



Article

A Dual Neighborhood Hypergraph Neural Network for Change Detection in VHR Remote Sensing Images

Junzheng Wu ^{1,2}, Ruigang Fu ¹, Qiang Liu ², Weiping Ni ^{2,*}, Kenan Cheng ², Biao Li ¹ and Yuli Sun ¹

¹ College of Electronic Science, National University of Defense Technology, Changsha 410073, China

² Northwest Institute of Nuclear Technology, Xi'an 710024, China

* Correspondence: niweiping@nint.ac.cn; Tel.: +86-29-84767440

Abstract: The very high spatial resolution (VHR) remote sensing images have been an extremely valuable source for monitoring changes occurring on the Earth's surface. However, precisely detecting relevant changes in VHR images still remains a challenge, due to the complexity of the relationships among ground objects. To address this limitation, a dual neighborhood hypergraph neural network is proposed in this article, which combines multiscale superpixel segmentation and hypergraph convolution to model and exploit the complex relationships. First, the bi-temporal image pairs are segmented under two scales and fed to a pre-trained U-net to obtain node features by treating each object under the fine scale as a node. The dual neighborhood is then defined using the father-child and adjacent relationships of the segmented objects to construct the hypergraph, which permits models to represent higher-order structured information far more complex than the conventional pairwise relationships. The hypergraph convolutions are conducted on the constructed hypergraph to propagate the label information from a small amount of labeled nodes to the other unlabeled ones by the node-edge-node transformation. Moreover, to alleviate the problem of imbalanced sampling, the focal loss function is adopted to train the hypergraph neural network. The experimental results on optical, SAR and heterogeneous optical/SAR data sets demonstrate that the proposed method offers better effectiveness and robustness compared to many state-of-the-art methods.

Keywords: change detection; hypergraph neural network; dual neighborhood; semisupervised; remote sensing images



Citation: Wu, J.; Fu, R.; Liu, Q.; Ni, W.; Cheng, K.; Li, B.; Sun, Y. A Dual Neighborhood Hypergraph Neural Network for Change Detection in VHR Remote Sensing Images. *Remote Sens.* **2023**, *15*, 694. <https://doi.org/10.3390/rs15030694>

Academic Editors: Gangyao Kuang, Xin Su and Olga Sykoti

Received: 23 November 2022

Revised: 18 January 2023

Accepted: 19 January 2023

Published: 24 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As a vital branch of remote sensing interpretation, change detection (CD) that aims at identifying the changes in images collected in the same area but at different times [1] has been widely applied in various fields, including urban planning, damage assessments, and environmental monitoring [2]. With the rapid development of Earth observation technology, very high spatial resolution (VHR) remote sensing images are now available, and they can provide abundant surface details and spatial distribution information. Consequently, CD with VHR images has drawn more attention. In the meantime, it remains a tough and challenging task [3].

In the last few decades, numerous CD methods have been developed for diverse circumstances, and they can be divided into pixel-based and object-based methods roughly according to the basic unit of processing [4]. The pixel-based CD (PBCD) methods employ the individual pixel as the basic analysis unit, and many widely-used methods belong to this category. For optical images CD, change vector analysis (CVA) [5] and its extensions, such as robust CVA (RCVA) [6], are the most commonly-used ones which can provide change intensity and change direction. Some transform-based methods, such as principal component analysis (PCA) [7] and slow feature analysis (SFA) [8], are also popular. To suppress the interference of speckle noise, ratio-based approaches [9,10] were devised for SAR CD. Some statistics-based methods [11] used specific distributions to model SAR

images and then obtained change maps through estimating the posteriori probabilities. Due to the great practical significance for the immediate evaluation and emergency disasters, CD for heterogeneous optical/SAR images has also attracted much attention. For instance, Liu et al. [12] defined a pixel-based transformation, which can transform the SAR and optical images into a common space, after which difference analysis can be conducted in the transformed space. Ferraris et al. [13] used the coupled dictionary learning framework to model the two heterogeneous images. To capture the information of the adjacent pixels, numerous methods used a sliding window to obtain patches as processing units [14]. In addition, various advanced techniques have been introduced into PBCD, such as Markov random fields [15], extreme learning machine (ELM) [16], and dictionary learning [17]. However, these PBCD methods are limited on VHR images, because the assumption of pixel independence is unreliable. When using a sliding window, it is challenging to obtain the most appropriate size. In addition, a fixed window can hardly represent a meaningful ground object. Furthermore, registration error and radiometric correction have a significant influence on the results [18].

With the spatial resolution increasing, the highly spectral variability and the difficulty of modeling the contextual information may further weaken the performance of PBCD methods. Under these circumstances, object-based CD (OBCD) techniques provide a unique way to overcome these limitations by using an object as a basic processing unit. An object (so-called parcel or superpixel in some literature) is a group of local pixel clusters that are obtained through segmentation using spectral, texture, and geometric (such as shape) features, as well as other information. A large number of OBCD methods have been proposed in the last decade. For instance, Yousif and Ban [19] developed a novel OBCD method for VHR SAR images, which can preserve meaningful detailed change information and mitigate the influence of noise. Multiple feature fusion strategy was designed to improve the performance of OBCD in [20]. Lv et al. [21] proposed an object-oriented key point vector distance to measure the changed degree for VHR images, which can reduce the number of pseudo changed points. These methods have shown extraordinary potential to weaken the effects of spectral variability, spatial georeferencing, and acquisition characteristics [22]. Nevertheless, most OBCD methods generally only use some hand-crafted features, which require much domain-specific knowledge and may be affected by noise and atmospheric conditions. Furthermore, to ensure the homogeneity within each object, most OBCD methods make the objects prone to over-segmentation, and the boundary fragmentation may lead to inadequate semantic integrity. It implies that the complex interactions among objects need to be further exploited, and this is a vital motivation of our method.

Recently, deep learning techniques have been employed in CD tasks, including both supervised/semi-supervised and unsupervised ways. Various models have been introduced into CD, such as convolutional neural networks (CNN), encoder [23] and transformer [24,25]. Among all deep learning models, CNN is the most widely used in CD tasks. For instance, Liu et al. [26] designed a Siamese auto-encoder framework for CD in optical aerial images, which can achieve transfer learning among different data sets. Chen et al. [27] presented a novel spatial-temporal attention neural network based on Siamese and designed a CD self-attention mechanism to calculate attention weights between any two pixels at various times and positions. Wu et al. [28] proposed a deep kernel PCA convolution to extract representative features for multi-temporal VHR images, and then built an unsupervised Siamese mapping network for binary and multiclass CD. An end-to-end CD method that combines a pixel-based convolutional network and superpixel-based operations was proposed in [29], which promoted the collaborative integration of OBCD and deep learning. Wang et al. [30] designed a fully convolutional Siamese network and improved the focal contrastive loss, which can reduce intra-class variance and enlarge inter-class difference. Additionally, some CNN-based generative adversarial networks (GAN) have been designed [31,32] for CD tasks in VHR images. Although these methods can obtain high performance under some circumstances, some weaknesses still

need to be addressed. First, CNNs assign weights to each pixel independently through convolution kernels, and it means that the complicated interactive relationships among pixels are ignored, which may result in insufficiently characterization of the ground objects. Second, CNNs tend to excel at extracting multi-level features from data with structures of regular rectangles, whereas the shapes of superpixels are usually irregular. Therefore, the geometrical information of superpixels may be inadequately exploited.

In view of the aforementioned issues, in this article we propose an object-level semi-supervised hypergraph neural network (HGNN) framework for VHR image CD tasks, acting upon both homogeneous and heterogeneous remote sensing images. The input images are firstly segmented under a fine scale and a coarse one, respectively. Treating each object under the fine scale as one node, the features of nodes can be obtained by combining segmentation under the fine scale and the outputs of the feature extractor, which is a pre-trained U-net. The hypergraph is then constructed by defining the dual neighborhood, to capture more comprehensive structural information among objects. Several hypergraph convolution layers are then sequentially conducted on the hypergraph to propagate information among nodes through the hypergraph structure. As changed and unchanged samples are usually imbalanced in CD, the focal loss function is adopted when training the network.

The main contributions of this article are as follows.

- (1) A novel dual neighborhood hypergraph neural network (DNHGNN) framework for CD is proposed, which can adequately exploit the complex relationships and interacting information of ground objects that commonly exist in VHR remote sensing images. To the best of our knowledge, it is the first HGNN-based method in the field of remote sensing CD.
- (2) A dual neighborhood is defined, which contains a spatial neighborhood according to the adjacent relationships of object under the fine scale and a structural neighborhood according to the father-child relationships between scales. Based on the dual neighborhood, reliable hyperedges can be obtained, which better represent the complicated interactive relationships of ground objects in VHR remote sensing images.
- (3) The multiscale object-based technique is integrated into hypergraph construction, which not only yields to high-level node features as inputs of HGNN, but also substantially reduces the number of nodes, making the hypergraph convolution on the image scene feasible and efficient.

The article is organized as follows. The related works are discussed in Section 2. In Section 3, we illustrate the proposed DNHGNN in detail. Then, the experimental results and discussions are reported in Section 4. Finally, the conclusion and future work are presented in Section 5.

2. Related Work

In this section, we provide a brief review of graph/hypergraph-based CD methods and graph neural networks, as they relate to this article.

2.1. Graph/Hypergraph-Based Change Detection

Figure 1 shows an example of an ordinary graph and hypergraph. The main difference is that the graph is represented using the adjacency matrix, in which each edge connects just two vertices, while the hypergraph can be represented by the incidence matrix, in which the hyperedges can encode high-order correlations by connecting more than two vertices.

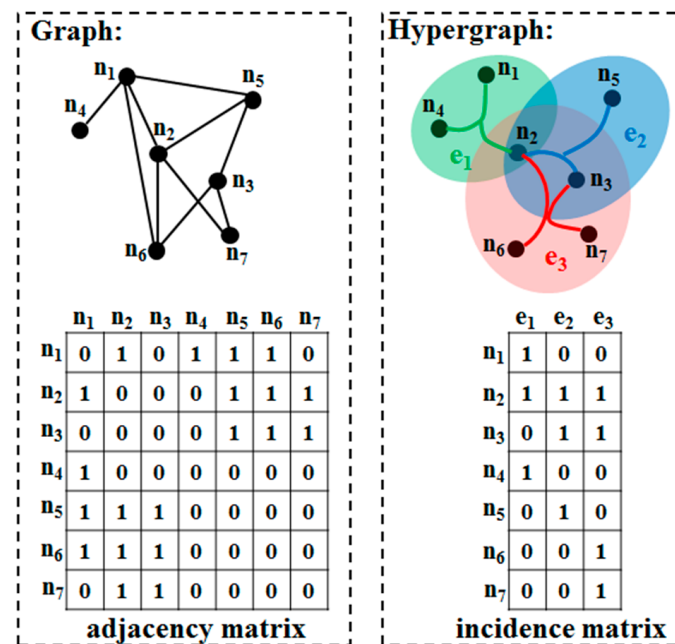


Figure 1. An example of ordinary graph and hypergraph.

The VHR images are usually full of structural features, and the graph/hypergraph models are feasible in representing these features. Thus, they have been utilized in CD. Pham et al. [33] proposed a weighted graph to capture contextual information of a set of characteristic points and measured the changed level by considering the coherence of the information carried by the two images. A hierarchical spatial-temporal graph kernel was designed to utilize the local and global structures for SAR image CD [34]. An object-based graph model is proposed for both optical and SAR image CD in [35]. Sun et al. [36] constructed a nonlocal patch similarity graph to establish a connection between heterogeneous images, and then determined the change level by measuring how much the graph structure of one image still conforms to that of the other image. As research into graph-based CD methods expands, several works have attempted to introduce hypergraph models into CD tasks. For instance, Wang et al. [37] formulated the CD as the problem of hypergraph matching and hypergraph partition, and constructed the hyperedges on the pixels and their coupling neighbors using the K nearest neighbor rule.

Although the aforementioned graph/hypergraph-based methods have validated their availability in many cases, some limitations still exist. In particular, they only utilize some low features (e.g., spectrum, intensity, texture), which may not only be affected by noise and atmospheric conditions, but also contain insufficiently semantic information. In addition, the complicated structural relationships in VHR images have not been fully exploited when constructing the graph/hypergraph models.

2.2. Graph/Hypergraph Neural Network

Motivated by CNNs, recurrent neural networks (RNNs), and graph embedding, many works have attempted to use deep learning methods to handle the data with graph structure. In such a situation, graph neural networks (GNNs) [38,39] emerge. The core ideas of GNN are graph structure and message propagations. To date, numerous GNN-based techniques have been developed, which can be divided into five categories: graph convolutional networks (GCN), graph attention networks, graph autoencoders, graph generative networks and graph spatial-temporal networks.

In the applications of image processing, GCN is the most widely-used and representative one among the above categories. Similar to the convolutional operations in CNN, GCN obtains new embedding for a centroid node through defining convolutional operations to aggregate information from its neighborhood. The convolutional operations in GCN can be divided into two categories: spectral-based and spatial-based approaches. This article adopts a spectral-based approach to carry out convolutions for data with graph structure. Bruna et al. [40] used the spectral graph theory to develop convolutional operations on graphs, which were defined based on the graph Laplacian matrix. To date, there have been numerous extensions on spectral-based approaches [41,42]. To reduce the computational complexity of eigendecomposition in the spectral-based approaches, ChebNet [43] and CayleyNet [44] approximated the decomposition by special polynomials. Kipf and Welling [45] further assumed that the order in ChebNet is 1, and proposed a fast layer-wise GCN. Since then, many works have made improvements over GCN from different aspects, such as the contextual graph Markov model (CGMM) [46] and graph wavelet neural network (GWNN) [47]. Due to the effective capability of learning representations from graph data, GCN has shown superior performance in many tasks of the RS community, such as classification in hyperspectral images [48,49]. In the last two years, several CD methods based on GCN have emerged. Saha et al. [50] first utilized GCN for semi-supervised CD tasks. In our previous works [51], a multiscale GCN for CD was proposed, which can comprehensively incorporate the information from several distinct scales. Tang et al. [52] combined GCN and metric learning to develop an unsupervised method for CD.

With each edge representing the interaction between two nodes, traditional GNN can effectively represent the pairwise relationships. However, in many real applications, such as social connections and recommendation systems, the relationships among objects are much more complex than pairwise. Under such circumstances, traditional GNN has intrinsic limitations when modeling these relationships. To tackle this challenging issue, Feng et al. [53] recently proposed the hypergraph neural network (HGNN), which used the hypergraph structure for data modeling, after which a hypergraph convolution operation was designed to better exploit the high-order data correlation for representation learning. An improvement named dynamic hypergraph neural networks (DHGNN) was proposed, which modified the hypergraph structure from adjusted feature embedding [54].

Based on the observations that relationships of ground objects in VHR remote sensing images may be beyond pairwise connections and even more complicated, in this article, we propose an HGNN framework for VHR image CD, which aims at improving performance by better exploiting the relationships among objects.

3. Proposed Method

In this section, the proposed method is illustrated in details, as shown in Figure 2. Firstly, we stack the two input images into one and then adopt fractal net evolution approach (FNEA) to segment the stacked image under a fine scale and a coarse scale. Meanwhile, the stacked image is fed into the feature extractor, which is a pre-trained U-net, to produce the high-level feature maps with the same lengths and heights as the input images. After that, the object-wise features of nodes are obtained by combining the segmentation under the fine scale and the high-level feature maps. The hypergraph can be constructed using the adjacent relationships under the fine scale and the father-child relationships between the two scales. Then, sequential hypergraph convolutional layers are conducted on the hypergraph, which exploit the high-order relationships carried by the hypergraph structure to cluster the objects potentially belonging to the same class (changed/unchanged) together in the embedding space.

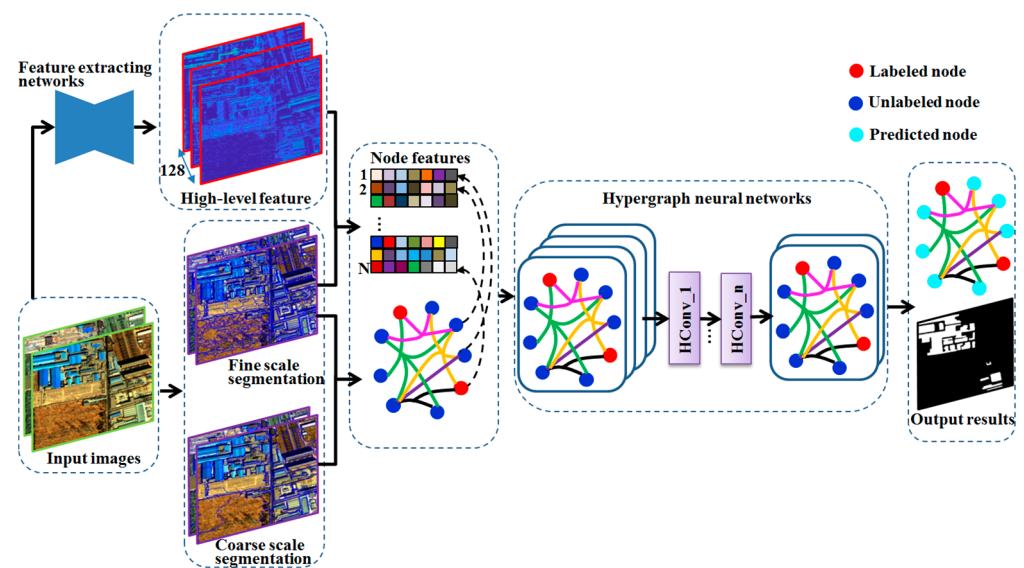


Figure 2. Flowchart of the proposed method.

3.1. Multiscale Segmentation and Feature Extraction

As mentioned above, the proposed method is object-based, which treats each object as a node of the hypergraph. Therefore, the input images need to be segmented into homogeneous regions—namely, objects. In this article, the FNEA is employed to obtain the objects. FNEA [55,56] segments the input images in a bottom-up way, which can get segmented results with hierarchical structure. For an input image, the segmented results of FNEA are decided by three parameters, namely, *scale*, *shape*, and *compactness*. The *scale* parameter, which is the most important parameter in FNEA, is used to control the minimum sizes of objects. Larger *scale* parameter would result in larger average size of objects, and vice versa. The *shape* parameter preserves the integrality of object contours. The *compactness* parameter is used to control the distinguishability of objects with similar spectral features. Different from other segmentation methods such as SLIC, the FNEA is a multilevel method, which means that an object under a coarse *scale* parameter is composed of several objects under a fine *scale* parameter and these father-child relationships actually contain abundant structural information of ground objects that can be exploited to construct the hypergraph.

As change detection involves two images simultaneously, to obtain objects in both images with the same boundaries, three segmented solutions [57] can be considered: (1) assigning, whereby one image is used as input for FNEA to generate object boundaries, then, assign the boundaries to the other image; (2) overlaying, whereby the two images are segmented by FNEA separately, and the two boundary sets are then overlaid to obtain coincident boundaries in two images; and (3) stacking, whereby the two images are stacked into one, and the stacked image is then segmented to generate boundaries. In this article, the stacking solution is adopted for the following reasons. Firstly, the assigning solution only utilizes information from one image while ignoring the other one, which may result in loss of object boundaries, especially in the changed regions. Secondly, the overlaying solution may generate excessively fragmentary boundaries, as ground objects in the two images usually appear with inconsistent shapes and sizes. The stacking solution considers information from the two images simultaneously, and thus, the generated boundaries are relatively consistent in the two images.

As most hand-crafted features depend on the professional knowledge of operators and are prone to inconsistency under diverse imaging conditions, we build high-level feature sets for hypergraph using a modified U-net based on the following situations. First, many versions of end-to-end U-nets have been effectively applied in CD tasks [58–60], which can extract high-level semantic features and obtain change maps with the same sizes of input images. Second, several optical data sets have been made available, such as the ONERA

Satellite Change Detection (OSCD) data set [61] and the AIST building change detection (ABCD) data set [62]. Therefore, we can obtain pixel-wise high-level features through a trained U-net using these available data sets.

The structure of the feature extracting network is shown in Figure 3, which consists of three parts that involve a contracting (or encoding) path, bottleneck and expanding (or decoding) paths. The OSCD data set is used to pre-train the U-net. When the pre-trained feature extracting network acts as the module shown in Figure 2, we treat the 128 feature maps $\mathbf{M} \in \mathbf{R}^{H \times W \times 128}$ as the output results, as marked by the red box in Figure 3. More details of the feature extracting network can be tracked in our previous works [51].

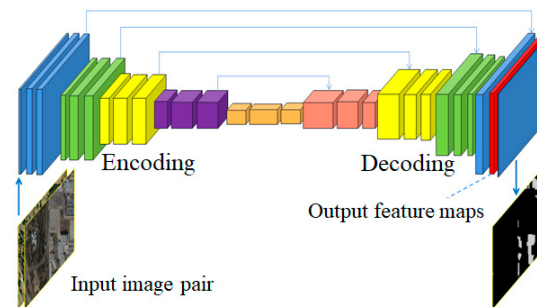


Figure 3. Illustration of the feature extracting network.

3.2. Hypergraph Construction

Supposing that the input images I_1, I_2 are stacked into one image I_S , then, I_S is segmented by the FNEA under a fine scale parameter S_1 and a coarse one S_2 with the same of *shape* and *compactness* parameters, respectively. Two object sets can be obtained: $\Omega_1 = \{F_1, F_2, \dots, F_N\}$, $\Omega_2 = \{C_1, C_2, \dots, C_M\}$, where F_i and C_j ($i = 1, 2, \dots, N$; $j = 1, 2, \dots, M$) denote the objects under the fine and coarse parameters, respectively, and naturally $N > M$. In this article, we treat each object under the fine scale parameter S_1 as a vertex, thus, a hypergraph can be defined as $\mathbf{G} = (\mathbf{V}, \varepsilon, w)$, which includes a vertex (node) set $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_N]^T \in \mathbf{R}^{N \times d}$, a hyperedge set $\varepsilon = [\mathbf{E}_1, \dots, \mathbf{E}_L]$, and each hyperedge $\mathbf{E} \in \varepsilon$ is assigned a positive weight $w(\mathbf{E})$, where d is the length of each feature vector, $d = 128$.

As $\mathbf{M} \in \mathbf{R}^{H \times W \times 128}$, the feature maps of input images have been obtained through the pre-trained feature extracting network, to build the high-level object-wise feature $\mathbf{v}_i \in \mathbf{R}^{1 \times 128}$ representing the i th node, we combine the $\mathbf{M} \in \mathbf{R}^{H \times W \times 128}$ which can represent high-level semantic information and the object set $\Omega_1 = \{F_1, F_2, \dots, F_N\}$ that contains abundantly spatial information to characterize the node. $\mathbf{v}_i$ is formulated by:

$$\mathbf{v}_i = \sum_{\{(j,k) | I_S(j,k) \in F_i\}} \mathbf{M}(j, k, :) / \#(F_i) \quad (1)$$

where $\#(F_i)$ denotes the number of pixels in F_i .

Most existing approaches of constructing hyperedge treat a centroid node with its K -nearest neighbors (KNN) in the spatial domain, or the centroid node and all nodes adjacent to it as the nodes within a hyperedge. The former may suffer the limitation that a large K would cause remarkable redundancy, while a small K would acquire deficient local relationships. Besides, a fixed K cannot be appropriate for all regions in VHR images due to the diversity of ground objects. The latter is based on the widely-used observation that the adjacent nodes (a node can present a pixel, a patch, an object and so on) have a greater probability to belong to the same class. However, this must recognize that the adjacent nodes by themselves may be insufficient to represent the structural relationships in some cases. Consequently, we define the dual neighborhood to construct informative hyperedges which can better exploit the relationships among objects.

As highlighted in Figure 4, a building can be presented by only one object (region) under the coarse scale, while it is over-segmented into many objects under the fine scale (marked by red boundaries). Supposing the red solid node in Figure 4b is the centroid

node, if the hyperedge is constructed using only itself and the adjacent nodes (yellow solid nodes), the structure of the building may not be fully represented. It can be observed that although the objects denoted by black solid nodes are not adjacent to the centroid one, they are still prone to belong to the same ground object, due to the fact that they are all children of the same object under the coarse scale. Thus, when constructing the hyperedge according to the red solid nodes, the structural information can be better exploited if taking both yellow and black solid nodes into account.

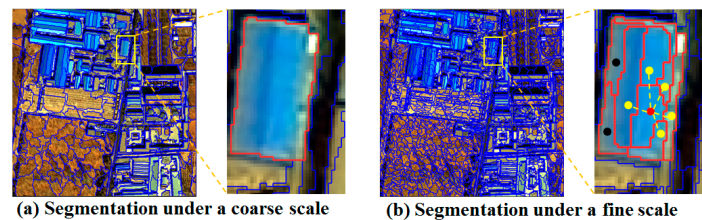


Figure 4. Illustration of dual neighborhood. (a) Segmentation under a coarse scale. (b) Segmentation under a fine scale.

According to the analysis above, we construct a hyperedge for each $\mathbf{v}_i$ based on the spatial neighborhood and structural neighborhood.

The spatial neighborhood of $\mathbf{v}_i$ is defined as follows:

$$N_1(\mathbf{v}_i) = \{\mathbf{v}_j | F_i, F_j \in \Omega_1, F_j \text{ is adjacent to } F_i\} \quad (2)$$

While the structural neighborhood of $\mathbf{v}_i$ is defined as follows:

$$N_2(\mathbf{v}_i) = \{\mathbf{v}_j | F_i \in C_k \& F_j \in C_k\} \quad (3)$$

Consequently, the hyperedge $\mathbf{E}_i$ centers on $\mathbf{v}_i$ is denoted as:

$$\mathbf{E}_i = \mathbf{v}_i \cup N_1(\mathbf{v}_i) \cup N_2(\mathbf{v}_i) \quad (4)$$

The incidence matrix $\mathbf{H} \in \mathbf{R}^{N \times L}$ ($L = N$ in this article) of the hypergraph is represented as:

$$\mathbf{H}(i, j) = h(\mathbf{v}_i, \mathbf{E}_j) = \begin{cases} 1, & \text{if } \mathbf{v}_i \in \mathbf{E}_j \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

The hyperedge weight is computed as:

$$w_i = w(\mathbf{E}_i) = \frac{2}{n_i(n_i - 1)} \sum_{j, k | \mathbf{v}_j \in \mathbf{E}_i, \mathbf{v}_k \in \mathbf{E}_i} \exp(-\|F_j - F_k\|) \quad (6)$$

where n_i is the number of vertices within $\mathbf{E}_i$. w_i measures the similarity of all vertices within the hyperedge. A large w_i means that the similar degree is high, thus the vertices within the hyperedge are prone to belong to the same class.

Based on $\mathbf{H}$ and w , the vertex degree of each $\mathbf{v}_i \in \mathbf{V}$ is:

$$d_i = d(\mathbf{v}_i) = \sum_{j=1}^N w_j \mathbf{H}(i, j) \quad (7)$$

and the edge degree of each $\mathbf{E}_i \in \varepsilon$ is:

$$\delta_i = \delta(\mathbf{E}_i) = \sum_{j=1}^N \mathbf{H}(j, i) \quad (8)$$

Further, $\mathbf{W} = \text{diag}(w_1, \dots, w_N)$, $\mathbf{D}_v = \text{diag}(d_1, \dots, d_N)$ and $\mathbf{D}_E = \text{diag}(\delta_1, \dots, \delta_N)$ denote the diagonal matrices of the hyperedge weights, the edge degrees and the vertex degrees, respectively.

Then, the normalized hypergraph Laplacian matrix can be computed as follows:

$$\Delta = \mathbf{I} - \mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_E^{-1} \mathbf{H}^T \mathbf{D}_v^{-1/2} \quad (9)$$

where Δ is positive semi-definite.

3.3. Hypergraph Convolution

Given a signal $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ with hypergraph structure, the approximation of spectral convolution of $\mathbf{x}$ and filter $\mathbf{g}$ can be denoted as [40]:

$$\mathbf{g} * \mathbf{x} \approx \sum_{k=0}^K \theta_k T_k(\tilde{\Delta}) \mathbf{x} \quad (10)$$

where $T_k(\tilde{\Delta})$ is the Chebyshev polynomial of order k with scaled Laplacian $\tilde{\Delta} = \frac{2}{\lambda_{\max}} \Delta - \mathbf{I}$, with $\lambda_{\max}$ being the largest eigenvalue of Δ . Equation (10) can avoid the complex computation of Laplacian eigenvectors by using only matrix powers, additions and multiplications. Feng et al. [52] further let $K = 1$ to limit the order of convolution operation as the Laplacian in hypergraph can already well represent the high-order relationships among nodes. Kipf and Welling [44] pointed out that $\lambda_{\max} \approx 2$ as the neural network parameters can adapt in scale during the training process. Based on the above assumptions, the convolution operation can be simplified to:

$$\mathbf{g} * \mathbf{x} \approx \theta_0 \mathbf{x} - \theta_1 \mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_E^{-1} \mathbf{H}^T \mathbf{D}_v^{-1/2} \mathbf{x} \quad (11)$$

where θ_0 and θ_1 are parameters of filters over all nodes. It can be beneficial to constrain the number of parameters to address over fitting and to minimize the number of operations (such as the matrix multiplications), by setting θ_0 and θ_1 as:

$$\begin{cases} \theta_1 = -\frac{1}{2} \theta \\ \theta_0 = \frac{1}{2} \theta \mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_E^{-1} \mathbf{H}^T \mathbf{D}_v^{-1/2} \end{cases} \quad (12)$$

Equation (12) can be simplified to the following expression:

$$\mathbf{g} * \mathbf{x} \approx \theta \mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_E^{-1} \mathbf{H}^T \mathbf{D}_v^{-1/2} \mathbf{x} \quad (13)$$

Then, given a hypergraph signal $\mathbf{X} \in \mathbf{R}^{N \times I_1}$ with N nodes and I_1 dimensional feature, the hypergraph convolution can be formulated by:

$$\mathbf{Y} = \mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_E^{-1} \mathbf{H}^T \mathbf{D}_v^{-1/2} \mathbf{X} \Theta \quad (14)$$

where $\Theta \in \mathbf{R}^{I_1 \times I_2}$ is the parameter to be learned during the training process. The output $\mathbf{Y} \in \mathbf{R}^{N \times I_2}$ can be used for clustering or classification.

Noting that $\mathbf{D}_v$ and $\mathbf{D}_E$ play a role of normalization in Equation (14), the hypergraph convolution actually propagates information by a node-edge-node manner, which can better refine the features using the hypergraph structure. Equation (14) can be interpreted from right to left as follows: first, the initial node feature $\mathbf{X}$ is processed by the filter matrix Θ to extract I_2 dimensional node feature. Then, the node feature is gathered according to the hyperedge by using the multiplication of $\mathbf{W} \mathbf{H}^T \in \mathbf{R}^{L \times N}$ to form the hyperedge feature, which can be denoted as $\mathbf{W} \mathbf{H}^T \mathbf{X} \Theta \in \mathbf{R}^{L \times I_2}$. After that, by multiplying matrix $\mathbf{H}$ to aggregate the related hyperedge feature, we can transform the hyperedge feature into output node feature $\mathbf{H} \mathbf{W} \mathbf{H}^T \mathbf{X} \Theta \in \mathbf{R}^{N \times I_2}$.

According to Equation (14), the output feature of the i th node can be expressed as:

$$\mathbf{Y}_i = \sum_{j=1}^N \sum_{k=1}^L \mathbf{H}(i, k) \mathbf{H}(j, k) w_k \mathbf{X}_j \Theta / (d_i \delta_i) \quad (15)$$

It can be concluded from Equation (15) that information is propagated among nodes within a common hyperedge. In addition, the hyperedge with a larger weight means that labels within the hyperedge are smoother [63].

Based on the above definitions, the hypergraph convolutional networks with L layers can be built using the following formulation:

$$\mathbf{X}^{(l+1)} = \sigma\left(\mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_E^{-1} \mathbf{H}^T \mathbf{D}_v^{-1/2} \mathbf{X}^{(l)} \Theta^{(l)}\right) \quad (16)$$

where $\mathbf{X}^{(l)}$ is the signal of hypergraph at the l th layer, and σ is the non-linear activation function like ReLU, $\mathbf{X}^{(0)} = \mathbf{V}$.

Imbalanced distribution of changed and unchanged samples is very typical in CD problems. Generally, the changed regions often account for a small part in the whole studied area. To solve the sample imbalance problem, in the training process, we adopt the focal loss [64], which has proved its availability in various detection and classification tasks. The loss function in this article can be defined as:

$$L_{HGNN} = \sum_{i=1}^{N_L} -\alpha \left(1 - p^{(i)}\right)^{\gamma} \log\left(p^{(i)}\right) \quad (17)$$

where α is used to control the contribution weight of positive (changed) and negative (unchanged) samples to the total loss. When γ is 0, it is equivalent to cross-entropy loss. N_L is the number of labeled samples. In this article, we set $\alpha = 0.2$ and $\gamma = 2$. $p^{(i)}$ is the prediction probability of the i th labeled sample and it is formulated as:

$$p_t^{(i)} = \begin{cases} p_i, & \text{if } y_i = 1 \text{ (changed)} \\ 1 - p_i, & \text{if } y_i = 0 \text{ (unchanged)} \end{cases} \quad (18)$$

where p_i is the predicted output of the networks.

4. Experiments and Analysis

In this section, we first introduce the data sets. Then, brief implementation and evaluation metrics are provided. Following that, experimental results are presented and discussed in detail.

4.1. Descriptions of Data Sets

We conducted the experiments on three openly optical VHR image data sets: the first one was the Sun Yat-Sen University CD data set (SYSU-CD) provided by Shi et al. [65]. SYSU-CD contains 20,000 pairs of aerial images with a spatial resolution of 0.5 m and size of 256×256 pixels. It was collected between 2007 and 2014 in Hong Kong. The data set included the following main types of changes: newly-built buildings, suburban expansion, changes in vegetation, road expansion, and sea construction. Several patch samples of the first data set are shown in Figure 5. The second one was the WHU-CD data set [66], as shown in Figure 6. It contained a pair of aerial images with a spatial resolution of 0.2 m, which were taken in 2012 and 2016. The image size is $15,354 \times 32,507$ pixels and the main types of changes labeled are construction and demolishing of buildings. We split the images into 6000 image pairs with a size of 256×256 pixels. Both data sets introduced variations derived from seasonal factors and illumination conditions, which could be used to test the robustness of methods. The third one was the LEVIR-CD data set released by Chen and Shi [27], which contained 637 VHR (0.5 m/pixel) Google Earth (GE) image patch pairs with a size of 1024×1024 pixels. The types of changes included several significant land-use changes, especially urban constructions. Some patch samples are shown in Figure 7.



Figure 5. Image patch examples and corresponding reference images of the SYSU-CD data set.



Figure 6. Image patch examples and corresponding reference images of the WHU-CD data set.



Figure 7. Image patch examples and corresponding reference images of the LEVIR-CD data set.

Two VHR SAR data sets were involved. The first one was the Wuhan-SAR data set, which was composed of a pair of images acquired by TerraSAR-X sensor with 1 m/pixel covering a suburban area of Wuhan, China. The image size is 780×870 pixels, as shown in the first row of Figure 8. The second one was the Shanghai-SAR data set, a pair of images covering an area in Shanghai, China, which were acquired by Gaofen-3 sensor with the size of 457×553 and 1 m/pixel, as shown in the second row of Figure 8.

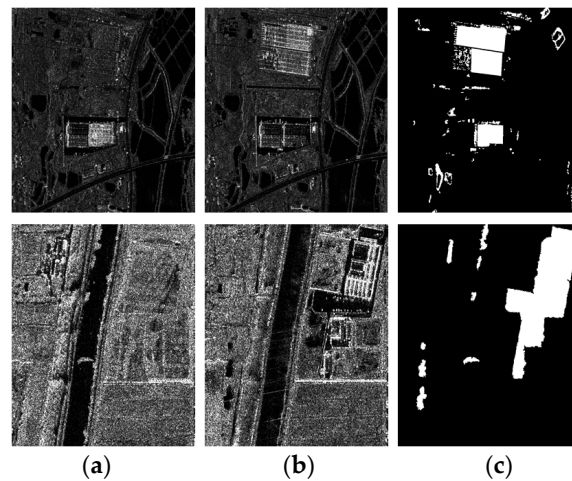


Figure 8. SAR data sets. (a) Image T1. (b) Image T2. (c) Reference change map.

Five VHR heterogeneous optical/SAR data sets (Data 1~5) were involved. The first one consisted of an optical image acquired by QuickBird in July 2006 and an SAR image captured by TerraSAR-X in July 2007, with a spatial resolution of 0.65 m. The mainly changed information was the flooding in Gloucester, England. The second to the fifth data sets all included a Google image and a Gaofen-3 SAR image, covering different cities in China, including Shanghai, Zhengzhou, Huizhou and Xiangshui, respectively. The spatial resolution was 1 m and the main types of changes were construction and demolition of buildings, and areas covered by water. These five data sets are shown in Figure 9.

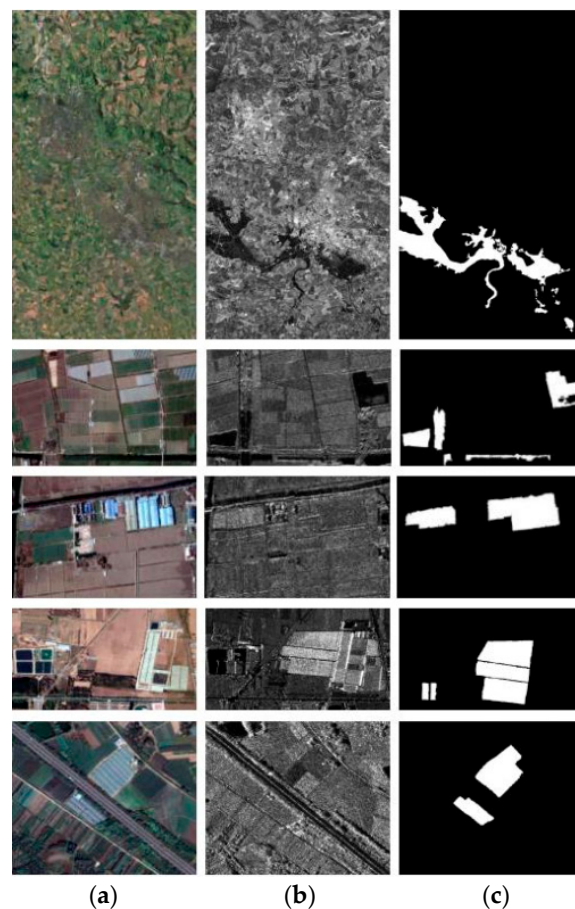


Figure 9. Heterogeneous optical/SAR data sets. (a) Optical image. (b) SAR image. (c) Reference change map.

4.2. Implementation and Evaluation Metrics

We implemented the proposed method via PyTorch library. In the training phase, we fixed the learning rate as 0.001 and the momentum as 0.9. As the proposed method was semi-supervised, for each image pair introduced in Section 5-A, we randomly selected 5% of the objects under the fine scale as labeled nodes after superpixel segmentation, while the other 95% of the objects were used as testing nodes. A hypergraph was constructed and trained for each image pair. The number of HGNN layers was set to 2, as per $L = 2$ in formula (16). The proposed networks were trained with 400 epochs and the dropout rate was 0.5, and the weight decay was 0.0005. All experiments were conducted on a single GeForce GTX 1080Ti GPU.

To evaluate the performance of the proposed method, we utilized four quantitative metrics: false alarm rate (FAR), missed alarm rate (MAR), overall accuracy (OA) and Kappa coefficient (Kappa), F1 score (F1), and intersection of union (IOU).

4.3. Experiments on Optical Images

To verify the superiority of DNHGNN on optical remote sensing CD tasks, the following seven benchmark methods were selected as competitors:

- (1) **FC-Siam-con**: The FC-Siam-con [58] uses a Siamese encoding stream to extract deep features from bi-temporal images. The features were then concatenated in the decoding stream for CD.
- (2) **FC-Siam-diff**: Different from the FC-Siam-con, FC-Siam-diff [58] uses absolute difference between the bi-temporal features to decide changed degrees.
- (3) **Siam-NestedUNet**: The Siam-NestedUNet [67] uses a Siamese semantic segmentation network UNet++ to extract features of different resolution. Following that, the features are fed into an ensemble channel attention module.
- (4) **DSIFN**: A deeply supervised image fusion network (DSIFN) [68] has been proposed which consists of a shared deep feature network and a difference discrimination network, which utilizes the channel attention module and spatial attention module.
- (5) **DSAMNet**: The deeply supervised attention metric-based network (DSAMNet) employs a metric module to learn change maps by means of deep metric learning, in which convolutional block attention modules are integrated to provide more discriminative features [65].
- (6) **GCNCD**: A GCN-based method that utilizes several types of hand-crafted features to build to feature sets of nodes [50].
- (7) **MSGCN**: A multiscale GCN has been proposed which can fuse the outputs of GCN under different segmented scales [51].

The first five methods are supervised; thus, we used 12,000 pairs of images for training, 4000 pairs for verification, 4000 pairs for testing on the SYSU-CD data set and 3600 pairs for training, 1200 pairs for verification, and 1200 pairs for testing on the WHU-CD data set, respectively. The parameters in these methods are set to be as consistent as possible with the original literature. The GCNCD and MSGCN are semi-supervised as DNHGNN. Thus, to ensure fair comparisons, we randomly selected 5% of the objects under the fine scale as labeled nodes. The two scale parameters of multiscale segmentation were set as 8 and 15 for both data sets.

Some typical results of SYSU-CD and WHU-CD are presented in Figures 10–12. The results generated by the three GNN-based methods (GCNCD, MSGCN and DNHGNN) can reflect main information regarding changes. In contrast, one can see that change maps provided by FC-Siam-con, FC-Siam-diff and DSIFN are affected by significant salt-and-pepper noise where numerous unchanged pixels are misclassified as changed ones (marked by the red boxes in Figures 10d,e,g and 11d,e,g). The reason is that they have limited immune capabilities on the spectral variations. Obvious false alarms also occur in the results of these three methods (see the red ellipses in Figures 10d,e,g and 11d,e). In addition, some changed regions in the results of DSIFN are not homogeneous enough, such as the image presented in the first row of column (g). The Siam-NestedUnet addresses these issues

well by introducing the ensemble channel attention module, and thus causes fewer false alarms. However, the boundaries between changed and unchanged regions are inaccurate in some results, as can be seen in the blue boxes of Figures 10f, 11f and 12f, as the results fail to partition the changed and unchanged regions. The DSAMNet achieves relatively good performance on some image pairs, such as the second row of Figure 10h and the fourth row of Figure 11h, but false alarms are still relatively high in some results, such as the red ellipses in Figure 11h. Moreover, DSAMNet loses some changed regions, which are marked by the blue boxes in Figure 10h. GCNCD can obtain results with relatively complete regions. Nevertheless, some boundaries are inaccurate (see the blue boxes in Figures 10i and 11i), as the discrimination of hand-crafted features is limited. MSGCN seems to obtain similar results as DNHGNN. Interpreted in detail, the results of DNHGNN are more consistent with the reference maps. In some regions with complex structures, DNHGNN achieves higher precision, as shown by the blue boxes in Figures 10j and 11j.

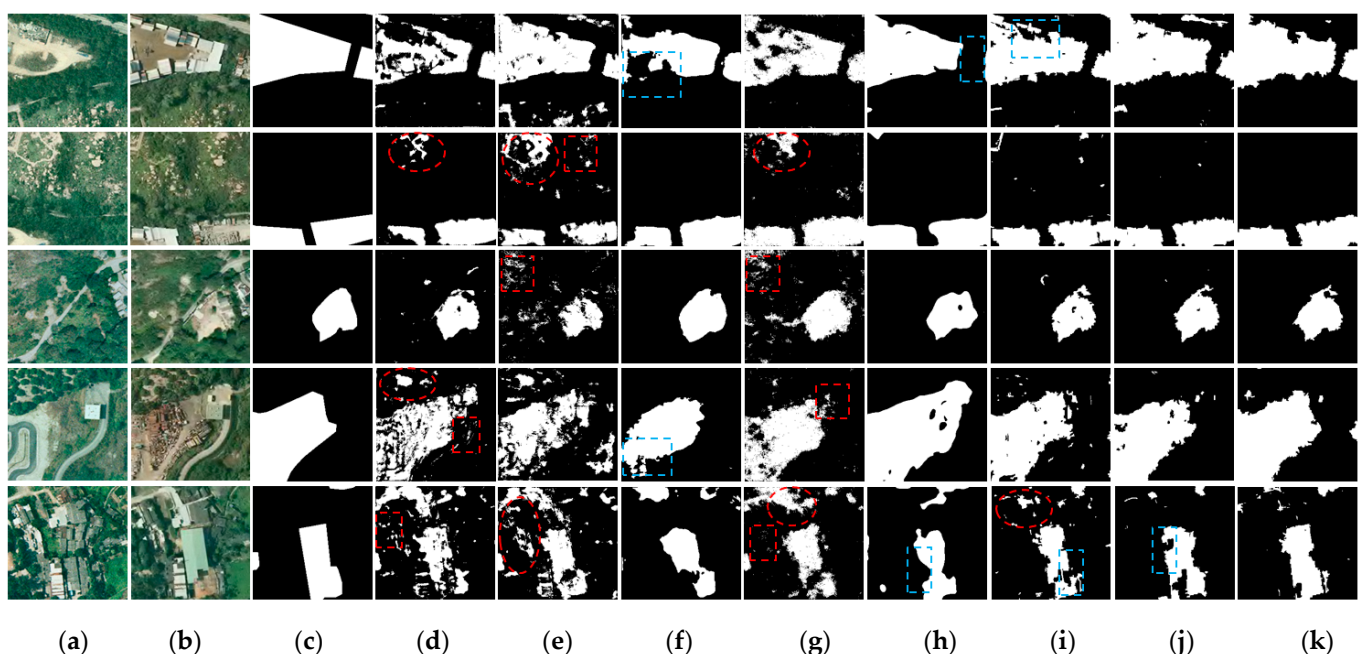


Figure 10. Some typical CD maps by different methods on the SYSU-CD data set. (a) Image T1. (b) Images T2. (c) Reference change map. (d) FC-Siam-con. (e) FC-Siam-diff. (f) Siam-NestedUNet. (g) DSFIN. (h) DSAMNet. (i) GCNCD. (j) MSGCN. (k) DNHGNN.

Table 1 reports the quantitative evaluation results of different methods. It can be concluded that the proposed DNHGNN outperforms all the compared methods in terms of OA, Kappa, F1 and IOU on all data sets. In addition, it clearly reaches the lowest MAR on the SYSU-CD and LEVIR-CD data sets, and the second-lowest on the WHU-CD data set. Regarding the FARs, those of DNHGNN are also relatively low. Compared with other methods, the DNHGNN yields an improvement of at least 4.92%, 0.78%, 2.66%, 1.49% and 2.27% for MAR, OA, Kappa, F1, and IOU, respectively, on the SYSU-CD data set. The improvement on the WHU-CD and LEVIR-CD data sets is also evident. For the SYSU-CD data set, GCNCD and MSGCN achieve slightly lower FARs than that of DNHGNN. However, they yield significantly higher MARs, resulting in the loss of some structural information in changed regions as the analysis above. On the whole, the DNHGNN can suppress false alarms and reduce missed detections simultaneously. The reasons for this behavior are: (1) the combination of pixel-wise high-level features with object-based extraction improves the robustness on the misregistration errors and spectral variations, and this can reduce the false alarms and missed detection; and (2) when constructing the hypergraph, the utilization of dual neighborhood helps capture the structural information of complex regions effectively and accurately.

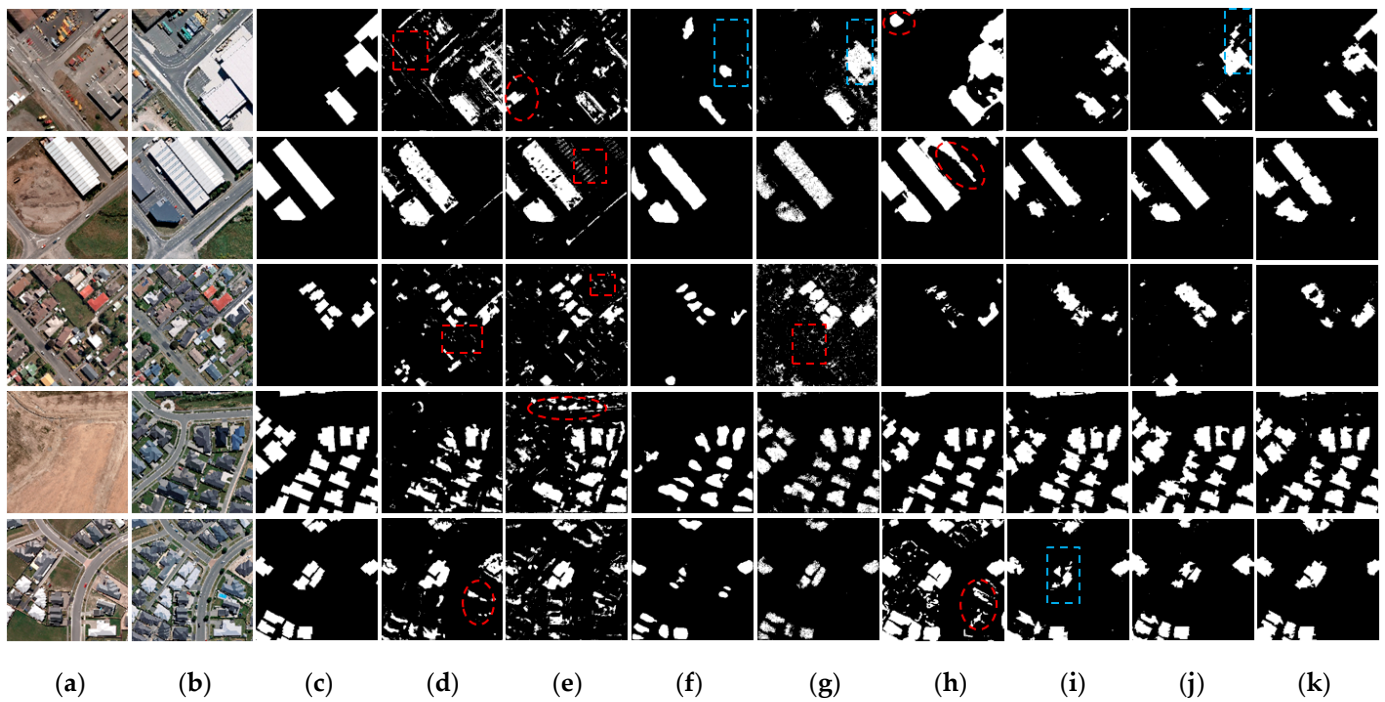


Figure 11. Some typical CD maps by different methods on the WHU-CD data set. (a) Image T1. (b) Images T2. (c) Reference change map. (d) FC-Siam-con. (e) FC-Siam-diff. (f) Siam-NestedUNet. (g) DSFIN. (h) DSAMNet. (i) GCNCD. (j) MSGCN. (k) DNHGNN.

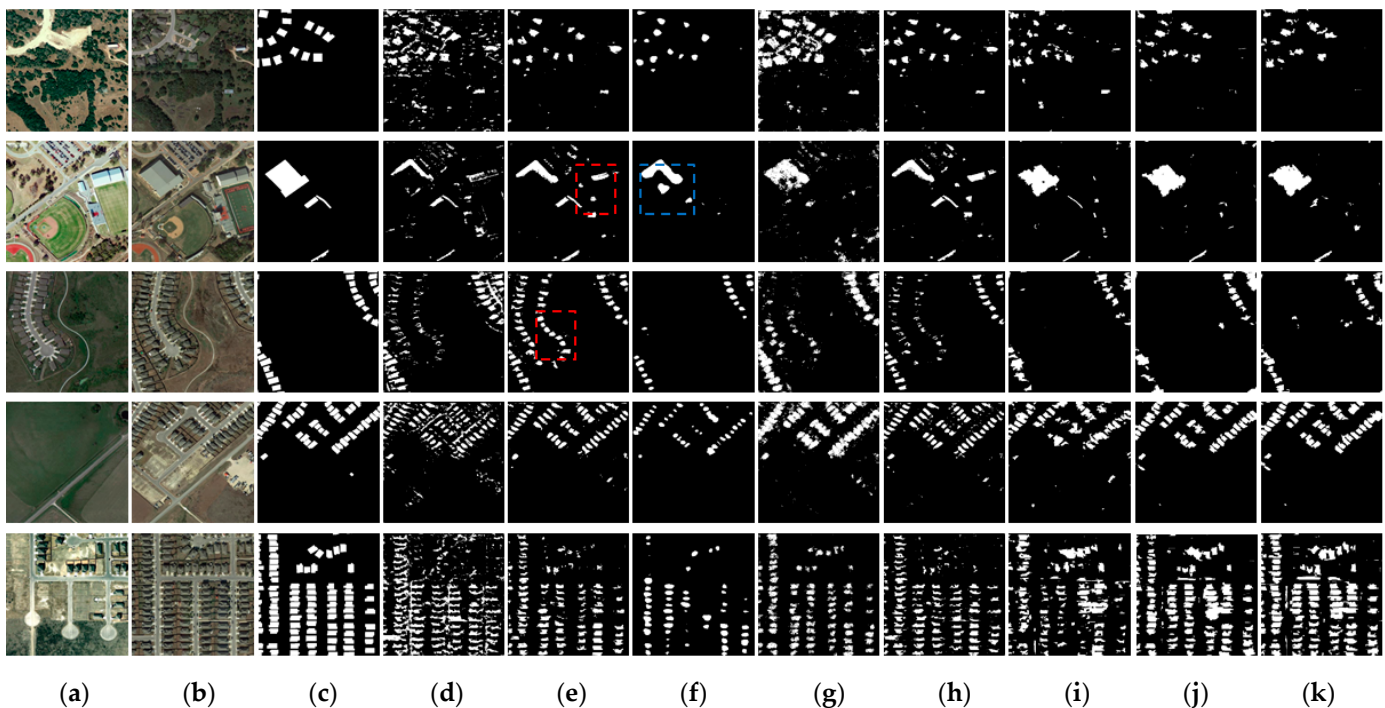


Figure 12. Some typical CD maps by different methods on the LEVIR-CD data set. (a) Image T1. (b) Images T2. (c) Reference change map. (d) FC-Siam-con. (e) FC-Siam-diff. (f) Siam-NestedUNet. (g) DSFIN. (h) DSAMNet. (i) GCNCD. (j) MSGCN. (k) DNHGNN.

Table 1. Quantitative accuracy results for different methods on the optical data sets (%).

Data set	Method	FAR	MAR	OA	Kappa	F1	IOU
SYSU-CD	FC-Siam-con	4.57	36.29	88.91	62.30	65.21	55.17
	FC-Siam-diff	7.77	39.27	86.51	54.03	58.30	48.65
	Siam-NestedUnet	3.26	24.41	92.12	73.47	77.02	65.12
	DSIFN	5.20	18.75	91.79	72.72	76.29	63.86
	DSAMNet	2.97	17.77	93.60	78.29	82.65	72.33
	GCNCD	1.92	20.02	95.01	82.21	88.01	80.13
	MSGCN	1.83	15.06	95.94	85.22	89.91	82.77
	DNHGNN	2.08	10.14	96.72	87.88	91.40	85.04
WHU-CD	FC-Siam-con	4.31	44.79	91.03	52.46	56.10	42.89
	FC-Siam-diff	4.63	44.07	90.88	49.60	53.22	41.77
	Siam-NestedUnet	0.89	53.55	93.07	54.54	60.05	47.32
	DSIFN	2.39	24.78	94.49	67.83	71.37	60.59
	DSAMNet	3.35	15.09	95.06	74.22	80.07	67.19
	GCNCD	1.70	34.13	95.30	70.74	75.10	62.54
	MSGCN	1.78	24.53	96.01	76.83	81.11	70.66
	DNHGNN	1.16	18.97	97.05	82.19	85.42	74.12
LEVIR-CD	FC-Siam-con	6.23	41.91	88.57	56.57	59.20	48.37
	FC-Siam-diff	3.97	31.24	90.69	63.41	67.45	55.44
	Siam-NestedUnet	0.54	42.98	89.88	61.74	63.54	52.83
	DSIFN	2.42	25.21	93.89	67.52	71.28	57.21
	DSAMNet	3.32	29.45	94.12	64.96	68.91	56.01
	GCNCD	2.31	26.79	92.80	66.88	70.77	56.84
	MSGCN	1.76	23.89	95.49	72.71	75.02	60.56
	DNHGNN	1.37	23.01	95.70	74.86	79.16	64.21

4.4. Experiments on SAR Images

The following six state-of-the-art methods are selected to compare with the proposed DNHGNN on SAR images:

- (1) PCA-Kmeans. A PCA-based method [7], in which the K-means approach was used for binary classification.
- (2) ELM. An extreme learning machine-based method for CD in SAR images [69].
- (3) S-PCA-Net. The S-PCA-Net [70] introduced the imbalanced learning process into PCA-Net [71].
- (4) CWNN. The method used convolutional wavelet neural network (CWNN) instead of CNN to extract robust features with better noise immunity for SAR image CD [72].
- (5) CNN. The method used a novel CNN framework without any preprocessing operations, which can automatically extract the spatial characteristics [73].
- (6) MSGCN [51].

The scale parameters of multiscale segmentation in MSGCN and DNHGNN are set as 10, 15 for the Wuhan-SAR data set and 15, 25 for the Shanghai-SAR data set, respectively.

The visual results on the two SAR data sets are shown in Figure 13. As can be seen, many false alarms appear in the results generated by PCA-Kmeans, ELM, S-PCA-Net and CWNN in both data sets, as per the red boxes in Figure 13a–d. The reason is that these methods are essential pixel-based methods, which may have poor immunity to the severe speckle noise that is commonly prevalent in VHR SAR images. Additionally, the homogeneity within the regions of changed buildings is not ideal, where the completeness of the buildings can hardly be represented by the results of these four methods (see the yellow boxes in Figure 13a–d). In contrast, CNN, MSGCN and DNHGNN can effectively suppress false alarms. However, more changed regions are missed in the results of CNN. For example, in the regions of yellow boxes in Figure 13e, the areas of changed buildings are fragmentary. The results of MSGCN lose more structural information compared with those of DNHGNN, as marked by the yellow box in Figure 13f.

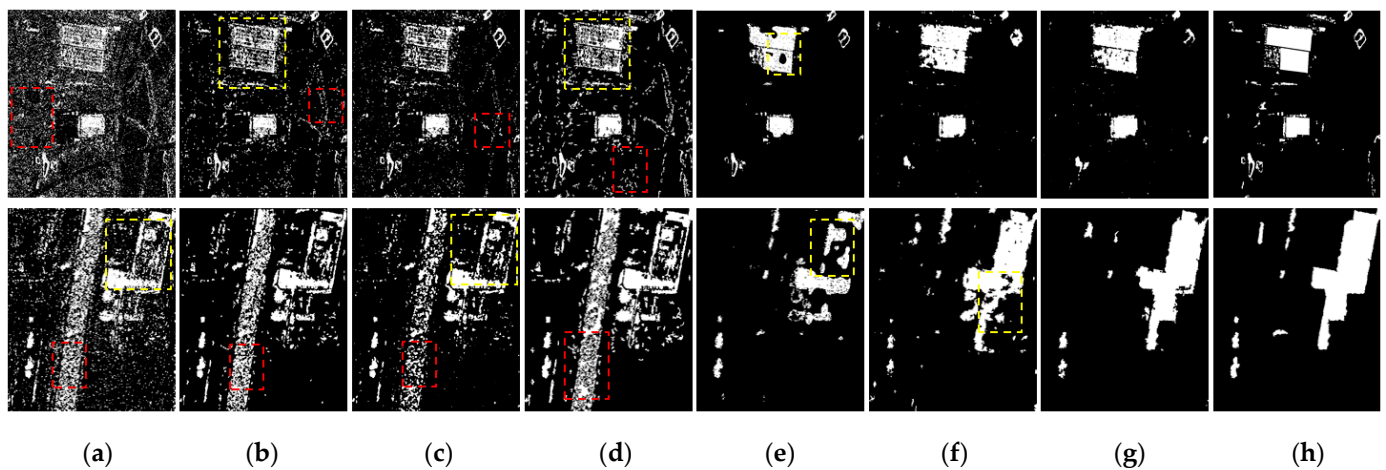


Figure 13. Change maps by different methods on the SAR data sets. (a) PCA-Kmeans. (b) ELM. (c) S-PCA-Net. (d) CWNN. (e) CNN. (f) MSGCN. (g) DNHGNN. (h) Reference change map.

Table 2 shows the quantitative evaluation results on the two SAR data sets. DNHGNN achieves the best in terms of all indices on the Wuhan-SAR data set. For the Shanghai-SAR data set, the MAR, OA, Kappa, F1 and UOI of DNHGNN are likewise the best. Although DNHGNN achieves a higher FAR than CNN, the MAR is much lower. Due to the complicated noise situation, the performances of the PCA-Kmeans, ELM, S-PCA-Net and CWNN are not ideal. Obviously, the FARs of these four methods are much higher than those of others. Although CNN achieves the lowest FAR, it yields significantly higher MAR, which is as high as 57.08%, and this means that many changed pixels are missed in the results of CNN, as per our analysis in the aforementioned visual results. Comparing the indices comprehensively, DNHGNN produces the best results on both data sets.

Table 2. Quantitative accuracy results for different methods on the SAR data sets (%).

Methods	Wuhan						Shanghai					
	FAR	MAR	OA	Kappa	F1	IOU	FAR	MAR	OA	Kappa	F1	IOU
PCA-Kmeans	11.80	37.22	84.20	49.11	54.11	45.85	11.51	40.11	85.56	38.52	42.22	35.10
ELM	4.15	37.21	90.65	62.45	66.03	50.36	8.89	39.80	88.06	43.33	46.21	36.97
S-PCA-Net	2.71	32.91	92.54	70.58	73.21	58.64	4.43	39.74	92.08	55.65	60.16	48.85
CWNN	3.61	28.92	92.41	71.21	75.87	61.77	10.97	36.50	86.51	40.93	44.37	35.81
CNN	3.41	28.87	92.59	71.78	77.10	63.74	0.76	57.08	93.68	54.27	59.54	46.22
MSGCN	2.97	16.73	94.87	80.57	83.30	71.92	0.77	42.71	94.89	65.49	70.09	59.94
DNHGNN	2.30	15.64	95.60	83.17	84.94	74.33	2.80	29.15	94.60	69.14	73.58	62.45

4.5. Experiments on Heterogeneous Optical/SAR Images

We validated proposed DNHGNN on heterogeneous images comparing with the following benchmark methods:

- (1) FPMS [74]. The fractal projection and Markovian segmentation-based method (FPMS) projects the pre-event image to the domain of post-event image by fractal projection. Then, an MRF segmentation model is employed to obtain change maps.
- (2) CICM [75]. A concentric circular invariant convolution model (CICM) is proposed to project one image into the imaging modality of the other.
- (3) IRG-MCS [76]. The authors define an Iterative robust similarity graph to measure changed degree.
- (4) SCASC. A method which uses a regression model with sparse constrained adaptive structure consistency [77].

- (5) GIR-MRF [78]. A structured graph learning based method, which first learns a robust graph to capture the local and global structure information of the image, and then projects the graph to the domain of the other image to complete the image regression.
- (6) MSGCN [51].

The scale parameters in MSGCN and DNHGNN are set as 25, 50 for the first data set and 12, 30 for the others.

Figure 14 shows the change maps of all comparing methods on the heterogeneous optical/SAR data sets. It can be observed that the performances of FPMS, CICM, and IRG-MCS are unsatisfactory, as they not only cause many false alarms (marked by red boxes in Figure 14a–c), but also miss some mainly changed regions, such as those shown by the yellow boxes in Figure 14a–c. SCASC seem to achieve better performances than those of the above three methods. Nevertheless, the missing of changed regions is still present, as can be seen in the yellow boxes in Figure 14d. Intuitively, GIR-MRF and MSGCN can capture the mainly changed information as DNHGNN does. However, interpreted in details, GIR-MRF causes more false alarms than DNHGNN (see the red boxes in Figure 14e). Compared with DNHGNN, MSGCN fails to accurately capture some shapes of changed regions, such as those marked by the yellow boxes in Figure 14f.

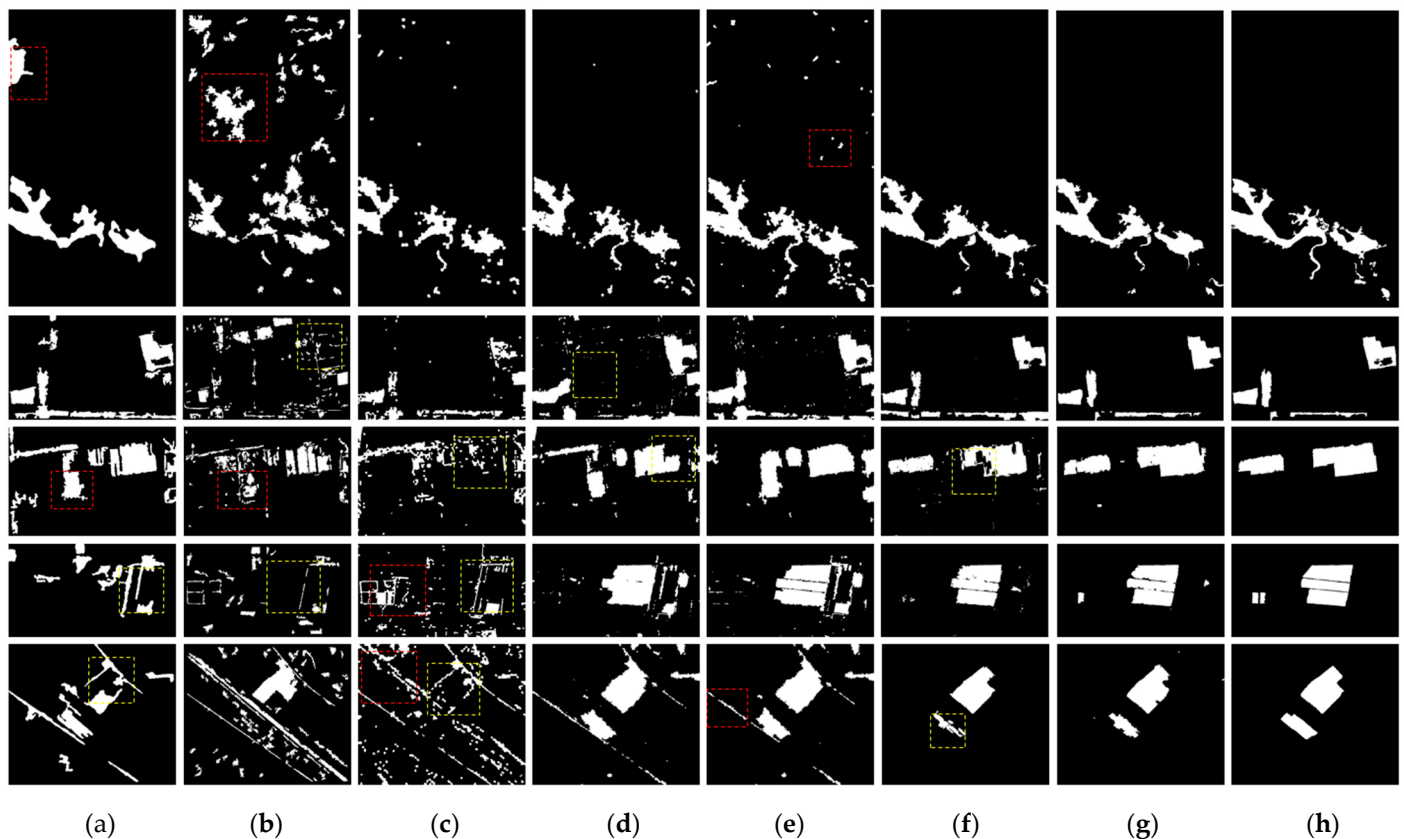


Figure 14. Change maps by different methods on the heterogeneous Optical/SAR data sets. (a) FPMS. (b) CICM. (c) IRG-MCS. (d) SCASC. (e) GIR-MRF. (f) MSGCN. (g) DNHGNN. (h) Reference change map.

The quantitative evaluation results on the two heterogeneous data sets are listed in Table 3. DNHGNN outperforms other methods significantly in terms of OA, Kappa, F1 and IOU. As it is consistent with the visual comparison of Figure 14, the relatively low MAR and FAR mean that DNHGNN can effectively suppress the false alarms and avoid the missed detection simultaneously. On the whole, the proposed DNHGNN outperforms the benchmark methods and the reasons of this behavior may be as follows:

Table 3. Quantitative accuracy results for different methods on the heterogeneous data sets (%).

Data set	Method	FAR	MAR	OA	Kappa	F1	IOU
Data 1	FPMS	2.55	10.28	97.01	75.73	78.86	65.36
	CICM	8.36	34.32	90.17	38.41	44.21	33.56
	IRG-MCS	1.18	27.74	97.21	74.36	76.89	64.13
	SCASC	1.21	17.98	97.65	80.47	82.24	71.20
	GIR-MRF	2.11	6.07	97.77	81.62	85.11	75.23
	MSGCN	1.11	13.03	98.15	84.34	86.45	76.80
	DNHGNN	0.56	13.26	98.66	88.10	89.82	79.88
Data 2	FPMS	6.57	11.24	93.03	64.96	68.66	52.28
	CICM	8.54	82.75	85.08	13.47	16.58	9.04
	IRG-MCS	4.00	45.50	92.43	51.19	55.32	38.24
	SCASC	6.63	30.40	91.33	53.30	57.99	40.83
	GIR-MRF	6.48	16.86	92.63	62.04	65.98	49.23
	MSGCN	3.30	14.58	95.73	75.14	77.48	63.23
	DNHGNN	1.05	9.93	98.19	88.53	89.52	81.03
Data 3	FPMS	10.27	44.02	85.66	40.30	48.45	31.97
	CICM	4.95	53.21	89.24	45.16	51.14	34.68
	IRG-MCS	8.93	80.23	81.82	12.14	13.23	8.08
	SCASC	6.58	47.76	88.47	45.60	52.16	35.28
	GIR-MRF	8.64	38.38	87.78	47.87	54.83	37.77
	MSGCN	1.29	17.10	96.81	84.40	85.20	75.75
	DNHGNN	1.57	4.05	98.13	91.44	92.51	86.06
Data 4	FPMS	7.96	82.12	83.86	11.76	14.11	9.23
	CICM	5.85	87.21	83.78	10.73	13.54	8.87
	IRG-MCS	8.79	88.12	81.56	10.12	13.01	8.24
	SCASC	4.89	14.28	94.03	73.13	76.69	62.19
	GIR-MRF	6.22	11.99	93.12	70.68	74.56	59.43
	MSGCN	0.53	10.82	98.29	91.34	92.29	85.69
	DNHGNN	0.38	6.19	98.95	94.77	95.36	91.13
Data 5	FPMS	5.24	51.11	91.73	39.43	43.84	28.07
	CICM	9.35	39.68	88.65	35.65	41.22	25.96
	IRG-MCS	9.92	84.25	85.47	14.10	17.20	9.02
	SCASC	3.86	3.95	96.14	74.63	76.64	62.13
	GIR-MRF	3.64	4.57	96.30	75.32	77.26	62.95
	MSGCN	0.13	17.54	98.72	88.79	89.46	80.94
	DNHGNN	0.19	9.56	99.19	93.22	93.65	88.06

(1) Different from the methods based on regression or feature projection, DNHGNN treats CD as a classification task and obtain features from the same feature space, which may avoid the influence of heterogeneity between images.

(2) The hypergraph neural networks make full use of the small amount of labeled samples (5%) to identify the labels of unlabeled nodes.

5. Discussion

In the following, we discuss the robustness to the scale parameters and the influence of the ratio of labeled samples.

5.1. Robustness to the Scale Parameters

In order to capture more comprehensive information of the ground objects and reduce the number of nodes, we segment the input image pair under one fine scale parameter S_1 and one coarse S_2 , respectively. Each object under S_1 is treated as a node and the father-child relationships between objects under S_1 and S_2 are used to construct hypergraphs. Thus, the scale parameters are pivotal to the performance of DNHGNN.

To ensure the homogeneity of each object, we set S_1 to a relatively small value to guarantee moderate over-segmentation. Nevertheless, S_2 needs to be further considered.

As an object under S_2 is merged by several objects under S_1 , different S_2 can obtain distinctly structural information. To be specific, when fixing the fine scale parameter, a larger S_2 would result in more nodes within a hyperedge, and this means that usable information within a hyperedge increases, but may lead to redundancy. Consequently, in this section, for each image pair, we fix the S_1 and set S_2 to different values to discuss the performance.

Figure 15 shows two samples of optical CD results with $S_1 = 8$ and different S_2 , in steps of 3. As can be seen in the first row of Figure 15, when S_2 increases from 15 to 18, the changed building in the up-left region is detected more precisely, while when S_2 increases from 21 to 24, the detected structure of this building is not integrated again due to the utilization of some redundant information. This phenomenon also occurs in the second row of Figure 15. On the whole, all results shown in Figure 15 can capture the main changed information, with some discrepancy in complex regions. Meanwhile, all these results can effectively suppress false alarms, as we can see that a small number of unchanged pixels are misclassified as changed ones.

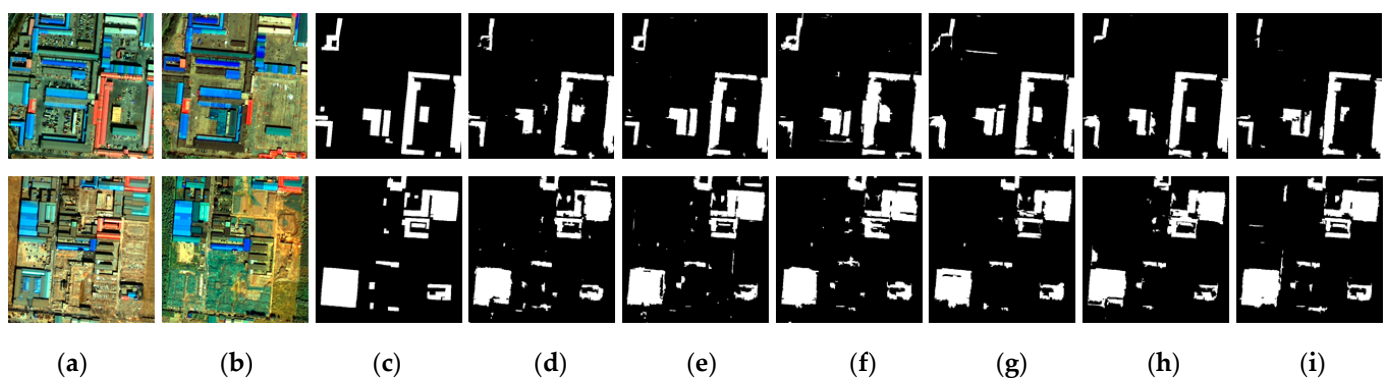


Figure 15. Two samples of CD results by DNHGNN with fixed $S_1 = 8$ and different S_2 . (a) Image T1. (b) Image T2. (c) Reference change map. (d) $S_2 = 15$. (e) $S_2 = 18$. (f) $S_2 = 21$. (g) $S_2 = 24$. (h) $S_2 = 27$. (i) $S_2 = 30$.

The corresponding OA and Kappa are displayed in Figure 16. It can be observed that both OA and Kappa maintain relatively high values and the discrepancy is not evident. Although the best S_2 is hard to determine, the performances under all $S_2 \in [15, 30]$ are acceptable. It means we can set S_2 in a relatively large range. In other words, the method has good robustness to the scale parameters.

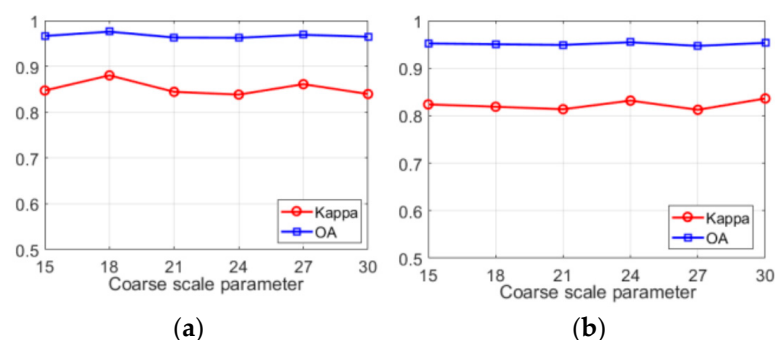


Figure 16. Kappa and OA under different S_2 (a) Sample 1. (b) Sample 2.

5.2. Influence of the Ratio of Labeled Samples

As a semi-supervised method, the motivation of DNHGNN is to infer the labels of nodes using only a few labeled nodes. Without doubt, the performance of DNHGNN would be influenced by the number of labeled samples. For each pair of input images, a part of the segmented objects are randomly selected as training samples, as per the red solid circles shown in Figure 2. In our experiments, the numbers of segmented objects

vary widely for different images. Thus, we utilize the ratio of labeled samples to evaluate the unavoidable influence. Figure 17 shows the OA and Kappa under different ratios of labeled samples using the image pairs of Figure 15 as input. It can be seen that higher ratios can obtain higher OA and Kappa, due to the increase of supervised information helping to more accurately infer the labels of nodes. It is noteworthy that the improvement of OA is not notable, as the DNHGNN has achieved a relatively high OA when the ratio is only 5%. From another point of view, this phenomenon demonstrates that the DNHGNN can achieve promising performance with relatively few labeled samples, which makes it valuable and feasible in many applications.

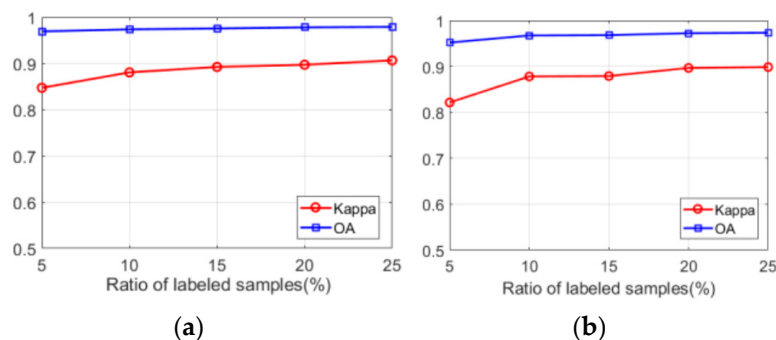


Figure 17. Kappa and OA with different labeled ratios. (a) Sample 1. (b) Sample 2.

6. Conclusions

In this article, a semi-supervised hypergraph neural network is proposed for change detection. The main idea of the presented framework is to model the high-order relationships commonly existing in VHR images using hypergraph and to propagate information through the defined hypergraph convolution. To reduce the number of nodes and obtain object-wise features with semantic information, the input image pair is firstly segmented through FNEA with one fine scale parameter and one coarse one, respectively. Treating each object under fine scale parameter as a node, the node features can be obtained by combining the segmentation with a pre-trained U-net. The hyperedges are constructed based on a defined dual neighborhood which uses not only spatial adjacent relationships under the fine scale, but also the father-child relationships between the two scales. After that, the constructed hyperedges can represent objects with complex structure more comprehensively. With the help of the structural information, the hypergraph convolution operations can accurately propagate the label information from labeled nodes to unlabeled ones. The experimental results on both homogeneous and heterogeneous remote sensing images have demonstrated the superiority of the proposed DNHGNN. Our future work is to extend the framework to time-series images.

Author Contributions: Conceptualization, J.W. and Y.S.; methodology, J.W.; software, J.W. and Q.L.; validation, J.W., W.N. and K.C.; formal analysis, Q.L.; investigation, J.W. and Y.S.; resources, J.W.; data curation, R.F. and K.C.; writing—original draft preparation, J.W.; writing—review and editing, J.W. and W.N.; visualization, J.W.; supervision, B.L.; project administration, B.L.; funding acquisition, R.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 62001482.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare they have no conflict of interest.

References

1. Singh, A. Review article digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* **2010**, *10*, 989–1003. [\[CrossRef\]](#)
2. Canuti, P.; Casagli, N.; Ermini, L.; Fanti, R.; Farina, P. Landslide activity as a geoinicator in Italy: Significance and new perspectives from remote sensing. *Environ. Geol.* **2004**, *45*, 907–919. [\[CrossRef\]](#)
3. Huang, X.; Gao, Y.; Li, J. An automatic change detection method for monitoring newly constructed building areas using time-series multi-view high-resolution optical satellite images. *Remote Sens. Environ.* **2020**, *244*, 111802. [\[CrossRef\]](#)
4. Hussain, M.; Chen, D.; Cheng, A.; Wei, H.; Stanley, D. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS J. Photogramm. Remote Sens.* **2013**, *80*, 91–106. [\[CrossRef\]](#)
5. Bruzzone, L.; Prieto, D.F. Automatic analysis of the difference image for unsupervised change detection. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 1171–1182. [\[CrossRef\]](#)
6. Thonfeld, F.; Feilhauer, H.; Braun, M.; Menz, G. Robust change vector analysis (RCVA) for multi-sensor very high resolution optical satellite data. *Int. J. Appl. Earth Observ. Geoinf.* **2016**, *50*, 131–140. [\[CrossRef\]](#)
7. Celik, T. Unsupervised change detection in satellite images using principal component analysis and K-means clustering. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 772–776. [\[CrossRef\]](#)
8. Wu, C.; Du, B.; Zhang, L. Slow feature analysis for change detection in multispectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2858–2874. [\[CrossRef\]](#)
9. Bovolo, F.; Bruzzone, L. A detail-preserving scale-driven approach to change detection in multitemporal SAR images. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2963–2972. [\[CrossRef\]](#)
10. Inglada, J.; Mercier, G. A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis. *IEEE Trans. Geosci. Remote Sens.* **2005**, *45*, 1432–1445. [\[CrossRef\]](#)
11. Yang, G.; Li, H.; Yang, W.; Fu, K.; Sun, Y.; Emery, W.J. Unsupervised change detection of SAR images based on variational multivariate Gaussian mixture model and Shannon entropy. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 826–830. [\[CrossRef\]](#)
12. Liu, Z.; Li, G.; Mercier, G.; He, Y.; Pan, Q. Change detection in heterogeneous remote sensing images via homogeneous pixel transformation. *IEEE Trans. Image Process.* **2018**, *27*, 1822–1834. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Ferraris, V.; Dobigeon, N.; Cavalcanti, Y.; Oberlin, T.; Chabert, M. Coupled dictionary learning for unsupervised change detection between multimodal remote sensing images. *Comput. Vis. Image Underst.* **2019**, *189*, 102817. [\[CrossRef\]](#)
14. Chatelain, F.; Tournet, J.Y.; Inglada, J. Change detection in Multisensor SAR images using bivariate Gamma distributions. *IEEE Trans. Image Process.* **2008**, *17*, 249–258. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Chen, Y.; Cao, Z. An improved MRF-based change detection approach for multitemporal remote sensing imagery. *Signal Process.* **2013**, *93*, 163–175. [\[CrossRef\]](#)
16. Chang, N.; Han, M.; Yao, W.; Chen, L.; Xu, S. Change detection of land use and land cover in an urban region with SPOT-5 images and partial Lanczos extreme learning machine. *J. Appl. Remote Sens.* **2010**, *4*, 043551.
17. Gong, M.; Zhang, P.; Liu, J. Coupled dictionary learning for change detection from multisource data. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7077–7091. [\[CrossRef\]](#)
18. Zhang, L.; Hu, X.; Zhang, M.; Shu, Z.; Zhou, H. Object-level change detection with a dual correlation attention-guided detector. *ISPRS J. Photogramm. Remote Sens.* **2021**, *177*, 147–160. [\[CrossRef\]](#)
19. Yousif, O.; Ban, Y. A novel approach for object-based change image generation using multitemporal high-resolution SAR images. *Int. J. Remote Sens.* **2017**, *38*, 1765–1787. [\[CrossRef\]](#)
20. Wang, X.; Liu, S.; Du, P.; Liang, H.; Xia, J.; Li, Y. Object-based change detection in urban areas from high spatial resolution images based on multiple features and ensemble learning. *Remote Sens.* **2018**, *10*, 276. [\[CrossRef\]](#)
21. Lv, Z.; Liu, T.; Benediktsson, J.A. Object-oriented key point vector distance for binary land cover change detection using VHR remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 6524–6533. [\[CrossRef\]](#)
22. Zhan, T.; Gong, M.; Jiang, X.; Zhang, M. Unsupervised scale-driven change detection with deep spatial-spectral feature for VHR images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 5653–5665. [\[CrossRef\]](#)
23. Chen, H.; Li, W.; Chen, S.; Shi, Z. Semantic-aware dense representation learning for remote sensing image change detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5630018. [\[CrossRef\]](#)
24. Bandara, W.; Patel, V.M. A transformer-based Siamese network for change detection. *IGARSS* **2022**, 207–210.
25. Chen, H.; Qi, Z.; Shi, Z. Remote sensing image change detection with transformers. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5607514. [\[CrossRef\]](#)
26. Liu, J.; Chen, K.; Xu, G.; Sun, X.; Yan, M.; Diao, W.; Han, H. Convolutional neural network-based transfer learning for optical aerial images change detection. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 127–131. [\[CrossRef\]](#)
27. Chen, H.; Shi, Z. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sens.* **2020**, *12*, 1662. [\[CrossRef\]](#)
28. Wu, C.; Chen, H.; Du, B.; Zhang, L. Unsupervised change detection in multitemporal VHR image based on deep kernel PCA convolutional mapping network. *IEEE Trans. Cybern.* **2021**, *52*, 12084–12098. [\[CrossRef\]](#)
29. Zhang, H.; Lin, M.; Yang, G.; Zhang, L. ESCNet: An end-to-end superpixel-enhanced change detection network for very-high-resolution remote sensing images. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *15*, 694. [\[CrossRef\]](#)

30. Wang, Z.; Peng, C.; Zhang, Y.; Wang, N.; Luo, L. Fully convolutional siamese networks based change detection for optical aerial images with focal contrastive loss. *Neurocomputing* **2021**, *457*, 155–167. [\[CrossRef\]](#)
31. Gong, M.; Niu, X.; Zhang, P.; Li, Z. Generative adversarial networks for change detection in multispectral imagery. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2310–2314. [\[CrossRef\]](#)
32. Liu, G.; Li, X.; Dong, Y. Style transformation-based spatial-spectral feature learning for unsupervised change detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *60*, 5401515. [\[CrossRef\]](#)
33. Pham, M.T.; Mercier, G.; Michel, J. Change detection between SAR image using a pointwise approach and graph theory. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2020–2032. [\[CrossRef\]](#)
34. Jia, L.; Wang, J.; Ai, J.; Jiang, Y. A hierarchical spatial-temporal graph-kernel for high-resolution SAR image change detection. *Int. J. Remote Sens.* **2020**, *41*, 3866–3885. [\[CrossRef\]](#)
35. Wu, J.; Li, B.; Qin, Y.; Ni, W.; Zhang, H. An object-based graph model for unsupervised change detection in high resolution remote sensing images. *Int. J. Remote Sens.* **2021**, *42*, 6212–6230. [\[CrossRef\]](#)
36. Sun, Y.; Lei, L.; Li, X.; Sun, H.; Kuang, G. Nonlocal patch similarity based heterogeneous remote sensing change detection. *Pattern Recognit.* **2021**, *109*, 107598. [\[CrossRef\]](#)
37. Wang, J.; Yang, X.; Yang, X.; Jia, L.; Fang, S. Unsupervised change detection between SAR images based on hypergraphs. *ISPRS J. Photogramm. Remote Sens.* **2020**, *164*, 61–72. [\[CrossRef\]](#)
38. Gori, M.; Monfardini, G.; Scarselli, F. A new model for learning in graph domains. in *Proc. IEEE Int. Joint Conf. Neural Netw.* **2005**, *2*, 729–734.
39. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 4–24. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Bruna, J.; Zaremba, W.; Szlam, A.; Lecun, Y. Spectral networks and locally connected networks on graphs. *arXiv* **2014**, arXiv:1312.6203.
41. Huang, Z.; Li, X.; Ye, Y.; Michael, K.N. MR-GCN: Multi-relational graph convolutional networks based on generalized tensor product. *IJCAI* **2020**, *20*, 1258–1264.
42. Shi, M.; Tang, Y.; Zhu, X.; Wilson, D.; Liu, J. Multi-class imbalanced graph convolutional network learning. *IJCAI* **2020**, *20*, 2879–2885.
43. Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. *arXiv* **2016**, arXiv:1606.09375.
44. Levie, R.; Monti, F.; Bresson, X.; Bronstein, M.M. CayleyNets: Graph convolutional neural networks with complex rational spectral filters. *IEEE Trans. Signal Process.* **2019**, *67*, 97–109. [\[CrossRef\]](#)
45. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2017**, arXiv:1609.02907.
46. Bacciu, D.; Errica, F.; Micheli, A. Contextual graph Markov model: A deep and generative approach to graph processing. *arXiv* **2018**, arXiv:1805.10636.
47. Xu, B.; Shen, H.; Cao, Q.; Qiu, Y.; Cheng, X. Graph wavelet neural network. *arXiv* **2019**, arXiv:1904.07785.
48. Qin, A.; Shang, Z.; Tian, J.; Wang, Y.; Zhang, T.; Tang, Y. Spectral-spatial graph convolutional networks for semisupervised hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 241–245. [\[CrossRef\]](#)
49. Wan, S.; Gong, C.; Zhong, P.; Du, B.; Zhang, L.; Yang, J. Multiscale dynamic graph convolutional network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3162–3177. [\[CrossRef\]](#)
50. Saha, S.; Mou, L.; Zhu, X.; Bovolo, F.; Bruzzone, L. Semisupervised change detection using graph convolutional network. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 607–611. [\[CrossRef\]](#)
51. Wu, J.; Li, B.; Qin, Y.; Ni, W.; Zhang, H.; Fu, R.; Sun, Y. A multiscale convolutional network for change detection in homogeneous and heterogeneous remote sensing images. *Int. J. Appl. Earth Observ. Geoinf.* **2021**, *105*, 102615. [\[CrossRef\]](#)
52. Tang, X.; Zhang, H.; Mou, L.; Liu, F.; Zhang, X.; Zhu, X.; Jiao, L. An unsupervised remote sensing change detection method based on multiscale graph convolutional network and metric learning. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–15. [\[CrossRef\]](#)
53. Feng, Y.; You, H.; Zhang, Z.; Ji, R.; Gao, Y. Hypergraph neural networks. *arXiv* **2019**, arXiv:1809.09401. [\[CrossRef\]](#)
54. Jiang, J.; Wei, Y.; Feng, Y.; Cao, J.; Gao, Y. Dynamic hypergraph neural networks. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*, Macao, China, 10–16 August 2019.
55. Baatz, M.; Schape, A. Multiresolution segmentation: An Optimization Approach for High Quality Multiscale Image Segmentation. Available online: <https://cir.nii.ac.jp/crid/1572261550679971840> (accessed on 17 January 2023).
56. Yang, Y.; Li, H.; Han, Y.; Gu, H. High resolution remote sensing image segmentation based on graph theory and fractal net evolution approach. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *40*, 197–201. [\[CrossRef\]](#)
57. Wan, L.; Xiang, Y.; You, H. An object-based hierarchical compound classification method for change detection in heterogeneous optical and SAR images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9941–9959. [\[CrossRef\]](#)
58. Daudt, R.C.; Saux, B.L.; Boulch, A. Fully convolutional Siamese networks for change detection. *arXiv* **2018**, arXiv:1810.08462.
59. Liu, R.; Jiang, D.; Zhang, L.; Zhang, Z. Deep depthwise separable convolutional network for change detection in optical aerial images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2020**, *13*, 1109–1118. [\[CrossRef\]](#)
60. Daudt, R.C.; Saux, B.L.; Boulch, A.; Gousseau, Y. Urban change detection for multispectral earth observation using convolutional neural networks. *arXiv* **2018**, arXiv:1810.08468.
61. Benedek, C.; Szirmay, T. Change detection in optical aerial images by a multilayer conditional mixed markov model. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3416–3430. [\[CrossRef\]](#)

62. Fujita, A.; Sakurada, K.; Imaizumi, T.; Ito, R.; Hikosaka, S.; Nakamura, R. Damage detection from aerial images via convolutional neural networks. In Proceedings of the 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), Nagoya, Japan, 8–12 May 2017; pp. 5–8.
63. Bai, S.; Zhang, F.; Philip. H.S. Torr. Hypergraph convolution and hypergraph attention. *Pattern Recognit.* **2021**, *110*, 107637. [\[CrossRef\]](#)
64. Lin, Y.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal loss for dense object detection. *arXiv* **2017**, arXiv:1708.02002.
65. Shi, Q.; Liu, M.; Li, S.; Liu, X.; Wang, F.; Zhang, L. A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5604816. [\[CrossRef\]](#)
66. Ji, S.; Wei, S.; Lu, M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Trans. Geosci. Remote Sens.* **2019**, *51*, 574–586. [\[CrossRef\]](#)
67. Li, K.; Li, Z.; Fang, S. Siamese NestedUnet networks for change detection of high resolution satellite image. In Proceedings of the 2020 1st International Conference on Control, Robotics and Intelligent System, Xiamen, China, 27–29 October 2020.
68. Zhang, C.; Yue, P.; Tapete, D.; Jiang, L.; Shangguan, B.; Huang, L.; Liu, G. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 183–200. [\[CrossRef\]](#)
69. Cao, F.; Dong, J.; Li, B.; Xu, Q.; Xie, C. Change detection from synthetic aperture radar images based on neighborhood-based ratio and extreme learning machine. *J. Appl. Remote sens.* **2016**, *10*, 046019.
70. Wang, R.; Zhang, J.; Chen, J.; Jiao, L.; Wang, M. Imbalanced learning-based automatic SAR images change detection by morphologically supervised PCA-net. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 554–558. [\[CrossRef\]](#)
71. Gao, F.; Dong, J.; Li, B.; Xu, Q. Automatic change detection in synthetic aperture radar images based on PCANet. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1792–1796. [\[CrossRef\]](#)
72. Gao, F.; Wang, X.; Gao, Y.; Dong, J.; Wang, S. Sea ice change detection in SAR images based on convolutional-wavelet neural networks. *IEEE Geosci. Remote Sens. Lett.* **2019**, *8*, 1240–1244. [\[CrossRef\]](#)
73. Li, Y.; Peng, C.; Chen, Y.; Jiao, L.; Zhou, L.; Shang, R. A deep learning method for change detection in synthetic aperture radar images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5751–5763. [\[CrossRef\]](#)
74. Mignotte, M. A fractal projection and Markovian segmentation-based approach for multimodal change detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8046–8058. [\[CrossRef\]](#)
75. Touati, R.; Mignotte, M.; Dahmane, M. Multimodal change detection in remote sensing images using an unsupervised pixel pairwise based Markov random field model. *IEEE Trans. Image Process.* **2019**, *29*, 757–767. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Sun, Y.; Lei, L.; Guan, D.; Kuang, G. Iterative robust graph for unsupervised change detection of heterogeneous remote sensing images. *IEEE Trans. Image Process.* **2021**, *30*, 6277–6291. [\[CrossRef\]](#) [\[PubMed\]](#)
77. Sun, Y.; Lei, L.; Li, M.; Kuang, G. Sparse-constrained adaptive structure consistency-based unsupervised image regression for heterogeneous remote-sensing change detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4405814. [\[CrossRef\]](#)
78. Sun, Y.; Lei, L.; Li, X.; Tan, X.; Kuang, G. Structured graph based image regression for unsupervised multimodal change detection. *ISPRS J. Photogramm. Remote Sens.* **2022**, *185*, 16–31. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.