Change Detection in SAR Images Based on Improved Non-Subsampled Shearlet Transform and Multi-Scale Feature Fusion CNN

Fangyu Shen, Yanfei Wang, and Chang Liu

***Abstract*—Traditional methods for change detection in synthetic aperture radar images have difficulty in obtaining results from the generated differential image (DI) owing to speckle noise. In recent years, many deep learning-based methods have emerged because of their outstanding anti-noise and self-learning ability. However, they are limited by the requirement of abundant high-precision labels. Therefore, in this article, we propose a novel unsupervised method based on improved non-subsampled shearlet transform (NSST) and multi-scale feature fusion convolutional neural network for change detection. First, this method improves the traditional NSST algorithm and proposes a novel pseudo-label generator to obtain more pseudo-labels with higher confidence. It is noteworthy that the more accurate the pseudo-labels are, the better the change detection results will be. Second, this method designs a multi-scale feature fusion block in the network to make the feature images contain more complete information and reduces the number of pooling layers to avoid losing feature image details. The main idea of this method is to eliminate the step of generating the DI and directly obtain results from the original images. The theoretical analysis and final results conducted on three real datasets prove its validity. Furthermore, to verify the generality and potential of the proposed method, we apply it to the cross-region change detection and compare it with the supervised method, which achieve satisfactory results.**

***Index Terms*—Change detection, convolutional neural network (CNN), non-subsampled shearlet transform (NSST), synthetic aperture radar (SAR).**

I. Introduction

CHANGE detection in synthetic aperture radar (SAR) images is an essential technology for analyzing and identifying the changed information in the same region at different times [1]. In contrast to optical sensors, SAR has several unique capabilities, such as working all day and making high resolution images without being restricted by harsh environments and clouds [2]. These capabilities make SAR widely used in the change-detection studies related to environmental monitoring, agricultural surveys, hazard assessment, land detection, and so on [3]–[7]. However, the unique coherent imaging principle of SAR inevitably makes the echo signal of the radar target fade, so that the signal received by the radar receiver fluctuates randomly. This fluctuation is eventually characterized in the SAR images and forms the multiplicative speckle noise, which causes image distortion [8] and brings difficulty to the subsequent interpretation of SAR images. Hence, the need to reduce the adverse impact of multiplicative speckle noise in current change-detection studies is urgent.

In the traditional change-detection methods, three steps are involved: 1) preprocess the original images; 2) generate the differential image (DI); and 3) analyze the generated DI and obtain the final results. The first step is the basis of the subsequent steps [9], and it includes image registration, radiation correction, and geometric correction, etc. Its main purpose is to effectively reduce the influence of factors such as the light intensity and pixel misalignment on the final experimental results. The second and third steps are the core steps of the change-detection method. In the second step, the ratio and difference operators are two classic techniques [10]. Because the difference operator has high requirements for the imaging quality and registration of SAR images, the most commonly used technique is the ratio operator. In addition, considering the effect of speckle noise, the logarithmic ratio (log-ratio) operator is extensively applied by many scholars recently [11]. For example, Solarna *et al.* [12], [13] apply the log-ratio operator in their change-detection study. Moreover, they propose methods that combine probabilistic modeling, Markov and conditional random fields, graph cuts and message passing inference methods and are able to generate change maps by fusing multiresolution and multimodality SAR data change detection. Garzelli *et al.* [14] optimize the change detection based on log-ratio features via a simple multi-scale preprocessing method, which is compared with classical despeckle filters on a 1-look amplitude SAR dataset. In addition to the above classic methods, some complex test statistics have also been proposed to generate DI in recent years. For example, Bouhlel *et al.* [15] propose a novel determinant ratio test (DRT) statistic for change detection in multilook PolSAR images. In [16], the generalized likelihood ratio test (GLRT) is reviewed, and the explicit expressions for Rao, Wald, Gradient,

Manuscript received August 15, 2021; revised October 13, 2021; accepted November 1, 2021. Date of publication November 10, 2021; date of current version December 8, 2021. This work was supported by the Ministry of Science and Technology of the People's Republic of China, in part by the National Key Research and Development Program under Grant 2017YFB0503001. *(Corresponding author: Fangyu Shen.)*

Fangyu Shen and Chang Liu are with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: shenfangyu17@mails.ucas.ac.cn; cliu@mail.ie.ac.cn).

Yanfei Wang is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: yfwang@mail.ie.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3126839

and Durbin tests are provided. The third step can be used to classify the results obtained by the second step. Its main purpose is to extract changed information from the DI and obtain the final results. Currently, there are two conventional methods in this step: the threshold method [17], [18] and the clustering method [19], [20]. The threshold method contains the expectation maximization (EM) based on the Gaussian distribution [21], and Generalized Kittler–Illingworth (GKI) method based on the generalized Gaussian distribution [22], etc. Although it has high simplicity and speed, it establishes an image-based model and applies the minimum error rate criterion of Bayes to obtain the optimal threshold. The threshold method is highly dependent on the image-based model, and a minor error threshold may cause larger errors in the results. Therefore, the most commonly used method in the third step is the clustering method, which is more flexible and has a higher accuracy in detecting changes. For example, the fuzzy C-means clustering (FCM) algorithm [19] and some improved algorithms based on the FCM, such as the fuzzy local information C-means clustering (FLICM) [23], fuzzy C-means based on Markov random field (MRFFCM) [24], and reformulated fuzzy local information C-means clustering (RFLICM) [20]. Although the traditional methods based on the above three steps can achieve good performance in detecting changes, the results can be improved because the changed information is hard to be acquired from the DI that are affected by the speckle noise, and the step of generating the DI inevitably introduces other noises.

In recent years, neural networks have attracted considerable attention because of their outstanding anti-noise and self-learning ability [25], [26], and many excellent deep learning(DL)-based methods for change detection in SAR images have emerged [27]–[29]. Among them, convolutional neural network (CNN) is widely used because of its unique ability to share local weights. However, it extracts image features layer-by-layer and ignores the feature connections of different layers and receptive fields. DL-based methods can eliminate the DI generation step (the second step mentioned above) and directly obtain the final results while avoiding the noise caused by the generated DI. Nevertheless, they are limited by the requirement of abundant high-precision labels for training. The methods for change detection in SAR images are usually applied to disaster evaluation or environmental monitoring in practical applications, and they are difficult to obtain prior knowledge. In view of this issue, several methods that use a large number of pseudo-labels with higher confidence to train the neural network have been proposed by many scholars. For example, Li *et al.* [8] propose an unsupervised method that uses spatial fuzzy clustering to obtain pseudo-labels and a CNN to detect changes. Gao *et al.* [30] use the FCM algorithm and Gabor wavelets to obtain pseudo-labels and propose a principal component analysis (PCA) network to successfully complete the change detection task. Gong *et al.* [31] propose a novel change detection method based on DL and use a joint classifier based on the FCM to obtain pseudo-labels. Although these methods can achieve better results, they have poor performance in suppressing the speckle noise and they lose many changed details while denoising during the process of obtaining the pseudo-labels. It is the fact that the capability of the neural network is restricted by accumulating the pseudo-labels's errors. So these methods can be further improved.

Table I summarizes the discussion of related work above. As can be seen from it, both the traditional-based and DL-based methods have shortcomings, which need to be improved. Therefore, in this article, we propose a novel unsupervised method based on improved NSST and multi-scale feature fusion CNN to detect changes. The method is an end-to-end structure that can eliminate the step of generating the DI and reduce the effect of noise caused by this step on the final results. It consists of three stages. First, we propose a novel pseudo-label generator based on the improved NSST. NSST is currently the best multi-scale geometric analysis tool originally developed from the wavelet transform (WT) [32]. It can not only represent images sparsely, but also express the lines more flexibly and accurately [33]. In this step, we use the NSST to divide the log-ratio DI into low-frequency part and high-frequency part. In the traditional NSST algorithm, the hard threshold is usually used after this decomposition. Although the traditional NSST algorithm is faster, it does not consider the neighborhood information, which leads to poor results. Therefore, inspired by the self-similar nature of images, we improved the traditional NSST algorithm in this study by adopting the linear enhancement [34] for the low-frequency part and the nonlocal means (NLM) filter for the high-frequency part. This is because the low-frequency part of the image contains its core information, whereas noise is mostly adhered to the high-frequency part. The NLM filter is an edge-preserving filter which can preserve more detailed information while denoising [35]. Then, we use the FCM algorithm to classify the DI. Following the above process, we can effectively retain more details of the DI while suppressing the noise and obtain more pseudo-labels with higher confidence. Second, we select the appropriate samples and build an easy-training network based on the multi-scale feature fusion. The network can integrate the spatial features of different receptive fields and different layer information to enhance the feature learning ability. Additionally, considering the size of the input samples, the number of pooling layers of the network is reduced to avoid losing the details of the feature images. Finally, the trained network is used to classify the original SAR images into two regions: changed and unchanged. To verify the superiority of our proposed method, we first compare our method with the supervised method. The experimental and statistical results demonstrate that our method can perform well for the change-detection tasks that lack prior knowledge and labeled data. After that, we select four change-detection methods, i.e., GKI method [36], neighborhood-based ratio (NR)-FCM method [37], PCA-Net method [30], and spatial fuzzy clustering(SFCM)-CNN method [8] to compare with our method. The experimental results show that the highest accuracy is acquired by our proposed method. Finally, for verifying the generality of our method, we apply it to the cross-region dataset, and the satisfactory results are also obtained.

There are two main contributions in this article.

1) For the current critical problem that the DL-based methods for change detection are limited by the requirement of abundant high-precision labels for training, we first

TABLE I
SUMMARY OF THE ABOVE RELATED WORK DISCUSSION

Method group	Technique		Implement in	Characteristics
Traditional-based (three steps)	Preprocessing		/	Difficult to obtain the changed information from the generated DI affected by the speckle noise. The step of generating the DI inevitably introduces other noises.
	Generate the DI	Ratio, difference, log-ratio, etc.	[11] [12] [13] [14]	
		Test statistic	[15] [16]	
	Analyze the DI	Threshold	[21] [22]	
		Clustering	[19] [23] [24] [20]	
Deep learning-based	Learn image features through the neural network		[8] [30] [31]	Show outstanding anti-noise ability but limited by the requirement of abundant high-precision labels for training.

propose a novel pseudo-label generator based on the improved NSST. In the proposed generator, the hard threshold method of both the low and high frequency part in the traditional NSST algorithm are replaced by the linear enhancement method and the NLM filter, respectively. This improvement can solve the problem that the traditional NSST method loses a lot of changed details while effectively suppressing the speckle noise. Experiment results show that the novel pseudo-label generator is effective and can obtain a large number of pseudo-labels with higher confidence for training.

2) For the problem that the CNN extracts image features layer-by-layer and ignores the feature connections in the different layers and different receptive fields, we build an easy-training network with high change-detection accuracy based on the multi-scale feature fusion. The network integrates the spatial features of different receptive fields and different layer information to make the feature images contain more complete information and reduces the number of pooling layers to avoid losing some feature image details.

The rest of this article is organized as follows. Section II shows the whole structure of the proposed method. Section III makes the introduction of the basic knowledge and significant ideas of the proposed method. The descriptions of the experimental datasets and the final results are drawn Section IV. The discussion of the experimental results is shown in Section V. Finally, Section VI concludes this article.

Notation: I_1 and I_2 denote two original images; LR_DI denotes the generated log-ratio DI; N_{iter} denotes the number of decomposition layers in the proposed novel pseudo-label generator; I_L denotes the low-frequency component of the image, and I_H denotes the high-frequency component of the image; I_{nsst} denotes the image to be processed by the improved NSST; $DI_{denoise}$ denotes the output image of the proposed pseudo-label generator. The synthetic expansion affine system is defined by $SE_{WV}(\varphi)$; $\varphi \in L^2(R^2)$ denotes the element in the synthetic expansion affine system, and L denotes the integrable space; W denotes the matrix of anisotropic expansion, and V denotes the matrix related to geometric transformation; $SE_{a,s,t}(\varphi)$ denotes the shearlet wave. $\hat{\varphi}(\zeta)$ denotes the Fourier transform of $\varphi(\zeta)$; D_0 denotes any point in the frequency domain; cof_{min} denotes the minimum value of the subband coefficients, and cof_{max} denotes the maximum value of the subband coefficients;

Algorithm 1: Pseudo-Label Generator Based on Improved NSST.

Input:
Set the number of decomposition layers, N_{iter};
Two original SAR images, I_1 and I_2;
Output:
1: Apply the log-ratio operator on the two input images to generate the DI;
2: Set DI as I_{nsst};
3: Initialize k =1;
4: **repeat**
5: Decompose I_{nsst} into low-frequency component I_L and high-frequency component I_H by NSLP;
6: Decompose I_H into directional detail subbands;
7: $I_{nsst} \leftarrow I_L$;
8: k=k + 1
9: **until** $k > N_{iter}$
10: Apply the linear enhancement to I_L;
11: Apply the NLM filter to I_H;
12: Perform the inverse NSST on I_H and I_L to obtain the $DI_{denoise}$;
13: Classify the $DI_{denoise}$ by the FCM algorithm.

$U(a, b)$ denotes the similarity between the pixel values at the positions of a and b; $L(I_{ij})$ denotes the label corresponding to the pixel I_{ij}, and N_{ij} denotes the neighborhood area of I_{ij}; $E(L(I_{ij}), L(N_{ij}))$ denotes the number of pixels in N_{ij} whose label is equal to I_{ij}.

II. OVERALL FRAMEWORK

Suppose that I_1 and I_2, which have the same size $P \times Q$, are two SAR images obtained from the same area but at different times. The binary image representing the changed information between I_1 and I_2 is our ultimate goal in this article.

The overall framework of this article is shown in Fig. 1. There are five steps: 1) generate the pseudo-labels of higher confidence by the novel pseudo-label generator; 2) obtain the training samples; 3) choose the reliable training samples; 4) train the CNN based on multi-scale feature fusion; 5) feed the two original images into the trained CNN and obtain the final results.

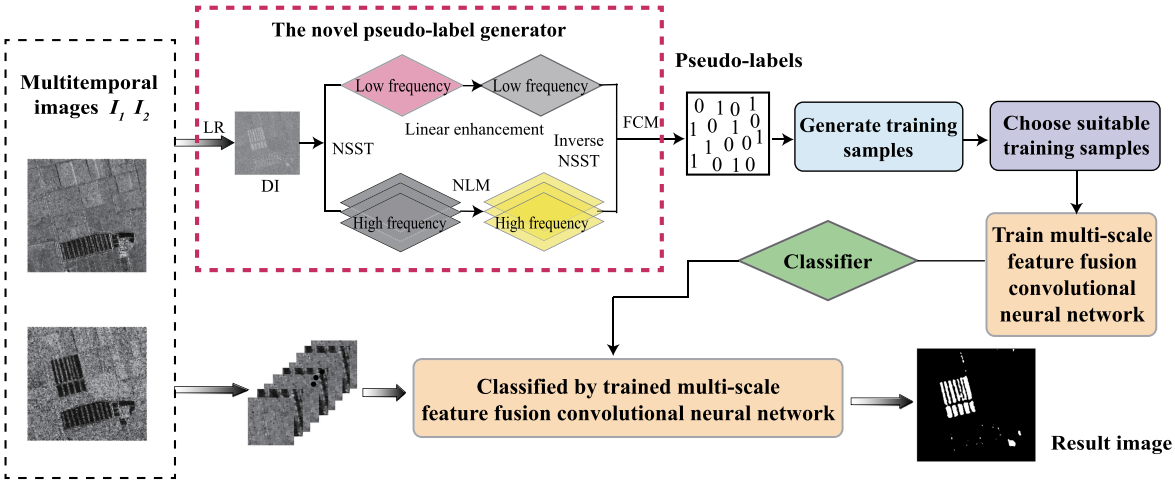

Fig. 1. Overall framework of our proposed method.

III. Methodology

A. Pseudo-labels Generation

For the problem that the change-detection methods in SAR images have difficulty in obtaining abundant true labels for training, the concept of pseudo-label is proposed. The pseudo-labels are used to replace the true labels when training the neural network, and they are the same as the true labels, except that their accuracy is not as good as that of true labels. In general, the more accurate the pseudo-labels are, the better change-detection results are. Therefore, we improve the traditional NSST algorithm and propose a novel pseudo-label generator in this article for obtaining more pseudo-labels with higher confidence. Algorithm 1 shows its specific steps. First, we use the following formula to get the log-ratio DI:

$$LR\_DI = |\ln \frac{I_1 + eps}{I_2 + eps}| \tag{1}$$

where the I_t represents the image to be detected at the time t, the eps is used to avoid the pixel values in the two images that are zero, and the LR_DI represents the generated log-ratio DI.

Second, since the generated DI has most of the noise, we choose the superior multi-scale analysis tool named NSST to make an improvement for suppressing the noise in the log-ratio DI. Let us give the detailed algorithm of the novel pseudo-label generator based on the improved NSST algorithm.

1) Traditional NSST Algorithm: The NSST algorithm is designed on the idea of synthetic expansion affine system, which is an improvement of the shearlet transform (ST). As in 2-D plane space, the synthetic expansion affine system can be defined by [38]

$$SE_{WV}(\varphi) = \{\varphi_{a,b,c}(\chi) = |det(W)|^{a/2}\varphi(V^b W^a \chi - c)\} \tag{2}$$

where $\varphi \in L^2(R^2)$ and L represents the integrable space. a, b and c represent the scale variable, the shear variable and the translation variable, respectively. $a, b \in Z$, $c \in Z^2$. W represents the matrix of anisotropic expansion, which is related to scale transformation. V represents the matrix related to geometric transformation, such as rotation and cutting.

Suppose that $N_{a,s} = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & \sqrt{a} \end{bmatrix} = \begin{bmatrix} a & \sqrt{a}s \\ 0 & \sqrt{a} \end{bmatrix}$, $N_{a,s}$ is the combined form of the matrix W and V, $t \in R^2$, and $(a, s) \in R^+ \times R$ satisfies the following affine system:

$$SE_{N_{a,s}}(\varphi) = SE_{a,s,t}(\varphi) = \{\varphi_{a,s,t}(\chi) a^{\frac{-3}{4}} (N_{a,s}^{-1}(\chi - t))\}. \tag{3}$$

Then $SE_{a,s,t}(\varphi)$ is called the shearlet wave. For any point $\forall \zeta = (\zeta_1, \zeta_2) \in \hat{R}^2$ in the domain of frequency, we define

$$\hat{\varphi}(\zeta) = \hat{\varphi}(\zeta_1, \zeta_2) = \hat{\varphi}_1(\zeta_1)\hat{\varphi}_2(\frac{\zeta_2}{\zeta_1}) \tag{4}$$

where $\hat{\varphi}(\zeta)$ is the Fourier transform of $\varphi(\zeta)$. $\hat{\varphi}_1$ and $\hat{\varphi}_2$ are wavelets satisfying the condition that $supp\hat{\varphi}_1 \in [-\frac{1}{2}, -\frac{1}{16}] \cup [\frac{1}{16}, \frac{1}{2}]$ and $supp\hat{\varphi}_2 \in [-1,1]$. On the assumption that

$$\begin{cases} \sum\limits_{a \geq 0} |\hat{\varphi}_1(2^{-2a}\omega)|^2 = 1 \\ |\omega| \geq \frac{1}{8}. \end{cases} \tag{5}$$

For $a \geq 0$, we can get the following:

$$\begin{aligned} &\sum_{a \geq 0} \sum_{b=-2^a}^{2^a - 1} |\hat{\varphi}(\zeta W^{-a} V^{-b})|^2 \\ &= \sum_{a \geq 0} \sum_{b=-2^a}^{2^a - 1} |\hat{\varphi}_1(2^{-2a}\zeta_1)|^2 |\hat{\varphi}_2(2^a \zeta_2/\zeta_1) - b|^2 = 1 \end{aligned} \tag{6}$$

where $D_0 = \{(\zeta_1, \zeta_2) \in \hat{R}^2, |\zeta_1| \geq \frac{1}{8}, |\frac{\zeta_2}{\zeta_1}| \leq 1\}$, and the formula $\{\hat{\varphi}(\zeta W^{-a} V^{-b})\}$ is a decomposition distribution of D_0 in the frequency domain.

As can be seen from the representation of D_0, the NSST has a high degree of directional sensitivity. Fig. 2 represents the process of its decomposition. It uses the non-subsampled Laplacian pyramid transform (NSLP) to achieve translation

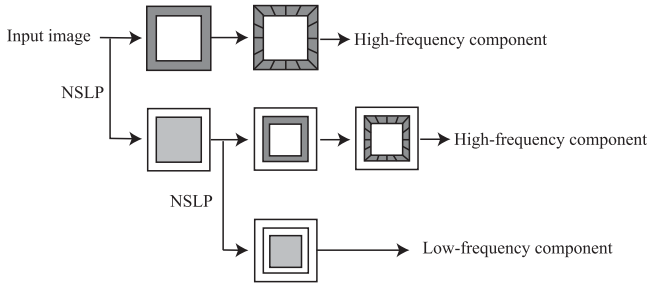

Fig. 2. Decomposition process of NSST.

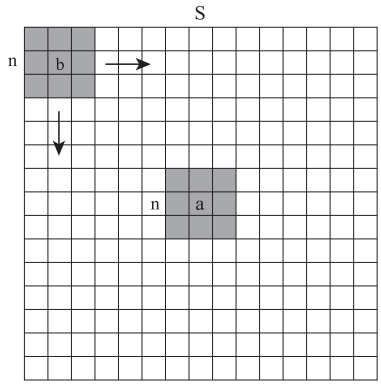

Fig. 3. NLM filter execution diagram.

invariance [39]. The specific process of the NSST is shown as follows.

1) Use the NSLP to divide the input image into the low-frequency and high-frequency parts. With the refinement of the decomposition layer, the low-frequency part is used as the input of the NSST again.

2) Use the directional filter to decompose the high-frequency part in order to achieve multidirectional decomposition. The specific steps are as follows. First, map the pseudopolarization coordinates back to Cartesian coordinates. Second, construct the window function and forming the directional filter. Last, get the high-frequency directional subbands.

2) Novel Pseudo-label Generator Based on the Improved NSST Algorithm: In the traditional NSST algorithm, the hard threshold usually becomes an usage after dividing the image into high-frequency and low-frequency parts. Although the speed is faster, it does not take into account the neighborhood information, which leads to poor results. Therefore, in this article, inspired by the self-similar nature of images, we improved the traditional NSST algorithm. We adopt the linear enhancement for the low-frequency part and the NLM filter for the high-frequency part.

The low-frequency part of the image represents its core information, which is the comprehensive evaluation of image intensity. Therefore, in this article, we adopt the linear enhancement for the low-frequency part. The calculation formula is as follows:

$$L(cof) = 255(cof - cof_{min})/(cof_{max} - cof_{min}) \quad (7)$$

where cof_{min} is the minimum value of the subband coefficients, and cof_{max} is the maximum value of the subband coefficients.

The high-frequency part mainly contains the edge information and the noise of the image. The traditional NSST algorithm must lose a lot of detailed characteristics while processing high-frequency part. Therefore, in this article, we adopt the NLM filter for the high-frequency part. The NLM filter considers the self-similarity of the image and takes full advantage of the image's redundant information. It can effectively suppress the noise while maintaining the most of detailed information of the image. The core idea of it is to replace the actual value with a similar patch information. In general, the NLM filter needs to obtain the similarity among all pixels in the image and the current pixel, so the two windows are set up, including the large searching window $(S \times S)$ centered on the target pixel and the small neighborhood window $(n \times n)$, as shown in Fig. 3.

The two small gray windows in the figure are neighborhood windows centered on a and b. Assume that I_n is the noisy image and I_{dn} is the denoised image. The calculation formula of the value at the position a in the image I_{dn} is shown as follows:

$$I_{dn}(a) = \sum_{b \in I} U(a,b) I_n(b) \quad (8)$$

where $U(a,b)$ is the similarity between the pixel values at the positions a and b. It can be computed as follows:

$$U(a,b) = \frac{1}{Z(a)} exp\left(-\frac{\|N(a) - N(b)\|^2}{h^2}\right) \quad (9)$$

$$\|N(a) - N(b)\|^2 = \frac{1}{n^2} \sum_{z \in S_n} \|I_n(a+z) - I_n(b+z)\|^2 \quad (10)$$

where $N(a)$ and $N(b)$ represent the neighborhood area centered on the positions a and b. Z is the normalization constant. h is associated with the noise variance of the image.

After using the improved NSST algorithm to process the log-ratio DI, we can get a new image $DI_{denoise}$. Finally, we perform the FCM algorithm on the $DI_{denoise}$ and obtain the pseudo-label map which can generate a large number of pseudo-labels with higher confidence.

B. Training Samples Generation

In most cases, the original images used for change detection are too large to be directly fed to the network for training. Therefore, in this study, we divide the two original images into several smaller image blocks of the same size and superimpose the smaller images located in the corresponding position to form a group of training samples.

In details, suppose that I_{ij}^1 and I_{ij}^2 in Fig. 4 represent the pixel values on the position (i,j) of the first and second SAR images, respectively. If we set the size of each image to $P \times Q$ and the neighborhood size of each pixel to $k \times k$, we can obtain $P \times Q$ numbers of smaller image blocks for each image, with each block having a size of $k \times k$. Then we superimpose the two smaller image blocks from the corresponding position (i.e., I_{ij}^1 and I_{ij}^2) to form a training sample, as shown in Fig. 5. Note that for the pixels on the edge that have no neighboring pixels, we use 0 around them to supplement smaller image block. After all

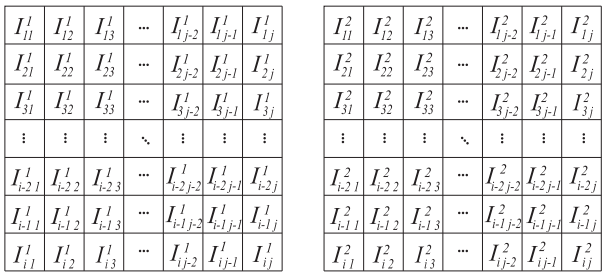

Fig. 4. (a) SAR image in the first time. (b) SAR image in the second time.

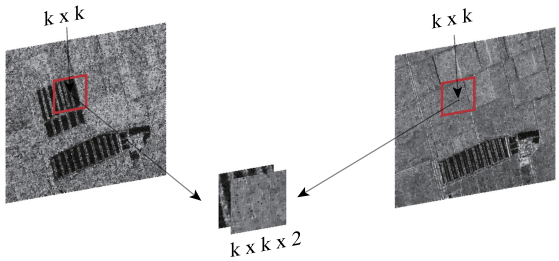

Fig. 5. Rule of generating the samples.

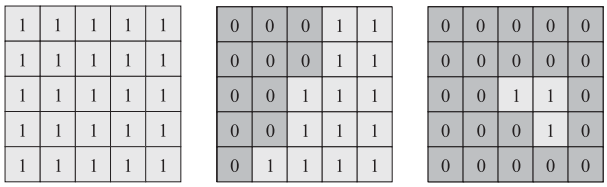

Fig. 6. Center pixel and its neighborhood pixel. (a) Similar label with all neighborhood pixels. (b) Similar label with most neighborhood pixels. (c) Similar label with a few neighborhood pixels.

the corresponding smaller image blocks from the two original image are superimposed, we can obtain $P \times Q$ training samples and the size of each sample is $k \times k$.

C. Reliable Samples Selection

Because the labels obtained by the proposed pseudo-label generator are not completely accurate, we need to select the training samples that are more likely to be correctly classified to train the neural network. Suppose that 0 and 1 represent two types of labels. Fig. 6 shows three situations that need to be considered when selecting suitable training samples. The pixels in the first situation are located in the changed or unchanged areas, as shown in Fig. 6(a). These pixels can be selected as training samples in this study. The pixels in the second situation are located at the edge of the image. This situation is complicated, and the labels of the center pixel and its neighborhood pixels should be considered. As shown in Fig. 6(b), if the center pixel has neighborhood pixels in which approximately half of the pixels have the same labels as it, this type of pixel can be selected as training samples. The pixels in the third situation are located in the noise area, as shown in Fig. 6(c). In this case, the center pixel has a neighborhood in which a few pixels have the same labels as it. Thus, this type of pixel needs to be discarded and cannot be chosen as training samples. All in all, the selection rules can be described by the following formula:

$$\frac{E(L(I_{ij}), L(N_{ij}))}{n \times n} \geq \eta \tag{11}$$

where $L(I_{ij})$ represents the label corresponding to pixel I_{ij}, and it is obtained by the novel pseudo-label generator above. N_{ij} represents the neighborhood area of I_{ij}. n is set to 3 in this article. E is used to determine the number of pixels in N_{ij} whose label is equal to I_{ij}. η is a parameter, which is more import in choosing suitable training samples. If η is set to be too small, then the result is less robust to noise, and if η is set to be too large, then the samples will lack of diversity. This kind of choosing idea is effective and has been applied by many scholars in the related research works [31].

D. Multi-scale Feature Fusion CNN Establishment

Training the classification network is an important part of the entire method. As a type of multilayer neural network specially designed to process 2-D data, CNN has been successfully used by many scholars to detect changes in SAR images. In this article, we present an easy-training but powerful neural network based on the multi-scale feature fusion.

First, considering the size of the training samples, we use only one pooling layer in the network. This is because too many pooling layers will cause the feature images in the CNN to be smaller, thereby losing a lot of details. Thus, one pooling layer is sufficient to meet the requirements of our study.

Second, we design a multi-scale feature fusion block in the CNN. The image features extracted by the convolutional kernels of different scales are different. A convolutional block with a larger kernel size can capture relatively general characteristics, whereas that with a small kernel can capture detailed features more effectively. In addition, the image features extracted from the different layers are also different. To make the feature images contain more complete information, we design a group of parallel convolutional blocks in the third layer with the kernel sizes of 1×1, 3×3 and 5×5, and then apply the concatenation operation on the first-layer results and them. Moreover, the 1×1 convolutional layer is used in the fifth layer to achieve cross-channel interaction and information integration. Fig. 7 shows the architecture of the proposed network, where "MP-2" indicates the size of the max-pooling kernel is 2×2 and the number n in "Conv-n" indicates the size of the convolutional kernel is $n \times n$.

The proposed network contains eight layers, except for the input and output layers. The size of the input layer is 11×11 pixels. The first and second layers are the convolutional layers, which have 16 and 32 filters, respectively. The third layer comprises several parallel convolutional blocks, each of which has 32 filters. The fourth layer is the convolutional layer, which has 64 filters. The fifth layer is the max-pooling layer, which has 64 filters.

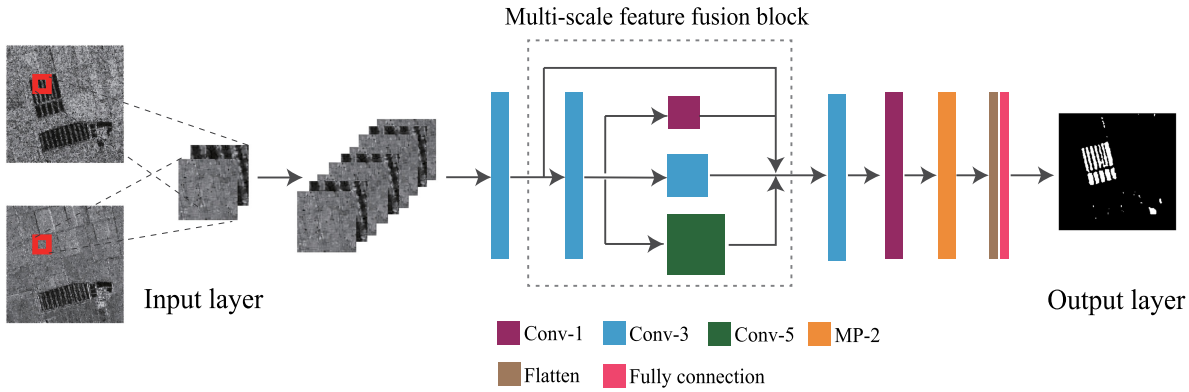

Fig. 7. Architecture of the proposed neural network.

Furthermore, the proposed network is an end-to-end framework. Because the proposed novel pseudo-label generator can generate a large number of high-accuracy pseudo-labels for training and the proposed network architecture is not particularly large, no overfitting phenomenon occurs in the actual training of our network. The trained network is obtained directly from suitable samples and their corresponding labels, and each branch of the network is not pretrained. The specific training steps are as follows. 1) Suitable training samples are fed to the input layer and the forward propagation stage begins. After the first convolutional layer, the results are sent to the multi-scale feature fusion block and the next layers in turn to obtain the network output. 2) The error between the output and the labels is calculated. If the error meets our expectations, a trained network is obtained. Otherwise, the back-propagation stage begins and the weights are updated. Afterward, the network re-enters the forward propagation stage.

After the network training is completed, we feed all the samples obtained from the two original images into the trained network for classification and obtain the final result image with only black and white colors. The white color represents the changed area, whereas the black color represents the unchanged area.

IV. Results

A. Description of Real SAR Image Datasets

1) Ottawa Dataset: The first dataset is obtained by Radarsat at the region of Canada in May 1997 and August 1997, named the Ottawa dataset. Fig. 8(a) and (b) represents their two original SAR images. The size of images is 290×350. The main factor causing the changed areas in this dataset is the flood disaster. Fig. 8(c) is the ground truth representing its changed area. The changed area of this dataset is relatively large.

2) Bern Dataset: The second dataset is obtained by ERS-2 at the region of Switzerland in April 1999 and May 1999, named the Bern dataset. Fig. 9(a) and (b) represents their two original SAR images. The size of images is 301×301. The main factor causing the changed areas in this dataset is also the flood disaster. Fig. 9(c) is the ground truth representing its changed area. The changed area of this dataset is relatively small.

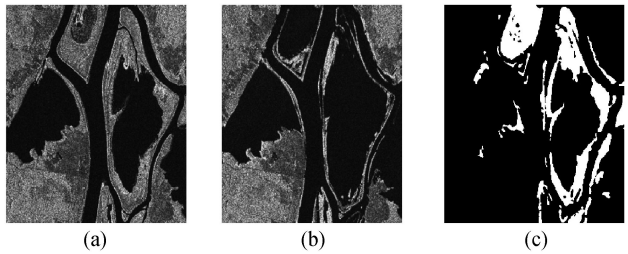

Fig. 8. Ottawa dataset. (a) May 1997. (b) August 1997. (c) Reference image.

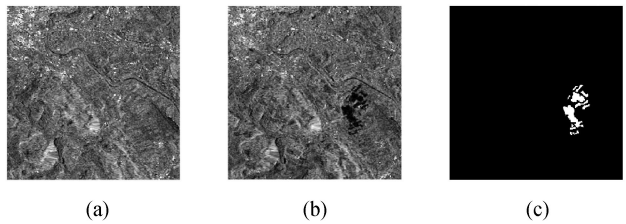

Fig. 9. Bern dataset. (a) April 1999. (b) May 1999. (c) Reference image.

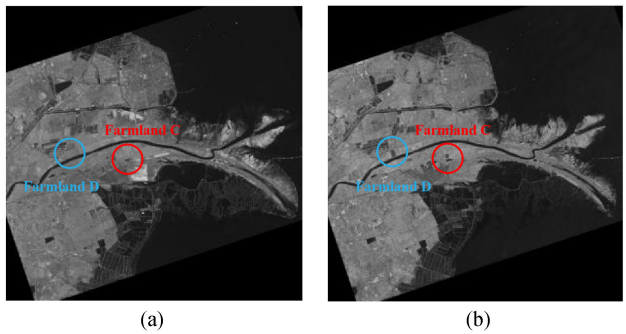

Fig. 10. Yellow River Estuary dataset. (a) June 2008. (b) June 2009.

3) Yellow River Estuary Dataset: The third dataset is obtained by Radarsat-2 at the region of China in June 2008 and June 2009, named the Yellow River Estuary dataset. The two original SAR images are shown in Fig. 10(a) and (b). The size of the original images is 7666×7692, which is too large to represent the detail of the original images. Therefore, we choose two

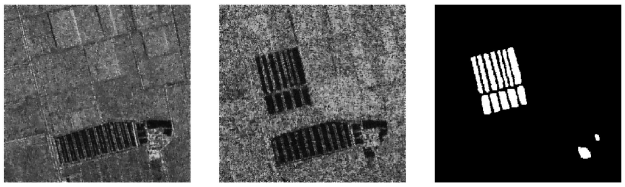

Fig. 11. Farmland C dataset. (a) June 2008. (b) June 2009. (c) Reference image.

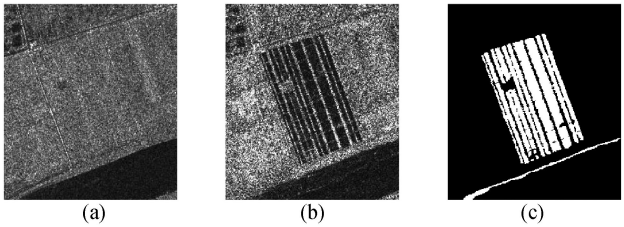

Fig. 12. Farmland D dataset. (a) June 2008. (b) June 2009. (c) Reference image.

typical areas of this dataset where the changed area is relatively irregular. The first area is the Farmland C dataset with a size of 306×291, as shown in Fig. 11. The other area is the Farmland D dataset, which has a size of 257×289, as shown in Fig. 12. The changed information in the Yellow River Estuary dataset is mainly caused by environmental monitoring rather than natural disasters. It should be noted that one of the Yellow River Estuary dataet is a single-look image, and the other is a four-look image. This situation causes serious inconsistencies between the noise degrees of the two original images. Therefore, this dataset can better reflect the robustness to noise of the proposed method.

B. Validation and Parameter Setting

This article employs five evaluation parameters to quantitatively analyze the change-detection results: kappa coefficient, overall error (OE), false positive (FP), false negative (FN), and percentage correct classification (PCC). The FP stands for the quantity of pixels whose detection results are changed but actually unchanged. The FN stands for the quantity of pixels whose detection results are unchanged but actually changed. The OE represents the total quantity of pixels that were incorrectly detected. The kappa coefficient is an indicator of accuracy evaluation. The larger the kappa coefficient is, the higher accuracy the change detection is. The kappa coefficient can be computed as follows:

$$kappa = \frac{p_i - p_j}{1 - p_j} \tag{12}$$

where p_i is equal to the PCC. Let TP stand for the quantity of pixels whose detection results are changed and actually changed. Let TN stand for the quantity of pixels whose detection results are unchanged and actually unchanged. Let $P \times Q$ be the total quantity of image pixels. p_i and p_j can be computed as follows:

$$p_i = \frac{TP + TN}{P * Q} \tag{13}$$

$$p_j = \frac{(TP + FP) * (TP + FN) + (FN + TN) * (TN + FP)}{P^2 * Q^2}. \tag{14}$$

Fig. 13. Results of the Bern dataset obtained by (a) supervised method. (b) Proposed method.

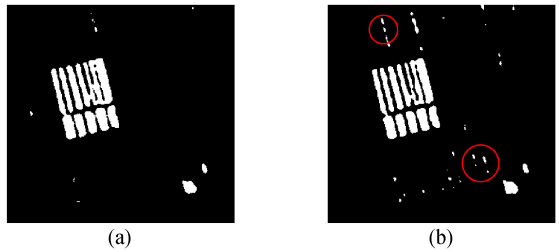

Fig. 14. Results of the Farmland C dataset obtained by (a) supervised method. (b) Proposed method.

In this article, the decomposition layer of the improved NSST algorithm is set to 2, and the numbers of directional subbands of each scale decomposition are set to 4 and 8, respectively. The parameter η controls the number of samples that are fed to the network. In this article, we set it to 0.45.

C. Experiments and Analysis

1) Results Under Proposed Method and Supervised Method: In the current change-detection field, most application tasks are focused on detecting natural disasters. Among them, tasks lack prior knowledge and labeled data. Therefore, in this study, we compare our method with the supervised method to demonstrate the performance of our method in the application engineering. The training and testing of these two methods are conducted on two neural networks with the same structure. However, the labels of these two methods are different. In our method, the labels are given by the novel pseudo-label generator. In the supervised method, the labels are given by the ground truth. We conducted experiments on two real datasets, and Figs. 13–14 show the experimental results. The figures show that the proposed method can detect the most changed area, and its result is generally close to that of the supervised method. However, owing to the inaccuracy of its labels, there are some missed or false detections in the proposed method, as indicated by the red circle in the figures. Table II represents the quantitative results of this experiment.

TABLE II
EVALUATION RESULTS OF THE BERN DATASET AND FARMLAND C DATASET OBTAINED BY SUPERVISED METHOD AND PROPOSED METHOD

Data set	Methods	FP(%)	FN(%)	OE(%)	PCC(%)	Kappa(%)
Bern	Supervised	0.11	0.17	0.28	99.72	88.45
	Proposed method	0.14	0.19	0.33	99.67	86.62
	Pseudo-label	0.16	0.24	0.40	99.60	85.23
Farmland C	Supervised	0.42	0.44	0.86	99.14	92.28
	Proposed method	0.61	0.68	1.29	98.71	88.35
	Pseudo-label	1.10	0.67	1.77	98.23	84.59

TABLE III
EVALUATION RESULTS OF DIFFERENT METHODS ON THE OTTAWA DATASET

Data set	Methods	FP(%)	FN(%)	OE(%)	PCC(%)	Kappa(%)
Ottawa	GKI	0.86	2.13	2.99	97.01	88.45
	NR-FCM	0.79	1.70	2.49	97.51	90.43
	PCA-Net	0.85	1.09	1.94	98.06	92.44
	SFCM-CNN	0.66	0.71	1.37	98.63	94.84
	Pseudo-label	0.88	1.24	2.12	97.88	91.97
	Proposed Method	0.47	0.84	1.31	98.69	95.02

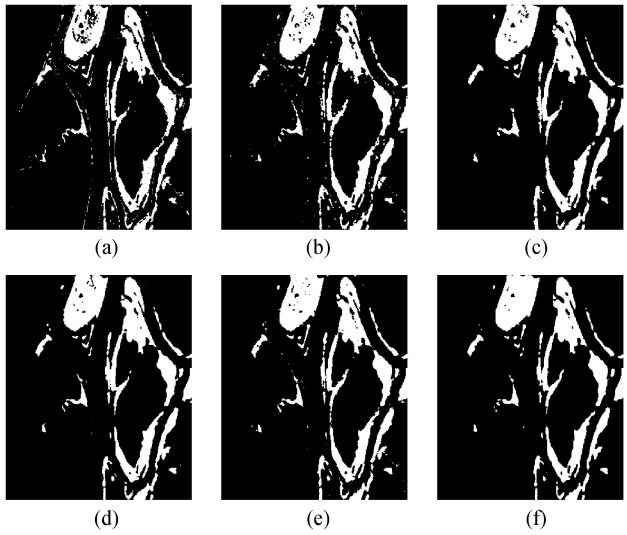

Fig. 15. Ottawa dataset, (a) GKI method. (b) NR-FCM method. (c) PCA-Net method. (d) SFCM-CNN method. (e) Pseudo-label method. (f) Proposed method.

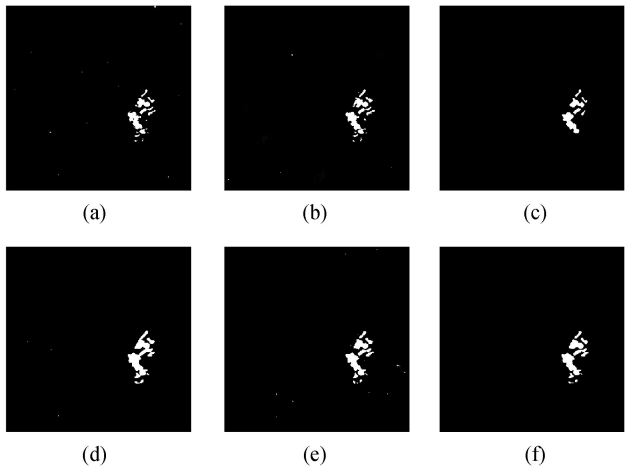

Fig. 16. Bern dataset: (a) GKI method. (b) NR-FCM method. (c) PCA-Net method. (d) SFCM-CNN method. (e) Pseudo-label method. (f) Proposed method.

In the Table II, we also give the quantitative evaluation results obtained by the proposed novel pseudo-label generator without the DL-based step. As we can see from it, the proposed novel pseudo-label generator performs well, and our proposed method is effective for the tasks lacking prior knowledge and labeled data.

2) Results on Real SAR Image Datasets: In order to evaluate the authenticity and efficiency of the proposed method, we compare our method with the GKI method [36], NR-FCM method [37], PCA-Net method [30], and SFCM-CNN method [8] on the three real datasets. Meanwhile, we evaluate the change-detection results of the proposed pseudo-label generator without the DL-based step on the three real datasets to demonstrate the good performance of the proposed generator and the contribution of the DL-based step to the results. The range of the changed regions and the degree of noise interference on the original image of the three real datasets are not the same. The changed areas of the Ottawa dataset are large, the changed areas of the Bern dataset are small, and the changed areas of the Farmland C dataset are irregular. The GKI method is a threshold method, and the NR-FCM method is a clustering method. The NR method is put forward by Gong *et al.* [37], and the experiments results have proved that it is superior to the log-ratio method. In this article, we combine the NR method with the FCM method to form a comparison method. The PCANet method and the SFCM-CNN method are two DL-based methods proposed in recent years.

a) Ottawa Dataset: The final results on the Ottawa dataset are represented in Fig. 15 and Table III. The changed regions in this dataset are very wide. As shown in Table III, GKI achieves the worst performance owing to the inaccuracy of the threshold selection, and the upper part of Fig. 15(a) shows that it contains more noise. NR-FCM performed better than GKI, but there is still some noise, as shown in the upper part of Fig. 15(b). The FN of the above two methods are higher, and there are many missed detections in these two methods. PCA-Net and SFCM-CNN are two DL-based methods for change detection. Owing to their good ability to learn features, they have a stronger anti-noise ability than the above methods. Therefore, as shown in Table III, their results are competitive. However, the pseudo-labels obtained by these two methods are not as accurate as the proposed method, so their results are lower than the propose method. In addition, as presented in Table III and Fig. 15(e), the PCC obtained by our proposed method is 0.81% higher than the pseudo-label generator, which illustrates the good performance of our proposed network structure.

b) Bern Dataset: The final results on the Bern dataset are represented in Fig. 16 and Table IV. Because the changed regions of this dataset are concentrated and small, all the methods tend to achieve high-accuracy results. As shown in Fig. 16, PCA-Net has the worst performance among all the methods, as it lost

TABLE IV
EVALUATION RESULTS OF DIFFERENT METHODS ON THE BERN DATASET

Data set	Methods	FP(%)	FN(%)	OE(%)	PCC(%)	Kappa(%)
Bern	GKI	0.06	0.38	0.44	99.56	80.37
	NR-FCM	0.09	0.33	0.42	99.58	82.08
	PCA-Net	0.06	0.48	0.54	99.46	74.47
	SFCM-CNN	0.24	0.15	0.39	99.61	85.37
	Pseudo-label	0.16	0.24	0.40	99.60	85.23
	Proposed Method	0.14	0.19	0.33	99.67	86.62

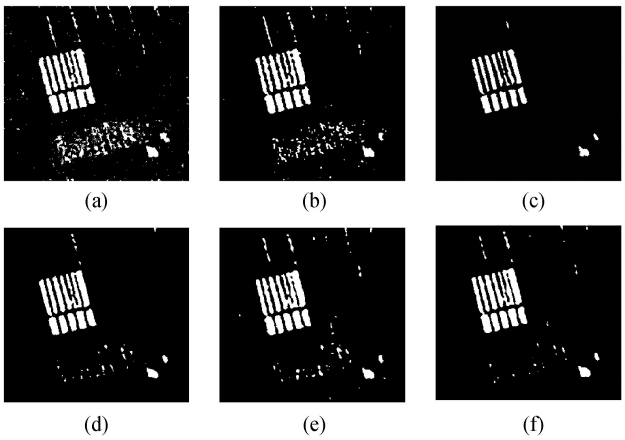

Fig. 17. Farmland C dataset. (a) GKI method. (b) NR-FCM method. (c) PCA-Net method. (d) SFCM-CNN method. (e) Pseudo-label method. (f) Proposed method.

TABLE V
EVALUATION RESULTS OF DIFFERENT METHODS ON THE FARMLAND C DATASET

Data set	Methods	FP(%)	FN(%)	OE(%)	PCC(%)	Kappa(%)
Farmland C	GKI	3.23	0.72	3.95	96.05	70.39
	NR-FCM	1.66	0.95	2.61	97.39	77.77
	PCA-Net	0.10	1.36	1.46	98.54	85.48
	SFCM-CNN	0.62	0.81	1.43	98.57	87.03
	Pseudo-label	1.10	0.67	1.77	98.23	84.59
	Proposed Method	0.61	0.68	1.29	98.71	88.35

most of the details of the changed areas. The remaining five methods have slight difference in performance on this dataset, and the accuracy rate of all methods can reach more than 99.5%, as shown in Table IV. However, the proposed method has an excellent pseudo-label generator and network structure with a stronger feature learning ability. The PCC of the proposed pseudo-label generator reaches 99.60%, and the proposed network structure improves the PCC by 0.07%. Therefore, the proposed method achieves the best results.

c) Farmland C Dataset: The final results on the Farmland C dataset are represented in Fig. 17 and Table V. The two images of this dataset are not disturbed by the same degree of noise. Therefore, this dataset can better reflect the robustness to noise of various methods. As shown in Fig. 17 and Table V, GKI and NR-FCM perform not well. These two methods do not perform special processing on the speckle noise, so they show poor performance when facing a scene such as the Farmland C dataset. Fig. 17(a) and (b) shows that the result of these two

Fig. 18. Farmland C dataset. (a) Results on individual dataset. (b) Results on cross-region dataset.

TABLE VI
EVALUATION RESULTS OF THE PROPOSED METHOD ON INDIVIDUAL DATASET AND CROSS-REGION DATASET

Data set	Methods	FP(%)	FN(%)	OE(%)	PCC(%)	Kappa(%)
Farmland C	Individual dataset	0.61	0.68	1.29	98.71	88.35
	Cross-region dataset	1.01	0.49	1.50	98.50	87.06

methods contain more speckle noise. PCA-Net, SFCM-CNN, and the proposed method are more suitable to deal with this type of dataset because of their powerful ability to learn image features. Therefore, as shown in Table V, their results are competitive. However, PCA-Net has a serious imbalance between the values of FP and FN, and its missed detection is too high. The pseudo-labels obtained by SFCM-CNN is not as accurate as the proposed method, and its ability to learn image features is not as good as the proposed method. So their performance is worse than the proposed method. In this dataset, the PCC of using the DL-based step in our method is 0.48% higher than that without the step.

Through the above analysis, we conclude that our method is applicable and competitive regardless of whether the changed regions are small (Bern dataset), large (Ottawa dataset), or irregular (Farmland C dataset).

3) Results on Cross-Region Dataset: The above performances on the individual dataset which the trained and the tested dataset are the same have proved that the proposed method is more competitive than the comparison methods. In this section, to prove the generalization of our method, we apply it on the cross-region dataset. In this case, the neural network is trained on the Farmland D dataset and tested on the Farmland C dataset. These two datasets all belong to the Yellow River Estuary dataset, and the characteristics of the images are similar. Fig. 18 and Table VI represent the results. As shown in Fig. 18, the proposed method based on the cross-region dataset also achieves a good performance. As presented in Table VI, its accuracy is only 0.21% lower than that of the individual dataset. Therefore, in the practical engineering, if the training model of the area to be detected cannot be obtained for some reasons, we can still achieve a good performance by using the trained model of the neighborhood area obtained by our proposed method.

V. DISCUSSION

In this section, we conduct experiments on the Farmland C dataset to discuss the parameters affecting the accuracy of the final results: the samples selection parameter η, the size of

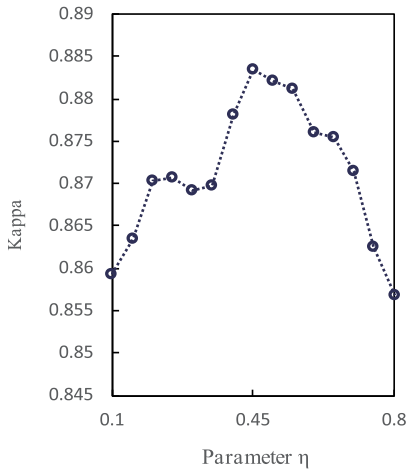

Fig. 19. Relationship between the kappa coefficient and the parameter η on the Farmland C dataset.

samples, and the multi-scale feature fusion block. In addition, we discuss the computational complexity involved in each step constituting the proposed method.

A. Influence of the Samples Selection Parameter η

In the step of selecting appropriate pseudo-labels for training, the parameter η is very important. If η is set too small, the result is poorly robust to noise, and if η is set too high, the samples lack diversity. In this article, we set η to 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, and 0.8 for testing. Fig. 19 shows the relationship between the parameter η and the kappa coefficient. As shown in Fig. 19, the kappa coefficient can obtain good results when η is between 0.45 and 0.55. When the parameter η is set too high or too low, the kappa coefficient becomes worse. In this article, we set the parameter η to 0.45, which can make the kappa coefficient obtain the highest value of 88.35%.

B. Influence of the Size of Samples

In the step of generating the training samples, we can obtain a group of training samples of size $k \times k$ according to the two original images. The value of k has an important effect on the final experimental results. In this article, the proposed network based on the multi-scale feature fusion has only eight layers, so our proposed method is more suitable for the case that the value range of k is small. When the value of k increases, the performance of our method will decrease because the network can not learn by more complex and rich features. In addition, the larger value of k may make the network fail to converge. Therefore, we choose a set of smaller value of $k \times k$, i.e., 5×5, $7 \times 7, 9 \times 9, 11 \times 11$, and 13×13 to test in this article, as shown in Fig. 20. Notably, after the value of k is determined, we add 0 to the outermost layer of each samples, which will be sent to the input layer of the network. It can be seen from the Fig. 20 that our proposed method is robust within the applicable range of this article. In other words, when the value range of k is small and no matter how it changes, our proposed method is robust,

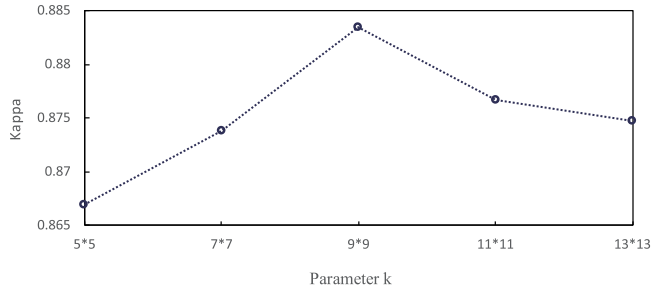

Fig. 20. Relationship between the kappa coefficient and the size of training samples on the Farmland C dataset.

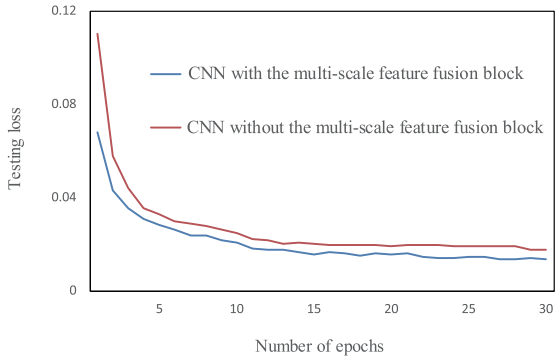

Fig. 21. Comparison of loss curve on the Farmland C dataset.

and the kappa coefficient can stay above 85%. When the value of $k \times k$ is set to be lower than 9×9, the kappa coefficient will rise. When the value of $k \times k$ is set to be higher than 9×9, the result begins to degrade. Therefore, in this article, we set the size of each training samples to 9×9, which can achieve the best performance.

C. Influence of the Multi-Scale Feature Fusion Block

In this article, to indicate the positive influence of the multi-scale feature fusion block in the network on the final results, we remove the multi-scale feature fusion block and form a new network. Fig. 21 shows the the loss curve between the new network and the proposed network in this article. As can be seen from the Fig. 21, the network with multi-scale feature fusion has a higher change-detection accuracy. Therefore, we can conclude that the multi-scale feature fusion block is effective for improving the performance of change detection.

D. Computational Complexity of the Proposed Method

In this article, to evaluate the performance of our method, we discuss the computational complexity involved in four steps constituting the method, as shown in Table VII. $P \times Q$ denotes the size of the original images. S and n denote the searching and neighborhood windows in the improved NSST algorithm, respectively. k denotes the size of the training samples. The computational complexity of one method generally refers to the time complexity and space complexity. In this article, we use the big-**O** notation to represent the computational complexity

TABLE VII
COMPUTATIONAL COMPLEXITY INVOLVED IN EACH STEP CONSTITUTING OUR METHOD

Step	Time complexity	Space complexity
Pseudo-labels Generation	$\mathbf{O}(P \times Q \times S^2 \times n^2)$	$\mathbf{O}(P \times Q)$
Training Samples Generation	$\mathbf{O}(P \times Q)$	$\mathbf{O}(k^2)$
Reliable Samples Selection	$\mathbf{O}(P \times Q)$	$\mathbf{O}(k^2)$
Convolutional Neural Network	FLOPs: 0.078×10^9	Params: 0.083×10^6

except for the network. In the first step (generating the pseudo-labels), as shown in Algorithm 1, the DI processing with the improved NSST algorithm accounts for the main weight, so we consider the time complexity of this process $\mathbf{O}(P \times Q \times S^2 \times n^2)$ as that of the whole step. The space complexity of this step is $\mathbf{O}(P \times Q)$. The second and third steps, namely, generating the training samples and selecting reliable training samples, require traversing every pixel in the image, so their time complexity are $\mathbf{O}(P \times Q)$. In addition, these two steps require the storage space for generating samples, so their space complexity are $\mathbf{O}(k \times k)$. In Table VII, the FLoating point OPerations (FLOPs) denotes the number of operations in the network, which means the time complexity. The $Params$ denotes the number of parameters in the network, which means the space complexity. In this article, the FLOPs of our network is 0.078×10^9, and the number of parameters in our network is 0.083×10^6.

VI. CONCLUSION

In this article, we propose a novel unsupervised method based on improved NSST and multi-scale feature fusion CNN for change detection. The method eliminates the step of generating the DI and directly obtains the final results from the two original images through an easy-training network without any preprocessing. There are two main contributions in this article. 1) It first proposes a novel pseudo-label generator based on improved NSST, which can effectively suppress noise while retaining more details of the DI, so as to obtain more pseudo-labels with higher confidence. It is mainly to solve the critical issue that the DL-based methods for change detection is limited by the requirement of abundant high-precision labels for training. 2) It builds an easy-training network with high change-detection accuracy based on the multi-scale feature fusion. The network integrates the spatial features of different receptive fields and different layer information to make the feature images contain more complete information and reduces the number of pooling layers to avoid losing feature image details. The theoretical analysis and final results conducted on three SAR real datasets prove the effectiveness of our method. Furthermore, to verify the generality and potential of the propose method, we apply it to the cross-region change detection and compare it with the supervised method, which also achieve satisfactory results.

In the future, the following works will be conducted. First, the rule of selecting suitable pseudo-labels is simple in our current method, and there is an imbalance between the two types of samples used for training. We will conduct further research and improvements in the proposed method. Second, our training and testing were performed on the same dataset, and the cross-dataset change detection is not implemented. We will strive to conduct research on the cross-dataset change detection. For example, we will treat the Bern dataset as the training dataset and the Ottawa dataset as the test dataset. Third, we will apply the pretraining which has been widely developed in the field of computer vision in recent years and achieved excellent results [40] to the field of change detection in hope of improving the accuracy of change detection.

REFERENCES

[1] K. Zhang, X. Fu, X. Lv, and J. Yuan, "Unsupervised multitemporal building change detection framework based on cosegmentation using time-series SAR," *Remote Sens.*, vol. 13, no. 3, 2021, Art. no. 471.

[2] S. Ahmadi and S. Homayouni, "A novel active contours model for environmental change detection from multitemporal synthetic aperture radar images," *Remote Sens.*, vol. 2020, no. 12, 2020, Art. no. 1976.

[3] H. Zhuang, K. Deng, Y. Yu, and H. Fan, "An approach based on discrete wavelet transform to unsupervised change detection in multispectral images," *Int. J. Remote Sens.*, vol. 38, no. 17, pp. 4914–4930, 2017.

[4] Y. Shu, W. Li, M. Yang, P. Cheng, and S. Han, "Patch-based change detection method for SAR images with label updating strategy," *Remote Sens.*, vol. 13, no. 7, 2021, Art. no. 1236.

[5] 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.

[6] H. Zhuang, H. Fan, K. Deng, and G. Yao, "A spatial-temporal adaptive neighborhood-based ratio approach for change detection in SAR images," *Remote Sens.*, vol. 10, no. 8, 2018, Art. no. 1295.

[7] A. A. Othman and R. Gloaguen, "River courses affected by landslides and implications for hazard assessment: A high resolution remote sensing case study in NE Iraq-W Iran," *Remote Sens.*, vol. 5, no. 3, pp. 1024–1044, 2013.

[8] Y. Li, C. Peng, Y. Chen, L. Jiao, L. Zhou, and R. Shang, "A deep learning method for change detection in synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5751–5763, Aug. 2019.

[9] 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.

[10] T. Celik, "Change detection in satellite images using a genetic algorithm approach," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 386–390, Apr. 2010.

[11] N. Lv, C. Chen, T. Qiu, and A. K. Sangaiah, "Deep learning and superpixel feature extraction based on contractive autoencoder for change detection in SAR images," *IEEE Trans. Ind. Informat.*, vol. 14, no. 12, pp. 5530–5538, Dec. 2018.

[12] D. Solarna, G. Moser, and S. B. Serpico, "A Markovian approach to unsupervised change detection with multiresolution and multimodality SAR data," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1671.

[13] D. Solarna, G. Moser, and S. B. Serpico, "Multiresolution and multimodality SAR data fusion based on Markov and conditional random fields for unsupervised change detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 29–32.

[14] A. Garzelli and C. Zoppetti, "Optimizing SAR change detection based on log-ratio features," in *Proc. 9th Int. Workshop Anal. Multitemporal Remote Sens. Images*, 2017, pp. 1–4.

[15] N. Bouhlel, V. Akbari, and S. Méric, "Change detection in multilook polarimetric SAR imagery with determinant ratio test statistic," *IEEE Trans. Geosci. Remote Sens*., pp. 1–15, Dec. 2020.

[16] D. Ciuonzo, V. Carotenuto, and A. De Maio, "On multiple covariance equality testing with application to SAR change detection," *IEEE Trans. Signal Process.*, vol. 65, no. 19, pp. 5078–5091, Oct. 2017.

[17] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognit.*, vol. 19, no. 1, pp. 41–47, 1986.

[18] M. Figueiredo and R. D. Nowak, "An EM algorthim for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.
[19] C. J. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern.*, vol. 3, no. 3, pp. 32–57, 1973.
[20] 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.
[21] G. Moser, J. Zerubia, and S. B. Serpico, "Dictionary-based stochastic expectation-maximization for SAR amplitude probability density function estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 1, pp. 188–200, Jan. 2006.
[22] 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.
[23] S. Krinidis and V. Chatzis, "A robust fuzzy local information C-means clustering algorithm MRF," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1328–1337, May 2010.
[24] M. Gong, L. Su, J. Meng, and W. Chen, "Fuzzy clustering with a modified MRF energy function for change detection in synthetic aperture radar images," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 1, pp. 98–109, Feb. 2014.
[25] R. Yang, Z. Pan, X. Jia, L. Zhang, and Y. Deng, "A novel CNN-based detector for ship detection based on rotatable bounding box in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1938–1958, Jan. 2021.
[26] J. Geng, X. Ma, X. Zhou, and H. Wang, "Saliency-guided deep neural networks for SAR image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7365–7377, Oct. 2019.
[27] J.-W. Chen, R. Wang, F. Ding, B. Liu, L. Jiao, and J. Zhang, "A convolutional neural network with parallel multi-scale spatial pooling to detect temporal changes in SAR images," *Remote Sens.*, vol. 12, no. 10, 2020, Art. no. 1619.
[28] Y. Sun, L. Lei, D. Guan, X. Li, and G. Kuang, "SAR image change detection based on nonlocal low-rank model and two-level clustering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 293–306, Jan. 2020.
[29] 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.
[30] 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.
[31] 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.
[32] L. Li and H. Ma, "Saliency-guided nonsubsampled shearlet transform for multisource remote sensing image fusion," *Sensors*, vol. 21, no. 5, 2021, Art. no. 1756.
[33] Y. Zhang, S. Wang, C. Wang, J. Li, and H. Zhang, "SAR image change detection using saliency extraction and shearlet transform," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4701–4710, Dec. 2018.
[34] E. Creusen, R. Duits, A. Vilanova, and L. Florack, "Numerical schemes for linear and non-linear enhancement of DW-MRI," *Numer. Math., Theory, Methods Appl.*, vol. 6, no. 1, pp. 138–168, 2013.
[35] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 60–65.
[36] G. Moser and B. S. Serpico, "Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2972–2982, Oct. 2006.
[37] 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.
[38] G. Kutyniok and W. Lim, "Compactly supported shearlets are optimally sparse," *J. Approximation Theory*, vol. 163, no. 11, pp. 1564–1589, 2012.
[39] G. Easley, K. Guo, and D. Labate, "Analysis of singularities and edge detection using the shearlet transform," in *Proc. Sampling Theory Appl.*, 2009, pp. 1–5.
[40] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapè, "MIMETIC: Mobile encrypted traffic classification using multimodal deep learning," *Comput. Netw.*, vol. 165, 2019, Art. no. 106944.

Fangyu Shen received the B.E. degree in computer science and technology from Northeast Normal University, Changchun, China, in 2017. She is currently working toward the Ph.D. degree in communication and information system with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

She is currently with the University of Chinese Academy of Sciences. Her research interests include synthetic aperture radar system, machine learning, remote sensing image processing, and deep learning.

Yanfei Wang received the B.S. degree in electronic engineering from Beijing Jiaotong University, Beijing, China, in 1984, and the M.S. and Ph.D. degrees in electronic engineering from the Institute of Electronics, Chinese Academy of Sciences, Beijing, China, in 1987 and 1998, respectively.

From 1992 to 1993, he was a Visiting Scholar with the Center for Remote Sensing, University of New South Wales, Sydney, NSW, Australia. He is currently a Research Fellow with the Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include synthetic aperture radar system, radar imaging techniques, and radar signal processing.

Chang Liu received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2000, and the M.S. and Ph.D. degrees in electronic engineering from the Institute of Electronics, Chinese Academy of Sciences, Beijing, China, in 2006.

He is currently a Research Fellow with the Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include synthetic aperture radar system, high-performance embedded radar signal processing, and synthetic aperture radar image application.

Dr. Liu was awarded the Second-Class Prize of the State Scientific and Technological Progress Award in 2010 and 2016, respectively.