PolSAR Image Classification Based on Low-Frequency and Contour Subbands-Driven Polarimetric SENet

Rui Qin 🆔, Xiongjun Fu 🆔, and Ping Lang 🆔

Abstract—In order to more efficiently mine the features of Pol-SAR images and build a more suitable classification model that combines the features of the polarimetric domain and the spatial domain, this article proposes a PolSAR image classification method, called low-frequency and contour subbands-driven polarimetric squeeze-and-excitation network (LC-PSENet). First, the proposed LC-PSENet introduces the nonsubsampled Laplacian pyramid to decompose polarimetric feature maps, so as to construct a multi-channel PolSAR image based on the low-frequency subband and contour subband of these maps. It guides the network to perform feature mining and selection in the subbands of each polarimetric map in a supervised way, automatically balancing the contributions of polarimetric features and their subbands and the influence of interference information such as noise, making the network learning more efficient. Second, the method introduces squeeze-and-excitation operation in the convolutional neural network (CNN) to perform channel modeling on the polarimetric feature subbands. It strengthens the learning of the contributions of local maps of the polarimetric features and subbands, thereby, effectively combining the features of the polarimetric domain and the spatial domain. Experiments on the datasets of Flevoland, The Netherlands, and Oberpfaffenhofen show that the proposed LC-PSENet achieves overall accuracies of 99.66%, 99.72%, and 95.89%, which are 0.87%, 0.27%, and 1.42% higher than the baseline CNN, respectively. The isolated points in the classification results are obviously reduced, and the distinction between boundary and nonboundary is more clear and delicate. Also, the method performs better than many current state-of-the-art methods in terms of classification accuracy.

Index Terms—Convolutional neural network (CNN), nonsubsampled laplacian pyramid (NSLP), polarimetric synthetic aperture radar (PolSAR) image classification, polarimetric feature, spatial domain, squeeze-and-excitation (SE) network.

I. INTRODUCTION

POLARIMETRIC synthetic aperture radar (PolSAR) is a kind of electromagnetic sensor that can work under all-weather and day-and-night conditions. It provides rich information of the observation area and has been widely used in the field of remote sensing including urban planning, agriculture

assessment, environment monitoring, and military surveillance [1]–[5]. These applications generate some specific research field like terrain classification, change detection, and target recognition [6]–[10]. Terrain classification of PolSAR images is a research hotspot in recent years, which requires the full mining and interpretation of PolSAR images.

Generally, terrain classification of PolSAR images contains two steps: feature extraction and classifier design. Classic feature extraction methods on PolSAR images are mainly based on polarimetric scattering mechanisms [11] and statistical characteristics [12], [13]. As to the polarimetric scattering mechanisms, polarimetric target decomposition is generally concerned by researchers. The Cloude–Pottier decomposition, Pauli decomposition, Freeman–Durden decomposition, and Krogager decomposition are some famous ones [11], [14]–[17]. Based on the basic features extracted by these decomposition methods, some multiple features have been constructed to improve the classification performance [6], [18]–[20]. In addition, some other advanced features like the roll-invariant features are explored by researchers [21]. During classification, unsupervised and supervised classifiers are commonly used, including the fuzzy k-means cluster classifier [22], [23], maximum likelihood classifier [5], support vector machine [24], decision tree [25], sparse representation classifier [26], [27], and neural networks [28], [29]. Considering the statistical distribution of the POLSAR image, some researchers have introduced a unique classifier, namely Wishart classifier [13], [30], [31]. Most PolSAR images classification methods based on the aforementioned feature extractions and classifiers are pixel-based, without considering the spatial relations between pixels, which limits the classification performance.

In recent years, deep learning has been widely used in the fields of image recognition and has achieved great success, including remote sensing image classification [32], [33]. As the most representative technology in deep learning, the convolutional neural network (CNN) automatically extracts features hierarchically from image, which can better represent the relations between pixels [34]–[39]. A large number of researchers have also introduced CNN-based methods to the research of the PolSAR image classification [18], [40]–[44]. Zhou *et al.* extracted 6-D real-value features from the **T** matrix to represent each pixel of the PolSAR image, and fed the six-channel real image into a tailored CNN for classification, which achieved better performance than directly using the real value of the

T matrix as the input representation of the CNN [18]. Chen and Tao [43] proposed a polarimetric-feature-driven deep CNN method for PolSAR image classification, which combines expert knowledge of target scattering mechanism interpretation and polarimetric feature mining to drive the CNN model. The method uses both classical roll-invariant polarimetric features and hidden polarimetric features in the model to achieve high classification accuracy. Considering the phase of the elements in the **T** matrix, Zhang *et al.* proposed to use complex-valued CNN for the PolSAR image classification [44]. Experiments have shown that the classification accuracy is improved compared to the real-valued CNN. In addition to the CNN, the convolutional autoencoder (CAE) has good feature representation learning capability [45]. Geng *et al.* proposed a PolSAR classification method based on the deep CAE, and verified its performance [46]. Considering the statistical characteristics of the **T** matrix of PolSAR data, Xie *et al.* introduced the Wishart distance as the reconstruction loss of the CAE, and proposed the W-CAE method [7]. The aforementioned methods mainly use polarimetric scattering features and statistical characteristics of each pixel to form several polarimetric maps, and feed them to the CNN-based network for the automatic feature extraction and classification. Therefore, both polarimetric features and spatial features are considered.

However, there are still two aspects to be improved. First, due to the influence of speckle noise and other factors, there are many isolated points in the classification result. Second, the commonly used CNN-based methods assume that the feature maps in one layer have the same importance to the next layer when extracting features. This type of model is not the most suitable for the PolSAR classification since features from different polarimetric feature maps may contribute to the classification inequality. Zhang *et al.* proposed a stacked sparse autoencoder (SSAE) using the local spatial information of the pixels by controlling the influences of the neighbor pixels on the central pixel referring their spatial distances, and they verified that the connectivity of the classification result is obviously improved [47]. Guo *et al.* [48] and Liu *et al.* [49] introduced a postprocessing technique called clean to filter out the isolated points of the classification map after classification using WRBM and W-DBN, which enhanced the classification accuracy. But these methods require manual modeling of pixel relationships or setting thresholds rely on experience. Liu *et al.* proposed a coding method for the complex scattering matrix **S** of each pixel, which expands one pixel of the original image to $4*4$ pixels so that the convolution kernel not only mine the spatial features between the original pixels, but also the relations of elements in the **S** matrix of each pixel [50]. But it is difficult to encode more polarimetric features, because the image becomes very large, and the size of the convolution kernel is difficult to determine.

To better deal with the aforementioned two problems, this article proposes a new solution, called low-frequency and contour subbands-driven polarimetric squeeze-and-excitation network (LC-PSENet). First, the proposed LC-PSENet introduces the nonsubsampled Laplacian pyramid (NSLP) to decompose each polarimetric feature map into a low-frequency subband and a contour subband, which helps to construct a multichannel PolSAR image based on these polarimetric subband maps. It can guide the network to perform feature mining and selection in the low-frequency and contour (LC) subbands, automatically balancing the contribution of polarimetric features and their subbands and the influence of interference information such as noise, making learning more efficient, and enhancing the continuity of the classification results. Second, the method introduces squeeze-and-excitation (SE) operation in the CNN to perform channel modeling on the multichannel image composed of the subbands of polarimetric feature maps. It strengthens the learning of the contributions of local maps of the polarimetric features and subbands, which more effectively combines the features of the polarimetric domain and the spatial domain, thereby improving the classification accuracy. The proposed method provides new ideas for reducing the isolated points in the classification results and efficiently combining the polarimetric domain features and spatial domain features for PolSAR image classification. Experiments on three real PolSAR datasets show that the proposed LC-PSENet significantly improves the classification accuracy of the baseline CNN, the isolated points in the classification results are significantly reduced, and the distinction between boundaries and nonboundaries is more clear and refined. In addition, the classification results are also superior to many recent state-of-the-art methods.

The main contributions of this article can be summarized as follows.

1) This article proposes to decompose the PolSAR image to obtain the LC subbands, and guide the network to mine the two types of subbands to varying degrees under supervision, making the learning of network more efficient. And the NSLP method distributes the noise unevenly to the subbands, which increases the flexibility of network, and helps to balance the impact of noise.

2) This article proposes to perform channel modeling on feature maps composed of polarimetric features and spatial subbands. It strengthens the balance between the contributions of polarimetric feature maps and their subbands, and the influence of the noise, thereby, effectively combining the features of the polarimetric domain and the spatial domain.

3) In this article, the supervised thinking in deep learning is combined with signal processing based image decomposition and enhancement, which provides a new idea for PolSAR image classification to efficiently combine the polarimetric domain and the spatial domain and reduce the isolated points of the classification results.

The remainder of this article is organized as follows. Section II presents an introduction and analysis of the proposed LC-PSENet. Experiments and results analysis on three real datasets are described in Section III. Section IV concludes this article.

II. METHODOLOGY

In this section, we present an introduction and analysis of the proposed LC-PSENet in detail. First, we introduce the representation of PolSAR data and extract polarimetric features. Then,

the PolSAR LC subbands image construction and polarimetric channel modeling methods are concerned and analyzed. Finally, we present an overview of the proposed method.

A. PolSAR Data and Polarimetric Feature Extraction

Polarimetric target decomposition method can extract features from each resolution cell of the PolSAR image. A variety of polarimetric decomposition features constitute a series of polarimetric feature maps with the same size as the original PolSAR image, which provides rich information for the PolSAR image classification. When these polarimetric feature maps are fed to a CNN-based model, features of polarimetric domain and spatial domain are combined.

1) Representation of PolSAR Images: PolSAR images are 2-D high-resolution radar images. In each resolution cell, the basic PolSAR data are expressed by a 2×2 complex scattering matrix

$$S = \begin{bmatrix} S_{\mathrm{HH}} & S_{\mathrm{HV}} \\ S_{\mathrm{VH}} & S_{\mathrm{VV}} \end{bmatrix} \tag{1}$$

where S_{HV} is the scattering element of the horizontal transmitting and vertical receiving polarization, and the other three elements are similarly defined. H and V denote the bases of horizontal polarization and vertical polarization, respectively. According to the reciprocity theorem and Pauli decomposition [14], the scattering matrix can be also represented as a 3-D vector

$$\boldsymbol{k} = \frac{1}{\sqrt{2}} [S_{\mathrm{HH}} + S_{\mathrm{VV}}, S_{\mathrm{HH}} - S_{\mathrm{VV}}, 2S_{\mathrm{HV}}]^{\mathrm{T}} \tag{2}$$

where the superscript T denotes the transpose. Then, the polarimetric coherency matrix is obtained as

$$\mathbf{T} = \frac{1}{L} \sum_{i=1}^{L} \boldsymbol{k}_i \boldsymbol{k}_i^{\mathrm{H}} = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix} \tag{3}$$

where L is the number of looks, the superscript H denotes conjugate transpose, respectively, and T is the element of $\mathbf{T}$. Apparently, it is a Hermitian matrix whose diagonal elements are real numbers, while off-diagonal elements are complex numbers. Hence, its upper triangular elements $\{T_{11}, T_{12}, T_{13}, T_{22}, T_{23}, T_{33}\}$ can fully represent the scattering characteristics of PolSAR image.

In the research of the CNN-based PolSAR image classification, Zhou et al. [18] proposed a new 6-D real vector representation of the PolSAR image extracted from $\mathbf{T}$, which is tailored for neural networks as follows:

$$A = 10\log_{10} (\mathrm{SPAN}) \tag{4}$$

$$B = T_{22}/\mathrm{SPAN} \tag{5}$$

$$C = T_{33}/\mathrm{SPAN} \tag{6}$$

$$D = |T_{12}| / \sqrt{T_{11} \cdot T_{22}} \tag{7}$$

$$E = |T_{13}| / \sqrt{T_{11} \cdot T_{33}} \tag{8}$$

$$F = |T_{23}| / \sqrt{T_{22} \cdot T_{33}} \tag{9}$$

where A is the total scattering power in decibel of all polarization channels, $\mathrm{SPAN} = T_{11} + T_{22} + T_{33}$, B and C are the normalized ratio of the power of T_{22} and T_{33}, respectively, and D, E, and F are the relative correlation coefficients.

All polarimetric features used in the following research are based on the aforementioned representations.

2) Polarimetric Feature Extraction: Based on the target decompositions (TD) theorems, many features with actual physical meaning can be extracted from the coherency matrix T. A number of works on PolSAR detection and classification have introduced these decomposition features [6], [21]. Cloude–Pottier decomposition [11] is a famous TD method based on eigenvalue–eigenvector, which is widely used in classification. Through this decomposition, the coherency matrix $\mathbf{T}$ can be parameterized as

$$\mathbf{T} = \mathbf{U}_3 \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \mathbf{U}_3^{\mathrm{H}} \tag{10}$$

where λ_1, λ_2, and λ_3 are the eigenvalues of the coherency matrix $\mathbf{T}$, and $\mathbf{U}_3$ represents a matrix composed of the three decomposed eigenvectors. Based on the decomposed eigenvalue–eigenvectors, three typical polarimetric features can be extracted, namely entropy H, mean alpha angle $\bar{\alpha}$, and anisotropy Ani.

$$H = -\sum_{n=1}^{3} P_n \log_3 P_n, \ \bar{\alpha} = \sum_{n=1}^{3} P_n \alpha_n, \ \mathrm{Ani} = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3} \tag{11}$$

where $P_n = \lambda_n/(\lambda_1 + \lambda_2 + \lambda_3)$, and $\cos(\alpha_n)$ is the magnitude of the first component of the nth coherency matrix eigenvector. Both the three eigenvalues and the derived features are commonly used in PolSAR image classification works [31].

Based on the aforementioned representation and feature extraction of the PolSAR data, we can use a 21-D real vector to represent each pixel of the PolSAR image. Obviously, there is information redundancy in this feature vector because the $\mathbf{T}$ matrix can completely represent the PolSAR image. However, the superposition of multiple polarimetric feature maps may enrich the spatial relations between pixels. Therefore, we conducted a preliminary experiment on the Flevoland dataset using the baseline CNN, and analyzed the 21-D features, thereby, selecting a set of features for the verification of the proposed method in this article. First, we divided the 21-D features into three categories for combined testing. These three categories are real value from $\mathbf{T}$ matrix, the 6-D real vector representation, and the Cloude–Pottier features, respectively, which are expressed as $\mathbf{T}$, 6-D, and Cloude. The combination test accuracy of $\mathbf{T}$, 6-D + Cloude, $\mathbf{T}$ + Cloude, $\mathbf{T}$ + 6-D, and $\mathbf{T}$ + 6-D + Cloude were 98.29%, 97.60%, 98.66%, 98.75%, and 98.80%, respectively. The results show that in the CNN-based classification, T matrix contributes the most among these three types of features, and the superimposed features have obvious advantages over the use of one type of these features alone. However, the test accuracy gap between $\mathbf{T}$ + 6-D and $\mathbf{T}$ + 6-D + Cloude is not very big. Therefore, we retain the $\mathbf{T}$ matrix and only consider removing a few features that are extremely

redundant with the **T** matrix. In these features, features B and C are almost the same as T_{11}, T_{22}, and T_{33}, and λ_3, $\bar{\alpha}$, and *Ani* can also fully express λ_1 and λ_2. In addition, [30] has proved that the mean scattering angle $\bar{\alpha}$ can help to improve the discrimination of pixels in the PolSAR image. Combining the aforementioned tests and analysis, we chose 16 features to form a 16-D vector to represent the PolSAR image, which can achieve classification accuracy comparable to that when using all 21-D features in the baseline CNN model, showing 98.79%. The selected 16-D feature is∗∗ expressed as $[T_{11}, T_{22}, T_{33}, R(T_{12}), I(T_{12}),$ $R(T_{13}), I(T_{13}), R(T_{23}), I(T_{23}), \lambda_3, \bar{\alpha}, \mathrm{Ani}, A, D, E, F]^{\mathrm{T}}$, where $R(\cdot)$ and $I(\cdot)$ are the real and imaginary parts of the complex $\cdot$, respectively.

B. PolSAR LC Subbands Image Construction Based on the NSLP

This section introduces the NSLP method to decompose the aforementioned 16 polarimetric feature maps to obtain subband maps, and analyzes the characteristics of each subband type. After that, it uses the LC subbands to expand the polarmetric channels, thereby, constructing a polarmetric subband-based multichannel PolSAR image.

1) Principle of the NSLP: The NSLP is an improved Laplacian pyramid (LP) decomposition algorithm. It retains the multiscale characteristics of the LP algorithm and can perform feature decomposition on images at different scales. They are both widely used in contour transformation of images. Since LP performs downsampling after low-pass filtering at each layer, the size of the decomposed image changes without translation invariance, which is not suitable for pixel-based classification. The NSLP draws on the idea of the atrous algorithm to use a nonsubsampled pyramid filter bank (NSPFB) to design a filter structure with translation invariance [51], which has multiscale properties similar to LP. Difference from LP decomposition, NSLP does not downsample the components after filtering, but upsamples the filter accordingly. That is to say, the secondary filter of the NSLP can be obtained by upsampling the filter of the previous stage with a step size of 2. When an image undergoes the L-level NSLP decomposition, $L + 1$ subband images with the same size as the original image can be obtained. Fig. 1 shows the structure of the three-layer NSLP decomposition, where H_0 represents the low-pass decomposition filter, H_1 represents the band-pass filter, and H_0^2 represents the filter obtained by up-sampling the filter H_0 with a step size of 2. The other filters in this figure also follow this representation style.

2) LC Subbands PolSAR Image Construction: Due to the translation invariance of the NSLP, a polarimetric feature map can be decomposed into multiple subband maps with the same size without changing the positional relationship of pixels [51]. The NSLP continuously decomposes the image at multiple scales, so each subband map can reflect features at different scales. Choosing the appropriate subband maps to replace the original polarimetric feature maps as the input of the CNN will be helpful to improve the learning efficiency of the network and obtain even better features.
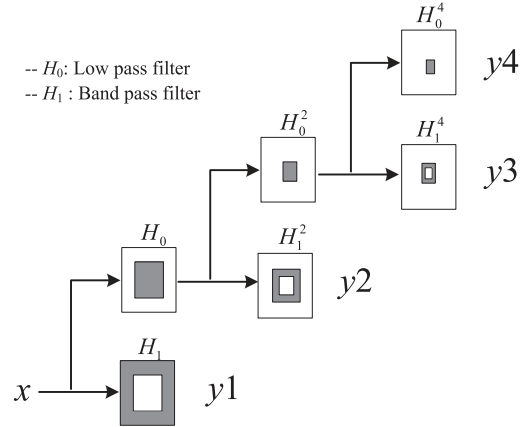

Fig. 1. Structure of three-layer NSLP decomposition.

Let the ith polarimetric feature map be x_i, it can be decomposed into $L + 1$ subband maps through the L-layer NSLP decomposition, expressed as

$$x_i(m, n) = y_i^1(m, n) + y_i^2(m, n) + \ldots + y_i^{L+1}(m, n) \tag{12}$$

where L is the number of layers of the NSLP decomposition, (m, n) represents the pixel position, that is, the pixel in the mth row and nth column. The decomposed $L + 1$ subband maps express as $\{y_i^1, y_i^2, \ldots, y_i^{L+1}\}$, where y_i^{L+1} is the low-frequency subband map, and $\{y_i^1, y_i^2, \ldots, y_i^L\}$ is the high-frequency component in each layer.

Taking the T_{11} feature map as an example, Fig. 2 shows the subband maps decomposed by the three-layer NSLP, where Fig. 2(a) is the T_{11} feature map, Fig. 2(b) is the decomposed low-frequency subband map, and Fig. 2(c)–(e) are the high-frequency sub-band maps of the decomposition output from the third layer to the first layer. As shown in the figure, Fig. 2(a) contains most of the information of the original image, the information of the homogeneous region is relatively complete, and the speckle noise in this map is much weaker than the original image. But the image becomes blurred, especially the boundaries between different categories. The reason is that during the NSLP decomposition process, the low-frequency subbands of each layer are continued to be decomposed in the next layer, and the scale of the filter is doubled for each additional layer of decomposition, making the low-frequency subband has the largest decomposition scale. The large-scale low-frequency filtering makes the speckle noise coefficient of the image very weak [52]. But for the high-frequency subband of each layer, the NSLP no longer decomposes them. As shown in Fig. 2(c)–(e), they contain less useful information than the low-frequency subband, and are accompanied by noise with different intensities. In other words, the main information and noise of the image are unequally distributed in different subbands. Moreover, Fig. 2(c) is the high-frequency part decomposed by the last layer. It has a larger filtering scale and can reflect the contour information of the PolSAR image well. And because part of the speckle noise energy is distributed in higher frequency subbands, this contour subband contains only the weakened noise. In Fig. 2(d), the

(a)

(b)   (c)

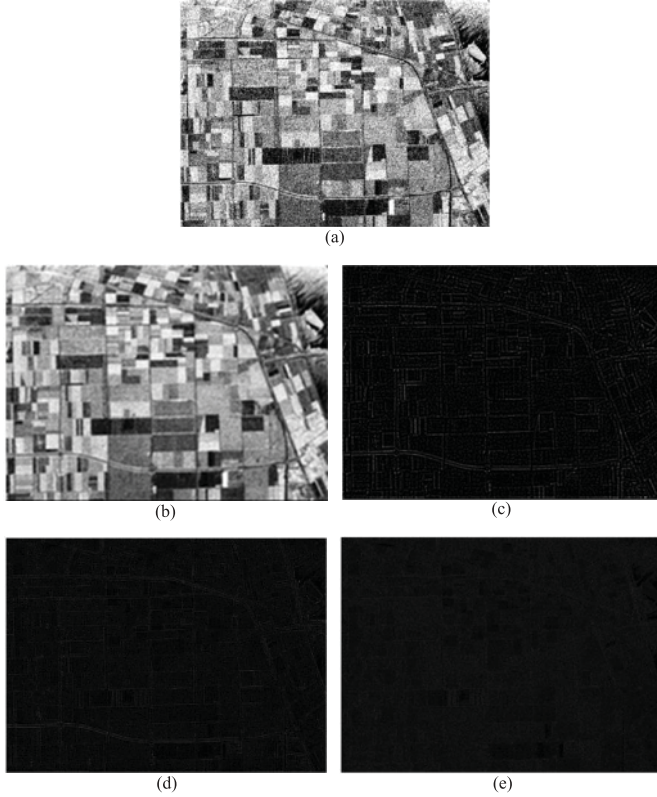(d)   (e)

Fig. 2. T_{11} map and the subband maps decomposed by three-layer NSLP. (a) T_{11} map. (b) Low-frequency subband map. (c) High-frequency subband map of the third layer. (d) High-frequency subband map of the second layer. (d) High-frequency subband map of the first layer.

contour and noises have similar power intensity, which is easy to be confused. Fig. 2(e) contains almost no useful information of the image, and is mainly manifested as noise.

Therefore, this article extracts the LC subbands of each polarimetric feature map to construct a PolSAR image based on these subband maps. That is to say, we choose the subband maps y_i^{L+1} and y_i^L to convert the 16 polarimetric feature maps into 32 subband feature maps to represent the PolSAR image, as shown in Fig. 3. Thus, a multichannel PolSAR image based on polarimetric subband maps is constructed, and it can be expressed as $\{y_1^{L+1}, y_1^L, y_2^{L+1}, y_2^L, \ldots, y_{16}^{L+1}, y_{16}^L\}$. When the subbands of polarimetric feature maps are used as the input of a CNN-based model, the features in the low-frequency subband and the contour subband will be mined to varying degrees under supervision, and the impact of noise on the classification will also be automatically weighed following the varying degrees of feature mining of the two types of subband. That is, the features and noise of the image will be reorganized under the supervision of the classification network. Compared with the direct use of polarimetric feature maps as input, this method is more flexible and advantageous.

C. Polarimetric Channel Modeling Based on Squeeze-and-Excitation Networks (SENet)

In addition to the overall contribution of polarimetric features and the subbands, the contribution and noise level of the local
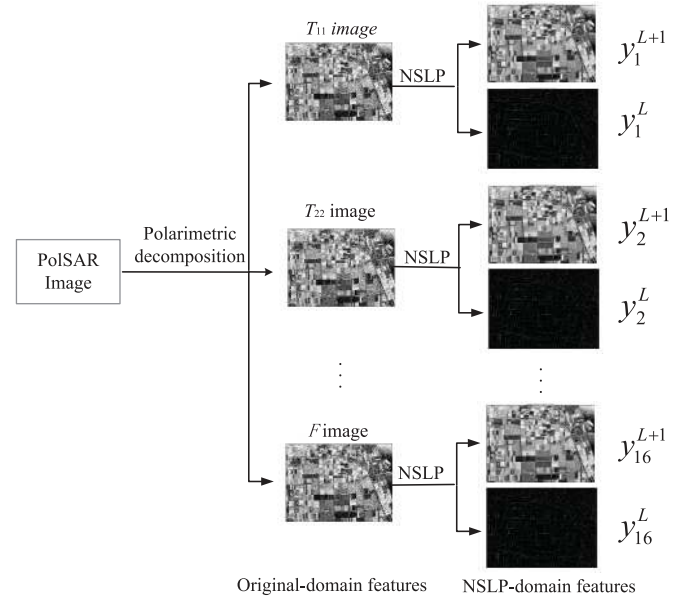

Fig. 3.   Schematic diagram of converting 16 polarimetric feature maps to 32 subband feature maps.

maps of the subbands are also unequal. However, a typical CNN mainly uses convolution kernels to extract image features hierarchically, it assumes that the feature maps in one layer have the same importance to the next layer. Such an assumption in this study is not the most appropriate. Here, we will discuss the modeling of polarimetric feature channels in the CNN, which helps to design a more suitable model for PolSAR image classification.

1) SE Modeling on Polarimetric Maps: In this research, we introduce the SE module into our network to modeling the polarimetric maps so that we can easily learn the weights of the polarimetric feature maps in the convolution layers. Thus, it strengthens the allocating of the contribution of each polarimetric feature and its sub-bands.

The SENet can model the channels of feature maps in the CNN [53]. Huang and Wang applied the network to the field of SAR image recognition, and proposed the group squeeze excitation sparsely connected convolutional networks [54] and enhanced SENet [55], respectively. Fig. 4 shows the network structure of an SE module. The input feature maps can be express as a tensor $\mathbf{U}, \mathbf{U} \in \mathbf{R}^{W \times H \times C}$, where $\mathbf{R}$ represent the set of real numbers, W and H represent the width and height of the feature map, and C represents the number of feature maps. As shown in the figure, the SE module constructs a branch network in the sequential model, which contains the global average pooling, the fully connected layers, and the corresponding activation functions, where r is a fixed hyperparameter in the module. The branch network first outputs a parameter with a size of $1 \times 1 \times C$, expressed as $\mathbf{S}, \mathbf{S} \in \mathbf{R}^{1 \times 1 \times C}$, which has the same number of channels as the input $\mathbf{U}$. After that, a scale operation multiplies the input $\mathbf{U}$ by the parameter $\mathbf{S}$ and outputs a new feature map tensor $\mathbf{X}$, where $\mathbf{X}$ shares the same size with $\mathbf{U}$, i.e., $\mathbf{X} \in \mathbf{R}^{W \times H \times C}$. This operation is expressed as $\mathbf{X} = F_{\text{scale}}(\mathbf{U}, \mathbf{S}) = \mathbf{U} \cdot \mathbf{S}$, and $\mathbf{S}$ is
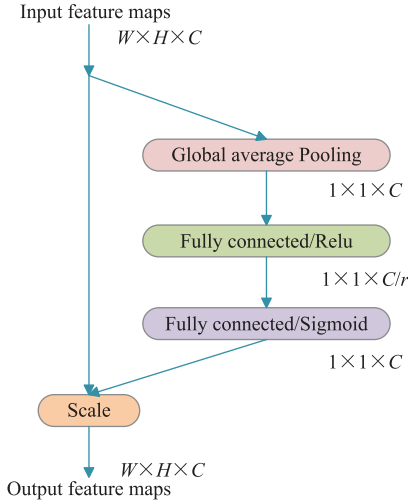
Fig. 4. Structure of the SE module.

Fig. 5. Designed PSENet.

equivalent to the weights of the feature maps, which allows the network to automatically learn the contribution of each feature map in the layer.

For the PolSAR image classification in this article, each channel of $\mathbf{U}$ represents the feature map learned from different polarimetric feature subbands, and the SE module can learn the importance of these feature maps, which indirectly learns the importance of polarimetric features and the LC subbands. Therefore, by introducing the SE module to modeling the polarimetric feature channel in a CNN model, a polarimetric SENet (PSENet) is tailored for PolSAR images classification. Thus, the network has the capability to efficiently mine features from the polarimetric domain and spatial domain, and is more suitable for PolSAR image classification than the baseline CNN. First, from the perspective of the spatial domain of the polarimetric feature maps, the convolution operation will learn the spatial relations between pixels such as the geometric texture of the polarimetric feature maps. Second, for polarimetric features, the SE module builds a 1-D deep neural network indirectly for learning the weight of the channels, which allocates weights to the polarimetric features to compute their contributions.

2) Design SENet for Polarimetric Maps: Since this article is a study of pixel-level classification of PolSAR images, pixel-centric patches will be used as the input of the network. Therefore, in this study, we design a general PSENet network, which is suitable for the classification of PolSAR images of different sizes. The detailed design of our PSENet is shown in Fig. 5. It contains two convolution layers, two fully connected layers and an output layer, and the first convolution layer follows a max pooling layer. Besides, a rectified linear unit (ReLu) activation function is added to each convolution layer and fully connected layer, and the output layer is activated by a softmax function with the crossentropy loss function. The convolution and pooling layers have no padding. In order to model the channels of different polarimetric feature maps and their subband maps, an SE module is added after the first convolution layer of the network. This operation includes two fully connected layers,
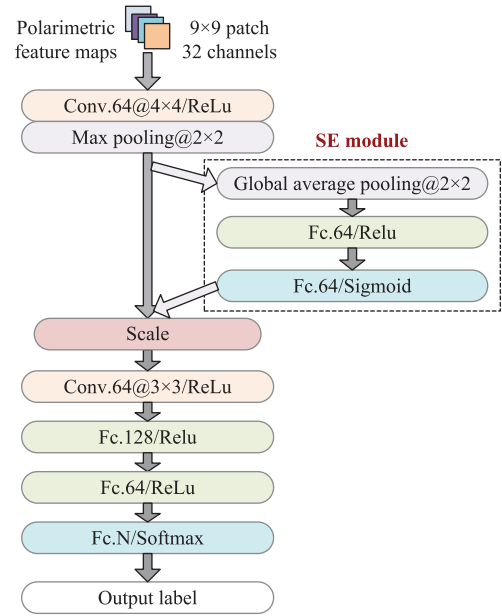
and they are activated by the ReLu function and the Sigmoid function, respectively. After that, a scale operation is used to fusion the outputs. The detailed parameters in each layer of the network are also shown in the figure, including kernel sizes, fully connected unit numbers, and channel numbers.

In this network, the convolution layer shares the weights to preserve the space locality by the sliding convolutional kernels, which considers the spatial relations of pixels in each subband maps. And the SE module allocates weights to the learned feature maps, that is, the network learns the contribution of different polarimetric feature maps and their LC subbands to classification.

D. Processing Flow of the Proposed LC-PSENet

In this section, we briefly present an overview of the proposed LC-PSENet. The processing procedure of the proposed algorithm is shown in Fig. 6. First, a number of features are extracted by polarimetric representation and the target decomposition algorithms. These features mainly reflect the polarimetric scattering mechanisms and statistical characteristics of PolSAR images in a resolution cell. According to the analysis in Section II-A, a total of 16 polarimetric features are selected including the real value of the coherent matrix in this research. Thus, 16 polarimetric feature maps with the same size and relative position of pixels as the original PolSAR image are obtained. Second, we introduce an NSLP to decompose each feature map and select two subband maps, namely LC subband. The nonsubsampled decomposition method ensures that the size of the LC subbands is the same as the size of the polarimetric feature maps, and does not change the positional relationship of the pixels. In this way, 16 polarimetric feature maps are decomposed to obtain 32 subband maps, which can be expressed as $\{y_1^{L+1}, y_1^L, y_2^{L+1}, y_2^L, \ldots, y_{16}^{L+1}, y_{16}^L\}$. We separately
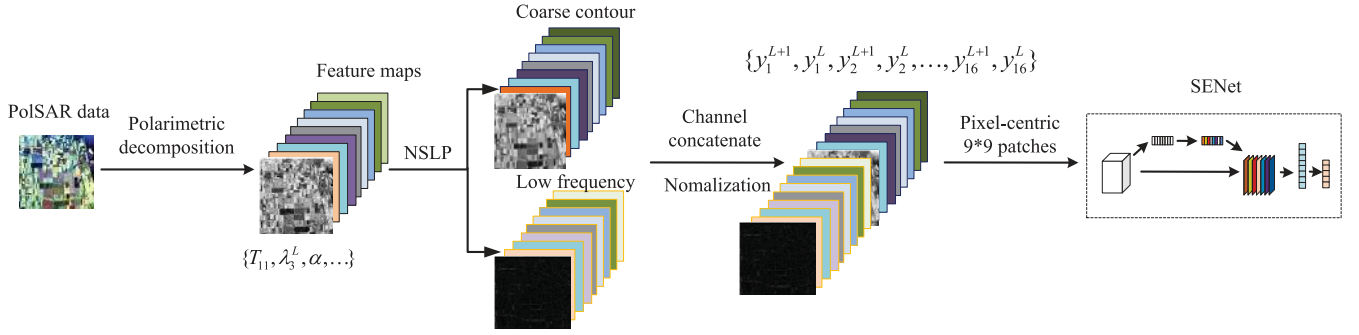
Fig. 6.　　Processing flow of the LC-PSENet.

normalize the 32 subband maps by using the Z-score technique to form the original input 32-channel image, which can guide neural networks to focus on the study of the homogeneous region and boundary region features of PolSAR images. Third, taking each pixel as the center and extracting a 9×9 patch, each pixel of the PolSAR image can be represented as a $9 \times 9 \times 32$ data, which we use as the direct input of the neural networks. In particular, this operation employs a zero-padding strategy for pixels near the image boundary. That is, a 9×9 patch centered at a pixel near the boundary will be zero padded for those pixels outside the image. Finally, an SENet tailored for the PolSAR image classification is designed. We randomly divide the samples into three parts according to a certain ratio, which are used as training set, verification set and test set, respectively. In this network, along the channel direction of the input image, pixel at any position (m, n) can be expressed as $\{y_1^{L+1}(m, n), y_1^L(m, n), y_2^{L+1}(m, n), \ldots, y_{16}^L(m, n)\}$, which represents the features of the polarimetric subbands within that pixel. The channel modeling of the SENet network automatically learns the contributions of polarimetric feature maps and the LC subbands. In our model, when the SE module acts on the output of the convolutional layer, the result of the channel modeling is

$$
\begin{aligned}
F_{eq} = &\left[\left(w_1^{y_1^L} m_1^{y_1^L} + \cdots + w_a^{y_1^L} m_a^{y_1^L} + \cdots \right) \right. \\
&\left. + \left(w_1^{y_1^{L+1}} m_1^{y_1^{L+1}} + \cdots + w_b^{y_1^L} m_b^{y_1^L} + \cdots \right) \right] + \cdots + \\
&\left[\left(w_i^{y_i^L} m_i^{y_i^L} + \cdots + w_p^{y_i^L} m_p^{y_i^L} + \cdots \right) \right. \\
&\left. + \left(w_i^{y_i^{L+1}} m_i^{y_i^{L+1}} + \cdots + w_p^{y_i^{L+1}} m_p^{y_i^{L+1}} + \cdots \right) \right] \\
&+ \cdots + \left[\left(w_{16}^{y_{16}^L} m_{16}^{y_{16}^L} + \cdots + w_u^{y_1^L} m_u^{y_{16}^L} + \cdots \right) \right. \\
&\left. + \left(w_{16}^{y_{16}^{L+1}} m_{16}^{y_{16}^{L+1}} + \cdots + w_v^{y_{16}^L} m_v^{y_{16}^L} + \cdots \right) \right]
\end{aligned}
\tag{13}
$$

where $m_p^{y_i^L}$ represents the pth feature map from the L (contour) subband of the ith polarimetric feature image, $w_p^{y_i^L}$ is the corresponding channel weight; $m_q^{y_i^{L+1}}$ is the qth feature map from the $L+1$ (low-frequency) subband of the ith polarimetric feature image, and $w_q^{y_i^{L+1}}$ is the corresponding channel weight. The other variables are similarly defined. It thus captures not only the weighted polarimetric features and LC subbands, but also the spatial relations surrounding the center pixel.

III. EXPERIMENT AND DISCUSSION

In this section, the performance of the proposed algorithm is evaluated and analyzed on three real POLSAR datasets. The overall accuracy (OA) and the classification result map are used to evaluate the performance of the proposed LC-PSENet. Moreover, the baseline CNN, PSENet, and LC-CNN is used to compare with the proposed LC-PSENet. This comparison experiment mainly demonstrates the effectiveness of LC subbands image construction and the SE operate in the CNN model. All these comparison algorithms have almost the similar experimental parameters in the classification model, including the number of convolution layers, the size of input patches, and the size of convolution filter in each convolution layer, which makes the comparison fair. Based on the network design in Section II, the degrees of freedom (DoF) of parameters of the aforementioned four comparison networks are 70927, 79247, 87311, and 95631, respectively. Also, to further demonstrate the competitive performance of the proposed LC-PSENet, some recent state-of-the-art methods are used for another comparison, which are CV-CNN [44], PCN [50], SSAE [47], W-DBN [49], SF-CNN [43], and WCAE [7].

In our experiment data, for example, in the coherency matrix **T** of the Flevoland dataset, T_{11}, T_{22}, and T_{33} of some pixels at the boundary are zero. These zero values may be measurement errors of some pixels at the boundary, which causes abnormal extraction of some features such as feature D in the pixel. In order not to affect feature extraction, we correct these abnormal data before feature extraction. Through the approach of near assignment, these data are corrected before feature extraction. Specifically, for the zero values of the four vertices of the image, we assign the corresponding values of the pixels in the diagonal direction of the three adjacent pixels to them, and for zero values
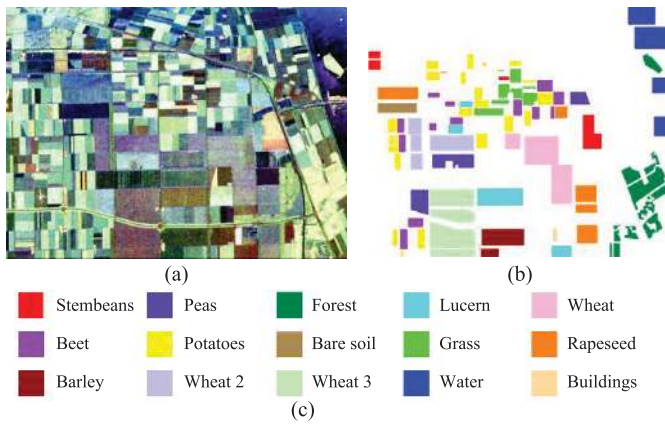
Fig. 7. Flevoland dataset. (a) Pauli RGB of the PolSAR image. (b) Ground-truth map. (c) Legend of the ground truth.

TABLE I
CLASSIFICATION ACCURACY OF THE FOUR DESIGNED METHODS ON FLEVOLAND DATASET

Class	CNN	PSENet	LC-CNN	LC-PSENet
Stem beans	99.92%	99.86%	99.90%	99.90%
Peas	99.69%	99.85%	99.88%	99.91%
Forest	98.36%	98.48%	99.23%	99.46%
Lucerne	99.48%	99.41%	99.87%	99.95%
Wheat	98.06%	98.44%	99.33%	99.59%
Beet	99.47%	99.51%	99.65%	99.73%
Potatoes	99.02%	99.10%	99.48%	99.69%
Bare soil	100%	100%	100%	100%
Grass	97.95%	98.10%	99.24%	99.16%
Rapeseed	96.51%	96.63%	98.61%	98.96%
Barley	99.69%	99.72%	100%	100%
Wheat2	98.24%	98.56%	99.56%	99.67%
Wheat3	98.96%	98.89%	99.50%	99.72%
Water	99.94%	99.91%	100%	99.98%
Buildings	100%	100%	100%	100%
OA	98.79%	98.89%	99.51%	**99.66%**

Bold is the maximum value of the experimental results.

of other pixels at the boundary, we assign them the nearest nonboundary values in their horizontal or vertical direction.

For parameter setting and optimization, the weight parameters of the kernel in each convolution layer are initialized from Gaussian distributions with zero mean and a standard deviation of $\sqrt{2/n}$, where n denotes the number of inputs to each units. We use the Adadelta algorithm to optimize the network during the training. The batch size is set to 64, and we initialize the learning rate η to 1 with the decay factor of 0.95 per iterations. In addition, we reduce the learning rate periodically with a factor of 0.5 per 25 epochs. The network is totally trained for 80 epochs.

The subband decomposition of PolSAR images ran on MATLAB platform. The networks were trained and evaluated on the Keras deep learning framework with Tensorflow as the backend, and ran on a Linux server with Intel Xeon CPU (2.2 GHz) and NVIDIA Tesla K80 GPU (24 GB).

A. Experiment on Flevoland Dataset

The Flevoland dataset is an L-band fully polarized SAR image of a farmland area in The Netherlands, which was obtained by the NASA / Jet Propulsion Laboratory AIRSAR platform in 1989 [56]. The PolSAR data have an azimuth resolution of 12.1 m and a slant angle resolution of 6.6 m. The Pauli RGB image of this dataset is shown in Fig. 7(a) with a size of 1024×750 pixels, and the corresponding ground truth is shown in Fig. 7(b). The area contains 15 types of land cover classes, which are mainly crops. Fig. 7(c) is the legend of the ground truth.

For training sample ratio in this experiment, Zhang *et al.* conducted a sensitivity analysis on the ratio of different training samples [44]. They demonstrated that after the training data exceeds 10%, the test accuracy tends to be stable. And in most literatures using this dataset, the proportion of training samples is between 5% and 10%. For the convenience of comparison, we refer to the literature [50] to randomly select 200 samples of Buildings and 1000 samples per type of the other 14 types. These samples accounted for about 9% of the total number of samples, of which about 8% as the training set, 1% as the validation set.

1) Accuracy: Table I lists the classification accuracy of each designed method in each category on the Flevoland dataset, and the OA is in the last line. The best results are shown in bold. As is shown in the table, the OA of the CNN, PSENet, LC-CNN, and LC-PSENet are 98.79%, 98.89%, 99.51%, and 99.66%, respectively. Apparently, the proposed LC-PSENet achieves the highest accuracy among the four comparison methods, which is 0.87% higher than the baseline CNN under the same experiment conditions. In detail, by decomposing the polarimetric feature maps, and then, using the CNN to fuse the decomposed LC subbands, the LC-CNN improves the testing accuracy by 0.72% compared with directly using the CNN to fuse the polarimetric feature maps. This result shows that the construction of LC subbands image significantly improves the classification accuracy. Based on the LC-CNN, the SE module is used to model the LC subbands of the polarimetric feature maps, and the LC-PSENet algorithm is obtained, with a testing accuracy of 99.66%, which is further improved than the LC-CNN by 0.15%. These results demonstrate the effectiveness of LC subbands image construction and SE channel modeling. Besides, LC-CNN's testing OA is 0.62% higher than PSENet, and its testing accuracy on each category is higher than PSEnet, which shows that the LC subbands image construction contributes more to classification accuracy improvement than SE operations. In addition, when analyzing the performance of each method on different categories, we found that the accuracy of LC-PSENet on the worst performing category Rapeseed is as high as 98.96%, which is 2.33% higher than that of CNN and PSENet. And the accuracies on the other 13 categories are all over 99%, showing a more balanced performance on various categories. But CNN and PSENet performed poorly in some categories, such as Rapeseed and Grass. There may be three reasons for the excellent performance of LC-PSENet. First, the selected LC subbands of the polarimetric feature maps can guide the neural network to focus on the boundary regions and homogeneous regions of the PolSAR image, respectively, providing *a priori* knowledge for the neural network to mine more useful features. Second, the LC subband maps contain only weakened speckle noise, and

TABLE II
CLASSIFICATION ACCURACY COMPARISON WITH THE OTHER STATE-OF-THE-ART ALGORITHMS

Network (Training ratio)	W-CAE (5%) (Xie et al.)	W-DBN (+Clean) (5%) (Liu, Jiao et al.)	PCN (9%) (Liu et al.)	CV-CNN (10%) (Zhang et al.)	SSAE (10%) (Zhang et al.)	SF-CNN (10%) (Chen et al.)	LC-PSENet (9%) (proposed)
Stem beans	88.09%	96.60% (96.71%)	96.44%	98.80%	-	99.42%	**99.90%**
Peas	96.11%	96.80% (98.68%)	95.29%	98.70%	-	99.48%	**99.91%**
Forest	97.89%	91.41% (96.45%)	95.00%	96.80%	-	**99.58%**	99.46%
Lucerne	97.42%	96.11% (98.47%)	93.51%	98.10%	-	99.22%	**99.95%**
Wheat	89.71%	92.32% (97.67%)	95.20%	95.00%	-	99.26%	**99.59%**
Beet	98.30%	95.81% (98.14%)	97.59%	97.60%	-	98.98%	**99.73%**
Potatoes	95.62%	88.80% (98.08%)	95.89%	96.70%	-	99.22%	**99.69%**
Bare soil	97.42%	96.99% (97.34%)	96.45%	98.80%	-	99.75%	**100%**
Grass	88.62%	90.93% (95.39%)	94.33%	90.00%	-	96.87%	**99.16%**
Rapeseed	77.83%	86.71% (95.90%)	93.89%	92.00%	-	**99.17%**	98.96%
Barley	99.27%	98.06% (99.49%)	94.51%	94.50%	-	99.54%	**100%**
Wheat2	88.11%	88.17% (94.79%)	95.03%	94.20%	-	99.16%	**99.67%**
Wheat3	98.05%	93.77% (98.55%)	95.43%	96.60%	-	**99.83%**	99.72%
Water	93.46%	99.60% (99.90%)	99.50%	99.40%	-	99.88%	**99.98%**
Buildings	75.25%	85.26% (88.56%)	95.58%	83.2%	-	93.08%	**100%**
OA	93.31%	93.08% (97.57%)	96.94%	96.20% (97.70%)	95.02%	99.30%	**99.66%**

the low-frequency map and the contour map contain different levels of noise. After training of the network, the proportion of low-frequency features and contour features that are mined is not the same, so the noise is also suppressed following the ratio of different feature types, which may finally effectively reduce the isolated points. Third, the SE module models the feature channel, effectively weighing the contribution of different polarimetric features, as well as the contribution of low-frequency maps and contour maps.

Table II lists the classification accuracy of the proposed LC-PSENet and some other state-of-the-art methods including W-CAE, W-DBN, PCN, CV-CNN, SSAE, and SF-CNN, where values in parentheses are the results after postprocessing. The testing OA are 93.31%, 97.57%, 96.94%, 97.70%, 95.02%, 99.30%, and 99.66%, respectively. It shows that the proposed LC-PSENet achieves the highest OA among the comparison methods, and performs best on most categories, demonstrating its superiority. Although there are fewer training samples for W-CAE and W-DBN, but their accuracies is much lower than LC-PSENet, and W-DBN needs an additional clean algorithm to maintain the accuracy of 97.57%. In addition, we trained our model with 5% data and the test result is 99.24%, which shows the LC-PSENet still have advantages.

Fig. 8 shows the classification results of the first four designed comparison methods on the whole Flevoland map and the areas with the ground truth. Compared with Fig. 8(a-1)–(c-1), Fig. 8(d-1) is more in line with the ground truth. As is shown in the figure, Fig. 8(a-1) and (b-1) have many isolated points in many homogeneous regions, and these isolated points are misclassified points. In Fig. 8(c-1) and (d-1), such isolated points are greatly reduced, as shown in the area marked by the red rectangular frame. And the isolated points in Fig. 8(d-1) are

slightly less than in Fig. 8(c-1). This result illustrates the effectiveness of the proposed LC-PSENet again, and demonstrates the aforementioned argument that it can suppress the effect of speckle noise through control the proportion of feature mining in the two types of subbands. By comparing Fig. 8(a-2)–(d-2), we can also find that the LC-PSENet has fewer isolated points in most homogeneous regions. And for the boundary, such as the green area wrapped by the blue area in the upper right corner of the maps, CNN and PSENet have thicker and uneven red pixels on their boundary, while LC-PSENet is more uniform at the boundary, which improves the discrimination and fineness of boundaries between different areas. This result proves that LC-PSENet can guide the neural network to learn more useful features for classification.

2) Convergence and Time Consumption: Fig. 9 shows the convergence curve of the four comparison networks, where Fig. 9(a) is the training loss curve and Fig. 9(b) is the validation accuracy curve. We can see from the training loss curve that LC-PSENet converges first, followed by the LC-CNN, finally, PSENet and CNN. And in the validation accuracy curve, LC-PSENet and LC-CNN obtain the highest validation accuracy. These results show us that the proposed LC-PSENet not only improves the recognition accuracy, but also accelerates the convergence speed. The improvement of the convergence speed is due to the subband maps providing *a priori* information for the network input, guiding the network to learn the main spatial features of the polar feature maps more effectively, and avoiding the network to directly mine features from PolSAR images containing noise and fuzzy boundaries.

As to time consumption, to make the comparison fair, the four comparison networks designed in this article have similar structure and parameter design. Because the differences in DoF of
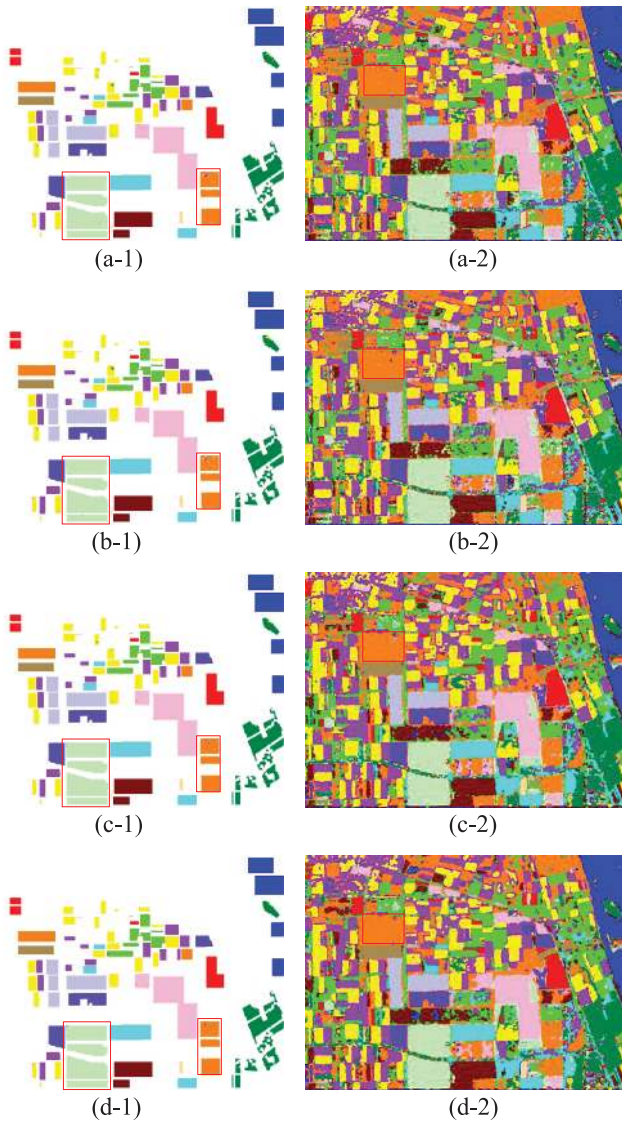
Fig. 8. Classification results on the Flevoland dataset. (a-1) Result overlaid with the ground truth map of CNN. (a-2) Result of whole map of CNN. (b-1) and (b-2) Results of the PSENet. (c-1) and (c-2) Results of the LC-CNN. (d-1) and (d-2) Results of the LC-PSENet.

Fig. 9. Convergence curve of the four comparison networks. (a) Training loss. (b) Validation accuracy.

the four comparison networks are very small, the four networks take 161.009, 161.011, 161.010, 161.012 s, respectively, for all 80 epochs in the training period of our experiment, with almost no difference. Also, the testing time of the comparison methods is basically the same. However, considering the convergence speed, LC-PSENet in Fig. 9(a) converges at least five epochs earlier than the CNN, which can save 10 s of the training time. Although LC-PSENet still needs to perform subband decomposition on the polarimetric feature maps before training, for a map of 750×1024 pixels, it only needs 1–2 s on the CPU to decompose the entire image using the NSLP. Therefore, LC-PSENet still has the advantage in time consuming. On the whole, LC-PSENet significantly improves the classification accuracy on the basis of the CNN, and it can save a certain amount of
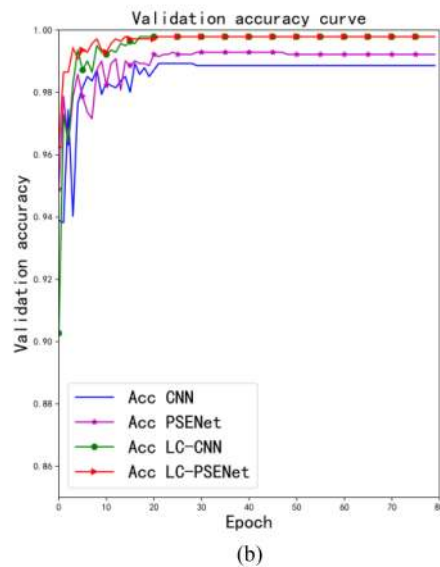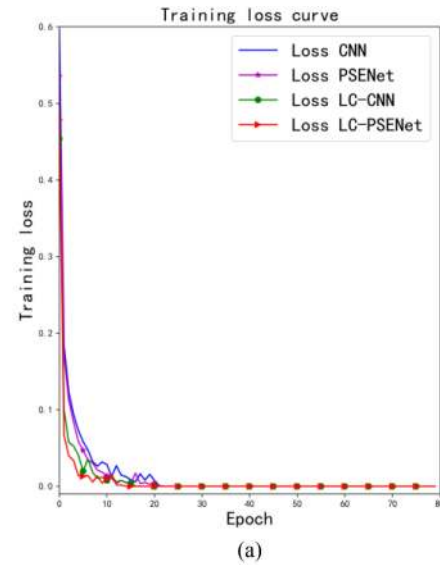
training time because of its faster convergence speed, further illustrating that it is an effective method.

3) Algorithm Complexity: In order to further verify the efficiency of the proposed method, the complexity of the proposed method will be compared with several methods. Because the parameters of the aforementioned partial comparison methods are not fully shown in the literature, this article only compares the proposed algorithm with the PCN, CV-CNN, and SF-CNN. The complexity comparison in this article includes two aspects: time complexity and space complexity. In convolutional networks, some operations such as pooling operations account for a very small proportion of the complexity, and the number of similar operations in the comparison network is also equivalent, so we only include convolutional layers and fully connected layers in our calculations. The time complexity of a single convolutional

TABLE III
ALGORITHM COMPLEXITY COMPARISON WITH OTHER METHODS

Method	PCN	CV-CNN	SF-CNN	LC-PSENet
Time complexity	2768128819200	45928944	56675520	**3375104**
Space complexity	7298780	7298780	56196720	**2228224**

TABLE IV
COMPARISON OF CLASSIFICATION ACCURACY OVER THE GROUND-TRUTH
AREA OF NETHERLANDS DATASET

class	CV-CNN (Zhang et al.)	CNN	PSENet	LC-CNN	LC-PSENet
Potato	99.8%	99.82%	99.78%	99.92%	99.81%
Fruit	98.3%	100%	100%	100%	100%
Oats	98.9%	99.45%	99.63%	99.82%	100%
Beet	96.1%	98.95%	98.88%	99.28%	99.33%
Barley	99.6%	98.56%	98.91%	99.24%	99.53%
Onions	93.2%	99.45%	99.61%	99.77%	100%
Wheat	99.9%	99.61%	99.67%	99.74%	99.77%
Beans	90.7%	100%	100%	100%	100%
Peas	99.0%	100%	100%	100%	100%
Maize	98.2%	99.77%	100%	100%	100%
Flax	97.0%	99.86%	99.97%	99.97%	99.97%
Rapeseed	99.8%	99.73%	99.69%	99.74%	99.76%
Grass	96.6%	99.88%	99.73%	99.97%	99.94%
Lucerne	99.2%	99.86%	99.95%	99.95%	99.95%
OA	99.0%	99.45%	99.51%	99.67%	**99.72%**

Bold is the maximum value of the experimental results.
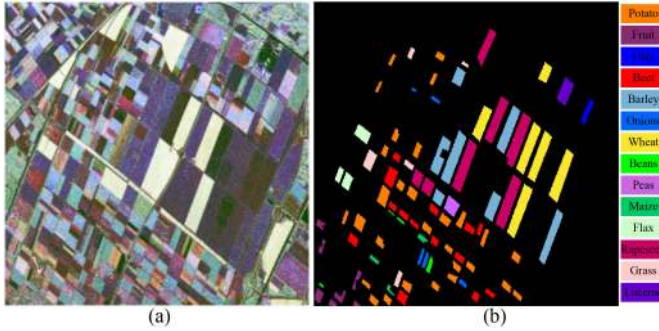

Fig. 10. Netherlands dataset. (a) Pauli RGB of the PolSAR image. (b) Ground-truth map and the legend.

layer is expressed as

$$T \sim O\left(M^2 \cdot K^2 \cdot C_{\text{in}} \cdot C_{\text{out}}\right) \tag{14}$$

where M represents the side length of the feature map output by the convolutional layer, K represents the side length of the convolution kernel of the layer, C_{in} is the number of input channels, and C_{out} is the number of output channels. The space complexity of a single convolutional layer is expressed as

$$S \sim O\left(K^2 \cdot C_{\text{in}} \cdot C_{\text{out}}\right). \tag{15}$$

For the fully connected layer, it can be regarded as a special convolutional layer whose convolution kernel size K is consistent with the size of the input matrix. The output feature maps of each convolutional layer is scalar points, that is, $M = 1$. Because NSLP takes very little time compared to neural network training, its complexity is not discussed mixed with neural networks.

Table III shows the time complexity and space complexity of the comparison algorithms. The figure shows that the time complexity and space complexity of the proposed algorithm are the smallest among the four networks. Among them, the time complexity of the PCN algorithm is extremely large, because the image size is enlarged by 16 times after the PCN encoding, and the network uses the entire image as the input so that the size of the feature maps of each layer is more than 10 000 times that of several other networks. The size of the input patch of the CV-CNN and SF-CNN is also a multiple of LC-PSENet, making LC-PSENet less complex.

B. Experiment on Netherlands Dataset

Another L-band fully polarized SAR data obtained by AIR-SAR in 1991 over Flevoland, the Netherlands was used to verify the effectiveness of the proposed method [56]. Fig. 10(a) shows the Pauli RGB image of this dataset with a size of 1020×1024, and Fig. 10(b) is the corresponding ground truth and its legend.

The dataset contains 14 types of land cover classes. Same as the literature [44], we randomly choose 850 samples per category as training and validation samples, where 10 800 samples are used for training and 1 100 samples for validation. The residual labeled data are used for testing. In this experiment, in addition to verifying its performance in classification accuracy, the performance of the method between the category boundaries is also discussed in detail.

Table IV lists the classification accuracy of each comparison method in each category on the Netherlands dataset, where the best OA is shown in bold. As is shown in Table IV, the testing OA of the CV-CNN, CNN, PSENet, LC-CNN, and LC-PSENet are 99.0%, 99.45%, 99.51%, 99.67%, and 99.72%, respectively. It shows that the proposed LC-PSENet achieves the highest classification accuracy among the comparison methods. In addition, in the classification results of LC-PSENet, the majority classes have a positive accuracy close to 100%, and all classes are higher than 99%. Since the parameter design of CNN, PSENet, LC-CNN, and LC-PSENet are similar, and they all use the same number of training samples as the CV-CNN in [44], the performance of LC-PSENet in the classification accuracy proves the superiority of the proposed method.

Fig. 11 shows the classification results of the designed four methods on the whole Flevoland map and the areas with the ground truth. At first glance, Fig. 11(a-1)–(d-1) are all in good agreement with the ground truth. But the classification accuracy of the LC-PSENet in some categories is significantly improved compared with other methods, such as the classification results of Barley. The red rectangular box in Fig. 11(a-1)–(d-1) marks the classification result of the four comparison algorithms in a same Barley area. As can be seen from the figure, the misclassified pixels in this area decrease sequentially from Fig. 11(a-1)–(d-1). The experimental results also prove the effectiveness of the LC-PSENet. In addition, as shown in the area marked by the black rectangular frame in Fig. 11(a-2)–(d-2), two phenomena occur in the classification result of the boundary. First, at the boundaries of different categories, Fig. 11(a-1) and (b-1) have significantly more heterogeneous pixels than Fig. 11(d-1), while
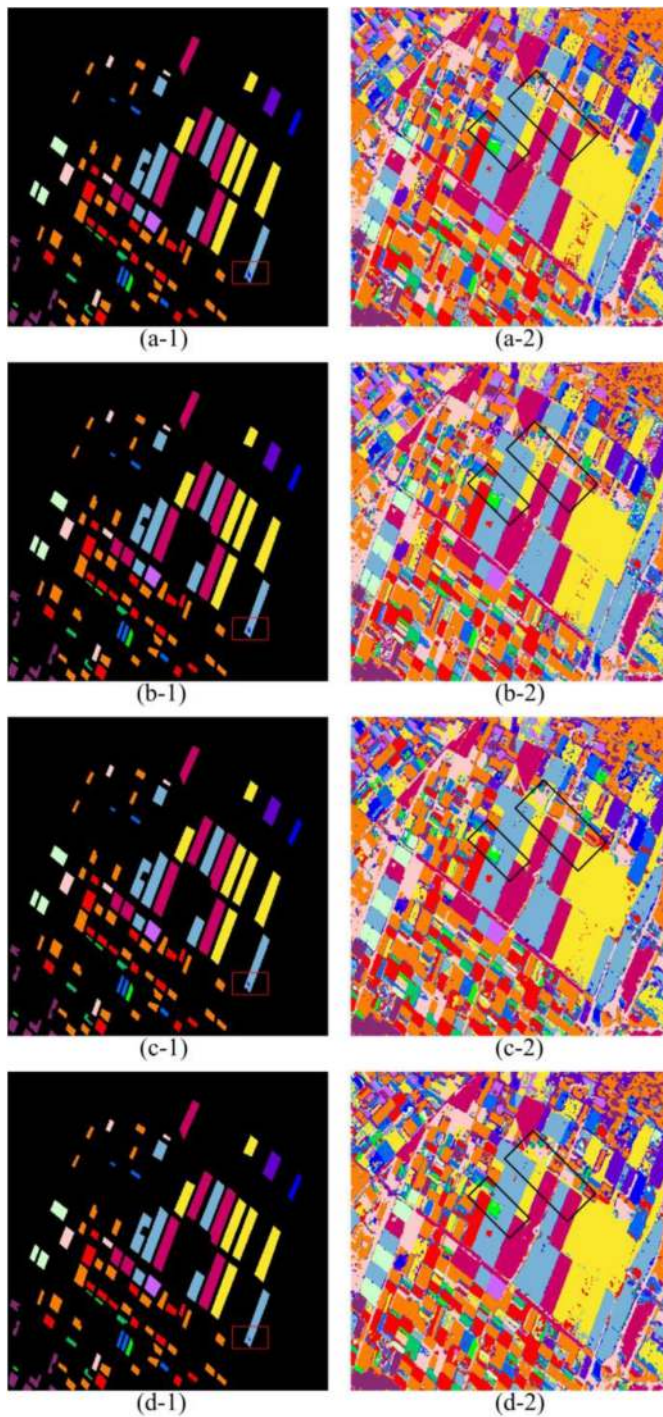
Fig. 11. Classification results on the Netherlands dataset. (a-1) Result overlaid with the ground-truth map of CNN. (a-2) Result of whole map of CNN. (b-1) and (b-2) Results of the PSENet. (c-1) and (c-2) Results of the LC-CNN. (d-1) and (d-2) Results of the LC-PSENet.

Fig. 11(d-1) appears as a fine line. Second, at the junction of two identical regions, Fig. 11(d-1) can be better classified as a homogeneous region without adding a heterogeneous boundary. These two phenomena show that the proposed algorithm is more delicate in boundary processing and reduces the probability of cutting a homogeneous region.
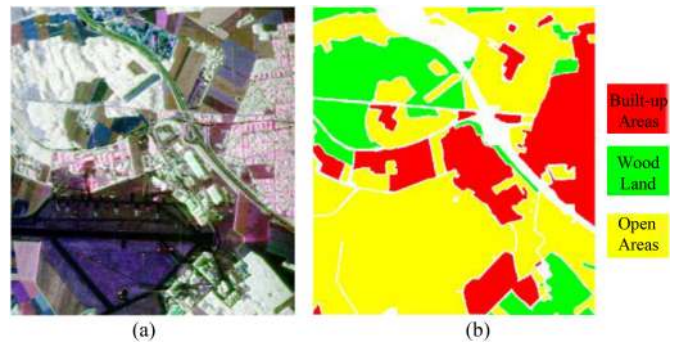

Fig. 12. Oberpfaffenhofen dataset. (a) Pauli RGB of the PolSAR image. (b) Ground-truth map and the legend.

TABLE V
COMPARISON OF CLASSIFICATION ACCURACY OVER THE GROUND-TRUTH
AREA OF OBERPFAFFENHOFEN DATASET

class	CV-CNN (Zhang et al.)	CNN	PSENet	LC-CNN	LC-PSENet
Built-up areas	91.3%	90.39%	90.73%	93.12%	93.76%
Wood land	92.2%	96.25%	96.41%	97.42%	97.78%
Open areas	94.6%	95.69%	95.76%	96.23%	96.20%
OA	93.4%	94.47%	94.62%	95.67%	**95.89%**

Bold is the maximum value of the experimental results.

C. Experiment on Oberpfaffenhofen Dataset

In order to further verify the universality of the method, an extra experiment was conducted on the ESAR L-band data over Oberpfaffenhofen, Germany [56]. Fig. 12(a) shows the Pauli RGB image of this dataset with a size of 1300×1200, and Fig. 12(b) is the corresponding ground truth with the legend. The dataset contains three types of land cover classes. Because of the large size and small number of classes, only 1% of the labeled samples are randomly sampled as the training dataset in this experiment. Thus, the training samples and the validation samples are 0.9% and 0.1%, respectively.

Table V lists the classification accuracy of each comparison method in each category on the Oberpfaffenhofen dataset. As is shown in Table V, the OA of the CV-CNN, CNN, PSENet, LC-CNN, and LC-PSENet are 93.4%, 94.47%, 94.62%, 95.67%, and 95.89%, respectively. So, we can judge that the proposed LC-PSENet performs best. And a detailed comparison of CNN and LC-CNN, and PSENet and LC-PSENet shows that the classification accuracy of the network using LC subbands as input is significantly higher than that of directly using polarimetric feature maps as the input, with increase more than 1%. In addition, the classification results of the LC-PSENet in built-up areas are the most improved compared with the CNN and PSENet. By observing the Pauli RGB image in Fig. 12, it is found that the texture of built-up areas of the Oberpfaffenhofen image is snowflake like, while the texture of wood land and open areas is more uniform. The substantial improvement in the classification results of built-up areas stems from the LC subbands' effective guidance of network learning.

Fig. 13 shows the classification results of the designed four methods on the areas with the ground truth. It can be seen that
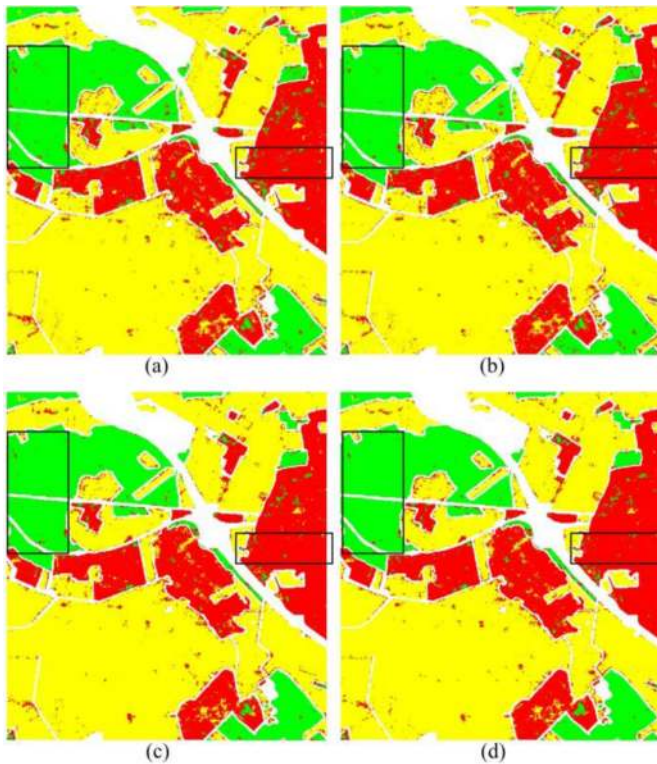
Fig. 13. Classification results on the Oberpfaffenhofen dataset. (a) Result overlaid with the ground truth map of CNN. (b) Results of the PSENet. (c) Results of the LC-CNN. (d) Results of the LC-PSENet.

this map is more complicated than the Flevoland, since there are similarities between the built-up areas and the wood land in some extent for Fig. 12(a). By comparing Fig. 13(a)–(d), we can also find that the LC-PSENet reduces the number of isolated pixels in many areas, as shown in the area marked by the black rectangular frame. The experimental results also prove the effectiveness of the LC-PSENet.

IV. CONCLUSION

In this article, a novel PolSAR image classification method named LC-PSENet is proposed. It aims to make full use of the polarimetric features and spatial features of PolSAR images to reduce the isolated points in the classification results and improve the classification accuracy. The proposed LC-PSENet first innovatively decomposes each polarimetric feature map into low-frequency subband and contour subband, and guides the network to perform feature mining and selection unevenly in the subbands in a supervised way, making the learning of the network more efficient and flexible. Then, by channel modeling in the CNN network, this method strengthens the network to learn the contribution of each polarimetric feature map and its LC subbands, which provides a more reasonable network for the PolSAR image classification with joint polarimetric features and spatial features. Experiments on three real PolSAR datasets have shown that this method significantly improves the classification accuracy of the CNN-based PolSAR image classification method, and effectively reduces the number of isolated points in the classification results, and makes the distinction between

boundaries more clear and delicate. At the same time, the classification accuracy of this method is better than many current state-of-the-art methods.

REFERENCES

[1] A. Lonnqvist, Y. Rauste, M. Molinier, and T. Hame, "Polarimetric SAR data in land cover mapping in boreal zone," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3652–3662, Oct. 2010.

[2] D. H. Hoekman and M. A. M. Vissers, "A new polarimetric classification approach evaluated for agricultural crops," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2881–2889, Dec. 2003.

[3] H. Wang, R. Magagi, and K. Goita, "Comparison of different polarimetric decompositions for soil moisture retrieval over vegetation covered agricultural area," *Remote Sens. Environ.*, vol. 199, pp. 120–136, Sep. 2017.

[4] K. Voormansik, T. Jagdhuber, K. Zalite, M. Noorma, and I. Hajnsek, "Observations of cutting practices in agricultural grasslands using polarimetric SAR," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 4, pp. 1382–1396, Apr. 2016.

[5] H. Skriver, "Crop classification by multitemporal C- and L-band single and dual-polarization and fully polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2138–2149, Jun. 2012.

[6] S. Uhlmann and S. Kiranyaz, "Integrating color features in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2197–2216, Apr. 2014.

[7] W. Xie *et al.*, "POLSAR image classification via Wishart-AE model or Wishart-CAE model," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 8, pp. 3604–3615, Aug. 2017.

[8] C. He, S. Li, Z. Liao, and M. Liao, "Texture classification of PolSAR data based on sparse coding of wavelet polarization textons," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 8, pp. 4576–4590, Aug. 2013.

[9] S. Chen, H. Wang, F. Xu, and Y. Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4806–4817, Aug. 2016.

[10] Z. Wang, L. Du, J. Mao, B. Liu, and D. Yang, "SAR target detection based on SSD with data augmentation and transfer learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 150–154, Jan. 2019.

[11] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 68–78, Jan. 1997.

[12] J. A. Kong, A. A. Swartz, H. A. Yueh, L. M. Novak, and R. T. Shin, "Identification of terrain cover using the optimum polarimetric classifier," *J. Electromagn. Waves Appl.*, vol. 2, no. 2, pp. 171–194, 1988.

[13] J. S. Lee, M. R. Grunes, and R. Kwok, "Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution," *Int. J. Remote Sens.*, vol. 15, no. 11, pp. 2299–2311, 1994.

[14] S. R. Cloude and E. Pottier, "A review of target decomposition theorems in radar polarimetry," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 498–518, Mar. 1996.

[15] Y. Yamaguchi, T. Moriyama, M. Ishido, and H. Yamada, "Four component scattering model for polarimetric SAR image decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 8, pp. 1699–1706, Aug. 2005.

[16] A. Freeman and S. L. Durden, "A three-component scattering model for polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 963–973, May 1998.

[17] E. Krogager, "Properties of the sphere, diplane, helix (target scattering matrix) decomposition," *Mol. Ecol.*, vol. 15, no. 11, pp. 3205–3217, 2006.

[18] Y. Zhou, H. Wang, F. Xu, and Y. Q. Jin, "Polarimetric SAR image classification using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1935–1939, Dec. 2016.

[19] T. Zou, W. Yang, D. Dai, and H. Sun, "Polarimetric SAR image classification using multifeatures combination and extremely randomized clustering forests," *EURASIP J. Adv. Signal Process.*, vol. 2010, Jan. 2010, Art. no. 4.

[20] B. Ren, B. Hou, J. Zhao, and L. Jiao, "Unsupervised classification of polarimetirc SAR image via improved manifold regularized low-rank representation with multiple features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 2, pp. 580–595, Feb. 2017.

[21] H. Aghababaee and M. R. Sahebi, "Incoherent target scattering decomposition of polarimetric SAR data based on vector model roll-invariant parameters," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4392–4401, Aug. 2016.

[22] L. Du and J. S. Lee, "Fuzzy classification of earth terrain covers using complex polarimetric SAR data," *Int. J. Remote Sens.*, vol. 17, no. 4, pp. 809–826, 1996.

[23] P. R. Kersten, J.-S. Lee, and T. L. Ainsworth, "Unsupervised classification of polarimetric synthetic aperture Radar images using fuzzy clustering and EM clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 519–527, Mar. 2005.

[24] C. Lardeux *et al.*, "Support vector machine for multifrequency SAR polarimetric data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 12, pp. 4143–4152, Dec. 2009.

[25] C.-S. Tao, S.-W. Chen, Y.-Z. Li, and S.-P. Xiao, "PolSAR land cover classification based on roll-invariant and selected hidden polarimetric features in the rotation domain," *Remote Sens.*, vol. 9, pp. 660–680, Jul. 2017.

[26] Y. Chen, L. Jiao, Y. Li, and J. Zhao, "Multilayer projective dictionary pair learning and sparse autoencoder for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 6683–6694, Dec. 2017.

[27] L. Zhang, L. Sun, B. Zou, and W. M. Moon, "Fully polarimetric SAR image classification via sparse representation and polarimetric features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 3923–3932, Aug. 2015.

[28] Y.-C. Tzeng and K.-S. Chen, "A fuzzy neural network to SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 1, pp. 301–307, Jan. 1998.

[29] C. Bentes, D. Velotto, and S. Lehner, "Target classification in oceanographic SAR images with deep neural networks: Architecture and initial results," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 3703–3706.

[30] J.-S. Lee, M. R. Grunes, T. L. Ainsworth, L.-J. Du, D. L. Schuler, and S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex Wishart classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2249–2258, Sep. 1999.

[31] L. Ferro-Famil, E. Pottier, and J.-S. Lee, "Unsupervised classification of multifrequency and fully polarimetric SAR images based on the H/A/Alpha Wishart classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2332–2342, Nov. 2001.

[32] B. Hou, H. D. Kou, and L. C. Jiao, "Classification of polarimetric SAR images using multilayer autoencoders and superpixels," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 3072–3081, Jul. 2016.

[33] Z. Wang, B. Du, Q. Shi, and W. Tu, "Domain adaptation with discriminative distribution and manifold embedding for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1155–1159, Jul. 2019.

[34] Y. Xu, L. Zhang, B. Du, and F. Zhang, "Spectral-spatial unified networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5893–5909, Oct. 2018.

[35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[36] W. Yang, X. Yin, and G.-S. Xia, "Learning high-level features for satellite image classification with limited labeled samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4472–4482, Aug. 2015.

[37] S. Yu, S. Jia, and C. Xu, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing*, vol. 219, pp. 88–98, Jan. 2017.

[38] L. Jiao and F. Liu, "Wishart deep stacking network for fast polsar image classification," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3273–3286, Jul. 2016.

[39] H. Petersson, D. Gustafsson, and D. Bergstrom, "Hyperspectral image analysis using deep learning—A review," in *Proc. 6th Int. Conf. Image Process. Theory Tools Appl. (IPTA)*, Dec. 2016, pp. 1–6.

[40] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 364–368, Mar. 2016.

[41] A. Romero, C. Gatta, and G. Camps-Valls, "Unsupervised deep feature extraction for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1349–1362, Mar. 2016.

[42] C. Yang, B. Hou, B. Ren, Y. Hu, and L. Jiao, "CNN-based polarimetric decomposition feature selection for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8796–8812, Nov. 2019.

[43] S. Chen and C. Tao, "PolSAR image classification using polarimetric-feature-driven deep convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 627–631, Apr. 2018.

[44] Z. Zhang, H. Wang, F. Xu, and Y. Q. Jin, "Complex-valued convolutional neural network and its application in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7177–7188, Dec. 2017.

[45] J. Masci, U. Meier, D. Cires, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Proc. 21st Int. Conf. Artif. Neural Netw.*, 2011, pp. 52–59.

[46] J. Geng, J. Fan, H. Wang, X. Ma, B. Li, and F. Chen, "High-resolution SAR image classification via deep convolutional autoencoders," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2351–2355, Nov. 2015.

[47] L. Zhang, W. Ma, and D. Zhang, "Stacked sparse autoencoder in PolSAR data classification using local spatial information," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1359–1363, Sep. 2016.

[48] Y. Guo, S. Wang, C. Gao, D. Shi, D. Zhang, and B. Hou, "Wishart RBM based DBN for polarimetric synthetic radar data classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2015, pp. 1841–1844.

[49] F. Liu, L. Jiao, B. Hou, and S. Yang, "POL-SAR image classification based on Wishart DBN and local spatial information," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3292–3308, Jun. 2016.

[50] X. Liu, L. Jiao, X. Tang, Q. Sun, and D. Zhang, "Polarimetric convolutional network for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 3040–3054, May 2019.

[51] A. L. Da Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3089–3101, Oct. 2006.

[52] Q. Sun and Y. Gao, "Unsupervised change detection in multitemporal SAR images via NSCT-domain feature clustering," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput.*, 2013, pp. 1–4.

[53] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," Sep. 2017. [Online]. Available: https://arxiv.org/abs/1709.01507

[54] G. Huang, X. Liu, J. Hui, Z. Wang, and Z. Zhang, "A novel group squeeze excitation sparsely connected convolutional networks for SAR target classification," *Int. J. Remote Sens.*, vol. 40, no. 11, pp. 4346–4360, 2019.

[55] L. Wang, X. Bai, and F. Zhou, "SAR ATR of ground vehicles based on ESENet," *Remote Sens.*, vol. 11, no. 11, pp. 1316–1330, Jun. 2019.

[56] Earth Online. [Online]. Available: http://envisat.esa.int/POLSARpro/datasets.html2, Accessed on: Nov. 2019.

Rui Qin was born in 1990. He received the master's degree from the Wuhan University of Technology, Wuhan, China, in 2015. He is currently working toward the Ph.D. degree with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China.

His research interests include deep learning, signal processing, radar target recognition, and synthetic aperture radar image classification.

Xiongjun Fu was born in 1978. He received the B.Eng. and Ph.D. degrees from the Beijing Institute of Technology (BIT), Beijing, China, in 2000 and 2005, respectively.

He is currently the Vice Dean of the School of Information and Electronics, BIT, and an Associate Professor, and a Ph.D. Supervisor with the BIT. His current research interests include radar system, radar signal processing, waveform design, and automatic target recognition.

Ping Lang received the B.E. degree in machine design engineering from the North University of China, Taiyuan, China, in 2015, and the M.E. degree in machine science and technology, in 2018 from the Beijing Institute of Technology, Beijing, China, where he is currently working toward the Ph.D. degree in signal and information processing with the School of Information and Electronics.

He was a Visiting Ph.D. Student with the School of Computing, University of Leeds, Leeds, U.K., from July 2019 to October 2019. His research interests include radar signal processing and machine learning, including explainable machine learning algorithm, radar target detection and recognition, and radar system design.