

Multiple Statistics Contributing to Few-Sample Deep Learning for Subtle Trace Detection in High-Resolution SAR Images

Jinsong Zhang^{ID}, *Graduate Student Member, IEEE*, Mengdao Xing^{ID}, *Fellow, IEEE*,
 Guang-Cai Sun^{ID}, *Senior Member, IEEE*, and Zhihao Wang^{ID}, *Graduate Student Member, IEEE*

Abstract—With the ability to locate subtle trace objects in the large-scale region, coherent change detection (CCD) has been vital research for a synthetic aperture radar (SAR) system. Finding the difference between repeat-pass repeat-geometry SAR image pair and extracting impressive trace pixels from difference image, the SAR CCD methods consist of a difference generation module and a difference analysis module. The previous CCD methods mainly pay attention to devising a sophisticated working system or an appropriate statistic model to generalize a well difference image. In this article, we introduce the deep learning method into the CCD algorithm and propose a novel trace detection paradigm, which works by hierarchically fusing the unsupervised coherent statistics model and supervised deep learning model. To be specific, the complex reflectance change detection estimator is introduced to generate a difference image and reduce the false alarm in the low clutter-to-noise region. Since the low correlation in a difference image caused by the natural factors severely affects the detection performance, the multiple statistics based on intensity summation and intensity difference are, respectively, proposed to extract water region and vegetation region and suppress the corresponding false alarm. Then the construction of the coarse-to-fine image makes use of land cover information and trace features while the compressed Unet improves the utilization efficiency of trace samples. Meanwhile, the inductive transfer learning based on unsupervised pretraining and few labeled trace samples helps to train a well detection model. Experiments on measured SAR data demonstrate the effectiveness of proposed methods.

Index Terms—Coherent change detection (CCD), fully convolutional network (FCN), statistical model, synthetic aperture radar (SAR), transfer learning.

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) can obtain high-resolution and 2-D remote sensing images by transmitting a large-bandwidth signal along range direction

Manuscript received January 24, 2021; revised March 13, 2021 and April 28, 2021; accepted April 29, 2021. Date of publication May 21, 2021; date of current version January 5, 2022. This work was supported in part by the National Key Research and Development Program of China, in part by the Joint Funds of the National Natural Science Foundation of China under Grant U19B2018, in part by the Foundation for Innovative Research Groups of the National Natural Science Foundation of China under Grant 61621005, and in part by the Key Scientific and Technological Innovation Team Foundation of Shanxi under Grant 2019TD-002. (*Corresponding author: Mengdao Xing*.)

The authors are with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China, and also with the Collaborative Innovation Center of Information Sensing and Understanding, Xidian University, Xi'an 710071, China (e-mail: xmd@xidian.edu.cn).

Digital Object Identifier 10.1109/TGRS.2021.3078741

and growing synthetic array along azimuth direction [1]–[3]. Due to the advantages of all-day and all-weather, SAR has extensive applications in civil and other fields. One of the critical applications is detecting the footprints, wheel prints, and other subtle traces for surveillance and searching purposes [4]–[6].

The imaging result of the SAR system is represented by complex-valued signals with the magnitude reflecting signal intensity and the phase indicating scattering properties [7]–[9]. The small scale of footprints, wheel prints, and other traces leads to the weak energy difference between these traces and the background region, which further leads to a great deal of difficulty to directly extract these trace pixels from SAR intensity image [10]–[12]. Instead, the change detection based on two or more registered images can be used to realize trace detection [5], [13]. The SAR change detection algorithms can be categorized into two types: incoherence change detection (ICCD) methods which utilize intensity changes in SAR magnitude images to indicate large-scale changes such as flood disaster and earthquake deformation [14], [15], and coherent change detection (CCD) methods which use SAR image's phase and magnitude information to estimate the coherence between repeat-pass repeat-geometry SAR collections [4], [5]. Since CCD makes full use of the coherent energy accumulation of SAR complex image pairs, it is more sensitive to the change of subtle traces, and its detection resolution can even reach one-fifth of the radar wavelength [16].

The SAR change detection methods are usually composed of two modules: a module for generating a difference image between repeat-pass repeat-geometry SAR images, and a module for analyzing this difference image to gain unchanged or changed interesting pixels [17], [18]. For the difference image generating module of CCD, a coherence measurement method with maximum-likelihood estimating based on SAR image pairs [19] was proposed. Although this method can detect the change information such as landfill activity, the background clutter, and thermal noise are ignored, which leads to a high false alarm in the difference images. Lots of articles try to use statistical methods to solve these problems. In [20], the generalized likelihood ratio is proposed as the statistic model instead of the original complex correlation coefficient, while in [21] the Stokes parameter is used as the test statistic. Wahl *et al.* [10] derive a complex reflectance change

detection (CRCD) metric for SAR coherent temporal change detection in areas with a low clutter-to-noise power ratio. Compared with traditional CCD methods, these novel methods can decrease the false alarm rate to a certain extent and achieve better performance. However, these correlation measurement methods cannot thoroughly remove the low correlation caused by natural environment variants. Villasensor *et al.* [19] proposed a multipass detection method to suppress the coherent change of background clutter. Ji *et al.* [22] proposed a multiband SAR CCD method, which can eliminate low coherent interference in each single band image and decrease false alarm probability. These methods obtain multidimensional information through multipass or multiband, which can effectively distinguish artificial traces from false alarm targets. However, multiple flights require high-quality experimental conditions such as strict interval time and orbit accuracy, and the produced baseline decorrelation will affect the detection performance. Meanwhile, the strong time-decorrelation of vegetation, rivers, and other natural areas makes them low correlation, further shows as false alarm areas in the final detection results. Thus, how to remove the interference region as well as retain change information as much as possible is still a practical problem to detect subtle traces between the SAR image pairs.

As the second part of the change detection algorithm, the difference image analyzing module tries to extract the changed information from the difference image and classify the pixels of this image as belonging to either a changed or unchanged region [8], [23]. These methods can be categorized into the nondeep learning method and the deep learning method. For the nondeep learning methods, thresholding [24], clustering [25]–[27], and active contours method [28]–[30] have been widely used to recognize the changed or unchanged pixels from the difference image. These methods work without any prior information for their unsupervised learning. In particular, the reformulated fuzzy local information C-means algorithm (RFLICM) is proposed to retain much more information than hard clustering in some cases [27]. For the deep learning method, Gong *et al.* [14] first established a restricted Boltzmann machine (RBM) to learn the relationships between SAR image pairs, then unrolled this RBM into a neural network to learn the concepts of the changed and unchanged pixels with supervised fine-tuning. Lv *et al.* [31] presented a feature learning method using a stacked autoencoder to extract the temporal change feature from superpixel with noise suppression. Gao *et al.* [32] exploited the representative neighborhood features extracted from each pixel by principal component analysis (PCA) filters as convolutional filters, which can generate change maps with fewer noise spots. To weaken the influence of speckle noise, Geng *et al.* [33] proposed saliency-guided deep neural networks (DNNs) based on unsupervised learning. Liu *et al.* [17] proposed a local restricted convolutional neural network (CNN) for change detection in polarimetric SAR images.

These methods presented for ICCD image analysis perform well for large-scale terrain variants such as flood disaster and earthquake deformation. However, when it comes to the CCD images, the natural environmental scene changes such as wind

blowing and river flowing generate decorrelation regions in the difference image. These regions bring great challenges to the detection methods for that they are easily classified as changed trace pixels. Meanwhile, the samples of subtle traces are hard to acquire due to the expensive operating costs of the repeat-pass SAR system, and the requirement of large-scale training data limits the learning efficiency and the detection performance. Therefore, how to train an excellent deep neural network model with a few samples is an important problem for SAR change detection, especially for the detection of subtle traces.

Considering that the difference image generating module and the difference image analyzing module are independent of each other, we propose a new method implemented via an unsupervised coherent statistics model and deep neural model for trace detection in high-resolution and large-scene SAR images. To be specific, the CRCD estimator [10] is first introduced to generate a difference image between repeat-pass repeat-geometry SAR images. Since the low correlation in difference image caused by the natural factors like wind-blown and river-flowing severely affects the trace detection results, the multiple statistics working with intensity summation and intensity difference of image pair are proposed to extract water region and vegetation region and eliminate the false alarm. Then coarse-to-fine (CTF) image, compressed Unet (CUnet), and corresponding inductive transfer learning are proposed to detect trace pixels. The main contributions of this article are listed as follows.

- 1) The introduction of the CRCD estimator integrates the clutter and noise power in SAR images into a unified statistics model, which can decrease the false alarm in the low clutter-to-noise region and improves the detection probability of subtle trace pixels. The generated difference image can be easily and effectively used in subsequent difference image analysis.
- 2) The novel multiple statistical methods are proposed by analyzing the radar cross section of different land covers and the distribution characteristic of the SAR image pair. Extracting water pixels with intensity summation and extracting vegetation pixels with intensity difference, this method has higher running speed and better accuracy than other unsupervised techniques.
- 3) The CTF image, CUnet, and corresponding inductive transfer learning are fused into an entire learning framework to tackle the few trace samples. In these methods, the construction of the CTF image does not only use land cover information but also weakens the false alarm region and highlights the subtle track regions. The CUnet can improve the utilization efficiency of trace samples and reduce the risk of over-fitting. Meanwhile, the inductive transfer learning based on unsupervised pretraining and few labeled trace samples helps to attain the best detection model than other methods.

The remainder of this article is organized as follows. Section II introduces related works of the proposed method. Section III describes the proposed framework for trace detection in detail. Section IV shows the experimental results on measured SAR data and Section V concludes this article.

II. RELATED WORK

A. Statistical Model of CCD

Given two spatially registered SAR complex images X_1 and X_2 , the corresponding pixels of these images form a 2-D joint vector $\mathbf{X} = [X_1, X_2]^T$. \mathbf{X} is subject to 2-D complex Gaussian distribution with zero mean, and its probability density function (PDF) is [4], [34]

$$F(\mathbf{X}) = \frac{1}{\pi^2 \det(\mathbf{C}_X)} \exp(-\mathbf{X}^H \mathbf{C}_X^{-1} \mathbf{X}) \quad (1)$$

where \det means the determinant. The superscript H indicates the complex conjugate transpose operation. \mathbf{C}_X is the covariance matrix of \mathbf{X} . In the CCD algorithm, the following operators [4] are usually used to estimate the complex correlation coefficient:

$$\hat{\gamma}_c = \frac{\left| \sum_{k=1}^N X_{1,k} X_{2,k}^* \right|}{\sqrt{\sum_{k=1}^N |X_{1,k}|^2 \sum_{k=1}^N |X_{2,k}|^2}} \quad (2)$$

where N is the size of the local neighborhood for joint estimation and its value should be consistent with the size of the actual changing area. $\hat{\gamma}_c$ falls into $[0,1]$, and the changed or unchanged area can be visually distinguished according to the value of $\hat{\gamma}_c$. The PDF of sample coherence $\hat{\gamma}_c$ is given by

$$p(\hat{\gamma}_c | \gamma, N) = 2(N-1)(1-\gamma^2)^N \hat{\gamma}_c \cdot (1-\hat{\gamma}_c^2)^{(N-2)} {}_2F_1(N, N; 1; \gamma^2 \hat{\gamma}_c^2) \quad (3)$$

where ${}_2F_1$ is the Gaussian hypergeometric function. This PDF is dominated by N and the underlying scene coherence γ .

However $\hat{\gamma}_c$ will provide a low value when the clutter-to-noise ratio (CNR) is small, indicating that temporal change has occurred [10]. This phenomenon makes it difficult to determine if a true change has actually taken place in the regions with low CNR. Meanwhile, the strong time-decorrelation of vegetation, rivers, and other natural areas makes them low correlation [34], [35], further shows as false alarm areas in the final detection results. Thus, the coherence estimation method should be discussed for higher detection performance.

B. Deep Learning for Difference Image Analysis

As shown in Section I, the deep learning methods have been widely used for the difference image analysis of SAR change detection. For the learning of DNNs, the method which first trains a DNN in an unsupervised way and then fine-tunes the pretrained network with a few labeled samples becomes a mature paradigm. Typical deep models include RBM, deep belief networks (DBNs) [14], stacked autoencoders [31], [36], recurrent neural networks (RNNs) [6], and CNN [17], [37], [38]. When classifying the difference images, these methods based on sliding-window first crop small patches centered on each pixel in the difference image, then use the well-trained networks to classify the central pixel of each patch as an unchanged or changed pixel, as shown in Fig. 1(a). The networks in this method consist of a few effective layers, which extract powerful contour features and semantic features from an input patch with fewer trainable

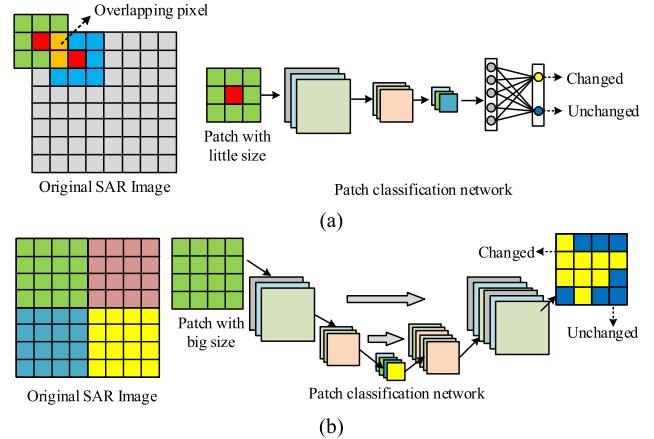


Fig. 1. Inference comparison of sliding-window detection methods and fully-convolutional detection methods. (a) Sliding-window method first crops fixed-size patches for each pixel from the original SAR image, then construct a deep neural network to classify these pixels as changed or unchanged pixels. (b) Fully-convolutional methods first divide the original SAR image into some nonoverlapping patches with big size then use an FCN to classify the pixels in these patches at once.

parameters than other very deep networks. What is more, when there are not sufficient training samples, these methods can also perform high accuracy. However, some disadvantages limit their applications on CCD difference image analysis: 1) each pixel in the original SAR image is cropped to multiple patches which are repeatedly computed by networks, thus the repeated computation of these overlapping pixels seriously causes the waste of computer resources and 2) the size of the patch is difficult to determine for that the patches with small size are sensitive to the speckle noise of SAR image and they may not provide enough context information, and the patches with big size usually require more computing resources and cost more running time.

Recently, the fully convolutional network (FCN) [39]–[41] is proposed to overcome the disadvantages of sliding window-based classification methods. A typical FCN, named Unet [40] is shown in Fig. 1(b) (simplified for viewing). Unet consists of an encoder module which can extract abstract features from the input image, and a decoder module which up-samples the high-dimensional and low-resolution feature maps to high-resolution pixel-wise classification result. Since Unet takes the nonoverlapping patches with a large size (such as 256×256 pixels and 512×512 pixels) as input, it has a smaller computation complexity and a faster running speed than sliding window methods, especially for the classification of large-scale SAR images. Meanwhile, based on the multiple convolution layers, Unet can extract more powerful features from input images with a large receptive field. However, the complex structure of Unet means that it requires lots of samples to train the weighting parameters. It is unrealistic for the difference image analysis of SAR CCD data. Therefore, how to balance the complexity of the classification network and the number of training samples is still a vital and essential problem for the difference image analysis of SAR CCD.

III. PROPOSED METHOD

A trace detection method based on multiple statistics and deep learning is proposed in this article. As shown in Fig. 2,

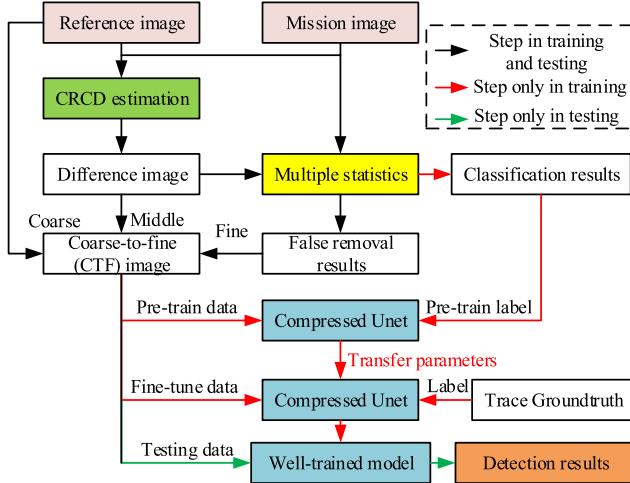


Fig. 2. Flowchart of the proposed method.

according to the training flow and testing flow, the proposed method can be divided into three steps, CRCD estimation for generating difference image, multiple statistics for false alarm removal and land cover classification, and the CUnet-based inductive transfer learning for model training and testing. The details of these steps are described as follows.

A. Difference Image Generation by CRCD

Although the original CCD method can detect subtle trace changes, the estimator $\hat{\gamma}_c$ will provide a low value when the CNR is low, indicating that temporal change has occurred. This makes it difficult to determine if a true change has actually taken place in regions of low CNR only according to threshold [10]. To tackle this problem, the CRCD estimator is introduced to suppress the false alarm in the low CNR region.

Assume two SAR complex images acquired at a different time under the same scene and collection geometry as X_1 and X_2 , respectively, then these images can be modeled as

$$X_k^1 = C_k^1 + n_k^1 \quad (4)$$

$$X_k^2 = \alpha C_k^1 e^{j\varphi} + \sqrt{(1 - \alpha)^2 C_k^2 + n_k^2} \quad (5)$$

where X_k^1 and X_k^2 are the k complex value of image X_1 and X_2 , respectively. α is a measure of the complex reflectivity change that has occurred between X_1 and X_2 , falling into $[0, 1]$. Term C^1 is the unchanged data while term C^2 is the changed data between image pairs. n is the additive system thermal noise. The phase φ represents the constant phase difference, which is treated as a nuisance parameter here. With a few algebraic manipulations, the CRCD estimator is as follows [10]:

$$\hat{\alpha} = \frac{2 \left| \sum_{k=1}^N X_k^1 * X_k^2 \right|}{\left(\sum_{k=1}^N |X_k^1|^2 + \sum_{k=1}^N |X_k^2|^2 - N\sigma_{n1}^2 - N\sigma_{n2}^2 \right)}. \quad (6)$$

Here, $\hat{\alpha} = 0$ indicates the complete change between these neighborhood pixels, and 1 represents unchanged. Compared with the estimator $\hat{\gamma}_c$ in (2), the estimator $\hat{\alpha}$ considers the clutter and thermal noise energy in a neighborhood pixels. Not only does the CRCD estimator raise the surrounding “no change” low-CNR coherence value, but it also increases the

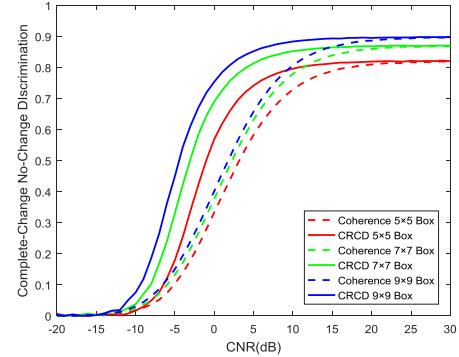


Fig. 3. Ability of the coherence estimator $\hat{\gamma}_c$ and CRCD estimator $\hat{\alpha}$ to differentiate between a complete-change and a no-change state.

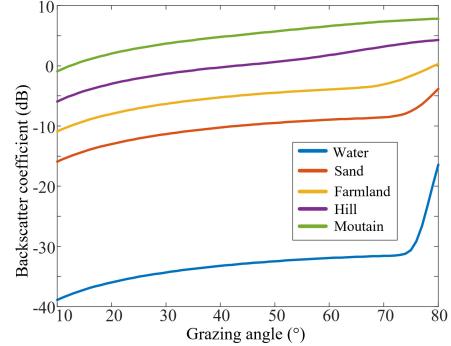


Fig. 4. Backscatter coefficients of water, sand, field, hill, and mountain under the wavelength of 8 mm and the grazing angle of 10° – 80° .

difference between the no-change value and subtle change value in the coherence result. It is noticeable that $(\sigma_{n1}^2, \sigma_{n2}^2)$ are known from the system design specifications or measured from the shadow region of SAR image pairs [42].

To understand the CRCD estimator, we construct an image pair using the Monte Carlo simulations of the image model in (4) and (5). The φ is set as a constant. The coherence α is set as 0 or 1, which represent complete-change or complete unchange, respectively. The differences in the means of the unchanged and change state are shown as a function of CNR. These differences by estimator $\hat{\gamma}_c$ and CRCD estimator $\hat{\alpha}$ are shown in Fig. 3. One can see that the CRCD estimator outperforms the estimator $\hat{\gamma}_c$ when the CNR is between -10 and 20 dB. The larger the size of the neighborhood window, the better the improvement. It is noticeable that too large window size also affects the detection performance [4], [11].

B. Unsupervised False Elimination and Preclassification Based on Multiple Statistics

Although the introduction of CRCD helps to remove the false alarm in low CNR regions, the strong time-decorrelations of water and vegetation make these areas low correlation, further show as false alarm areas in the CRCD difference image. Based on the statistical characteristics of these natural areas, two false elimination or preclassification methods are proposed.

1) *Water Detection Based on the Intensity Summation of Image Pair:* We first use the Morchin model [43] to analyze the scattering characteristics of various land covers in SAR

images. In this model, the backscatter coefficients can be expressed as follows:

$$\sigma_0 = \frac{A\sigma_c \sin \theta_g}{\lambda} + u \cot^2 \beta_0 \exp \left[\frac{-\tan^2(B - \theta_g)}{\tan^2 \beta_0} \right] \quad (7)$$

where λ is wavelength and θ_g is the grazing angle. A , B , u , and β_0 are characteristic parameters, which are related to the type of land covers. The backscatter coefficients of rivers, sands, fields, hills, and mountains under a wavelength of 8 mm and the grazing angle of 0°–80° are shown in Fig. 4, where the water is approximated by the first-level sea condition. Compared with those of other land covers, the σ_0 of water is greatly reduced, which leads to its intensity in SAR images is lower than that of other areas. Based on this analysis, we can choose an appropriate threshold to extract water areas from repeat SAR images. Since the incoherent summation between the intensities of repeat-pass SAR images helps to suppress the speckle noise and highlight the energy differences between different land covers, we take it as the statistic criterion

$$S = \sum_{k=1}^N |X_k^1|^2 + \sum_{k=1}^N |X_k^2|^2 \quad (8)$$

where $|X_k^1|$ and $|X_k^2|$ are the corresponding pixel intensity of the SAR complex image pair, and the mean filter is used to suppress the speckle noise during the estimation. N is the size of the mean filter. After getting the statistic S , the global threshold value τ_s is gotten by the OTSU method [24] to determine whether the pixel m belongs water region or not, like in the following equation:

$$S_m \begin{cases} \leq \tau_s, & \text{water region} \\ > \tau_s, & \text{other region} \end{cases}. \quad (9)$$

2) Vegetation Detection Based on the Intensity Difference of Image Pair: When the thermal noise powers of repeat-pass repeat-geometry images are approximate to each other, the total observed correlation can be written [4], [34]

$$\rho_{\text{total}} = \rho_{\text{spatial}} \cdot \rho_{\text{thermal}} \cdot \rho_{\text{temporal}} \quad (10)$$

where ρ_{thermal} and ρ_{spatial} are thermal decorrelation and spatial baseline decorrelation, respectively. Temporal decorrelation ρ_{temporal} follows from a physical change in the surface over the time period between observations, which can be written as

$$\rho_{\text{temporal}} = \exp \left\{ -\frac{1}{2} \left(\frac{4\pi}{\lambda} \right)^2 ((\Delta y)^2 \sin^2 \theta + (\Delta z)^2 \cos^2 \theta) \right\} \quad (11)$$

where θ is the nominal incident angle. Δy and Δz are the changes of scatter position along vertical and height directions, respectively. For an SAR system with millimeter wavelength, a subtle displacement of the scatter generates a tiny ρ_{temporal} , further causes the severe reduction of ρ_{total} . Thus, both the irregular swing of vegetation and the generation of subtle traces will lead to a low coefficient in the difference image, and it is essential to remove the false alarm caused by the vegetation region.

The motion of vegetation does not only affect the observed correlation but also results in varying backscatter over time. The detailed ground measurements [44] show that the scattering coefficient varies rapidly with the change of grazing angle, which means that there is a significant intensity difference of vegetation region between the reference image and mission image. Different from vegetation, the trace area is not easily affected by natural factors, e.g., wind blowing, so the intensity difference of trace region between image pair is much lower than that of vegetation. Based on these analyses, the intensity difference between image pair is taken as the statistic criterion

$$D = \left| \sum_{k=1}^N |X_k^1|^2 - \sum_{k=1}^N |X_k^2|^2 \right|. \quad (12)$$

For a pixel m in image pair, the pixel satisfied follow condition is decided as vegetation region:

$$D_m \geq \tau_D \text{ & } \hat{\alpha}_m \leq \tau_\alpha \quad (13)$$

where τ_D and τ_α are the threshold of statics D and estimator $\hat{\alpha}$, respectively, which should be determined manually.

Constructing the statistics criterion of intensity summation and intensity difference between the reference image and mission image, the water region and vegetation region can be automatically extracted from large-scene SAR images. At the same time, the false alarm caused by these regions can also be removed. Thus, the proposed multiple statistics can not only suppress the false alarm in the difference image but can also provide a pretraining label for subsequent model learning.

C. Trace Detection Based on CUnet and Inductive Transfer Learning

To realize the change detection of SAR image based on deep networks, a vast number of labeled samples are required to optimize the loss function during the training process, which leads to the labor-intensive workload of manual labeling. What is more, the trace samples are hard to be acquired due to the expensive operating costs of the repeat-pass high-resolution SAR system. To solve these problems, a CUnet based on inductive transfer learning is proposed.

1) Generation of CTF Image: Since the objects to be detected in the ICCD method are mostly the large-scale surface deformation such as earthquake and flood, the intensity between repeat-pass SAR images perform obviously different, and most of deep learning methods in ICCD use the intensity of repeat-pass SAR complex images or the difference image between them as the input of the neural network [27], [32], [37]. Unlike ICCD, most of the objects to be detected in CCD are footprints, vehicles, and other weak traces [5], [11] that cannot be directly extracted from intensity image for their subtle feature variants between image pairs. To make use of multilevel and multidimensional features of a difference image and false removal results, we construct a CTF image in this section and take it as the input of the proposed detection network. The generation of a CTF image is as follows:

- 1) First assuming an intensity image of the repeat-pass SAR image as I_1 , it is chosen as the “coarse image,”

which reflects abundant land cover information of the observation region.

- 2) Then let the coherence coefficient generated by CRCD estimator as the “middle image” I_2 , which reflects the variants between the repeat-pass SAR images caused by the subtle trace and natural phenomena.
- 3) Then assuming the result of false alarm elimination as the “fine image” I_3 , it highlights the trace area and weakens the false alarm area caused by natural phenomena.
- 4) Finally concatenating image I_1 , I_2 , and I_3 as a whole image along the channel dimension. The combined image is named the CTF image (shorten for CTF).

Different from every single image, the CTF image not only covers various land cover information but also weakens the false alarm region and highlights the subtle track regions. The CTF image in Section IV shows that the feature of land covers and trace regions in the CTF image is easy to be extracted than that in a single image. Meanwhile, CTF image constructs a bridge between source task and target task in transfer learning by associating the source task and target task with different labels, which can significantly improve the performance of trace detection. The detail is described in the third part of this section.

2) *Architecture of CUnet*: The Unet structure has been widely used in image segmentation of optical images, hyperspectral images, and SAR images [40], [45], [46]. The reasons mainly lie in that the encoder can extract powerful features from the input image, and the skip-connection-based decoder can effectively upsample different-level feature maps to the accurate segmentation results. However, when it comes to the analysis of a difference image of SAR CCD data, there are too many weighting parameters to be trained and too many pooling layers in Unet. To tackle the problem of the few numbers and small scale of trace samples, the CUnet is proposed in this section.

As shown in Fig. 5, the CUnet takes the CTF image patch with a size of 128×128 pixels as input. Both the encoder and decoder in CUnet consist of four convolution module. There are two convolutional layers in each module and the kernel size of these layers is 3×3 pixels. Each convolutional layer is followed by an activation layer and a batch normalization (BN) layer [47]. Specifically, the leaky ReLU (LReLU) is taken as the activation function

$$y_i = \begin{cases} x_i, & \text{if } x_i \geq 0 \\ x_i/a_i, & \text{if } x_i < 0 \end{cases} \quad (14)$$

where $a_i \in (1, +\infty)$. x_i is the value of the neuron unit. To prevent over-fitting, the dropout layer is introduced to the first convolutional layer of each module. Three maxpooling layers with size 2×2 pixels are used to reduce the size of feature maps and increase the receptive field. A convolutional layer with a size of 1×1 is added in the final layer of the decoder to transform the feature maps into category probability. The number of the trainable weighting parameters of Unet is 7846723 while that of CUnet is 1947763. Fewer parameters mean that CUnet can improve the utilization efficiency of trace samples and reduce the risk of over-fitting.

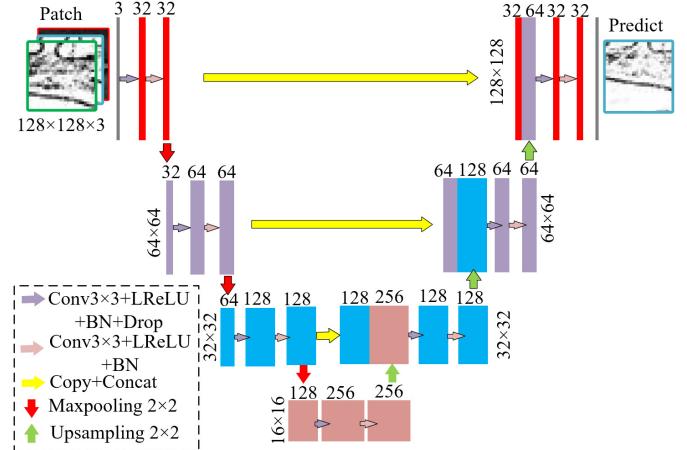


Fig. 5. Structure of the proposed CUnet.

Experiments in Section IV verify the effectiveness of CUnet for trace detection.

3) *Model Training Based on Inductive Transfer Learning*: Although the construction of the CTF image and CUnet can help to train a classification model well, the small number of trace samples is still a vital problem for model training. Recently transfer learning has been widely used for few-shot learning [22], [48], [49]. Since the number of training samples of trace detection is not enough to train an effective CUnet, it is essential to construct a source domain with large labeled samples that are related to our trace data and can help to train our trace detection task. This learning strategy belongs to the inductive transfer learning [50]. Generally, the widely used ImageNet [51], Pascal VOC, or COCO dataset are taken as the source domain. However, the SAR images are obviously different from these optical images in that SAR images are sparse scattering center-based images that are not as intuitive and clear as optical images [1], [52]. Under this condition, brute-force transfer may be unsuccessful. In the worst case, it may even hurt the performance of learning in trace detection, a situation which is referred to as “negative transfer” [50].

The SAR CCD data are usually acquired by the spotlight SAR system for their high-resolution, e.g., from a few centimeters to dozens of centimeters. The high-resolution means large-scale image scene and large image size. The trace pixels only occupy a small portion of this SAR image, while most of the regions are various land covers. Since both the trace regions and land covers share the same imaging condition and imaging algorithm, the correlation between them is more significant than that between optical images and trace regions. Based on these analyses, we define these various land covers in the CTF image as the source domain data and propose a novel training strategy of CUnet. The proposed strategy consists of a pretraining stage and a fine-tuning stage. The detailed descriptions are as follows.

- 1) Cropping the CTF image excluding trace pixels into patches with size 128×128 pixels to build the source data $X_S = \{x_{S_1}, x_{S_2}, \dots, x_{S_n}\}$ and building source label $Y_S = \{y_{S_1}, y_{S_2}, \dots, y_{S_n}\}$ by classifying each pixel in X_S as water, vegetation, or background pixels according to

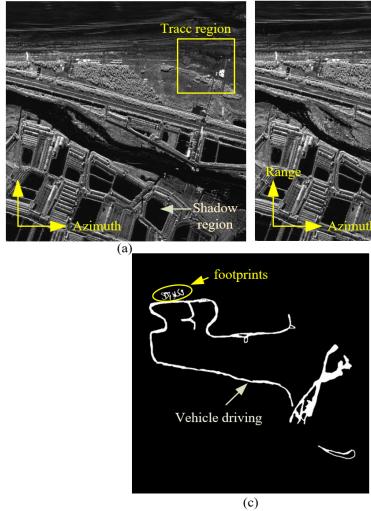


Fig. 6. Repeat-pass repeat-geometry SAR images. (a) Reference image. (b) Mission image. The images bracketed a vehicle traveling on the sand on the top right of the image (a). (c) Ground truth of trace region.

the proposed unsupervised multiple statistics method or manual annotation.

- 2) Cropping the trace pixels in CTF image into the patches with size 128×128 pixels to build the target's data $X_T = \{x_{T_1}, x_{T_2}, \dots, x_{T_m}\}$ and label $Y_T = \{y_{T_1}, y_{T_2}, \dots, y_{T_m}\}$ where y_{T_m} is 0 or 1, which means the trace pixels or background pixels, respectively.
- 3) Modeling the predictive function $f_S(\cdot)$ of the source task with the proposed CUnet using the training samples $\{x_{S_i}, y_{S_i}\}$ in source data. This step is named as “pretrain-ing.”
- 4) Modeling the predictive function $f_T(\cdot)$ of trace detection task with the proposed CUnet, initializing the weighting in $f_T(\cdot)$ with the well-trained $f_S(\cdot)$, and training $f_T(\cdot)$ using the training samples $\{x_{T_i}, y_{T_i}\}$ in target data. This step is named “fine-tune.”

With the above inductive transfer learning, we will get a well-trained CUnet model. Given an image patch, this model can automatically classify each pixel in the patch as a trace or background. Among these steps, the CTF image bridges the source task of land cover classification and target task of trace detection where the maximization of the correlation between source task and target task effectively prompts the learning of trace detection with small samples.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In our proposed framework, the CRCD estimation is utilized for generating a difference image, the multiple statistic model is utilized for false alarm removal, and the CUnet-based inductive transfer learning is proposed for trace detection. To verify the effectiveness of our method, these three parts are, respectively, analyzed in follows to illustrate the contributions of each module to the whole framework.

A. Data Description and Evaluation Criteria

Measured SAR data are collected in this article. These data consist of two repeat-pass repeat-geometry SAR images with

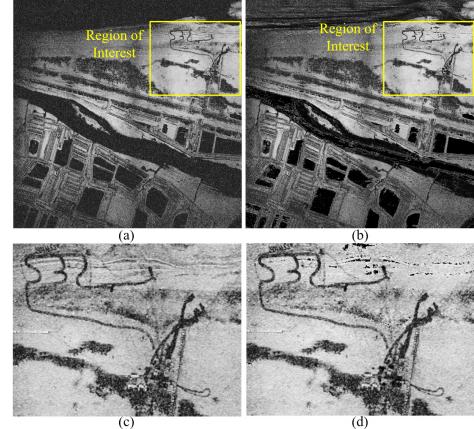


Fig. 7. Results by different detection estimators. (a) Detection result of CCD coefficient. (b) Detection results of CRCD coefficient. (c) Region of interesting of (a). (d) Region of interesting of (b).

a size of 4096×4096 pixels. These images were acquired in Yantai City, Shandong Province of China by a Ka-band SAR system under the spotlight imaging mode and HH polarization. The acquired interval between the two images is 4 h. The range resolution is 0.13 m and the azimuth resolution is 0.21 m. As shown in Fig. 6(a), there are various land covers in this measured SAR data, e.g., water, vegetation, and other backgrounds. A vehicle was driven from top to bottom on the upper right region of Fig. 6(a). Meanwhile, we also arrange some subtle footprints in these regions. The generated mission region is shown in Fig. 6(b). The groundtruth of the trace region is shown in Fig. 6(c), where white pixels mean trace pixels and black pixels mean background regions.

Define the number of undetected trace pixels as false negative (FN), the number of wrongly detected background pixels as false positive (FP), the number of correctly detected trace pixels as true positive (TP), and the number of correctly undetected background pixels as true negative (TN). A few evaluation criteria are listed as follows:

$$P_{FP} = FP / (TN + FP) \times 100\% \quad (15)$$

$$P_{FN} = FN / (TP + FN) \times 100\% \quad (16)$$

$$POE = (FP + FN) / (TP + FP + TN + FN) \times 100\% \quad (17)$$

$$P_{CC} = (TP + TN) / (TP + FP + TN + FN) \times 100\% \quad (18)$$

where POE is the percentage of overall error and P_{CC} is the percentage of correct classification. P_{FP} , P_{FN} , POE , and P_{CC} fall into $[0,1]$ while the smaller the value of P_{FP} , P_{FN} , and POE , the lower the detection false rate. The bigger the value of P_{CC} , the higher the detection accuracy. Kappa coefficients are calculated as follows:

$$\text{Kappa} = (P_{CC} - PRE) / (1 - PRE) \times 100\% \quad (19)$$

where

$$PRE = \frac{(TP + FP) \times (TP + FN) + (FN + TN) \times (FP + TN)}{(TP + TN + FP + FN)^2}. \quad (20)$$

Kappa usually falls into $[0,1]$, and the bigger the value of Kappa, the higher the detection accuracy.

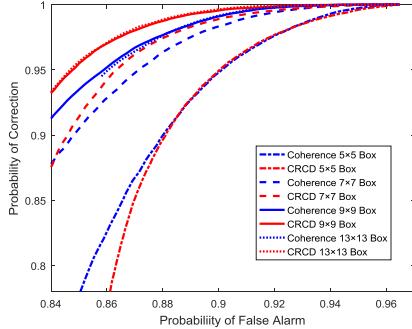


Fig. 8. ROC curves of Fig. 7(c) and (d) of coherence estimator $\hat{\gamma}_c$ and CRCD estimator $\hat{\alpha}$ with different widow size.

B. Experimental Results of CRCD

As shown in Section III, the CRCD estimator $\hat{\alpha}$ is introduced to produce a difference image with low false alarm and high detection alarm. The shadow regions in Fig. 6(a) and (b) are, respectively, selected to compute the thermal noise term σ_{n1}^2 and σ_{n2}^2 in (6). The results of the CCD estimator $\hat{\gamma}_c$ and CRCD estimator $\hat{\alpha}$ are shown in Fig. 7(a) and (b), respectively. The neighborhood size is set 9×9 pixels. The brighter the coefficient pixels, the higher the probability that they are classified as unchanged pixels, while the darker the coefficient pixels, the more likely they are changed pixels. One can see that besides the changed trace pixels, there are lots of dark pixels in Fig. 7(a) and (b). If we directly classify these images as trace region or background according to the value of the coefficient, the vegetation, water, and other low-correlated pixels will cause a severe false alarm.

To understand the detection performance of the CCD estimator $\hat{\gamma}_c$ and CRCD estimator $\hat{\alpha}$, the whole trace regions are cropped from Fig. 7(a) and (b). With the different window sizes and decision thresholds, the receiver operating characteristic (ROC) curves are attained, as shown in Fig. 8. Under the same false alarm probability, both the correction probabilities of the CCD estimator and the CRCD estimator have been improved by the increase of the window size. Notably, using the 9×9 window and 13×13 window, we achieve the best detection accuracy. Since an oversize window will lead to the blurry boundary of trace regions and the huge computational load [10], [11], we select the 9×9 window, instead of the 13×13 window to generate the difference image. The CRCD estimator has a higher correction probability than the CCD estimator under the identity false alarm probability. Meanwhile, the CRCD estimator has a lower false alarm probability than the CCD estimator under the identity correction probability. Thus, the introduction of the CRCD method helps us to get the difference image with low false alarm and high detection accuracy.

C. Experimental Results of False Elimination and Image Preclassification

The effectiveness of unsupervised false elimination and preclassification is verified in this section.

1) *Results of Water Detection:* We first analyze the performance of the water removal method based on the intensity

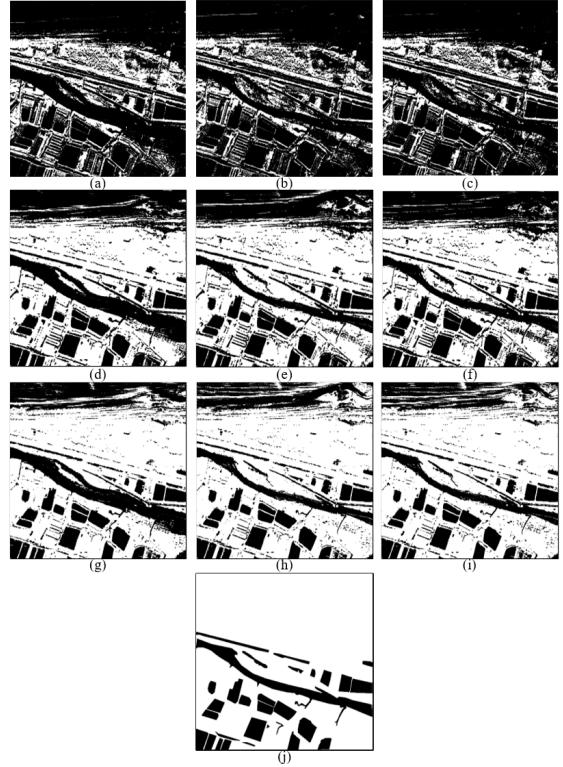


Fig. 9. Water detection results by different methods. (a) MRF based on reference image. (b) MRF based on mission image. (c) MRF based on image pair. (d) LevelSet based on reference image. (e) LevelSet based on mission image. (f) LevelSet based on image pair. (g) OTSU based on reference image. (h) OTSU based on mission image. (i) OTSU based on image pair. (j) Ground truth by manual annotation.

summation. The window size of the mean filter is set as 3×3 pixels. The proposed method takes the intensity summation of image pair as statistics and determines the global detection threshold by the OTSU method [24]. To validate the effectiveness of these steps, the intensity of the reference image and mission image is, respectively, taken as the comparison statistics, and the Markov random field (MRF) [53] and LevelSet [54] are chosen as the comparison methods. The visualization results of these methods are shown in Fig. 9. Fig. 9(j) shows the ground truth of the water region in Fig. 6(a). It is noticeable that the top region in Fig. 6(a) is the sand covered with seawater, which varies significantly with time flowing and we do not annotate this region in the groundtruth. Fig. 9(a)–(c) shows that lots of land pixels are wrongly classified as water regions by the MRF method. The results in Fig. 9(d)–(f) show that although most water regions are classified correctly, there are still lots of isolated false alarm pixels. Fig. 9(g)–(i) shows that the false alarm pixels are significantly reduced by the OTSU method. In particular, the intensity summation helps to suppress the speckle noise in the reference image and mission image. To further understand the performance of different methods, the quantitative results corresponding to Fig. 9 are shown in Table I. The proposed method has achieved the highest P_{CC} and the highest Kappa. Meanwhile, this method only costs 0.2 s for the image pair with 4096×4096 pixels.

2) *Results of Vegetation Detection:* The performance of the vegetation removal method based on the intensity difference

TABLE I
WATER DETECTION RESULTS BY DIFFERENT METHODS

Statistics	Criteria	Method		
		MRF	LevelSet	OTSU
Reference Image	PFP (%)	0.88	1.52	2.39
	PFN (%)	67.06	41.56	28.97
	POE (%)	37.22	13.46	8.62
	PCC (%)	62.78	86.54	91.38
	Kappa (%)	29.90	63.90	74.07
	Time (s)	283.29	317.45	0.19
Mission Image	PFP (%)	1.27	2.11	3.09
	PFN (%)	69.17	35.73	14.72
Intensity Summation of Image Pair	POE (%)	40.81	10.97	5.27
	PCC (%)	59.19	89.03	94.73
	Kappa (%)	26.05	68.75	82.61
	Time (s)	281.71	317.43	0.18
	PFP (%)	1.01	1.77	2.91
	PFN (%)	69.32	39.53	14.65
	POE (%)	41.24	12.53	5.13
	PCC (%)	58.76	87.47	94.87
	Kappa (%)	25.83	65.62	83.13
	Time (s)	289.66	319.30	0.20

TABLE II

VEGETATION DETECTION RESULTS BY DIFFERENT METHODS

Criteria	Method		
	MRF	GA-FCM	Proposed method
PFP (%)	2.13	3.05	2.49
PFN (%)	75.54	77.38	72.58
POE (%)	27.15	25.11	24.18
PCC (%)	72.85	74.89	75.82
Kappa (%)	26.98	24.54	30.58
Time (s)	340.02	635.56	0.68

is analyzed in this section. The window size of the mean filter is set as 3×3 pixels. τ_D is set as 0.73 while τ_a is set as 0.35 after a few trials. The unsupervised MRF method and genetic algorithm for fuzzy C means clustering (GA-FCM for shortening) [55] are chosen as comparison methods. The mission image of the SAR image pair is taken as the input for MRF and GA-FCM. The potential is set as 0.8, while the maximum iteration is set as 30 for MRF. For GA-FCM, the number of generations is 30, the crossover probability is 0.6, and the mutation probability is 0.1. The vegetation detection results of these methods are shown in Fig. 10. The manually annotated groundtruth of the vegetation region is shown in Fig. 10(d). It is noticeable that this ground truth is taken as the evaluation reference here and it does not affect the final trace detection result. One can see that MRF and GA-FCM extract most vegetation regions from the mission image. However, there are lots of false alarms in Fig. 10(a) and (b) caused by the shadow regions and building regions. Compared with these methods, the proposed intensity difference method retains most of the vegetation regions while eliminating lots of false alarm. The quantitative results corresponding to Fig. 10 are shown in Table II.

As shown in Table II, the proposed method has a lower overall error and a higher correct accuracy than MRF and GA-FCM. What is more, it can also run inference at a very fast speed (0.68 s for an image pair with 4096×4096 pixels). Based on the above analysis, the proposed method effectively extracts the vegetation region by utilizing the intensity difference between image pairs and the low-correlation characteristics of vegetation regions.

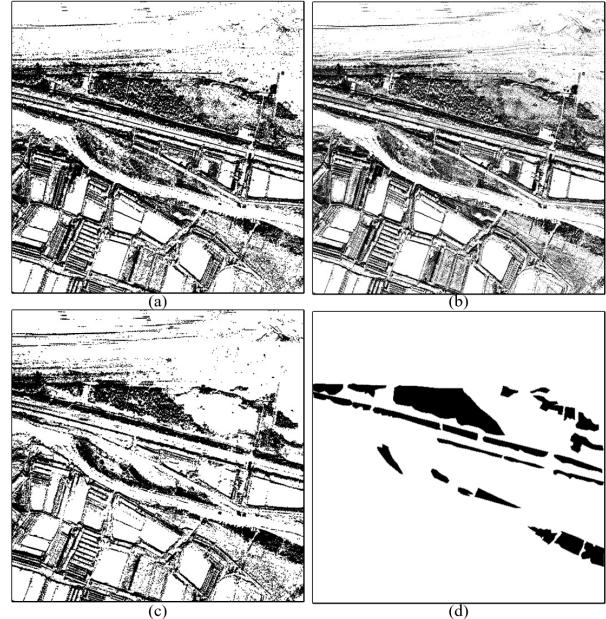


Fig. 10. Vegetation detection results by different methods. (a) MRF method. (b) GA-FCM method. (c) Proposed intensity difference method. (d) Ground truth by manual annotation.

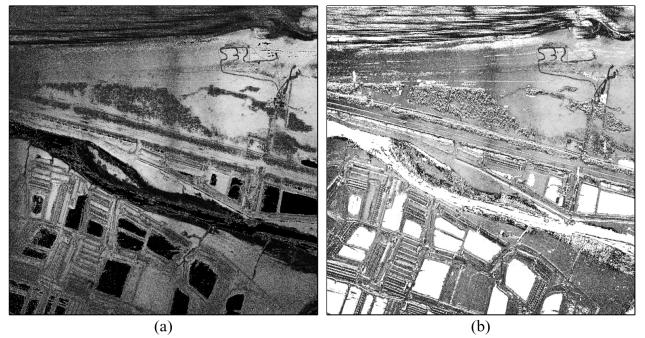


Fig. 11. False alarm removal result. (a) CRCD result. (b) False removal of (a).

3) *Results of False Elimination:* After the extraction of water and vegetation regions from the SAR image pair, the corresponding pixels in a difference image are set as 1, representing unchanged pixels. The CRCD results without false elimination and with false removal by the proposed method are shown in Fig. 11. One can see that the most low-correlation pixels in Fig. 11(a) including water and vegetation have been removed in Fig. 11(b), where the trace regions can be easily discriminated from the background region. It is noticeable that although these false alarm regions have been removed, the other low-correlation regions in Fig. 11(b) still affect the performance of trace detection.

D. Experimental Results of CUnet and Inductive Transfer Learning

The experimental results of the proposed CUnet and inductive transfer learning are analyzed as follows:

1) *Description of Pretraining and Fine-Tune Dataset:* By concatenating the original SAR image in Fig. 6(b), the coherence map estimated by CRCD method in Fig. 7(b) and the false alarm elimination results in Fig. 11(b) along the channel

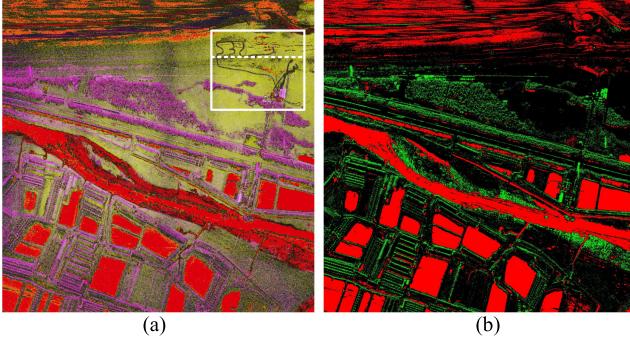


Fig. 12. Generated CTF image and corresponding label. (a) CTF image. (b) Land cover label corresponding to CTF image generated by multiple statistics.

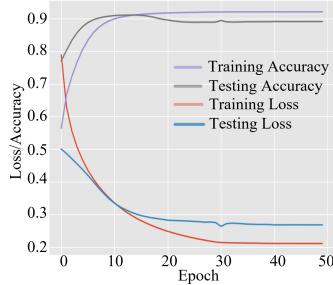


Fig. 13. Pretraining process of CUnet for source task.

dimension, we get a CTF image, as shown in Fig. 12(a). One can see that the trace pixels can be more easily distinguished from other regions in Fig. 12(a), which is a benefit for training a robust detection network. The preclassification results by multiple statistics methods are shown in Fig. 12(b), where red pixels mean water regions and green pixels mean vegetation pixels. In Section III-C, we propose the inductive transfer learning strategy, which consists of a pretraining step based on the source task and a fine-tuning step based on the target task. To provide training and testing data for each task, the pixels outside the rectangular region of Fig. 12(a) are chosen as the source data and the corresponding labels in Fig. 12(b) are chosen as the label of source data for pretraining a network. The pixels inside the rectangular region of Fig. 12(a) are selected as the target data and the corresponding labels in Fig. 7(c) are chosen as the label of target data to fine-tune the trace detection network.

For the pretraining step, we first crop the source data of the CTF image into 128×128 pixels patches with a stride of 112 pixels. The number of image patches of source data is 1330. We choose 80% of them as training samples and 20% of them as testing samples. The stochastic gradient descent (SGD) [56] strategy is used to optimize the cross-entropy loss of CUnet. The batch size is 32 and the epoch size is 50. The initialization learning rate is 0.001, and the momentum is 0.9 for SGD. The weights of CUnet are randomly initialized. It is noticeable that we do not use any data augmentation strategy in the model training. This pretraining process is shown in Fig. 13. One can see that CUnet has a stable convergence process when using the source dataset as training images and the unsupervised classification results as the training labels.

For the fine-tuning step, we crop the white region in Fig. 12(a) into 128×128 pixels patches with a stride

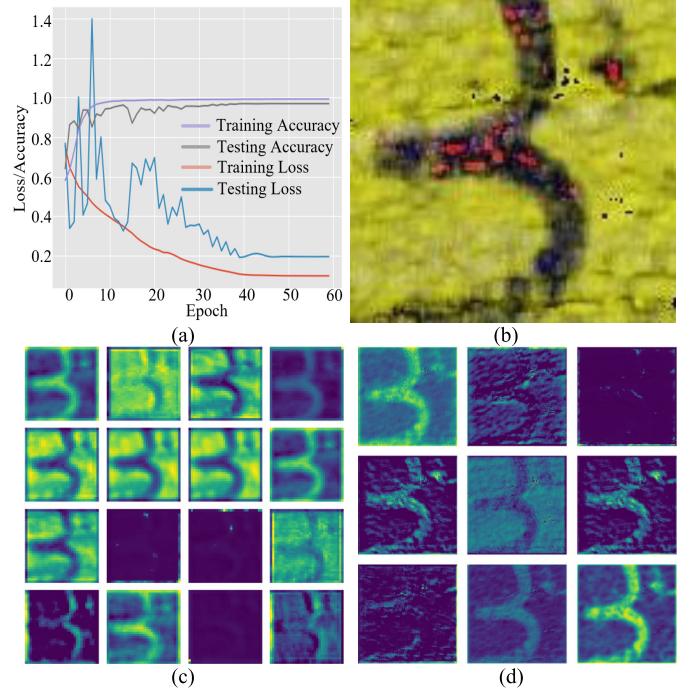


Fig. 14. Fine-tuning process of CUnet for trace detection. (a) Training curve. (b) Testing image patch. (c) Feature maps of the second layer of the third convolution module. (d) Feature maps of the second layer of the sixth convolution module.

of 64 pixels. Different from the cropping stride of source data, the smaller cropping stride here makes more training and testing samples. To understand the detection results on the entire image, the patches corresponding to lower pixels in the white region are taken as the training samples while the patches corresponding to the upper pixels in the white region are considered as the testing samples. What is more, some patches outside the white region are chosen as the negative samples to improve the robustness of the proposed method. With the above strategies, we get 138 patches (49 patches cover trace pixels) for training and 102 patches (69 patches cover trace pixels). Different from the random initialization of CUnet for source task, the weights of CUnet for trace detection are initialized by the well-trained source-task model. The batch size is 32, and the epoch size is 60. The Adam method is used to optimize the cross-entropy loss function. The training curve in Fig. 14(a) shows that the inductive transfer learning based on land cover classification promotes the learning of CUnet for trace detection with a few trace patches. Giving a testing sample in Fig. 14(b), the feature maps extracted by well-trained CUnet are shown in Fig. 14(c) and (d). Fig. 14(c) shows that the encoder in CUnet can extract powerful and abstract features from the trace region. Fig. 14(d) shows that the decoder in CUnet can effectively discriminate the background pixels and trace pixels with the extracted features.

2) Comparisons of Different Detection Methods: To validate the effectiveness of proposed CUnet, a few unsupervised and supervised detection methods are chosen as the comparison methods. For unsupervised-based methods, the most widely used LevelSet [57], OTSU [24] and GA-FCM [55] are selected to detect trace pixels from the CTF image. These methods

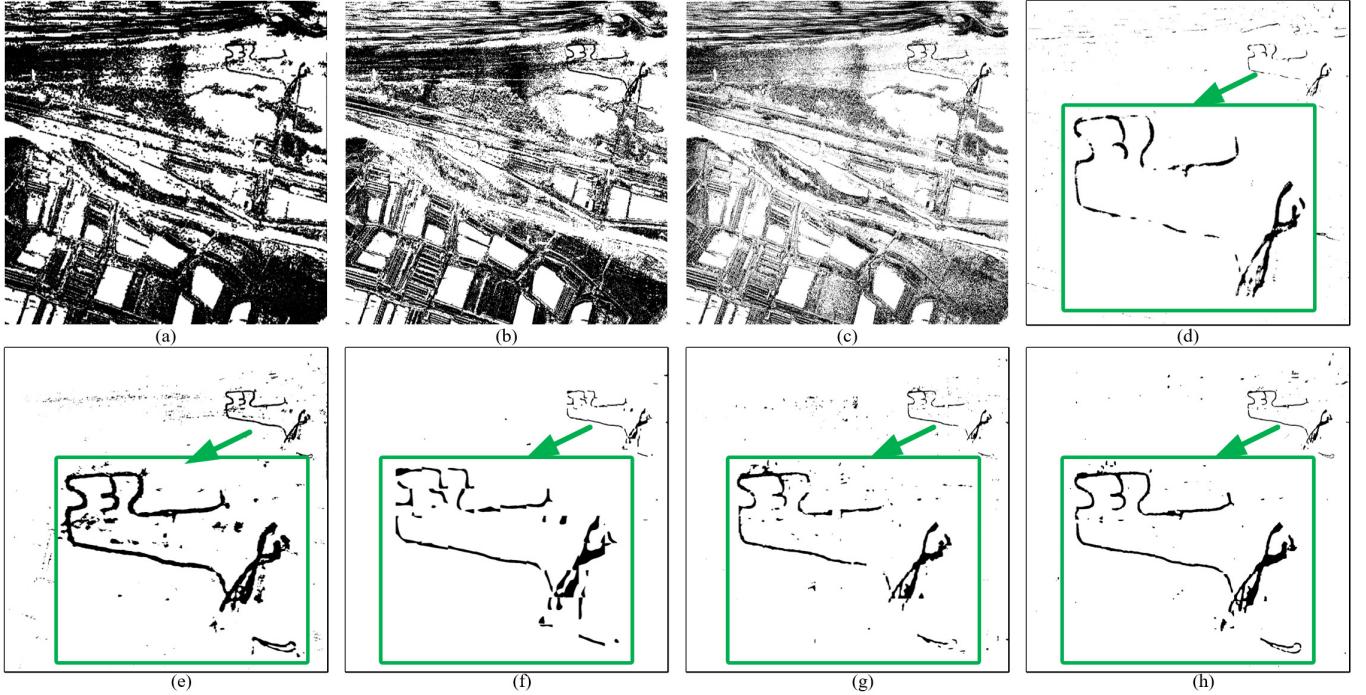


Fig. 15. Trace detection results by different methods. (a) LevelSet. (b) OTSU. (c) GA-FCM. (d) DBN. (e) CNN. (f) SegNet. (g) Unet. (h) CUnet.

TABLE III
QUANTITATIVE ANALYSIS OF DIFFERENT DETECTION METHODS

Learning strategy	Method	Input source	Input Size	Evaluation Criteria					
				P _{FP} (%)	P _{FN} (%)	P _{OE} (%)	P _{CC} (%)	Kappa (%)	Time (s)
Unsupervised	LevelSet	Fine image	4096×4096	0.07	99.05	29.19	70.81	1.24	318.80
	OTSU	Fine image	4096×4096	0.03	99.24	41.20	58.79	0.86	1.07
	GA-FCM	Fine image	4096×4096	0.11	99.12	27.78	72.22	1.10	1682.40
Supervised deep learning	DBN	CTF image	45×45	0.15	67.43	0.52	99.48	40.29	2694.35
	CNN	CTF image	45×45	0.02	70.35	0.75	99.25	44.70	2374.47
	SegNet	CTF image	128×128	0.10	43.47	0.28	99.72	62.70	13.49
	Unet	CTF image	128×128	0.14	49.11	0.32	99.68	58.62	8.57
	CUnet	CTF image	128×128	0.04	40.34	0.24	99.76	70.50	8.23

take the whole CTF image as input and try to use different features, e.g., intensity feature and contour feature to detect interesting objects. For supervised deep learning methods, the sliding window-based DBN [14] and CNN [17] crop patches for each pixel in CTF image and classify the pixels as trace region or background region. To be specific, the size of the cropped patches is 45×45 pixels. Every hidden layer in DBN is pretrained 40 passes through the entire training set, and a 300-250-100-2 network is used. Three convolution layers with a kernel size of 3×3 pixels and two max-pooling layers with a kernel size of 2×2 pixels are used in CNN methods. For the FCNs, the lightweight SegNet [39] and Unet [40] are taken as the comparison methods. These deep learning methods share the same training set. The trace detection results by the proposed method and other comparison methods are presented in Fig. 15. As shown in Fig. 15(a)–(c), although most trace pixels are effectively extracted by unsupervised methods, the detection results are polluted by many noise spots and complicated background regions. The reason mainly lies in that these unsupervised methods cannot extract discriminatory features to differentiate trace and background pixels. By contrast, the results in Fig. 15(d)–(h) show that

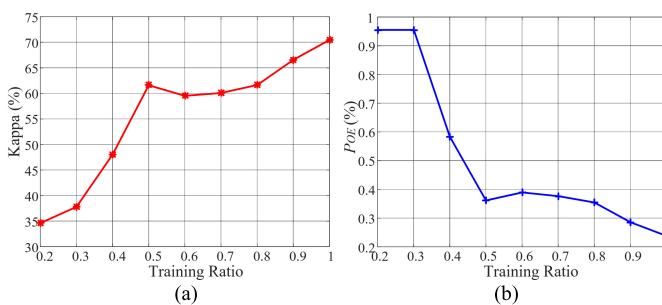
supervised deep-learning methods have greatly decreased the false alarm caused by background pixels, further validating that the introduction of deep learning helps to differentiate trace and other pixels. Among these methods, CNN, Unet, and the proposed CUnet retain clearer trace contour than other methods while there is the least false alarm in the CUnet detection result. Moreover, the quantitative analysis in Table III also declares this point. The P_{CC} yielded by the proposed method equals 99.76%, which is higher than those of other methods, while P_{FN} yielded by the proposed method equals 40.34%, which is lower than those of other methods. Serving as an overall evaluation, the Kappa 70.50% of the proposed method has a big promotion than other methods. What is more, the proposed CUnet runs inference with a fast speed of 8.23 s for an image with 4096×4096 pixels.

3) Comparisons of Different Pretraining Strategies: To verify the performance of the proposed unsupervised pretraining method for trace detection, the other two methods are used to pretrain the CUnet. On the one hand, the CTF image is still taken as the input image of the source task, what is different from proposed unsupervised pretraining is that the manually annotated land cover labels including water, vegetation, and

TABLE IV

TRACE DETECTION RESULTS BY DIFFERENT PRETRAINING METHODS

Criteria	Pre-training Method			
	Non-pre-training	Pre-training by VOC data	Supervised pre-training	Unsupervised pre-training
P_{FP} (%)	0.04	0.05	0.05	0.04
P_{FN} (%)	66.38	47.17	40.13	40.34
POE (%)	0.61	0.28	0.24	0.24
PCC (%)	99.39	99.72	99.76	99.76
Kappa (%)	48.30	66.59	69.91	70.50

Fig. 16. Detection results with different training ratio. (a) Kappa. (b) P_{OE} .

background classes are taken as ground truth of source task. On the other hand, the data in Pascal VOC semantic segmentation task is used to pretrain the proposed CUUnet. Using the well-trained CUUnet model by these pretraining strategies to initialize the CUUnet for trace detection, we can get three corresponding track detection models. Meanwhile, we also train the trace detection model with non-pretraining by random initialization. The performance of these models is shown in Table IV. One can see that the non-pretraining method achieves the lowest classification accuracy and Kappa coefficient. The pretraining by VOC data significantly improves the Kappa coefficient and decreases the detection overall error. Replacing the VOC data with land cover pixels of CTF image, the supervised pretraining and unsupervised pretraining methods achieve higher detection accuracy than other pretraining methods. Notably, the unsupervised pretraining method detects trace pixels with the highest P_{CC} of 99.76% and the highest Kappa of 70.50%, which demonstrates the effectiveness of proposed inductive transfer learning based on unsupervised pretraining on CTF image.

4) Comparisons of Different Number of Training Samples: The CUUnet and inductive transfer learning strategy are proposed to overcome the few samples of trace regions in the CTF image. To validate the adaptability of these methods, we randomly choose some samples from the trace training set to fine-tune the CUUnet. The Kappa and P_{OE} under different ratios of the selected samples to all trace training samples are shown in Fig. 16(a) and (b). One can see that the Kappa has been improved while the P_{OE} has been reduced by increasing the number of training samples. In particular, when the ratio is 0.5, the Kappa and P_{OE} have achieved 61.62% and 0.36%, respectively. Thus, the proposed CUUnet and inductive transfer learning methods have strong adaptability to the number of training samples, and the more training samples, the better detection results.

V. CONCLUSION

To extract subtle trace regions from high-resolution and large-scene SAR images, a CCD method based on multiple statistics and a few-sample deep learning model is proposed in this article. Given two registered repeat-pass repeat-geometry SAR images, this method can accurately locate trace pixels by hierarchically fusing the statistic model and deep learning model. In this method, we first introduce the CRCD estimator to generate the coherence map or difference image between the image pair. Compared with the original CCD estimator, the CRCD estimator can suppress the false alarm and improve the detection probability in the low CNR region. Then, the unsupervised multiple statistics based on intensity summation and intensity difference are proposed to extract water regions and vegetation regions from image pair and remove the low correlation in difference image caused by these regions. Then the CTF image is constructed by combining the original SAR image, the difference image, and the false alarm removal image. Next, we proposed the CUUnet and corresponding inductive transfer learning for trace detection with a few samples. The unsupervised pretraining on source task and supervised fine-tuning on target task can significantly improve the performance of trace detection. Finally, experiment results verify the effectiveness of the proposed method for trace detection. In the future, more measured SAR data will be used to evaluate the detection performance, and more studies will be carried out on the extension of our method.

REFERENCES

- [1] 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.
- [2] M. Kang, K. Ji, X. Leng, X. Xing, and H. Zou, "Synthetic aperture radar target recognition with feature fusion based on a stacked autoencoder," *Sensors*, vol. 17, no. 1, pp. 1–16, 2017.
- [3] J. Chen, J. Zhang, Y. Jin, H. Yu, B. Liang, and D. G. Yang, "Real-time processing of spaceborne SAR data with nonlinear trajectory based on variable PRF," *IEEE Trans. Geosci. Remote Sens.*, early access, Apr. 1, 2021, doi: [10.1109/TGRS.2021.3067945](https://doi.org/10.1109/TGRS.2021.3067945).
- [4] R. Touzi, A. Lopes, J. Bruniquel, and P. W. Vachon, "Coherence estimation for SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 135–149, Jan. 1999.
- [5] R. D. Phillips, "Activity detection in SAR CCD," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2013, pp. 2998–3001.
- [6] H. Chen, C. Wu, B. Du, L. Zhang, and L. Wang, "Change detection in multisource VHR images via deep siamese convolutional multiple-layers recurrent neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2848–2864, Apr. 2020.
- [7] B. Ding, G. Wen, C. Ma, and X. Yang, "An efficient and robust framework for SAR target recognition by hierarchically fusing global and local features," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5983–5995, Dec. 2018.
- [8] M. Cha, R. D. Phillips, P. J. Wolfe, and C. D. Richmond, "Two-stage change detection for synthetic aperture radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6547–6560, Dec. 2015.
- [9] J. Chen, M. Xing, X.-G. Xia, J. Zhang, B. Liang, and D.-G. Yang, "SVD-based ambiguity function analysis for nonlinear trajectory SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3072–3087, Apr. 2021.
- [10] D. E. Wahl, D. A. Yocky, C. V. Jakowatz, and K. M. Simonson, "A new maximum-likelihood change estimator for two-pass SAR coherent change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2460–2469, Apr. 2016.
- [11] J. Jung, D.-J. Kim, M. Lavalle, and S.-H. Yun, "Coherent change detection using InSAR temporal decorrelation model: A case study for volcanic ash detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5765–5775, Oct. 2016.

- [12] G.-C. Sun, Y. Liu, M. Xing, S. Wang, L. Guo, and J. Yang, "A real-time imaging algorithm based on sub-aperture CS-dechirp for GF3-SAR data," *Sensors*, vol. 18, no. 8, p. 2562, Aug. 2018.
- [13] R. Sabry, "Improving SAR-based coherent change detection products by using an alternate coherency formalism," *IEEE Geosci. Remote Sens. Lett.*, early access, May 14, 2020, doi: [10.1109/LGRS.2020.2991760](https://doi.org/10.1109/LGRS.2020.2991760).
- [14] 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.
- [15] 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, Mar. 2019.
- [16] G. Gao and G. Shi, "CFAR ship detection in nonhomogeneous sea clutter using polarimetric SAR data based on the notch filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4811–4824, Aug. 2017.
- [17] 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.
- [18] L. I. Kuncheva and W. J. Faithfull, "PCA feature extraction for change detection in multidimensional unlabeled data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 1, pp. 69–80, Jan. 2014.
- [19] J. D. Villasensor, D. R. Fatland, and L. D. Hinzman, "Change detection on Alaska's North Slope using repeat-pass ERS-1 SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 1, pp. 227–236, Jan. 1993.
- [20] M. Newey, G. Benitz, and S. Kogon, "A generalized likelihood ratio test for SAR CCD," in *Proc. Conf. Rec. 46th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Nov. 2012, pp. 1727–1730.
- [21] R. Touzi and A. Lopes, "Statistics of the Stokes parameters and of the complex coherence parameters in one-look and multilook speckle fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 519–531, Mar. 1996.
- [22] G. Ji, Y. Dong, Y. Bu, Y. Li, L. Zhou, and X. Liang, "Multi-band SAR coherent change detection method based on coherent representation of targets," *J. Radars*, vol. 7, no. 4, pp. 455–464, 2018.
- [23] L. Li, Z. Yang, L. Jiao, F. Liu, and X. Liu, "High-resolution SAR change detection based on ROI and SPP net," *IEEE Access*, vol. 7, pp. 177009–177022, 2019.
- [24] J. Jennifer Ranjani and S. J. Thiruvengadam, "Fast threshold selection algorithm for segmentation of synthetic aperture radar images," *IET Radar, Sonar Navigat.*, vol. 6, no. 8, pp. 788–795, Oct. 2012.
- [25] Y. Zheng, X. Zhang, B. Hou, and G. Liu, "Using combined difference image and k -means clustering for SAR image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 3, pp. 691–695, Mar. 2014.
- [26] O. Yousif and Y. Ban, "Improving SAR-based urban change detection by combining MAP-MRF classifier and nonlocal means similarity weights," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 10, pp. 4288–4300, Oct. 2014.
- [27] 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.
- [28] M.-T. Pham, G. Mercier, and J. Michel, "Change detection between SAR images using a pointwise approach and graph theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2020–2032, Apr. 2016.
- [29] G. Yang, H.-C. Li, W. Yang, K. Fu, Y.-J. Sun, and W. J. Emery, "Unsupervised change detection of SAR images based on variational multivariate Gaussian mixture model and Shannon entropy," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 826–830, May 2019.
- [30] R. Wang, J.-W. Chen, Y. Wang, L. Jiao, and M. Wang, "SAR image change detection via spatial metric learning with an improved Mahalanobis distance," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 1, pp. 77–81, Jan. 2020.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] H. A. Zebker and J. Villasenor, "Decorrelation in interferometric radar echoes," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 5, pp. 950–959, Sep. 1992.
- [35] R. Abdelfattah and J. M. Nicolas, "Mixture model for the segmentation of the InSAR coherence map," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2007, pp. 4479–4482.
- [36] C. Pratola, F. D. Frate, G. Schiavon, and D. Solimini, "Toward fully automatic detection of changes in suburban areas from VHR SAR images by combining multiple neural-network models," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2055–2066, Apr. 2013.
- [37] Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu, "Change detection based on deep siamese convolutional network for optical aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1845–1849, Oct. 2017.
- [38] 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.
- [39] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [40] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Munich, Germany: Springer, Oct. 2015, pp. 234–241.
- [41] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- [42] O. Haddad, R. Abdelfattah, and H. Ajili, "Extracting radar shadow from SAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 2101–2104.
- [43] J. Clarke, "Airborne early warning radar," *Proc. IEEE*, vol. 73, no. 2, pp. 312–324, Feb. 1985.
- [44] F. Ulaby, "Vegetation clutter model," *IEEE Trans. Antennas Propag.*, vol. AP-28, no. 4, pp. 538–545, Jul. 1980.
- [45] W. Feng, H. Sui, W. Huang, C. Xu, and K. An, "Water body extraction from very high-resolution remote sensing imagery using deep U-Net and a superpixel-based conditional random field model," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 618–622, Apr. 2019.
- [46] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, Feb. 2017.
- [47] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 448–456.
- [48] F. Hu, G.-S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14680–14707, Nov. 2015.
- [49] D. Malmgren-Hansen, A. Kusk, J. Dall, A. A. Nielsen, R. Engholm, and H. Skriver, "Improving SAR automatic target recognition models with transfer learning from simulated data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1484–1488, Sep. 2017.
- [50] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [51] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1097–1105.
- [52] G. Sun, X. Jiang, M. Xing, Z.-J. Qiao, Y. Wu, and Z. Bao, "Focus improvement of highly squinted data based on azimuth nonlinear scaling," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2308–2322, Jun. 2011.
- [53] Z. Liu, F. Li, N. Li, R. Wang, and H. Zhang, "A novel region-merging approach for coastline extraction from Sentinel-1A IW mode SAR imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 324–328, Mar. 2016.
- [54] H. McNairn, A. Kross, D. Lapen, R. Caves, and J. Shang, "Early season monitoring of corn and soybeans with TerraSAR-X and RADARSAT-2," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 28, pp. 252–259, May 2014.
- [55] Y. Ding and X. Fu, "Kernel-based fuzzy C-means clustering algorithm based on genetic algorithm," *Neurocomputing*, vol. 188, pp. 233–238, May 2016.
- [56] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Netw.*, vol. 12, no. 1, pp. 145–151, Jan. 1999.
- [57] U. Kandaswamy, D. A. Adjerooh, and M. C. Lee, "Efficient texture analysis of SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2075–2083, Sep. 2005.



Jinsong Zhang (Graduate Student Member, IEEE) was born in Shandong, China, in 1995. He received the B.S. degree in remote sensing science and technology from Xidian University, Xi'an, China, in 2017, where he is pursuing the Ph.D. degree in signal processing with the National Laboratory of Radar Signal Processing.

His research interests include synthetic aperture radar image processing, computer vision, and deep learning.



Mengdao Xing (Fellow, IEEE) received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 1997 and 2002, respectively.

He is a Professor with the National Laboratory of Radar Signal Processing, Xidian University, where he holds the appointment of Dean with the Academy of Advanced Interdisciplinary Research Department. He has authored or coauthored more than 200 refereed scientific journal articles and 2 books about synthetic aperture radar (SAR) signal processing. The total citation times of his research are greater

than 10 000 (H-index 50). He was rated as Most Cited Chinese Researchers by Elsevier. He has achieved over 50 authorized China patents. His research has been supported by various funding programs, such as National Science Fund for Distinguished Young Scholars. His research interests are SAR, SAR interferometry (InSAR), inverse synthetic aperture radar (ISAR), sparse signal processing, and microwave remote sensing.

Dr. Xing has held several special issues on IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE (GRSM) and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He serves as an Associate Editor for radar remote sensing of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the Editor-in-Chief of *MDPI Sensors*.



Guang-Cai Sun (Senior Member, IEEE) received the master's degree in communications engineering from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2006, and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, in 2012.

He is a Professor with the National Laboratory of Radar Signal Processing and also with the Collaborative Innovation Center of Information Sensing and Understanding, Xidian University. He has authored or coauthored 1 book and over 50 articles. His

research interests include imaging of several synthetic aperture radar (SAR) modes, moving target detection, and imaging.



Zhihao Wang (Graduate Student Member, IEEE) was born in Anhui, China, in 1995. He received the B.S. degree in electronic information engineering from Xidian University, Xi'an, China, in 2017, where he is pursuing the Ph.D. degree in signal processing with the National Laboratory of Radar Signal Processing.

His research interests include synthetic aperture radar ground moving target indication (SAR-GMTI), clutter suppression, and moving target detection for spaceborne surveillance radar.