Saliency-Guided Deep Neural Networks for SAR Image Change Detection

Jie Geng, Xiaorui Ma, *Member, IEEE*, Xiaojun Zhou, and Hongyu Wang, *Member, IEEE*

Abstract—Change detection is an important task to identify land-cover changes between the acquisitions at different times. For synthetic aperture radar (SAR) images, inherent speckle noise of the images can lead to false changed points, which affects the change detection performance. Besides, the supervised classifier in change detection framework requires numerous training samples, which are generally obtained by manual labeling. In this paper, a novel unsupervised method named saliency-guided deep neural networks (SGDNNs) is proposed for SAR image change detection. In the proposed method, to weaken the influence of speckle noise, a salient region that probably belongs to the changed object is extracted from the difference image. To obtain pseudotraining samples automatically, hierarchical fuzzy C-means (HFCM) clustering is developed to select samples with higher probabilities to be changed and unchanged. Moreover, to enhance the discrimination of sample features, DNNs based on the nonnegative- and Fisher-constrained autoencoder are applied for final detection. Experimental results on five real SAR data sets demonstrate the effectiveness of the proposed approach.

Index Terms—Change detection, deep neural networks (DNNs), synthetic aperture radar (SAR) image, unsupervised learning.

I. INTRODUCTION

CHANGE detection is an important technique to identify land-cover changes, which analyzes two or more images acquired over the same area at different times [1], [2]. It is able to provide insights into environmental monitoring, urban planning, resource exploring, and so on [3]. Multimodal remote sensing images, such as multispectral, hyperspectral, and synthetic aperture radar (SAR) images, have been applied for change detection. For an SAR image, it is captured from an active microwave sensor that reflects backscattering information of land cover in all weather and all time [4], [5]. Therefore, the SAR image has its advantage for change detection task due to its independence of sunlight condition.

Change detection is an essential procedure in SAR image interpretation, which has been researched in the past few decades [6]–[9].

According to the stage of change detection in which the classification step is performed, two major approaches are presented: postcomparison analysis, where the multitemporal data are considered simultaneously for classification [10] and post-classification comparison, where classification is conducted independently at each acquisition, and then, the changes are analyzed by comparing classification maps [11]. The front approach has been the current mainstream with superior performance, which consists of three steps: 1) preprocessing the images; 2) generating a difference image (DI); and 3) analyzing the DI to identify the changed regions [12], [13].

The first step mainly includes coregistration and denoising. First, the multitemporal images must be aligned with a coordinate frame. For SAR images, speckle noise seriously degrades the quality of the images, which can lead to false positive (FP) results. Therefore, numerous denoising techniques have been developed to remove speckle noise on SAR images, such as Lee filtering [14], Frost filtering [15], and SAR-oriented version of block-matching 3-D (SAR-BM3D) algorithm [16].

The second step is to generate the DI from a pair of multitemporal images. For SAR images, the widely applied methods to generate the DI are the ratio operators, such as a log-ratio operator and a mean-ratio operator. The Log-ratio operator converts the pixel values of the images into a logarithmic dimension, and then, the ratio of the image pair is calculated [17]. The Mean-ratio operator integrates the local neighborhood pixels and calculates the mean value ratio of the image pair. Gong *et al.* [18] proposed a neighborhood-based ratio operator to combine spatial information and gray-level information, which produces superior results than traditional methods.

The final step is analyzing the DI to obtain changed and unchanged pixels, which can be divided into supervised and unsupervised methods. In unsupervised change detection, thresholding approaches [19], [20] and clustering approaches [21], [22] are widely utilized, where prior information of training set is not required. In supervised methods, training set with ground truth is utilized as prior information to train the classifiers, such as support vector machine and extreme learning machine (ELM) [23].

According to the aforementioned discussion, there are several issues that hamper the performance of SAR image change detection. First, the inherent speckle noise of SAR images can

lead to false changed points. If denoising methods are utilized directly, useful information may also be removed during denoising. Second, the DI has a great influence on detection performance. At the same time, changed information may be lost during generating the DI. Finally, supervised methods tend to yield superior results than unsupervised methods since prior information is utilized in supervised methods to train the classifier. However, prior information is generally obtained by manual labeling, which requires lots of manual work and influences the generalization of the model. Therefore, effective denoising, feature extraction, and prior information acquisition should be considered for SAR image change detection. In order to suppress speckle noise and keep useful information, a saliency detection model is developed to extract appealing regions that probably belong to the changed objects. To extract changed information, deep neural networks (DNNs) are applied to learn features from the two original images and the DI, which can obtain discriminative features. Moreover, in order to obtain prior information automatically, the clustering model is proposed to yield pseudosamples that are possible to be changed and unchanged.

Based on the above-mentioned analysis, a novel unsupervised method named saliency-guided DNNs (SGDNNs) is proposed for SAR image change detection. In the proposed framework, first, the salient region that probably belongs to the changed object is extracted from the DI, which can remove the background pixels and weaken the influence of speckle noise. Then, hierarchical fuzzy C-means (HFCM) clustering is proposed to yield pseudotraining samples, where samples with higher probabilities to be changed and unchanged are obtained automatically. Finally, DNNs based on the nonnegative- and Fisher-constrained autoencoder (NFCAE) are developed for final change detection, which aim to discover effective changed information from the two original images and the DI. The major novel contributions of this paper are summarized as follows.

1) A simplified context-aware saliency detection model is developed in the SGDNN, which aims to obtain salient areas that probably belong to changed objects. The saliency detection model is able to extract appealing and compact salient regions from the DI with simple operation. It has the capability to remove the background pixels and mitigate the influence of speckle noise.

2) A hierarchical FCM clustering model is proposed in SGDNN, which aims to automatically select credible samples to train the deep network. Two-level clustering is utilized to divide the DI into three groups, which are the changed pixels, unchanged pixels, and intermediate pixels. Only samples from the changed and unchanged classes are chosen as the training set.

3) DNNs based on the NFCAE are developed for change detection, which aim to enhance the discrimination between the changed and unchanged objects. Patch features of the two original images and the DI are adopted as sample features, and the corresponding pseudolabels are utilized as sample labels. Pseudosamples are applied for training the DNN, which prefers to discover effective changed information.

The rest of this paper is organized as follows. Section II presents the related work. The proposed SGDNN for SAR image change detection is described in Section III. In Section IV, the performance of the proposed method is analyzed in the experiments. Conclusions are finally summarized in Section V.

II. RELATED WORK

In SAR image change detection, postcomparison analysis has been the current mainstream with superior performance. Based on whether prior information is required or not, postcomparison methods can be divided into unsupervised methods and supervised methods.

In the unsupervised approaches, the thresholding analysis and the clustering analysis are two mainstreams. Intuitively, the thresholding analysis is generally adopted to segment the DI. Bruzzone and Prieto [24] proposed an unsupervised model based on the Bayes theory, where expectation–maximization algorithm is applied to search the optimal threshold. Bazi et al. [25] developed a Kittler–Illingworth threshold selection criterion, where the distributions of changed and unchanged classes are assumed to be Gaussian distributions. Moser and Serpico [19] utilized a generalized Kittler–Illingworth minimum-error thresholding algorithm for change detection of SAR amplitude images, where the distributions of SAR amplitude images are assumed to be non-Gaussian distributions. These thresholding approaches produce superior performance when the changed and unchanged classes belong to distinct modes in the histogram of DI. However, statistical distributions of the changed and unchanged classes cannot be accurately modeled by Gaussian or non-Gaussian distributions, and thus, the thresholding models cannot always produce superior results [26]. In clustering approaches, a local spatial feature is taken into account for unsupervised change detection. Celik [21] combined k-means clustering and principal component analysis for multitemporal images change detection. Gong et al. [3] developed a reformulated fuzzy local-information C-means clustering algorithm to classify the changed and unchanged pixels, where a mean-ratio image and a log-ratio image are fused as the DI. Li et al. [27] proposed a cascade clustering algorithm that integrates contextual information during feature extraction. Unsupervised methods cannot utilize prior information, which may reduce the discrimination between the changed class and the unchanged class.

In the supervised methods, training samples can be utilized as the prior information for model learning. Li et al. [28] employed a matching pursuit algorithm to analyze the DI, where the dictionary is constructed to represent each patch of the DI. Later, a deformable dictionary learning was proposed in [9] for change detection of multitemporal images, where deformation similarity is utilized for patch matching. Compared with the unsupervised approaches, supervised change detection can take advantage of information from training samples to improve detection performance. However, the training data are generally obtained by manual labeling, which tends to need lots of manual work and weaken the generalization capability.
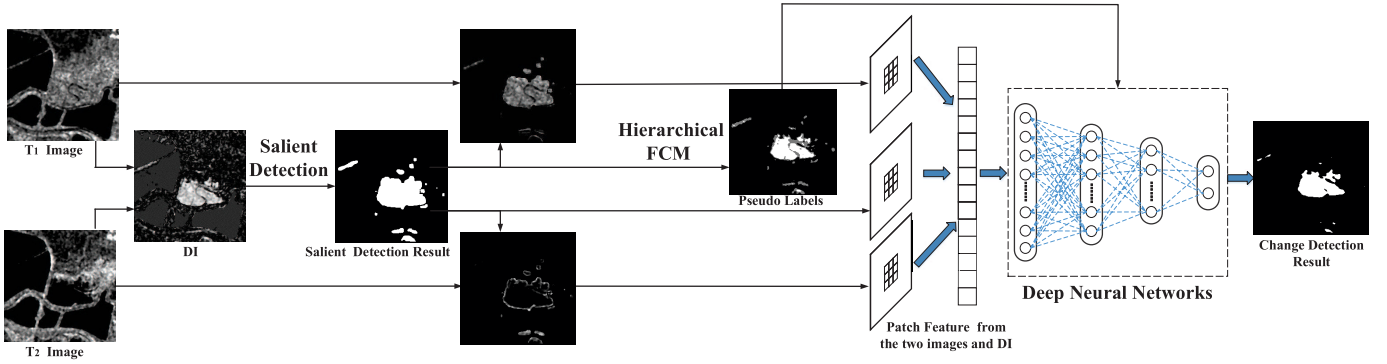
Fig. 1. Framework of SGDNNs for SAR image change detection.

Most recently, preclassification approaches have been proposed to automatically add prior information for model learning in the unsupervised frameworks. Gao *et al.* [23] applied fuzzy C-means (FCM) for the DI preclassification, which aims to select pixels with high probabilities to be changed and unchanged. Later, a change detection framework based on the frequency-domain analysis and random multigraphs was proposed in [29], where FCM is also utilized for preclassification. Zhan *et al.* [30] developed a two-layer neural network for the identification of changed areas, where the fuzzy local-information C-means algorithm is adopted for preclassification.

Deep learning models have been developed to learn the latent structure in the data recently [31]. Compared with the traditional shallow neural networks, DNN can automatically learn a feature hierarchy from the input data [32], [33]. Typical deep models include deep belief networks [34], convolutional neural networks (CNNs) [35], stacked autoencoders (SAEs) [36], [37], and recurrent neural networks [38]. DNN has been successfully developed in various remote sensing tasks [39], [40], which has also been applied for SAR image change detection. Gong *et al.* [41] combined SAE, CNN, and unsupervised clustering to learn robust difference representations for change detection. Liu *et al.* [42] developed an unsupervised deep convolutional coupling network for change detection of heterogeneous images. Mou *et al.* [43] proposed a recurrent CNN (ReCNN) for change detection of multispectral images. Liu *et al.* [44] proposed a local restricted CNN (LRCNN) for change detection of polarimetric SAR images, which can not only recognize different changed types but also reduce the influence of the noise. In our previous work, supervised contractive autoencoders (SCAEs) combined with FCM clustering [13] were developed for SAR image change detection, where neighborhood pixel features of two original SAR images and DI are connected as the input of SCAE. Since Lee filter is used for denoising in our framework, some useful information has been removed before generating the DI. A more reasonable method based on the DNN should be explored to improve the change detection accuracies.

In order to take advantage of DNN in feature learning, SGDNNs are proposed in this paper. Denoising methods are not utilized in our framework since useful changed information may be removed during denoising. A simplified context-aware saliency detection model is developed, which can remove the background pixels and weaken the influence of speckle noise. To utilize prior information for training the deep network, HFCM is developed to yield pseudotraining samples automatically. Considering that information may be lost during generating the DI, patch features of the two original images and the DI are combined as the input of the deep network. Moreover, DNNs based on the NFCAE are developed to enhance the discrimination between the changed and unchanged objects, where the patch features of the two original images and the DI are adopted as input features. The whole SGDNN aims to improve the final performance of SAR image change detection.

III. METHODOLOGY

In this section, SGDNNs are proposed for SAR image change detection. Our proposed framework consists of four modules: DI generation by the log-ratio operator, salient region detection for original images filtering, hierarchical FCM clustering to yield pseudolabels, and DNNs to obtain the final change map. The proposed framework is shown in Fig. 1. The detailed descriptions of salient region detection, hierarchical FCM clustering, and DNNs are reported in Sections III-A–III-C.

A. Salient Region Detection

As for SAR image change detection, difference information can be weakened by speckle noise. To reduce the influence of speckle noise, the salient region that probably belongs to the changed object is extracted before final precise detection. The salient region reflects distinctive information such as visual sense, which can be utilized to obtain potential changed areas and remove the background pixels. Traditional DI reveals the difference between two images, which has strong contrast regions reflecting changed information. Thus, the DI can be utilized for salient region detection to obtain potential changed areas. Therefore, in our developed framework, the DI is first generated by the log-ratio operator. Then, the saliency detection approach is adopted to extract salient regions from the DI.

Context-aware saliency detection [45] evaluates the salient values of each pixel in multiple scales, which is able to extract appealing and compact salient regions. A simplified context-aware saliency detection model is developed in this paper to extract potential changed areas. The Log-ratio operator is

adopted to generate the DI $\boldsymbol{D}$. $\boldsymbol{D}$ is resized into K scales, and salient values of different scales are calculated. Set each pixel of the kth DI $\boldsymbol{D}^k$ as the center of the patch, and compute the distance of patches to represent the dissimilarity. Then, set $\boldsymbol{x}_i^k$ and $\boldsymbol{x}_j^k$ as two vectorized patches corresponding to pixels i and j, and the dissimilarity can be defined as

$$d\left(\boldsymbol{x}_i^k, \boldsymbol{x}_j^k\right) = \frac{\left\| \boldsymbol{x}_i^k - \boldsymbol{x}_j^k \right\|_2}{1 + \lambda_0 \cdot d_{\text{position}}\left(\boldsymbol{x}_i^k, \boldsymbol{x}_j^k\right)} \tag{1}$$

where $d_{\text{position}}(\cdot)$ denotes the positional distance of two pixels and λ_0 is the balance factor. Here, the purpose to calculate the dissimilarity is to obtain the salient value of each pixel.

The saliency of the pixel i with scale k is measured by the salient value of the patch $\boldsymbol{x}_i^k$. To calculate the salient value of $\boldsymbol{x}_i^k$, the N most similar patches are selected from the images whose scales are $\{k, 1/2k, 1/4k\}$. Here, the dissimilarities between $\boldsymbol{x}_i^k$ and other patches in the image are calculated to obtain the N most similar patches. Thus, the salient value of $\boldsymbol{x}_i^k$ is defined as

$$S_i^k = \frac{1}{N} \sum_{j=1}^{N} d\left(\boldsymbol{x}_i^k, \boldsymbol{x}_j^{k_t}\right) \tag{2}$$

where $k_t \in \{k, 1/2k, 1/4k\}$. Equation (2) is a simplified version compared with the original model.

After calculating S_i^k, the saliency map with scale k can be denoted as S^k, which is normalized between 0 and 1. The saliency maps of K scales are combined together with interpolating to the original image size, which is written as

$$\bar{S}_i = \frac{1}{K} \sum_{k=1}^{K} \left[S_i^k \right] \tag{3}$$

where $[\cdot]$ denotes the interpolation operator to resize the map back to the original image size. According to visual contextual effect, areas surrounding the most attentional position have higher probabilities to be salient than far-away regions. Therefore, each pixel outside the attentional areas is weighted by the positional distance between the pixel and the closest attentional pixel, which aims to include the visual contextual effect and obtain more effective saliency result. The salient value of the pixel i is finally calculated as

$$\hat{S}_i = \bar{S}_i \cdot \left(1 - d_{\text{position}}(i, i_{\text{attentional}})\right) \tag{4}$$

where $i_{\text{attentional}} \in \{i : \bar{S}_i / \max(\bar{S}_i) > \tau\}$ and τ is a threshold to select the attentional pixels.

Based on the above-mentioned procedures, the saliency map $\hat{S}$ can be extracted from the DI. To process the original SAR images, the saliency map is divided into a 0-1 map by a thresholding function, which is written as

$$S_T = \begin{cases} 1, & \hat{S} > \gamma \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where γ denotes a threshold. S_T denotes the 0-1 map, in which 1 stands for potential changed area and 0 represents the unchanged area. The 0-1 map is multiplied with the original SAR images, which is defined as follows:

$$I_{Ti} = I_i \odot S_T \tag{6}$$

where I_{Ti} denotes the filtered images, I_i denotes the original SAR image, $\odot$ is the dot product operator, and $i = 1$ or 2. With the utilization of S_T, it is able to remove the background pixels in the DI and reduce the influence of speckle noise.

B. Hierarchical FCM Clustering

After extracting salient areas, hierarchical FCM clustering is developed for preclassification, which aims to select credible samples to train the DNN. The DI is generated for preclassification, where the filtered images I_{Ti} are applied, and thus, just pixels potential to be changed are taken into consideration. Here, the neighborhood-based ratio operator [18] is adopted to obtain DI_T from the filtered images I_{T1} and I_{T2}, which can be defined as

$$
\begin{aligned}
DI_T&(m, n) \\
&= \theta \cdot \frac{\min(I_{T1}(m, n), I_{T2}(m, n))}{\max(I_{T1}(m, n), I_{T2}(m, n))} \\
&\quad + (1-\theta) \cdot \frac{\sum_{i,j \in \Omega \wedge i \neq m, j \neq n} \min(I_{T1}(i, j), I_{T2}(i, j))}{\sum_{i,j \in \Omega \wedge i \neq m, j \neq n} \max(I_{T1}(i, j), I_{T2}(i, j))}
\end{aligned}
\tag{7}
$$

where θ denotes a coefficient to measure the local heterogeneity, which is the ratio between the variance and the mean in the neighborhood Ω.

In order to choose credible samples from the whole image, the HFCM model is developed to divide the DI into three groups, which are the changed class, unchanged class, and intermediate class. The pixels with higher probabilities to be changed and unchanged are classified into the changed class and the unchanged class, respectively. Other pixels are classified into the intermediate class, which cannot be adopted as the training samples. Our hierarchical clustering is composed by two-level FCM clustering, which is conducted as follows. First, the FCM algorithm is applied to separate pixels into three clusters: changed pixels Ω_c, unchanged pixels Ω_u, and intermediate pixels Ω_i. The pixel numbers of these three clusters are denoted as N_c, N_u, and N_i, respectively. Second, the FCM algorithm is utilized again to separate pixels into M clusters ($M > 3$), which are sorted in descending order based on the average values. Third, set the threshold numbers of changed class and unchanged class as $T_c = \alpha_c \cdot N_c$ and $T_u = \alpha_u \cdot N_u$, respectively. Finally, select the first m_c clusters of the second-level clustering as the changed class, where the total number of these clusters is less than T_c. Similarly, select the last m_u clusters of the second-level clustering as the unchanged class, where the total number of these clusters is less than T_u. Other clusters of the second-level clustering are set as the intermediate class. The detailed implementation of the HFCM model is described in Algorithm 1.

C. Deep Neural Networks for Change Detection

In order to obtain more precise change detection results, DNNs are developed for final classification. DNN aims to enhance the discrimination between the changed and unchanged objects. In the proposed framework, the NFCAE [46] is adopted as the basic module of DNN, where nonnegative constraint can improve the reconstruction

Algorithm 1 HFCM Model

Input: Difference image DI_T, clustering number M, threshold parameter α_c and α_u.

Step 1: Conduct FCM algorithm to separate pixels into three clusters Ω_c, Ω_u, and Ω_i. The corresponding numbers are N_c, N_u, and N_i.

Step 2: Conduct FCM algorithm again separate pixels into M clusters.

Step 3: Sort M clusters in descending order based on average values.

Step 4: Set the threshold number of the changed class $T_c = \alpha_c \cdot N_c$, and set the threshold number of the unchanged class $T_u = \alpha_u \cdot N_u$.

Step 5: Select the first m_c clusters from M clusters as the changed class Ω'_c, where the total number of these clusters is less than T_c.

Step 6: Select the last m_u clusters from M clusters as the unchanged class Ω'_u, where the total number of these clusters is less than T_u.

Step 7: Other clusters of the M clusters are set as the intermediate class Ω'_i.

Output: The pre-classification results with three classes $\{\Omega'_c, \Omega'_u, \Omega'_i\}$.

---

quality of sample features [47] and Fisher constraint is able to enhance the intraclass compactness. Credible samples selected by the HFCM model are used to train the DNN, where a percentage of the pseudochanged and pseudounchanged pixels are utilized. As for each pixel, the corresponding patch values from the two original images and the DI are combined into a feature vector. Features and corresponding pseudolabels of credible samples are applied as an initial input of DNN. The optimization of DNN is described as follows.

DNN is composed of L layers NFCAE, which is pretrained layer-by-layer and fine-tuned from the top layer toward the bottom layer. A softmax classifier is connected with the top layer NFCAE for fine-tuning DNN. Features of training samples are denoted as $X = [x_1, x_2, \cdots, x_{N_\text{train}}]$, and the pseudolabels of training samples are represented as $Y = [y_1, y_2, \cdots, y_{N_\text{train}}]$. As for the lth layer NFCAE, the weights and biases are written as $\theta^l = \{W_1^l, W_2^l, b_1^l, b_2^l\}$. The encoding and decoding process is defined as follows:

$$h_i^l = W_1^l \cdot h_i^{l-1} + b_1^l \tag{8}$$

$$\hat{h}_i^l = W_2^l \cdot h_i^l + b_2^l \tag{9}$$

where h_i^{l-1}, h_i^l, and $\hat{h}_i^l$ denote the input, hidden representation, and output of the lth layer NFCAE, respectively. The parameters are optimized by the following objective function:

$$J_\text{pretrain}(\theta^l) = \frac{1}{N_\text{train}} \sum_{i=1}^{N_\text{train}} \|\hat{h}_i^l - h_i^{l-1}\|_2^2 + \beta J_{KL}(\rho\|\hat{\rho})$$
$$+ \frac{\lambda}{2}\left(\|\delta(W_1^l)\|_F^2 + \|\delta(W_2^l)\|_F^2\right) \tag{10}$$

where the first term denotes average representation error between encoding and decoding on the whole training samples,

the second term stands for the sparsity penalty term, and the third term denotes the nonnegative constraint to restrict negative weights. The sparsity penalty term is calculated as $J_{KL}(\rho\|\hat{\rho}) = \sum_{j=1}^{n_k} \rho \log \rho/\hat{\rho}_j + (1-\rho)\log 1 - \rho/1 - \hat{\rho}_j$, where n_k denotes the number of the hidden units and $\hat{\rho}$ denotes the average activation of the hidden units over the whole training set, i.e., $\hat{\rho} = \sum_{i=1}^{N_\text{train}} h_i^l/N_\text{train}$. $\delta(W) = \max(-W, 0)$. The nonnegative constraint is developed to enhance the representation quality of features [46]. A backpropagation algorithm is utilized to solve the above-mentioned objective function [48].

After pretraining, the DNN is fine-tuned from the top layer to the bottom layer, where a softmax classifier is connected to add pseudotraining labels. The global objective function for fine-tuning is defined as follows:

$$J_\text{finetune}(\theta) = \frac{1}{N_\text{train}} \sum_{i=1}^{N_\text{train}} \sum_{l=1}^{L} \|h_i^l - h_i^{l-1}\|_2^2$$
$$+ \frac{\lambda}{2} \sum_{l=1}^{L} \|\delta(W_1^l)\|_F^2 + \eta J_\text{Fisher}(\theta) \tag{11}$$

where the first term is the total representation error of the encoding process, the second term denotes the nonnegative constraint of the whole network, and the third term represents the Fisher restriction to enhance the intraclass compactness. The Fisher restriction $J_\text{Fisher}(\theta)$ is defined as

$$J_\text{Fisher}(\theta) = \frac{1}{N_\text{train}^1} \sum_{i=1}^{N_\text{train}^1} 1\{y(h_i^L) = 1\} \|h_i^L - \bar{h}_1^L\|_2^2$$
$$+ \frac{1}{N_\text{train}^0} \sum_{i=1}^{N_\text{train}^0} 1\{y(h_i^L) = 0\} \|h_i^L - \bar{h}_0^L\|_2^2 \tag{12}$$

where the first term is the top layer feature compactness of the changed class and the second term is the top layer feature compactness of the unchanged class. N_train^1 and N_train^0 denote the training number with respect to the changed class and unchanged class, $\bar{h}_1^L$ and $\bar{h}_0^L$ represent the average feature of the top layer with respect to the changed class and the unchanged class, and h_i^L is the top layer feature. The limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm [49] is utilized to solve (11), and the weights and the biases of the whole network are optimized.

Finally, the softmax classifier is applied for change detection. The predicted probability of the sample x_i belonging to the changed class is written as

$$p(\hat{y}_i = 1 | h_i^L; W^L, b^L) = \frac{e^{W^{L,1} h_i^L + b^{L,1}}}{\sum_{j=0}^{1} e^{W^{L,j} h_i^L + b^{L,j}}} \tag{13}$$

where $W^{L,j}$ represents the portion weight belonging to the jth class and $b^{L,j}$ denotes the portion bias belonging to the jth class. The label of the sample x_i is determined as follows:

$$Label(x_i) = \arg \max_{j=0 \text{ or } 1} p(\hat{y}_i = j | h_i^L; W^L, b^L). \tag{14}$$
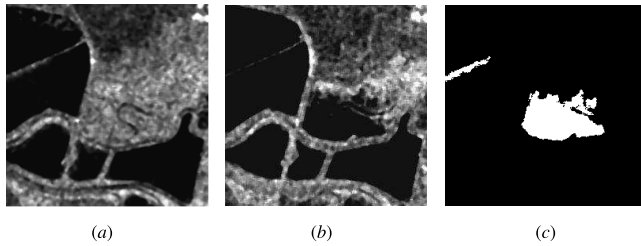
Fig. 2.   San Francisco data. (a) Image acquired in August 2003. (b) Image acquired in May 2004. (c) Ground-truth image.

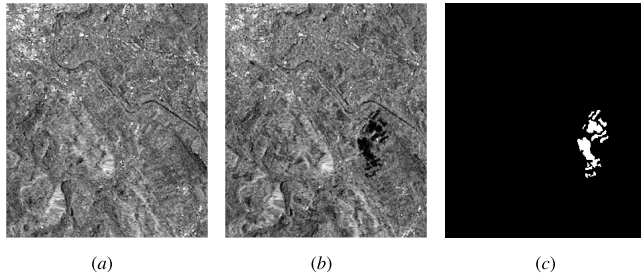Fig. 3.   Bern data. (a) Image acquired in April 1999. (b) Image acquired in May 1999. (c) Ground-truth image.
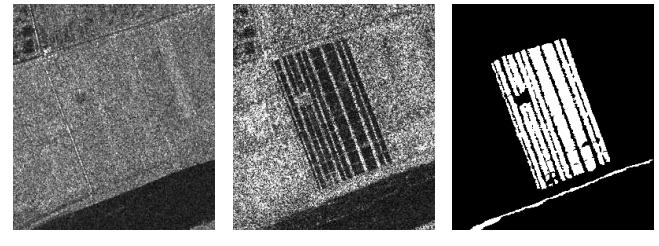

Fig. 4.   Yellow River data. (a) Image acquired in June 2008. (b) Image acquired in June 2009. (c) Ground-truth image.
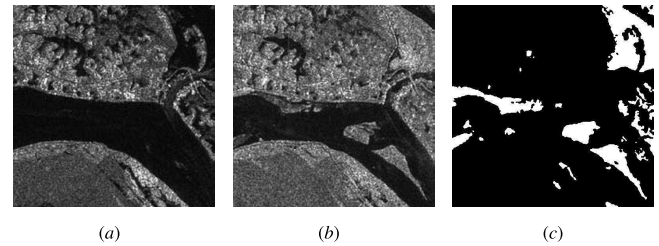

Fig. 5.   Red River data. (a) Image acquired in August 1996. (b) Image acquired in August 1999. (c) Ground-truth image.
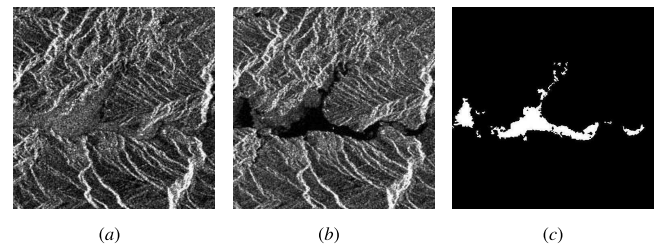

Fig. 6.   Shihmen Dam data. (a) Image acquired in October 1998. (b) Image acquired in September 1999. (c) Ground-truth image.

IV. Experiments

In order to verify the effectiveness of the proposed approach, five real SAR data sets are utilized in the experiments. Five evaluation indices are adopted to assess the change detection performance, which are FPs, false negatives (FNs), overall errors (OEs), percentage correct classification (PCC), and kappa coefficient (Kappa) [3], [8], [26]. The experimental data, the analysis of each module of the proposed deep model, and the experimental comparison are reported in Sections IV-A–IV-F, respectively.

A. Experimental Data

The first data used in the experiment is the San Francisco data, which represents the city of San Francisco with a size of 256×256 pixels. The original images were captured by the ERS-2 SAR sensor in August 2003 and May 2004, which are shown in Fig. 2(a) and (b), respectively. The corresponding ground-truth change map is shown in Fig. 2(c).

The second data is the Bern data, whose spatial size is 301×301 pixels. It represents an area near the city of Bern, Switzerland. The original images were captured in April and May 1999, respectively. Fig. 3 shows the two images and the corresponding ground-truth image.

The third data is the Yellow River data, which shows a region of the Yellow River Estuary in China. The original images were acquired by Radarsat-2 sensor with a size of 257×289 pixels. They were obtained in June 2008 and June 2009, respectively. The two original images and the ground-truth image are shown in Fig. 4. It is clearly seen that the original images are seriously contaminated by speckle noise.

The fourth data is the Red River data, which captures the land cover of the Red River, Vietnam. The original images were acquired in August 1996 and August 1999, respectively. The change was caused by the receding of the flood and the

appearing of the sandbank. The size of the two images is 512×512 pixels. The two original images and the corresponding ground-truth image are shown in Fig. 5.

The fifth data is the Shihmen Dam data, which is shown in Fig. 6. They were obtained in October 1998 and September 1999, which reflect the changes before and after the sediment cleaning of the Shihmen Dam. The size of the two images is 500×500 pixels. The original images and the ground-truth image are shown in Fig. 6.

B. Analysis of Salient Region Detection

In the proposed approach, various parameters have great effects on the change detection results. The proposed framework mainly consists of three modules: the salient region detection, HFCM model, and DNN, which all have some parameters for analysis. Other parameters are set based on experience. The balance factor λ_0, the scale K, and the attentional threshold τ are set to 3, 4, and 0.5, respectively. The threshold parameters α_c and α_u are set to 0.9 and 1.1, respectively. In addition, the balance parameters β, λ, and η are set to 0.01, 0.1, and 1, respectively. In order to analyze the influence of parameters to change detection performance, San Francisco data are utilized as an example.

First, the number of similar patches N in the saliency detection model is discussed. Fig. 7 shows the change detection
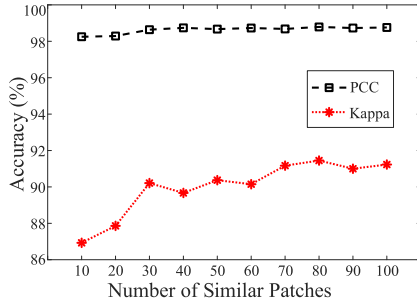
Fig. 7. Change detection results with different numbers of similar patches in a saliency detection model.
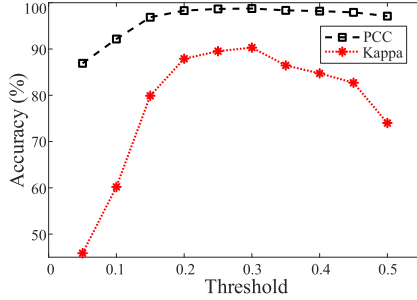

Fig. 8. Change detection results with different thresholds in a saliency detection model.

TABLE I

CHANGE DETECTION RESULTS WITH SALIENCY DETECTION AND
WITHOUT SALIENCY DETECTION ON SAN FRANCISCO DATA

Saliency Detection	FP	FN	OE	PCC(%)	Kappa
Without	773	298	1071	98.27	0.8789
With	321	509	830	98.73	0.9058

accuracies with the changing of N, which is ranged from 10 to 100. It can be observed from Fig. 7 that the detection accuracies increase with the rising of N. The detection accuracies achieve the best when $N = 80$. Therefore, the number of similar patches is set to 80 on the San Francisco data.

The threshold γ in our simplified saliency detection model is also analyzed. The change detection results with different thresholds are shown in Fig. 8, where the threshold is changed from 0.05 to 0.5. According to Fig. 8, the detection accuracies increase first and then decrease with the growth of the threshold. Moreover, when the threshold is equal to 0.3, the proposed model yields the optimal change detection accuracies. Thus, the threshold γ is set to 0.3 on the San Francisco data.

To illustrate the effect of saliency detection, change detection accuracies of the proposed framework with and without saliency detection are reported in Table I. In Table I, FP, FN, OE, PCC, and Kappa are adopted as evaluation indices. It is observed that the accuracies of the framework with saliency detection are much higher than the framework without saliency detection. Therefore, the saliency detection model is able to weaken the influence of speckle noise and improve the change detection performance.

C. Analysis of HFCM

In our HFCM model, clustering number M of the second clustering is a key parameter. Fig. 9 displays the change
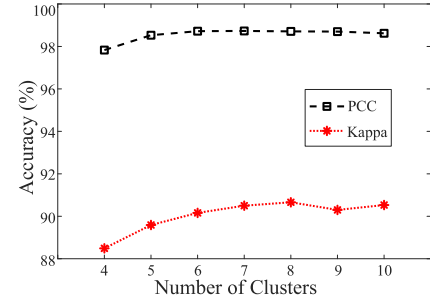

Fig. 9. Change detection results with different numbers of clusters in an HFCM model.

TABLE II

CHANGE DETECTION RESULTS WITH DIFFERENT CLUSTERING
MODELS ON SAN FRANCISCO DATA

Features	FP	FN	OE	PCC(%)	Kappa
FCM-2	1605	245	1850	97.18	0.8125
FCM-3	779	593	1372	97.46	0.8544
HFCM	321	509	830	98.73	0.9058

detection results with different clustering numbers, where the number is ranged from 4 to 10. It is clear that the detection accuracies monotonously increase with the growth of clustering number. More clusters can obtain a better hierarchical distribution of samples, which improve the clustering performance. In addition, the curves tend to be flat when the clustering number is larger than 8, which indicates that eight clusters are enough to produce effective clustering result.

In order to verify the effectiveness of the developed HFCM, the traditional FCM with just once clustering is applied for comparison. FCM with two-class clustering (changed and unchanged) is denoted as FCM-2, and FCM with three-class clustering (changed, unchanged, and intermediate) is denoted as FCM-3. FCM-2 and FCM-3 are compared with our HFCM in Table II. It is noted that the developed HFCM produces the highest detection precisions in PCC and Kappa and yields the least error pixels. Compared with FCM-2, the developed HFCM and FCM-3 obtain much better detection results, which illustrates that three-class clustering is able to select more effective training samples of the changed class and unchanged class. Compared with FCM-3, our HFCM yields higher detection accuracies. It can be demonstrated that two-level clustering is superior to just once clustering since the two-level clustering can select more credible samples with higher probabilities to be changed and unchanged classes.

D. Analysis of DNN

In the experiments, the parameters of DNN are analyzed to obtain the optimal change detection results. The key parameters, including hidden layer number and unit number of each layer, are investigated based on the change detection performance. To analyze the parameters, first, superior situations of DNN with different depths are searched, and then, architectures with different depths are compared to acquire the final deep network. The 30% samples of pseudochanged and pseudounchanged classes from HFCM are applied for training DNN.

TABLE III

CHANGE DETECTION COMPARISON WITH DIFFERENT ARCHITECTURES OF DNN ON SAN FRANCISCO DATA

Hidden layer number	1	2	3	4
Hidden unit number	50	100-50	150-100-50	200-150-100-50
PCC(%)	97.89	98.01	98.73	98.15
Kappa	0.8475	0.8502	0.9058	0.8541

TABLE IV

CHANGE DETECTION RESULTS OF DIFFERENT FEATURES ON SAN FRANCISCO DATA

Features	FP	FN	OE	PCC(%)	Kappa
DI	196	982	1178	98.20	0.8500
DI+original images	321	509	830	98.73	0.9058

TABLE V

CHANGE DETECTION RESULTS OF DIFFERENT CONSTRAINTS IN DNN ON SAN FRANCISCO DATA

Constraint	FP	FN	OE	PCC(%)	Kappa
nonnegative	405	534	939	98.59	0.8997
Fisher	463	420	883	98.67	0.9041
nonnegative+Fisher	321	509	830	98.73	0.9058

TABLE VI

COMPUTATIONAL TIME (s) OF EACH MODULE IN THE PROPOSED SGDNN

Data	Saliency Detection	HFCM	DNN
San Francisco	424.02	23.84	1150.39
Bern	426.97	35.41	1875.78
Yellow River	339.58	30.83	1204.76
Red River	479.49	107.99	4780.52
Shihmen Dam	484.07	88.03	4106.13

TABLE VII

CHANGE DETECTION RESULTS OF EACH MODEL ON SAN FRANCISCO DATA

	FP	FN	OE	PCC(%)	Kappa
FCM-ELM [23]	1190	516	1706	97.40	0.8161
FCM-SAE	1911	304	2215	96.62	0.7802
FCM-SCAE [13]	669	300	969	98.52	0.8925
SGK [50]	561	487	1048	98.40	0.8804
FCM-CNN-SAE [41]	**231**	604	835	98.70	0.9004
DC-CNN [51]	1217	**121**	1338	97.96	0.8612
SGDNN	321	509	**830**	**98.73**	**0.9058**

Table III shows the detection accuracies of superior situations with different layers. It is worth noting that the DNN with three hidden layers performs much higher accuracies, where there is no overfitting in this architecture. We can conclude that the detection performance is not appreciable when the DNN is not deep since it cannot extract discriminative features with high level. At the same time, the change detection accuracies become worse when the DNN is much deeper since the deeper network has more parameters for optimization but training samples are limited to optimize parameters. Therefore, the DNN with three hidden layers is constructed for SAR image change detection, where the hidden unit number of each layer is set to 150, 100, and 50, respectively.

To illustrate the effectiveness of features from the two original images and the DI, features just extracted from the DI are utilized for comparison. The compared results are reported in Table IV. It is clear that using features from the two original images and the DI can produce higher detection precisions in PCC and Kappa. It is illustrated that effective changed information can be discovered from two original images, and combining features of two original images and DI can improve the discrimination between the changed and unchanged objects.

Moreover, in order to verify the effect of nonnegative constraint and Fisher constraint in our DNN, Table V reports the results of the proposed model with different constraints. From Table V, it can be observed that the model combining two constraints yields higher accuracies than the models just using one constraint. Therefore, it is reliable to develop nonnegative and Fisher constraints in our deep model that can improve the change detection accuracies.

E. Analysis of Computational Time

In the experiments, the computational time of each module in the proposed SGDNN is reported in Table VI. In our SGDNN, the preprocessing of saliency detection spends much time, which cannot be real-time processing. The HFCM

model needs less running time, but the training of DNN spends more time on each data. It can be concluded that the proposed network has high computational complexity.

F. Experimental Comparison

In order to demonstrate the effectiveness of the proposed SGDNN, six related approaches are utilized for comparison, which are ELM with FCM (FCM-ELM) [23], SAEs with FCM (FCM-SAEs), SCAEs with FCM (FCM-SCAEs) [13], saliency-guided detection with k-means clustering (SGK) [50], FCM-CNN-SAE [41], and dual-channel CNN (DC-CNN) [51]. The parameters of each method are set to yield the optimal change detection results. Five data are utilized in the experiments to compare the change detection performance. The visual results and the statistical values of FP, FN, OE, PCC, and Kappa are reported on each data.

1) Experiments on San Francisco Data: The experimental results on San Francisco data are reported in Table VII. It is observed that our SGDNN model obtains the optimal precisions in SAR image change detection. The PCC and Kappa of the SGDNN model are much higher than other models, and the OE pixels are fewer than other models. The results of FCM-SCAE and SGK are slightly worse than the proposed SGDNN method. Compared with FCM-CNN-SAE and DC-CNN, the proposed SGDNN produces higher PCC and Kappa, which illustrates that our deep model has better capability than the CNN-based models in SAR image change detection. In addition, FCM-ELM and FCM-SAE produce worse results than other approaches.

Change detection maps are shown in Fig. 10. According to the visual results, it is observed that SGDNN outperforms other approaches, which produces better local consistency and fewer isolated pixels. There are many error pixels in the change maps of FCM-ELM and FCM-SAE, which illustrates that these methods have the worse capability in overcoming the speckle noise of SAR images. The visual effect of SGDNN is much better, which illustrates that saliency detection can reduce the influence of speckle noise. However, the isolated pixels are not well processed in the results of SGK, FCM-CNN-SAE, and DC-CNN. The results of Fig. 10
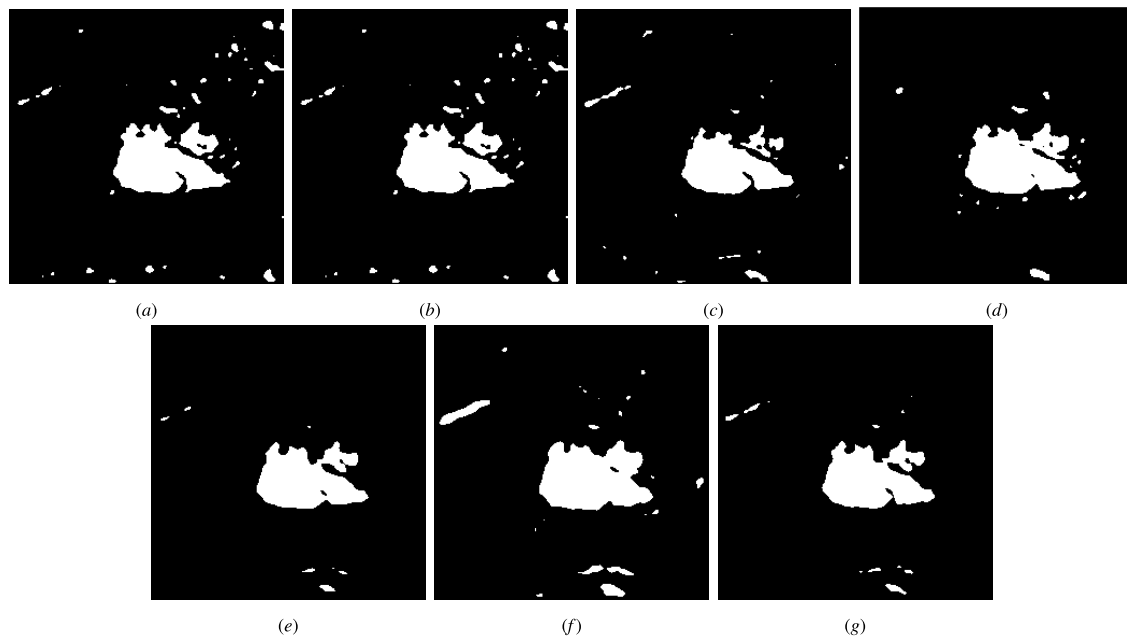
Fig. 10. Change detection maps of each approach on San Francisco data. (a) FCM-ELM. (b) FCM-SAE. (c) FCM-SCAE. (d) SGK. (e) FCM-CNN-SAE. (f) DC-CNN. (g) SGDNN.
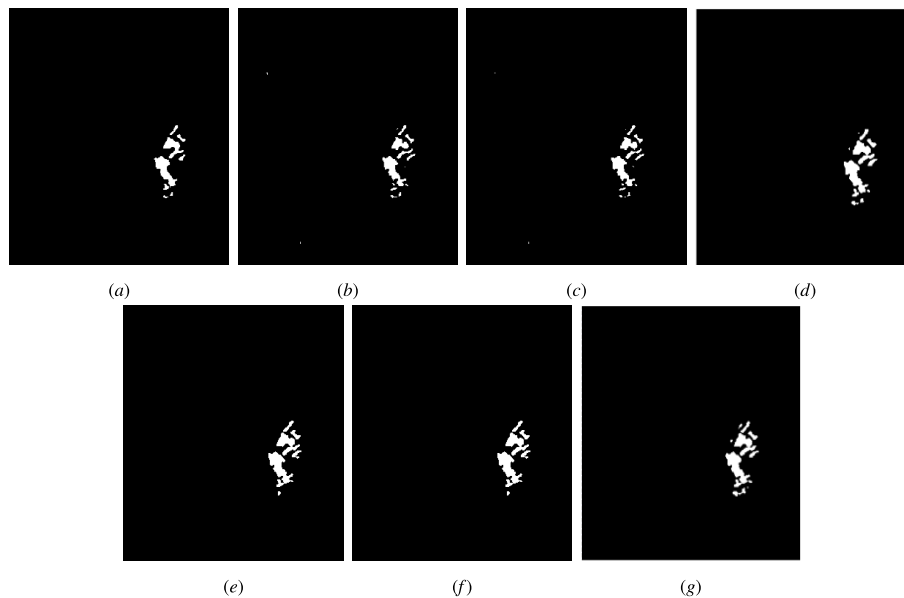

Fig. 11. Change detection maps of each approach on Bern data. (a) FCM-ELM. (b) FCM-SAE. (c) FCM-SCAE. (d) SGK. (e) FCM-CNN-SAE. (f) DC-CNN. (g) SGDNN.

are coincident with those of Table VII. Therefore, SGDNN is effective to yield an appreciable result for SAR image change detection.

2) Experiments on Bern Data: Bern data are also adopted to verify the effectiveness of the proposed approach. The statistical results are listed in Table VIII. As we can see, the detection accuracies of our SGDNN method are higher than other methods. The OE of SGDNN is improved to 268, which is much lower than that of the other four models. The PCC and Kappa of SGDNN are slightly better than other methods. Therefore, the proposed SGDNN model yields the best change detection performance on Bern data.

TABLE VIII
CHANGE DETECTION RESULTS OF EACH MODEL ON BERN DATA

	FP	FN	OE	PCC(%)	Kappa
FCM-ELM [23]	76	308	384	99.54	0.8206
FCM-SAE	61	295	356	99.62	0.8296
FCM-SCAE [13]	**48**	278	326	99.64	0.8332
SGK [50]	132	**151**	283	99.69	0.8749
FCM-CNN-SAE [41]	169	165	334	99.63	0.8538
DC-CNN [51]	183	178	361	99.60	0.8420
SGDNN	103	165	**268**	**99.70**	**0.8793**

Fig. 11 shows the change detection maps of each approach. Since the changes occur in a nearly continuous region and speckle noise is not serious in the images, the isolated pixels
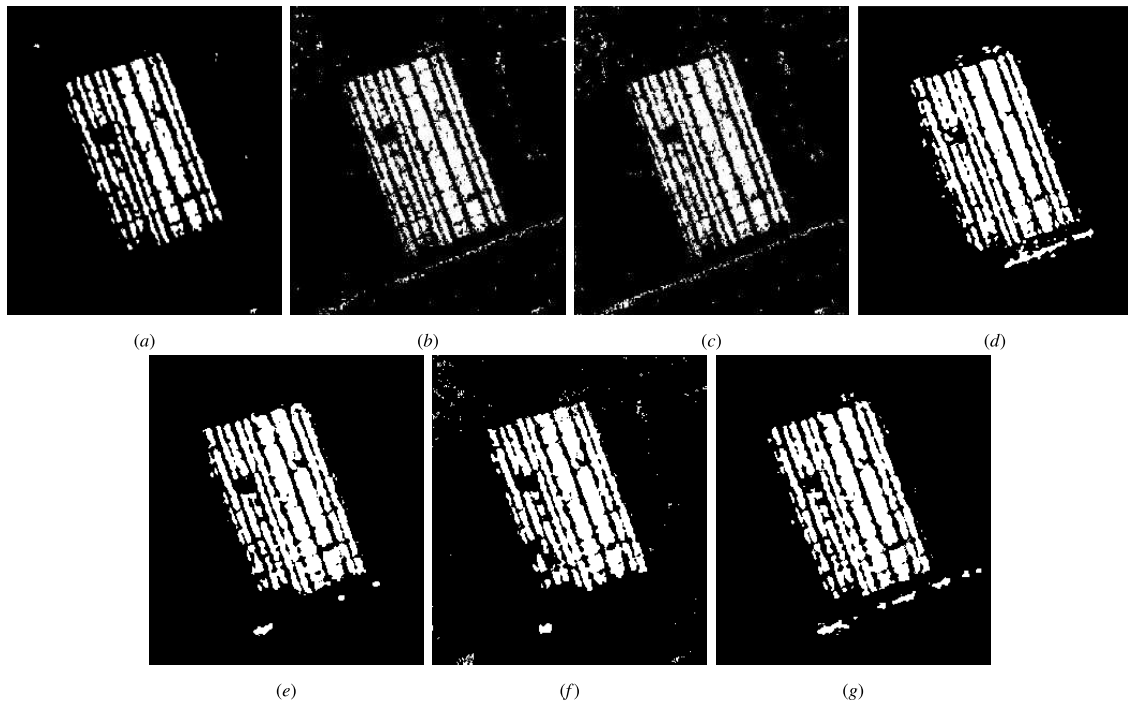
Fig. 12.   Change detection maps of each approach on Yellow River data. (a) FCM-ELM. (b) FCM-SAE. (c) FCM-SCAE. (d) SGK. (e) FCM-CNN-SAE. (f) DC-CNN. (g) SGDNN.

TABLE IX

CHANGE DETECTION RESULTS OF EACH MODEL ON YELLOW RIVER DATA

	FP	FN	OE	PCC(%)	Kappa
FCM-ELM [23]	**62**	5531	5593	92.44	0.6962
FCM-SAE	453	4388	4841	93.50	0.7516
FCM-SCAE [13]	557	4138	4695	93.62	0.7640
SGK [50]	942	**2403**	**3345**	**95.20**	**0.8345**
FCM-CNN-SAE [41]	787	3140	3927	94.71	0.8085
DC-CNN [51]	899	3462	4361	94.13	0.7859
SGDNN	862	2794	3656	94.94	0.8286

TABLE X

CHANGE DETECTION RESULTS OF EACH MODEL ON RED RIVER DATA

	FP	FN	OE	PCC(%)	Kappa
FCM-ELM [23]	27387	805	28192	89.25	0.6413
FCM-SAE	28610	**783**	29393	88.79	0.6304
FCM-SCAE [13]	24680	1228	25908	90.12	0.6603
SGK [50]	9620	3577	13197	94.97	0.7917
FCM-CNN-SAE [41]	**1296**	9759	11055	95.48	0.7898
DC-CNN [51]	8134	3492	11626	95.44	0.8134
SGDNN	1887	8074	**9961**	**95.59**	**0.8162**

are much fewer in the results. Compared the change detection maps with the ground-truth map, the proposed SGDNN yields better visual effect, which is coincident with the results of Table VIII.

3) Experiments on Yellow River Data: The statistical values and the visual results on Yellow River data are reported in Table IX and shown in Fig. 12, respectively. As listed in Table IX, the precisions of our SGDNN model and SGK model outperform other models, where the PCC is improved more than 1% and the Kappa is improved more than 0.06. Since Yellow River data contains a strong level of speckle noise, the proposed SGDNN applies saliency detection to mitigate the influence of speckle noise.

According to the visual results in Fig. 12, the change maps of FCM-ELM, FCM-SAE, and FCM-SCAE contain many error pixels, which reflects that these methods cannot overcome the influence of speckle noise. Since our SGDNN model contains saliency detection process, the change detection map produces better local consistency and fewer isolated pixels. Moreover, it is observed that SGDNN can detect the changed line of the image, which indicates the superiority of our deep model. Based on the above-mentioned analysis, it can be

concluded that SGDNN is able to produce an appreciable performance for SAR image change detection.

4) Experiments on Red River Data: In order to broadly verify the performance of the proposed SGDNN, Red River data are also applied in the experiments. Table X lists the statistical results of each approach. Due to the strong level of speckle noise on the original images, the performances of FCM-ELM, FCM-SAE, and FCM-SCAE are not appreciable. The proposed SGDNN, SGK, FCM-CNN-SAE, and DC-CNN yield much higher accuracies than other three models, where the PCC is improved more than 4% and the Kappa is improved more than 0.12. The results of SGDNN are slightly higher than SGK, FCM-CNN-SAE, and DC-CNN, which illustrates that our deep network can improve the discrimination between the changed and unchanged objects.

Change detection maps of each approach are shown in Fig. 13. It is worth noting that the proposed SGDNN yields the best visual effect, in which the error pixels are fewer than other methods. In the change maps of FCM-ELM, FCM-SAE, and FCM-SCAE, there are numerous error pixels since these methods have the worse capability in overcoming the speckle noise of SAR images. The visual effects of SGK and SGDNN perform better, which illustrates that saliency detection is
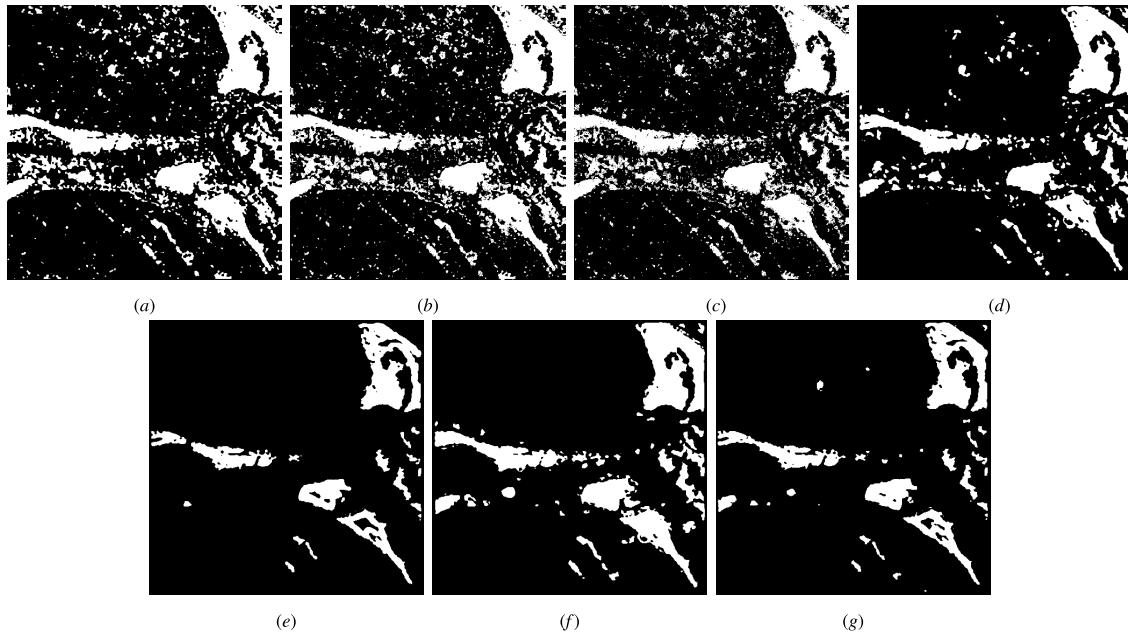
Fig. 13. Change detection maps of each approach on Red River data. (a) FCM-ELM. (b) FCM-SAE. (c) FCM-SCAE. (d) SGK. (e) FCM-CNN-SAE. (f) DC-CNN. (g) SGDNN.
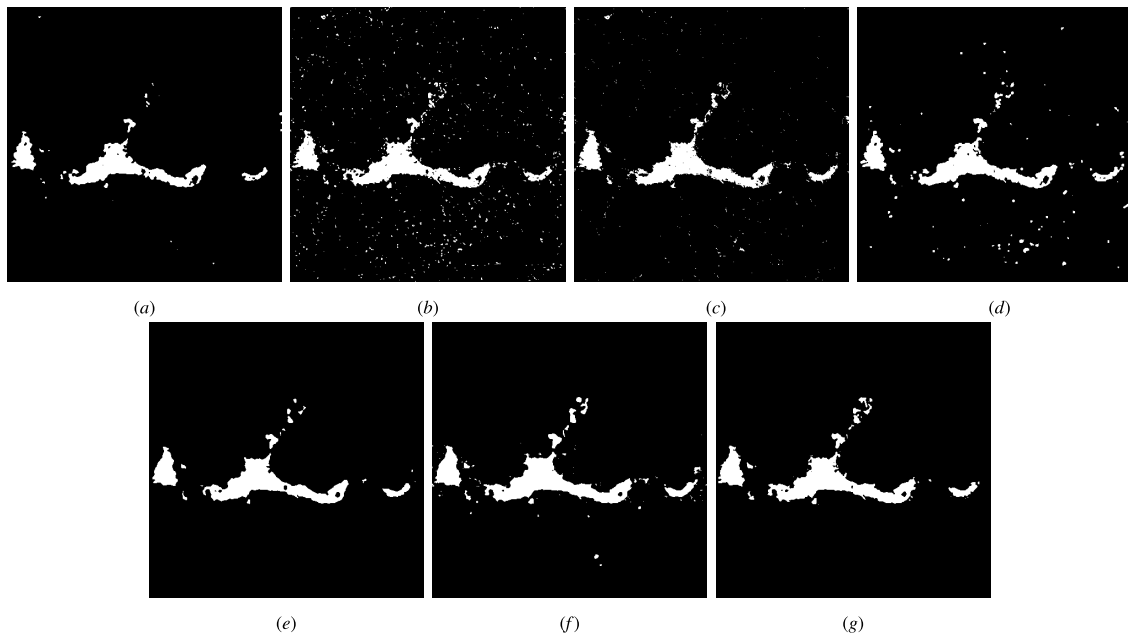

Fig. 14. Change detection maps of each approach on Shihmen Dam data. (a) FCM-ELM. (b) FCM-SAE. (c) FCM-SCAE. (d) SGK. (e) FCM-CNN-SAE. (f) DC-CNN. (g) SGDNN.

able to mitigate the influence of speckle noise. In addition, compared with SGK, our SGDNN has fewer FP pixels since our DNN has the better capability in classification than k-means clustering. In addition, the proposed SGDNN yields superior visual effect than FCM-CNN-SAE and DC-CNN, which indicates that our deep model has better capability than the CNN-based models. Therefore, SGDNN is able to obtain superior change detection result on Red River data.

5) Experiments on Shihmen Dam Data: To confirm the robustness of the proposed SGDNN, the Shihmen Dam data are also utilized in the experiments to compare the performance of each method. The parameters of each method

are set to produce the optimal results. The change detection accuracies on Shihmen Dam data are listed in Table XI. It is observed that SGDNN performs the highest precisions on Shihmen Dam data. In addition, the results of FCM-SAE are much worse than other approaches.

The visual results of each method are shown in Fig. 14. It is clear that SGDNN has the optimal visual performance since isolated pixels are fewer in the change detection map of the SGDNN model. There are many error pixels in the change maps of other approaches. It is illustrated that FCM-ELM, FCM-SAE, and FCM-SCAE have the worse capability in overcoming the speckle noise of SAR images, and SGK

TABLE XI
CHANGE DETECTION RESULTS OF EACH MODEL ON SHIHMEN DAM DATA

	FP	FN	OE	PCC(%)	Kappa
FCM-ELM [23]	**133**	1850	1983	99.12	0.8986
FCM-SAE	3082	429	3511	98.60	0.8509
FCM-SCAE [13]	1447	967	2414	99.03	0.8880
SGK [50]	1942	974	2916	98.83	0.8675
FCM-CNN-SAE [41]	712	1105	1817	99.26	0.9167
DC-CNN [51]	2311	**221**	2532	98.99	0.8901
SGDNN	1450	288	**1738**	**99.30**	**0.9217**

cannot effectively handle the isolated pixels. Moreover, our SGDNN has better capability than the CNN-based models. The results of Fig. 14 are coincident with those of Table XI. To summarize, the proposed SGDNN has the capability to improve the accuracies of SAR image change detection.

V. CONCLUSION

In this paper, the SGDNNs are proposed for SAR image change detection. Experiments on five real SAR data demonstrate that SGDNN can produce superior change detection performance. It is verified that salient region detection in our framework is able to extract the potential changed areas and remove the background pixels from the DI, which can weaken the influence of speckle noise on SAR images. The HFCM model is able to select more credible samples as the training set automatically, which makes the proposed framework to be an unsupervised method. Moreover, DNN has the capability to obtain the discriminative features from the two original SAR images and the DI. Therefore, the proposed SGDNN is verified to perform better than other related change detection methods.

Since the traditional saliency detection model and the clustering model are utilized in our method, it is not an end-to-end deep learning framework for SAR image change detection. We prefer to develop an end-to-end DNN for change detection in our future work.

REFERENCES

[1] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 294–307, Mar. 2005.

[2] E. M. Domínguez, E. Meier, D. Small, M. E. Schaepman, L. Bruzzone, and D. Henke, "A multisquint framework for change detection in high-resolution multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3611–3623, Jun. 2018.

[3] 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.

[4] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.

[5] A. Reigber *et al.*, "Very-high-resolution airborne synthetic aperture radar imaging: Signal processing and applications," *Proc. IEEE*, vol. 101, no. 3, pp. 759–783, Mar. 2013.

[6] F. Bovolo and L. Bruzzone, "A detail-preserving scale-driven approach to change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 12, pp. 2963–2972, Dec. 2005.

[7] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1432–1445, May 2007.

[8] 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.

[9] L. Li *et al.*, "Deformable dictionary learning for SAR image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4605–4617, Aug. 2018.

[10] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Review ArticleDigital change detection methods in ecosystem monitoring: A review," *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1565–1596, May 2004.

[11] M. D. Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson, "Challenges and opportunities of multimodality and data fusion in remote sensing," *Proc. IEEE*, vol. 103, no. 9, pp. 1585–1601, Sep. 2015.

[12] M. G. Gong, P. Z. Zhang, L. Su, and J. Liu, "Coupled dictionary learning for change detection from multisource data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7077–7091, Dec. 2016.

[13] J. Geng, H. Wang, J. Fan, and X. Ma, "Change detection of SAR images based on supervised contractive autoencoders and fuzzy clustering," in *Proc. Int. Workshop Remote Sens. Intell. Process. (RSIP)*, May 2017, pp. 1–3.

[14] J.-S. Lee, "Speckle suppression and analysis for synthetic aperture radar images," *Proc. SPIE*, vol. 25, no. 5, May 1986, Art. no. 255636.

[15] V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 2, pp. 157–166, Mar. 1982.

[16] S. Parrilli, M. Poderico, C. V. Angelino, and L. Verdoliva, "A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 606–616, Feb. 2012.

[17] Y. Bazi, L. Bruzzone, and F. Melgani, "Automatic identification of the number and values of decision thresholds in the log-ratio image for change detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 3, pp. 349–353, Jul. 2006.

[18] 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.

[19] G. Moser and S. B. 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.

[20] Y. Bazi, L. Bruzzone, and F. Melgani, "Image thresholding based on the EM algorithm and the generalized Gaussian distribution," *Pattern Recognit.*, vol. 40, no. 2, pp. 619–634, Feb. 2007.

[21] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and *k*-means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, Oct. 2009.

[22] S. Krinidis and V. Chatzis, "A robust fuzzy local information C-means clustering algorithm," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1328–1337, May 2010.

[23] F. Gao, J. Dong, B. Li, Q. Xu, and C. Xie, "Change detection from synthetic aperture radar images based on neighborhood-based ratio and extreme learning machine," *J. Appl. Remote Sens.*, vol. 10, no. 4, 2016, Art. no. 046019.

[24] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.

[25] 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.

[26] G. Feng, X. Liu, J. Dong, G. Zhong, and M. Jian, "Change detection in SAR images based on deep semi-NMF and SVD networks," *Remote Sens.*, vol. 9, no. 5, p. 435, 2017.

[27] H.-C. Li, T. Celik, N. Longbotham, and W. J. Emery, "Gabor feature based unsupervised change detection of multitemporal SAR images based on two-level clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2458–2462, Dec. 2015.

[28] Y. Li, M. Gong, L. Jiao, L. Li, and R. Stolkin, "Change-detection map learning using matching pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4712–4723, Aug. 2015.

[29] F. Gao, X. Wang, J. Dong, and S. Wang, "SAR image change detection based on frequency domain analysis and random multi-graphs," *J. Appl. Remote Sens.*, vol. 12, no. 1, pp. 1–17, Dec. 2017.

[30] T. Zhan, M. Gong, X. Jiang, and S. Li, "Log-based transformation feature learning for change detection in heterogeneous images," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 9, pp. 1352–1356, Sep. 2018.

[31] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, and Y. Bengio, "An empirical evaluation of deep architectures on problems with many factors of variation," in *Proc. 24th Int. Conf. Mach. Learn.*, Jun. 2007, pp. 473–480.

[32] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[33] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[34] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[35] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proc. IEEE Int. Symp. Circuits Syst.*, May/Jun. 2010, pp. 253–256.

[36] J. Geng, H. Wang, J. Fan, and X. Ma, "Deep supervised and contractive neural network for SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2442–2459, Apr. 2017.

[37] X. Zhang, Y. Liang, C. Li, N. Huyan, L. Jiao, and H. Zhou, "Recursive autoencoders-based unsupervised feature learning for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 1928–1932, Nov. 2017.

[38] Y. Du, W. Wang, and L. Wang, "Hierarchical recurrent neural network for skeleton based action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1110–1118.

[39] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the Art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

[40] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[41] M. Gong, H. Yang, and P. Zhang, "Feature learning and change feature classification based on deep learning for ternary change detection in SAR images," *J. Photogramm. Remote Sens.*, vol. 129, pp. 212–225, Jul. 2017.

[42] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 545–559, Mar. 2018.

[43] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, Feb. 2019.

[44] F. Liu, L. Jiao, X. Tang, S. Yang, W. Ma, and B. Hou, "Local restricted convolutional neural network for change detection in polarimetric SAR images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 818–833, Mar. 2019.

[45] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, Oct. 2012.

[46] J. Geng, H. Wang, J. Fan, and X. Ma, "SAR image classification via deep recurrent encoding neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2255–2269, Apr. 2018.

[47] E. Hosseini-Asl, J. M. Zurada, and O. Nasraoui, "Deep learning of part-based representation of data using sparse autoencoders with nonnegativity constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 12, pp. 2486–2498, Dec. 2016.

[48] L. Bottou, "Stochastic gradient descent tricks," in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 2012, pp. 421–436.

[49] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, 1995.

[50] Y. Zheng, L. Jiao, H. Liu, X. Zhang, B. Hou, and S. Wang, "Unsupervised saliency-guided SAR image change detection," *Pattern Recognit.*, vol. 61, pp. 309–326, Jan. 2017.

[51] T. Liu, Y. Li, Y. Cao, and Q. Shen, "Change detection in multitemporal synthetic aperture radar images using dual-channel convolutional neural network," *J. Appl. Remote Sens.*, vol. 11, no. 4, Oct. 2017, Art. no. 042615.

Jie Geng received the B.S. degree in electronic and information engineering from the Dalian University of Technology, Dalian, China, in 2013, and the Ph.D. degree from the School of Information and Communication Engineering, Dalian University of Technology, in 2018.

He is currently an Associate Professor with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China. His research interests include SAR image processing and machine learning.

Xiaorui Ma (M'17) received the B.S. degree in applied mathematics from the School of Mathematics and Statistics, Lanzhou University, Lanzhou, China, in 2008, and the Ph.D. degree from the School of Information and Communication Engineering, Dalian University of Technology, Dalian, China, in 2017.

She is currently a Lecturer with the School of Information and Communication Engineering, Dalian University of Technology. Her research interests include remote sensing image classification and machine learning.

Xiaojun Zhou is currently pursuing the master's degree with the School of Information and Communication Engineering, Dalian University of Technology, Dalian, China.

Her research interests include remote sensing image change detection and deep learning.

Hongyu Wang (M'02) received the B.S. degree from Jilin University, Changchun, China, in 1990, the M.S. degree from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, in 1993, and the Ph.D. degree in precision instrument and optoelectronics engineering from Tianjin University, Tianjin, China, in 1997.

He is currently a Professor with the School of Information and Communication Engineering, Dalian University of Technology, Dalian, China. His research interests include signal processing, remote sensing image processing, and wireless communication.