmetrically normalized adjacency matrix and $\mathrm{ReLU}(\cdot) = \max(0, \cdot)$. $\mathbf{W}^{(0)}$ and $\mathbf{W}^{(1)} = \{\mathbf{W}_{\boldsymbol{\mu}}^{(1)}, \mathbf{W}_{\boldsymbol{\sigma}}^{(1)}\}$ are the weight matrices of the first and second layers of the GCN, respectively. The decoder adopts an inner product between the latent variables:

$$\begin{aligned} p(\mathbf{A} \mid \mathbf{Z}) &= \prod_{i=1}^{N}\prod_{j=1}^{N} p(a_{ij} \mid \mathbf{z}_i, \mathbf{z}_j) \\ p(a_{ij} \mid \mathbf{z}_i, \mathbf{z}_j) &= sig(\mathbf{z}_i^{\mathrm{T}}\mathbf{z}_j) \end{aligned} \tag{19}$$

where $i, j = 1, 2, 3..., N$ and the $sig(\cdot)$ is the logistic sigmoid function.

The loss function is designed to optimize the model parameters $\mathbf{W}^{(l)}$, and is defined as:

$$\mathrm{L} = \mathrm{E}_{q(\mathbf{Z}|\mathbf{X}^c,\mathbf{A})}[\log p(\mathbf{A} \mid \mathbf{Z})] - \mathrm{KL}[q(\mathbf{Z} \mid \mathbf{X}^c, \mathbf{A}) \,\|\, p(\mathbf{Z})] \tag{20}$$

The loss function consists of two parts. The first part is $\mathrm{E}_{q(\mathbf{Z}|\mathbf{X}^c,\mathbf{A})}[\log p(\mathbf{A} \mid \mathbf{Z})]$, which is used to measure the reconstruction error aiming to maintain the global relationships and dependencies between nodes. The second part $\mathrm{KL}[q(\mathbf{Z} \mid \mathbf{X}^c, \mathbf{A}) \,\|\, p(\mathbf{Z})]$ calculates the Kullback-Leibler divergence of $q(\mathbf{Z} \mid \mathbf{X}^c, \mathbf{A})$ and $p(\mathbf{Z})$, where $p(\mathbf{Z}) = \prod_i N(\mathbf{z}_i \mid 0, \mathbf{I})$ is the Gaussian prior. $\mathrm{KL}[q(\mathbf{Z} \mid \mathbf{X}^c, \mathbf{A}) \,\|\, p(\mathbf{Z})]$ enforces the distribution of the samples learned by the encoder being an approximation to the standard normal distribution, by measuring how well $q(\mathbf{Z} \mid \mathbf{X}^c, \mathbf{A})$ matches $p(\mathbf{Z})$. Full-batch gradient descent is used for training.

The adjacency matrix $\mathbf{A}$ and the three channel vectors $\{\mathbf{X}^1, \mathbf{X}^2, \mathbf{X}^3\}$ obtained in the previous steps are used to train VGAE. The fused embedded representation is denoted as $\mathbf{X} = (\mathbf{X}^1 + \mathbf{X}^2 + \mathbf{X}^3)/3$. Finally, the *k*-means algorithm is employed to classify the nodes into the changed or unchanged classes.

4. Experimental study

4.1 Introduction to datasets

Five sets of real bi-temporal SAR images were used in the experiments, namely, three extremely imbalanced datasets and two available benchmark datasets. The three extremely imbalanced datasets were collected by the COSMO-SkyMed SAR sensor at Guizhou Province, China in June 2016 and April 2017. The first of these three datasets, called dataset GZ-A, presents mainly some mountains and a river, as shown in Fig. 3. The second, called dataset GZ-B, is composed mainly of hills, plains and some buildings, as shown in Fig. 4. The third, called dataset GZ-C, exhibits mainly plains and hills, as shown in Fig. 5. The fourth dataset, San Francisco, records mainly the urban land coverage of San Francisco, the United States. The SAR images were captured by the ERS-2 SAR sensor satellite in August 2003 and May 2004, as shown in Fig. 6. The fifth dataset, Inland, is a scene of the Yellow River exhibiting a S-shaped bend, captured by the Radarsat-2 satellite in June 2008 and June 2009, as shown in Fig. 7. The corresponding ground reference map (GRM) was obtained by manual marking, where white represents the changed area and black represents the unchanged area. The pixel-level detailed information of all GRMs is listed in Table 1.

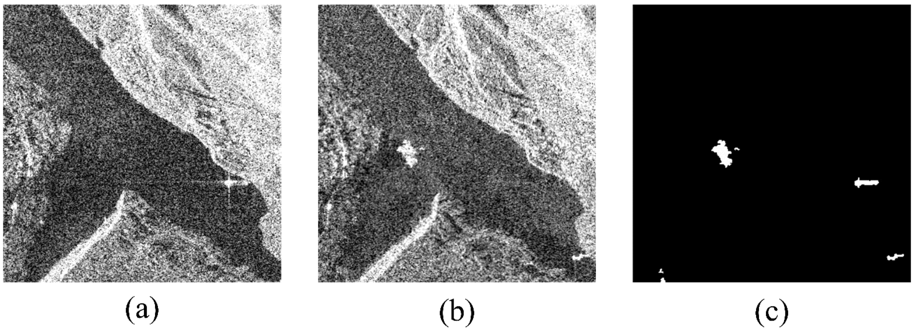


Fig. 3 GZ-A. (a) Acquired in April 2016, (b) Acquired in April 2017, (c) GRM.

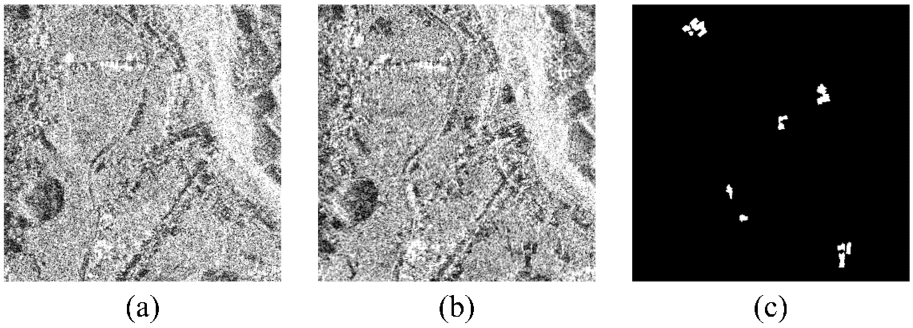


Fig. 4 GZ-B. (a) Acquired in April 2016, (b) Acquired in April 2017, (c) GRM.

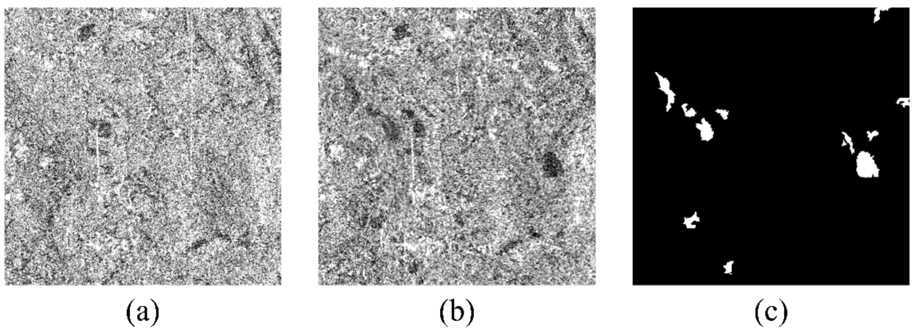


Fig. 5 GZ-C. (a) Acquired in April 2016, (b) Acquired in April 2017, (c) GRM.

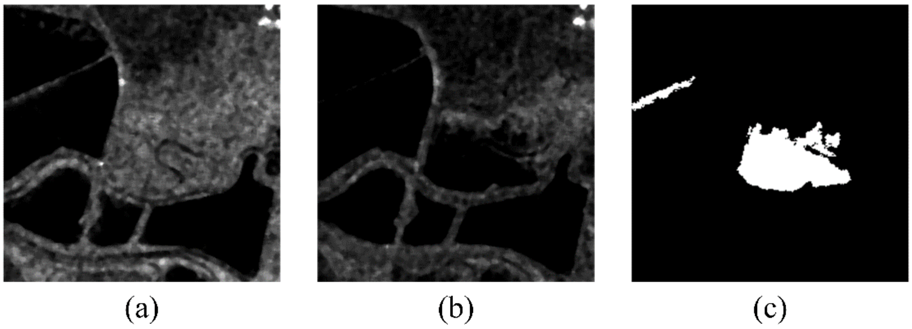


Fig. 6 San Francisco, (a) Acquired in August 2003, (b) Acquired in May 2004, (c) GRM.

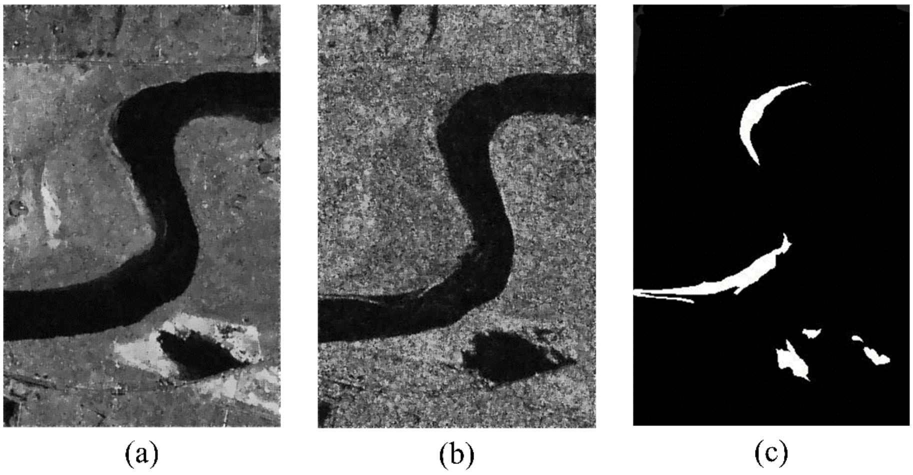


Fig. 7 Inland, (a) Acquired in June 2008, (b) Acquired in June 2009, (c) GRM.

Table 1. The details of experimental datasets. N_c and N_{uc} refer to the number of changed and unchanged pixels, respectively.

Datasets	size	N_c	N_{uc}	$N_c : N_{uc}$
GZ-A	400×400	1066	158934	1:149
GZ-B	400×400	1492	158508	1:106
GZ-C	400×400	3467	156533	1:45
Inland	443×291	4255	124658	1:30
San	256×256	4685	60851	1:13

It can be noted from Table 1 that the three image pairs of datasets GZ exhibit significant imbalances, that is, the number of changed pixels is much smaller than that of unchanged pixels. From Figs. 3 to 5 it can be seen that these three datasets suffer from strong speckle noise, making change detection very challenging. The other two datasets have a relatively balanced sample distribution and suffer from speckle noise pollution to a low degree. Therefore, these datasets were used for benchmark testing to test the generalization ability of the proposed method.

4.2 Evaluation criteria and experimental setting

The following indicators were adopted to evaluate the change detection methods: false alarm (FA) rate, missed detection (MD) rate, percentage correct classification (PCC), Kappa coefficient (KC), and F_1 score. True negative (TN) and true positive (TP), respectively, refer to the number of unchanged pixels and changed pixels classified correctly. False negative (FN) and false positive (FP), respectively, indicate the number of changed pixels and unchanged pixels that are misclassified.

(1) **FA**: The false alarm rate is given by:

$$P_{FA} = \frac{FP}{FP+TP} \times 100\% \tag{21}$$

(2) **MD**: The missed detection rate is calculated as:

$$P_{MD} = \frac{FN}{FN+TP} \times 100\% \tag{22}$$

(3) **PCC**: Accuracy of pixel-based classification can be expressed as:

$$PCC = \frac{TP+TN}{TP+FP+TN+FN} \tag{23}$$

(4) **KC**: Kappa coefficient is used for consistency checks, defined as:

$$KC = \frac{PCC-PRE}{1-PRE} \tag{24}$$

$$PRE = \frac{(TP+FN)\times(TP+FP)+(TN+FP)\times(TN+FN)}{(TP+FN+TN+FP)^2} \tag{25}$$

(5) **F_1**: F_1 score is an essential indicator of classification performance, which is defined as:

$$F_1 = \frac{2\times precision\times recall}{precision+recall}, precision = \frac{TP}{TP+FP},\ recall = \frac{TP}{TP+FN} \tag{26}$$

In the experiments, the hyperparameters were set as: the number of superpixels divided by SLIC $N = 5000$, and the control parameter : $\lambda = 0.1$, $\eta = 0.01$, $\alpha_1 = 0.5$, $\alpha_2 = 0.25$, $\alpha_3 = 0.25$. In the process of training VGAE, Adam was used for 200 iterations with a learning rate of 0.01. In addition, three training sets in $\{\mathbf{X}^1, \mathbf{X}^2, \mathbf{X}^3\}$ were fed sequentially to VGAE, and 18-dim hidden layer and 6-dim latent variables were used respectively. All experiments were implemented on a PC with a 3.3-GHz four-core CPU and 24-GB memory. The VGAE training were implemented with NVIDIA GeForce RTX 2080s GPU with 8-GB memory and PyTorch1.7.0.

4.3 Comparative experiments

Five pixel-based change detection methods were used in the comparative experiments, including: PCAKM (Celik., 2009), SLRDI+Gabor+FCM, NRELM (Gao et al., 2016), GFPCANet (Gao et al., 2016) and GFCWNN (Gao et al., 2019). Among them, SLRDI represents the filtered log ratio DI. Five object-based methods were developed for comparison. SLIC was used to perform superpixel segmentation on TCFDI, and the change detection task was transformed into the classification of the objects (superpixels). *K*-means clustering (KM) was used to classify directly the superpixels, that is, SLIC+KM. PCA, AE and SAE were used to perform feature learning on the superpixels to build low-dimensional embedding representations. We also evaluated the stacked contractive autoencoder (SCAE), a relatively new object-based change detection approach that uses SLIC to perform superpixel segmentation (Lv et al., 2018). Thus, there were four methods for superpixel classification using the KM algorithm: SLIC+PCA+KM, SLIC+AE+KM, SLIC+SAE+KM and SCAE. Three state-of-the-art evaluation methods, applied to the benchmark datasets, were utilized to improve the comparison of the four methods, including two pixel-based approaches: nonlocal low-rank PCA and two-level clustering (NLR-PCATLC) (Sun et al., 2020), fuzzy local information *c*-means based on multiple features (MFFLICM) (Meng et al., 2020), and one object-based method, heterogeneous graph (HG) (Wang et al., 2022). The proposed Nonlocal Learning-Based Small Area Change Detection (NLBSACD) framework adopted the forementioned two methods (Eq. 13 and 14) to build two graphs; the developed versions are NLBSACD^1 and NLBSACD^2. The experimental results on the five real SAR datasets are recorded in Tables 2 - 6, and the change detection maps are listed in Figs. 8 - 12.

Table 2. Comparative experimental results based on the GZ-A dataset. Best results are shown in bold.

	Methods	P_{FA} (%)	P_{MD} (%)	PCC (%)	KC (%)	F_1 (%)
Pixel-based	PCAK	96.45	**0.1**	82.61	5.69	6.86
	SLRDI+Gabor+KM	92.93	0.39	91.61	12.16	13.31
	NRELM	92.00	4.10	92.90	13.74	14.76
	GFPCANet	95.37	0.88	86.91	7.71	8.84
	GFCWNN	96.58	37.37	88.44	5.34	6.49
Object-based	SLIC+KM	96.79	1.56	80.97	5.04	6.22
	SLIC+PCA+KM	96.73	1.56	81.29	5.14	6.32
	SLIC+AE+KM	96.74	2.73	81.52	5.14	6.32
	SLIC+SAE+KM	97.15	2.93	78.82	4.35	5.54
	SCAE	96.19	2.72	75.90	4.07	5.26
	NLBSACD^1	**23.05**	32.58	99.66	71.69	71.86
	NLBSACD^2	23.51	28.58	**99.68**	**73.71**	**73.86**

Table 3. Comparative experimental results based on the GZ-B dataset. Best results are shown in bold.

	Methods	P_{FA} (%)	P_{MD} (%)	PCC (%)	KC (%)	F_1 (%)
Pixel-based	PCAK	96.76	**3.15**	73.04	4.56	6.28
	SLRDI+Gabor+KM	63.91	12.47	98.44	50.45	51.11
	NRELM	86.27	13.94	94.83	22.44	23.69

	GFPCANet	91.28	6.84	90.84	14.49	15.94
	GFCWNN	73.19	17.36	94.74	38.47	40.48
Object-based	SLIC+KM	95.29	7.04	82.38	7.32	8.96
	SLIC+PCA+KM	95.32	7.04	82.29	7.27	8.92
	SLIC+AE+KM	95.59	9.99	81.72	9.75	8.41
	SLIC+SAE+KM	95.56	95.84	81.78	6.81	8.47
	SCAE	96.26	6.29	77.41	5.49	7.19
	NLBSACD [1]	26.18	42.56	99.41	64.31	64.61
	NLBSACD [2]	**24.67**	42.09	**99.44**	**65.21**	**65.49**

Table 4. Comparative experimental results based on the GZ-C dataset. Best results are shown in bold.

	Methods	P_{FA} (%)	P_{MD} (%)	PCC (%)	KC (%)	F_1 (%)
Pixel-based	PCAK	90.59	6.95	80.44	13.70	17.10
	SLRDI+Gabor+KM	63.17	11.83	96.47	50.44	51.95
	NRELM	70.45	21.03	95.47	41.15	43.01
	GFPCANet	78.69	9.55	92.55	32.11	34.49
	GFCWNN	87.00	27.95	95.24	20.77	22.03
Object-based	SLIC+KM	85.41	7.67	88.11	22.28	25.19
	SLIC+PCA+KM	85.42	7.67	88.11	22.28	25.19
	SLIC+AE+KM	87.98	6.86	85.08	18.15	21.29
	SLIC+SAE+KM	89.89	**6.14**	81.76	14.92	18.24
	SCAE	90.72	8.65	80.44	13.43	16.85
	NLBSACD [1]	**14.94**	47.79	**98.76**	64.10	64.70
	NLBSACD [2]	23.73	40.21	98.73	**66.39**	**67.03**

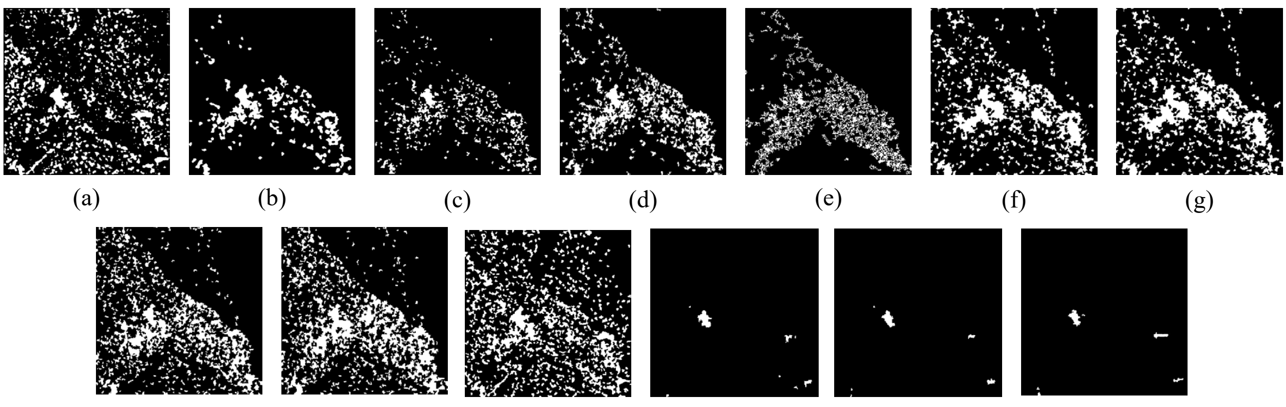


Fig. 8. Change detection maps of GZ-A dataset. (a) PCAK; (b) GFCM; (c) NRELM; (d) GFPCANet; (e) GFCWNN; (f) SLIC+KM; (g) SLIC+PCA+KM; (h) SLIC+AE+KM; (i) SLIC+SAE+KM; (j) SCAE; (k) NLBSACD[1]; (l) NLBSACD[2]; (m) GRM.

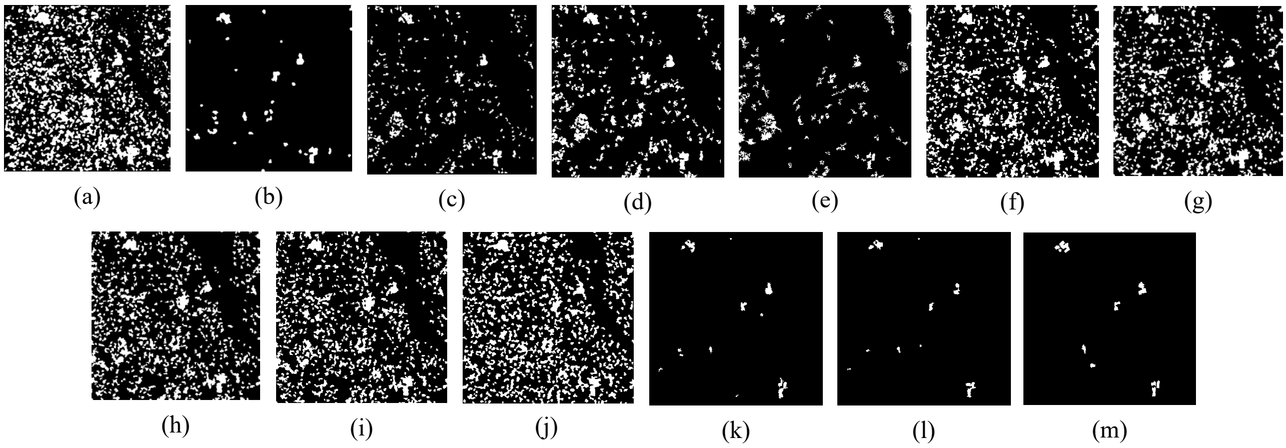


Fig. 9. Change detection maps of GZ-B dataset. (a) PCAK; (b) GFCM; (c) NRELM; (d) GFPCANet; (e) GFCWNN; (f) SLIC+KM; (g) SLIC+PCA+KM; (h) SLIC+AE+KM; (i) SLIC+SAE+KM; (j) SCAE; (k) NLBSACD^1; (l) NLBSACD^2; (m) GRM.

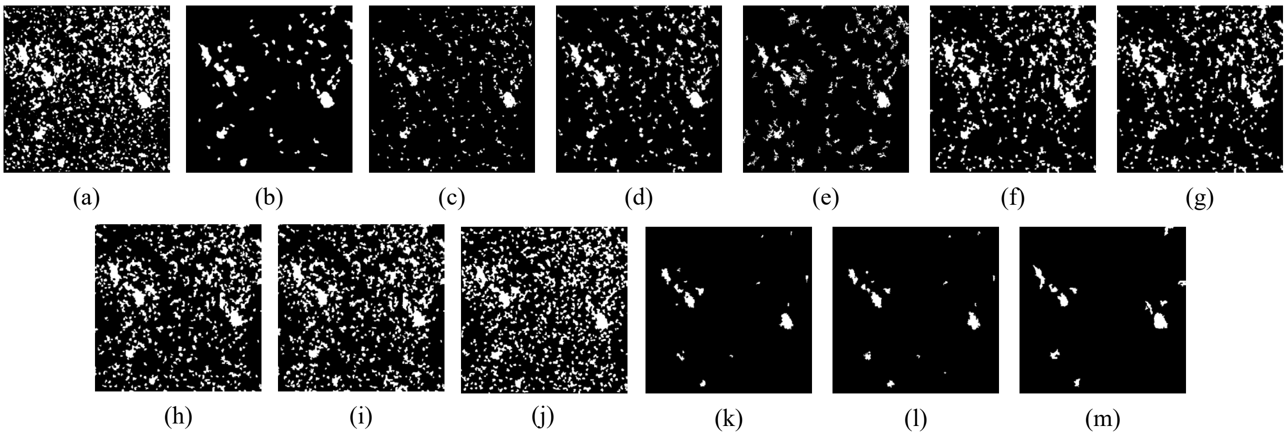


Fig. 10. Change detection maps of GZ-C dataset. (a) PCAK; (b) GFCM; (c) NRELM; (d) GFPCANet; (e) GFCWNN; (f) SLIC+KM; (g) SLIC+PCA+KM; (h) SLIC+AE+KM; (i) SLIC+SAE+KM; (j) SCAE; (k) NLBSACD^1; (l) NLBSACD^2; (m) GRM.

As can be seen from Tables 2, 3 and 4, the change detection map for datasets GZ-A, GZ-B, and GZ-C using NLBSACD are significantly more accurate than those produced by other methods, with the lowest FA (23.51%, 24.67% and 23.73%). It is worth noting that the results of those pixel-based methods are limited for the three GZ datasets, in which more than 60%, or even 90% of the detected changes are false alarms. This is due mainly to the following two reasons: 1) pixel-based methods are sensitive to speckle noise, and local patch-based methods have limited ability to suppress speckle noise. Both methods are prone to produce numerous false alarms when faced with strong speckle noise. 2) there exist significant imbalances in the GZ datasets, which bring great challenges to learning based on pixel-based and local patch-based methods.

Similarly, the object-based methods that do not consider nonlocal information, such as SLIC+AE+KM, SLIC+SAE+KM, and SCAE also produce low accuracy. The objective functions of PCA, AE, SAE and SCAE optimization are global in the process of learning and feature extraction, so it is hard for these methods to pay attention to the minority (changed) class features due to the significant imbalance. In this case, the key features belonging to the changed class will be regarded as redundant noise and discarded, which makes the learned latent representations no longer discriminative and leads to numerous false alarms.

The proposed NLBSACD approach, compared with other methods, can accurately learn key features, and the detection accuracy on the GZ datasets reached 99.68%, 99.44% and 98.76%. The experimental results show that NLBSACD has excellent noise robustness, which is mainly attributed to the connection between nonlocal superpixels during the feature learning process. It ensures that the key information of the changed samples is captured and removes speckle noise. In addition, VGAE establishes collaborative representation between homogeneous samples in the latent embedding space, which further enhances the discrimination between changed and unchanged samples. The numerical experiments on imbalanced datasets illustrates the effectiveness of NLBSACD for small area change detection.

Table 5. Comparative experimental results based on the San Francisco dataset. Best results are shown in bold.

	Methods	P_{FA} (%)	P_{MD} (%)	PCC (%)	KC (%)	F_1 (%)
Pixel-based	PCAK	75.38	3.88	78.69	31.40	39.20
	SLRDI+Gabor+KM	14.70	4.23	98.52	89.44	90.24
	NRELM	49.08	1.09	93.11	63.81	67.23
	GFPCANet	7.18	7.77	98.93	91.95	92.43
	GFCWNN	10.73	3.20	98.94	**92.06**	92.54
	NLR-PCATLC	8.06	8.00	98.85	91.35	91.94
Object-based	SLIC+KM	19.34	1.60	98.20	87.60	88.65
	SLIC+PCA+KM	19.34	1.60	98.20	87.68	88.65
	SLIC+AE+KM	25.97	2.59	97.37	82.72	84.13
	SLIC+SAE+KM	33.14	**0.92**	96.43	77.96	79.84
	SCAE	65.73	5.08	86.61	43.13	50.35
	HG	11.56	4.93	98.84	91.57	91.63
	NLBSACD [1]	5.58	10.50	98.87	91.29	91.89
	NLBSACD [2]	**4.04**	10.78	**98.96**	91.94	**92.67**

Table 6. Comparative experimental results based on the Inland dataset. Best results are shown in bold.

	Methods	P_{FA} (%)	P_{MD} (%)	PCC (%)	KC (%)	F_1 (%)
Pixel-based	PCAK	86.20	**6.04**	80.43	19.43	24.07
	SLRDI+Gabor+KM	19.54	25.88	**98.51**	76.12	76.58
	NRELM	19.67	28.77	98.47	74.73	75.51

	GFPCANet	**18.96**	27.78	98.21	75.62	76.38
	GFCWNN	29.62	13.68	98.35	**76.69**	**77.54**
	MFFLICM	31.53	13.27	98.47	73.24	76.46
Object-based	SLIC+KM	42.11	16.48	97.45	67.10	68.39
	SLIC+PCA+KM	40.29	17.13	97.58	68.19	69.41
	SLIC+AE+KM	42.71	13.81	97.42	67.53	68.81
	SLIC+SAE+KM	47.86	12.76	96.93	63.76	65.26
	SCAE	85.73	6.08	81.61	20.74	25.26
	NLBSACD [1]	19.47	30.60	98.44	73.75	74.55
	NLBSACD [2]	19.41	30.59	98.45	73.94	75.01

Benchmark tests were implemented on the datasets San Francisco and Inland, aiming to validate the performance of the proposed approach in general scenarios. It was found that the F_1 scores of NLBSACD on the two datasets reached 0.927 and 0.75, respectively. NLBSACD is competitive to the state-of-the-art pixel-based methods, and achieves a slightly higher detection accuracy than state-of-the-art object-based methods on the San Francisco dataset. The detection result of NLBSACD on the Inland dataset is slightly less accurate than that of GFCWNN, which is mainly because the superpixels produced by SLIC segmentation inevitably contain some heterogeneous pixels. The fine-grained basic unit used for the object-based methods is coarser compared with the pixel-based methods, leading to some misclassification. However, NLBSACD, as an object-based method, has more advantages in terms of efficiency for fine-resolution SAR images. The computational time of NLBSACD on the San Francisco dataset is 49.17s. In comparison, the time costs of GFCWNN and GFPCANet are 1023.42s and 831.68s. It is obvious that the proposed method is more efficient than pixel-based deep learning methods.

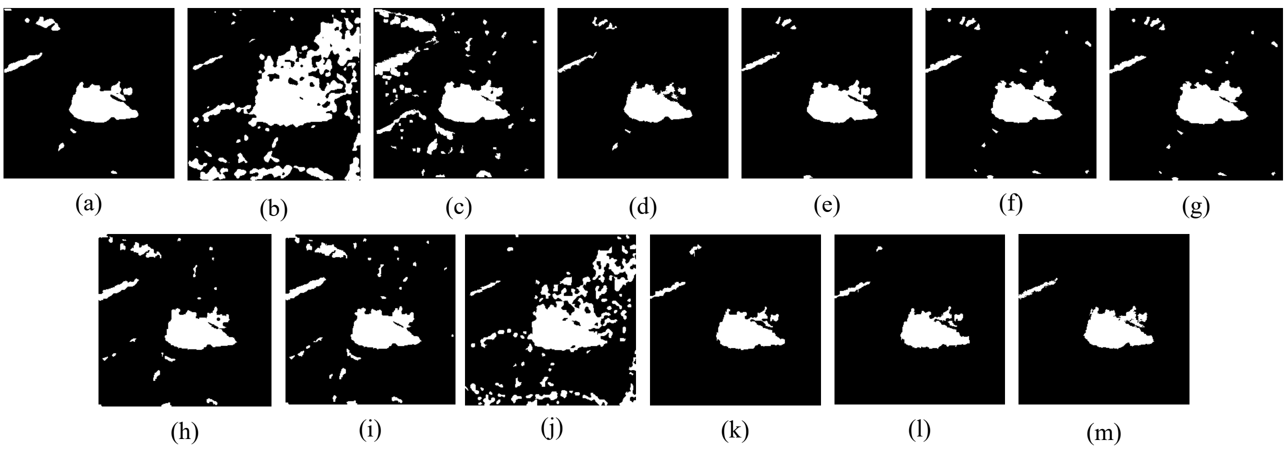


Fig. 11. Change detection maps of San Francisco dataset. (a) PCAK; (b) GFCM; (c) NRELM; (d) GFPCANet; (e) GFCWNN; (f) SLIC+KM; (g) SLIC+PCA+KM; (h) SLIC+AE+KM; (i) SLIC+SAE+KM; (j) SCAE; (k) NLBSACD[1]; (l) NLBSACD[2]; (m) GRM.

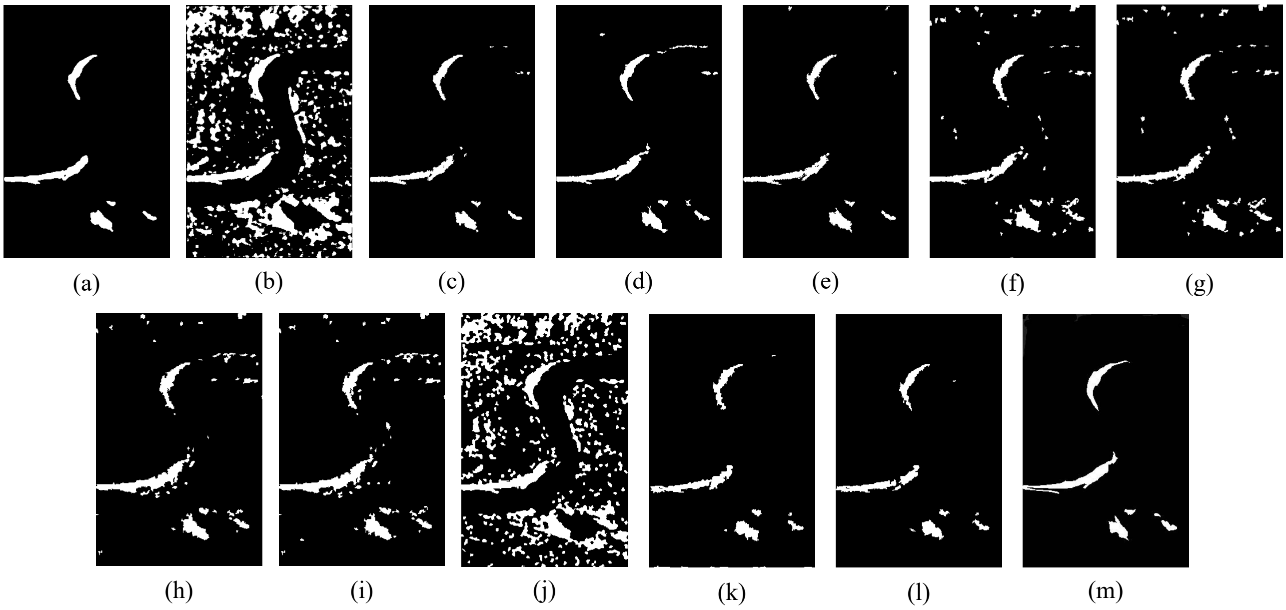


Fig. 12. Change detection maps of Inland dataset. (a) PCAK; (b) GFCM; (c) NRELM; (d) GFPCANet; (e) GFCWNN; (f) SLIC+KM; (g) SLIC+PCA+KM; (h) SLIC+AE+KM; (i) SLIC+SAE+KM; (j) SCAE; (k) NLBSACD^1; (l) NLBSACD^2; (m) GRM.

4.4 Ablation experiments

In the ablation experiments, CDI, LRDI and MSRDI were implemented in NLBSACD, rather than TCFDI. The F_1 score was used as the evaluation criteria, and the experimental results are exhibited in Fig. 13. NLBSACD using TCFDI achieved the highest score on the five datasets. The advantages of TCFDI lie mainly in the following three points: 1) Using TCFDI has better edge adhesion when performing superpixel segmentation, which is conducive to obtaining more homogeneous superpixels. 2) TCFDI is beneficial to establish a more accurate and reliable graph structure by fusing CDI, LRDI and MSRDI. 3) In the learning process of VGAE, TCFDI provides richer characteristic information, depending on the reliable graph structure to capture the key discriminative knowledge between the changed and unchanged classes. The experimental results demonstrate the effectiveness of TCFDI and its importance in NLBSACD.

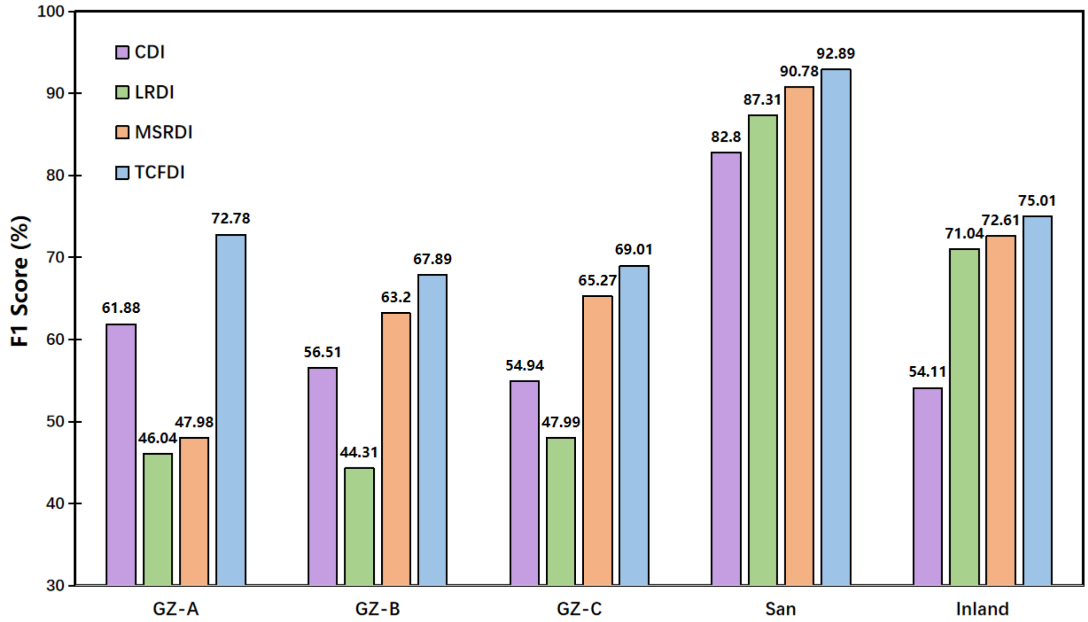
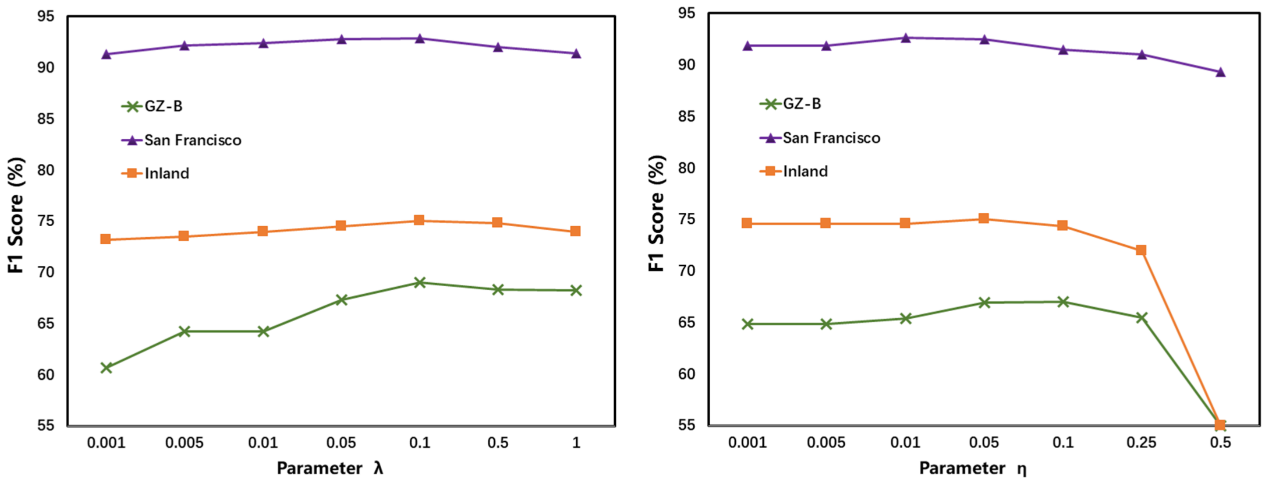


Fig. 13. Results of ablation experiments based on four kinds of Dis.

4.5 Analysis of parameters

The key parameters in NLBSACD are α, λ and η. As shown in Fig. 13, among MSRDI, CDI and LRDI, the former produced the greatest accuracy, so the contributing parameters were simply set as follows: $\alpha_1 = 0.5$, $\alpha_2 = 0.25$, $\alpha_3 = 0.25$.



(a) (b)

Fig. 14 Relationship between parameter λ, η and change detection accuracy.

In this section, a set of different values were used to construct the graph structure by GRBF for analyzing the parameter λ in depth. The relationship between the detection result and the parameter λ is shown in Fig. 14(a). It can be noticed that the control parameter λ has little effect on the Inland and San Francisco datasets, and the highest accuracy is achieved when $\lambda \in [0.1, 0.5]$. The highest accuracy is achieved when $\lambda = 0.1$ for dataset GZ-B. The experimental results show that the selection of 0.1 can build a reliable graph structure accurately, and can better describe the relationship between the superpixels of the DIs.

The developed similarity metric function in Eq. 14 was used to construct the graph to analyze the influence of the parameter η on the detection accuracy. As can be seen from Fig. 14, high F_1 scores are obtained for three datasets when $\eta \in [0.005, 0.1]$. Moreover, a rapid fall occurs when the value of the parameter η is greater than 0.25. The reason is that the similarity structure of the node will be destroyed when the similarity measurement focuses too much on geospatial position. Consequently, an unreliable graph structure is produced, which makes it difficult for VGAE to learn discriminative features of the changed and unchanged classes. Hence, the detection accuracy is naturally decreased.

5. Discussion

In this article, we developed a NLBSACD framework, and tested empirically in different datasets of SAR imagery with outstanding accuracy. The method itself is novel in several ways and we identify the following points to discuss further.

The proposed NLBSACD is an object-based approach, demonstrating strong robustness for discrete speckle noise. The GNN in NLBSACD, compared with early Hopfield Neural Network (HNN) (Tatem et al., 2001) and state-of-the-art CNN applied at per pixel level, is suitable to capture and identify irregular changes at an object level. Besides, NLBSACD increases the computational efficiency significantly. The basic unit in NLBSACD is at superpixel level, which reduces the number of recognition units than pixel-based approaches. For example, the 5K units of NLBSACD were classified, rather than the 160K units of pixels-based methods for the same SAR imagery. Meanwhile, we also notice some misclassification since within-object is not entirely homogeneous. Shrinking the size of superpixel (increasing the number of objects *N*) could increase the within-object homogeneity. Outside this paper, we found the F_1 score rose significantly as *N* increased when $N < 3500$. And, reached 0.91 and gradually tended to increase slightly when *N* increasing from 3500 to 5000 on the San Francisco datasets, with similar phenomenon occurring on the Inland dataset. However, increasing *N* resulted in huge computational cost of building graph and VGAE training. Indeed, the value of *N* can be adjusted based on practical application requirements and the computational power etc.

The reliability and accuracy of the graph structure is shown by comparing the two versions of NLBSACD. The proposed GRBF constrained by geospatial distance, compared with standard GRBF, established a more generalized connection between objects by combining feature information and geospatial correlation. These analyses further motivates the consideration of optimizing the graph structure, which can be explored from the following two aspects. On one hand, the relationships between nodes can be modeled mathematically using prior knowledge and spatial characteristics. On the other hand, supervised or semi-supervised strategy can be introduced to automatically update the graph structure to establish more reliable and accurate relationships amongst different nodes.

Although the median value filling strategy may introduce a small amount of noise, it hardly affects the accuracy of change detection. It was found in the experiments that the elements of the feature vector are highly homogeneous in intensity and the discrepancy between feature vector length is small, benefitting from SLIC superpixel segmentation. Thus, the amount of noise introduced is small. Further, the constructed embedded feature representation, randomly sampled from the learned distribution, has strong noise robustness and stability. Therefore, the introduction of a small amount of noise will not degrade the discrimination of the constructed embedding representation.

Finally, the proposed scheme exhibits excellent change detection performance on five real SAR datasets with significant differences. We would like to further extend the proposed method to other application fields, such as target identification (Tatem et al., 2002) and PolSAR image classification (Zou et al., 2018; Tang et al., 2021), as well as change detection in optical sensor imagery such as Landsat and Sentinel-2 satellite images.

6. Conclusion

In this paper, we developed a VGAE-based approach to learn nonlocal features for bi-temporal SAR image change detection. A three-channel fused difference image, called TCFDI, was used to obtain homogeneous superpixel objects with SLIC for presentation to subsequent modules. The TCFDI integrates the advantages of the three DIs to ensure the accuracy of superpixel segmentation and maintain abundant characteristic information of objects. Crucially, the GRBF combining the intensity spatial feature and the visual spatial position information was proposed to establish the graph structure between superpixels, which laid the foundation for graph-based learning. Nonlocal feature learning using VGAE was able to suppress speckle noise effectively, and build discriminative high-level representations in latent space, leading to superior accuracy and robustness compared to a range of benchmark local feature learning methods. Numerical experimental results confirmed the effectiveness and robustness of the proposed approach for small area change detection, especially where imbalance exists. Moreover, it maintains competitive detection accuracy in general scenarios, illustrating the practical value for SAR remote sensing application.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 61301224) and the Natural Science Foundation of Chongqing (Grant No. cstc2021jcyj-msxmX0174). Dr Ce Zhang was supported in part by the Natural Environment Research Council (Grant No. NE/T004002/1).

Declaration of Competing Interest

The authors declare that they have no conflicts of interest to disclose.

Reference

Bazi, Y., Bruzzone, L., Melgani, F., 2005. An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. IEEE Geosci. Remote Sens. Lett. 43(4), 874-887.

Cai, L., Li, J., Wang, J., Ji, S., 2021. Line graph neural network for link prediction. IEEE Trans. Pattern Anal. Mach. Intell. Early Access.

Celik, T., 2009. Unsupervised change detection in satellite images using principal component analysis and k-means clustering. IEEE Geosci. Remote Sens. Lett. 6(4), 772-776.

Cheng, G., Yang, C., Yao, X., Guo. L., Han, J., 2018. When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs. IEEE Trans. Geosci. Remote Sens. 56(5), 2811-2821.

Dai, Y., Jin, T., Li, H., Song, Y., Hu, J., 2021. Imaging enhancement via CNN in MIMO virtual array-based radar. IEEE Trans. Geosci. Remote Sens. 59(9), 7449-7458.

Defferrard, M., Bresson, X., Vandergheynst, P., 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In: Adv. Conf. Neural Netw. Inf. Proc. Syst. (NIPS). 29.

Dong, H., Ma, W., Jiao, L., Liu F., Li, L., 2022. A Multiscale Self-Attention Deep Clustering for Change Detection in SAR Images. IEEE Transactions on Geoscience and Remote Sensing. 60, 1-16.

Gao, F., Dong, J., Li. B., Xu, Q., Xie, C., 2016. Change detection from synthetic aperture radar images based on neighborhood-based ratio and extreme learning machine. J. Appl. Remote Sens, 10(4), 046019.

Gao, F., Dong, J., Li, B., Xu, Q., 2016. Automatic change detection in synthetic aperture radar images based on PCANet. IEEE Geosci. Remote Sens. Lett. 13(12), 1792–1796.

Gao, F., Wang, X., Gao, Y., Dong, J., Wang, S., 2019. Sea ice change detection in SAR images based on convolutional-wavelet neural networks. IEEE Geosci. Remote Sens. Lett. 16(8), 1240–1244.

Gong, M., Cao, Y., Wu, Q., 2012. A neighborhood-based ratio approach for change detection in SAR images. IEEE Geosci. Remote Sens. Lett. 9(2), 307-311.

Gong, M., Yu, L., Jiao, L., Jia, M., Su, L., 2014. SAR change detection based on intensity and texture changes. ISPRS J. Photogramm. Remote Sens. 93, 123–135

Gong, M., Yang, H., Zhang, P., 2017. Feature learning and change feature classification based on deep learning for ternary change detection in SAR images. ISPRS J. Photogramm. Remote Sens. 129, 212-225.

Grover, A., Zweig, A., Ermon, S., 2019. Graphite: iterative generative modeling of graphs. In: Int. Conf. Mach. Learn (ICML). 97.

Han, H., Chen, Y., Hsiao, P., Fu, L., 2021. Using channel-wise attention for deep CNN based real-time semantic segmentation with class-aware edge information. IEEE Trans. Intell. Transp. Syst. 22(2), 1041-1051.

Hussain, M., Chen, D., Cheng, A., Wei, H., Stanley, D., 2013. Change detection from remotely sensed images: From pixel-based to object-based approaches. ISPRS J. Photogramm. Remote Sens. 80, 91-106.

Jia, L., Li, M., Zhang, P., Wu, Y., Zhu, H., 2016. SAR image change detection based on multiple kernel k-means clustering with local-neighborhood information. IEEE Geosci. Remote. Sens. Lett.13(6), 856-860.

Jaswanth, A., Gupta, N., Mishra, A. K., Hum, Y. C., 2022. Change Detection of SAR images based on Convolution Neural Network with Curvelet Transform. 2022 2nd International Conference on Artificial Intelligence and Signal Processing (AISP). 1-6.

Kipf, T. N., Welling, M., 2016. Semi-supervised classification with graph convolutional networks. In: Int. Conf. Learn. Representations. (ICML).

Kipf, T. N., Welling, M., 2016. Variational graph auto-encoder. In: 2016 NIPS workshop on Bayesian deep learning.

Li, H, C., Celik, T., Longbotham, N., Emery, W, J., 2015. Gabor feature based unsupervised change detection of multitemporal SAR images based on two-level clustering. IEEE Geosci. Remote Sens. Lett. 12(12), 2458-2462.

Li, Y., Peng, C., Chen, Y., Jiao, L., Zhou, L., Shang, R., 2019. A deep learning method for change detection in synthetic aperture radar images. IEEE Trans. Geosci. Remote Sens. 57(8), 5751–5763.

Li, L., Wang, C., Zhang, H., Zhang, B., Wu, F., 2019. Urban building change detection in SAR images using combined differential image and residual U-Net network. Remote Sens. 11, 1091.

Li, C., Xia, W., Yan, Y., Luo, B., Tang, J., 2021. Segmenting object in day and night: Edge-conditioned CNN for thermal image semantic segmentation. IEEE Trans. Intell. Neural Netw. Learn. Syst. 32(7), 3069-3082.

Liu, R., Ma, L., Wang, Y., Zhang, L., 2021. Learning converged propagations with deep prior ensemble for image enhancement. IEEE Trans. Image Proces. 28(3): 1528-1543.

Liu, G., Li, L., Jiao, L., Dong, Y., Li, X., Stacked Fisher auto-encoder for SAR change detection. Pattern Recognit. 96 (106971).

Lu, D., Moran, E., Hetrick, S., 2011. Detection of impervious surface change with multitemporal Landsat images in an urban-rural frontier. ISPRS J. Photogramm. Remote Sens. 66, 298-306.

Lv, N., Chen, C., Qiu, T., Sangai, A., 2018. Deep learning and superpixel feature extraction based on contractive auto-encoder for change detection in SAR images. IEEE Trans. Ind. Inform. 14(12), 5530-5538.

Muster, S., Langer, M., Abnizova, A., Young, K., Boike, J., 2015. Spatio-temporal sensitivity of MODIS land surface temperature anomalies indicates high potential for large-scale land cover change detection in Arctic permafrost landscapes. Remote Sens. Environ. 168, 1-12.

Meng, W., Wang, L., Du, A., Li, Y., 2020. SAR Image Change Detection Based on Data Optimization and Self-Supervised Learning. IEEE Access. 8, 217290-217305.

Pantze, A., Santoro, M., Fransson, J., 2014. Change detection of boreal forest using bi-temporal ALOS PALSAR backscatter data. Remote Sens. Environ. 155, 120-128.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: IEEE Conf. Comput. Vis. Pattern Recognit (CVPR). 779-788.

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39(6), 1137-1149.

Salha, G., Hennequin, R., Tran, V., Vazirgiannis, M., 2019. In: 28-th Int. Jiont Conf. Artificial Intelligence. (IJCAI).

Sun, Y., Lei, L., Guan, D., Li, X., Kuang, G., 2020. SAR Image Change Detection Based on Nonlocal Low-Rank Model and Two-Level Clustering. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 13, 293-306.

Tajbakhsh, N., Shin, J., Gurudu, S., Hurst, R.T., Kendall, C., Gotway, M., Liang, J., 2016. Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Medical. Imaging. 35(5), 1299-1312.

Tatem, A., Lewis, H., Atkinson, P., Nixon, M., 2002. Super-resolution land cover pattern prediction using a Hopfield neural network. Remote Sens. Environ. 79, 1-14.

Tatem, A., Lewis, H., Atkinson, P., Nixon, M., 2001. Super-resolution target identification from remotely sensed images using a Hopfield neural network. IEEE Trans. Geosci. Remote Sens. 39(4), 781-796.

Tang, R., Xu, X., Yang, R., Gui, R., 2021. Deep graph cluster based unsupervised representation learning for PolSAR image classification. In: 2021 IEEE Int. Geosci. Remote Sens. Symp. 4252-4255.

Wang, C., Pan, S., Long, G., Zhu, X., Jiang, J., 2017. MGAE: marginalized graph auto-encoder for graph clustering. In: 2017 ACM. Conf. Info. Knowledge. Management. 889-898.

Wang, J., Yang, X., Yang, X. Jia, L., Fang, S., 2020. Unsupervised change detection between SAR images based on hypergraphs. ISPRS J. Photogramm. Remote Sens. 164, 61–72.

Wang, G., Zuluaga, M., Li, W., Pratt, R., Patel, P., Aertsen, M., Doel, T., David, A., Deprest, J., Ourselin, S., Vercauteren, T., 2019. DeeplGeoS: a deep interactive geodesic framework for medical image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 41(7), 1559-1572.

Wang, J., Zhang, A., 2022. SAR Image Change Detection Based on Heterogeneous Graph With Multiattributes and Multirelationships. IEEE Access. 10, 44347-44361.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P., 2021. A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems. 32(1), 4-24.

Yang, L., Cheung, N-M., Li, J., Fang, J., 2019. Deep clustering by gaussian mixture variational auto-encoders with graph embedding. In: 2019 IEEE/CVF Int. Conf. Comput. Vis (ICCV). 6439-6448.

Zhang, P., Gong, M., Su, L., Liu, J., Li, Z., 2016. Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images. ISPRS J. Photogramm. Remote Sens. 116, 24-41.

Zhang, X., Su, H., Zhang, C., Gu, X., Tan, X., Atkinson, P., 2021. Robust unsupervised small area change detection from SAR imagery using deep learning. ISPRS J. Photogramm. Remote Sens. 173, 79-94.

Zhang, X., Chen, J., Meng, H., 2013. A novel SAR image change detection based on graph-cut and generalized gaussian model. IEEE Geosci. Remote Sens. Lett. 10(1), 14-18.

Zhang, W., Jiao, L., Liu, F., Yang, S., Song, W., Liu, J., 2022. Sparse Feature Clustering Network for Unsupervised SAR Image Change Detection. IEEE Transactions on Geoscience and Remote Sensing. 60, 1-13.

Zheng, Y., Zhang, X., Hou, B., Liu, G., 2014. Using combined difference image and *k*-means clustering for SAR

image change detection. IEEE Geosci. Remote Sens. Lett. 11(3), 691-695.
Zou, H., Shao, N., Li, M., Chen, C., Qin, X., 2018. Superpixel-based unsupervised classification of PolSAR images with adaptive number of terrain classes. In: 2018 IEEE Int. Geosci. Remote Sens. Symp. 2390-2393.
Zhuang, H., Tan, Z., Deng, K., Yao, G., 2020. Adaptive generalized likelihood ratio test for change detection in SAR image. IEEE Geosci. Remote Sens. Lett. 17(3), 416-420.
Zhuang, H., Fan, H., Deng, K., Yao, G., 2018. A spatial-temporal adaptive neighborhood-based ratio approach for change detection in SAR images. Remote Sens. 10(8), 1295.