

Synthetic Aperture Radar Image Change Detection via Siamese Adaptive Fusion Network

Yunhao Gao, Feng Gao , Junyu Dong , Qian Du , *Fellow, IEEE*, and Heng-Chao Li , *Senior Member, IEEE*

Abstract—Synthetic aperture radar (SAR) image change detection is a critical yet challenging task in the field of remote sensing image analysis. The task is nontrivial due to the following challenges: First, intrinsic speckle noise of SAR images inevitably degrades the neural network because of error gradient accumulation. Furthermore, the correlation among various levels or scales of feature maps is difficult to be achieved through summation or concatenation. Toward this end, we proposed a siamese adaptive fusion (AF) network for SAR image change detection. To be more specific, two-branch CNN is utilized to extract high-level semantic features of multitemporal SAR images. Besides, an AF module is designed to adaptively combine multiscale responses in convolutional layers. Therefore, the complementary information is exploited, and feature learning in change detection is further improved. Moreover, a correlation layer is designed to further explore the correlation between multitemporal images. Thereafter, robust feature representation is utilized for classification through a fully connected layer with softmax. Experimental results on four real SAR datasets demonstrate that the proposed method exhibits superior performance against several state-of-the-art methods. Our codes are available at https://github.com/summitgao/SAR_CD_SAFNet.

Index Terms—Attention mechanism, change detection, deep learning, siamese adaptive fusion network (SAFNet), synthetic aperture radar (SAR).

I. INTRODUCTION

REMOTE sensing image change detection aims to find the changed information between two multitemporal images acquired in the same area at different times [1]. It provides valuable information for many applications, such as target detection [2], natural resource supervision [3], and agricultural development [4]. When a natural disaster suddenly occurs, a robust change detection algorithm can efficiently detect subtle

changes, and corresponding measures can be quickly taken by local governments to reduce the loss of life and property [5], [6]. Therefore, change detection has attracted extensive research attention.

Synthetic aperture radar (SAR) images are produced by an active system that sends a signal to the ground, and then receives the reflected signal. Different objects exhibit different characteristics between the scattering and polarized signals, which is beneficial to further accurate interpretation [7]. The SAR sensor has an all-weather and all-time imaging capability. It can penetrate smoke, cloud, and haze to acquire high-quality images [8]. Especially when a natural disaster occurs, SAR images make up the shortcomings of other data sources, such as LiDAR, optical, and multispectral data. Therefore, SAR imagery is a very useful data source for change detection.

It is worth mentioning that some pioneer efforts have been devoted to tackling the SAR change detection task. These methods can be classified into two broad categories: Supervised and unsupervised approaches [9]. The supervised approaches usually achieve better performance than the unsupervised ones by learning from labeled samples. However, it is generally difficult to collect high-quality labeled samples or acquire prior knowledge of the region to be studied. Therefore, the supervised approaches are commonly combined with an unsupervised one. The unsupervised approaches are employed for reliable sample generation, and the tedious task of manually labeling samples is eliminated.

In this article, we focus on unsupervised SAR image change detection. Due to speckle noise, it is very challenging to identify changes accurately in SAR images. To overcome the challenges, researchers designed a framework with three steps: Image preprocessing, difference image (DI) generation, and DI classification [10]. The first step involves denoising and coregistration. Denoising reduces speckle noise to some extent, but it may cause undesired degradation of the geometric details. Coregistration with subpixel accuracy is critical to generate a robust DI. In the second step, the ratio method [11] is commonly used to generate a DI. Many efficient operators are proposed, such as the log-ratio operator [12], Gauss-ratio operator [13], and neighborhood-based ratio operator [14]. In the final step of change detection, thresholding, maximum expectation, and clustering methods are generally involved for classification. In [15], fuzzy local information c-means clustering algorithm was proposed to provide robustness to noisy images. In [16], changed and unchanged pixels were clustered by the fuzzy c-means (FCM) based on Markov random field energy function.

Manuscript received June 25, 2021; revised September 15, 2021; accepted October 8, 2021. Date of publication October 15, 2021; date of current version November 2, 2021. This work was supported in part by the National Key Research, and Development Program of China under Grant 2018AAA0100602, in part by the National Natural Science Foundation of China under Grant U1706218 and Grant 61871335, and in part by the Key Research, and Development Program of Shandong Province under Grant 2019GHY112048. (*Corresponding author: Feng Gao.*)

Yunhao Gao, Feng Gao, and Junyu Dong are with the Qingdao Key Laboratory of Mixed Reality, and Virtual Ocean, School of Information Science, and Engineering, Ocean University of China, Qingdao 266100, China (e-mail: 914283361@qq.com; gaofeng@ouc.edu.cn; dongjunyu@ouc.edu.cn).

Qian Du is with the Department of Electrical, and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA (e-mail: du@ece.msstate.edu).

Heng-Chao Li is with the Sichuan Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University, Chengdu 610031, China (e-mail: hcli@swjtu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3120381

Gao *et al.* [17] proposed the PCANet to further classify the preclassification results. That achieves excellent performance in speckle noise suppression. In addition, many advanced methods have been employed for change detection, such as extreme learning machine (ELM) [18] and support vector machines [19]. However, it is difficult to capture high-level semantic features for change detection automatically and effectively.

Recently, deep learning-based methods have become increasingly popular and shown their superiority in remote sensing communities [20]. Many attempts in remote sensing image analysis based on deep learning methods have been motivated by these successful applications [21]–[24]. Some researchers make efforts to design deep models for the change detection task. In these models, unsupervised clustering methods are first employed to preclassify the input SAR images, and training samples are selected from reliable labels based upon preclassification. These samples are considered as prior knowledge and fed into a deep model for training [25]. Finally, the model forms its interpretation of the input image, and then the final change map is obtained. Gong *et al.* [26] first presented a SAR change detection method based on deep learning. The stacked restricted Boltzmann machines are utilized for feature extraction. Liu *et al.* [27] presented a change detection framework based on a convolutional coupling network. The network is symmetric with each side consisting of one convolutional layer and several coupling layers. Some recent breakthroughs in change detection were achieved by the convolutional neural network (CNN). Mou *et al.* [28] proposed a change detection framework that combines CNN and recurrent neural network into an end-to-end network. Gao *et al.* [29] detected changed information from sea ice SAR images by transferred deep learning. In [30], a general end-to-end 2D CNN named GETNET was designed for hyperspectral image change detection. Liu *et al.* [31] presented a local restricted CNN framework for SAR change detection, in which the original CNN is improved with a local spatial constraint. In [32], a noise modeling-based unsupervised fully convolutional network framework was presented for HSI change detection, which was proved with powerful learning features. However, a single-branch neural network inevitably results in the accumulation of error gradients. In other words, unstable feature representation limits its performance.

It is nontrivial to build an effective SAR change detection model, due to the following two challenges.

1) *Unstable Feature Representation*: Intrinsic speckle noise of SAR image inevitably degrades the neural network because of error gradient accumulation. It affects the performance of change detection to some extent, which results in an unstable network. Moreover, it is difficult to make full use of complementary information between different levels of features. Therefore, the extracted features may not well describe the changed information between multitemporal SAR images.

2) *Insufficient Feature Correlation*: Feature representations from two images are employed to represent the changed information, but the correlation may not be well explored through fusion operations such as summation and concatenation.

To solve the above challenges, we establish a deep siamese adaptive fusion network (SAFNet) for SAR image change

detection, which exploits stable feature representation based on siamese architecture. Specifically, we employ two-branch networks to exploit the high-level semantic features, which are rarely considered in SAR image change detection tasks. Furthermore, a correlation layer is designed to integrate features from two-branch networks. The attention mechanism is introduced to adaptively choose features among different scales. Ultimately, we introduce conditionally parameterized convolutions to enhance feature representation. Extensive experiments on four real SAR datasets demonstrate the superiority of the proposed SAFNet over state-of-the-art methods. Meanwhile, we have released our codes to facilitate other researchers.

The main contributions of the proposed SAFNet are summarized as follows.

- 1) We explore the SAR change detection task via a well-designed siamese neural network. The two branch network independently extracts the features of multitemporal images, and then the discrimination of features is improved through similarity measurement.
- 2) For stable feature representation, an AF module is utilized to combine the outputs of different layers. Since features from different layers contain complementary information, the attention-based fusion mechanism is introduced to use such information to improve feature representation. Therefore, the multiscale responses from convolutional layers are adaptively fused.
- 3) To avoid the loss of correlation caused by traditional feature integration methods, such as summation and concatenation, a correlation layer is designed for feature integration.

The rest of this article is organized as follows. In Section II, the proposed method is described in detail. Section III presents the experimental results on real multitemporal SAR images to validate the proposed method. Finally, Section IV concludes this article.

II. METHODOLOGY

Given two coregistered SAR images I_1 and I_2 captured at different periods, we aim to generate a change map that represents the changed information between two images. The proposed change detection method is comprised of two steps: First, high-level semantic features of multitemporal images are extracted through two-branch networks, and a similarity measure is used to optimize the process of feature extraction. Second, a correlation layer is employed to integrate the features for classification and change map generation.

A. Feature Extraction and Reliable Sample Generation

In this work, spatial neighborhood information of each pixel is analyzed for label prediction, thus detecting the changed region (the pixel with label “1”) in the image. Given two multitemporal SAR images I_1 and I_2 , we first obtain the pixelwise patch-pairs $(x_i^{t_1}, x_i^{t_2}, y_i)$, where $x_i^{t_1} \in \mathbb{R}^{r \times r}$ and $x_i^{t_2} \in \mathbb{R}^{r \times r}$ are the image patches centered the i th pixel in I_1 , and y_i is the ground truth label. Therefore, the spatial information of the i th pixel is fed into the CNN for feature extraction.

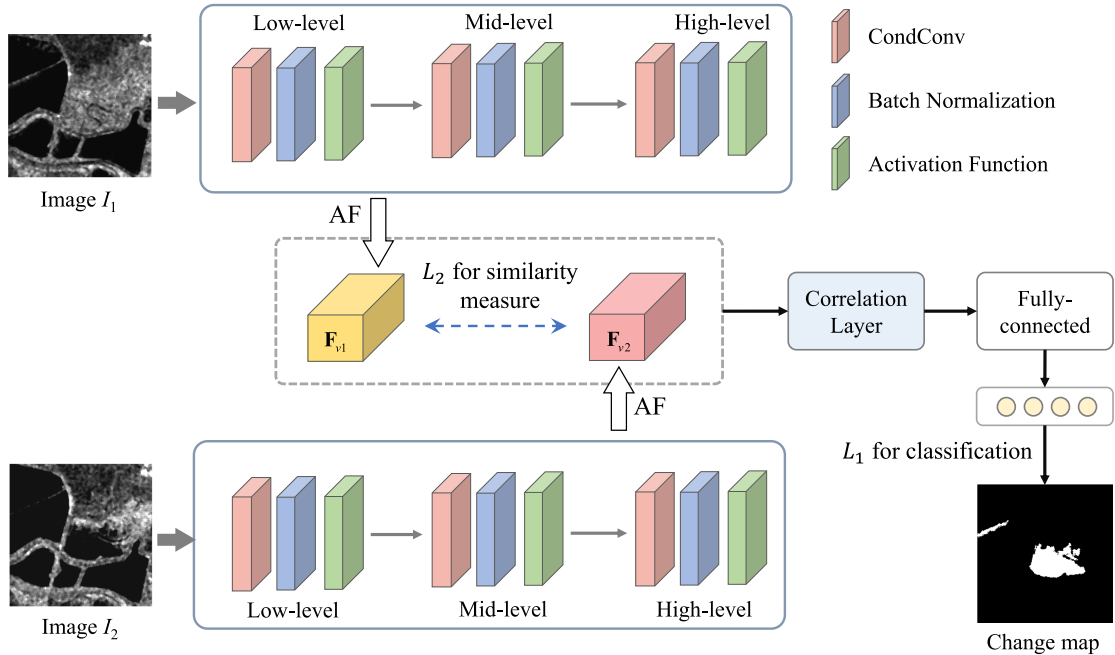


Fig. 1. Framework of the proposed SAFNet. Low-level, mid-level, and high-level CondConv blocks are utilized to extract three levels of features. Then fuse them by AF module. Finally, the final label prediction is achieved by the correlation layer and fully connected layer.

The traditional change detection model usually analyzes the DI generated by log-ratio or neighborhood ratio. Therefore, change detection results depend heavily on the quality of the DI. Especially, the performance always deteriorates when SAR images are seriously disturbed by speckle noise. In this article, feature representation is directly extracted from the original SAR images by a two-branch CNN model, which is less sensitive to speckle noise. Therefore, stable feature representation can be achieved. The proposed SAFNet is comprised of two branches (S_1, S_2) that accept patch-pairs as input. The feature extraction step is expressed as

$$\mathcal{F}_{\text{merge}}^i = \mathcal{C}(\mathcal{F}_1^i, \mathcal{F}_2^i) = \mathcal{C}(S_1(x_i^{t_1}), S_2(x_i^{t_2})) \quad (1)$$

where \mathcal{F}_1^i denote the features of image patch $x_i^{t_1}$ extracted by S_1 , \mathcal{F}_2^i denote the features of image patch $x_i^{t_2}$ extracted by S_2 , and $\mathcal{C}(\cdot)$ denotes the correlation analysis. It is implemented by a group convolution operation, where \mathcal{F}_1^i and \mathcal{F}_2^i denote the input and kernels, respectively. Then, the final fused features $\mathcal{F}_{\text{merge}}^i$ are obtained for classification by a fully connected layer with softmax. In the process of network training, similarity measure, and classification loss are applied to optimize the SAFNet. The pseudolabel samples are generated by the FCM algorithm in an unsupervised manner [18]. We randomly select a certain proportion of samples from the pseudolabel set for training, and the rest for testing. The details of pseudolabel sample generation are described as follows.

- 1) The FCM algorithm is performed on DI to generate the changed and unchanged clusters: Ω_c^1 and Ω_u^1 . Here, the number of changed clusters is T_c^1 . The upper limit of the change class is no more than $T_c^1 \cdot \theta$, $\theta = 1.2$.

- 2) The FCM algorithm is reperformed on DI to generate five clusters: $\Omega_1^2, \Omega_2^2, \Omega_3^2, \Omega_4^2, \Omega_5^2$. Ω_1^2 has the higher probability of changed clusters, and so on. The number of clusters is $T_1^2, T_2^2, T_3^2, T_4^2, T_5^2$. The pixels from Ω_1^2 is assigned to Ω_c .
- 3) The clusters are assigned to Ω_i when $\sum_{i=1} T_i^2 < T_c^1 \cdot \theta$, and the rest clusters are denoted as Ω_u . Thus, the preclassification map consisted of $[\Omega_c, \Omega_i, \Omega_u]$ is generated.

B. Siamese Adaptive Fusion Network

In this article, we propose SAFNet to compare image patches from SAR images, and it achieves robust feature discrimination power through AF module and correlation layer. The framework of the proposed SAFNet is illustrated in Fig. 1.

Image patches centered at selected sample pixels are extracted from I_1 and I_2 , respectively. Two groups of patch-pairs are treated as the input of two branches. The network optimization in SAFNet is achieved by weight sharing and residual learning. Each branch of the SAFNet is comprised of three CondConv blocks and the AF module, which are merged by the correlation layer and fully connected layer. In the following, we will describe the SAFNet in detail.

1) *CondConv Blocks*: Although feature extraction is greatly improved by aggregating multiple convolutional branches, the computational cost increases dramatically [33]. To address the problem, conditional computation was performed by activating only a portion of the entire network [34]. In [35], efficient inference was performed by conditionally parameterized convolutions.

The structure of a typical conditionally parameterized convolutions (CondConv) block is shown in Fig. 2, which generates

TABLE I
DETAILS OF THE PROPOSED SAFNET

Layer	Type	Kernel Size	Stride	Padding	Output	Output Size
CondConvBlock1 (Low-level)	Convolution	3×3	1	Yes	16	28×28×16
Conv1 (Low-level)	Convolution	1×1	2	No	32	14×14×32
CondConvBlock2 (Mid-level)	Convolution	3×3	1	Yes	32	14×14×32
Conv2	Convolution	1×1	2	No	64	7×7×64
CondConvBlock3 (High-level)	Convolution	3×3	1	Yes	64	7×7×64
Correlation layer	Convolution	-	-	-	64	1×1×64
Fully Connected	FC	-	-	-	2	-

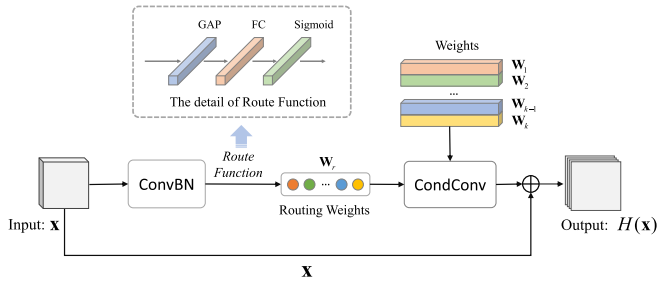


Fig. 2. Detailed components of the CondConv block. ConvBN denotes three operations: Convolution, batch normalization, and activation layer. The conditionally parameterized convolutions (CondConv) are achieved by combining initialization weight and routing weight from route function. Therefore, the ability of feature extraction is significant with one kernel computation.

several groups of convolution kernels through routing function and initialization weights, and the details of CondConv are described in [35]. ConvBN is the combination of three operations: Convolution, batch normalization, and activation layer.

The routing weights \mathbf{W}_r generated by routing function is calculated as

$$\mathbf{W}_r = R(\mathbf{F}_{c1}) = \delta(\text{GAP}(\mathbf{F}_{c1}) * \mathbf{W}_f) \quad (2)$$

where GAP denotes the global average pooling, δ is the sigmoid function, and $*$ is the convolution operation. \mathbf{W}_f denotes the matrix of fully connected layer, which mapping the global feature to k routing weights. Therefore, the output of CondConv through residual learning [36] is defined as

$$H(x) = \sigma((\alpha_1 \cdot \mathbf{W}_1 + \dots + \alpha_k \cdot \mathbf{W}_k) * \mathbf{F}_{c1}) + x \quad (3)$$

where $[\alpha_1, \dots, \alpha_k]$ are the weights from the routing weights \mathbf{W}_r . Consequently, the CondConv is mathematically equivalent to a linear mixture of multiple convolutions.

2) *Feature Fusion by Adaptive Fusion Module*: Feature fusion exhibits good performance in many image classification and object detection tasks. Feature fusion is considered as an effective method for complementarity, however, the redundant features are introduced which degrade the discrimination. In the proposed SAFNet, an adaptive feature fusion mechanism is introduced to extract the complementary information among different CondConv blocks. As illustrated in Fig. 1, multiple

CondConv blocks are used to capture features of different levels, including low-level, mid-level, and high-level features.

The important part of each layer of features is given higher weight through a nonmutually exclusive attention vector, so the discrimination of the final fused features is stronger.

The output features of the three levels of CondConv blocks (i.e., CondConvBlock1, CondConvBlock2, and CondConvBlock3 in Table I) are denoted as \mathbf{F}_1 , \mathbf{F}_2 , and \mathbf{F}_3 , respectively. One CondConv block is employed to extract the features at each level. In this article, \mathbf{F}_1 contains 16 feature maps, \mathbf{F}_2 contains 32 feature maps, and \mathbf{F}_3 contains 64 feature maps. Since \mathbf{F}_1 , \mathbf{F}_2 , and \mathbf{F}_3 contain different number of feature maps, dimension matching is essential for feature fusion. To achieve this, 64 kernels of size 1×1 with different strides are employed to convolute \mathbf{F}_1 , \mathbf{F}_2 , and \mathbf{F}_3 . After such convolution, the feature maps from three levels all become 64 dimensions with the same spatial size for fusion.

Recently, computer vision adaptively encodes informative context from a long-range region to select more critical information for the current task. Hu *et al.* [37] proposed the squeeze-and-excitation network (SENet) to explore the channel relationship of features. In [38], selective kernel network (SKNet) adaptively selects the kernel size through attention mechanism. Inspired by these methods, the AF module is designed to merge feature maps from different levels, where an attention mechanism is introduced to emphasize important features while suppressing unnecessary ones. As illustrated in Fig. 3, three feature maps \mathbf{F}_1 , \mathbf{F}_2 , and \mathbf{F}_3 are obtained from multiple CondConv blocks. Recognizing that not all the features are essential for the final classification, the attention mechanism is introduced to adaptively choose features from suitable scales. First, the input feature maps are fused by elementwise summation after dimension matching as

$$\mathbf{F} = D(\mathbf{F}_1) + D(\mathbf{F}_2) + D(\mathbf{F}_3) \quad (4)$$

where $\mathbf{F} \in \mathbb{R}^{w \times w \times c}$ is the fused feature. $D(\cdot)$ is the function of dimension matching. Then GAP is employed to capture the global information of \mathbf{F}_s as follows:

$$\mathbf{F}_s = \frac{1}{w \times w} \sum_{i=1}^w \sum_{j=1}^w \mathbf{F}(i, j). \quad (5)$$

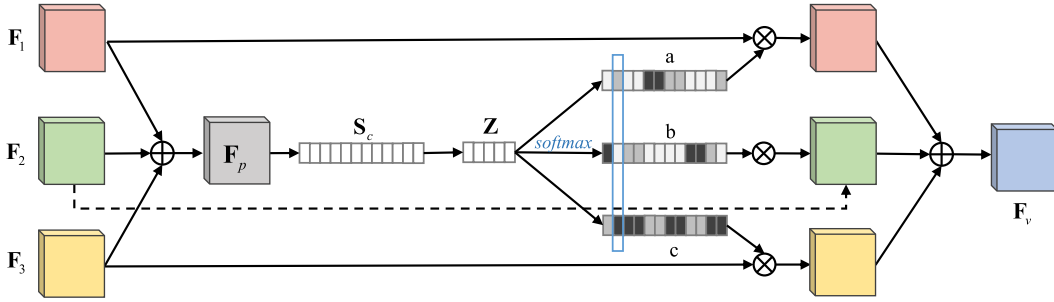


Fig. 3. Detail components of the adaptive fusion module.

Then, information is aggregated in $\mathbf{F}_s \in \mathbb{R}^{c \times 1}$. After that, \mathbf{F}_s is further squeezed into a compact feature $\mathbf{F}_z \in \mathbb{R}^{\frac{c}{\gamma} \times 1}$ by fully connected layer as

$$\mathbf{F}_z = \sigma(\mathbf{W} * \mathbf{F}_s) \quad (6)$$

where σ is the ReLU activation, and $\mathbf{W} \in \mathbb{R}^{\frac{c}{\gamma} \times c}$ is the weighting matrix ($\gamma = 8$ in this article). Therefore, the model complexity is reduced significantly.

Fully connected layer and soft attention (softmax layer) are used to adaptively select features from suitable scales, which is guided by the compact feature descriptor \mathbf{F}_z . Let $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^{c \times 1}$ represent the soft attention vector obtained by the softmax layer. Note that \mathbf{a}_i is the i th element of \mathbf{a} , likewise \mathbf{b}_i and \mathbf{c}_i . We can have $\mathbf{a}_i + \mathbf{b}_i + \mathbf{c}_i = 1$ owing to the intrinsic feature of the softmax layer. Finally, the feature map \mathbf{F}_v is obtained through the attention weights on various scales

$$\mathbf{F}_v = \mathbf{a} \cdot \mathbf{F}_1 + \mathbf{b} \cdot \mathbf{F}_2 + \mathbf{c} \cdot \mathbf{F}_3. \quad (7)$$

\mathbf{F}_v is utilized to generate the final features of one branch of the SAFNet which is denoted by feat.

3) *Correlation Layer*: The two branches of networks are employed to extract the semantic features from multitemporal SAR images, and then changed information prediction is achieved after feature integration. Generally speaking, it is easy to ignore the correlation through simple integration operations such as summation and concatenation. Therefore, a feature correlation operation is developed for feature integration. The features from AF modules are denoted as \mathbf{F}_{v1} and \mathbf{F}_{v2} , which can be further integrated through the correlation layer as shown in Fig. 4. The integrated features are computed by

$$\mathbf{F}_c = \mathbf{F}_{v1} * \mathbf{F}_{v2} \quad (8)$$

where $*$ denotes the convolution operation (conv), and $\mathbf{F}_c \in \mathbb{R}^{b \times 1 \times 1 \times c}$ is the final feature representation. It should be noted that the group convolution strategy is applied to the correlation layer. In other words, feature integration through convolution participates in the parameter optimization under the end-to-end training. For the correlation layer, $\mathbf{F}_{v1} \in \mathbb{R}^{b \times w \times w \times c}$ and $\mathbf{F}_{v2} \in \mathbb{R}^{b \times w \times w \times c}$ are first reshaped to $\mathbf{F}_{v1} \in \mathbb{R}^{1 \times k \times w \times w}$ and $\mathbf{F}_{v2} \in \mathbb{R}^{k \times 1 \times w \times w}$, respectively. Here, b is the batch size, and $k = b \times c$ denotes the number of kernels. It should be noted that the group convolution strategy is also applied, and the number of groups is k .

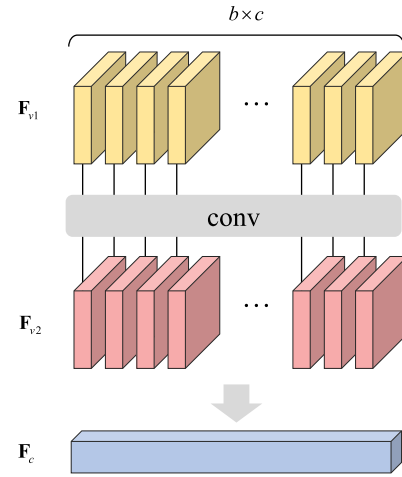


Fig. 4. Detailed components of the correlation layer.

C. SAFNet Optimization and Change Map Generation

Table I shows the implementation details of the proposed SAFNet (a typical branch). The input image patch is resized to 28×28 pixels. In each branch of the network, three levels of CondConv blocks are implemented. Conv1 and Conv2 are the transitional convolution operators among different levels of CondConv blocks. Each CondConv block contains two convolutions, as illustrated in Fig. 2. After feature extraction, the AF module is employed to emphasize meaningful features. As mentioned before, the SAFNet contains two branches of neural network. Therefore, the SAFNet generates a feature pair \mathbf{F}_{v1} and \mathbf{F}_{v2} . feat_0 and feat_1 are computed as

$$\text{feat}_0 = \mathbf{F}_{v1} * \mathbf{W}_{v1} \quad (9)$$

$$\text{feat}_1 = \mathbf{F}_{v2} * \mathbf{W}_{v2}. \quad (10)$$

After obtaining feat_0 and feat_1 , the Euclidean distance metric is employed to produce a value that is used to measure the feature similarity of two branches. To customize an appropriate metric where patch-pairs have stronger discrimination, the contrastive loss is employed. Contrastive loss is formulated as

$$L_2 = \begin{cases} D(\text{feat}_0, \text{feat}_1)^2 & y = 0 \\ \max(0, m - D(\text{feat}_0 - \text{feat}_1))^2 & y = 1 \end{cases} \quad (11)$$

where feat_0 and feat_1 denote the feature maps extracted from two branches of convolutional network, respectively. m is a margin, which is set as 1 in this article, and $D(\text{feat}_0 - \text{feat}_1)$ measures the distance between feat_0 and feat_1 using Euclidean distance. In this article, y is the ground truth label that measures the similarity of image pairs. $y = 0$ represents higher similarity, and there is no change in the patch-pairs. $y = 1$ indicates that the land cover is changed.

To obtain the final classification label in the change detection task, the cross-entropy loss is utilized to optimize the network. The features obtained by the AF module are integrated by correlation operation, and label prediction is achieved through the fully connected layer and softmax layer. Therefore, the loss value between label prediction and the ground truth is formulated as

$$L_1 = \sum_{i=1}^N [y_t \log(y) + (1 - y_t) \log(1 - y)] \quad (12)$$

where y_t is the ground truth while y is the label mapping from the last fully connected layer. L_1 is designed to supervise the learning process of the fused feature between multitemporal SAR images. The final loss value is represented as the combination of L_1 and L_2

$$L = L_1 + \lambda L_2 \quad (13)$$

where λ is the weight parameters for the contrastive loss L_2 . In the experiments, λ is empirically set to 0.5.

Image patch pairs centered at selected samples are used for parameter optimization of the SAFNet. The same as most CNN models, L is optimized using the backpropagation algorithm. Note that L_2 can also be considered as regularization terms for L_1 to standardize the training process. Finally, the trained SAFNet is utilized for prediction. Each test sample is assigned a label according to the results of SAFNet via a feed-forward propagation (“0” denotes unchanged and “1” denotes changed). Then, the final change map is obtained.

III. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, experiment data and evaluation criteria are introduced first. Then we will investigate the factors that may influence the performance of SAFNet. Finally, the effectiveness of SAFNet will be evaluated by comparison with several state-of-the-art methods.

A. Experiment Data and Evaluation Criteria

To verify the effectiveness of SAFNet for SAR image change detection, experiments are conducted on four real SAR datasets. The first dataset is the San Francisco dataset. It was acquired by the ERS-2 SAR sensor (the spatial resolution is 30 m) over the city of San Francisco. The size of the original image is 7749×7713 pixels. We select a typical region (256×256 pixels) to evaluate the proposed method. The images were captured in August 2003 and May 2004, respectively. The ground truth image was created by integrating prior information with photo interpretation. The dataset is shown in Fig. 5.

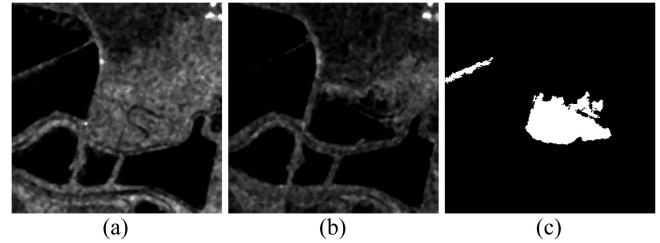


Fig. 5. San Francisco dataset. (a) Image acquired in August 2003. (b) Image acquired in May 2004. (c) Ground truth image.

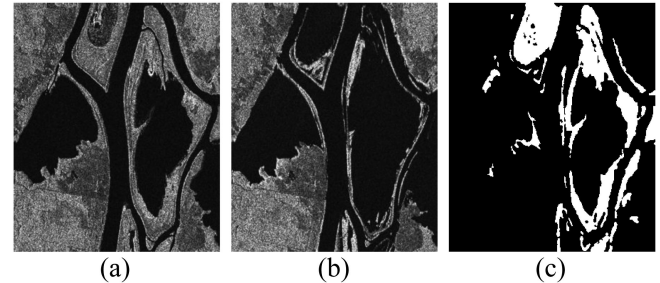


Fig. 6. Ottawa dataset. (a) Image acquired in May 1997. (b) Image acquired in August 1997. (c) Ground truth image.

The second dataset is the Ottawa dataset, as shown in Fig. 6. Two images were acquired by the Radarsat sensor in May 1997 and August 1997, respectively. The National Defense Research and Development Canada provides the dataset, and the dataset shows the changed information in areas affected by floods. These images were registered by the automatic registration algorithm from A.U.G. Signals Ltd. that is available through distributed computing at www.signalfusion.com. The size of Ottawa dataset is 290×350 pixels and the spatial resolution is 10 m. The available ground truth was created by integrating prior information and photo interpretation.

The third dataset is the Yellow River dataset. It was selected from two large SAR images collected in the Yellow River Estuary area of China. Both images were taken by Radarsat-2 (the spatial resolution is 8 m) in June 2008 and June 2009, respectively. The original size of the dataset is 7666×7692 pixels. The size is too large to describe details. Therefore, two typical regions (306×291 pixels and 289×257 pixels) were used in the experiment to verify the effectiveness of the proposed SAFNet. The changed regions showed newly cultivated farmland. The dataset is shown in Fig. 7. In particular, this dataset is polluted by noise with different characteristics. Specifically, one image in the dataset is a single-look image while the other is a four-look image. Therefore, it is very challenging to perform change detection on this dataset.

Both quantitative measures and qualitative analysis are performed on the four datasets. Specifically, the qualitative analysis is achieved by visual comparison between the change map generated by the proposed method and the ground truth image. Besides, false positives (FP), false negatives (FN), overall errors (OE), percentage correct classification (PCC), and

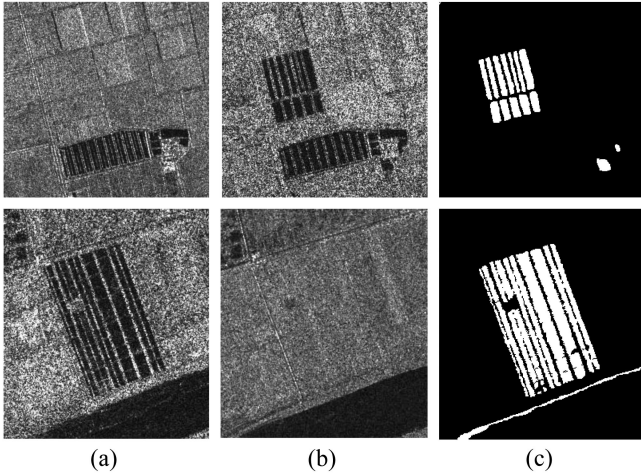


Fig. 7. Yellow River dataset. (a) Image acquired in June 2008. (b) Image acquired in June 2009. (c) Ground truth image.

Kappa coefficient (KC) are utilized as the quantitative measures to assess the change detection performance. FP denotes the number of unchanged pixels that are mistakenly detected as the changed ones. FN is the number of pixels belonging to the changed class but is incorrectly detected as the unchanged ones. OE denotes the total number of pixels that are incorrectly detected, i.e., the sum of FP and FN. The PCC is computed by

$$PCC = \frac{N_u + N_c - OE}{N_u + N_c} \times 100\% \quad (14)$$

where N_u denotes the total number of unchanged pixels in the ground truth image while N_c denotes the total number of changed pixels. The KC is computed by

$$KC = \frac{PCC - PRE}{1 - PRE} \times 100\% \quad (15)$$

where

$$PRE = \frac{(N_c + FP - FN) \times N_c + (N_u + FN - FP) \times N_u}{(N_u + N_c) \times (N_u + N_c)} \quad (16)$$

Here, we can observe that the value of KC is determined by FP and FN rather than by OE alone. Therefore, KC reflects the balance between FP and FN to some extent. More detailed information should be taken into account to obtain suitable KC values, and KC is a more persuasive indicator of the change detection result than PCC or OE.

B. Parameter Analysis

1) *Analysis of the Sample Image Size:* We first investigate the change detection performance by tuning the values of r , which denotes the size of the input patch-pairs employed in SAFNet. We set r to, 7, 9, 11, 13, and 15 to indicate the relationship between r and PCC on four datasets. As shown in Fig. 8, the PCC values do not perform well when $r = 7$. It is evident that deep learning methods can not extract robust features in such small image patches. On the Ottawa dataset, the proposed SAFNet obtains the best performance when $r = 9$.

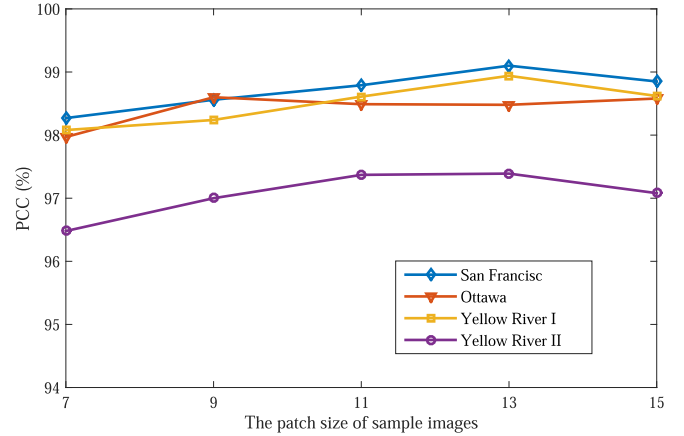


Fig. 8. Relationship between patch size r and PCC values.

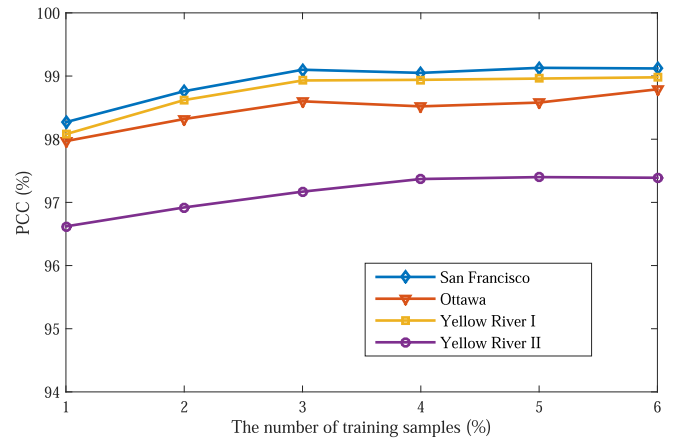


Fig. 9. Relationship between the number of training samples N_t and PCC values.

When $r > 9$, change detection results tend to deteriorate. On the San Francisco and the Yellow River I datasets, the SAFNet achieves the best PCC values when $r = 13$. When $r > 13$, PCC values tend to get worse. It indicates that it is difficult to describe the changed information of the center pixel by large image patches. Therefore, in our implementation, we set $r = 9$ and $r = 11$ on the Ottawa and Yellow River II datasets. On the San Francisco and Yellow River I datasets, $r = 13$ is employed as the best choice.

2) *Analysis of the Number of Training Samples:* Next, we discuss the relationship between the number of training samples N_t and the experimental results, because the training samples are essential. Training samples are randomly selected from pseudolabel set. $N_t = [1\%, 2\%, 3\%, 4\%, 5\%, 6\%]$. From Fig. 9, we can observe that it is difficult to obtain excellent classification performance with fewer training samples. Because insufficient training samples lead to overfitting. PCC values commonly decrease with the reduction of N_t , especially when $N_t < 3\%$ has a curve of obvious decline. When $N_t > 3\%$, PCC value tends to be stable. Considering that more training samples will affect the efficiency and generalization to a certain extent. Therefore,

TABLE II
RELATIONSHIP BETWEEN PCC AND THE REGULARIZATION FACTOR OF LOSS FUNCTIONS

Dataset	San Francisco	Ottawa	Yellow River I	Yellow River II
$\lambda=1$	98.21	97.68	96.85	96.74
$\lambda=0.5$	99.10	98.60	98.94	97.37
$\lambda=0.1$	98.69	98.26	98.49	96.91
$\lambda=0.01$	98.13	97.95	97.79	96.10

TABLE III
RELATIONSHIP AMONG AF MODULE, CORRELATION LAYER AND PCC VALUES

Method	San Francisco	Ottawa	Yellow River I	Yellow River II
Full model	99.10	98.60	98.94	97.37
w/o AF	99.01	98.02	98.45	97.04
w/o correlation	98.89	98.24	98.61	97.22
w/o AF and correlation	98.80	97.96	98.19	96.70

we choose $N_t = 4\%$ on the Yellow River dataset, and $N_t = 4\%$ on the San Francisco and Ottawa datasets.

3) *Analysis of the Hyperparameter*: In this article, the combination of similarity measure and classification loss is used to jointly optimize the model parameters. We use a hyperparameter to balance the similarity measure and classification prediction. Here, λ is set to $[0.01, 0.1, 0.5, 1]$. It is found from Table II that when $\lambda = 1$, it is difficult to focus on the model prediction, which is utilized for the measurement of final results. When $\lambda > 0.5$, the PCC values began to decrease gradually, due to the similarity measurement is reduced. It reflects that similarity measure improves discrimination to a certain extent. Therefore, $\lambda = 0.5$ is selected as the most appropriate hyperparameter.

4) *Ablation Experiment*: We present an ablation experiment in which importance of model components is evaluated. Table III shows the relationship among AF module, correlation layer, and PCC values. We can observe that the full model (SAFNet with AF module and correlation layer) achieves better performance. Specifically, there are 0.09%, 0.58%, 0.49%, and 0.33% improvements in PCC on four datasets compared with the full model excluding AF module, respectively. That is because multiscale responses from convolutional layers are adaptively combined by AF module. Therefore, it is shown that the AF module effectively alleviates the problem of unstable feature representation. In addition, an ablation experiment between the correlation layer and PCC values is exhibited. Table III reveals that the PCC value of the full model is higher than the full model without the correlation layer. That is because the features from two-branch networks further integrate for classification. To sum up, the proposed SAFNet benefits from the AF module and correlation analysis.

C. Experimental Results and Discussion

To verify the effectiveness of the proposed SAFNet, several existing change detection methods are used for comparison,

TABLE IV
CHANGE DETECTION RESULTS OF DIFFERENT METHODS ON THE SAN FRANCISCO DATASET

Method	FP	FN	OE	PCC(%)	KC(%)
PCAKM [39]	1618	25	1643	97.49	83.68
NBRELM [18]	393	362	755	98.85	91.35
GaborPCANet [17]	333	342	675	98.97	92.23
RMG-FDA [40]	267	428	695	98.94	91.88
ResNet [36]	104	938	1042	98.41	86.95
DBN [26]	204	478	682	98.96	91.94
DCNet [41]	276	468	744	98.86	91.28
ESNet [42]	554	290	844	98.71	90.54
Proposed SAFNet	212	377	589	99.10	93.12

i.e., PCAKM [39], NBRELM [18], GaborPCANet [17], RMG-FDA [40], ResNet [36], DBN [26], DCNet [41], and ESCNet [42]. PCAKM uses PCA filters for feature extraction, and then the extracted features are classified by k -means clustering. In NBRELM, the neighborhood-based ratio operator is adopted on DI generation and feature extraction. The extracted features are then classified by the ELM. GaborPCANet utilizes PCANet to extract discriminant features. PCANet is a simple deep learning network whose convolution filters are chosen from PCA filters. In RMG-FDA, the graph-based method is chosen as the classifier for changed pixel identification. In ResNet, ResNet-18 is employed for pixel-wise classification. DBN applies a deep belief network to complete the SAR change detection task. DCNet is a deep neural network, which cascades multiple channel-weighting-based residual blocks. In ESCNet, two weight-share superpixel sampling networks and a siamese neural network based on U-NET are utilized to mine the information between multitemporal images. Specifically, for the PCAKM, the size of neighborhood patch is set to be $S = 3$, and the number of principal components selected by PCA is set as $h = 5$. In GaborPCANet, $f = \sqrt{2}$, $V = 5$, $U = 8$, and $k_{\max} = 2\pi$ are utilized for feature extraction with PCANet. For the NBRELM and GaborPCANet, the size of neighborhood patch is set to be 7×7 . In RMG-FDA, the threshold for frequency-domain analysis is set as $\text{SmoothVal} = 0.1$ and $t_{\text{sal}} = 0.3$, and six graphs are used in the random multigraphs algorithm.

1) *Results on the San Francisco Dataset*: Fig. 10 provides the visual comparison of the change maps generated by different methods on the San Francisco dataset. The quantitative evaluation is given in Table IV. It can be observed that in the change maps generated by PCAKM, many unchanged pixels are falsely detected as changed ones. Therefore, the FP value of PCAKM is relatively high. RMG-FDA and ResNet achieve better performance with less noise, but many changed pixels are missed. For NBRELM and GaborPCANet, balanced FP and FN values are obtained. However, there is still some white noise in the change maps, and some changed regions are ignored in the upper left corner. More acceptable visualized results are shown in Fig. 10(f)–(i). Compared with the Fig. 10(f) and (g), DCNet obtains the details of the change regions, while DBN suppresses the noise. Furthermore, the proposed SAFNet achieves the best

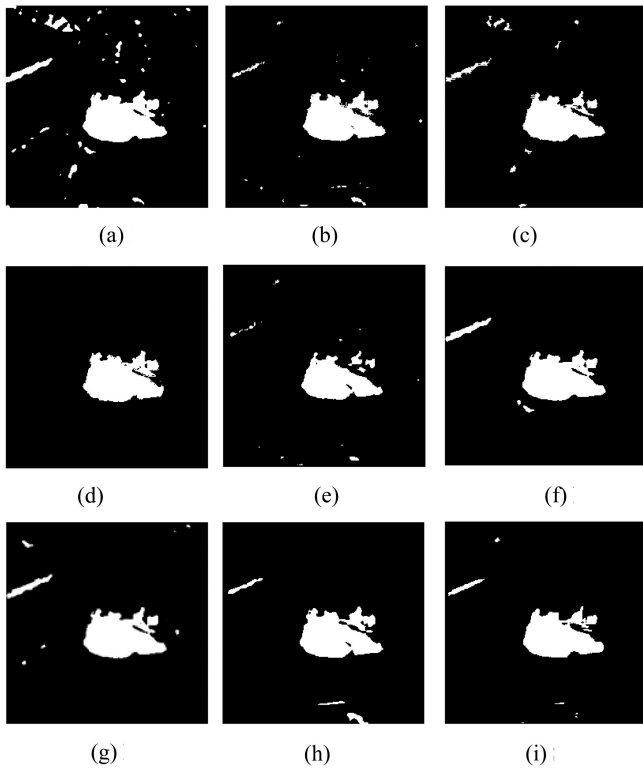


Fig. 10. Visualized results of change detection methods on the San Francisco dataset. (a) Result by PCAKM. (b) Result by NBRELM. (c) Result by GaborPCANet. (d) Result by RMG-FDA. (e) Result by ResNet. (f) Result by DBN. (g) Result by DCNet. (h) Result by the proposed ESCNet. (i) Result by the proposed SAFNet.

TABLE V
CHANGE DETECTION RESULTS OF DIFFERENT METHODS ON THE
OTTAWA DATASET

Method	FP	FN	OE	PCC(%)	KC(%)
PCAKM [39]	955	1515	2470	97.57	90.73
NBRELM [18]	963	1192	2155	97.88	91.98
GaborPCANet [17]	953	942	1895	98.13	92.99
RMG-FDA [40]	198	1883	2081	97.95	91.96
ResNet [36]	551	1623	2174	97.86	91.73
DBN [26]	995	704	1699	98.33	93.76
DCNet [41]	679	1051	1730	98.30	93.54
ESCNet [42]	10	2231	2241	97.79	86.06
Proposed SAFNet	882	534	1416	98.60	94.81

PCC and KC values. It is evident that the AF module and correlation analysis improve the performance of change detection.

2) *Results on the Ottawa Dataset:* Fig. 11 illustrates the change maps by different methods on the Ottawa dataset. Table V lists the detailed evaluation criteria. It can be observed that the PCAKM and RMG-FDA missed many small changed regions. Therefore, these methods suffer from high FN values. NBRELM, GaborPCANet, and the proposed SAFNet perform better. From Table V, we can observe that the SAFNet is superior to NBRELM, and is about 0.72% higher in PCC values.

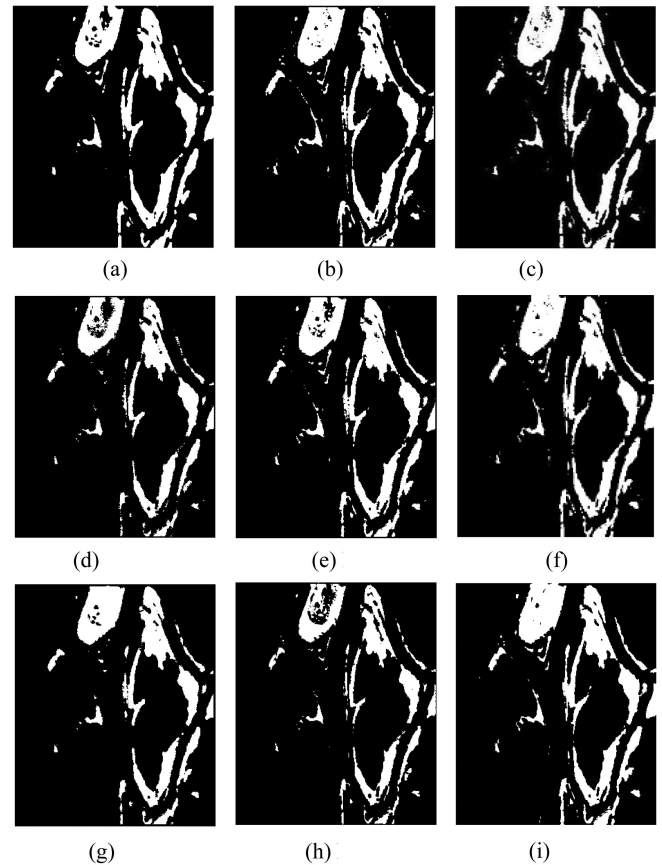


Fig. 11. Visualized results of change detection methods on the Ottawa dataset. (a) Result by PCAKM. (b) Result by NBRELM. (c) Result by GaborPCANet. (d) Result by RMG-FDA. (e) Result by ResNet. (f) Result by DBN. (g) Result by DCNet. (h) Result by the proposed ESCNet. (i) Result by the proposed SAFNet.

Meanwhile, the SAFNet surpasses GaborPCANet by 0.47% in PCC. In ESCNet's results, FP value is the lowest while FN value is high, because many subtle change regions are eliminated as noise. In addition, the PCC of SAFNet is slightly higher than DBN and DCNet. That is because SAFNet has better detail retention resulting in lower FN values. In general, the proposed SAFNet is predominant on the Ottawa dataset compared with some state-of-the-art methods.

3) *Results on the Yellow River Dataset:* The change maps generated by different methods on the Yellow River dataset is shown in Figs. 12 and 13, and the corresponding evaluation criteria are listed in Tables VI and VII. The Yellow River datasets are seriously interfered by speckle noise. Therefore, it is challenging for traditional techniques to obtain satisfying results. For Yellow River I dataset, the change maps of PCAKM and GaborPCANet exhibit many noise regions, and hence both methods suffer from very high FP values. For NBRELM, many changed pixels are missed, and therefore the FN value is relatively high. Despite the noise interference is suppressed, much important change information is ignored and results in a high FN value. Fig. 12(f)–(i) perform better, and it is evident that deep learning-based methods can explore contextual information more effectively. Especially, ESCNet and SAFNet suppress the noise at the bottom of the change map.

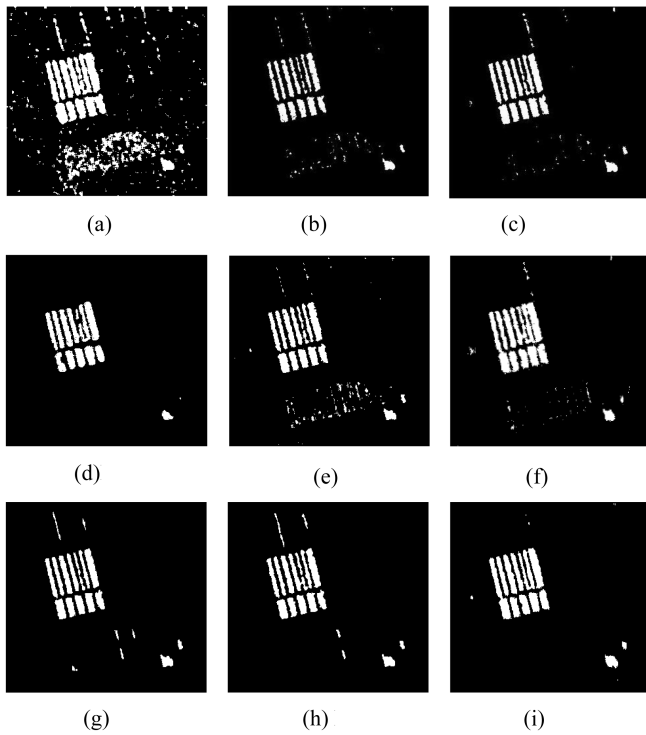


Fig. 12. Visualized results of change detection methods on the Yellow River I dataset. (a) Result by PCAKM. (b) Result by NBRELM. (c) Result by GaborPCANet. (d) Result by RMG-FDA. (e) Result by ResNet. (f) Result by DBN. (g) Result by DCNet. (h) Result by the proposed ESCNet. (i) Result by the proposed SAFNet.

TABLE VI
CHANGE DETECTION RESULTS OF DIFFERENT METHODS ON THE YELLOW RIVER I DATASET

Method	FP	FN	OE	PCC(%)	KC(%)
PCAKM [39]	5158	155	5313	94.03	62.92
NBRELM [18]	256	1794	2050	97.70	76.05
GaborPCANet [17]	745	1113	1858	97.91	80.63
RMG-FDA [40]	169	1614	1783	98.00	79.37
ResNet [36]	1267	457	1724	98.06	83.78
DBN [26]	561	668	1229	98.62	87.49
DCNet [41]	493	658	1151	98.71	88.33
ESCNet [42]	335	916	1251	98.60	86.70
Proposed SAFNet	359	582	941	98.94	90.32

For Yellow River II dataset, we can observe a large number of noise regions in Fig. 13(a)–(c). In addition, many changed information is missed at the bottom of changed regions. As a result, the change detection results are seriously deteriorated. For RMG-FDA, noise is effectively suppressed by frequency-domain analysis, but much changed change information is lost. In contrast, Fig. 13(f)–(h) generated by deep learning-based methods are more similar to ground truth image. However, the changed information is ignored on the left of the change map resulted by DCNet. Therefore, the FN values are relatively high. Generally speaking, the proposed SAFNet achieves the best

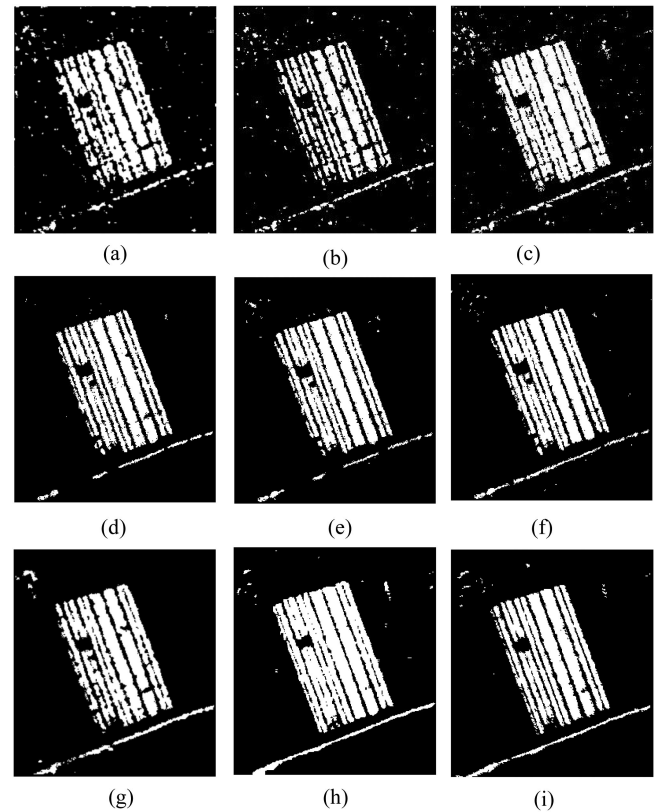


Fig. 13. Visualized results of change detection methods on the Yellow River II dataset. (a) Result by PCAKM. (b) Result by NBRELM. (c) Result by GaborPCANet. (d) Result by RMG-FDA. (e) Result by ResNet. (f) Result by DBN. (g) Result by DCNet. (h) Result by the proposed ESCNet. (i) Result by the proposed SAFNet.

TABLE VII
CHANGE DETECTION RESULTS OF DIFFERENT METHODS ON THE YELLOW RIVER II DATASET

Method	FP	FN	OE	PCC(%)	KC(%)
PCAKM [39]	2137	2663	4800	93.54	77.85
NBRELM [18]	1059	3777	4836	93.49	76.14
GaborPCANet [17]	2435	1533	3968	94.66	82.43
RMG-FDA [40]	380	1906	2286	96.92	89.13
ResNet [36]	1031	1870	2901	96.09	86.49
DBN [26]	861	1266	2127	97.14	90.22
DCNet [41]	934	1809	2743	96.31	87.21
ESCNet [42]	181	1981	2162	97.09	89.63
Proposed SAFNet	734	1222	1956	97.37	90.98

performance of change detection. The PCC value of SAFNet is 0.23%, 1.06%, and 0.28% ahead of DBN, DCNet, and ESCNet. According to the visualized result and PCC, SAFNet exhibits the best performance in Yellow River datasets.

Based on the above experiments on four real SAR datasets, the proposed SAFNet has superior performance over several traditional shallow classification models. Besides, by employing the dual network with similarity loss and classification loss to extract the high-level semantic features, the proposed SAFNet

TABLE VIII
CHANGE DETECTION RESULTS ON NOISE POLLUTED DATASETS

Dataset	PCC value on noise-polluted datasets(%)						
	$Var = 0$	$Var = 20$	$Var = 40$	$Var = 60$	$Var = 80$	$Var = 100$	$Ray = \sqrt{\pi/2}$
San Francisco	99.10	98.99	98.76	98.72	98.44	98.06	98.65
Ottawa	98.60	98.23	97.95	97.51	97.20	96.75	97.82
Yellow River I	98.94	98.75	98.63	98.57	98.29	98.17	98.33
Yellow River II	97.37	97.10	96.37	95.89	94.65	93.70	96.09

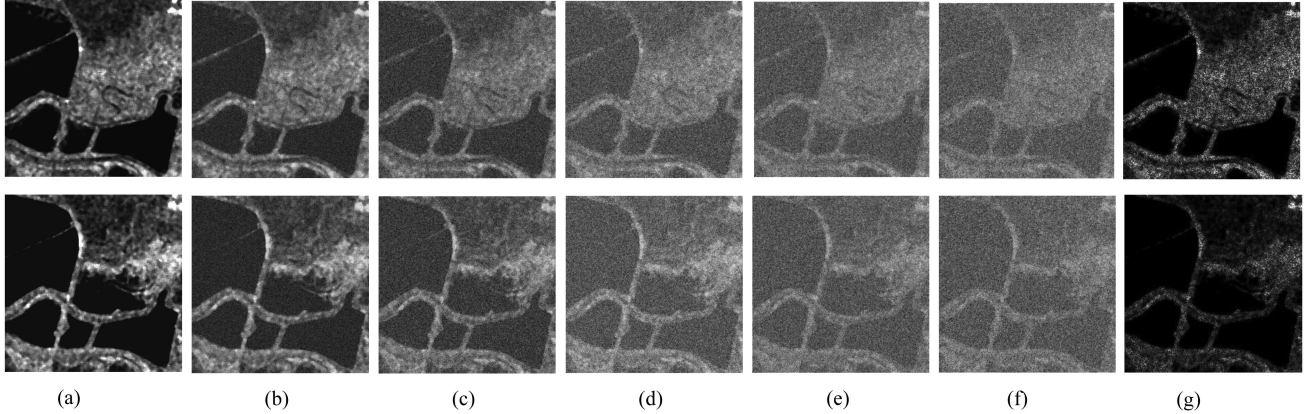


Fig. 14. Noise-polluted San Francisco dataset.

achieves better performance than other deep learning-based methods on four real SAR datasets. Furthermore, AF module improves the change detection performance by adaptively fusing the multilevel features for classification. Moreover, the proposed SAFNet has a strong capacity in feature learning. It is a powerful and useful tool for SAR image change detection.

D. Noise Robustness Analysis

Speckle noise degrades the quality of SAR images. It causes the entanglement between noise and signal. Therefore, we further compare and analyze the noise robustness of the proposed method in this experiment. First, speckle noise is added to the original SAR dataset $I_n = I \cdot n$. Here, I is the original SAR image, n is the noise characteristics. As shown in Tables VIII and VIII, $Var = 40$ represents the noise with Gaussian distribution variance of 40, and so on. $Ray = \sqrt{\pi/2}$ represents Rayleigh distribution noise with parameter $\sqrt{\pi/2}$. Larger variance means image degradation more seriously. The noise-polluted San Francisco dataset is typically shown in Fig. 14, Gaussian noise reduces the signal-to-noise ratio of the image, and thus affects the visualization. Rayleigh distribution noise aggravates the interference of multiplicative noise which is difficult to filter out.

We can observe from Table VIII that as the noise level increases, the performance of the change detection method deteriorates rapidly. Especially when $Var > 60$, the PCC values drop sharply. In addition, the results become worse under the interference of Rayleigh distribution noise. Although the

original images suffer from different levels of noise, the SAFNet still reaches acceptable PCC values in most cases. In conclusion, the performance of SAFNet declined with along a small scope. That is because the stable feature representation is directly obtained from the input images, thus the noise interference is suppressed to some extent.

IV. CONCLUSION

In the past few years, many deep learning methods for SAR image change detection have attracted a lot of attention. However, discriminative feature representation still needs improvement. In this article, features from different levels are adaptively fused by AF module. Therefore, meaningful features are emphasized and irrelevant ones are suppressed. In addition, the correlation layer is introduced to further integrate the features from two-branch networks. Besides, feature extraction of two-branch networks is guided by the similarity measure. In the projection space, the changed image pairs still maintain the similarity, while the unchanged pixels are far away from each other. After the correlation layer, the classification loss is employed to optimize the whole network. Therefore, distinctive feature representation is achieved for change detection. Compared with the state-of-the-art methods, the SAFNet exhibits superior performance in terms of quantitative metrics and visual comparison.

With the development of high-resolution Earth observation technology, change detection technology will serve high-resolution, long-time series, and large-scale scenarios in the

future. In addition, multisource image change detection is also deserved extensive attention.

REFERENCES

- [1] G. Quin, B. Pinel-Puysegur, J.-M. Nicolas, and P. Loreaux, "MIMOSA: An automatic change detection method for SAR time series," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5349–5363, Sep. 2014.
- [2] X. Bai and F. Zhou, "Analysis of new top-hat transformation and the application for infrared dim small target detection," *Pattern Recognit.*, vol. 43, no. 6, pp. 2145–2156, 2010.
- [3] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 858–867, Jul. 1997.
- [4] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 294–307, Mar. 2005.
- [5] M. Jian, K.-M. Lam, J. Dong, and L. Shen, "Visual-patch-attention-aware saliency detection," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1575–1586, Aug. 2015.
- [6] M. Zhang, M. Gong, H. He, and S. Zhu, "Symmetric all convolutional neural-network-based unsupervised feature extraction for hyperspectral images classification," *IEEE Trans. Cybern.*, to be published, doi: [10.1109/TCYB.2020.3020540](https://doi.org/10.1109/TCYB.2020.3020540).
- [7] J. Liu, M. Gong, A. K. Qin, and K. C. Tan, "Bipartite differential neural network for unsupervised image change detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 3, pp. 876–890, Mar. 2020.
- [8] H.-C. Li, T. Celik, N. Longbotham, and W. J. Emery, "Gabor feature based unsupervised change detection of multitemporal SAR images based on two-level clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2458–2462, Dec. 2015.
- [9] H. Chen, L. Jiao, M. Liang, F. Liu, S. Yang, and B. Hou, "Fast unsupervised deep fusion network for change detection of multitemporal SAR images," *Neurocomputing*, vol. 332, pp. 56–70, 2019.
- [10] L. Bruzzone and D. F. Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 452–466, Apr. 2002.
- [11] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874–887, Apr. 2005.
- [12] R. Dekker, "Speckle filtering in satellite SAR change detection imagery," *Int. J. Remote Sens.*, vol. 19, no. 6, pp. 1133–1146, 1998.
- [13] B. Hou, Q. Wei, Y. Zheng, and S. Wang, "Unsupervised change detection in SAR image based on Gauss-log ratio image fusion and compressed projection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3297–3317, Aug. 2014.
- [14] M. Gong, Y. Cao, and Q. Wu, "A neighborhood-based ratio approach for change detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 307–311, Mar. 2012.
- [15] S. Krinidis and V. Chatzis, "A robust fuzzy local information c-means clustering algorithm," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1328–1337, May 2010.
- [16] M. Gong, Z. Zhou, and J. Ma, "Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2141–2151, Apr. 2012.
- [17] F. Gao, J. Dong, B. Li, and Q. Xu, "Automatic change detection in synthetic aperture radar images based on PCANet," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1792–1796, Dec. 2016.
- [18] F. Gao, J. Dong, B. Li, Q. Xu, and C. Xie, "Change detection from synthetic aperture radar images based on neighborhood-based ratio and extreme learning machine," *J. Appl. Remote Sens.*, vol. 10, no. 4, 2016, Art. no. 046019.
- [19] S. Wang, S. Yang, and L. Jiao, "Saliency-guided change detection for SAR imagery using a semi-supervised Laplacian SVM," *Remote Sens. Lett.*, vol. 7, no. 11, pp. 1043–1052, 2016.
- [20] Y. Gao *et al.*, "Hyperspectral and multispectral classification for coastal wetland using depthwise feature interaction network," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2021.3097093](https://doi.org/10.1109/TGRS.2021.3097093).
- [21] M. Gong, H. Li, E. Luo, J. Liu, and J. Liu, "A multiobjective cooperative coevolutionary algorithm for hyperspectral sparse unmixing," *IEEE Trans. Evol. Comput.*, vol. 21, no. 2, pp. 234–248, Apr. 2017.
- [22] M. Jian *et al.*, "Saliency detection based on directional patches extraction and principal local color contrast," *J. Vis. Commun. Image Representation*, vol. 57, pp. 1–11, 2018.
- [23] Q. Wang, W. Huang, Z. Xiong, and X. Li, "Looking closer at the scene: Multiscale representation learning for remote sensing image scene classification," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: [10.1109/TNNLS.2020.3042276](https://doi.org/10.1109/TNNLS.2020.3042276).
- [24] Q. Wang, Q. Li, and X. Li, "A fast neighborhood grouping method for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5028–5039, Jun. 2021.
- [25] M. Jian *et al.*, "Visual saliency detection by integrating spatial position prior of object with background cues," *Expert Syst. with Appl.*, vol. 168, 2021, Art. no. 114219.
- [26] 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.
- [27] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 545–559, Mar. 2018.
- [28] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, Feb. 2019.
- [29] Y. Gao, F. Gao, J. Dong, and S. Wang, "Transferred deep learning for sea ice change detection from synthetic-aperture radar images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 10, pp. 1655–1659, Oct. 2019.
- [30] Q. Wang, Z. Yuan, Q. Du, and X. Li, "Getnet: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2018.
- [31] F. Liu, L. Jiao, X. Tang, S. Yang, W. Ma, and B. Hou, "Local restricted convolutional neural network for change detection in polarimetric SAR images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 818–833, Mar. 2019.
- [32] X. Li, Z. Yuan, and Q. Wang, "Unsupervised deep noise modeling for hyperspectral image change detection," *Remote Sens.*, vol. 11, no. 3, 2019, Art. no. 258.
- [33] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1492–1500.
- [34] Y. Bengio, N. Léonard, and A. Courville, "Estimating or propagating gradients through stochastic neurons for conditional computation," 2013, *arXiv:1308.3432*.
- [35] B. Yang, G. Bender, Q. V. Le, and J. Ngiam, "Condconv: Conditionally parameterized convolutions for efficient inference," 2019, *arXiv:1904.04971*.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 770–778.
- [37] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Vision Pattern Recognit.*, 2018, pp. 7132–7141.
- [38] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 510–519.
- [39] 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.
- [40] F. Gao, X. Wang, J. Dong, and S. Wang, "Synthetic aperture radar image change detection based on frequency-domain analysis and random multigraphs," *J. Appl. Remote Sens.*, vol. 12, no. 1, 2018, Art. no. 016010.
- [41] Y. Gao, F. Gao, J. Dong, and S. Wang, "Change detection from synthetic aperture radar images based on channel weighting-based deep cascade network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 11, pp. 4517–4529, Nov. 2019.
- [42] H. Zhang, M. Lin, G. Yang, and L. Zhang, "Escnet: An end-to-end superpixel-enhanced change detection network for very-high-resolution remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: [10.1109/TNNLS.2021.3089332](https://doi.org/10.1109/TNNLS.2021.3089332).



Yunhao Gao received the B.Sc. degree in electronic information engineering, in 2017, from Ocean University of China, Qingdao, China, where he is currently working toward the M.Sc. degree in computer science and applied remote sensing with the School of Information Science and Technology.

His research interests include computer vision and remote sensing image processing.



Feng Gao received the B.Sc. degree in software engineering from Chongqing University, Chongqing, China, in 2008, and the Ph.D. degree in computer science and technology from Beihang University, Beijing, China, in 2015.

He is currently an Associate Professor with the School of Information Science and Engineering, Ocean University of China, Qingdao, China. His research interests include remote sensing image analysis, pattern recognition, and machine learning.



Junyu Dong received the B.Sc. and M.Sc. degrees in applied mathematics from Ocean University of China, Qingdao, China, in 1993 and 1999, respectively, and the Ph.D. degree in image processing from the Department of Computer Science, Heriot-Watt University, Edinburgh, U.K., in 2003.

He is currently a Professor and Vice Dean with the School of Information Science and Engineering, Ocean University of China. His research interests include visual information analysis and understanding, machine learning, and underwater image processing.



Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland, Baltimore, MD, USA, in 2000.

She is currently the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is also a Fellow of the SPIE-International Society for Optics and Photonics. She was the Chair of the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition, from 2010 to 2014. She was the Co-Chair of the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society, from 2009 to 2013. She was an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, from 2011 to 2015, the *Journal of Applied Remote Sensing*, from 2014 to 2015, and the IEEE SIGNAL PROCESSING LETTERS, from 2012 to 2015. She is the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, from 2016 to 2020.



Heng-Chao Li (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from Southwest Jiaotong University, Chengdu, China, in 2001 and 2004, respectively, and the Ph.D. degree from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2008, all in information and communication engineering.

He is currently a Full Professor with the School of Information Science and Technology, Southwest Jiaotong University. His research interests include statistical analysis of synthetic aperture radar (SAR) images, remote sensing image processing, and pattern recognition.

Dr. Li was the recipient of the 2018 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society for his service to the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He is an Editorial Board Member of the *Journal of Southwest Jiaotong University* and the *Journal of Radars*. Moreover, he has served as a Guest Editor for the Special Issues of the *Journal of Real-Time Image Processing*, the IEEE JSTARS, and the IEEE JOURNAL ON MINIATURIZATION FOR AIR AND SPACE SYSTEMS. He is serving as an Associate Editor for the IEEE JSTARS. He was the Program Committee Member for the 26th International Joint Conference on Artificial Intelligence (IJCAI-2017), and the 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images2019, and the Session Chair for the 2017 International Geoscience and Remote Sensing Symposium, and the 2019 Asia-Pacific Conference on Synthetic Aperture Radar.