Change Detection in Synthetic Aperture Radar Images Using a Dual-Domain Network

Xiaofan Qu, Feng Gao, Member, IEEE, Junyu Dong, Member, IEEE, Qian Du, Fellow, IEEE, and Heng-Chao Li

Abstract—Change detection from synthetic aperture radar (SAR) imagery is a critical yet challenging task. Existing methods mainly focus on feature extraction in the spatial domain, and little attention has been paid to the frequency domain. Furthermore, in patch-wise feature analysis, some noisy features in the marginal region may be introduced. To tackle the above two challenges, we propose a dual-domain network (DDNet). Specifically, we take features from the discrete cosine transform (DCT) domain into consideration and the reshaped DCT coefficients are integrated into the proposed model as the frequency domain branch. Feature representations from both frequency and spatial domain are exploited to alleviate the speckle noise. In addition, we further propose a multi-region convolution (MRC) module, which emphasizes the central region of each patch. The contextual information and central region features are modeled adaptively. The experimental results on three SAR data sets demonstrate the effectiveness of the proposed model. Our codes are available at https://github.com/summitgao/SAR_CD_DDNet.

Index Terms—Change detection, deep learning, frequency domain, neural network, synthetic aperture radar (SAR).

I. INTRODUCTION

RECENT years have witnessed the rapid growth of synthetic aperture radar (SAR) sensors, inducing many researchers to work on SAR image segmentation, texture analysis and change detection [1]–[3]. As an essential task of SAR image interpretation, change detection, focusing on identifying the changed regions between multitemporal SAR images, has attracted increasing attention in remote sensing communities.

Change detection using SAR images is more challenging than optical ones due to the existence of speckle noise [4]. Some pioneer efforts have been made to tackle the problem of noise in multitemporal image analysis [5]. Traditionally, the mainstream methods commonly compare multitemporal images to generate a difference image (DI) and analyze the DI to obtain the change map [6]. Although some pixel-wise changed information can be captured, these methods can hardly adaptively exploit the rich feature representations in the original data.

Currently, due to the popularity of deep neural networks, remote sensing image change detection methods have achieved a performance boost. Samadi et al. [7] employed deep belief network (DBN) for SAR image feature learning. The change detection performance has been greatly improved compared with traditional DI clustering methods. Wang et al. [8] presented a general end-to-end convolutional neural network (CNN) for hyperspectral image change detection. Li et al. [9] proposed a well-designed CNN to learn the spatial characteristics from raw images. In [10], a local restricted CNN was proposed to enforce local similarity between neighboring pixels in change maps.

The abovementioned deep learning-based methods have achieved great success by exploiting deep feature representations. However, it is nontrivial to build a robust SAR change detection model, due to the following two challenges.

1) Mutual Reinforcement of Spatial and Frequency Features: Existing models are mainly based on feature extraction in the spatial domain, and little attention has been paid to the frequency domain. It has been recently shown that the compressed representation in the frequency domain is capable of suppressing the noise in the spatial domain and enriching the patterns for image understanding [11], [12]. Thereby, reinforcing spatial and frequency features in a unified framework should be considered.

2) Enhancement of the Central Region Features: The contextual information is crucial to the performance of SAR image change detection, and hence existing methods commonly use patch-wise features for classification. However, some noisy features in the marginal region of each patch may be introduced. Therefore, how to emphasize the central region in each patch while retaining the contextual information is a tough challenge.

To tackle the above two issues, we propose a Dual-Domain Network, DDNet for short, which jointly exploits the spatial and frequency features for SAR change detection task. Specifically, we first take into consideration features from the discrete cosine transform (DCT) domain. Reshaped DCT coefficients...
are integrated into a CNN model as a branch for inference. Therefore, feature representations from both frequency and spatial domain are exploited to alleviate the speckle noise. In addition, we further proposed a multi-region convolution (MRC) module, which emphasizes the central region of each patch. Intuitively, the contextual information and central region features are adaptively modeled efficiently.

The contributions of this letter can be summarized as follows.

1) To the best of our knowledge, we are the first to introduce features in the DCT domain to solve the SAR image change detection problem. Features from both frequency and spatial domain are exploited. Therefore, the speckle noise is suppressed efficiently.

2) We propose an MRC module, which jointly emphasizes the central area of each image patch while keeping the contextual information. The central region features and contextual information are adaptively organized for the classification task.

3) Experimental results on three real SAR data sets demonstrate the effectiveness of the proposed dual-domain network (DDNet) and promising results are observed. To benefit other research studies along remote sensing image change detection, the PyTorch code of the proposed DDNet has been made publicly available.

II. METHODOLOGY

Given two SAR images $I_1$ and $I_2$ captured at different times over the same geographical region, the objective is to generate a binary change map in which the changed pixels are marked as “1” and unchanged pixels are marked as “0.”

The proposed model works in an unsupervised manner. The preclassification is the first step. The main purpose of this step is to find samples which have a high probability to be changed or unchanged. The log-ratio operator [13] is first employed to generate a DI. Then, hierarchical FCM clustering [14] is implemented to classify the DI into three clusters: $\Omega_c$, $\Omega_r$, and $\Omega_u$. Pixels belonging to $\Omega_c$ and $\Omega_u$ are reliable pixels that have a high probability to be changed or unchanged, respectively. Pixels in $\Omega_r$ are uncertain and need to be further classified. 10% of the image patches centered at pixels in $\Omega_r$ and $\Omega_u$ are randomly chosen as the training samples for DDNet. It should be noted that the number of positive samples and negative samples are equal. For a given pixel, image patches centered at the pixel are extracted from $I_1$ and $I_2$, respectively. The size of each patch is $r \times r \times r$ ($r = 7$ in this work). Both patches are combined to form a new image patch with the size of $2 \times r \times r$. The generated new patches are fed into the DDNet for training. After training, image patches centered at pixels in $\Omega_r$ will be classified by the network. The whole process is unsupervised.

The framework of the proposed DDNet is illustrated in Fig. 1. The network comprises a spatial domain branch and a frequency domain branch. The spatial domain branch is designed to capture multi-region features, while the frequency domain branch is used to encode the DCT coefficients. Next, we introduce the spatial and frequency domain feature extraction, respectively. Afterward, we show how to combine the features from both domains for further classification.

A. Spatial Domain Feature Extraction

In the spatial domain, the network contains four MRC modules as illustrated in Fig. 1. Details of the MRC module are shown in Fig. 2. Since the contextual information is essential for SAR image change detection, existing methods commonly adopt a window with fixed size ($3 \times 3$, $5 \times 5$, $7 \times 7$, etc.) to determine whether the position has changed. We argue that if we abandon some marginal regions in feature extraction, the central region can be emphasized and the noise in marginal regions can be potentially eliminated. Toward this end, we propose to extract features of multi-regions to enhance the feature representation in SAR change detection.

The following are three representative situations.

1) Global Region: As illustrated in Fig. 2, it is a candidate of the square-shaped patch, where the CNN is guided to capture the global contextual information of the central pixel.

2) Horizontal Middle Region: The image patch is designed to concentrate on the center area, and pixels on the top and bottom have been removed.

3) Vertical Middle Region: The image patch is designed to concentrate on the center area, and pixels on the left and right have been removed.
If the abovementioned regions are considered in the CNN model, the central region will be emphasized and the noisy pixels on the margin will be effectively suppressed.

Given an image patch \( A \in \mathbb{R}^{2 \times r \times r} \), we feed it into a \( 1 \times 1 \) convolution layer to generate new feature map \( F \in \mathbb{R}^{C \times r \times r} \). Then, \( F \) is divided into three groups \( F_g, F_h, \) and \( F_v \) according to the channel dimension. Hence, the shape of \( F_g, F_h, \) and \( F_v \) is \((C/3) \times r \times r\), respectively. \( F_g \) represents the global region features, \( F_h \) denotes the horizontal middle region feature, where several top and bottom rows are set to 0. \( F_v \) denotes the vertical middle region feature, where several left and right columns are set to 0. It should be noted that \( r \) is set to 7 in this work, two top rows and two bottom rows are set to 0 in \( F_h \), while two left and two right columns are set to 0 in \( F_v \).

After the \( 3 \times 3 \) convolution layer, we obtain three groups of features \( F'_{g}, F'_{h}, \) and \( F'_{v} \), respectively. These features are fused by element-wise summation as

\[
F_{\text{fus}} = F'_{g} + F'_{h} + F'_{v}
\]

where \( F_{\text{fus}} \in \mathbb{R}^{(C/3) \times r \times r} \) represents the spatial fused features. \( C \) is set to be 15 in our implementations, and we therefore obtain the final spatial fused feature map with the size of \( 5 \times 7 \times 7 \). The features are then reshaped to a vector \( V_{s} \), and the vector is \( 5 \times 7 \times 7 = 245 \) in length. Hence, \( V_{s} \) has a global context view and the center region information has been enhanced.

### B. Frequency Domain Feature Extraction

Due to the presence of speckle noise, it is difficult to extract robust discriminative features from the DI in the spatial domain. Recently, researchers have verified that the compressed representation in the frequency domain is capable of suppressing the speckle noise [11]. Inspired by Xu’s work [12], DCT is employed for frequency feature extraction.

An input image patch of size \( 2 \times r \times r \) is resized to \( 2 \times 8 \times 8 \) by bilinear interpolation. Then, the image patch is converted to the frequency domain by DCT. Afterward, the obtained DCT coefficient vector \( v \) is \( 2 \times 64 = 128 \) in length. It has been confirmed that DCT proves highly effective in noise suppression in the spatial domain. In order to further select the critical components of the DCT coefficients, we employ an “ON-OFF switch” which generates an informative vector \( i \) and an attention gate \( g \) with two linear transformations, as illustrated in Fig. 1. The informative vector \( i \) and attention gate \( g \) are generated as

\[
i = W^i v + b^i
\]

\[
g = \sigma(W^g v + b^g)
\]

where \( W^i \) and \( W^g \) are the weight matrix, \( b^i \) and \( b^g \) are the bias in linear transformation, and \( \sigma \) denotes the sigmoid activation function.

Then, the DCTB employs the attention gate to the informative vector using element-wise multiplication and obtains the final frequency feature vector \( V_f \) as

\[
V_f = g \odot i
\]

where \( \odot \) is the element-wise multiplication.

### C. Final Change Map Generation

After obtaining the spatial domain feature \( V_s \) and frequency domain feature \( V_f \), they are concatenated and fed into a fully connected (FC) layer. Then, the possibility of changed or unchanged is calculated by a softmax layer to generate the output. After training, pixels in \( \Omega \) will be classified by the network, and the final change map can be obtained.

## III. Experimental Results and Analysis

In this section, extensive experiments are conducted using three real SAR data sets. First, we describe the data sets used in the experiment and give the evaluation criteria. Then, the critical parameters that affect the change detection result are discussed. Finally, the DDNet is compared with several state-of-the-art methods.

### A. Data Set and Evaluation Metric

To verify the effectiveness of the proposed DDNet, we use three co-registered and geometrically corrected multitemporal SAR data sets acquired by different sensors. The first data set is the Ottawa data set, which was captured by the RADARSAT SAR sensor in May and August 1997. The size of each image is \( 290 \times 350 \) pixels. The second data set is the Sulzberger data set with the size of \( 256 \times 256 \) pixels. It is a part of the Sulzberger Ice Shelf which was provided by the European Space Agency’s ENVISAT satellite. This image shows the breakup of an ice shelf caused by a tsunami in March 2011. The last data set is the Yellow River data set, which is from the Yellow River Estuary in China. It is acquired by the Radarsat-2 satellite in June 2008 and June 2009. The size of the data set is \( 291 \times 306 \) pixels. It is worth noting that it is difficult to identify the changed regions accurately since speckle noise is much stronger in the Yellow River data set.

We measure the performance of the proposed DDNet with five common evaluation metrics in change detection, including false-positives (FP), false-negatives (FN), overall error (OE), percentage of correct classification (PCC), and kappa coefficient (KC).

### B. Analysis of the Patch Size

The spatial contextual information is captured by an image patch with the size of \( r \). We evaluate the performance of the proposed DDNet by taking \( r = 5, 7, 9, 11, 13, \) and 15, respectively. Fig. 3 shows that the relationship between \( r \) and PCC. It can be observed that, as \( r \) gradually increases, the value of PCC first rises and then tends to be stable. The PCC curves show that the spatial contextual information is significant for the change detection task. However, a large patch size increases the computational burden, and it may introduce some noise information that affects the change detection performance. Therefore, we take \( r = 7 \) in our following implementation.

### C. Ablation Study

In order to explore the effectiveness of the MRC module and frequency domain features, ablation studies are conducted on...
We designed three variants: 1) CNN denotes the traditional CNN; 2) w/o DCT refers to the model removing the frequency domain branch; and 3) w/o MRC denotes the model excluding the spatial domain branch. From Table I, it can be observed that removing either the DCT or MRC module affects the feature expressions and degrades the change detection performance.

To further demonstrate the effectiveness of the MRC and DCT, we visualize the features of both domains in Fig. 4. The representations learned in the dual domain are significantly better than the input with more separated and clearly bounded clusters.

D. Performance Comparison

To verify the effectiveness of the proposed model, we compared DDNet with the following state-of-the-art baselines: PCAKM [6], NR-ELM [15], DBN [16], DCNet [17] and MSAPNet [18]. It is worth mentioning that the above methods are implemented with the default parameters described in their works. Table II and Fig. 5 summarize the quantitative and visualized results of different methods, respectively.

On the Ottawa data set (first row of Fig. 5), we can observe that the result of PCAKM suffers from a high FN value, which means that many changed regions are missed. The proposed DDNet achieves the best performance. The KC value of the proposed method is improved by 3.04%, 0.39%, 0.01%, 0.3% over PCAKM, NR-ELM, DBN, DCNet, and MSAPNet, respectively. Meanwhile, on the Sulzberger data set (second row of Fig. 5), it can be seen that the change maps generated by PCAKM and NR-ELM contain many noisy areas, and thereby the overall performance is affected. For NR-ELM, DBN, and DCNet, the final change maps miss many changed regions, and they are afflicted with high FN values. Furthermore, deep learning-based methods have better performance than shallow model methods. The KC value of the proposed DDNet increased by 0.99%, 0.53%, and 0.57% over DBN, DCNet, and MSAPNet on the Sulzberger data set, respectively. From the above comparisons, it is evident that the proposed DDNet is capable of extracting features from the spatial and frequency domains in parallel effectively on the Ottawa and Sulzberger data sets.

The speckle noise on the Yellow River data set is much stronger, and hence it is rather challenging to identify the changed regions accurately. On the Yellow River data set (the third row of Fig. 5), there are many small noisy areas in the final changed maps generated by PCAKM and NR-ELM. Therefore, the FN values of both methods are extremely high in Table II. Compared with other methods, the proposed DDNet obtains 8.63%, 10.18%, 3.04%, 0.79%, and 1.01% gains in KC over PCAKM, NR-ELM, DBN, DCNet, and MSAPNet, respectively.

Experimental results on the three data sets demonstrate that the proposed DDNet achieves the best performance, substantially surpassing all baselines. Besides, our model gains an advantage over traditional CNN-based models, which is mainly due to the joint learning of spatial and frequency domain features.

IV. CONCLUSION AND FUTURE WORK

In this letter, we present a novel DDNet to tackle the SAR change detection task. In the DDNet, the features of spatial and frequency domains are fused to improve the classification performance. In the spatial domain, we design an MRC module to enhance the central region features in the input image patches. In the frequency domain, we employ DCT and gating mechanism to extract the frequency features. Experimental results on three data sets demonstrated that the proposed DDNet is superior to several excellent change detection methods. In the future, we plan to work on large-scale data set to verify our change detection method.
TABLE II

<table>
<thead>
<tr>
<th>Method</th>
<th>Results on the Ottawa dataset</th>
<th>Results on the Sulzberger dataset</th>
<th>Results on the Yellow River dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FP</td>
<td>FN</td>
<td>OE</td>
</tr>
<tr>
<td>PCAKM [6]</td>
<td>955</td>
<td>1515</td>
<td>2470</td>
</tr>
<tr>
<td>NR ELM [15]</td>
<td>695</td>
<td>1076</td>
<td>1771</td>
</tr>
<tr>
<td>DBN [16]</td>
<td>995</td>
<td>704</td>
<td>1699</td>
</tr>
<tr>
<td>DCNet [17]</td>
<td>679</td>
<td>1051</td>
<td>1730</td>
</tr>
<tr>
<td>MSAPNet [18]</td>
<td>805</td>
<td>952</td>
<td>1757</td>
</tr>
<tr>
<td>Proposed DDNet</td>
<td>641</td>
<td>1027</td>
<td>1668</td>
</tr>
</tbody>
</table>

Fig. 5. Visualized results of different change detection methods on the Ottawa data set (first row), Sulzberger data set (second row), and Yellow River data set (third row). (a) Image captured at \( t_1 \). (b) Image captured at \( t_2 \). (c) Ground truth image. (d) Result by PCAKM [6]. (e) Result by NR-ELM [15]. (f) Result by DBN [16]. (g) Result by DCNet [17]. (h) Result by MSAPNet [18]. (i) Result by the proposed DDNet.

REFERENCES


