Ultralightweight Spatial–Spectral Feature Cooperation Network for Change Detection in Remote Sensing Images

Tao Lei, *Senior Member, IEEE*, Xinzhe Geng, Hailong Ning, Zhiyong Lv, *Senior Member, IEEE*, Maoguo Gong, *Senior Member, IEEE*, Yaochu Jin, *Fellow, IEEE*, and Asoke K. Nandi, *Life Fellow, IEEE*

***Abstract*— Deep convolutional neural networks (CNNs) have achieved much success in remote sensing image change detection (CD) but still suffer from two main problems. First, the existing multiscale feature fusion methods often use redundant feature extraction and fusion strategies, which often lead to high computational costs and memory usage. Second, the regular attention mechanism in CD is difficult to model spatial–spectral features and generate 3-D attention weights at the same time, ignoring the cooperation between spatial features and spectral features. To address the above issues, an efficient ultralightweight spatial–spectral feature cooperation network (USSFC-Net) is proposed for CD in this article. The proposed USSFC-Net has two main advantages. First, a multiscale decoupled convolution (MSDConv) is designed, which is clearly different from the popular atrous spatial pyramid pooling (ASPP) module and its variants since it can flexibly capture the multiscale features of changed objects using cyclic multiscale convolution. Meanwhile, the design of MSDConv can greatly reduce the number of parameters and computational redundancy. Second, an efficient spatial–spectral feature cooperation (SSFC) strategy is introduced to obtain richer features. The SSFC differs from the existing 2-D attention mechanisms since it learns 3-D spatial–spectral attention weights without adding any parameters. The experiments on three datasets for remote sensing image CD demonstrate that the proposed USSFC-Net achieves better CD accuracy than most CNNs-based methods and requires lower computational costs and fewer parameters, even it is superior to some Transformer-based methods. The code is available at https://github.com/SUST-reynole/USSFC-Net.**

***Index Terms*— Change detection (CD), convolutional neural network (CNN), multiscale feature extraction, spatial–spectral feature cooperation (SSFC).**

Manuscript received 7 December 2022; revised 24 February 2023; accepted 15 March 2023. Date of publication 24 March 2023; date of current version 7 April 2023. This work was supported in part by the Natural Science Basic Research Program of Shaanxi under Grant 2021JC-47 and Grant 2022JQ-592; in part by the National Natural Science Foundation of China under Grant 62271296, Grant 62201452, and Grant 62201334; in part by the Key Research and Development Program of Shaanxi Province under Grant 2022GY-436, Grant 2021ZDLGY08-07, and Grant 2021GY-181; in part by the Shaanxi Joint Laboratory of Artificial Intelligence under Grant 2020SS-03; and in part by the Scientific Research Program Funded by the Shaanxi Provincial Education Department under Grant 22JK0568. *(Corresponding author: Hailong Ning.)*

Tao Lei and Xinzhe Geng are with the Shaanxi Joint Laboratory of Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, China (e-mail: leitao@sust.edu.cn; 201606020514@sust.edu.cn).

Hailong Ning is with the School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China (e-mail: ninghailong93@gmail.com).

Zhiyong Lv is with the School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China (e-mail: Lvzhiyong_fly@hotmail.com).

Maoguo Gong is with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Xidian University, Xi'an 710071, China (e-mail: gong@ieee.org).

Yaochu Jin is with the Faculty of Technology, Bielefeld University, 33619 Bielefeld, Germany (e-mail: yaochu.jin@uni-bielefeld.de).

Asoke K. Nandi is with the Department of Electronic and Electrical Engineering, Brunel University London, UB8 3PH Uxbridge, U.K., and also with the School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: asoke.nandi@brunel.ac.uk).

Digital Object Identifier 10.1109/TGRS.2023.3261273

I. Introduction

THE goal of remote sensing image change detection (CD) is to identify differences between two images of the same geographical location taken at different periods [1]. It is of great significance in many fields, including disaster monitoring [2], urban planning [3], environmental investigation [4], to name a few. In recent years, these applications have become more crucial due to the deterioration of the natural environment. Therefore, a large number of CD methods have emerged, which can be roughly categorized into two groups: traditional methods and deep-learning-based methods.

Most of the traditional methods rely on manual feature extraction, such as principal component analysis (PCA) [5], [6], Gabor filter [7], multivariate alteration detection (MAD) [8], and change vector analysis (CVA) [9]. These methods can achieve CD to a certain extent, but they suffer from the following weaknesses. On one hand, the image features extracted by traditional methods are susceptible to seasonal changes, lighting conditions, and satellite sensors, making them less robust for achieving high CD accuracy. On the other hand, although some methods [7], [10] can reduce false detection by combining shape and texture features, such strategies usually require intensive computation and have many hyperparameters, leading to low robustness and high computational cost. In addition, manually extracted features rely heavily on prior domain knowledge, which limits the generalization ability of models.

In recent years, increased interest has led to more applications of convolutional neural networks (CNNs) [11] to remote sensing image CD. Compared with traditional methods, deep-learning-based methods require less human intervention and can automatically learn features from annotated data. In addition, deep-learning-based methods can better

understand complex scenes due to their excellent feature extraction abilities, and they perform much better than traditional methods. Since CD can be regarded as an image segmentation task, most of the current CD backbones based on deep learning use encoder–decoder structures such as fully convolutional network (FCN) [12] and U-shape networks (U-Net) [13]. However, unlike the general image segmentation task, the input of CD is a pair of bitemporal images. To fuse effectively bitemporal images information, the Siamese structure is applied as the benchmark for CD in remote sensing images [14].

Due to the scale variation and complex background in remote sensing images, various multiscale feature fusion modules and attention mechanisms [65], [66] have been introduced into deep neural networks for CD in remote sensing images [16]. However, they still face the following challenges. First, introducing the existing multiscale feature fusion module directly may lead to a large amount of feature redundancy; one of the reasons is that using multiple atrous convolutions in parallel requires redundant learnable parameters. Second, although both spatial attention and channel attention can improve the CD accuracy to different extents, they ignore the comodeling of spatial features and spectral features and cannot reason 3-D attention weights directly. Since the spectral information of the bitemporal remote sensing image is contained in the feature maps of the multidimensional channels, the cascade of spatial and channel attention is often used to model the spatial–spectral dependence of changed objects, requiring a lot of additional memory and computational costs [37].

To address the above issues, an ultralightweight spatial–spectral feature cooperation network (USSFC-Net) is proposed. This network takes a pseudo-Siamese U-Net [13] as the backbone. It uses multiscale decoupled convolution (MSDConv) instead of the vanilla convolution for feature extraction. MSDConv has two advantages. First, it decouples the convolution into the concatenation of the spatial and channel correlations, where the channel correlation is calculated by point convolution and the spatial correlation is obtained by depthwise convolution. This decoupling significantly reduces the computational costs and parameter redundancy of the convolution. Second, to capture changed objects with different scales, MSDConv cyclically uses combination of dilation rates for depthwise convolution to perform spatial correlation expansion. This cyclic multiscale structure avoids increasing additional parameters. Thus, MSDConv stands out as a lightweight and efficient multiscale feature extraction module. Besides, a spatial–spectral feature cooperation (SSFC) strategy is designed to capture better change-related features. It is implemented by generating spatial and spectral cooperation 3-D weights using Gaussian modeling. This strategy does not require any additional learnable parameters and can be efficiently embedded into the MSDConv to obtain richer features in the feature extraction stage.

The main contributions of this article can be summarized as follows.

1) An MSDConv is designed for CD networks. It can effectively capture the multiscale features of changed objects in remote sensing images using a compact design with decoupled spatial and channel correlations. Different from popular multiscale feature extraction methods such as Inception [18], atrous spatial pyramid pooling (ASPP) [65], and feature pyramid network (FPN) [19], the MSDConv is more lightweight and efficient.
2) An SSFC is introduced into the MSDConv to obtain richer features. It is a low-cost yet high-performance attention mechanism for CD. Compared with popular 2-D attention mechanisms such as spatial attention, channel attention, and their variants, the SSFC achieves spatial–spectral cooperation 3-D attention without any additional parameters.
3) An efficient USSFC-Net is proposed based on the use of MSDConv and SSFC. Extensive experiments are conducted, and the results show that the proposed USSFC-Net achieves higher CD accuracy and requires fewer parameters than most popular CD networks.

The rest of this article is organized as follows. The related work is reviewed in Section II. A detailed description of the proposed method is provided in Section III. The experimental results are reported in Section IV. Discussions of key issues are given in Section V. Conclusions are included in Section VI.

II. Related Work

A. Backbone Network for Remote Sensing Image CD

Owing to the development of deep learning and computer vision, more and more deep learning methods are used for remote sensing image CD. These methods can be roughly categorized into two groups. The first group learns the changed features directly from the difference images that are generated by pretemporal and posttemporal images [14], [29]. However, they do not consider the specific characteristic of CD task and simply use generic CNNs to achieve CD, which leads to low detection accuracy.

The second group of methods performs feature extraction on bitemporal images separately, and then fully compares and fuses different spatial–temporal features at different stages of the network to obtain difference images. Daudt et al. [14] first applied the Siamese network [20] to remote sensing image CD. The Siamese network usually consists of two weight-sharing branches for feature extraction on pretemporal and posttemporal images, separately. When applying the Siamese network to bitemporal remote sensing images, the bitemporal features are extracted and used to generate changed images, which is more conducive to improving the CD accuracy. Besides, more improved CD frameworks based on the Siamese structure can be seen in [15], [21], [22], [23], [24], [25], [26], [27], and [28]. However, as many application scenarios are constrained by computing resources, more and more compact networks are designed to achieve CD in a low-cost manner. Wang et al. [61] used bottleneck and dilated convolutions to replace the vanilla convolutional layers, which effectively reduced the parameters and computational costs. To obtain rich contextual information, Han et al. [62] introduced an artificial padding convolution and designed a new loss function for CD in optical remote sensing imagery. By introducing the vision Transformer into CD, Dai et al. [63] used MobileViT to achieve high-precision CD at a faster inference speed.

The advantage of the Siamese network is that it has fewer parameters. However, a weight-sharing encoder may lead to weak feature extraction ability, thereby affecting the classification accuracy for changed objects. Using a nonweight-sharing encoder may address this problem, but this usually leads to an increase in the number of parameters. To address this problem, it is necessary to design a ultralightweight network for CD tasks.

B. Multiscale Feature Fusion for Remote Sensing Image CD

Among the current popular neural networks, CNN has two distinct advantages over multilayer perceptron and Transformer: parameter-sharing and sparse connectivity. Such computational characteristics lead to the fact that the size of the receptive field determines the performance of feature extraction [64]. Therefore, designing an effective receptive field scale for convolutional layers is crucial to CNN-based CD methods. To solve this problem, researchers have designed multiscale fusion modules to extend efficiently the receptive field. A more intuitive method to extend the receptive field is to simply increase the size of convolution kernels. Lei et al. [29] proposed the pyramid pooling module to extract deep features and fuse them using three different sizes of convolution kernels for difference images, which can effectively capture the multiscale features of remote sensing images. Shen et al. [30] used a similar multiscale feature extraction strategy and applied point convolution to dimensionality reduction before multiscale feature fusion. Hou et al. [31] proposed the dynamic inception module, which introduces dynamic convolution into a multiscale feature fusion module to improve the feature representation ability of networks.

Although the above methods can solve the problem of extending the receptive field using large kernel sizes or multiscale feature fusion, these strategies will cause the increase in parameters and computational cost. To address this problem, the asymmetric convolution [32] reduces the number of parameters by decomposing the convolution kernels of size $k \times k$ into a superposition of two 1-D convolution kernels of size $1 \times k$ and size $k \times 1$. However, this operation leads to offset of pixels in feature maps. To achieve receptive field extension more efficiently, the atrous convolution [33] extends the receptive field using irregular convolutional kernels with null values. This approach is widely used for dense prediction tasks such as image semantic segmentation, e.g., DeepLab V2-V3+ [34], [52], [65] achieved efficient multiscale feature fusion by designing the ASPP module. Besides the design of multiscale convolution kernels, the structure of FPN [19] is also popular for feature fusion. Inspired by FPN and nonlocal [35], Chen et al. proposed NL-FPN [36] for remote sensing image CD, which can effectively fuse multiscale features while capturing the long-range dependence of image.

However, these aforementioned methods need to reuse a large number of convolution or pooling operations at different scales, which results in abundant feature redundancy and computational burden. So far, there is almost no single method that can capture and fuse image multiscale features with a more efficient way.

C. Attention Mechanism for Remote Sensing Image CD

In recent years, attention mechanisms have been proven to be effective in capturing important differences in feature spatial and channel for various computer vision tasks [37]. In particular, self-attention [38] and its variants can model global spatial relationship, which can effectively help networks identify changed and unchanged objected.

Inspired by cognitive science, attention mechanisms in neural networks can usually be categorized into two types, channel attention and spatial attention. Channel attention [66] first uses the fully connected layer on feature maps to model channel relationship, and then applies channel attention weights to the original feature maps to obtain feature maps with different importance. On this basis, Li et al. [39] reapplied the attention weights from squeeze-and-excitation (SE) module to convolutional kernels to select autonomously suitable convolutional kernels of different sizes. To simply the computation of the SE module, Wang et al. [40] captured local cross-channel interactions by considering each channel and its nearest neighbors, and then improved the efficiency of channel attention. Besides channel attention, many methods focus more on the local and global relationships on feature maps. The convolution block attention module (CBAM) [37] initially realizes the cooperation of spatial and channel attention by sequentially cascading them. Zhang et al. [24] and Shi et al. [28] introduced CBAM to remote sensing image CD. They fused spatial attention and channel attention to reconstruct difference images, thus achieving higher CD accuracy. The nonlocal [35] made a network pay more attention to the long-range dependence of spatial features. Based on this idea, Chen et al. [16] introduced a dual attention module to capture long-range dependence, thus improving the feature representation of network. Lei et al. [27] propose a spatial–spectral nonlocal (SSN) strategy for remote sensing image CD. It is different from the vanilla nonlocal module, because spatial multiscale features are incorporated to model the large-scale variation in objects during the process of CD. The module can be used to strengthen the edge integrity and internal tightness of changed objects.

Since the boom of the Transformer-based networks in computer vision tasks, more and more studies have introduced self-attention into image classification and image semantic segmentation [41], [42], [43]. Image semantic segmentation, which is similar to CD, is a kind of dense classification tasks. Based on self-attention, some excellent works have also emerged in the field of remote sensing image CD. Among them, Chen et al. [44] proposed an efficient modeling of global semantic relationship in spatial–temporal, which facilitates the feature representation of changed objects in spatial interest area. Bandara and Patel [45] proposed a Transformer-based Siamese CD framework to model efficiently the long-range dependence required for changed objects. Furthermore, Zhang et al. [59] used the popular Swin Transformer [43] to model the global dependence of bitemporal features. To address the insensitivity of the Transformer to position, Feng et al. [46] introduced depthwise convolutional relative position coding and proposed a CD network combining the Transformer and CNN using the strategy of local and global feature fusion, achieving better CD results.

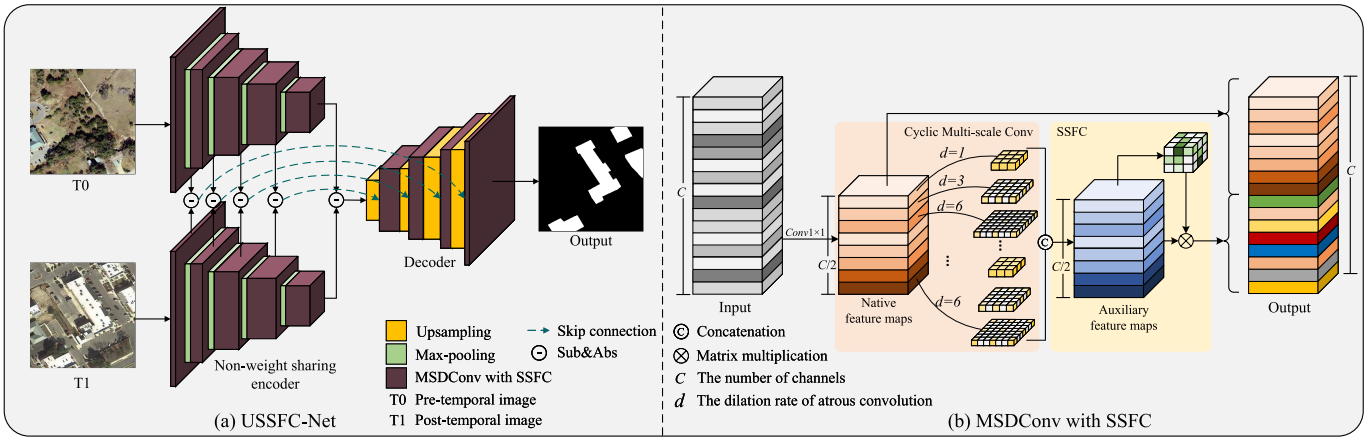


Fig. 1. Architecture of the proposed USSFC-Net. (a) Overview architecture. A pair of bitemporal images are input into a nonweight-sharing encoder, and the difference images at each stage are fused to the decoder. The MSDConv with SSFC strategy is proposed as the basic component of the feature extractor. (b) Proposed MSDConv with SSFC. The MSDConv can capture multiscale feature representations of changed objects. The SSFC generates richer features for MSDConv. Both of them can cooperatively improve the CD accuracy.

To model contextual dependence among feature maps at different stages, Zhang et al. [60] proposed a multilevel change-aware deformable attention.

The above attention-based CD networks can enhance the semantic representation of networks by modeling spatial or spectral relationships, respectively. But these methods not only increase the complexity of models but also ignore the cooperation between spatial and spectral information. It is clear that the current attention-based CD networks cannot effectively model spatial–spectral dependence at the same time.pt

III. Methods

A. Overview

In this section, we first give a brief overview of the proposed USSFC-Net. As shown in Fig. 1, the USSFC-Net consists of a dual branch nonweight-sharing encoder and decoder. We first put a set of bitemporal images into the dual branch encoder which consists of MSDConv and SSFC for feature extraction, respectively. In this stage, each MSDConv block efficiently captures multiscale features of bitemporal images. To enrich the features generated by MSDConv, we use the SSFC strategy for feature enhancement in the spatial correlation expansion phase. This is followed by a decoder consisting of a deconvolutional upsampling layer and a feature recovery layer using proposed MSDConv. At each stage of the encoder, we acquire a difference image and connect it to the corresponding position of the decoder to obtain richer feature maps of changed objects. Finally, the network performs the dimensionality reduction and normalization operations using point convolution to output the final CD results.

As can be seen from Fig. 1(a), compared with other popular CD networks, the proposed USSFC-Net makes the following changes in the Siamese structure [14].

1) A nonweight-sharing pseudo-Siamese encoder is used to achieve better feature extraction by increasing few parameters.
2) The proposed MSDConv is introduced to replace the vanilla convolution in the encoder–decoder structure. It is a compact feature extraction module to obtain multiscale feature representations of changed objects. The main idea is to capture efficiently multiscale changed objects by separable cyclic multiscale feature extraction while reducing the number of vanilla convolution parameters.
3) The USSFC-Net introduces an SSFC strategy. It can obtain 3-D attention weights without adding parameters and effectively improve the feature representation ability of the network. The MSDConv preserves integrity of changed objects from a multiscale perspective, and the SSFC helps the MSDConv generate richer features from an attention perspective. Clearly, by combining the MSDConv with the SSFC, we can achieve better CD results for remote sensing images.

B. Efficient Spatial Correlation Extension Using MSDConv

As mentioned previously, we argue that the existing multiscale feature fusion methods reuse a large number of convolution kernels or pooling operations at different scales. To improve the efficiency of multiscale feature extraction and fusion, we propose the MSDConv that can effectively capture multiscale features of images without adding any additional parameters and computational costs. The structure of MSDConv is shown in Fig. 1(b). The MSDConv is inspired by Xception [17] but is different from it. The MSDConv not only obeys the conclusion that the spatial and channel correlation can be sufficiently decoupled as proposed in [17] but also additionally implements an efficient spatial correlation expansion strategy. Specifically, let $\mathbf{X}$ be input feature maps, $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$, where C, H, and W denote the number of channels of the feature maps, the height of the feature maps, and the width of the feature maps, respectively. The process of generating feature maps using the vanilla convolution can be expressed as

$$\mathbf{Y}_{h,w,c'} = \sum_{i,j,c} \widetilde{\mathbf{W}}_{i,j,c,c'} \times \mathbf{X}_{h+i-1,w+j-1,c} \tag{1}$$

where $\mathbf{Y} \in \mathbb{R}^{C'\times H\times W}$ is the output feature map, and $\widetilde{\mathbf{W}} \in \mathbb{R}^{K\times K\times C\times C'}$ denotes the vanilla convolution kernel. K denotes the kernel size. C denotes the number of channels of the input

feature maps. C' denotes the number of channels of the output feature maps. c and c' denote the index of one channel of one input and output feature map, respectively. i and j denote the spatial positions of convolution kernels. h and w denote the spatial positions of input feature maps. According to (1), when using the vanilla convolution to learn features, the number of parameters denoted by P and the computational costs denoted by Q are expressed as

$$P = K \times K \times C \times C' \tag{2}$$

$$Q = K \times K \times C \times C' \times H \times W. \tag{3}$$

To reduce the parameter redundancy of the vanilla convolution operation, a novel convolution operation with decoupling spatial and channel correlations is proposed. Specifically, to obtain C' feature maps, we first generate $C'/2$ native feature maps using point convolution. Thus, the native feature maps are generated without any mapping of spatial correlations, and only to obtain tight channel correlations by dimensionality reduction. In the second stage, we use a cyclic multiscale convolution to extend the spatial correlation of the native feature maps, thus obtaining the auxiliary feature maps. As shown in Fig. 1(b), the cyclic multiscale convolution is achieved by atrous convolution with combination of dilation rates such as (1, 3, 6). It is worth noting that cyclic multiscale convolution expands different dilation rates to the corresponding convolution kernels at the same convolution layer. This ensures that the MSDConv can capture the multiscale features of changed objects through only one convolutional layer and fuse the multiscale features through iterations of layers. Analogously to (1), cyclic multiscale convolution with dilation rate d can be expressed as

$$\mathbf{Y}'_{h,w,c'} = \sum_{i,j} \widetilde{\mathbf{W}}'_{i,j,c} \times \mathbf{X}_{h+id-1,w+jd-1,c} \tag{4}$$

where $\widetilde{\mathbf{W}}' \in R^{K\times K\times (C'/2)}$ is a cyclic multiscale convolution, in which the cth convolution kernel is computed with the cth channel of the feature map $\mathbf{X}$ to obtain the cth output feature map $\mathbf{Y}'$. In fact, we use point convolution for aggregation between channels and cyclic multiscale convolution to generate feature maps, and the number of parameters denoted by P_m and the computational cost denoted by Q_m are expressed as

$$P_m = C \times \frac{C'}{2} + K \times K \times \frac{C'}{2} \tag{5}$$

$$Q_m = C \times \frac{C'}{2} \times H \times W + K \times K \times \frac{C'}{2} \times H \times W \tag{6}$$

where Q_m is the sum of the computations of 1×1 pointwise convolution and cyclic multiscale convolution. Compared with the vanilla convolution, the proposed MSDConv can effectively reduce the number of parameters and computations as follows:

$$r = \frac{C \times \frac{C'}{2} + K \times K \times \frac{C'}{2}}{K \times K \times C \times C'} = \frac{1}{2K^2} + \frac{1}{2C} \tag{7}$$

where r denotes the ratio of the number of parameters and computations required for MSDConv and vanilla convolution. From (7), the parameters of the proposed MSDConv are only $[1/(2K^2) + 1/(2C)]$ of the vanilla convolution. Meanwhile, the MSDConv can capture the multiscale features of changed objects. As far as we know, this article is the first study to capture and fuse the multiscale information of changed objects by designing compact convolution kernels.

C. Spatial–Spectral Feature Cooperation

In Section III-B, we proposed MSDConv to obtain the multiscale features of changed objects. However, if the auxiliary feature maps are directly fused with the native feature maps, the spatial–spectral dependency will be ignored. As a result, we design an SSFC strategy to model spatial–spectral dependence to obtain richer features.

In cognitive science, the human brain generates attention through intentional or unintentional focusing on an object. The correspondence between intention and targets is based on three elements in the attention mechanism: query, key, and value. We can understand query as intention and key as a target. The attention mechanism is to find the relationship between query and key and map it into value to refine the feature maps. To get the attention relationship between query and key, we are inspired by Nadaraya–Watson kernel regression [56], [57] and design a Gaussian-kernel-based SSFC strategy. The Nadaraya–Watson kernel regression is expressed in terms of three elements of the attention mechanism as

$$\widetilde{\mathbf{Y}} = \sum_{i=1}^{n} \frac{F(\mathbf{Q} - \mathbf{K}_i)}{\sum_{j=1}^{n} F(\mathbf{Q} - \mathbf{K}_j)} \times \mathbf{V}_i \tag{8}$$

where $\widetilde{\mathbf{Y}}$ denotes the output feature map of attention mechanism, n denotes the feature vector dimension, $F(\bullet)$ is the kernel function, and $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ denote the query, key, and value, respectively. If a Gaussian kernel function is adopted here, the Nadaraya–Watson kernel regression can be expressed as

$$F(x) = e^{-\frac{x^2}{2}} \tag{9}$$

$$\widetilde{\mathbf{Y}} = \sum_{i=1}^{n} \text{Softmax}\left(-\frac{(\mathbf{Q} - \mathbf{K}_i)^2}{2}\right) \times \mathbf{V}_i. \tag{10}$$

According to (8)–(10), a more generalized model of attention mechanism is defined as

$$\widetilde{\mathbf{Y}} = N(\varpi(\mathbf{Q}, \mathbf{K})) \times \mathbf{V} \tag{11}$$

where $\varpi(\mathbf{Q}, \mathbf{K})$ denotes the attention weights obtained by modeling the relationship between query and key, and $N(\bullet)$ denotes the normalization function. Therefore, we can extend the Nadaraya–Watson kernel regression to a higher dimensional tensor and then design the SSFC strategy using the Gaussian kernel function as follows:

$$\widetilde{\mathbf{Y}} = \text{Sigmoid}\left(\frac{(\mathbf{Q} - \mathbf{K})^2}{2\sigma^2} + \frac{1}{2}\right) \times \mathbf{V}. \tag{12}$$

In practical applications, both $\mathbf{K}$ and $\mathbf{V}$ are the input feature maps $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$, where C, H, and W denote the number of channels of the feature maps, the height of the feature maps, and the width of the feature maps, respectively. $\mathbf{Q}$ is the mean value of channel dimension $\overline{\mathbf{X}} \in \mathbb{R}^{C\times 1\times 1}$, and

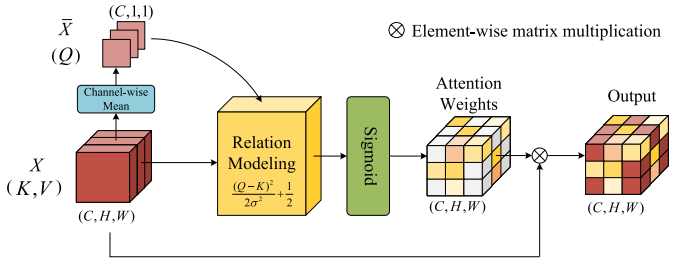


Fig. 2. Overview of the proposed SSFC strategy. This strategy is clearly different from the popular 2-D attention mechanism since it does not require any learnable parameters and generates 3-D attention weights on feature maps only through heuristic computation.

σ^2 is the channel dimensional variance. The value of σ^2 influences the richness of feature maps, and a larger value of σ^2 means that the variance of the feature map is larger, which corresponds to richer contextual information in the feature map. To make the attention weights be positive facilitation, we add 1/2 to the original attention score and normalize it using Sigmoid function to obtain the attention weights. Finally, the obtained attention weights are multiplied with input feature maps $\mathbf{X}$(value) to obtain the output feature maps $\widetilde{\mathbf{Y}}$. The SSFC strategy is shown in Fig. 2.

Our proposed SSFC strategy can generate 3-D attention weights using the idea of SSFC. By modeling spatial–spectral dependence, the SSFC can enhance the edge and internal details of changed objects in remote sensing images. Compared with the existing attention mechanisms [37], [50], our proposed SSFC does not add any learnable parameters, which is simpler and more efficient. Finally, we embed the SSFC into the MSDConv, as shown in Fig. 1(b).

D. Building Ultralightweight CD Network

1) Architecture: We improve the popular Siamese structure by building a nonweight-sharing pseudo-Siamese structure based on the U-shaped network. The nonweight-sharing encoder allows more flexibility in learning feature coding weights. Compared with the weight-sharing structure, our network only increases parameters by 0.33 M. We conducted a comparison experiment between pseudo-Siamese and Siamese structure in Section V. Compared with other complex network designs, a simple feature extraction using the pseudo-Siamese structure is presented, relying only on difference images and skip-connections for temporal difference information interaction.

2) Encoder: We use a pseudo-Siamese network to extract bitemporal image features. Specifically, the encoder of USSFC-Net uses five successive down sampling steps, i.e., stages 0–4. In stage 0, we use the vanilla convolution to ensure adequate edge and texture information of changed objects. In the range stages, we use the proposed MSDConv with the SSFC to encode semantic information efficiently. We set the number of feature map channels to 512 at stage 4 to extract sufficient semantic information while maintaining the lightweight of the network.

3) Decoder: To recover the semantic features generated by the encoder, we design a simple yet efficient decoder to obtain change maps. The decoder uses an approximately symmetric structure with the encoder and requires four successive deconvolutions to achieve feature maps upsampling. At the end of each stage, we use the proposed MSDConv to recover the feature of changed objects. Finally, the 1×1 convolution and activation are used to obtain the predicted change map.

TABLE I

OVERALL ARCHITECTURE OF OUR USSFC-NET. THE OUT-CHS DENOTES OUTPUT CHANNELS, THE C-BLOCK DENOTES DOUBLE 2-D CONVOLUTIONS, THE M-BLOCK DENOTES DOUBLE MSDCONV WITH SSFCS, THE MAXPOOL DENOTES MAX POOLING, AND THE TRANSCONV DENOTES TRANSPOSE CONVOLUTION

Component	Stage	Input size	Operator	Out-chs
Encoder	0	256×256×3	C-Block 3×3	32
		256×256×32	MaxPool 2×2	32
	1	128×128×32	M-Block 3×3	64
		128×128×64	MaxPool 2×2	64
	2	64×64×64	M-Block 3×3	128
		64×64×128	MaxPool 2×2	128
	3	32×32×128	M-Block 3×3	256
		32×32×256	MaxPool 2×2	256
	4	16×16×256	M-Block 3×3	512
Decoder	5	16×16×512	TransConv 2×2	256
		32×32×256	M-Block 3×3	256
	6	32×32×256	TransConv 2×2	128
		64×64×128	M-Block 3×3	128
	7	64×64×128	TransConv 2×2	64
		128×128×64	M-Block 3×3	64
	8	128×128×64	TransConv 2×2	32
		256×256×32	M-Block 3×3	32
Classifier	9	256×256×32	Conv 2d 1×1	1

4) Details: According to the above settings, we build an ultralightweight CD network with a Siamese U-Net as backbone. Different from the Siamese U-Net, we also make some changes to it by considering the CD task. First, we use a nonweight-sharing two-branch network as the encoder, which makes feature extraction more flexible. Second, we halve the number of channels of feature maps at each stage of the network to make the network more compact. The specific network structure is shown in Table I.

5) Loss Function: Remote sensing image CD is essentially a pixel-level classification task. In the network training stage, we use binary cross-entropy loss to optimize the network weights. Formally, the loss function is defined as

$$L = -\frac{1}{N_s}\sum_{i} y_i \cdot \log x_i \tag{13}$$

where N_s denotes the total number of training samples, y_i denotes the label of the ith sample, and x_i denotes the predicted value of the ith sample.

IV. EXPERIMENTS

We conduct a series of comparative experiments and ablation studies to evaluate our proposed USSFC-Net for remote sensing image CD. We analyze the effectiveness of the proposed USSFC-Net according to the accuracy of CD and we compare the efficiency of the proposed USSFC-Net and state-of-the-art networks according to the number of parameters and computational costs.

A. Experimental Setup

The experiments are conducted on three remote sensing image CD datasets. They are LEVIR-CD [25], CDD [51], and DSIFN-CD [24].

LEVIR-CD is a large public CD dataset covering a variety of complex change features. It contains 637 pairs of remote sensing images of size 1024 × 1024 with 0.5-m resolution. To make full use of GPU memory and prevent overfitting, we crop the images into 13 072 patches of size 256 × 256. Finally, the dataset is divided into three parts: 10 000/1024/2048 for training/validation/test, respectively.

CDD is a public CD dataset of seasonal changes in the same area obtained from Google Earth. It contains 11 pairs of multispectral images with resolutions ranging from 0.03 to 1 m. In all, 16 000 patches of size 256 × 256 are obtained from the original images by cropping and rotation operations. The final dataset is divided into three parts: 10 000/3000/3000 for training/validation/test, respectively.

DSIFN-CD is a public CD dataset manually collected from Google Earth. It consists of six high-resolution images from different cities in China. The authors provide cropping the Xi'an image pair into 48 patches of size 512 × 512 for model testing. The other five city images were cropped into 3940 patches of the same size for training and validation. The final obtained dataset is divided into three parts: 3600/340/48 for training/validation/test, respectively.

Evaluation Metrics: To evaluate the performance of the proposed method, we mainly use three evaluation metrics for comprehensive evaluation of the experimental results, including Precision (Pre), Recall (Rec), and F1-score (F1). Specifically, Pre reflects the proportion of correct predictions in the positive samples predicted by the model, Rec reflects the correct proportion of model predictions in all the positive samples, and F1 is the weighted harmonic mean of both. In general, a higher F1 indicates a better detection accuracy of model. These metrics are defined as

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{14}$$

$$\text{Rec} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{15}$$

$$\text{F1} = 2 \times \frac{\text{Pre} \times \text{Rec}}{\text{Pre} + \text{Rec}} \tag{16}$$

where TP, FP, and FN represent the number of true positive, false positive, and false negative, respectively.

Implementation Details: The proposed USSFC-Net is implemented by PyTorch and trained using NVIDIA GeForce RTX 3090 GPU for 200 epochs. During the training process, we apply Kaiming initialization [58] training from scratch and use the Adam [53] optimization algorithm to optimize the model, setting the momentum to 0.99. The weight decay is set to 0.0005. The batch size is 32 and the initial learning rate is 0.0001. For MSDConv, the combination of dilation rates is (1, 3, 6) which proved to be optimal in Section V.

B. Comparison With State-of-the-Art Methods

To verify the superiority of our USSFC-Net, several state-of-the-art methods for remote sensing image CD are adopted as comparative methods, including fully convolutional early fusion network (FC-EF) [14], fully convolutional Siamese-difference network (FC-Siam-Di) [14], fully convolutional Siamese-concatenation network (FC-Siam-Conc) [14], FCN with pyramid pooling (FCN-PP) [29], spatial-temporal attention-based network (STANet) [25], deeply supervised image fusion network (IFNet) [24], feature difference convolutional neural network (FDCNN) [55], densely connected Siamese neasted U-shape network (SNUNet) [54], deeply supervised attention metric-based network (DSAMNet) [28], and bitemporal image transformer (BIT) [44].

FC-EF fuses bitemporal images at an early stage and then performs CD using a FCN.

FC-Siam-Di uses the multilayer difference features of the Siamese network to fuse the bitemporal information, and then achieves the CD.

FC-Siam-Conc achieves feature long-range mapping by fusing bitemporal features through skip-connections via a fully convolutional Siamese network.

FCN-PP proposes a pyramid pooling module that uses multiple convolutions to explore efficiently the context of bitemporal remote sensing images.

STANet designs a multiscale attention mechanism to model the spatial–temporal relationships of bitemporal remote sensing images, generating better feature representations for changed objects of different sizes.

IFNet proposes a deeply supervised difference network for CD and reconstructs change maps using a strategy of multiscale feature fusion.

FDCNN improves CD accuracy by generating multiscale and multidepth difference maps.

SNUNet designs a densely connected U-shaped Siamese network and refines features on different semantic information using an integrated channel attention module. To be fair, among the multiple networks provided by the authors, we choose SNUNet-16, which has the same magnitude of the number of parameters as the proposed USSFC-Net.

DSAMNet organically combines metric-based and classification-based CD methods and introduces the deeply supervised module to enhance the learning ability of the feature extractor and generate more useful features.

BIT represents high-level semantic features by context-rich tokens and introduces a Transformer-based encoder to model context-based space-time.

The quantitative analysis of the experimental results on the LEVIR-CD dataset is shown in Table II, where the best values are in bold. Compared with CNN-based methods, our proposed USSFC-Net obtains the best results with significant advantages. For example, compared with SNUNet, our method achieves 5.0%/1.1%/3.1% higher in Pre, Rec, and F1, respectively. The experimental results on the CDD dataset are shown in Table III, where our method is 1.8%/1.3%/1.5% higher compared with DSAMNet. In addition to LEVIR-CD and CDD, we test the proposed USSFC-Net on a smaller public dataset DSIFN-CD, and the experimental results are shown in Table IV. Compared with IFNet, our USSFC-Net is 12.4%/1.6% higher in Rec and F1, respectively.

In particular, compared with the latest Transformer-based method BIT, our proposed USSFC-Net leads 0.5%/3.0%/1.7%

TABLE II

COMPARISON RESULTS ON THE LEVIR-CD TEST SET. LARGER VALUES OF BOTH PRE AND REC INDICATE A BETTER MODEL. F1 TAKES BOTH INDICES INTO ACCOUNT. A LARGER VALUE OF F1 INDICATES A BETTER MODEL. THE BEST VALUES ARE IN BOLD. OUR USSFC-NET IS SUPERIOR TO BOTH CNN-BASED METHODS AND TRANSFORMER-BASED METHOD

Method Type	Network	Pre (%)	Rec (%)	$F1$ (%)
CNN	FC-EF [14]	74.96	90.53	82.01
	FC-Siam-Di [14]	78.18	92.92	84.92
	FC-Siam-Conc [14]	74.32	91.63	82.07
	FCN-PP [29]	80.31	89.48	84.64
	STANet [25]	86.14	89.39	87.73
	IFNet [24]	87.55	86.52	87.03
	FDCNN [57]	82.99	88.71	85.76
	SNUNet [56]	84.66	91.34	87.87
	DSAMNet [29]	82.75	88.39	85.48
	USSFC-Net (ours)	89.70	92.42	**91.04**
Transformer	BIT [46]	89.24	89.37	89.31

TABLE III

COMPARISON RESULTS ON THE CDD TEST SET. LARGER VALUES OF BOTH PRE AND REC INDICATE A BETTER MODEL. F1 TAKES BOTH INDICES INTO ACCOUNT. A LARGER VALUE OF F1 INDICATES A BETTER MODEL. THE BEST VALUES ARE IN BOLD. OUR USSFC-NET IS SUPERIOR TO BOTH CNN-BASED METHODS AND TRANSFORMER-BASED METHOD

Method Type	Network	Pre (%)	Rec (%)	$F1$ (%)
CNN	FC-EF [14]	52.67	84.20	64.80
	FC-Siam-Di [14]	61.85	76.69	68.48
	FC-Siam-Conc [14]	44.07	80.44	56.94
	FCN-PP [29]	81.69	90.31	85.78
	STANet [25]	88.98	93.11	91.00
	IFNet [24]	85.33	91.76	88.43
	FDCNN [57]	83.61	91.70	87.47
	SNUNet [56]	90.92	94.75	92.79
	DSAMNet [29]	91.67	94.83	93.22
	USSFC-Net (ours)	93.45	96.08	**94.74**
Transformer	BIT [46]	92.89	94.02	93.45

TABLE IV

COMPARISON RESULTS ON THE DSIFN-CD TEST SET. LARGER VALUES OF BOTH PRE AND REC INDICATE A BETTER MODEL. F1 TAKES BOTH INDICES INTO ACCOUNT. A LARGER VALUE OF F1 INDICATES A BETTER MODEL. THE BEST VALUES ARE IN BOLD. OUR USSFC-NET IS SUPERIOR TO BOTH CNN-BASED METHODS AND TRANSFORMER-BASED METHOD

Method Type	Network	Pre (%)	Rec (%)	$F1$ (%)
CNN	FC-EF [14]	50.01	55.99	52.84
	FC-Siam-Di [14]	52.62	56.94	54.69
	FC-Siam-Conc [14]	48.67	56.19	52.16
	FCN-PP [29]	56.42	59.25	57.80
	STANet [25]	66.22	67.16	66.69
	IFNet [24]	72.36	63.86	67.85
	FDCNN [57]	64.42	68.38	66.34
	SNUNet [56]	62.47	69.74	65.90
	DSAMNet [29]	61.28	75.41	67.62
	USSFC-Net (ours)	63.73	76.32	**69.47**
Transformer	BIT [46]	68.36	70.18	69.26

on the LEVIR-CD dataset in Pre, Rec, and F1, respectively, and 0.6%/2.1%/1.3% on the CDD dataset, and 6.1%/0.2% on the DSIFN-CD dataset in Rec and F1. In addition, as shown in Table V, we compare the parameters and computational costs of USSFC-Net and BIT. The results show that the proposed USSFC-Net can reduce Params. by 78.1% and floating-point operations (FLOPs) by 43.8%. We think this result can be summarized as follows. First, convolution has an advantage in the low-level visual feature extraction due to its localization and translation invariance. Different from the concatenation of attention and convolution layers, we embed both into the one module to model spatial–spectral dependence. Second, self-attention always tends to have a higher FLOPs than convolution. It adds more extra computational costs to the model.pt

The quantitative results on three public datasets illustrate that our proposed USSFC-Net can achieve the best accuracy on CD datasets.

The visual analysis of the experimental results on the LEVIR-CD, CDD, and DSIFN-CD datasets is shown in Fig. 3. In the case of large-scale building detection, for instance, the first row in LEVIR-CD, the boundaries of the detection result of most comparative methods are not smooth enough, and there is a lack of internal areas of the buildings. However, our proposed USSFC-Net not only provides good results at the edges of large buildings but also adequately detects interior areas. In areas where change objects are dense, for example, the second row in LEVIR-CD, our USSFC-Net provides more complete areas in detection than the comparative methods, without any missed or false detections. It is worth noting that in the third row of CDD, our USSFC-Net accurately detects changed inconspicuous objects, while most comparative methods fail to detect these inconspicuous objects, indicating their poor robustness to shadows and noise contained in remote sensing images. In brief, our proposed USSFC-Net shows obvious advantages in handling marginally irregular and sparse small changing objects, achieving the best performance for remote sensing image CD.

C. Model Efficiency

We tested and analyzed the computational efficiency of the proposed USSFC-Net and comparative methods for remote sensing image CD from different perspectives. Three metrics are included: F1, number of parameters (Params.), and FLOPs. The detailed results are shown in Table V and Fig. 4. Since the methods [14] use fewer layers of feature extraction and their detection accuracy is low, we do not compare with them. It is obvious that our method with the highest F1 requires the lowest parameters and computational costs.

D. Ablation Studies

To validate the effectiveness of the proposed MSDConv and SSFC, we conduct ablation experiments for these two components separately. Due to the comprehensive consideration of the dataset size and the number of changed objects, the ablation experiments are conducted on LEVIR-CD, and the results are shown in Table VI. Our method uses the nonweight-sharing Siamese U-Net as the baseline. It can be seen from the experimental results that the proposed MSDConv, by fully capturing and fusing the multiscale information of changed

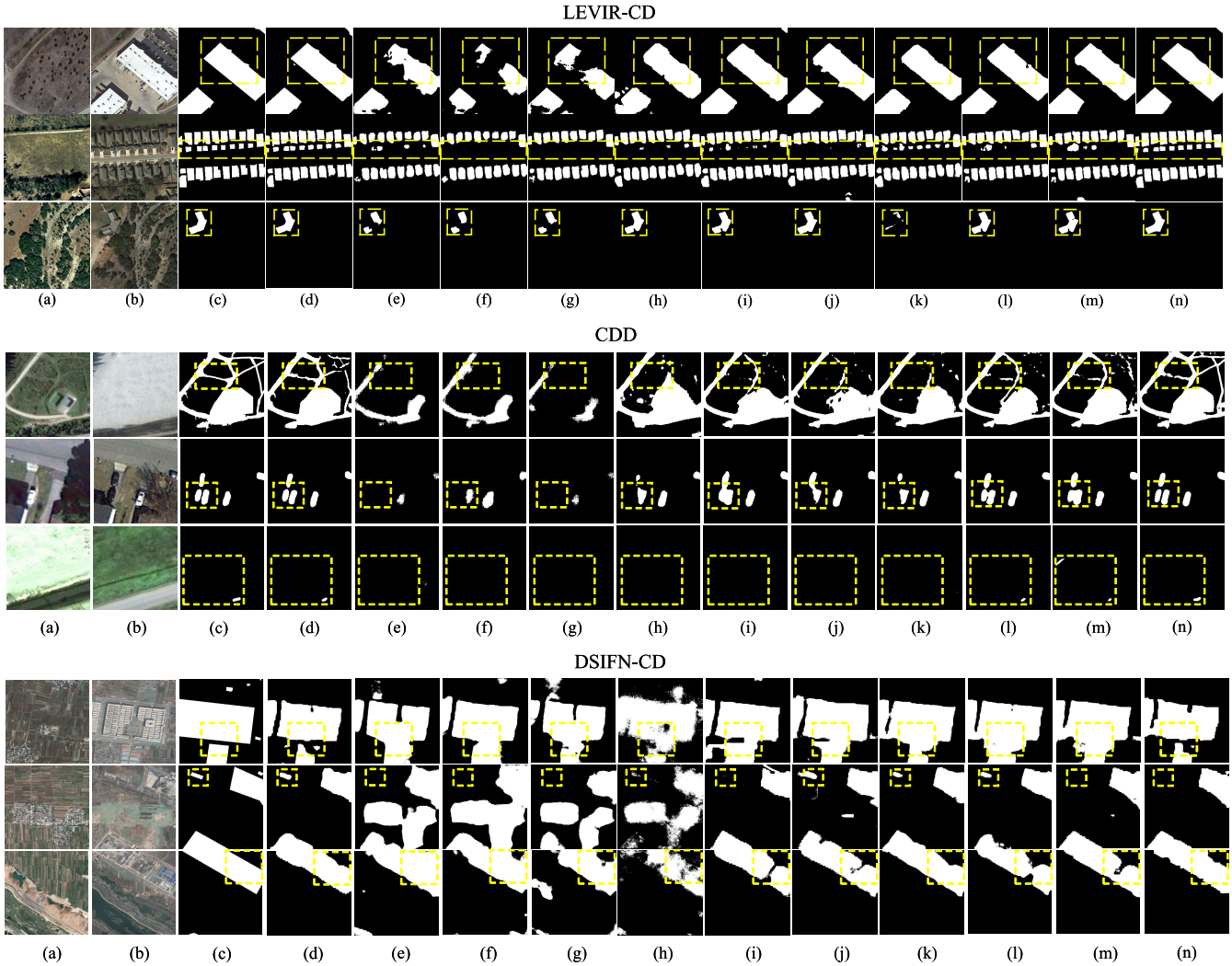


Fig. 3. Visualization comparison of results on the LEVIR-CD, CDD, and DSIFN-CD test sets. White is true positive and black is true negative. (a) Pretemporal image. (b) Posttemporal image. (c) Ground truth. (d) Our USSFC-Net. (e) FC-EF. (f) FC-Siam-Di. (g) FC-Siam-Conc. (h) FCN-PP. (i) STANet. (j) IFNet. (k) FDCNN. (l) SNUNet. (m) DSAMNet. (n) BIT.

TABLE V

COMPARISON RESULTS OF COMPUTATIONAL EFFICIENCY. WE REPORT PARAMETERS (PARAMS.) AND FLOPS, AS WELL AS F1 ON THREE CD TEST SETS. A LARGER VALUE OF F1 INDICATES A BETTER MODEL. THE BEST VALUES ARE IN BOLD

Method Type	Network	$F1$ (%)			Params. (M)	FLOPs (G)
		LEVIR-CD	CDD	DSIFN-CD		
CNN	FC-EF [14]	82.01	64.80	52.84	**0.85**	3.34
	FC-Siam-Di [14]	84.92	68.48	54.69	**0.85**	**3.33**
	FC-Siam-Conc [14]	82.07	56.94	52.16	1.07	4.08
	FCN-PP [29]	84.64	85.78	57.80	28.13	34.65
	STANet [25]	87.73	91.00	66.69	16.93	6.58
	IFNet [24]	87.03	88.43	67.85	50.71	41.18
	FDCNN [57]	85.76	87.47	66.34	1.86	32.40
	SNUNet [56]	87.87	92.79	65.90	3.01	27.44
	DSAMNet [29]	85.48	93.22	67.62	16.95	75.29
	USSFC-Net (ours)	**91.04**	**94.74**	**69.47**	1.52	4.86
Transformer	BIT [46]	89.31	93.45	69.26	6.93	8.44

objects, can improve F1 by 1.2% on top of the baseline. After we embed the SSFC into MSDConv, we can improve F1 by 2.5% on top of the baseline due to the richer features obtained by the cooperation of the spatial–spectral features. Compared with other multiscale feature fusion and attention modules, our method achieves significantly better detection accuracy.

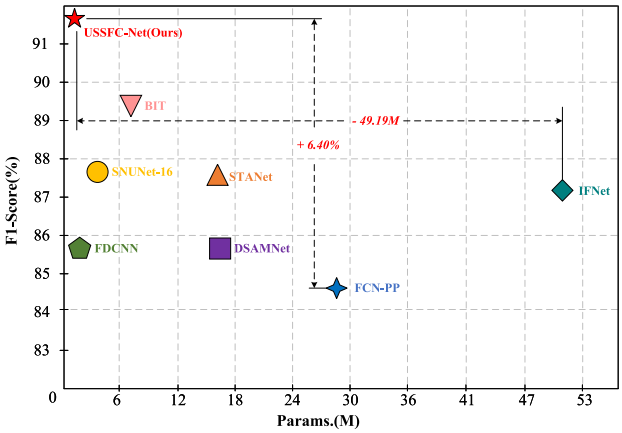


Fig. 4. Comparison of different methods on computational efficiency (Params.) and accuracy (F1).

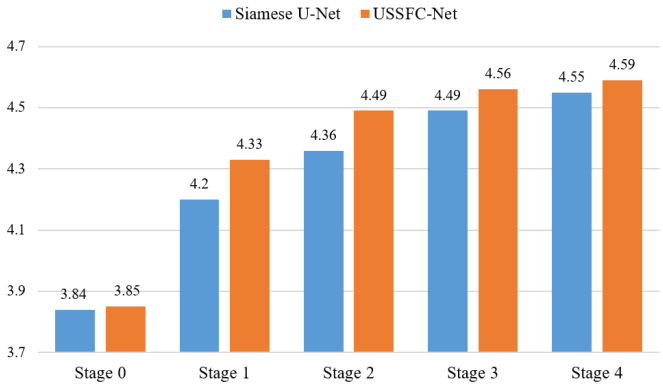


Fig. 5. Entropy of convolution layers of different networks on the LEVIR-CD test set.

TABLE VI

ABLATION STUDY ON THE PROPOSED MSDCONV AND SSFC ON THE LEVIR-CD TEST SET. WE REPORT PARAMETERS (PARAMS.) AND FLOPS, AS WELL AS F1. A LARGER VALUE OF F1 INDICATES A BETTER MODEL. THE BEST VALUES ARE IN BOLD

Network	Params. (M)	FLOPs (G)	$F1$ (%)
Siamese U-Net	12.48	18.06	88.50
+ PP [29]	22.71	20.48	89.10
+ ASPP [65]	16.94	19.17	89.42
+ **MSDConv (ours)**	**1.52**	**4.86**	**89.68**
+ MSDConv + SE [66]	1.56	4.87	90.37
+ MSDConv + CBAM [37]	1.57	4.89	90.27
+ **MSDConv + SSFC (ours)**	**1.52**	**4.86**	**91.04**

TABLE VII

COMPARISON OF DIFFERENT LIGHTWEIGHT CONVOLUTIONS ON THE LEVIR-CD TEST SET. WE REPORT PARAMETERS (PARAMS.) AND FLOPS, AS WELL AS F1. A LARGER VALUE OF F1 INDICATES A BETTER MODEL. THE BEST VALUES ARE IN BOLD

Network	$F1$ (%)	Params.(M)	FLOPs (G)
Siamese U-Net	88.50	12.48	18.06
+ SPConv [48]	89.96	7.92	12.52
+ Ghost [47]	90.02	**1.40**	**4.44**
+ **MSDConv (ours)**	**91.04**	1.52	4.86

The ablation experiments demonstrate the effectiveness of our proposed MSDConv and SSFC.

V. DISCUSSION

A. Effectiveness of MSDConv

To verify that the proposed MSDConv contributes more to the compression rate and accuracy of the network than the currently popular lightweight convolutional operations, we conduct experiments on the LEVIR-CD dataset, which includes the performance comparison of the proposed MSDConv with similar lightweight convolutional operations including SPConv [48] and Ghost Module [47], as well as visualization results for the abundance of generating features.

Table VII shows the F1-score (F1), parameters (Params.), and FLOPs using MSDConv, SPConv, and Ghost module on the LEVIR-CD test set. The Siamese U-Net is considered as a backbone network, and we use MSDConv, SPConv, and Ghost module instead of the vanilla convolution used by the Siamese U-Net except the first stage of the network. The results show that our proposed MSDConv performs significantly better than SPConv and Ghost module. The proposed MSDConv significantly improves F1 by 1.1% and 1.02% compared with SPConv and Ghost module. Moreover, MSDConv has similar performance with the Ghost module in reducing the network parameters.

There are two main reasons why our proposed MSDConv with SSFC can provide good experimental results. One is that the MSDConv can adequately capture multiscale features of objects and the other is the SSFC strategy generates richer feature maps by modeling the spatial–spectral features. To demonstrate this, we introduce information entropy to estimating information stored in convolution layers. In general, a higher information entropy means that the convolution layer contains richer information. Since it is very difficult to calculate the entropy of the continuous distribution in a convolution layer, we divide the continuous distribution into several different discrete zones and then calculate the probability of each zone. Fig. 5 shows the comparison of information entropy from each stage of the Siamese U-Net encoder and the USSFC-Net encoder. We can see that our USSFC-Net obtains higher entropy values than Siamese U-Net at each stage, which means that our method can provide richer features. Furthermore, we visualize the network interlayer feature maps at the second convolutional layer as shown in Fig. 6. The results show that the proposed MSDConv generates richer features, which can effectively help networks improve the CD accuracy.

The concatenation of multiple atrous convolutions can usually improve the feature representation ability of a network, thereby improving the accuracy of CD in remote sensing images. However, blindly increasing the value of dilation rates may have an opposite effect on improving the network performance. For example, Chen et al. [52] mentioned that when the size of the feature map was similar to the dilation rate, the atrous convolution would degenerate into a 1×1 convolution. Experiments demonstrate that the best segmentation results can be achieved when the multigrid size is (1, 2, 4). For the

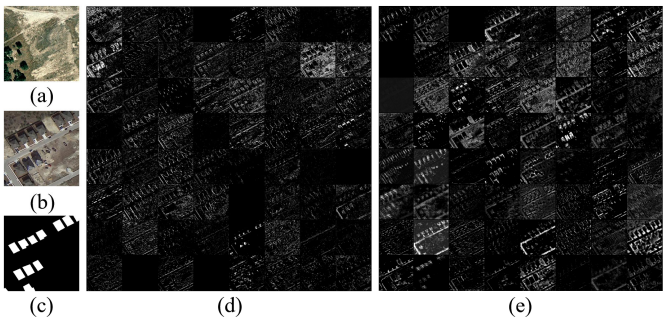


Fig. 6. Feature visualization of the proposed MSDConv on the LEVIR-CD test set. (a) Pretemporal image. (b) Posttemporal image. (c) Ground truth. (d) Siamese U-Net. (e) USSFC-Net.

TABLE VIII

COMPARISON OF DIFFERENT DILATION RATE COMBINATIONS IN MSDCONV ON THE LEVIR-CD TEST SET. LARGER VALUES OF BOTH PRE AND REC INDICATE A BETTER MODEL. F1 TAKES BOTH INDICES INTO ACCOUNT. A LARGER VALUE OF F1 INDICATES A BETTER MODEL. THE BEST VALUES ARE IN BOLD

Network	Pre (%)	Rec (%)	$F1$ (%)
Siamese U-Net	87.26	89.77	88.50
+ MSDConv -dilation rates = (1)	89.92	88.31	89.11
+ MSDConv -dilation rates = (1, 2)	90.15	88.02	90.08
+ MSDConv -dilation rates = (1, 2, 4)	90.93	90.75	90.84
+ MSDConv -dilation rates = (1, 3, 6)	89.70	92.42	**91.04**
+ MSDConv -dilation rates = (1, 4, 8)	89.98	91.59	90.78
+ MSDConv -dilation rates = (1, 2, 4, 8)	89.26	90.27	90.16

CD task, we also conduct some experiments to choose the dilation rate set of MSDConv more scientifically, as shown in Table VIII. First, we set atrous convolution in MSDConv to be a cyclic combination of dilation rates based on the experience of previous works containing (1), (1, 2), (1, 2, 4), and (1, 2, 4, 8). Second, according to the previous experimental results, we find that the accuracy of the three dilation rates is the highest. Therefore, we build a series of dilation rate combinations containing (1, 2, 4), (1, 3, 6), and (1, 4, 8). We conduct ablation experiments on different combinations, from which the optimal dilation rate combination (1, 3, 6) is finally selected.

B. Necessity of SSFC Strategy

As mentioned earlier, the SSFC strategy uses the spatial–spectral feature to refine the feature maps. In fact, the direct fusion of native feature maps and auxiliary feature maps easily causes over-redundant features [47]. Therefore, we adopt the SSFC strategy to suppress the redundancy of the auxiliary feature maps and generate abundant features. In addition, the SSFC strategy can efficiently model the spatial–spectral dependence and guide the network learning to focus on significantly changed objects without adding any learnable parameters. Thus, the advantages of MSDConv and SSFC complement each other. It is also demonstrated in ablation studies. We demonstrate the interpretability of SSFC by visualizing the feature activation of USSFC-Net. Fig. 7 shows the attention activation maps on the LEVIR-CD test set.

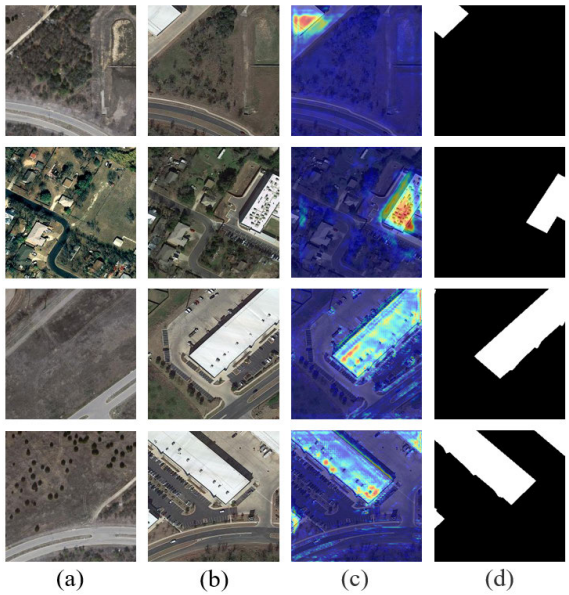


Fig. 7. Feature activation after SSFC on the LEVIR-CD test set. (a) Pretemporal image. (b) Posttemporal image. (c) Feature activation map. (d) Ground truth. Red and yellow in (c) denote higher attention values.

TABLE IX

COMPARISONS OF SIAMESE USSFC-NET (SIAM) AND PSEUDO-SIAMESE ONE (P-SIAM) ON THREE TEST SETS. THE BEST VALUES (F1-SCORE) ARE IN BOLD

Network	LEVIR-CD (%)	CDD (%)	DSIFN-CD (%)	Params. (M)
Siam	89.58	94.06	57.80	**1.19**
P-Siam	**91.04**	**94.74**	**69.47**	1.52

C. Discussion on Siamese Network

The weight-sharing Siamese structure can map bitemporal images to the same feature space, which has spatial–spectral feature uniformity for metric-based CD methods. Since our CD method is based on pixel-level classification, we consider whether using a nonweight-sharing pseudo-Siamese structure for bitemporal image feature extraction can improve the detection accuracy. The pseudo-Siamese structure can independently perform feature extraction for bitemporal images, and the additional parameters introduced can achieve more complex feature representation, which promotes the CD accuracy of the network. We train the weight-sharing Siamese USSFC-Net (Siam) and the nonweight-sharing pseudo-Siamese USSFC-Net (P-Siam) for the U-shaped network backbone, respectively. As shown in Table IX, the experiments demonstrate that the P-Siam can obtain better detection accuracy. Thanks to our proposed MSDConv and SSFC strategies, the pseudo-Siamese structure only increases the number of parameters by 0.33 M.

VI. CONCLUSION

In this article, we have proposed a USSFC-Net for CD in remote sensing images. The proposed USSFC-Net solves the main problems of current CD by introducing MSDConv and SSFC. Specifically, MSDConv can effectively extract multiscale features of changed objects by designing a compact structure. The SSFC strategy effectively captures global

contextual information to refine features by comodeling spatial and spectral features and does not require any additional parameters. We tested USSFC-Net on three CD public datasets. The experimental results indicate that our method outperforms other competitive methods based on CNN or Transformer in terms of CD accuracy, parameters, and FLOPs.

It is worth noting that with the popularity of deep learning models, industrial deployments have become an important challenge for the practical applications of current deep learning models. It is hoped that the proposed USSFC-Net can effectively address the challenge of deployment of remote sensing image CD on low-resource devices with improved CD accuracy while achieving efficient simplification of the model.

REFERENCES

[1] A. Sebastian et al., "Temporal correlation detection using computational phase-change memory," *Nature Commun.*, vol. 8, no. 1, pp. 1–10, Oct. 2017.

[2] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 125–138, Jan. 2016.

[3] B. Demir, F. Bovolo, and L. Bruzzone, "Updating land-cover maps by classification of image time series: A novel change-detection-driven transfer learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 300–312, Jan. 2013.

[4] X. Huang, L. Zhang, and T. Zhu, "Building change detection from multitemporal high-resolution remotely sensed images based on a morphological building index," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, pp. 105–115, 2013.

[5] 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.

[6] G. F. Byrne, P. F. Crapper, and K. K. Mayo, "Monitoring land-cover change by principal component analysis of multitemporal Landsat data," *Remote Sens. Environ.*, vol. 10, no. 3, pp. 175–184, Nov. 1980.

[7] Z. Li, W. Shi, H. Zhang, and M. Hao, "Change detection based on Gabor wavelet features for very high resolution remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 783–787, May 2017.

[8] 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, Apr. 1998.

[9] 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. Photogramm. Remote Sens.*, vol. 80, pp. 91–106, Jun. 2013.

[10] D. Xue, T. Lei, X. Jia, X. Wang, T. Chen, and A. K. Nandi, "Unsupervised change detection using multiscale and multiresolution Gaussian-mixture-model guided by saliency enhancement," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1796–1809, 2021.

[11] Y. LeCun et al., "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.

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

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

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

[15] X. Tang et al., "An unsupervised remote sensing change detection method based on multiscale graph convolutional network and metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5609715.

[16] 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, 2021.

[17] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.

[18] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[19] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.

[20] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. San Diego, CA, USA, Jun. 2005, pp. 539–546.

[21] C. Wu, H. Chen, B. Du, and L. Zhang, "Unsupervised change detection in multitemporal VHR images based on deep kernel PCA convolutional mapping network," *IEEE Trans. Cybern.*, vol. 52, no. 11, pp. 12084–12098, Nov. 2022.

[22] H. Lee et al., "Local similarity Siamese network for urban land change detection on remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4139–4149, 2021.

[23] Z. Lv, F. Wang, G. Cui, J. A. Benediktsson, T. Lei, and W. Sun, "Spatial–spectral attention network guided with change magnitude image for land cover change detection using remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4412712.

[24] 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.

[25] H. Chen and Z. Shi, "A spatial–temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sens.*, vol. 12, no. 10, p. 1662, May 2020.

[26] H. Chen, C. Wu, B. Du, L. Zhang, and L. Wang, "Change detection in multisource VHR images via deep Siamese convolutional multiple-layers recurrent neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2848–2864, Jul. 2020.

[27] T. Lei et al., "Difference enhancement and spatial–spectral nonlocal network for change detection in VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4507013.

[28] Q. Shi, M. Liu, S. Li, X. Liu, F. Wang, and L. Zhang, "A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5604816.

[29] T. Lei, Y. Zhang, Z. Lv, S. Li, S. Liu, and A. K. Nandi, "Landslide inventory mapping from bitemporal images using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 982–986, Jun. 2019.

[30] F. Shen, Y. Wang, and C. Liu, "Change detection in SAR images based on improved non-subsampled shearlet transform and multi-scale feature fusion CNN," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 12174–12186, 2021.

[31] X. Hou, Y. Bai, Y. Li, C. Shang, and Q. Shen, "High-resolution triplet network with dynamic multiscale feature for change detection on satellite images," *ISPRS J. Photogramm. Remote Sens.*, vol. 177, pp. 103–115, Jul. 2021.

[32] X. Ding, Y. Guo, G. Ding, and J. Han, "ACNet: Strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1911–1920.

[33] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learn. Represent (ICLR)*, May 2016, pp. 1–13.

[34] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2017.

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

[36] P. Chen, B. Zhang, D. Hong, Z. Chen, X. Yang, and B. Li, "FCCDN: Feature constraint network for VHR image change detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 187, pp. 101–119, May 2022.

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

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

[39] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 510–519.

[40] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11534–11542.

[41] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, and X. Zhai, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021, pp. 1–22.

[42] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jegou, "Training data-efficient image transformers & distillation through attention," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2021, pp. 10347–10357.

[43] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.

[44] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5607514.

[45] W. G. C. Bandara and V. M. Patel, "A transformer-based Siamese network for change detection," 2022, *arXiv:2201.01293*.

[46] Y. Feng, H. Xu, J. Jiang, H. Liu, and J. Zheng, "ICIF-Net: Intra-scale cross-interaction and inter-scale feature fusion network for bitemporal remote sensing images change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4410213.

[47] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1580–1589.

[48] Q. Zhang et al., "Split to be slim: An overlooked redundancy in vanilla convolution," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, Jul. 2020, pp. 3195–3201.

[49] C. Zhang, Y. Xu, and Y. Shen, "CompConv: A compact convolution module for efficient feature learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 3012–3021.

[50] J. Fu et al., "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3146–3154.

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

[52] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[53] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[54] 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, 2022, Art. no. 8007805.

[55] M. Zhang and W. Shi, "A feature difference convolutional neural network-based change detection method," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7232–7246, Apr. 2020.

[56] E. A. Nadaraya, "On estimating regression," *Theory Probab. Appl.*, vol. 9, no. 1, pp. 141–142, 1964.

[57] G. S. Watson, "Smooth regression analysis," *Sankhyā, Indian J. Statist., Ser. A*, vol. 26, no. 4, pp. 359–372, 1964.

[58] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[59] C. Zhang, L. Wang, S. Cheng, and Y. Li, "SwinSUNet: Pure transformer network for remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5224713.

[60] X. Zhang, W. Yu, and M.-O. Pun, "Multilevel deformable attention-aggregated networks for change detection in bitemporal remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5621518.

[61] R. Wang, F. Ding, J.-W. Chen, L. Jiao, and L. Wang, "A lightweight convolutional neural network for bitemporal image change detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2020, pp. 2551–2554.

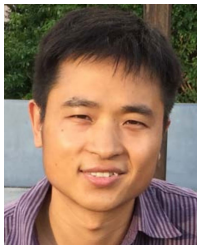
[62] M. Han, R. Li, and C. Zhang, "LWCDNet: A lightweight fully convolution network for change detection in optical remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 5621518.

[63] Y. Dai, T. Zheng, C. Xue, and L. Zhou, "MViT-PCD: A lightweight ViT-based network for Martian surface topographic change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2023.3234645.

[64] R. Liu, L. Mi, and Z. Chen, "AFNet: Adaptive fusion network for remote sensing image semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7871–7886, Sep. 2021.

[65] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 801–818.

[66] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.



Tao Lei (Senior Member, IEEE) received the Ph.D. degree in information and communication engineering from Northwestern Polytechnical University, Xi'an, China, in 2011.

From 2012 to 2014, he was a Post-Doctoral Research Fellow with the School of Electronics and Information, Northwestern Polytechnical University. From 2015 to 2016, he was a Visiting Scholar with the Quantum Computation and Intelligent Systems Group, University of Technology Sydney, Sydney, NSW, Australia. He is currently a Professor with the School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an. He has authored or coauthored 80+ research articles, including the IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP), IEEE TRANSACTIONS ON FUZZY SYSTEMS (TFS), and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS). His research interests include image processing, pattern recognition, and machine learning.

Xinzhe Geng received the B.S. degree from the Shaanxi University of Science and Technology, Xi'an, China, in 2020, where he is going to pursue the M.S. degree with the School of Electronic Information and Artificial Intelligence.

His research interests include image processing and pattern recognition.

Hailong Ning received the Ph.D. degree in signal and information processing from the University of Chinese Academy of Sciences (UCAS), Beijing, China, in 2021.

He is currently an Associate Professor with the School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an, China. His main research interests include pattern recognition, machine learning, computer vision, and multimodal learning.

Zhiyong Lv (Senior Member, IEEE) received the M.S. and Ph.D. degrees from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2008 and 2014, respectively.

He was an Engineer of surveying and worked at the First Institute of Photogrammetry and Remote Sensing, Xi'an, China, from 2008 to 2011. He is currently with the School of Computer Science and Engineering, Xi'an University of Technology, Xi'an. His research interests include multihyperspectral and high-resolution remotely sensed image processing, spatial feature extraction, neural networks, pattern recognition, deep learning, and remote sensing applications.

Maoguo Gong (Senior Member, IEEE) received the B.S. degree (Hons.) in electronic engineering and the Ph.D. degree in electronic science and technology from Xidian University, Xi'an, China, in 2003 and 2009, respectively.

Since 2006, he has been a Teacher with Xidian University, where he was promoted to an Associate Professor and a Full Professor in 2008 and 2010, respectively, with exceptive admission. His research interests include computational intelligence with applications to optimization, learning, data mining, and image understanding.

Dr. Gong was a recipient of the Prestigious National Program for the support of Top-Notch Young Professionals from the Central Organization Department of China, the Excellent Young Scientist Foundation from the National Natural Science Foundation of China, and the New Century Excellent Talent in University from the Ministry of Education of China. He is also an Associate Editor of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION and the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.

Yaochu Jin (Fellow, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in automatic control from Zhejiang University, Hangzhou, China, in 1988, 1991, and 1996, respectively, and the Dr.-Ing. degree in neuroinformatics from Ruhr University Bochum, Bochum, Germany, in 2001.

He was a "Finland Distinguished Professor" with the University of Jyväskylä, Jyväskylä, Finland; a "Changjiang Distinguished Visiting Professor" with Northeastern University, Shenyang, China; and a "Distinguished Visiting Scholar" with the University of Technology Sydney, Sydney, NSW, Australia. He is currently an Alexander von Humboldt Professor of artificial intelligence at the Chair of Nature Inspired Computing and Engineering, Faculty of Technology, Bielefeld University, Bielefeld, Germany. He is also a Distinguished Chair and a Professor in computational intelligence with the Department of Computer Science, University of Surrey, Guildford, U.K. His main research interests include evolutionary optimization, evolutionary learning, trustworthy machine learning, and evolutionary developmental systems.

Dr. Jin is a member of the Academia Europaea. He was a recipient of the 2018 and 2021 IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION Outstanding Paper Award, and the 2015, 2017, and 2020 *IEEE Computational Intelligence Magazine* Outstanding Paper Award. He is currently the Editor-in-Chief of *Complex and Intelligent Systems*. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS from 2016 to 2021, an IEEE Distinguished Lecturer from 2013 to 2015 and from 2017 to 2019, and the Vice President for Technical Activities of the IEEE Computational Intelligence Society from 2015 to 2016. He was named by the Web of Science as a Highly Cited Researcher from 2019 to 2021 consecutively.

Asoke K. Nandi (Life Fellow, IEEE) received the Ph.D. degree in physics from the University of Cambridge (Trinity College), Cambridge, U.K., in 1978.

He held academic positions in several universities, including the University of Oxford, Oxford, U.K.; Imperial College London, London, U.K.; the University of Strathclyde, Glasgow, U.K.; and the University of Liverpool, Liverpool, U.K.; and a Finland Distinguished Professorship at the University of Jyväskylä, Jyväskylä, Finland. In 2013, he moved to Brunel University London, Uxbridge, U.K., to become the Chair and the Head of electronic and computer engineering. He is currently a Distinguished Visiting Professor with Xi'an Jiaotong University, Xi'an, China. In 1983, he codiscovered the three fundamental particles known as W^+, W^-, and Z^0 (by the UA1 Team at CERN), providing the evidence for the unification of the electromagnetic and weak forces, for which the Nobel Committee for Physics in 1984 awarded the prize to his two team leaders for their decisive contributions. He has authored over 600 technical publications, including 270 journal articles as well as five books, entitled *Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines* (Wiley, 2020), *Automatic Modulation Classification: Principles, Algorithms and Applications* (Wiley, 2015), *Integrative Cluster Analysis in Bioinformatics* (Wiley, 2015), *Blind Estimation Using Higher-Order Statistics* (Springer, 1999), and *Automatic Modulation Recognition of Communications Signals* (Springer, 1996). His research interests include signal processing and machine learning, with applications to communications, image segmentations, and biomedical data. He has made many fundamental theoretical and algorithmic contributions to many aspects of signal processing and machine learning. He has much expertise in "big and heterogeneous data," dealing with modeling, classification, estimation, and prediction. The H-index of his publications is 80 (Google Scholar) and his ERDOS number is 2.

Dr. Nandi is a fellow of the Royal Academy of Engineering, U.K., and a fellow of seven other institutions, including the IEEE and the IET. Among the many awards he received are the Institute of Electrical and Electronics Engineers (USA) Heinrich Hertz Award in 2012, the Glory of Bengal Award for his outstanding achievements in scientific research in 2010, the Water Arbitration Prize of the Institution of Mechanical Engineers, U.K., in 1999, and the Mountbatten Premium, Division Award of the Electronics and Communications Division, the Institution of Electrical Engineers, U.K., in 1998. He is an IEEE EMBS Distinguished Lecturer from 2018 to 2019.