Spatial–Spectral Decoupling Interaction Network for Multispectral Imagery Change Detection

Jie Fang, Guanghua He, Zhijie Zhu, Bahari Issa M. Attaher, and Jian Xue, *Member, IEEE*

***Abstract*— We present a spatial–spectral decoupling interaction network for multispectral imagery change detection, which can exploit the underlying information of the multispectral imagery adequately through simultaneously considering the discriminative attribute of each pixel and robust spatial structure of the corresponding patch. Specifically, a 1-D convolutional neural network (1D-CNN) is applied to the spectral vector of each pixel to extract its discriminative feature, while a 2D-CNN is applied to the patch centering on the corresponding pixel to explore the spatial structure information. In addition, an interaction mechanism is incorporated into the feature fusion module to enhance the spatial–spectral consistency.**

***Index Terms*— Multispectral imagery change detection, spatial–spectral decoupling interaction network.**

I. Introduction

DUE to its widespread applications such as resource management [1] and damage assessment [2], multispectral imagery change detection has increasingly attracted the attention of the researchers recently. During the last decade, many approaches have been proposed for multispectral imagery change detection. For instance, change vector analysis (CVA) [3] measures the differences of two images through calculating the Euclidean distance of each corresponding pixel pair. Bruzzone and Bovolo [4] proposed a top-down approach to the design of the architecture of novel change-detection systems. Bovolo *et al.* [5] proposed an automatic and unsupervised multiple change detection method. Liu *et al.* [6] proposed a multiscale morphological compressed CVA strategy for unsupervised multiple change detection.

The aforementioned methods contribute to the development of multispectral imagery change detection algorithms to a large extent, but most of them are based on manual features, which cannot exploit the intrinsic characteristics of the multispectral images effectively. Recently, deep neural networks (DNNs), especially deep convolutional neural networks (DCNNs), have achieved significant performances on many computer vision tasks such as scene recognition semantic segmentation [7]. As for multispectral imagery change detection, Liu *et al.* [8] proposed a symmetric convolutional coupling network (SCCN), which is based on the deep belief network (DBN). Gong *et al.* [9] proposed to incorporate the superpixel-based feature extraction and difference feature learning modules into the network to consider the spatial and spectral information simultaneously and further improve the performance of change detection. Zhan *et al.* [10] proposed to use an Siamese convolutional network to exploit the differences between each corresponding paired pixels in two multispectral images. Fang *et al.* [11] proposed a densely connected Siamese network for multispectral image change detection, which can exploit more sufficient latent information of images through dense connections. In addition, Liu *et al.* [12] reviewed existing DNN-based change detection approaches for multitemporal hyperspectral images and pointed out some development direction of related fields.

Even though the existing CNN-based methods have achieved relatively satisfactory performances, they ignore the differences between attribute information, which reflects the independent feature of the very pixel, and spatial structure information, which reflects the relationships among the very pixel and ones distributed in its neighbors. Specifically, they tackle these two kinds of information in a unified way but do not consider the differences between them, which may result in the spatial diffusion problem, and the predicted saliency map often loses some important details further. The main reasons contain two aspects: 1) patches with smaller size cannot obtain sufficient spatial information and 2) ones with bigger size cannot reflect the intrinsic attribute information of their center pixels well.

In these cases, we propose a spatial–spectral decoupling interaction network for multispectral imagery change detection, which can obtain intrinsic attribute information and sufficient spatial structure information of each pixel simultaneously. Specifically, we first divide the multispectral image into several patches with size d in an overlapping way with stride one. Then, we apply a weight-shared Siamese 1D-CNN to the center pixels of each paired patch to extract the discriminative spectral attribute differential signal. In addition, we utilize a different method like local binary pattern (LBP) to obtain the whitened structure of each paired patch and then apply a weight-shared Siamese 2D-CNN to corresponding whitened structure patches to obtain the spatial structural differential signal. Besides, we incorporate a dynamic weight guidance-based interaction mechanism into the aforementioned 1D-CNN and 2D-CNN to enhance the consistency of the representations

Manuscript received October 18, 2021; revised December 23, 2021; accepted February 23, 2022. Date of publication February 25, 2022; date of current version March 23, 2022. *(Corresponding author: Jie Fang.)*

Jie Fang is with the School of Communications and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China, and also with Corporation of Shaanxi Wukong Cloud Information and Technology, Xi'an, Shaanxi 710000, China (e-mail: 2443952262@qq.com).

Guanghua He, Zhijie Zhu, and Jian Xue are with the School of Communications and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China.

Bahari Issa M. Attaher is with the Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710048, China, and also with Corporation of Shaanxi Wukong Cloud Information and Technology, Xi'an, Shaanxi 710000, China.

Digital Object Identifier 10.1109/LGRS.2022.3154745

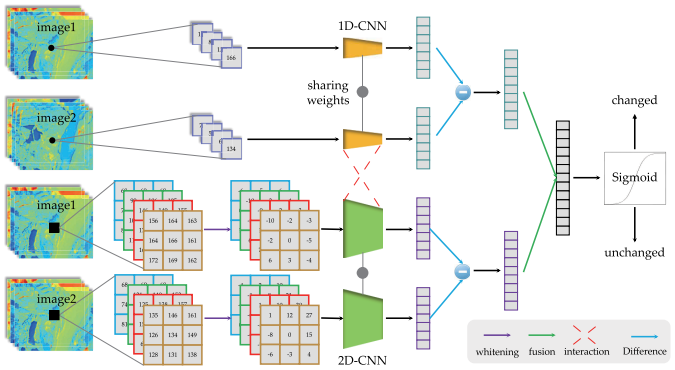

Fig. 1. Proposed multispectral imagery change detection framework.

from the two branches. Finally, we fuse the two differential signals as a spatial–spectral one to finalize the decision of changed or not. In summary, the contributions of this letter can be concluded as follows.

1) We propose a spatial–spectral decoupling interaction network for multispectral imagery change detection, which can exploit the intrinsic spectral and detailed spatial information of the image effectively.
2) We propose a dynamic weight guidance-based interaction mechanism to enhance the consistency of spatial and attribute representations, which improves the robustness of the final spatial–spectral feature.

II. Proposed Method

A. Overview

Following the Siamese network-based spatial–spectral joint learning method, we propose a spatial–spectral decoupling interaction network for multispectral imagery change detection. The flowchart of the proposed method is shown in Fig. 1, which mainly contains two components, including data preparation and spatial–spectral decoupling interaction network.

B. Data Preparation

As is shown in Fig. 1, the inputs of the proposed spatial–spectral decoupling interaction network contain two components: spectral vectors and spatial patches. The former is simply putting out the independent spectrum vector of each pixel as the input of spectral attribute representation network, and here we emphatically introduce the latter. Specifically, we denote a multispectral image as $\mathbf{X} \in \mathbf{R}^{W\times H\times C}$, where W, H, and C denote its width, height, and band number, respectively. The (i, j)th pixel in image $\mathbf{X}$ can be defined as an intensity vector $\mathbf{x}_{i,j} \in \mathbf{R}^C = [x_{i,j,1}, x_{i,j,2}, \ldots, x_{i,j,C}]$, where $x_{i,j,k}$ represents the element of location with coordinate (x, y) in the kth channel of image $\mathbf{X}$. Then, we define a neighboring patch $\mathbf{p}_{i,j} \in \mathbf{R}^{d\times d\times C}$ centered on $\mathbf{x}_{i,j}$ to depict its spatial structural information. Additionally, we apply a whitening procedure like LBP to each patch. Specifically, we obtain the intensity-independent spatial structural patch $\mathbf{p}^I_{i,j}$ through $\mathbf{p}^I_{i,j} = \mathbf{p}_{i,j} - \mathbf{p}^c_{i,j} \odot \mathbf{I}^{d\times d\times C}$, where $\mathbf{p}^c_{i,j}$ denotes the center pixel of $\mathbf{p}_{i,j}$, $\mathbf{I}^{d\times d\times C}$ denotes a matrix with all ones, and $\odot$ denotes the expanded multiplication. Compared to $\mathbf{p}_{i,j}$, on the one hand, $\mathbf{p}^I_{i,j}$ decouples the spatial structural representation from the attribute representation, which avoids the information interferences between the spatial and spectral domains. On the other hand, $\mathbf{p}^I_{i,j}$ is more robust for varied illuminations. It is noteworthy that, neighboring regions of the boundary and near-boundary pixels cannot constructed with the aforementioned strategy. In this case, we propose a new neighboring region construction strategy, which is described as follows.

1) Expand the original multispectral image $\mathbf{X} \in \mathbf{R}^{W\times H\times C}$ to $\mathbf{X}_e \in \mathbf{R}^{(W+K-1)\times(H+K-1)\times C}$ in a conventional zero padding way, where K is the size of neighboring region.
2) For the padding set $S^e = \{\mathbf{x}^e_{i,j} | i, j \le t \text{ or } i \ge W + t + 1 \text{ or } j \ge H + t + 1\}$, we utilize the symmetric fulling strategy to replace its 0 values. Here, $t = (K - 1/2)$, and $\mathbf{x}^e_{i,j}$ represents the spectral vector in (i, j)th location of image $\mathbf{X}_e$.

C. Spatial–Spectral Decoupling Interaction Network

The proposed spatial–spectral network mainly contains four subparts: spectral representation, spatial representation network, interaction mechanism, and feature fusion, which are detailed as follows.

1) Spectral Representation: The multispectral imagery change detection is actually a pixel-wise classification problem. Consequently, it is important to construct a discriminative differential signal for each corresponding paired pixels. In this letter, considering the correlations of adjacent bands and the natural strong relation representation capability of CNNs, we utilize a Siamese 1D-CNN to exploit this differential signal. First, two 1D-CNNs are used to extract the intrinsic attribute features of two pixels. It is noteworthy that these two 1D-CNNs are weight-shared. On the one hand, compared to two independent networks, fewer parameters of weight-shared Siamese network can alleviate the overfitting problem. On the other hand, only the shared weight can reflect the differential information of two pixels sufficiently. For the sake of simplicity, we use a common identifier SpecNet to depict the aforementioned 1D-CNN. We donate the feature vectors of paired pixels as spec_1 and spec_2, which are defined in the following equations:

$$\text{spec}_1^{(n)} = \text{SpecNet}\left(\mathbf{v}_1^{(n)}\right) \tag{1}$$

$$\text{spec}_2^{(n)} = \text{SpecNet}\left(\mathbf{v}_2^{(n)}\right). \tag{2}$$

In (1) and (2), $\mathbf{v}_1^{(n)}$ and $\mathbf{v}_2^n$ denote the spectral vectors of nth pixels in two images, respectively. After obtaining $\text{spec}_1^{(n)}$ and $\text{spec}_2^{(n)}$, we calculate their differential values with the following equation:

$$\text{spec}_d^{(n)} = \text{spec}_1^{(n)} - \text{spec}_2^{(n)} \tag{3}$$

and consider $\text{spec}_d^{(n)}$ as the final discriminative differential signal of the nth paired pixels in two images.

TABLE I
SPECTRAL REPRESENTATION NETWORK

Layers	Types	Input Dim.	Output Dim.	KS	S
Conv1	C	$B \times 1$	$(B-2) \times 64$	3	1
Conv2	C	$(B-2) \times 64$	$(B-4) \times 9$	3	1
Pool	P	$(B-4) \times 9$	$(B/4-1) \times 9$	4	4
Fc1	F	$(B/4-1) \times 9$	128	/	/
Fc2	F	128	84	/	/

TABLE II
SPATIAL REPRESENTATION NETWORK

Layers	Types	Input Dim.	Output Dim.	KS	S
Conv1	C	$9 \times 9 \times 3$	$7 \times 7 \times 64$	3	1
Conv2	C	$7 \times 7 \times 64$	$5 \times 5 \times 32$	3	1
Pool	P	$7 \times 7 \times 32$	$3 \times 3 \times 32$	2	2
Fc1	F	$3 \times 3 \times 32$	128	/	/
Fc2	F	128	84	/	/

2) Spatial Representation: Besides the intrinsic spectral vector of each paired pixels themselves, the spatial structural information in their surroundings is also important for the difference estimation. This point can be explained in two aspects.

1) Single spectral vector may be interfered by the noise sometimes and cannot depict its intrinsic information effectively, and pixels in their surrounding regions can assist the feature representation in this condition.
2) Besides the spectral vectors of each paired pixels themselves, spatial structural relations among them and their surroundings can help the decision of whether changed or not.

Specifically, because of the spatial continuity of multispectral imagery, even the spectral vectors of paired pixels are similar, we still have sufficient confidence to believe the information of this location is changed if their spatial structural relations have a relatively large gap. In these cases, similar to the spectral representation process, we utilize a weight-shared Siamese 2D-CNN (SpatNet) to exploit the spatial structural differential signals of each corresponding paired pixel. First, we use SpatNet to extract the spatial structural features of paired pixels, which are defined in the following equations:

$$\text{spat}_1^{(n)} = \text{SpatNet}\left(\mathbf{p}_1^{I(n)}\right) \tag{4}$$

$$\text{spat}_2^{(n)} = \text{SpatNet}\left(\mathbf{p}_2^{I(n)}\right). \tag{5}$$

In (4) and (5), $\mathbf{p}_1^{I(n)}$ and $\mathbf{p}_1^{I(n)}$ denote the intensity-independent spatial structural patch of the nth pixel in two images, respectively. Then, we can obtain the differential value with the following equation:

$$\text{spat}_d^{(n)} = \text{spat}_1^{(n)} - \text{spat}_2^{(n)} \tag{6}$$

and consider $\text{spat}_d^{(n)}$ as the final spatial structural differential signal of n_{th} pixels in two images.

3) Network and Interaction Mechanism: The inner architectures of spectral and spatial representation networks are shown in Table I, where B denotes the number of spectrum bands. KS and S denote the kernel size and stride, respectively.

The inner architectures of the spatial representation network are shown in Table II, where KS and S have similar meanings with corresponding ones in Table I.

In addition, we incorporate a dynamic weight guidance-based interaction mechanism into two representation networks to enhance their consistency, which uses the hidden latent features of the spatial domain to guide those of the spectral domain. Specifically, we first calculate the guidance weight with the following equation:

$$\boldsymbol{\omega}_g = \mathbf{I} + \sigma\left(P_l\left(\mathbf{F}^{\text{Spa}}\right)\right) \tag{7}$$

where $\mathbf{I}$ denotes an unit matrix. $\sigma(\cdot)$ denotes the sigmoid function. $\mathbf{F}^{\text{Spa}}$ denotes the output from the pool layer of the spatial representation network. P_l represents the location max pooling operation, which is defined in the equation

$$P_l\left(\mathbf{F}^{M\times N\times D}\right) = \left\{\mathbf{T}^{M\times N} | \mathbf{T}_{i,j} = \max_{1\le d\le D} \mathbf{F}_{i,j,d}\right\}. \tag{8}$$

We then flatten $\boldsymbol{\omega}_g$ to w_g using the following equation:

$$w_g = \text{vec}\left(\boldsymbol{\omega}_g\right) \tag{9}$$

where $\text{vec}(\cdot)$ denotes the vectorized operation. After that, we obtain the weighted spectral hidden representation in the following equation:

$$\mathbf{F}_{\text{weighted}}^{\text{Spe}} = w_g \odot \mathbf{F}_{\text{original}}^{\text{Spe}} \tag{10}$$

where $\mathbf{F}_{\text{original}}^{\text{Spe}}$ and $\mathbf{F}_{\text{weighted}}^{\text{Spe}}$ denote the original and weighted spectral features, respectively. $\odot$ denotes the element-wise expanded multiplication, which is defined in the following equation:

$$\mathbf{R}^{1\times D} \odot \mathbf{Q}^{N\times D} = \left\{\mathbf{O}^{N\times D} | \mathbf{O}_i = \mathbf{R}\cdot\mathbf{Q}_i\right\} \tag{11}$$

where $\mathbf{O}_i$ and $\mathbf{Q}_i$ denote the ith row of $\mathbf{O}$ and $\mathbf{Q}$, respectively.

Through the aforementioned operation series, we can ensure the consistency of representation in spectral and spatial domains to a certain extent. The reason is that the learning procedure of each feature channel in the spectral domain is guided by a specific location-based feature in the spatial domain, which enhances the interactions of two domains.

4) Feature Fusion: After obtaining the spectral discriminative differential signal $\text{spec}_d^{(n)}$ and spatial structural differential signal $\text{spat}_d^{(n)}$, we fuse them together to finalize the decision. First, we concatenate $\text{spec}_d^{(n)}$ and $\text{spat}_d^{(n)}$ as $\text{spatspec}_d^{(n)}$ with the following equation:

$$\text{spatspec}_d^{(n)} = \text{Concatenate}\left(\text{spec}_d^{(n)}, \text{spat}_d^{(n)}\right). \tag{12}$$

Second, we apply FusionNet to $\text{spatspec}_d^{(n)}$ to improve its representation and discrimination capability, which is defined in the following equation:

$$\text{feature}^{(n)} = \text{FusionNet}\left(\text{spatspec}_d^{(n)}\right) \tag{13}$$

where FusionNet is actually a three-layer fully connected network, and its architecture is shown in Table III, where KS and S denote the kernel size and stride, respectively.

TABLE III
FEATURE FUSION NETWORK

Layers	Types	Input Dim.	Output Dim.	KS	S
Fc1	F	168	84	/	/
Fc2	F	84	84	/	/
Fc3	F	84	2	/	/

Third, we use the conventional cross entropy loss to optimize the network, which is defined in the following equation:

$$\mathcal{L} = -\frac{1}{N}\sum_{n=1}^{N}\left(y^{(n)}\log\hat{y}^{(n)} + \left(1-y^{(n)}\right)\log\left(1-\hat{y}^{(n)}\right)\right) \quad (14)$$

where $\mathcal{L}$ represents the loss of our spatial–spectral decoupling interaction network, N represents the number of training samples, $y^{(n)}$ represents the label of the nth sample, and $\hat{y}^{(n)}$ represents the corresponding predicted saliency score.

III. EXPERIMENTS

A. Datasets

We use GF-1 [13] and ETM+ datasets [14] to measure the performances of different methods in this letter.

The ETM+ dataset [13] contains Taizhou and Kunshan datasets, which are collected by the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) sensor. The Taizhou dataset contains 400×400 pixels, while the Kunshan dataset contains 800×800 pixels. Both the Taizhou and Kunshan datasets are with a total of six spectrum bands and a spatial resolution of 30 m per pixel. Additionally, two images of the Taizhou datasets were collected on March 17, 2000, and February 6, 2003, respectively, and those of the Kunshan dataset were obtained at the same time.

The GF-1 dataset [14] contains the Minfeng and Hongqi Canal datasets, which are collected by the GF-1 satellite. The Minfeng dataset contains 651×461 pixels, while the Hongqi Canal dataset contains 539×543 pixels. Both the Minfeng and Hongqi Canal datasets are with a total of four spectrum bands and spatial resolution of 2 m per pixel. Additionally, two images of the Minfeng dataset were collected on December 9, 2013, and October 16, 2015, respectively, and those of the Hongqi Canal dataset were acquired at the same time.

B. Contrasting Methods and Evaluation Metrics

To validate the effectiveness of the proposed method, we compare it with five state-of-the-arts ones, including SCCN [8], S-CNN+SVM [15], DSIFUN [16], DASNet [17], and SNUNet-CD [11].

In addition, we use the overall accuracy (OA) and Kappa coefficient to measure the performances of different methods. OA is defined in the following equation:

$$\mathrm{OA} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \quad (15)$$

where TP, TN, FP, and FN represent the true positive, true negative, false positive, and false negative, respectively. Kappa is defined in the following equation:

$$\mathrm{Kappa} = \frac{\mathrm{OA}-\mathrm{PRE}}{1-\mathrm{PRR}} \quad (16)$$

where PRE is defined in the following equation:

$$\mathrm{PRE} = \frac{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})+(\mathrm{FN}+\mathrm{TN})(\mathrm{FP}+\mathrm{TN})}{(\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN})^2}. \quad (17)$$

C. Experimental Results

This section reports the experimental results. In order to validate the effectiveness of each proposed module and the superiority of the proposed method, we design the ablation experiments and contrasting methods, which are detailed as follows.

1) Ablation Experiments: The ablation experimental results are shown in Table IV, where spectral represents the method only with the spectral-domain representation. Spatial represents the method only with the spatial-domain representation. SpaWh represents the method only with whitened spatial domain representation. SpeSpaCat represents the method with the concatenate spectral and spatial representations. SpeSpaWhiCat represents the method with the concatenate spectral and whitened spatial representations. SpecSpatWhiCatInt represents the method with spectral representation, whitened spatial representation, and the proposed dynamic weight guidance interaction mechanism. SpeSpaWhiFuInt represents the method with spectral representation, whitened spatial representation, the proposed dynamic weight guidance interaction mechanism, and the proposed feature fusion module. From Table IV, we can see that each of the proposed components contributes to the improvement of the detection performance. For instance, when the spatial domain representation is concatenated to the spectral domain one as the final representation, it achieves 0.0139 OA improvement and 0.0227 Kappa improvement. The reason is that spatial structural information can improve the representation stability of each pixel due to the local similarity of the image in the spatial domain.

2) Contrasting Experiments: This section reports the contrasting experiments on four subdatasets, and all the experimental results are presented in Table V. From which we can see that the proposed method achieves relatively satisfactory performances. For instance, compared with SCCN [8], the proposed method achieves 0.0461 OA increment on the Taizhou set. Besides that, compared with S-CNN+SVM [15], the proposed method surpasses it by 0.0186 in terms of the Kappa indicator on the Kunshan set. In addition, the proposed method obtains 0.0181 OA increment and 0.0203 Kappa increment, respectively, compared with DSIFUN on the Minfeng set, which reflects its superiority to a certain extent. More importantly, even compared with the latest approaches DASNet and SNUNet, our method still shows significant advantages. The reasons can be summarized as follows.

1) The proposed method simultaneously considers the independent discriminative attribute information of each pixel and dependent robust spatial structure information of the corresponding patch sufficiently, which can help the model to construct discriminative and stable feature representation of each corresponding paired pixel in two temporal images.

TABLE IV
ABLATION EXPERIMENTAL RESULTS ON THE HONGQI DATASET

Spectral	Spatial	SpaWhi	SpeSpaCat	SpeSpaWhiCat	SpeSpaWhiCatInt	SpeSpaWhiFuInt	OA	Kappa
√	–	–	–	–	–	–	0.9252	0.8969
–	√	–	–	–	–	–	0.9335	0.9099
–	–	√	–	–	–	–	0.8486	0.7887
–	–	–	√	–	–	–	0.9391	0.9196
–	–	–	–	√	–	–	0.9457	0.9271
–	–	–	–	–	√	–	0.9494	0.9326
–	–	–	–	–	–	√	**0.9536**	**0.9380**

TABLE V
CONTRASTING EXPERIMENTAL RESULTS ON FOUR SUB-DATASETS

Dataset	Metric	SCCN [8]	S-CNN+SVM [15]	DSIFUN [16]	DASNet [17]	SNUNet-CD[11]	Ours
Taizhou	OA	0.9217	0.9422	0.9485	0.9549	0.9614	**0.9678**
	Kappa	0.9058	0.9284	0.9301	0.9352	0.9387	**0.9433**
Kunshan	OA	0.9244	0.9387	0.9402	0.9451	0.9470	**0.9524**
	Kappa	0.9105	0.9179	0.9188	0.9237	0.9288	**0.9365**
Minfeng	OA	0.9189	0.9405	0.9457	0.9512	0.9575	**0.9638**
	Kappa	0.9027	0.9264	0.9298	0.9364	0.9427	**0.9501**
Hongqi Canal	OA	0.9147	0.9369	0.9401	0.9452	0.9488	**0.9536**
	Kappa	0.8942	0.9205	0.9246	0.9279	0.9352	**0.9380**

2) The proposed dynamic weight guidance-based interaction mechanism can enhance the consistency of attribute and spatial representations and further improve the effectiveness of the feature.

IV. CONCLUSION

In this letter, we propose a spatial–spectral decoupling interaction network for multispectral imagery change detection, which takes full consideration of spatial and spectral information of multispectral imagery. Specifically, we use a Siamese 1D-CNN and a Siamese 2D-CNN to extract the spectral discriminative and spatial structural information, respectively, in a decoupling way. Besides, we incorporate an interaction mechanism into the feature extraction process to enhance the spatial–spectral consistency. Then, we calculate the spatial and spectral differential signals of each corresponding paired pixel in two images, respectively. Additionally, we fuse the differential signals as a spatial–spectral feature to finalize the change detection. Finally, the experimental results on four public datasets validate the superiority of our method.

REFERENCES

[1] K. Rokni, A. Ahmad, A. Selamat, and S. Hazini, "Water feature extraction and change detection using multitemporal Landsat imagery," *Remote Sens.*, vol. 6, no. 5, pp. 4173–4189, 2014.
[2] L. Gueguen and R. Hamid, "Toward a generalizable image representation for large-scale change detection: Application to generic damage analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3378–3387, Jun. 2016.
[3] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 218–236, Jan. 2006.
[4] 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.
[5] F. Bovolo, S. Marchesi, and L. Bruzzone, "A framework for automatic and unsupervised detection of multiple changes in multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2196–2212, May 2012.
[6] S. Liu, Q. Du, X. Tong, A. Samat, L. Bruzzone, and F. Bovolo, "Multiscale morphological compressed change vector analysis for unsupervised multiple change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4124–4137, Sep. 2017.
[7] J. Fang and X. Cao, "GAN and DCN based multi-step supervised learning for image semantic segmentation," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*. Cham, Switzerland: Springer, 2018, pp. 28–40.
[8] 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, Dec. 2016.
[9] M. Gong, T. Zhan, P. Zhang, and Q. Miao, "Superpixel-based difference representation learning for change detection in multispectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2658–2673, May 2017.
[10] 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.
[11] S. Fang, K. Li, J. Shao, and Z. Li, "SNUNet-CD: A densely connected Siamese network for change detection of VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
[12] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, Jun. 2019.
[13] S. Zhou, "GaoFen-1 remote sensing image dataset," Tech. Rep., Feb. 2019, doi: 10.11888/Geogra.tpdc.271227.
[14] X. Lu, Y. Yuan, and X. Zheng, "Joint dictionary learning for multispectral change detection," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 884–897, Apr. 2017.
[15] B. Liu, X. Yu, P. Zhang, A. Yu, Q. Fu, and X. Wei, "Supervised deep feature extraction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 1909–1921, Apr. 2018.
[16] C. Zhang *et al.*, "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 183–200, Aug. 2020.
[17] J. Chen *et al.*, "DASNet: Dual attentive fully convolutional Siamese networks for change detection in high-resolution satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1194–1206, 2020.