DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images

Jie Chen [ID], Ziyang Yuan, Jian Peng, Li Chen, Haozhe Huang, Jiawei Zhu, Yu Liu [ID], and Haifeng Li [ID], *Member, IEEE*

Abstract—Change detection is a basic task of remote sensing image processing. The research objective is to identify the change information of interest and filter out the irrelevant change information as interference factors. Recently, the rise in deep learning has provided new tools for change detection, which have yielded impressive results. However, the available methods focus mainly on the difference information between multitemporal remote sensing images and lack robustness to pseudochange information. To overcome the lack of resistance in current methods to pseudochanges, in this article, we propose a new method, namely, dual attentive fully convolutional Siamese networks, for change detection in high-resolution images. Through the dual attention mechanism, long-range dependencies are captured to obtain more discriminant feature representations to enhance the recognition performance of the model. Moreover, the imbalanced sample is a serious problem in change detection, i.e., unchanged samples are much more abundant than changed samples, which is one of the main reasons for pseudochanges. We propose the weighted double-margin contrastive loss to address this problem by punishing attention to unchanged feature pairs and increasing attention to changed feature pairs. The experimental results of our method on the change detection dataset and the building change detection dataset demonstrate that compared with other baseline methods, the proposed method realizes maximum improvements of 2.9% and 4.2%, respectively, in the $F1$ score. Our PyTorch implementation is available at https://github.com/lehaifeng/DASNet.

Index Terms—Change detection, dual attention, high-resolution images, Siamese network, weighted double-margin contrastive (WDMC) loss.

I. INTRODUCTION

REMOTE sensing change detection is a technique for detecting and extracting the change information in a geographical entity or phenomenon through multiple observations

[1]. Remote sensing image change detection research plays a crucial role in land use and land cover change analysis, forest and vegetation change monitoring, ecosystem monitoring, urban expansion research, resource management, and damage assessment [2]–[9]. With the development of satellite imaging technology, many high-resolution remote sensing images can be obtained more easily. In high-resolution remote sensing images, ground objects have more space and shape features; hence, high-resolution remote sensing images become an important data source for change detection [10]. The effective extraction and learning of the rich feature information in high-scoring remote sensing images, reduction in the interference of pseudochanges, which means changes that are not truly occurring and changes that do not interest us, and further improvement in the accuracy of change detection are important issues in the field of remote sensing change detection.

Traditional change detection methods can be divided into two categories according to the research objects: pixel-based change detection methods and object-based change detection methods [11]. Pixel-based change detection methods typically generate a difference graph by directly comparing the spectral information or texture information of the pixels and obtain the final result graph via threshold segmentation or clustering [12]–[18]. Although this method is simple to implement, it ignores the spatial context information and generates a substantial amount of salt-and-pepper noise during processing. Another type of method divides the remote sensing image into disjoint objects and uses the rich spectral, textural, structural, and geometric information in the image to analyze the differences of temporal images [19]–[23]. Although this type of method uses the spatial context information of high-resolution remote sensing images, traditional manual feature extraction is complex and exhibits poor robustness.

In recent years, deep learning has yielded impressive results in image analysis, natural language processing, and other fields [24]–[26]. Neural networks with FCN [27] structures are widely used in remote sensing change detection tasks [28]–[33]. These methods can be divided into two categories: methods in the first category fuse unchanged images and changed images and input them into a network with FCN structure to detect the changes by maximizing the boundaries instead of directly measuring the changes. The methods in the other category detect changes by

measuring the distances between feature pairs that are extracted from images.

However, the available methods have low robustness to pseudochanges. There are two main reasons for this:

First, due to the lack of satisfactory features that can effectively distinguish between changed areas and unchanged areas, the available features are often sensitive to factors such as noise, angle, shadow, and context.

Second, the distributions of the changed and unchanged data are often severely unbalanced, namely, the number of unchanged samples is much larger than the number of changed samples.

To address the first problem, we propose the dual attentive fully convolutional Siamese networks (DASNet) framework. The basic strategy is to use a dual attention mechanism to locate the changed areas and to obtain more discriminant feature representations, which renders the learned features more robust to changes. Many previous studies [34], [35] have demonstrated that more distinguishable feature representations can be obtained by capturing long-range dependencies[35]. Attention mechanisms can model long-range dependencies and have been widely used in many tasks [36]–[41]. Among them, the self-attention mechanism [42], [43] explores the performances of nonlocal operations of images and videos in the space-time dimension. The self-attention mechanism is of substantial significance for the long-range dependencies of images. Additionally, many researchers[44]–[46] began to combine spatial attention and channel attention so that the network could not only focus on the region of interest but also improve the discrimination of features. Therefore, we introduce an extended attention mechanism [46], which is composed of a channel attention module and a spatial attention module, for obtaining more discriminative image features to enhance the model's performance in recognizing changes and to improve its robustness to pseudochanges.

Aiming at overcoming the second problem, due to the imbalance of data in change detection and the unbalanced contributions to the network of the changed feature pairs and the unchanged feature pairs in the contrastive loss [47] of traditional Siamese networks, we established with weighted double-margin contrastive (WDMC) loss. The WDMC loss can mitigate the effects of having more unchanged regions than changed regions in the original data by setting weight coefficients. It can also alleviate the imbalance in the punishment between the unchanged feature pairs and the changed feature pairs during network training by setting double margins.

The main contributions of this article are as follows.

1) We propose a new remote sensing change detection method that is based on deep metric learning, which uses dual attention modules to improve feature discrimination to more robustly distinguish changes.

2) We propose the WDMC loss, which increases the distance between changed feature pairs and reduces the distance between unchanged feature pairs while balancing the impacts of changed regions and unchanged regions on the network. Thereby enhancing the network's performance in recognizing change information and its robustness to pseudochanges.

3) Compared with the selected baselines, the proposed method yielded SOTA results on the change detection dataset (CDD) [48] and the building change detection dataset (BCDD) [49], and the $F1$ scores reached 91.9% and 89.8%, respectively. Compared with other baseline methods, the maximum increases were 2.9% and 4.2%, respectively.

The rest of this article is organized as follows. Section II reviews the related works. Section III describes the proposed method in detail. To evaluate our method, experiments are designed in Section IV, and the proposed method is discussed. Finally, Section V concludes this article.

II. RELATED WORKS

Change detection is a basic task in the field of remote sensing, and researchers have developed many change detection technologies for this task. Remote sensing change detection methods typically include feature extraction and changing area identification. The objective of the former is to extract meaningful features, such as color distribution, texture characteristics, and context information. The objective of the latter is to use technical algorithms to analyze previously extracted features to identify the changing regions in multitemporal remote sensing images. Based on the techniques that are utilized in these change detection methods, we divide the remote sensing change detection methods into traditional change detection methods and deep learning change detection methods.

Traditional change detection methods mainly accept the feature differences and ratios of pairs of pixels or objects as input and detect changes by determining thresholds. Change vector analysis [12] conducts different calculations on the data of each band of images in various periods to obtain the change amount in each pixel in each band to form a change vector, and it identifies the changed and unchanged regions using thresholds. Principal component analysis [14], which is a classical mathematical transformation method, obtains the difference image by extracting the effective information in the image bands, performs the difference calculation with the first principal component and obtains the change map via threshold segmentation. Multivariate alteration detection [50] extracts change information by maximizing the variance in the image difference. Slow feature analysis [51] can extract invariant features from multitemporal remote sensing image data, reduce the radiation differences of unchanged pixels, and improve the separability between changed and unchanged pixels.

The remote sensing change detection methods that are based on deep learning use the features that are extracted from multitemporal images by deep neural networks to determine the change information in ground objects. In the field of natural image change detection, deep learning methods perform well [52], [53]. In the field of remote sensing, the strategy of using deep learning for change detection has also been utilized. SCCN [28] uses a deep symmetrical network to study changes in remote sensing images, and DSCN [29] uses two branch networks that share weights for feature extraction and uses the features that
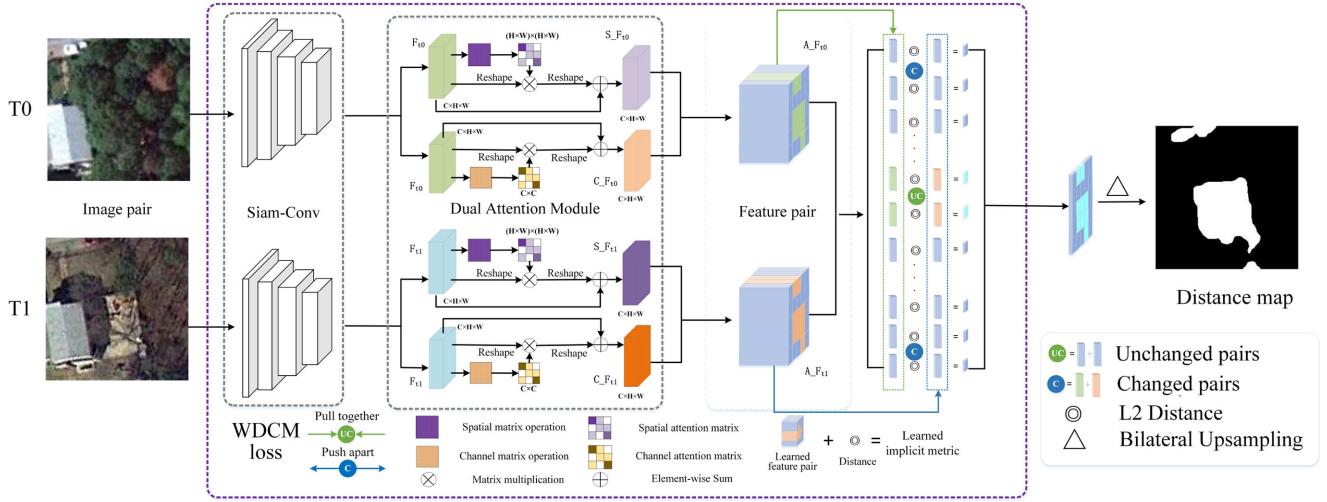
Fig. 1.   Overview of the DASNet.

are obtained by the last layer of the two branches for threshold segmentation to obtain a binary change map. CDNet [30], FC-EF, FC-Siam-Conc, FC-Siam-Diff in [31], and BiDateNet [32] implement end-to-end training on the CDD and learn the decision boundary to obtain the change map.

In contrast to the above methods, the proposed method directly measures changes and obtains a more discriminative feature representation through the dual attention module, which can capture long-range dependencies. Our proposed method uses the WDMC loss to improve the degree of intraclass compactness and to balance the influences of the changed regions and the unchanged regions on the network to enhance the recognition performance of the network for the changed information.

III. METHODOLOGY

In this section, we introduce the proposed network in detail. First, the overall structure of the network that is proposed in this article is introduced (see Fig. 1). The spatial attention and channel attention modules are explained subsequently. Finally, the proposed WDMC loss function is introduced.

A. Overview

Compared with general optical remote sensing images, high-resolution remote sensing images have more abundant information and higher requirements for feature extraction. We first use the Siam-Conv module to generate local features (F_{t0}, F_{t1}) by inputting high-resolution image pairs that are obtained at different times in the same region. Then, the dual attention mechanism is used to establish the connections between local features, which are used to obtain global context information to better distinguish the changing area from the unchanged area.

Consider spatial attention as an example, which generates new features that contain spatial long-range contextual information via the following three steps. The first step is to use a spatial attention matrix to model the spatial relationship between any two pixels of the features obtained from Siam-Conv. The next step

is to conduct matrix multiplication between the spatial attention matrix and the original features. Third, the final representations are obtained by applying the element-by-element summation operation on the matrix that was obtained in the second step, and the original features. The channel attention module captures the long-range context information in the channel dimension.

The process of capturing channel relationships is similar to that of spatial attention modules. The first step is to compute the channel attention matrix in the channel dimension. The second step is to conduct matrix multiplication between the channel attentional matrix and the original features. Finally, the matrix that is obtained in the second step and the original features are summed element-by-element. After that, the outputs of the two attention modules are aggregated to obtain a better representation of the features.

Then, the features A_F_{t0} and A_F_{t1} that are obtained through the dual attention mechanism are mapped to the feature space. We use the WDMC loss function to decrease the pixel pair distance of the unchanged area and to increase the pixel pair distance of the changing area. Our method uses metric learning and uses the distance between deep features as the basis for discrimination, thereby substantially increasing the accuracy of change detection.

B. Spatial Attention Mechanism

As discussed previously, discriminant feature representations are important for determining the types of features and, thus, for identifying changed and unchanged areas. However, many studies [54], [55] have shown that using only local features that are generated by traditional FCNs may lead to misclassification. To model the rich context of local features, we introduce a spatial attention module. The spatial attention module encodes the context information of a long range into local features, thereby enhancing the feature representations.

As illustrated in Fig. 2, the feature $F \in \mathbb{R}^{C \times H \times W}$, where C represents the number of channels of feature F, whereas H
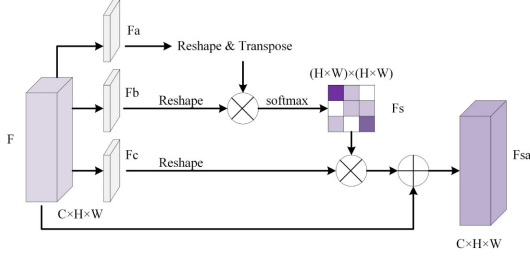
Fig. 2. Spatial attention mechanism.

Fig. 3. Channel attention mechanism.

and W represent the height and width of feature F, respectively, that is obtained by Siam-Conv is input into three convolutional layers that have the same structure to obtain three new features, namely, Fa, Fb, and Fc, where $\{Fa, Fb, Fc\} \in \mathbb{R}^{C \times H \times W}$. Then, we reshape Fa and Fb to $\mathbb{R}^{C \times N}$, where $N = H \times W$. After that, we conduct matrix multiplication between the transpose of Fb and Fa and obtain the spatial attention map $Fs \in \mathbb{R}^{N \times N}$ through a softmax layer

$$Fs_{ji} = \frac{\exp(Fa_i \cdot Fb_j)}{\sum_{i=1}^{N} \exp(Fa_i \cdot Fb_j)} \quad . \tag{1}$$

Fs_{ji} can be used to measure the effectiveness of the feature at position i on the feature at position j. The stronger the connection between the two features, the larger the value of Fs_{ji}.

We reshape Fc to $\mathbb{R}^{C \times N}$ and conduct matrix multiplication with Fs to obtain the result and reshape it to $\mathbb{R}^{C \times H \times W}$. Finally, we multiply the result from the previous step by a scale parameter η and perform an elementwise summation operation with F to obtain the final output

$$Fsa_j = \eta \sum_{i=1}^{N} (Fs_{ji}Fc_j) + F_j \tag{2}$$

where η is initialized as 0 and gradually learns to assign more weight. From formula (2), it can be concluded that the resulting feature Fsa at each position is the result of a weighted sum of the features at all positions and the original features. Therefore, it has a global context view and selectively aggregates contexts based on spatial attention maps. Similar semantic features promote each other, which improves the compactness and semantic consistency within the class and enables the network to better distinguish between changes and pseudochanges.

C. Channel Attention Mechanism

Each high-level feature channel map can be regarded as a response to a ground object, and the semantic responses are related to each other. By utilizing the correlation between channel maps, interdependent feature maps can be enhanced, and feature representations with specified semantics can be improved to better distinguish changes. Therefore, we construct a channel attention module for establishing the relationships between channels.

As illustrated in Fig. 3, in contrast to the spatial attention module, the convolution operation is not used to obtain the new features in the channel attention module. Feature $F \in \mathbb{R}^{C \times H \times W}$ that was obtained by Siam-Conv is reshaped to $\mathbb{R}^{C \times N}$, where
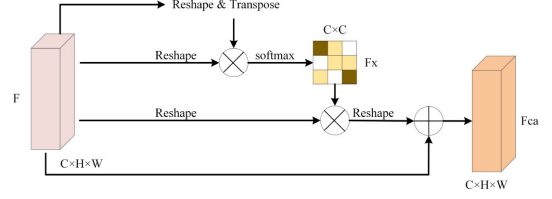
$N = H \times W$. After that, matrix multiplication is performed between the transpose of F and F to obtain the channel attention map $Fx \in \mathbb{R}^{N \times N}$ through a softmax layer

$$Fx_{ji} = \frac{\exp(F_i \cdot F_j)}{\sum_{i=1}^{C} \exp(F_i \cdot F_j)} \quad . \tag{3}$$

Fx_{ji} can be used to measure the impact of the ith channel on the jth channel. Similarly, the stronger the connection between the two channels, the larger the value of Fx_{ji}.

We reshape F to $\mathbb{R}^{C \times N}$ and conduct matrix multiplication with Fx to obtain the result. Finally, we multiply the result from the previous step by a scale parameter γ and conduct an elementwise summation operation with F to obtain the final output

$$Fca_j = \gamma \sum_{i=1}^{C} (Fc_{ji}F_i) + F_j \tag{4}$$

where γ is initialized as 0 and gradually learns to assign more weight. From formula (4), it is concluded that the final feature of each channel is the result of a weighted sum of the features of all channels and the original feature, which models the long-term semantic dependence between the feature graphs. It enhances the identifiability of the feature and highlights the feature representation of the region of change.

D. WDMC Loss Function

In the traditional contrastive loss, a problem of imbalance is encountered in the punishment for the changed feature pairs and the unchanged feature pairs during training. The traditional contrastive loss can be formulated as

$$\text{contrastive loss} = \sum_{i,j} \frac{1}{2} [(1 - y_{i,j}) \, d_{i,j}^2 + y_{i,j} \max (d_{i,j} - m, 0)^2]. \tag{5}$$

We denote the feature map of the unchanged image through our model as f_0 and the feature map of the changed image through our model as f_1, $d_{i,j}$ is the distance between the feature vectors of f_0 and f_1 at (i,j), m is the margin that is enforced for changed feature pairs, and $y \in \{0, 1\}$, where $y = 0$ if the corresponding pixel pair is deemed unchanged, and $y = 1$ if it is deemed changed.

According to formula (5), the traditional contrastive loss does not contribute to the loss value only if the distance between the unchanged feature pairs is 0. However, in the remote sensing change detection task, the unchanged area is affected by various

imaging conditions, and the imaging difference is sometimes very large; hence, many noise changes will hinder the optimization of the distance between the unchanged feature pairs to 0. However, the changed feature pairs can only contribute to the loss if the distance exceeds the margin, which results in the imbalance of the punishment degree in the training process and affects the network's judgment of the change. In addition, the unchanged area in remote sensing change detection tasks is much larger than the changed area; hence, there is a severe imbalance between the unchanged samples and the changed samples. In response to these problems, we propose the WDMC loss function.

The proposed WDMC loss can be formulated as follows:

$$\text{WDMC loss} = \sum_{i,j} \frac{1}{2} \left[w_1 (1 - y_{i,j}) \max(d_{i,j} - m_1, 0)^2 \right.$$
$$\left. + w_2 \, y_{i,j} \max(d_{i,j} - m_2, 0)^2 \right] \quad (6)$$

where m_1 and m_2 represent the margins of the unchanged sample pairs and the changed sample pairs, respectively, and w_1 and w_2 represent the weights of the unchanged feature pairs and the changed feature pairs, respectively

$$w_1 = \frac{1}{P_U} \quad (7)$$

$$w_2 = \frac{1}{P_C} \quad (8)$$

where P_U and P_C are the frequencies of the changed and unchanged pixel pairs, respectively.

According to formula (6), by setting the margins for the unchanged sample pairs, we can alleviate the imbalance in the punishment between the unchanged feature pairs and the changed feature pairs during network training. We set the weight coefficients to mitigate the effects of having more unchanged regions than changed regions in the original data. Finally, these parameters balance the network's interest in the changed area and the unchanged area.

Inspired by the strategy of deep supervision [56], [57], we calculate the WDMC loss for the feature pairs that are obtained through the spatial attention module, the feature pairs that are obtained by the channel attention module and the final output feature pairs. λ_i represents the weight for each loss, L_{sa} represents the WDMC loss that is calculated using the feature pairs that are extracted from the spatial attention module, L_{ca} represents the WDMC loss that is calculated using the feature pairs that are extracted from the channel attention module, and L_e represents the WDMC loss that is calculated using the final output feature pairs.

Thus, our loss function is ultimately expressed as

$$\text{Loss} = \lambda_1 L_{sa} + \lambda_2 L_{ca} + \lambda_3 L_e. \quad (9)$$

In summary, we balance the contributions of the unchanged feature pairs and the changed feature pairs to the loss value during the training process and use the strategy of deep supervision to enhance the feature representation performance of the hidden layers. The discriminative feature representations improve the network's performance in recognizing changes.

E. Implementation Details

Now, we introduce in detail the relevant parameter design in the model design process.

First, we describe the design of the network structure. In terms of Siam-Conv network structure selection, we chose two basic networks: VGG16 [58] and ResNet50 [59]. For VGG16, we keep only the first five convolution modules, and we remove the max-pooling layer of the last module. In the first five convolution modules, the size of the convolution kernel is 3×3. For ResNet50, we remove the downsampling operations and employ dilated convolutions in the last two ResNet blocks. Then, the dual attention module composed of the spatial attention module and the channel attention module is connected to Siam-Conv to form the complete DSANet.

In terms of loss design, we set four parameters to balance the contributions of the unchanged region feature pairs and the changed region feature pairs to the network and to enhance the network's performance in the identification of changed regions. The values of parameters w_1 and w_2 are the pixel ratios of the changed area and the unchanged area in the dataset. The parameters m_1 and m_2 must be manually adjusted to optimize the performance of the model.

IV. EXPERIMENTS AND DISCUSSION

To evaluate the performance of the proposed method, we compared other change detection methods on the CDD and BCDD datasets, and we designed ablation experiments for evaluating the proposed structure and improved loss function. Finally, the experimental results are analyzed comprehensively.

A. Databases

1) CDD Dataset: CDD (see Fig. 4) is an open remote sensing change detection dataset. The dataset is composed of multisource remote sensing images with 11 original image pairs, which include seven pairs of seasonal change images with a size of 4725×2200 pixels and four pairs of images with a size of 1900×1000 pixels. The resolution varies from 3 to 100 cm per pixel, with the seasons varying widely among the bitemporal images. In [49], Ji *et al.* processed the original data to generate datasets with training sets of size 10 000 and test and validation sets of size 3000 via clipping and rotation.

2) BCDD Dataset: The BCDD dataset (see Fig. 5) covers the area of the 6.3 magnitude earthquake that struck Christchurch, New Zealand, in February 2011. The dataset contains two image scenes that were captured at the same location in 2012 and 2016, along with semantic labels and change detection labels for buildings. Since the size of each image is $32\,507 \times 15\,354$ pixels, we divided the two images into nonoverlapping 256×256 pixel image pairs. Then, we used the cropped images to form the training set, the verification set, and the test set according to the ratio of 8:1:1.

We counted the number of changed pixels and the number of unchanged pixels in the CDD dataset and the BCDD dataset, and the result can be seen in Table I. In the CDD dataset, the ratio of changed pixels to unchanged pixels was 0.147. In the

Fig. 4.   Multitemporal images that were selected as the training set from the CDD dataset. First row: unchanged images, second row: changed images, and third row: labels.
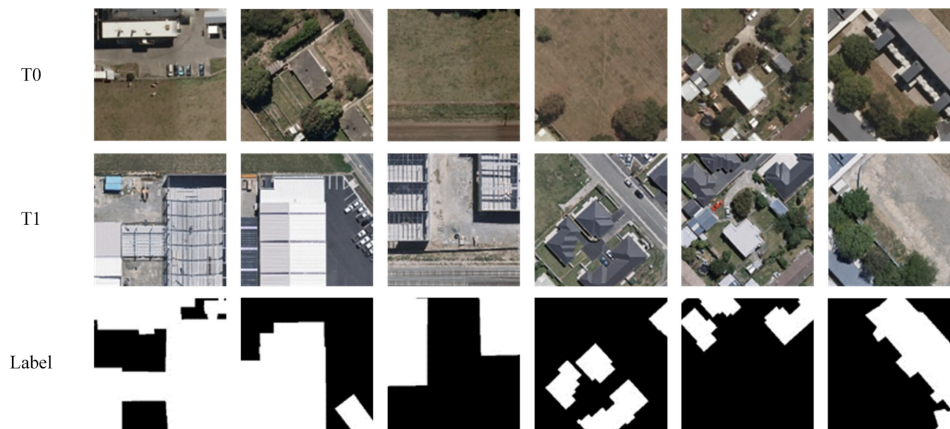

Fig. 5.   Multitemporal images that were selected as the training set from the BCDD dataset. First row: unchanged images, second row: changed images, and third row: labels.

TABLE I
STATISTICS ON THE NUMBER OF CHANGED PIXELS AND THE NUMBER OF
UNCHANGED PIXELS IN THE CDD DATASET AND BCDD DATASET

Dataset		changed pixels	unchanged pixels	c/uc
CDD	train	83,981,014	571,378,986	0.147
	val	24,676,497	171,800,431	0.144
	test	25,411,239	171,196,761	0.148
	total	134,068,750	914,376,178	0.147
BCDD	train	17,071,534	382,435,922	0.044
	val	1,854,764	48,083,668	0.039
	test	2,426,517	47,511,915	0.051
	total	21,352,815	478,031,505	0.045

BCDD dataset, the ratio was 0.045. This means that in these two datasets, the area that changed is much smaller than the area that has not changed.

B. Metrics and Implementation Details

To evaluate the performance of the proposed method, we used four evaluation indicators: precision (P), recall (R), $F1$ score ($F1$), and overall accuracy (OA). In the change detection task, the higher the precision value, the fewer false detections of

the predicted result occur, and the larger the recall value, the fewer predicted results were missed. $F1$ score and OA are the overall evaluation metrics of the prediction results. The larger their values, the better the prediction results. They are expressed as follows:

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{10}$$

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{11}$$

$$f_1 = \frac{2PR}{P + R} \tag{12}$$

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{13}$$

where TP is the number of true positives, FP is the number of false positives, TN is the number of true negatives, and FN is the number of false negatives.

The proposed method was implemented with PyTorch as the backend, which is powered by $3 \times$ GTX TITAN XP. The Adam optimizer was adopted with a learning rate of 1e−4 as an
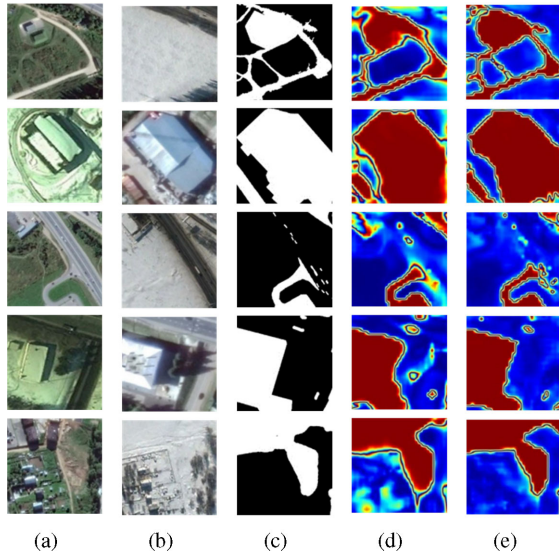
Fig. 6. Distance map visualizations for various layers: (a) unchanged image, (b) changed image, (c) label, (d) distance map of DASNet (VGG16) with the cosine similarity, (e) distance map of DASNet (VGG16) with the l2 distance. The changed parts are shown in white, whereas the unchanged parts are shown in black.

Fig. 7. Influence of the values of m_1 and m_2 on the performance of DSANet.

TABLE II
INFLUENCE OF DIFFERENT VALUES OF λ_1, λ_2, AND λ_3

λ_1	λ_2	λ_3	Rec	Pre	F1	OA
0	0	1	0.924	0.921	0.922	0.981
0.25	0.25	0.5	0.927	0.925	0.926	0.982
1	1	1	0.932	0.922	0.927	0.982

TABLE III
LOSS COMPARISON STUDIES ON THE CDD DATASET

Method	BaseNet	Rec	Pre	F1	OA
Siam-Conv with contrastive loss	VGG16	0.822	0.913	0.859	0.967
Siam-Conv with WDMC loss(cos)	VGG16	0.889	0.827	0.856	0.963
Siam-Conv with WDMC loss(l2)	VGG16	0.896	0.888	0.892	0.973
DASNet with contrastive loss	VGG16	0.844	0.915	0.878	0.971
DASNet with WDMC loss(cos)	VGG16	0.896	0.841	0.871	0.970
DASNet with WDMC loss(l2)	VGG16	**0.925**	**0.914**	**0.919**	**0.980**
Siam-Conv with contrastive loss	ResNet50	0.841	0.915	0.876	0.971
Siam-Conv with WDMC loss(cos)	ResNet50	0.893	0.841	0.866	0.969
Siam-Conv with WDMC loss(l2)	ResNet50	0.902	0.898	0.900	0.975
DASNet with contrastive loss	ResNet50	0.878	0.905	0.891	0.973
DASNet with WDMC loss(cos)	ResNet50	0.901	0.879	0.890	0.973
DASNet with WDMC loss(l2)	ResNet50	**0.932**	**0.922**	**0.927**	**0.982**

optimization algorithm, and the batch size of the training data was set to 8. According to the experimental performance, the value of m_1 was set to 0.3, the value of m_2 was set to 2.2, and the values of λ_1, λ_3, and λ_2 were set to 1.

C. Effect of the WDMC Loss Function

The measurement values differ between feature pairs that are obtained by different distance metrics. In the task of change detection, a suitable distance metric must increase the distance between the changed feature pairs and decrease the distance between the unchanged feature pairs.

To obtain a suitable distance metric, we evaluated the impacts of the l2 distance and the cosine similarity on the results of the change detection task. We visualize the results from various distance metrics. As shown in Fig. 6, we visualized the output of two trained DASNet networks, which were all set up in the same way except for different distance metrics, and we obtained the distance maps based on various distance metrics. For the same feature layer, the background of the receiving map that was obtained using the cosine similarity was noisier, and its performance in describing the changing boundary was weaker. Thus, the performance of this metric in excluding noisy changes was weak. The resulting graph that was obtained by using the second-order Euclidean distance had a cleaner background and a higher contrast in the front background, which endowed the model with stronger change recognition performance.

In the change detection task, there were far more unchanging sample pairs than changing sample pairs. The traditional contrastive loss has an unbalanced penalty for unchanged feature pairs and changed feature pairs. As a result, the network with the traditional contrastive loss performed weakly in discerning change information. We propose the WDMC loss to strengthen the network's performance in the identification of change information.
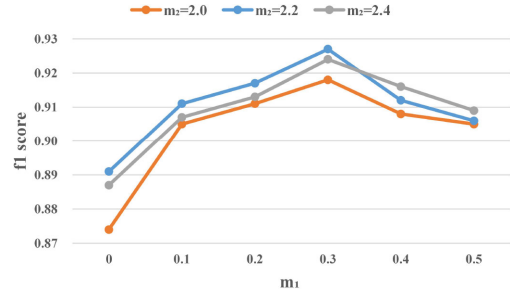
Since the selection of the values of m_1 and m_2 will have a great influence on the results of the network, many experiments were designed to find the values of m_1 and m_2 with the best performance. Fig. 7 shows the performance of a DSANet network using ResNet50 as the backbone on the CDD dataset with different $m1$ and $m2$ values. We also tested the effect of different values of λ_1, λ_2 and λ_3 on the CDD data set to the final result. As listed in Table II, the appropriate choice of λ_1, λ_2, and λ_3 improved the accuracy of the model to a certain extent. Table III lists the comparison between the improved loss function and the traditional contrastive loss, and precision, recall, $F1$ score, and OA were used as judgment indicators to evaluate the performance.

According to the results in Table III, when we used VGG16 as the backbone, compared with the traditional contrastive loss, the recall and $F1$ score of our proposed WDMC loss function with l2 distance improved by 7.4% and 3.3%, respectively, and OA improved by 0.6%, whereas other experimental conditions remained unchanged. When we used ResNet50 as the backbone, for Siam-Conv, compared with the traditional contrastive loss, the recall and $F1$ score of our proposed WDMC loss function with l2 distance improved by 6.1% and 2.4%, respectively, and
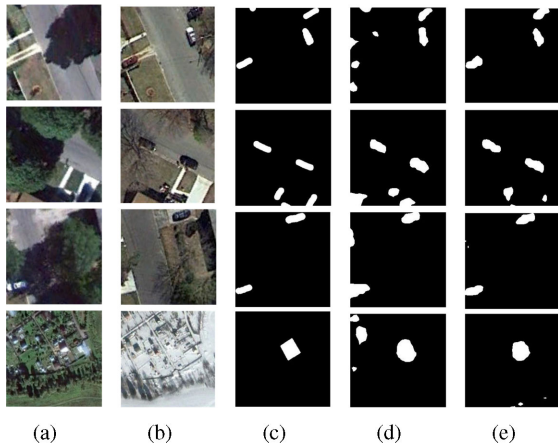
Fig. 8. Visualization results of Siam-Conv (ResNet50) using different losses: (a) unchanged image, (b) changed image, (c) label, (d) result of Siam-Conv (ResNet50) with contrastive loss, and (e) result of Siam-Conv (ResNet50) with WDMC loss. The changed parts are shown in white, whereas the unchanged parts are shown in black.

Fig. 9. Change process of the two parameters η and γ in 30 epochs.

TABLE IV
ATTENTION MECHANISM ABLATION RESEARCH ON THE CDD DATASET, WHERE CAM DENOTES CHANNEL ATTENTION AND SAM DENOTES SPATIAL ATTENTION

Method	BaseNet	SAM	CAM	Rec	Pre	F1	OA
Siam-Conv	VGG16			0.896	0.888	0.892	0.973
DASNet	VGG16	✓		0.919	0.901	0.910	0.978
DASNet	VGG16		✓	0.922	0.892	0.906	0.976
DASNet	VGG16	✓	✓	**0.925**	**0.914**	**0.919**	**0.980**
Siam-Conv	ResNet50			0.902	0.898	0.900	0.974
DASNet	ResNet50	✓		0.927	0.903	0.916	0.979
DASNet	ResNet50		✓	0.930	0.894	0.912	0.978
DASNet	ResNet50	✓	✓	**0.932**	**0.922**	**0.927**	**0.982**

OA improved by 0.4%, whereas other experimental conditions remained unchanged. We also carried out experiments on the improved loss and traditional contrastive loss based on the proposed method, and they also proved the effectiveness of the proposed WDMC loss. Therefore, the use of the proposed WDMC loss can improve the performance of the network.

To verify the robustness of the proposed loss function to pseudochanges, we visualized the output results of Siam-Conv using traditional contrastive loss. Similarly, we also visualized the output results of Siam-Conv using WDMC loss. The use of Siam-Conv is to avoid the influence of the dual attention mechanism on the results. It can be clearly seen in Fig. 8 that the network trained with WDMC loss was more resistant to pseudochanges. The network trained with WDMC loss was better able to distinguish the pseudochanges due to seasonal changes, sensor changes, etc. However, the traditional contrastive loss could not do this well.

D. Ablation Study for Attention Modules

To better focus on the change regions and to improve the recognition accuracy of the network, on the basis of the WDMC loss, a dual attention module was incorporated for the identification of long-range dependencies to obtain better feature representations. We designed ablation experiments to evaluate the performance of the dual attention mechanism. As Fig. 9 shows, we also give the change process of the two parameters η and γ in 30 epochs.

According to Table IV, the dual attention module improved the models performance comprehensively. Compared with the baseline Siamese network (a Siamese network that is based on VGG16 or ResNet50), for using VGG16 as the backbone, the recall, precision, $F1$ score and OA of the network with the spatial attention module were 0.919, 0.901, 0.910, and 0.978, respectively, which correspond to increases of 2.3%, 1.3%, 1.8%, and 0.5%, respectively. For using ResNet50 as
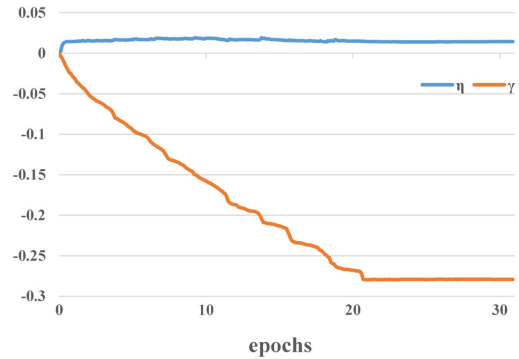
the backbone, the recall, precision, $F1$ score, and OA of the network with the spatial attention module were 0.927, 0.903, 0.916, and 0.979, respectively, which correspond to increases of 2.5%, 0.5%, 1.6%, and 0.5%, respectively. The recall, precision, $F1$ score, and OA of the network that used VGG16 as the backbone with the channel attention module were 0.922, 0.892, 0.906, and 0.976, respectively, which correspond to increases of 2.6%, 0.4%, 1.4%, and 0.3%, respectively. The recall, precision, $F1$ score, and OA of the network that used ResNet50 as the backbone with the channel attention module were 0.930, 0.894, 0.912, and 0.978, respectively, which correspond to increases of 2.8%, 0.4%, 1.2%, and 0.4%, respectively. After both the spatial attention module and the channel attention module were added into the network, the network that used VGG16 performance was further improved. The recall, precision, $F1$ score and OA were 0.925, 0.914, 0.919, and 0.980, respectively, which correspond to increases of 2.9%, 2.6%, 3.7%, and 0.7%, respectively. The network that used ResNet50 performance was further improved. The recall, precision, $F1$ score and OA were 0.932, 0.922, 0.927, and 0.982, respectively, which correspond to increases of 3.0%, 2.4%, 2.7%, and 0.6%, respectively.

The addition of a dual attention module to the Siamese network overcomes the problem that the WDMC loss function cannot improve the precision while improving the recall. Both the spatial attention mechanism and the channel attention mechanism improve the accuracy of the network. When these two attention mechanisms are combined, the spatial information and channel information of the features are fully utilized, and the comprehensive performance of the network is improved.
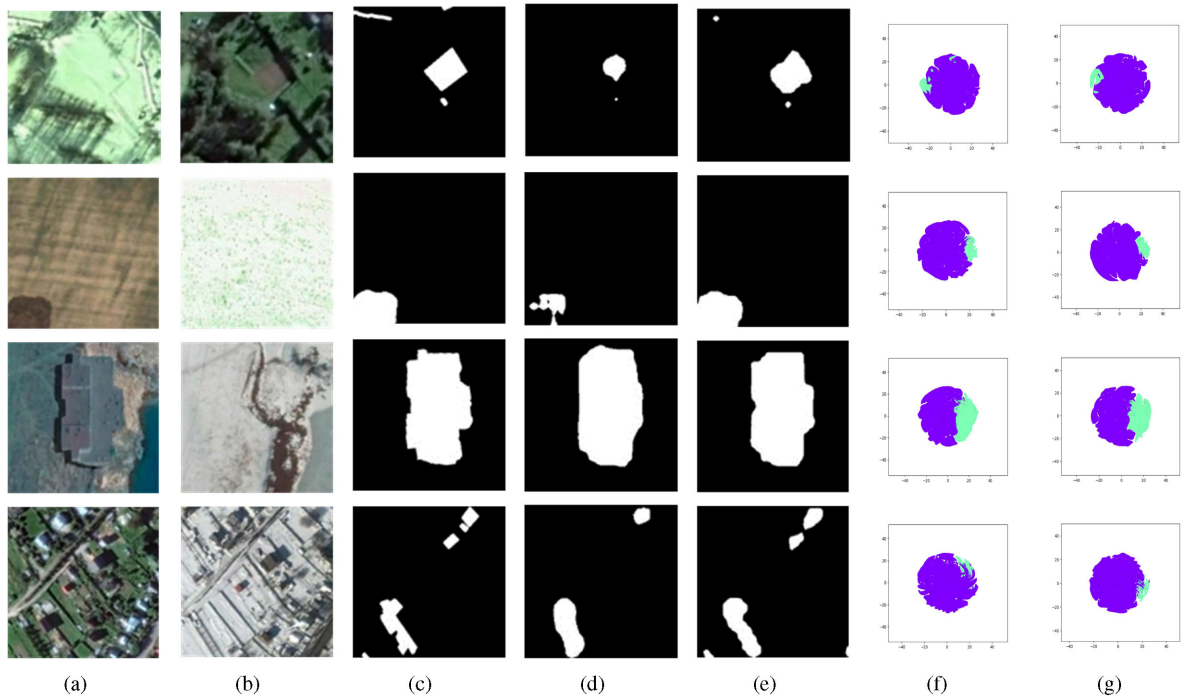
Fig. 10. t-SNE visual comparison diagram: (a) unchanged image, (b) changed image, (c) label, (d) result of Siam-Conv, (e) result of DASNet, (f) the last feature layer of Siam-Conv, and (g) the last feature layer of DASNet. The changed parts are shown in white, whereas the unchanged parts are shown in black.

E. Visualization of the Dual Attention Mechanism Effect

To more intuitively reflect the role of the dual attention module, we used the t-SNE [60] algorithm to visualize the last feature layer of the Siamese net without adding the dual attention mechanism (Siam-Conv) and the last feature layer of the proposed model (see Fig. 10). In the t-SNE diagram, the purple parts represent the unchanged feature vectors in 2 D, and the green parts represent the changed feature vectors in 2 D. Compared with Siam-Conv, the main advantage of the DASNet model was that changed and unchanged features were both well clustered, with strong discrimination between them. This enlarged the margin between the types of features and effectively increased the accuracy of feature classification.

To further illustrate the effectiveness of the attention mechanism, we applied the Grad-CAM [61] to different networks using images from the CDD dataset. Grad-CAM is a visualization method that uses gradients to calculate the importance of spatial positions in the convolutional layer. We compared the visualization results of Siam-Conv and DASNet. Fig. 11 illustrates the visualization results. In Fig. 11, we can clearly see that the Grad-CAM mask of DASNet covers the target area better than Siam-Conv. In other words, DASNet can make good use of the information in the target object area and aggregate characteristics from it.

Fig. 11 also demonstrates the robustness of the dual attention mechanism to pseudochanges. Siam-Conv focuses not only on change information but also on pseudochange due to seasonal changes in trees, agricultural fields, etc. This is a manifestation of the network's lack of resistance to the interference of
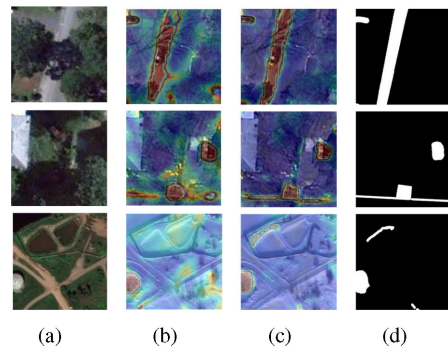

Fig. 11. Grad-CAM visualization results: (a) unchanged image, (b) result of the fusion of the changed image with Grad-CAM mask of the Siam-Conv, (c) result of the fusion of the changed image with Grad-CAM mask of the DASNet, and (d) label. The changed parts are shown in white, whereas the unchanged parts are shown in black.

pseudochange information. DASNet, however, is more concerned with real changes.

F. Performance Experiment

To evaluate the performance of the proposed method, we compared it with several image-based deep learning change detection methods: CDnet [30] is used in the study of street scene change detection. It is composed of contraction blocks and expansion blocks, and the change map is obtained through a softmax layer. FC-EF [31] stacks image pairs as the input images. It uses the skip connection structure to fuse the shallow
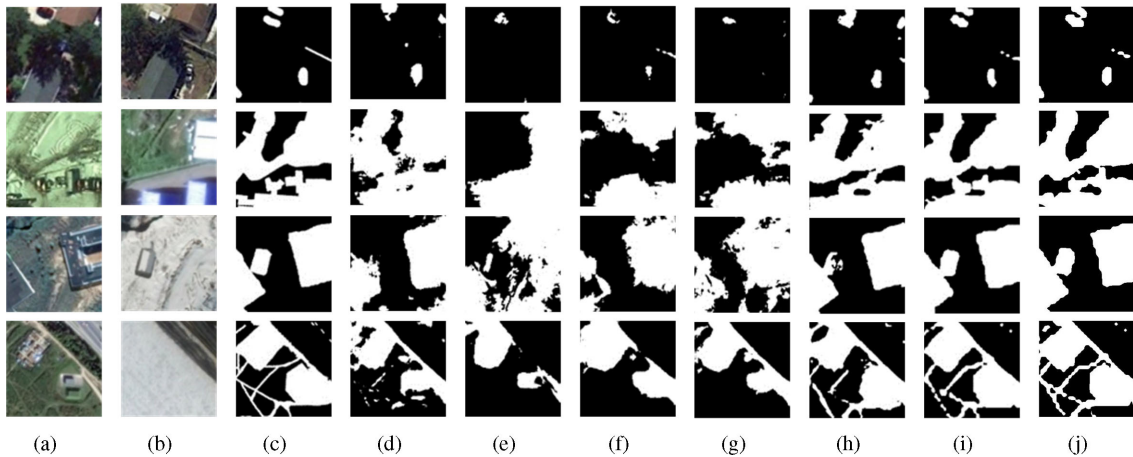
Fig. 12. Visualized comparison of the results of various change detection methods on the CDD dataset: (a) unchanged image, (b) changed image, (c) label, (d) CDnet, (e) FC-EF, (f) FC-Siam-Diff, (g) FC-Siam-Con, (h) BiDateNet, (i) DASNet(VGG16), and (j) DASNet(ResNet50). The changed parts are shown in white, whereas the unchanged parts are shown in black.
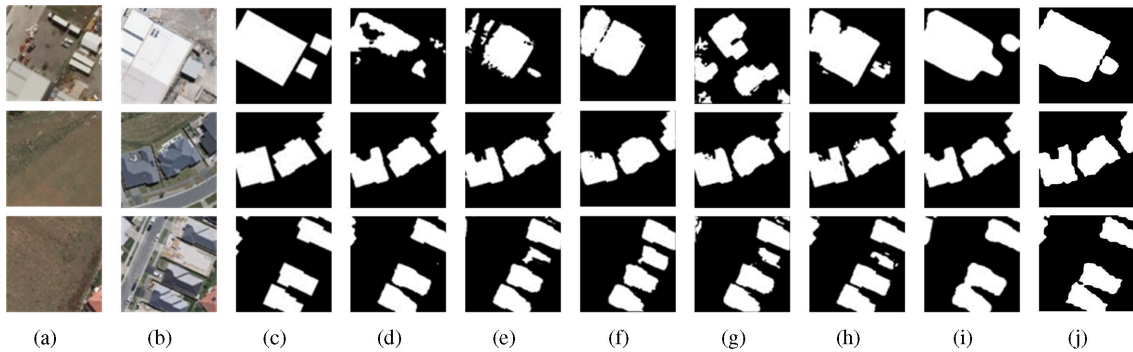

Fig. 13. Visualized comparison of the results of various change detection methods on the BCDD dataset: (a) unchanged image, (b) changed image, (c) label, (d) CDnet, (e) FC-EF, (f) FC-Siam-Diff, (g) FC-Siam-Con, (h) BiDateNet, (i) DASNet(VGG16), and (j) DASNet(ResNet50). The changed parts are shown in white, whereas the unchanged parts are shown in black.

and deep features and finally obtains the change map through the softmax layer. FC-Siam-Conc [31] is an extension of FC-EF. Its encoding layers are separated into two streams with equal structure and shared weights. Then, the skip connections are concatenated in the decoder part. FC-Siam-Diff [31] is another extension of FC-EF. It differs from FC-Siam-Conc in that it does not fuse the feature results that are obtained from the encoder streams in the decoder part but calculates the absolute value of the difference between the feature results. BidateNET [32] integrates the convolution module in LSTM [62] into UNet [63] so that the model can better learn the temporal change pattern.

We conducted comparative experiments with the above methods on the CDD and BCDD datasets.

From the data in Tables V and VI, the proposed method performed significantly better than the other change detection methods. For a more intuitive evaluation, we visualized the experimental results (see Figs. 12 and 13). According to Table IV, the proposed method realized satisfactory performance in remote sensing change detection. Compared with the other optimal change detection methods, the $F1$ score improved by 2.9% on the CDD dataset and by 4.2% on the BCDD dataset.

TABLE V
RESULTS ON THE CDD DATASET

Method	Rec	Pre	F1	OA
CDnet	0.817	0.827	0.822	0.964
FC-EF	0.528	0.684	0.596	0.911
FC-Siam-Diff	0.703	0.677	0.691	0.931
FC-Siam-Conc	0.666	0.710	0.687	0.925
BiDateNet	0.894	0.901	0.898	0.975
DSANet(VGG16)	**0.925**	**0.914**	**0.919**	**0.980**
DSANet(ResNet50)	**0.932**	**0.922**	**0.927**	**0.982**

To evaluate the performance of our method more intuitively, we visualized the experimental results. According to Figs. 12 and 13, the proposed method realized satisfactory results on both the CDD and BCDD datasets.

In Fig. 12, the first row demonstrates our model's performance in recognizing small changes in complex scenarios. The colors of images before and after the change in the second row vary substantially, but this has a highly limited impact on our network; there are many pseudochange interferences in the images before and after the change in the third row, whereas

TABLE VI
RESULTS ON THE BCDD DATASET

Method	Rec	Pre	F1	OA
CDnet	0.821	0.908	0.862	0.989
FC-EF	0.746	0.841	0.791	0.981
FC-Siam-Diff	0.710	0.700	0.704	0.971
FC-Siam-Conc	0.736	0.631	0.679	0.966
BiDateNet	0.819	0.889	0.852	0.986
DASNet(VGG16)	**0.905**	**0.892**	**0.898**	**0.990**
DASNet(ResNet50)	**0.905**	**0.900**	**0.910**	**0.991**

the change in the bottom-left corner of the fourth row is not readily observable. Our method performed well under both conditions.

According to Fig. 13, compared with other change detection methods, the proposed method focuses more on the changing area of interest. Since the BCDD is a dataset for building change detection, it is only necessary to identify building changes. Other changes, such as road changes, can be regarded as pseudochanges. The first row and the third row clearly show that the method proposed in this article not only performs well in identifying changes but also exhibits satisfactory resistance to the interference of pseudochanges.

V. CONCLUSION AND FUTURE WORK

In this article, we proposed a DASNet for high-score remote sensing image change detection, which directly measures changes by learning implicit metrics. To minimize the distance between the unchanged areas and maximize the distance between the changed areas, we used spatial attention and channel attention to obtain better feature representations and used the WDMC loss to balance the influences of the changed regions and the unchanged regions on the network. Compared with other baseline methods, our proposed network performed well on both the CDD and BCDD datasets. The Siamese network structure can learn the change representations of remote sensing images well, and the attention mechanism can describe the local features of the changes and recognize the pseudochanges.

In the future, we will conduct further research on small samples and on open-world and noisy environments to improve the mobility and robustness of change detection.

ACKNOWLEDGMENT

REFERENCES

[1] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989.

[2] G. Xian, C. Homer, and J. Fry, "Updating the 2001 National Land Cover Database land cover classification to 2006 by using Landsat imagery change detection methods," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1133–1147, 2009.

[3] Z. Y. Lv, W. Shi, X. Zhang, and J. A. Benediktsson, "Landslide inventory mapping from bitemporal high-resolution remote sensing images using change detection and multiscale segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1520–1532, May 2018.

[4] N. Sofina and M. Ehlers, "Building change detection using high resolution remotely sensed data and GIS," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 8, pp. 3430–3438, Aug. 2016.

[5] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Review ArticleDigital change detection methods in ecosystem monitoring: A review," *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1565–1596, 2004.

[6] H. Luo, C. Liu, C. Wu, and X. Guo, "Urban change detection based on Dempster–Shafer theory for multitemporal very high-resolution imagery," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 980.

[7] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Change detection techniques," *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2401, 2004.

[8] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using VHR optical and SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2403–2420, May 2010.

[9] M. E. Zelinski, J. Henderson, and M. Smith, "Use of Landsat 5 for change detection at 1998 Indian and Pakistani nuclear test sites," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3453–3460, Aug. 2014.

[10] L. Bruzzone and F. Bovolo, "A novel framework for the design of change-detection systems for very-high-resolution remote sensing images," *Proc. IEEE*, vol. 101, no. 3, pp. 609–630, Mar. 2013.

[11] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogrammetry Remote Sens.*, vol. 80, pp. 91–106, 2013.

[12] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.

[13] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and *k*-means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, Oct. 2009.

[14] J. Deng, K. Wang, Y. Deng, and G. Qi, "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data," *Int. J. Remote Sens.*, vol. 29, no. 16, pp. 4823–4838, 2008.

[15] C. Wu, B. Du, X. Cui, and L. Zhang, "A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion," *Remote Sens. Environ.*, vol. 199, pp. 241–255, 2017.

[16] C. Huang *et al.*, "Use of a dark object concept and support vector machines to automate forest cover change analysis," *Remote Sens. Environ.*, vol. 112, no. 3, pp. 970–985, 2008.

[17] G. Cao, Y. Li, Y. Liu, and Y. Shang, "Automatic change detection in high-resolution remote-sensing images by means of level set evolution and support vector machine classification," *Int. J. Remote Sens.*, vol. 35, no. 16, pp. 6255–6270, 2014.

[18] M. Volpi, D. Tuia, F. Bovolo, M. Kanevski, and L. Bruzzone, "Supervised change detection in VHR images using contextual information and support vector machines," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 20, pp. 77–85, 2013.

[19] Y. Zhang, D. Peng, and X. Huang, "Object-based change detection for VHR images based on multiscale uncertainty analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 1, pp. 13–17, Jan. 2018.

[20] C. Zhang, G. Li, and W. Cui, "High-resolution remote sensing image change detection by statistical-object-based method," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 7, pp. 2440–2447, Jul. 2018.

[21] J. L. Gil-Yepes, L. A. Ruiz, J. A. Recio, Á. Balaguer-Beser, and T. Hermosilla, "Description and validation of a new set of object-based temporal geostatistical features for land-use/land-cover change detection," *ISPRS J. Photogrammetry Remote Sens.*, vol. 121, pp. 77–91, 2016.

[22] Y. Qin, Z. Niu, F. Chen, B. Li, and Y. Ban, "Object-based land cover change detection for cross-sensor images," *Int. J. Remote Sens.*, vol. 34, no. 19, pp. 6723–6737, 2013.

[23] L. Ma *et al.*, "Object-based change detection in urban areas: The effects of segmentation strategy, scale, and feature space on unsupervised methods," *Remote Sens.*, vol. 8, no. 9, 2016, Art. no. 761.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[25] Y. Sun, H. Song, A. J. Jara, and R. Bie, "Internet of things and big data analytics for smart and connected communities," *IEEE Access*, vol. 4, pp. 766–773, 2016.

[26] J. Yang, T. Liu, B. Jiang, H. Song, and W. Lu, "3D panoramic virtual reality video quality assessment based on 3D convolutional neural networks," *IEEE Access*, vol. 6, pp. 38669–38682, 2018.

[27] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[28] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 545–559, Mar. 2018.

[29] Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu, "Change detection based on deep Siamese convolutional network for optical aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1845–1849, Oct. 2017.

[30] P. F. Alcantarilla, S. Stent, G. Ros, R. Arroyo, and R. Gherardi, "Street-view change detection with deconvolutional networks," *Auton. Robots*, vol. 42, no. 7, pp. 1301–1322, 2018.

[31] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional Siamese networks for change detection," in *Proc. 25th IEEE Int. Conf. Image Process.*, 2018, pp. 4063–4067.

[32] M. Papadomanolaki, S. Verma, M. Vakalopoulou, S. Gupta, and K. Karantzalos, "Detecting urban changes with recurrent neural networks from multitemporal Sentinel-2 data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 214–217.

[33] L. Xu, W. Jing, H. Song, and G. Chen, "High-resolution remote sensing image change detection combined with pixel-level and object-level," *IEEE Access*, vol. 7, pp. 78909–78918, 2019.

[34] W. Byeon, T. M. Breuel, F. Raue, and M. Liwicki, "Scene labeling with LSTM recurrent neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3547–3555.

[35] B. Shuai, Z. Zuo, B. Wang, and G. Wang, "Scene segmentation with DAG-recurrent neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1480–1493, Jun. 2018.

[36] G. Lin, C. Shen, A. Van Den Hengel, and I. Reid, "Efficient piecewise training of deep structured models for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3194–3203.

[37] Z. Lin *et al.*, "A structured self-attentive sentence embedding," in *Proc. Int. Conf. Learn. Represent.*, 2017, pp. 1–14.

[38] T. Shen, T. Zhou, G. Long, J. Jiang, S. Pan, and C. Zhang, "DiSAN: Directional self-attention network for RNN/CNN-free language understanding," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 5446–5455.

[39] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[40] J. Tang, R. Hong, S. Yan, T.-S. Chua, G.-J. Qi, and R. Jain, "Image annotation by kNN-sparse graph-based label propagation over noisily tagged web images," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 2, pp. 1–15, 2011.

[41] J. Tang, L. Jin, Z. Li, and S. Gao, "RGB-D object recognition via incorporating latent data structure and prior knowledge," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 1899–1908, Nov. 2015.

[42] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7794–7803.

[43] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn. PMLR*, 2019, pp. 7354–7363.

[44] L. Chen *et al.*, "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6298–6306.

[45] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.

[46] J. Fu *et al.*, "Dual attention network for scene segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3146–3154.

[47] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2006, vol. 2, pp. 1735–1742.

[48] M. Lebedev, Y. V. Vizilter, O. Vygolov, V. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 42, no. 2, pp. 565–571, 2018.

[49] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 574–586, Jan. 2019.

[50] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sens. Environ.*, vol. 64, no. 1, pp. 1–19, 1998.

[51] C. Wu, B. Du, and L. Zhang, "Slow feature analysis for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2858–2874, May 2014.

[52] E. Guo *et al.*, "Learning to measure change: Fully convolutional Siamese metric networks for scene change detection," 2018, *arXiv:1810.09111*.

[53] A. Varghese, J. Gubbi, A. Ramaswamy, and P. Balamuralidhar, "ChangeNet: A deep learning architecture for visual change detection," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 129–145.

[54] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters—Improve semantic segmentation by global convolutional network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4353–4361.

[55] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2881–2890.

[56] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in *Proc. Artif. Intell. Statist.*, 2015, pp. 562–570.

[57] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1395–1403.

[58] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.

[59] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2016, pp. 770–778.

[60] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[61] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "GRAD-CAM: Why did you say that? Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 618–626.

[62] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[63] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.

Jie Chen received the M.S. degree in cartography and geographic information engineering from East China University of Technology, Nanchang, China, in 2006, and the Ph.D. degree in photogrammetry and remote sensing from Central South University, Changsha, China, in 2010.

He is currently an Associate Professor with the Department of Geo-Informatics, Central South University. His research interests include intelligent analysis and understanding of remote sensing images. He is a Reviewer for several journals in the field of remote sensing.

Ziyang Yuan received the B.Sc. degree in geographic information science, in 2018, from the School of Geosciences and Info-Physics, Central South University, Changsha, China, where he is currently working toward the M.Sc. degree in surveying and mapping.

His research interests include remote sensing and deep learning.

Jian Peng received the B.S. degree in cartography and geographic information engineering, in 2015. He is currently working toward the Ph.D. degree in remote sensing and artificial intelligence with the School of Geosciences and Info-Physics, Central South University, Changsha, China.

He research interests include memory models, deep learning, and artificial intelligence.

Li Chen received the B.Sc. and M.Sc. degrees in computer science from the School of Software, Central South University, Changsha, China, in 2015 and 2018, respectively. He is currently working toward the Ph.D. degree in remote sensing and artificial intelligence with the School of Geosciences and Info-Physics, Central South University.

His research interests include transfer learning, deep learning, and visualization.

Yu Liu received the B.S. degree in computer science from Northwestern Polytechnical University, Xi'an, China, in 2005, and the M.Sc. degree in image processing and Ph.D. degree in computer graphics from the University of East Anglia, Norwich, U.K., in 2007 and 2011, respectively.

He is currently an Associate Professor with the Department of Systems Engineering, National University of Defense Technology, Changsha, China. His research interests include image/video processing, computer graphics, and visual haptic technology.
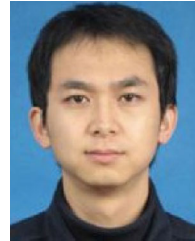
Haozhe Huang is currently working toward the Ph.D. degree in remote sensing and artificial intelligence with Central South University, Changsha, China.

His current research focuses on remote sensing image understanding and machine learning.

Haifeng Li (Member, IEEE) received the master's degree in transportation engineering from the South China University of Technology, Guangzhou, China, in 2005, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009.

He is currently a Professor with the School of Geosciences and Info-Physics, Central South University, Changsha, China. He was a Research Associate with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, in 2011, and a Visiting Scholar with the University of Illinois at Urbana-Champaign, Urbana, IL, USA, from 2013 to 2014. He is a Reviewer for many journals. He has authored more than 30 journal papers. His current research interests include geo/remote sensing big data, machine/deep learning, and artificial/brain-inspired intelligence.

Jiawei Zhu received the B.S. degree in the geographic information system in 2013 and the M.S. degree in cartography and geographic information science in 2016 from Central South University, Changsha, China, where she is currently working toward the Ph.D. degree in surveying and mapping.

Her research interests include network representation learning, topological data analysis, and spatiotemporal data mining.