Synthetic Aperture Radar Image Change Detection via Layer Attention-Based Noise-Tolerant Network

Desen Meng, Feng Gao, Member, IEEE, Junyu Dong, Member, IEEE, Qian Du, Fellow, IEEE, and Heng-Chao Li

Abstract—Recently, change detection methods for synthetic aperture radar (SAR) images based on convolutional neural networks (CNNs) have gained increasing research attention. However, existing CNN-based methods neglect the interactions among multilayer convolutions, and errors involved in the pre-classification restrict the network optimization. To this end, we proposed a layer attention-based noise-tolerant network, termed LANTNet. In particular, we design a layer attention module that adaptively weights the feature of different convolution layers. In addition, we design a noise-tolerant loss function that effectively suppresses the impact of noisy labels. Therefore, the model is insensitive to noisy labels in the preclassification results. The experimental results on three SAR datasets show that the proposed LANTNet performs better compared to several state-of-the-art methods. The source codes are available at https://github.com/summitgao/LANTNet.

Index Terms—Change detection, layer attention, noise-tolerant loss, synthetic aperture radar (SAR).

I. INTRODUCTION

REMOTE sensing image change detection aims to identify regions that changed in the same area using multitemporal images. In fact, it facilitates various applications, such as disaster assessment, deforestation monitoring, and object detection [1], [2], [3]. The synthetic aperture radar (SAR) is an ideal source for change detection since the SAR sensor is independent of sunlight illumination and weather conditions. Therefore, SAR change detection has received extensive research attention recently.

Many supervised and unsupervised methods have been devoted to detect changes in multitemporal SAR images. In fact, it commonly costs a lot of effort to collect high-quality labeled samples that reflect the changed information, and learning the changed information in an unsupervised manner is more efficient. Therefore, in this letter, we focus on unsupervised SAR change detection.

Traditionally, the task of SAR image change detection are solved by clustering methods, such as k-means [4], fuzzy c-means (FCM) [5], and hierarchical clustering [6]. Recently, leveraging the representation power of convolutional neural networks (CNNs) and the attention mechanism to capture changed information has become a common solution for change detection. Gao et al. [7] designed an adaptive fusion convolution module to encode the input, allowing the network to focus on the more informative channels. Zhang et al. [8] proposed a sparse feature clustering network for SAR image change detection. Geng et al. [9] presented a saliency-guided network to suppress the influence of speckle noise and improve the SAR change detection performance. In [10], transfer learning is employed for the parameter initialization for change detection deep networks. Furthermore, wavelet-CNN [11], discrete cosine transform [12], and ensemble learning [13] are also used to enhance the representation power of deep networks.

Indeed, the CNN-based models have shown impressive performance in change detection task. However, they still show limitations in the following aspects.

1) Feature interaction among convolution layers. Research in human vision system finds that visual representations in the cortex are activated in parallel fashion [14]. Existing self-attention-enhanced CNNs commonly incorporate an attention block after a convolution layer, which overlooks the interactions among multilayer convolutions. In light of this, it is essential to explore the information flow among the convolutional layers.

2) Errors involved in the pseudo-labels. Existing unsupervised change detection methods generally use clustering to generate pseudo-labels with high confidence for training. However, the errors involved in the pseudolabels restrict network optimization [15]. Therefore, it is vital to build a noise-tolerant model for unsupervised change detection task.

To address the aforementioned issues, we propose a layer attention-based noise-tolerant network, dubbed as LANTNet, which interconnects convolutional layers and is less sensitive to noisy labels. Specifically, we design a layer attention module for change detection, which adaptively weights features from different layers, as illustrated in Fig. 1. In addition, we design a noise-tolerant loss function that can effectively suppress the impact of noisy labels. It combines the cross entropy (CE) loss and the mean absolute error (MAE). Hence, the network is less sensitive to noisy labels and has a fast convergence rate. We conducted extensive experiments on three SAR datasets, and the results demonstrate that the
proposed LANTNet yields better performance as compared to state-of-the-art methods.

In summary, the contributions of this letter are threefold as follows.

1) We propose a layer attention module, which exploits the correlations of multilayer convolutions. Features from different convolution layers cooperate with each other, effectively improving the representation power of the network.

2) We introduce a noise-tolerant loss function to alleviate the influence of noisy labels in pseudo-labeled samples. It makes the model less sensitive to noisy labels and has fast convergence rate.

3) Experimental results confirm the effectiveness of the proposed LANTNet. Furthermore, we have made our data and codes publicly available to benefit other research of the remote sensing community.

II. METHODOLOGY

Given the multitemporal SAR images $I_1$ and $I_2$ acquired in the same geographical zone, the goal of change detection is to generate a change map where each spatial location is assigned to one type of change. In this work, we consider the generation of change map as a binary classification task, which means that the label is either 1 (changed) or 0 (not changed).

As shown in Fig. 2, the proposed LANTNet comprises three components: 1) preclassification module; 2) convolutional stem; and 3) layer attention module. First, multitemporal coregistered SAR images $I_1$ and $I_2$ are fed into the preclassification module to obtain samples that are highly likely to be changed or unchanged. New image patches will be generated and then fed into the convolutional stem to compute feature groups $X$. The feature groups are then updated to attention features $Y$ by the layer attention module. Finally, $Y$ and the features of the convolutional stem are combined for classification to calculate the final change map.

A. Preclassification and Convolutional Stem

In the preclassification module, the logarithmic ratio operator and the hierarchical FCM algorithm are employed, and details of the module can be found in [6]. After preclassification, the image patches ($R \times R$ pixels) with pseudolabels are extracted from $I_1$, $I_2$, and the difference image, respectively. They are concatenated to form new image patches ($3 \times R \times R$ pixels) as training samples.

The convolutional stem consists of four convolution layers. $1 \times 1$ convolution is first used to extract the shallow feature $F_0$. Then, three $3 \times 3$ convolution layers are employed to extract intermediate features $F_i$ as follows:

$$F_i = \text{Conv}(F_{i-1}), \quad i = 1, 2, 3. \quad (1)$$

For the shallow feature $F_0$, we set the channel numbers to 16, and for intermediate features $F_1$, $F_2$, and $F_3$, we set the channel numbers to 32. Then, $1 \times 1$ convolution is employed to increase the channel dimensions of $F_0$ to 32. After that, $F_0$, $F_1$, $F_2$, and $F_3$ are combined to form a feature group, and fed into the layer attention module.

B. Layer Attention Module

The attention mechanism helps the network focus on the most important part and neglect irrelevant information. However, the correlations among multilayer convolutions are rarely taken into account. To this end, we design a layer attention module to weigh the features $F_0$, $F_1$, $F_2$, and $F_3$ from multiple layers by learning the correlation matrix. Therefore, layer attention effectively fuses the spatial information of low-level layers and the semantic information from high-level layers. The informative layers are emphasized, and the redundant ones are suppressed. Fig. 2 shows the details of the proposed layer attention module. The input of the module is the feature group $X \in \mathbb{R}^{N \times C \times R \times R}$ from the convolution stem. In our implementation, $N$ is set to 4. Then we reshape $X$ to a feature matrix with the dimension of $N \times CRR$. Subsequently, we apply matrix multiplication between the new feature matrix and a diagonal matrix $W$ to weigh the features from different layers. Note that $W$ is initialized to identity matrix and only the diagonal elements of $W$ are updated in each epoch. The weight matrix adaptively assigns weights to the input feature groups. After that, a weighted feature matrix $\hat{X} \in \mathbb{R}^{N \times CRR}$ can be obtained.

Next, we calculate an attention matrix $A \in \mathbb{R}^{N \times N}$ as the correlation matrix as follows:

$$A = \text{Softmax}(\hat{X} \otimes \hat{X}^T) \quad (2)$$

where the element $A_{i,j}$ in the attention matrix denotes the correlation index between the $i$th and $j$th layer features, and $\otimes$ denotes the matrix multiplication. Matrix multiplication is performed between the weighted feature matrix $\hat{X}$ and its transpose matrix $\hat{X}^T$. Softmax operation is applied to generate the attention matrix $A$.

Finally, we multiply the weighted feature matrix $\hat{X}$ by the attention matrix $A$ and reshape the output to the dimension of $N \times C \times R \times R$. The input $X$ are combined to form the final
output $Y$ as follows:

$$Y = \phi(A \otimes \hat{X}) \oplus X$$

where $\phi$ denotes the reshape operation.

To compute the final change map, $Y$ is reshaped to the dimension of $NC \times R \times R$ and then fed into a $1 \times 1$ convolution layer for channel reduction, as illustrated in Fig. 2. After that, the combined features are handled by a fully connected (FC) layer. The softmax function is used to calculate the possibility of changed or unchanged. After training, the final change map can be obtained by classifying all pixels from the multi-temporal SAR images.

### C. Noise-Tolerant Loss Function

Like most existing unsupervised methods, we employ clustering to generate pseudo-labels with high probability for network training. However, errors are commonly involved in these pseudolabels and restrict the network optimization [15]. Therefore, it is critical to build a noise-tolerant model.

Toward this end, we design a noise-tolerant loss which combines the CE loss and the MAE. Suppose the dataset for change detection is represented as $\{(x_i, y_i) | 1 \leq i \leq m\}$, where $x_i$ is the input image patch, $y_i$ is the label associated with $x_i$, and $m$ is the total number of samples. The MAE loss is defined as follows:

$$L_{MAE}(f(x), y) = \|e_y - f(x)\|_1$$

where $f(\cdot)$ represents the network and $e_y$ is a one-hot vector with $e_{yj} = 1$ if $j = y$, otherwise $e_{yj} = 0$.

The CE loss is computed as follows:

$$L_{CE}(f(x), y) = -e_y \log f(x).$$

Ghosh et al. [16] empirically demonstrated that MAE is robust to noisy labels, but it often costs more time to converge. To the contrary, CE tends to overfit noisy labels, but it performs better than MAE when training with clean data. To leverage the good convergence performance of CE, and exploit the noise robustness of provided by MAE, we design a robust loss function which combines the CE and MAE. Formally, the robust loss function is defined as follows:

$$L = \alpha L_{CE} + \beta L_{MAE}$$

where $\alpha$ and $\beta$ are two decoupled hyperparameters. In our implementation, we set $\alpha$ to 0.1 and $\beta$ to 0.9 for a tradeoff between convergence efficiency and the noise robustness.

### III. Experiments and Analysis

#### A. Datasets and Evaluation Metrics

To evaluate the proposed LANTNet, we conducted extensive experiments on three real SAR datasets captured by different sensors. The first dataset is Chao Lake dataset with a size of $384 \times 384$ pixels. It was captured by the Sentinel-1 sensor in May and July 2020 in China, respectively. The second dataset is the Ottawa dataset with the size of $290 \times 350$ pixels, which was acquired by the Radarsat sensor in May and August 1997, respectively. The Ottawa dataset shows the changed areas afflicted by floods. The third dataset is the Sulzberger dataset which was captured by the ENVISAT satellite of the European Space Agency on March 11 and 16, 2011. Both images show a breakup of the ice shelf caused by a tsunami.

Five commonly used evaluation metrics include: 1) false positive (FP); 2) false negative (FN); 3) overall error (OE); 4) percentage of correct classification (PCC); and 5) Kappa coefficient (KC) are used to evaluate the change detection performance.
Fig. 3. Relationship between image patch size and the PCC value.

Fig. 4. Visualization of the feature representations on the Ottawa dataset. (a) Features before the layer attention module. (b) Features after the layer attention module.

B. Analysis of the Patch Size

The image patch size $R$ is a critical parameter. To explore the relationship between the patch size $R$ and PCC, we set $R$ from 5 to 15, and the experimental results are shown in Fig. 3. It can be seen that all curves tend to increase first and then decrease. When $R$ is small, the contextual information around each pixel is insufficient. However, if $R$ becomes too large, irrelevant neighborhood information would be taken into account. Therefore, we set $R = 9$ for the Chao Lake dataset and $R = 7$ for the Ottawa and Sulzberger datasets.

C. Ablation Study

To verify the effectiveness of the proposed layer attention module and the noise-tolerant loss, we conduct ablation experiments on three datasets. Three variants are designed for comparison. First, we design a basic CNN with the same structure as LANTNet, which removes the layer attention module and uses CE as loss function instead of the noise-tolerant loss. Then, we run our model without layer attention module (w/o layer attention) and without the noise-tolerant loss (w/o noise-tolerant loss). The results are shown in Table I. It can be observed that both the layer attention module and the noise-tolerant loss are effective for change detection performance improvement.

Furthermore, to gain deeper insights toward the layer attention, we visualized the features before and after the layer attention module with the tool of t-distributed stochastic neighbor embedding (t-SNE). As shown in Fig. 4, the distance between features clusters derived by the layer attention is larger, which reflects the feature representations by our model are indeed more discriminative.

D. Experimental Results and Comparison

To verify the effectiveness of the proposed LANTNet, five closely related change detection methods are used for comparison include: 1) principal component analysis and K-means (PCAKM) [4]; 2) nearest neighborhood-extreme learning machine (NR-ELM) [17]; 3) GaborPCANet [6]; 4) convolutional-wavelet neural network (CWNN) [11]; and 5) DDNet [12]. Note that the methods mentioned above are implemented with the default parameters provided in their work. All visual results are illustrated in Fig. 5, and the corresponding quantitative results are shown in Table II.

On the Chao Lake dataset (the first row of Fig. 5), PCAKM has the highest FP value among all methods and the FP value affects the overall performance of PCAKM. At the same time, GaborPCANet and DDNet also suffer from high FP values, while NR-ELM and CWNN suffer from high FN values. CWNN and DDNet are two CNN-based methods and have strong feature representation power, as shown in Table II. However, the pseudolabels utilized by CWNN and DDNet are not accurate enough, and their models may overfit the noisy labels. The KC value of the proposed
LANTNet is improved by 26.07%, 10.38%, 8.99%, 5.02%, and 3.34% over PCAKM, NR-ELM, GaborPCANet, CWN, and DDNet, respectively. It is evident that the proposed method achieves the best performance on the Chao Lake dataset.

On the Ottawa dataset (the second row of Fig. 5), PCAKM, NR-ELM, GaborPCANet, and DDNet suffer from high FN values and many change pixels are not correctly classified. CWN suffers from a high FP value and generates many noisy spots in the change map. The proposed LANTNet achieves the best PCC value, and it demonstrates that layer attention effectively improves the feature representation. The comparisons demonstrate the good performance of the proposed LANTNet on the Ottawa dataset.

On the Sulzberger dataset (the third row of Fig. 5), all the methods suffer from high FP values, and the proposed LANTNet achieves the lowest FP value. It can be seen from Fig. 5 that the proposed LANTNet is effective in suppressing the label noise and reduces the FP values. The KC value of the proposed LANTNet is improved by 7.38%, 3.8%, 3.53%, 2.89%, and 0.77% over PCAKM, NR-ELM, GaborPCANet, CWN, and DDNet, respectively. The comparison shows that the LANTNet is powerful in noise suppression on the Sulzberger dataset.

IV. CONCLUSION

In this letter, we propose a layer attention-based noise-tolerant network (LANTNet) for SAR change detection. Compared with existing CNN-based methods, our LANTNet improves in two aspects. First, we design a layer attention module that adaptively weights the feature of different convolution layers. In addition, since the pseudolabels obtained by preclassification are not accurate enough, we design a noise-tolerant loss to help the network avoid overfitting to the noisy labels. Experimental results on three multitemporal SAR datasets demonstrate that the proposed LANTNet is superior to several state-of-the-art change detection methods.

REFERENCES


