Polarimetric SAR Image Classification Based on Ensemble Dual-Branch CNN and Superpixel Algorithm

Wenqiang Hua, *Member, IEEE*, Cong Zhang, Wen Xie, and Xiaomin Jin

***Abstract*—Recently, convolutional neural networks (CNNs) have been successfully utilized in polarimetric synthetic aperture radar (PolSAR) image classification and obtained promising results. However, most CNN-based classification methods require a large number of labeled samples and it is difficult to obtain sufficient labeled samples. For this reason, an ensemble dual-branch CNN (EDb-CNN) is proposed for PolSAR image classification with small samples. First, to solve the problem of the small sample in PolSAR image classification, a new data enhancement method based on the superpixel algorithm is proposed to expand the number of labeled samples. Second, to obtain different scales of features from PolSAR images, a Db-CNN model is proposed. This model contains two parallel CNN structures. One CNN branch is used to extract the polarization features from the complex coherency matrix. The other branch is utilized to extract the spatial features based on weighted spatial neighborhood. On the top of these two branches, a feature fusion model is adopted to combine these two deep features, and a weighted loss function is employed to improve the learning procedure. Then, the ensemble learning algorithm is used for each CNN branch and Db-CNN network to obtain the better classification results. Finally, a postprocess algorithm based on the superpixel algorithm is proposed to improve the consistency of classification results. Experiments on two PolSAR datasets show that the proposed method achieves a much better performance than other classification methods, especially when only a few labeled samples are available.**

***Index Terms*—Dual-branch convolutional neural network (Db-CNN), polarimetric SAR, superpixels, terrain classification.**

I. Introduction

POLARIMETRIC synthetic aperture radar (PolSAR) is a powerful microwave imaging technology, which can provide the target information on Earth surface under all-weather and all-time conditions. Compared with other remote sensing images, PolSAR images are able to provide richer information as they transmit and receive electromagnetic waves in four polarization combinations (HH, HV, VH, and VV). Due to these characteristics, PolSAR images have been widely applied in remote sensing field, such as image classification [1], target recognition [2], and detection task [3], among which, PolSAR image classification has been extensively studied as the basis of understanding and interpretation of remote sensing [4]–[6].

Therefore, PolSAR image classification has attracted extensive research attention and more and more methods are proposed to accomplish this task. The majority of available methods are based on feature extraction and representation [7]–[9], and classifier designing and optimization [10]–[12]. In general, polarimetric target decomposition is one of most powerful and widely used methods for feature extraction of PolSAR images, such as Krogager decomposition [13], Pauli decomposition [14], H/a/A decomposition [15], Freeman three-component decomposition [16], and Huynen decomposition [17], Yamaguchi decomposition [18], and the extensions of the these mentioned methods [19], [20]. In addition, statistic features, backscattering elements, color features, texture features, and other popular features are used for PolSAR image classification. However, these extracted features are mostly class-specific, and it is difficult to achieve satisfactory classification results by using only one feature extraction method. But if multiple feature extraction methods are adopted, it will not only increase the time cost but also need effective feature selection methods to avoid the influence of redundant features.

After obtaining the features, the design of the classifier is also one of the important research content to solve the classification of PolSAR images. Traditional classifiers mainly include the complex Wishart classifier [21] and machine learning based methods, such as k-nearest neighbor [21], support vector machine (SVM) [23], sparse representation [24], and decision tree [25], etc. Many researchers have successfully used these methods for PolSAR image classification and have obtained the satisfactory classification results. However, in these methods, the classifier is designed according to the extracted features. Not only it takes a lot of time to extract features and design classifiers, but also the generalization performance is poor. Therefore, traditional methods cannot make full use of the relationship between features and classifiers to improve the classification accuracy of PolSAR images.

With the advancements in computing technology, the theory of deep learning has become increasingly important in

Manuscript received July 30, 2021; revised February 25, 2022; accepted March 24, 2022. Date of publication March 29, 2022; date of current version April 18, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61901368, in part by the Natural Science Foundation of Shaanxi Province under Grant 2019JQ-377, and in part by the Key Special project of China High Resolution Earth Observation System Young Scholar Innovation Fund under Grant GFZX04061502. *(Corresponding author: Wenqiang Hua.)*

The authors are with the Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China (e-mail: huawenqiang@xupt.edu.cn; zhangcong1402@163.com; xiewen@xupt.edu.cn; xmjin@xupt.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3162953

the field of pattern recognition. Among them, convolutional neural network (CNN) has demonstrated remarkable learning ability in many tasks, such as action recognition [26], [27], image classification [28], [29], semantic segmentation [30], [31], and scene labeling [32], [33]. Unlike the traditional machine learning methods, CNN is driven by data and can automatically extract features needed by the corresponding classifier through back-propagation. Moreover, CNN usually contains multiple convolution layers and activation layers, which can provide the deep nonlinear features. Therefore, to solve the problem of PolSAR image classification, some scholars introduce the CNN method to extract deep features of PolSAR images and classify them. In 2017, Zhou *et al.* [28] applied CNN based method to solve the problem of PolSAR image classification for the first time. In this method, they proposed a three-layer CNN architecture to classify PolSAR images and obtained the good classification results. After that, the classification methods based on deep learning are gradually applied to the classification of PolSAR images and achieved remarkable results, such as graph-based architecture [34], fully convolutional networks [35], complex-valued CNN (CV-CNN) [36], and other advanced convolutional networks [37], [38]. However, these success of CNN-based PolSAR image classification methods depend on adequate labeled samples. With the development of remote sensing technology, it is becoming easier to obtain the large number of unlabeled PolSAR data. But marked PolSAR images still takes a lot of manpower and material resources. Therefore, PolSAR image classification methods based on small samples has become one of the research hot topics in recent years.

As mentioned above, designing a suitable CNN architecture for PolSAR image classification with small samples is very necessary. In the emerging preliminary research, learning from the traditional machine learning method to solve the small sample problem, which provides a good research idea for the deep learning method to solve the small sample problem of PolSAR image classification. PolSAR image classification is a pixel-level classification task and aims to assign each pixel with a class label of given categories. Therefore, how to use a small number of labeled samples and the related information between pixels in PolSAR images to transfer the labeled information to unlabeled samples is an effective solution. Superpixels algorithm [39] can divide an image into many small regions of approximately the same size and shape. Combined with the spatial relationship between PolSAR pixels, this feature of superpixels can help the marker propagation of labeled samples in PolSAR images. Therefore, this article proposes a data enhancement (DE) strategy based on superpixel algorithm to expand the number of labeled samples.

In addition, pixel-based CNN classification [35]–[38] usually takes each pixel as the center and selects pixel blocks as the input information. Although this input information selected method also achieves good classification results, it does not make good use of the spatial relationship between pixels in the pixel block. In order to effectively use this spatial relationship, a spatial weighted pixel block selection method is proposed, which can better express the characteristics of the center pixel. At the same time, in order to obtain more abundant multiscale features, a dual-branch CNN (Db-CNN) model is proposed. This Db-CNN architecture contains two parallel deep CNN architectures. In these two parallel deep CNN architectures, one CNN uses directly selected pixel blocks as input, and the other uses spatial weighted pixel blocks as input. This inputting method not only increases the diversity of input features but also makes up for the lack of information caused by the single input mode. Then, on the top of these two branches, a feature fusion model is used to combine these extracted features and a weighted loss function is used to improve the learning procedure of the proposed Db-CNN model. Moreover, in the proposed Db-CNN model, there are three outputs: one output is the result of feature fusion of two parallel deep CNNs. The other two outputs are the output of each branch CNN architecture, and each branch has an output result. Then, in order to further improve the classification effect, the idea of ensemble learning is used to decide the final output results by majority voting. Finally, a postprocess algorithm based on superpixels is proposed to improve the regional consistency of classification results.

Therefore, the main contributions and advantages of this article are as follows:

1) Aiming at the small sample problem of PolSAR image classification, a new ensemble dual-branch CNN (EDb-CNN) model is proposed. It contains a new weighted loss function and an ensemble module, both of which can effectively improve the network classification performance.
2) The superpixel-based DE method is proposed to increase the number of training samples such that the classification accuracy can be improved with the limited samples.

The rest of this article is organized as follows: Section II describes the proposed method. Section III shows the experimental results and analysis with two real PolSAR datasets. Finally, Section IV concludes this article.

II. Proposed Method

A. Feature Representation

In PolSAR images, each resolution cell is expressed as the complex coherency matrix or complex covariance matrix, which is a symmetrical matrix and the diagonal elements are real numbers. The form of complex coherency matrix T is defined as

$$T = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix}. \tag{1}$$

To adapt the input format of the CNN, we utilize the coherency matrix T to convert into a real 6-dimensional (6-D) vector as shown below:

$$\begin{aligned} A &= 10\log_{10}(SPAN) \\ B &= T_{22}/SPAN \\ C &= T_{33}/SPAN \\ D &= |T_{12}|/\sqrt{T_{11}\cdot T_{22}} \\ E &= |T_{13}|/\sqrt{T_{11}\cdot T_{33}} \\ F &= |T_{23}|/\sqrt{T_{33}\cdot T_{22}} \end{aligned} \tag{2}$$

where A denotes the total scattering power, $SPAN = T_{11} + T_{22} + T_{33}$ denotes the sum of the principal diagonal elements of T. B and C denote the normalized ratio of the power of T_{22} andT_{33}, respectively. D, E, and F denote the relative correlation coefficients. Therefore, PolSAR data can be converted into a 6-D to form a $6 \times M \times N$ dataset, where M and N denote the number of rows and columns of PolSAR images, respectively.

B. Pixel Block Selection Based on Spatial Weighted

In the CNN-based PolSAR image classification methods, each pixel is used as the center to select the pixel block, then the pixel block is classified, and finally, the category of the pixel block is used as the category of the central pixel. This method is very suitable for uniform areas, because the pixel blocks contain pixels with the same category. But, in the nonuniform region, each pixel block may contain multiple categories, which seriously affect the classification of the whole pixel and center pixel. Aiming at this phenomenon, a pixel block selection based on spatial weight is proposed in this article.

This method weights the feature information of pixels in the pixel block by calculating the distance between the central pixel and the pixel block, which is expressed as follows:

$$W_{ij} = \frac{w_{ii}}{w_{ij}} \tag{3}$$

$$w_{ij} = \frac{1}{2} Tr((T_i)^{-1} T_j) + \ln |T_i| \tag{4}$$

where the range of W_{ij} value is (0, 1], w_{ij} denotes the distance between the center pixel and its neighborhood pixel, $Tr()$ is the trace of a matrix, T_i and T_j represent the coherency matrix of center pixel and its neighborhood pixel, respectively. The smaller the value of w_{ij} means that the pixel is more similar to the central pixel, the greater the possibility of belonging to the same category, the greatest contribution to the final judgment of the central pixel category, and the greater the weight W_{ij}. On the contrary, the greater the value of w_{ij} means that the greater the difference between the pixel and the central pixel, the less likely it is to belong to the same category, the smaller the contribution to the final judgment of the central pixel category, and the smaller the weight of its feature.

C. DE Method Based on Superpixels

Superpixels method is an image segmentation technique, which segments the image into some homogeneous pixel areas depending on pixels' distance and feature domain. Turbopixels segmentation algorithm [39] is one of the best superpixel methods, which can divide an image into many small regions with the similar size and shape. In the turbopixel segmentation method, pixels in any superpixel with high similarity are considered to belong to the same classes. For PolSAR images, pixels in the same region usually have high similarity, so they are more likely to belong to the same category. Therefore, combining the statistical characteristics and spatial information of PolSAR images, a DE method based on turbopixels algorithm is proposed to expand the number of labeled samples by using the information

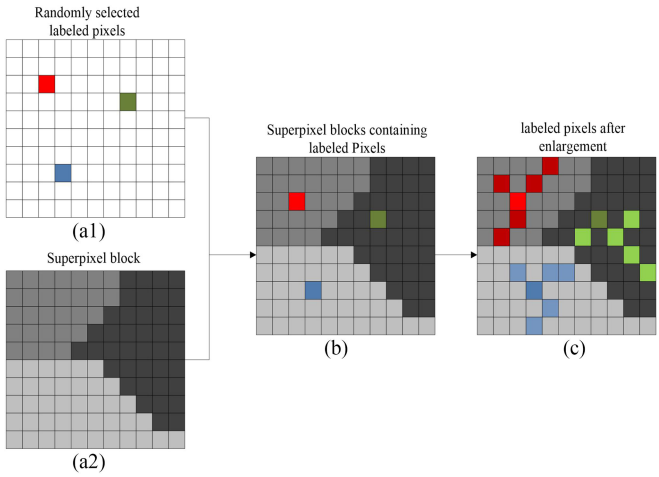

Fig. 1. Flowchart of the DE method based on superpixels.

of a small number of labeled samples and a large number of unlabeled samples. This method consists of four steps.

1) A small number of labeled samples are randomly selected as the initial training samples U_1.
2) Apply turbopixels algorithm divide the PolSAR images into N_t superpixels.
3) Select the superpixel block that contains the labeled pixels, and calculate the distance between the labeled pixels and other unlabeled pixels in each superpixel block in (4).
4) Select the m_L unlabeled samples closest to the labeled samples in superpixel block, and take them as the extended training samples.

Fig. 1 shows the flowchart of the DE method based on superpixels. In Fig. 1(a1), the red, green, and blue represent three categories of labeled pixels randomly selected. Fig. 1(a2) shows that the image is divided into three superpixel blocks. Fig. 1(b) represents the superpixel block containing labeled pixels. Fig. 1(c) shows that the 5 closest pixels of each category to the labeled pixels are selected and labeled in the surperpixel block of Fig. 1(b). As can be seen from Fig. 1(a1) and (c), the number of labeled pixels in each class has increased from 1 to 6, realizing the expansion of the label sample set.

D. Architecture of the Db-CNN Model

In recent years, deep neural networks have achieved great success in the field of image processing. Compared with the traditional machine learning methods, deep learning methods can learn the deep hidden information of the data actively. As one of the best deep learning networks, CNN has made remarkable achievements in various fields of image processing, such as object detection, image segmentation, terrain classification, etc.

The Db-CNN model is based on a basic deep CNN, and the architecture of Db-CNN model mainly consists of four parts: data augment processing, training process, testing process, and postprocessing, as shown in Fig. 2. In the process of DE, the number of labeled samples is expanded by using the superpixels method and the similarity between pixels. The training process consists of two parallel CNN structures, a feature fusion module,

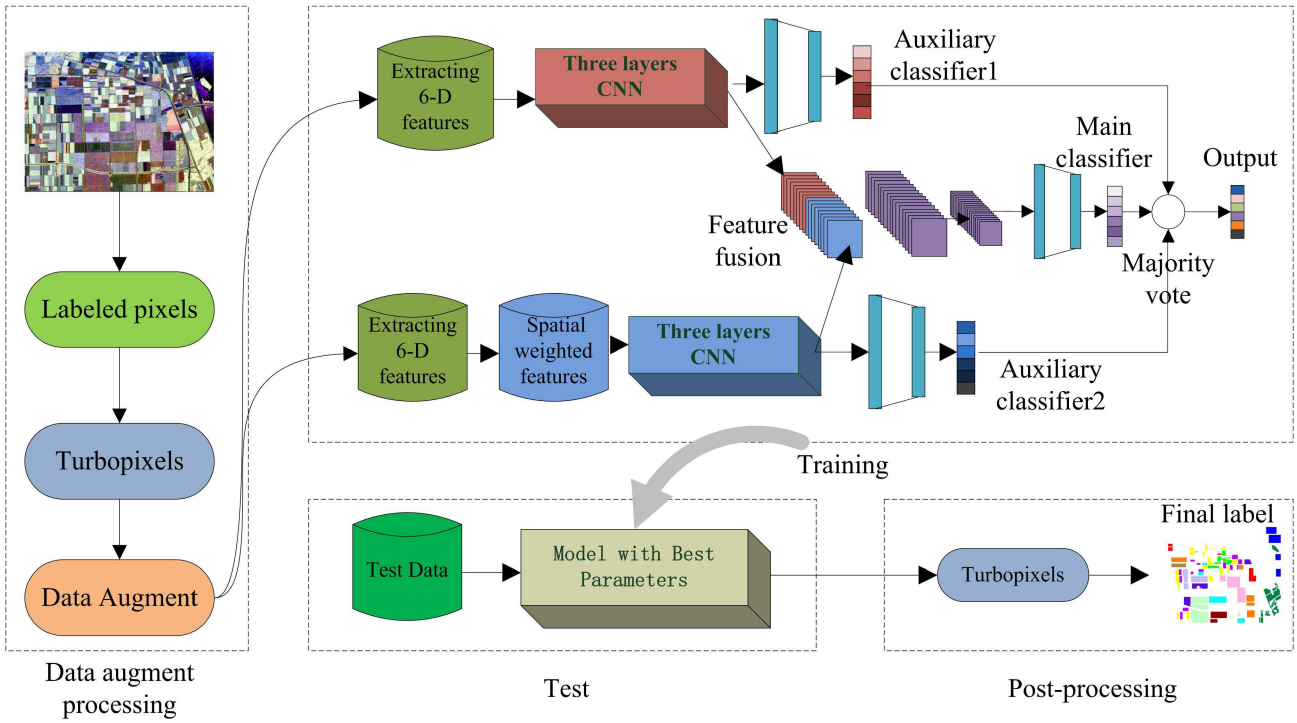

Fig. 2. Network architecture of the proposed method.

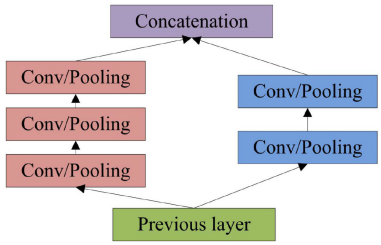

Fig. 3. Inception structure.

two auxiliary softmax classifiers, a main softmax classifier, and an ensemble learning module. In two parallel CNN structures, one CNN channel uses the extracted 6d-information as the input data, and the other channel uses the spatial weighted features as the input data. Then, an inception structure is used to fuse these two depth features. This inception structure is shown in Fig. 3. In this structure, different branches represent two convolution channels of the proposed method, and each channel uses different size convolution kernels to obtain different receptive fields. Then, the subsequent connections mean the fusion of different scale features. In the ensemble learning module, the majority voting method is used to vote the classification results of the two auxiliary classifiers and the main classifier. Finally, apply the trained model to predict the test data, and the postprocessing method based on superpixel is used to improve the prediction results.

The weighted cascade loss function (CLF) of the proposed method consists of three terms as shown in the following equation:

$$\text{Loss} = L_{\text{main}} + \lambda_1 L_{\text{auxiliary1}} + \lambda_2 L_{\text{auxiliary2}} \tag{5}$$

where L_{main} denotes the loss function after the fusion of two branch channel features, $L_{\text{auxiliary1}}$ and $L_{\text{auxiliary2}}$ denote represent the loss function of each branch CNN respectively. To emphasize the importance of these two branch networks, we use put weights for these two losses λ_1 and λ_2.

E. Procedure of the EDb-CNN Algorithm

The EDb-CNN Algorithm is presented in Algorithm 1.

F. Postprocessing Algorithm

Postprocessing algorithm is significant in pixels-based classification methods. It can be used to reduce the effect of speckle noises and achieve better performance. In this section, we introduce the postprocessing methods based on turbopixels algorithm. The specific process is as follows.

1) Computer the categories number $(R_i, i = 1, 2, \ldots k)$ of pixels in each superpixel block and the pixels number $(N_{R_i}, i = 1, 2, \ldots k)$ in each category based on the classification results of EDb-CNN and the segmentation result of turbopixels algorithm.
2) Select the main category R_w in each superpixel block, which represents more than half of the total number of pixel categories in each superpixel block.
3) For the pixels that do not belong to category R_w, calculate the wishart distance $d(T_{Ri}, V_{Rw})$ from the pixels to the center of category R_w. If this distance is less than the average wishart distance dc of the category R_w, the category of this pixel is corrected to category R_w. $d(T_{Ri}, V_{Rw})$ and

Algorithm 1: EDb-CNN.

Input: PolSAR image, randomly select the initial labeled set U_1

1. Use the data enhancement method to enlarge the labeled sample set. (Section II, Part C)
2. Construct labeled and unlabeled sample sets (U_{train} and U_{test}). The first branch only uses extracted 6-D features; The second branch first extracts 6-D features and then spatially weights them (Section II, Part A and B).
3. Train the EDb-CNN model using the labeled samples U_{train}. (Section II, Part C)
4. Use the trained model to predict the unlabeled samples U_{test}, and obtain the three predict results (class1, class2, and class3) of two auxiliary classifiers and the main classifier. (Section II, Part D)
5. The majority voting method is used to vote the classification result class1, class2, and class3, and the classification result of each pixel is output. (Section II, Part D)

Output: classification of PolSAR image

dc are defined as

$$d(T_{Ri}, V_{Rw}) = \frac{1}{2}\{\ln(|T_{Ri}|) + \ln(|V_{Rw}|) + Tr({T_{Ri}}^{-1}V_{Rw} + {V_{Rw}}^{-1}T_{Ri})\} \quad (6)$$

$$dc = \frac{1}{N_{Rw}}\sum_{i=1}^{N_{Rw}} d(T_{Ri}, V_{Rw}) \quad (7)$$

where T_{Ri} denotes the coherence matrix of pixels in the superpixel block that do not belong to category R_w. V_{Rw} represents the average of coherence matrix in category R_w. N_{Rw} represents the number of all pixels in category R_w.

III. Experimental Design and Analysis

A. Data Description

Two real PolSAR datasets are used to verify the proposed method: First, the Flevoland I farmland dataset was acquired by the NASA/JPL AIRSAR system in August 1989. Fig. 4(a) shows the Pauli RGB image of Flevoland I, which mainly contains 15 classes, and has 750×1024 pixel. Fig. 4(b) shows the ground truth map of Fig. 4(a) [40], [41]. The second dataset is Flevoland II area, which was acquired from the C-band by the RADARSAT-2 system in April 2008. Fig. 5(a) shows the Pauli RGB image of Flevoland II area. This data has 1400×1200 pixel, and mainly contain four different terrain classes, including urban, water, forest, and cropland. Fig. 5(b) shows the ground truth map of Fig. 5(a) [40], [41].

B. Experimental Design

In the proposed method, we designed an EDb-CNN method based on ensemble learning and superpixels algorithms. In the EDb-CNN method, there are three different outputs, including two auxiliary classifiers (classifier1 and classifier2) and one main classifier. In order to verify the effectiveness of this method, each part of this method is compared in the experiment.

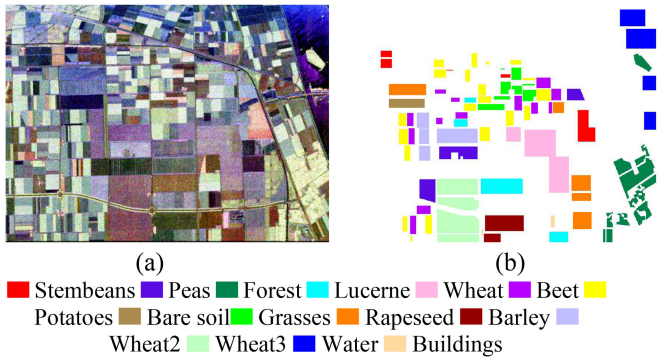

Fig. 4. AIRSAR dataset in Felvoland I. (a) Pauli image. (b) Ground truth image.

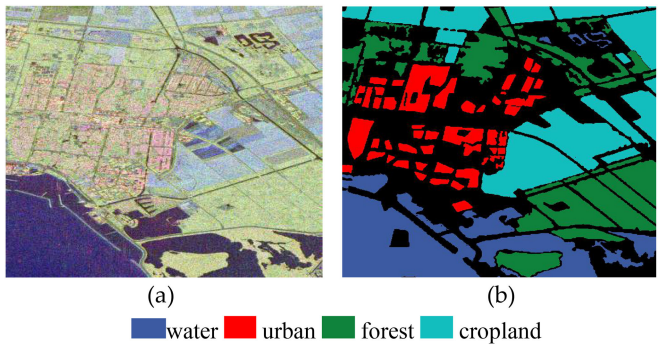

Fig. 5. RADARSAT-2 data in Felvoland II. (a) Pauli image. (b) Ground truth image.

In order to verify the effectiveness of the dual-branch, we compared the original Db-CNN with the traditional CNN method. In order to verify the effectiveness of DE based on superpixels, we compared the Db-CNN-DE with original Db-CNN. Db-CNN-DE denotes that the original Db-CNN method is combined with the DE processing. In order to verify the effectiveness of CLF, we combined the Db-CNN-DE with Db-CNN-DE-CLF. Db-CNN-DE-CLF denotes that strategy of CLF is used in Db-CNN-DE method. In order to verify the effect of ensemble learning, we compared the EDb-CNN-DE-CLF with auxiliary classifier1, auxiliary classifier2, and Db-CNN-DE-CLF method. EDb-CNN-DE-CLF denotes that the ensemble learning strategy is used to integrate the results of the two auxiliary classifiers and the main classifier of Db-CNN-DE-CLF method. Finally, in order to verify the effect of postprocessing, we compared the proposed method with EDb-CNN-DE-CLF method. The proposed method includes all the functions of the four modules (EDb-CNN, DE, CLF, and postprocessing).

In this article, all PolSAR datasets were tested with randomly selected training samples. Before the classification, the refined Lee filter [42] with a 7×7 window size is used to reduce the effects of speckle noise. For the data augment processing, each dataset was divided into 3000 superpixels based on turbopixels

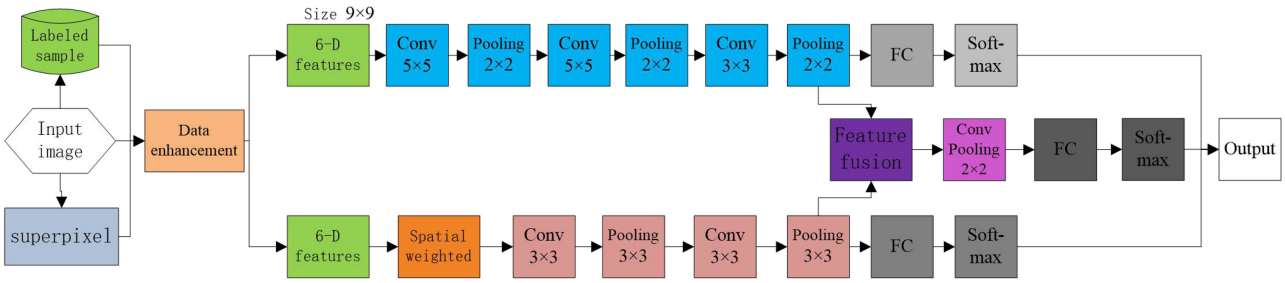

Fig. 6. Structure and parameter setting of the proposed method.

algorithm. The size of the input pixel block is set as 9×9, the learning rate is 0.001, ReLU is used as nonlinearity activation function, and other parameters of the proposed method are shown in Fig. 6. By comparing the performance of different classification methods, overall accuracy (OA) and kappa statistics of classification results are adopted to measurements of the classification performance. In order to compare the fairness of the experiment, all experiments are carried out with Tensorflow 2.0 on a desktop computer that runs Windows 10 operating system with an Intel(R) Core (TM) CPU processor (3.20 GHz) and 16 G memory.

C. Parameter Analysis

In the proposed method, there are three key parameters, weight parameters of the cascaded loss function, number of training samples, and size of pixel block. In order to select the appropriate parameters, the first dataset Flevoland I area is used as an example in this section.

1) *Effects of Weight Parameters of CLF:* In this section, we only consider the influence of two weights parameters of (5).

The loss function is a necessary component for deep learning models in network training. In this article, the proposed method is designed based on the dual-branch architecture, and the final results depend on two different features: polarimetric features and spatial weighted features. Because of the complexity of this structure, the traditional single loss function at the end of the network is insufficient to optimize the proposed model. Therefore, a weighted CLF is proposed to better guide the training procedure. In this article, λ_1 and λ_2 are used to represent the weight coefficient of cascaded loss function. In order to select the optimal weight coefficient, we did comparison experiments with different value of λ_1 and λ_2. Table I shows the OA and kappa statistics of the classification results of the proposed method, and 10 labeled samples of each class are used. As shown in Table I, different weight coefficients have a significant impact on the classification results. From Table I, it can be found that when $\lambda_1 = 0.7$ and $\lambda_2 = 0.7$, the classification accuracy is 93.41%, which is higher than other values. When $\lambda_1 = 0.6$ and $\lambda_2 = 0.3$, the classification accuracy is 85.95%, which is lower than other values. This shows that the selection of different weight parameters has a great influence on the classification results. It also shows that the cascaded loss function proposed in this article is an effective strategy, which can effectively increase the classification accuracy of the proposed method.

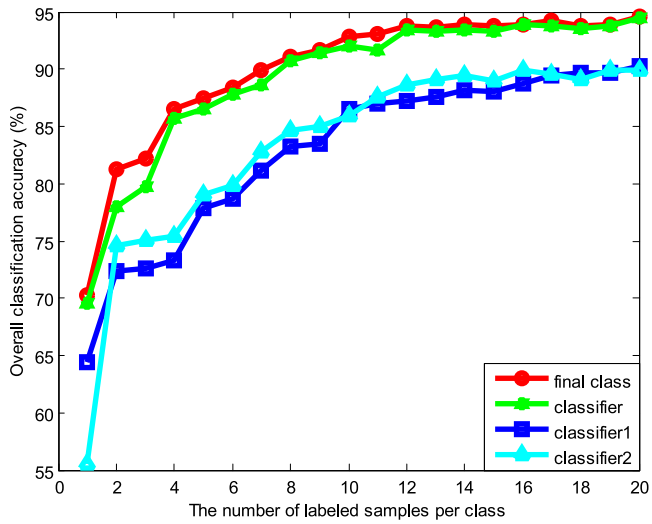

Fig. 7. Classification accuracy of Flevoland I datasets vary with the different number of labeled samples.

2) *Effects of The Number of Labeled Samples:* Different number of training samples usually have a great influence on supervised classification methods. It can reflect the effectiveness and stability of the classification method in a different number of training samples. Therefore, this section only considers the influence of different training samples on the proposed method, and the weights of the cascaded loss function are set to $\lambda_1 = 0.7$ and $\lambda_2 = 0.7$.

Fig. 7 shows the change curve of classification accuracy of the proposed method when the different number of training samples are selected for each class in Flevoland I area. There are four curves in Fig. 7. The dark blue and light blue curves represent the classification accuracy of the two auxiliary classifiers of the proposed method in different training samples. The green curve shows the main classifiers of the proposed method in different training samples. The red curve represents the final classification result of the proposed method without postprocessing.

From Fig. 7, it can be found that the OA of the proposed method is 70.23% when 1 labeled sample is selected as training samples for each class. Moreover, with the increase of the number training samples, the classification accuracy of the proposed

TABLE I
COMPARISON RESULTS OF THE PROPOSED METHOD WITH DIFFERENT WEIGHTS PARAMETERS OF CLF

Results	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
	0.9	0.9	0.8	0.8	0.7	0.7	0.6	0.6	0.5	0.5	0.4	0.4	0.3	0.3	0.2	0.2
OA	93.02		93.01		**93.41**		93.00		87.08		93.03		86.65		87.21	
Kappa	0.9239		0.9238		**0.9282**		0.9237		0.8590		0.9240		0.8543		0.8604	
Results	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
	0.1	0.1	0.9	0.8	0.9	0.7	0.9	0.6	0.9	0.5	0.9	0.4	0.9	0.3	0.9	0.2
OA	92.91		93.06		93.33		93.31		87.31		92.84		86.83		88.52	
Kappa	0.9153		0.9244		0.9270		0.9271		0.8615		0.9219		0.8564		0.8747	
Results	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
	0.9	0.1	0.8	0.7	0.8	0.6	0.8	0.5	0.8	0.4	0.8	0.3	0.8	0.2	0.8	0.1
OA	87.13		86.52		87.15		86.34		93.18		87.09		86.72		86.59	
Kappa	0.8596		0.8529		0.8597		0.8506		0.9257		0.8591		0.8549		0.8537	
Results	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
	0.7	0.6	0.7	0.5	0.7	0.4	0.7	0.3	0.7	0.2	0.7	0.1	0.6	0.5	0.6	0.4
OA	86.18		86.77		87.07		86.96		86.74		86.72		89.40		87.12	
Kappa	0.8487		0.8556		0.8589		0.8577		0.8553		0.8550		0.8516		0.8594	
Results	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
	0.6	0.5	0.6	0.4	0.6	0.3	0.6	0.2	0.6	0.1	0.5	0.4	0.5	0.3	0.5	0.2
OA	86.40		87.12		85.95		86.36		86.97		86.76		87.06		86.83	
Kappa	0.8516		0.8594		0.8468		0.8510		0.8578		0.8555		0.8588		0.8564	
Results	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
	0.5	0.1	0.4	0.3	0.4	0.2	0.4	0.1	0.3	0.2	0.3	0.1	0.2	0.1	0.1	0
OA	83.92		85.05		87.08		88.09		87.04		86.99		93.21		92.54	
Kappa	0.8252		0.8375		0.8590		0.8700		0.8586		0.8581		0.9259		0.9189	

The bold number indicates the best result in Table I.

method is increasing. When 10 labeled samples are selected as training samples for each category, the classification accuracy of the proposed method is 93.41%. This shows that the OA increases by 23.18% as the number of each labeled samples used increases from 1 to 10. This also shows that the proposed method is an effective method, with the increase of training samples, the classification accuracy also gradually increases. When each class selects 20 labeled samples as training samples, the OA of the proposed method is 94.62%. Thus, it can be seen that the training sample increases from 10 to 20, the classification accuracy of the proposed method only increases by 1.12%. This shows that the proposed method is a reasonable classification method with the increase of training samples, the increase of classification accuracy gradually decreases. From the red curve in Fig. 7, we can also find that when the training samples are very small, the change of classification accuracy increased significantly with the increase of training samples. When the training samples increase to 10 labeled samples each class, the change of classification accuracy rate gradually slows down. In addition, it can be seen from Fig. 7 that both the classification accuracy of the two auxiliary classifiers and the classification accuracy of the main classifier increase with the increase of training samples. This further shows that the method proposed in this article is an effective and reasonable method.

Therefore, to increase the efficiency of the proposed method and obtain satisfactory classification results, we suggest that 10 number of each labeled samples are used in this article.

3) *Effects of The Size of Pixel Block:* In the proposed method, the central pixel and its neighborhood pixels as a whole represent the feature information of the central pixel. Therefore, the size of pixel is an important parameter of the proposed method. In order to only analyze the size of pixel block, the weights of the cascaded loss function are set to $\lambda_1 = 0.7$ and $\lambda_2 = 0.7$, the number of labeled samples per class is set 10.

Fig. 8(a) and (b) shows the OA and time cost of proposed method with different size of pixel block on Flevoland I data set. From Fig. 8(a), it can be found that the OA first increases gradually with the increase in the size of the pixel block. This is mainly because when the pixel block is comparatively small, with the increase in the size of the pixel block, the neighborhood pixels will also increase, resulting in the corresponding increase of the spatial information, thus improving the final classification accuracy. However, when the pixel block is larger than 9×9, the neighborhood pixels are relatively far away from the central pixel, resulting in the redundant spatial information cannot effectively improve the classification results, and even reducing the classification accuracy with an increase in the size of the pixel block. From Fig. 8(a) and (b), it can be found that when

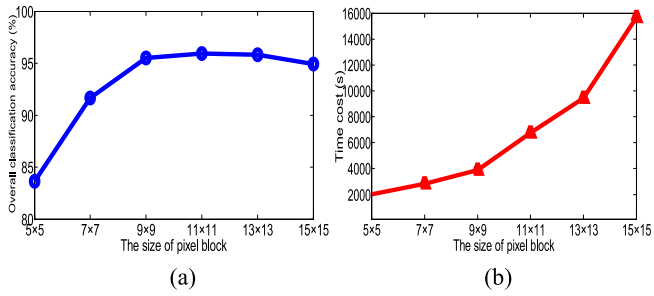

Fig. 8. Classification results of Flevoland I dataset vary with the different size of pixel block. (a) Classification accuracy. (b) Time cost.

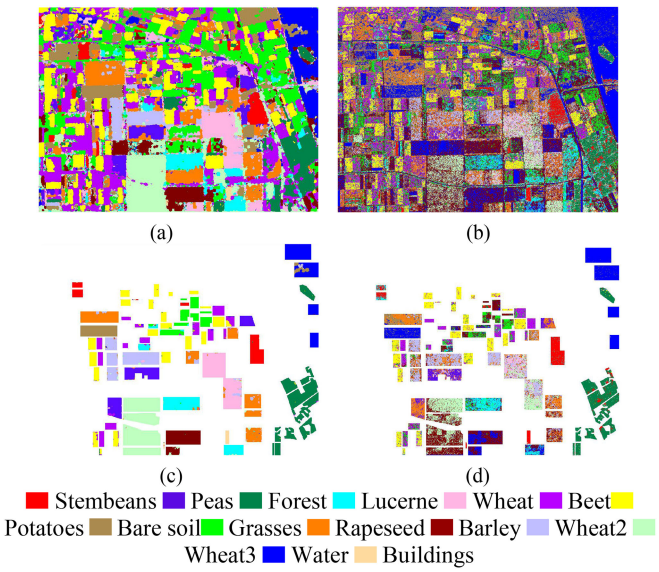

Fig. 9. Classification results of the Flevoland data set. (a) Proposed method without filter. (b) Proposed method with filter. (c) and (d) Masked results according to the ground truth of (a) and (b) respectively.

the size of pixel block is 9×9, the OA is 95.49% and the time consumed is 3890 s, while the size of pixel block is 11×11, the OA is 95.92% and the time consumed is 6742 s. The OA only increased by 0.43%, but the time spent increased by 2852 s. Therefore, considering the OA and time cost comprehensively, the size of pixel block is set to 9×9 throughout the following experiments.

D. Effect of Speckle Filtering

In order to investigate the effect of speckle noise suppression on the proposed method, the first Flevoland I data is used as an example to analyze the effect of the Lee filter on the classification accuracy, and each category selected 10 labeled samples as training samples. Fig. 9 and Table II show the classification results of the proposed method using and without using Lee filter method.

From Fig. 9 and Table II, it can be found that the OA of the proposed method is only 53.68% without using Lee filter, which is much lower than the classification accuracy of the proposed method using Lee filter. This result shows that the proposed method is very sensitive to noise. The main reason is that when the filtering method is not used, due to the influence of speckle noise, the unreliable samples are increasing by DE method in the training samples, so that this large number of unreliable samples are used for training network, resulting in the poor performance. Therefore, we filter the PolSAR data before network training.

TABLE II
CLASSIFICATION ACCURACY (%) OF THE FLEVOLAND DATA SET WITH AND WITHOUT FILTER

Method / region	With Filter	Without Filter
Stem beans	**98.69**	83.34
Rapeseed	**93.71**	43.19
Bare soil	**100.00**	9.92
Potatoes	**97.44**	85.88
Beet	**92.68**	68.19
Wheat 2	**95.76**	60.86
Peas	**97.00**	45.11
Wheat 3	**99.82**	48.22
Lucerne	**93.60**	52.75
Barley	**96.25**	54.40
Wheat	**81.52**	24.26
Grasses	**97.59**	39.46
Forest	**98.86**	89.27
Water	91.70	**99.06**
Building	**97.69**	1.36
OA	**95.49**	53.68
Kappa	**0.9508**	0.4984

The bold number indicates the best result in each row.

E. Experimental Results

In this section, the weights of the cascaded loss function are set to $\lambda_1 = 0.7$ and $\lambda_2 = 0.7$, and each category selected 10 labeled samples as training samples. The first Flevoland I data set contains 15 categories. So, the total number of training samples in this data set is 150. Similar, the second Flevoland II data set use 40 labeled samples (four categories) are used as training samples.

1) *Classification Results of the Flevoland I Acquired by Airsar in 1989:* In this part, Fig. 10 shows the classification maps of different classification methods, and Table III shows the classification accuracy of each category of crops by different methods. Fig. 10(a) shows the classification result of the traditional CNN method. Fig. 10(b) shows the classification result of the Db-CNN method. Fig. 10(c) shows the classification result of the Db-CNN-DE method. Fig. 10(d) shows the classification result of the Db-CNN-DE-CLF method. Fig. 10(e) and (f) shows the classification results of the auxiliary classifier1 and auxiliary classifier2, respectively. Fig. 10(g) shows the classification result of the EDb-CNN-DE-CLF method. Fig. 10(h) shows the classification of the proposed method.

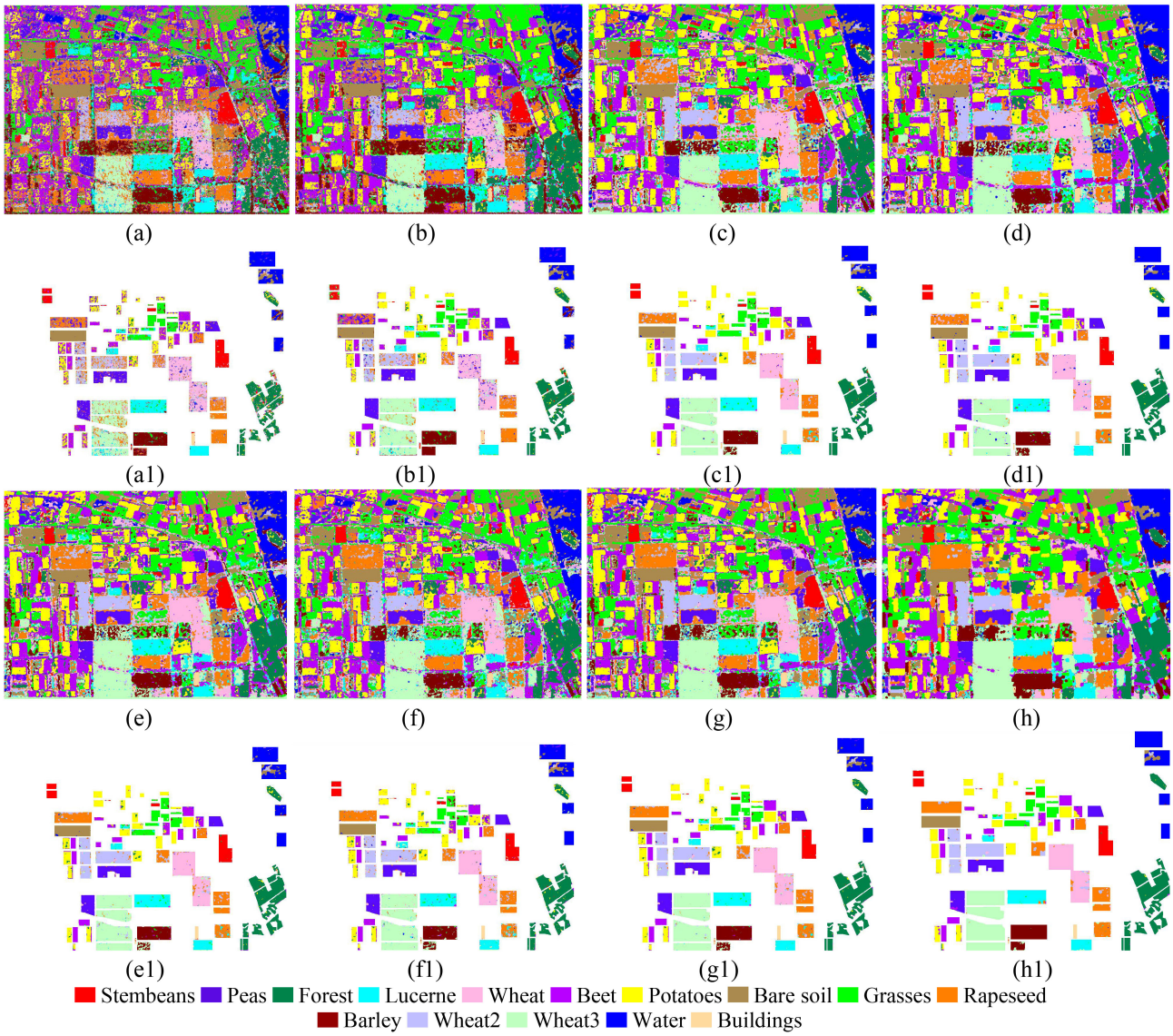

Fig. 10. Classification results of the Flevoland I data set acquired by Airsar. (a) CNN. (b) Db-CNN. (c) Db-CNN-DE. (d) Db-CNN-DE-CLF. (e) Classifier1. (f) Classifier2. (g) EDb-CNN-DE-CLF. (h) The proposed method. (a1), (b1), (c1), (d1), (e1), (f1), (g1) and (h1) are masked results according to the ground truth of (a)–(h), respectively.

From Table III, it can be found that the classification accuracy and kappa statistics of the proposed method are better than other comparison methods in the majority of categories. In the proposed method, the classification accuracy of 14 different crops is higher than 90%, and only one crops has a classification accuracy of 81.52%. For CNN method, only 2 crops have a classification accuracy greater than 90%, and there are 7 crops of classification accuracy below 80%. For Db-CNN method, there are 8 crops with classification accuracy higher than 90%. For Db-CNN-DE method, there are 9 crops with classification accuracy higher than 90%. For Db-CNN-DE-CLF method, there are 11 crops with classification accuracy higher than 90%. For EDb-CNN-DE-CLF method, there are 12 categories with classification accuracy higher than 90%. Moreover, the OA of the proposed method is 95.49%, which is significantly higher than other comparison methods.

In addition, compared the Db-CNN method and CNN method, it can be found that OA of Db-CNN method is 5.64% higher than CNN method. This indicates that the classification result of Db-CNN method is better than that of the CNN method. This is mainly because the Db-CNN method extracts the richer and more scale feature information by two CNN branch structure, and these richer and more scale features can effectively improve the classification results of PolSAR images. Compared the Db-CNN method and Db-CNN-DE method, it can be found that OA of the Db-CNN-DE method is 6.07% higher than the Db-CNN method. This indicates that the proposed DE is an effective method, which combines the superpixel method and the spatial information of PolSAR images, effectively uses a large number of unlabeled samples and expands the number of training samples. Compared to the Db-CNN-DE-CLF method and Db-CNN-DE method, it can be found that OA of

TABLE III
CLASSIFICATION ACCURACY (%) OF THE FLEVOLAND AREA ACQUIRED BY AIRSAR

Method / region	CNN	Db-CNN	Db-CNN -DE	Db-CNN-DE-CLF	Classifier1	Classifier2	EDb-CNN-DE-CLF	Proposed Method
Stem	93.45	91.64	96.58	97.57	97.22	95.31	98.04	**98.69**
Rapeseed	74.69	63.51	82.69	83.50	85.14	84.08	89.96	**93.71**
Bare soil	99.02	98.63	99.31	98.67	98.71	99.24	99.35	**100**
Potatoes	62.31	84.38	94.72	94.12	91.25	89.86	93.16	**97.44**
Beet	81.91	87.79	88.38	87.47	88.47	86.77	90.39	**92.68**
Wheat 2	58.67	70.76	89.72	91.50	91.27	88.72	92.92	**95.76**
Peas	87.28	95.83	96.28	96.68	97.18	96.71	**98.29**	97.00
Wheat 3	72.66	90.12	96.93	96.09	96.24	96.39	98.35	**99.82**
Lucerne	91.11	93.51	91.00	90.84	91.72	88.32	93.57	**93.60**
Barley	89.94	94.11	87.95	89.84	86.29	86.85	91.36	**96.25**
Wheat	68.91	69.16	78.39	80.06	78.24	78.24	79.78	**81.52**
Grasses	81.37	90.11	91.97	91.88	92.19	87.89	93.02	**97.59**
Forest	78.22	93.35	95.40	96.71	95.88	94.87	97.15	**98.86**
Water	78.77	79.29	88.45	93.39	87.72	86.26	89.22	**91.70**
Building	79.46	80.14	95.65	97.82	93.33	97.14	96.87	**97.69**
OA	79.85	85.49	91.56	92.41	91.39	90.44	93.43	**95.49**
Kappa	0.7815	0.8423	0.9081	0.9173	0.9062	0.8959	0.9284	**0.9508**

The bold number indicates the best result in each row.

TABLE IV
CLASSIFICATION ACCURACY (%) OF THE FLEVOLAND AREA ACQUIRED BY RADARSAT-2

Method / region	CNN	Db-CNN	Db-CNN -DE	Db-CNN-DE-CLF	Classifier1	Classifier2	EDb-CNN -DE-CLF	Proposed Method
Urban	83.67	83.72	91.29	90.54	93.45	87.38	91.97	**97.56**
Water	**99.37**	98.44	97.92	98.93	97.04	98.27	98.20	97.97
Forest	75.27	93.24	88.67	91.24	86.30	91.01	92.11	**95.79**
Cropland	55.30	67.80	86.44	86.43	83.20	83.91	86.12	**89.51**
OA	78.40	85.80	91.08	91.78	89.99	90.14	92.10	**95.21**
Kappa	0.7101	0.8074	0.8788	0.8882	0.8648	0.8656	0.8926	**0.9348**

The bold number indicates the best result in each row.

Db-CNN-DE-CLF method is 0.85% higher than Db-CNN-DE method. This indicates that the weighted CLF is better than non-CLF. Compared to the EDb-CNN-DE-CLF method, classifier1, classifier2, and Db-CNN-DE-CLF method, it can be found that OA of the EDb-CNN-DE-CLF method is 2.04%, 2.99%, and 1.02% higher than classifier1, classifier2, and Db-CNN-DE-CLF method, respectively. This indicates that combined with the proposed network structure, the introduction of ensemble learning strategy can improve the classification accuracy of the proposed method. Compared to the proposed method and the EDb-CNN-DE-CLF method, it can be found that OA of the proposed method is 2.06% higher than EDb-CNN-DE-CLF method. This indicates that the proposed postprocessing strategy is an effective method, which can improve the classification results. Through these comparisons, it is found that each part of the proposed method can effectively improve the classification results of PolSAR images. Similar, compared with the kappa coefficients of different methods, the same conclusion is obtained: the proposed method is much better than the traditional CNN and Db-CNN methods, and each module of the proposed method has a significant contribution to improving the classification accuracy.

Comparing the kappa statistics of these methods in Table IV, it can also be seen that the proposed method is 0.9508, which is obviously better than other comparison methods. Moreover, from the classification map, it can be seen that the regional consistency of the proposed method is obviously better than other comparison methods, especially in bare soil, rapesee, barely, and grasses area. Therefore, we can also conclude that the proposed method is an effective method, and better than other comparison methods.

2) *Classification Results of the Flevoland II Acquired by Radarsat-2 in 2008:* In this part, Fig. 11(a)–(h) shows that the classification maps of different methods for

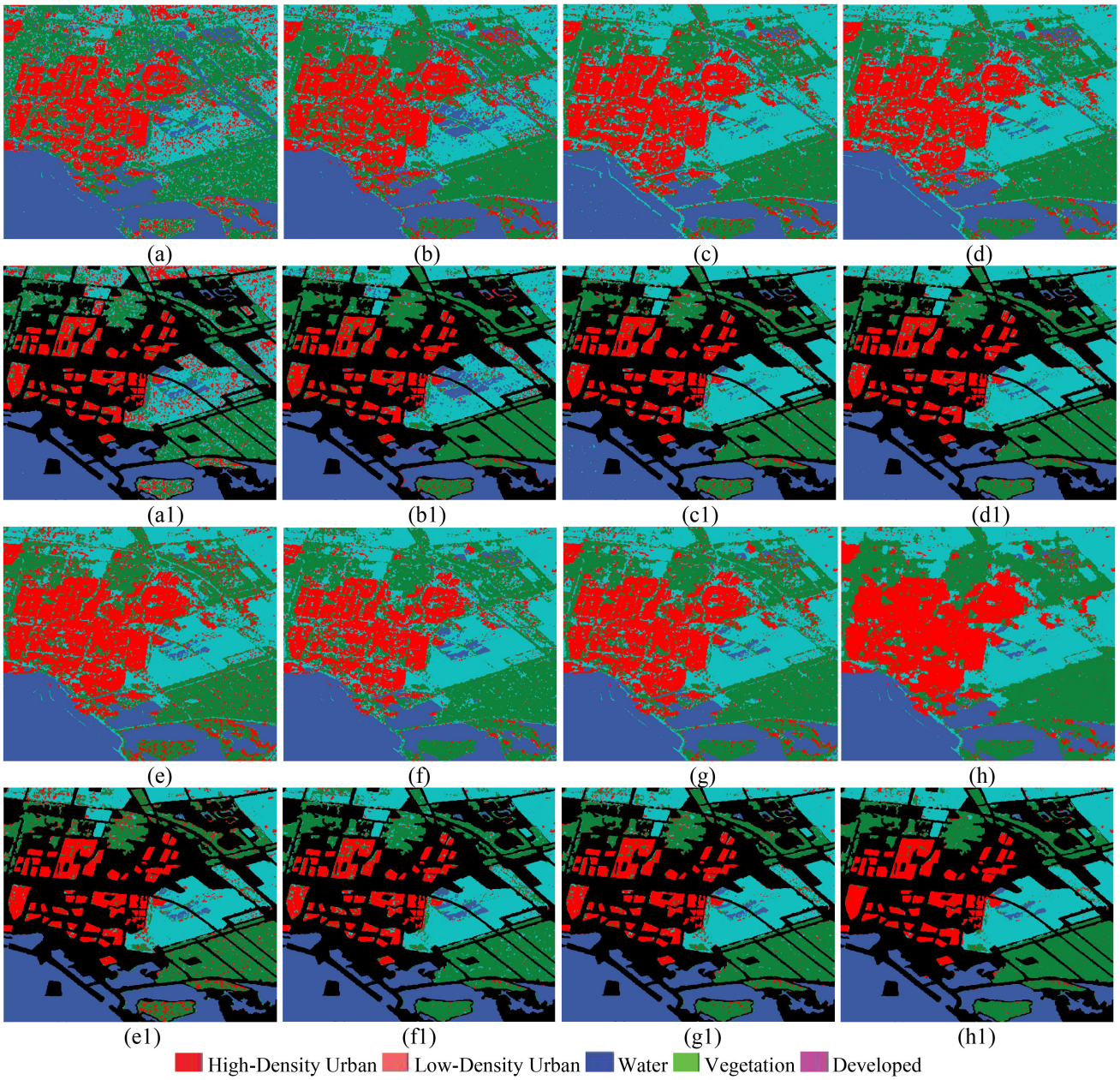

Fig. 11. Classification result of the Flevoland II data set acquired by radarsat-2. (a) CNN. (b) Db-CNN. (c) Db-CNN-DE. (d) Db-CNN-DE-CLF. (e) Classifier1. (f) Classifier2. (g) EDb-CNN-DE-CLF. (h) The proposed method. (a1), (b1), (c1), (d1), (e1), (f1), (g1) and (h1) are masked results according to the ground truth of (a), (b), (c), (d), (e), (f), (g) and (h) respectively.

Flevoland II area, and Table IV shows the classification accuracy of each category, OA, and kappa statistics by different methods.

From Table IV, it can be found that the OA of the proposed method is 95.21%, which is 16.81%, 9.41%, 4.13%, 3.43%, and 3.11% higher than CNN, Db-CNN, Db-CNN-DE, Db-CNN-DE-CLF, and EDb-CNN-DE-CLF methods, respectively. Compared to the Db-CNN-DE and Db-CNN method, the OA of Db-CNN-DE method is 5.28% higher than that of Db-CNN method after combining with the proposed DE. Compared to the proposed method and EDb-CNN-DE-CLF method, the OA of the proposed method is 3.11% higher than that of EDb-CNN-DE-CLF method after combined with postprocessing. This conclusion also proves the validity of the proposed DE method and the postprocessing method based on super-pixel, and further indicates that each part of the proposed method is effectiveness and can improve the classification accuracy for PolSAR images. In addition, From Fig. 11 and Table IV, it also can be found that the classification accuracy of cropland area and forest area are 55.30% and 75.27% by using the traditional CNN method. However, the classification accuracy of cropland area and forest area are 89.51% and 95.79% by using the proposed method, which is higher 34.21% and 20.52% than the CNN method, respectively. This is mainly because the training

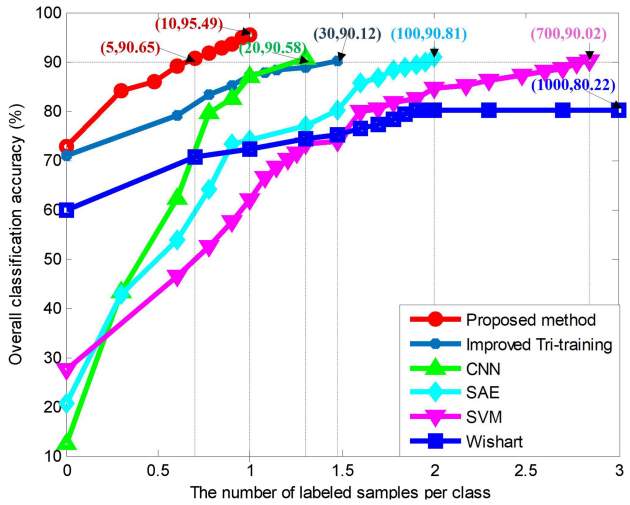

Fig. 12. Classification result of different classification methods varying with the number of labeled samples for the Flevoland I data set. [In (N_l, OA), N_l denotes the number of labeled samples per class, OA denotes classification accuracy.]

samples selected under the same conditions are very few, and the CNN network cannot show good classification ability under the lack of sufficient samples. This also shows that the proposed method is an effective method for PolSAR image classification with the small samples.

From Fig. 11, the classification map of the proposed method is better than other comparison methods, especially in urban and cropland area. Moreover, by comparing the kappa statistics, it also can be found that the kappa statistics of the proposed method is much higher than other comparison methods. This further shows the superiority of the proposed method.

3) *Comparisons With Existing Methods in terms of the Number of Labeled Samples:* In this part, we take the first Flevoland I dataset as an example to compare the classification accuracy of the proposed method with other traditional classification methods under different training samples. Fig. 12 shows the number of training samples (N_l) required for different classification methods when the classification accuracy is above 90%. Because the number of training samples required for different distribution varies greatly. Therefore, in order to better display the classification results, the abscissa of Fig. 12 is represented by logarithm. In the experiment, we compare the proposed method with the improved Tri-training [11], CNN [28], SAE [43], SVM [23] and supervised Wishart [21] methods. In Fig. 12, different color curves represent the classification accuracy of different methods in different training samples. The colored annotation (N_l, OA) represents the coordinates of the corresponding points. The first number N_l represents the number of training samples for each category, and the second number OA represents the classification accuracy of the classification method under the current number of samples.

From Fig. 12, it can find that the proposed method only needs 5 $(5 \approx 10^{0.6990})$ labeled samples for each category, and the classification accuracy can reach 90.65%, and the OA is 72.67% when there is only one training sample for each category. The OA of CNN, SAE, and SVM methods are less than 30%, the OA of the improved tri-training method and the supervised Wishart method is 71.04% and 59.78%, when there is only one training sample in each category. Moreover, it can be found that the improved tri-training method needs 30 $(30 \approx 10^{1.4771})$ labeled samples per category, CNN method needs 20 $(20 \approx 10^{1.3010})$ labeled samples per category, SAE method needs 100 $(100 = 10^2)$ labeled samples per category, SVM method needs 700 $(700 \approx 10^{2.8451})$ labeled samples per category when the OA reaches 90%, and OA of Wishart classifier is 80.22% when the number of training samples is 1000 $(1000 = 10^3)$. From this, it can find that the accuracy of the proposed method is obviously better than other comparison methods when training samples are few. In the case of very few samples (only 1 training samples per category), the proposed method can also obtain more than 70% classification accuracy. Therefore, we can conclude that the proposed method is better than other comparison methods, and can effectively improve the classification accuracy of PolSAR images with small labeled samples.

IV. Conclusion

In this article, we presented an effective deep network model for PolSAR image classification with small samples, which combines the CNN, superpixels algorithm, and ensemble learning. First, a DE method based on superpixels algorithm is proposed to enlarge the labeled samples. Second, to obtain more scale and deep polarization information, a Db-CNN model is proposed. Third, in order to increase the difference between two branch networks and improve the performance of the network, a spatial weighted method is proposed in one CNN branch to enhance the difference between two branches. This spatial weighted method can enhance the feature weight similar to the center pixel and weaken the feature weight different from the center pixel in the selected pixel block. Then, combined with the idea of ensemble learning, the results of main dual-network classifiers and two auxiliary classifiers are integrated. Finally, a postprocessing method based on super-pixel is proposed to reduce the speckle noise and improve the homogeneity of classification results.

Experimental results and analysis with two real PolSAR datasets show that the proposed method is an effective PolSAR image classification method, especially within few training samples. The experimental results demonstrate that the proposed method only needs five labeled samples to get more than 90% classification accuracy, while other comparison methods need dozens or hundreds of labeled samples to achieve the same classification accuracy. In addition, we analyzed the influence of the DE and postprocessing method based on super-pixel for the classification results. Figs. 10 and 11 indicate that the proposed DE method and the postprocessing method are two effective methods to improve the performance of classification results of PolSAR images. In our future work, we will focus on physically explainable deep learning for PolSAR classification with limited labeled samples.

REFERENCES

[1] G. N. Li *et al.*, "Marine oil slick detection using improved polarimetric parameters based on polarimetric synthetic aperture radar data," *Remote Sens.*, vol. 13, no. 9, 2021, Art. no. 1607.

[2] G. W. Liu, X. Zhang, and J. M. Meng, "A small ship target detection method based on polarimetric SAR," *Remote Sens.*, vol. 11, no. 24, 2019, Art. no. 2938.

[3] R. Qin, X. J. Fu, and P. Lang, "PolSAR image classification based on low-frequency and contour subbands-driven polarimetric SENet," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4760–4773, Aug. 2020.

[4] B. Ren, B. Hou, J. Chanussot, and L. Jiao, "Modified tensor distance-based multiview spectral embedding for PolSAR land cover classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 12, pp. 2095–2099, Dec. 2020.

[5] H. Y. Liu, T. W. Zhu, F. H. Shang, Y. Liu, D. Lv, and S. Yang, "Deep fuzzy graph convolutional networks for PolSAR imagery pixelwise classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 504–514, Dec. 2021.

[6] J. L. Wang *et al.*, "Representative learning via span-based mutual information for PolSAR image classification," *Remote Sens.*, vol. 13, no. 9, 2021, Art. no. 1609.

[7] Z. D. Wen, Q. Wu, Z. G. Liu, and Q. Pan, "Polar-spatial feature fusion learning with variational generative-discriminative network for PolSAR classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8914–8927, Nov. 2019.

[8] H. W. Dong, L. M. Zhang, D. Lu, and B. Zou, "Attention-based polarimetric feature selection convoulutional network for PolSAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Sep. 2022, Art. no. 4001705.

[9] Q. Yin, W. Hong, F. Zhang, and E. Pottier, "Optimal combination of polarimetric features for vegetation classification in PolSAR image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 10, pp. 3919–1031, Oct. 2019.

[10] W. Q. Hua, S. Wang, H. Y. Liu, K. Liu, Y. Guo, and L. Jiao, "Semi-supervised PolSAR image classification based on improved co-training," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 11, pp. 4971–4986, Nov. 2017.

[11] S. Wang *et al.*, "Semi-supervised PolSAR image classification based on improved tri-training with a minimum spanning tree," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8583–8597, Dec. 2020.

[12] R. Qin, X. J. Fu, and P. Lang, "PolSAR image classification based on low-frequency and contour subbands-driven polarimetric SENet," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4760–4773, Aug. 2020.

[13] E. Krogager, "New decomposition of the radar target scattering matrix," *Electron. Lett.*, vol. 26, no. 18, pp. 1525–1527, 1990.

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

[15] 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.

[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] J. R. Huynen, "Phenomenological theory of radar targets," Ph.D. dissertation, Drukkerij Bronder-Offset N. V., Rotterdam, The Netherlands, 1970.

[18] Y. Yamaguchi, M. I. Moriyama, 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.

[19] S. Zhang, S. Wang, B. Chen, and S. Mao, "Classification method for fully PolSAR data based on three novel parameters," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 39–43, Jan. 2014.

[20] S. Wang, K. Liu, J. J. Pei, M. Gong, and Y. Liu, "Unsupervised classification of polarimetric SAR images based on scattering power entropy and copolarized ratio," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 622–626, May 2014.

[21] 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.

[22] L. A. V. Isuhuaylas, Y. Hirata, L. V. Santos, and N. S. Torobeo, "Natural forest mapping in the Andes (Peru): A comparison of the performance of machine-learning algorithms," *Remote Sens.*, vol. 10, no. 5, 2018, Art. no. 782.

[23] L. Long and X. Liu, "SVM lithological classification of PolSAR image in Yushigou area, Qilian," *Sci. J. Earth Sci.*, vol. 3, no. 4, pp. 128–132, 2013.

[24] L. M. Zhang, X. Wang, B. Zou, and S. Zhu, "Independent target detection of PolSAR image joint polarimetric and spatial features based on adaptive convolution sparse," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 9, pp. 1533–1537, Sep. 2020.

[25] W. Q. Hua, S. Wang, Y. Zhao, B. Yue, and Y. Guo, "Semi-supervised PolSAR classification based on improved tri-training," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2017, pp. 2153–7003.

[26] X. H. Wang, L. L. Gao, J. K.Song, and H. Shen, "Beyond frame-level CNN: Saliency-aware 3-D CNN with LSTM for video action recognition," *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 510–514, Apr. 2017.

[27] H. Yang, C. F. Yuan, L. Zhang, Y. Sun, W. Hu, and S. J. Maybank, "STA-CNN: Convolutional spatial-temporal attention learning for action recognition," *IEEE Trans. Image Process.*, vol. 29, pp. 5783–5973, Apr. 2020.

[28] Y. Zhou, F. X. Wang, 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.

[29] W. Q. Hua, W. Xie, and X. M. Jin, "Three-channel convolutional neural network for polarimetric SAR images classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4895–4907, Aug. 2020.

[30] F. Ma *et al.*, "Attention graph convolution network for image segmentation in big SAR imagery data," *Remote Sens.*, vol. 11, no. 21, 2019, Art. no. 2586.

[31] S. C. Guo *et al.*, "Learnable gated convolutional neural network for semantic segmentation in remote-sensing images," *Remote Sens.*, vol. 11, no. 16, 2019, Art. no. 1922.

[32] S. Akodad *et al.*, "Ensemble learning approaches based on covariance pooling of CNN features for high resolution remote sensing scene classification," *Remote Sens.*, vol. 12, no. 20, 2020, Art. no. 3292.

[33] X. B. Han *et al.*, "Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification," *Remote Sens.*, vol. 9, no. 8, 2017, Art. no. 848.

[34] H. Y. Liu, D. R. Xu, and T. W. Zhu, "Graph convolutional networks by architecture search for PolSAR image classification," *Remote Sens.*, vol. 13, no. 7, 2021, Art. no. 1404.

[35] A. G. Mullissa, C. Persello, and A. Stein, "PolSARNet: A deep fully convolutional network for polarimetric SAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 5300–5309, Dec. 2019.

[36] X. X. Qin, H. X. Zou, W. S. Yu, and P. Wang, "Superpixel-oriented classification of PolSAR images using complex-valued convolutional neural network driven by hybrid data," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10094–10111, Dec. 2021.

[37] X. F. Tan, M. Li, P. Zhang, Y. Wu, and W. Song, "Complex-valued 3-D convolutional neural network for polSAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1022–1026, Jun. 2020.

[38] Y. W. Guo *et al.*, "Fuzzy superpixels based semi-supervised similarity-constrained CNN for PolSAR image classification," *Remote Sens.*, vol. 12, no. 10, 2020, Art. no. 1694.

[39] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "Turbopixels: Fast superpixels using geometric flows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 12, pp. 2290–2297, Dec. 2009.

[40] 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.

[41] M. C. G. Sanpedro, "Optical and radar remote sensing applied to agricultural areas in Europe," Ph.D. dissertation, Dept. Earth, Phys. Thermodyn., Univ. Toulouse, Valencia, Spain, 2008.

[42] J. S. Lee, M. R. Grunes, and G. Grandi, "Polarimetric SAR speckle filtering and its implication for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2363–2373, May 1999.

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

Wenqiang Hua (Member, IEEE) received the B.E degree from the University of Electronic Science and Technology of China, Chengdou, China, in 2012, and the Ph.D. degree in circuits and systems from the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education, Xidian University, Xi'an, China, in 2018.

He is currently a Lecturer with the School of Computer Science and Technology, Xi'an University of Post and Telecommunication. His research interests include machine learning, deep learning, and PolSAR image processing.

Cong Zhang received the B.S. degree from Xi'an Aeronautical University, Xi'an, China, in 2020. He is currently working toward the M.S. degree with School of Computer Science and Technology, Xi'an University of Post and Telecommunication, Xi'an, China.

His research interests include deep learning and PolSAR terrain classification.

Wen Xie (Member, IEEE) received the B.S. degree in automation from Xidian University, Xi'an, China, in 2011, and the Ph.D. degree in circuits and systems from the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education, Xidian University, in 2017.

She is currently a Lecturer with the School of Computer Science and Technology, Xi'an University of Post and Telecommunication. Her research interests include machine learning, deep learning, and image classification.

Xiaomin Jin received the Ph.D. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2018.

He is currently a Lecturer with the School of Computer Science and Technology, Xi'an University of Post and Telecommunications, Hangzhou, China. His research interests include machine learning and deep learning.