Semisupervised Complex Network With Spatial Statistics Fusion for PolSAR Image Classification

Yinyin Jiang, Ming Li, *Member, IEEE*, Peng Zhang, *Member, IEEE*, and Wanying Song

***Abstract*—Deep learning has achieved satisfactory results in polarimetric synthetic aperture radar (PolSAR) image classification, which requires a large number of labeled samples for training. However, in practice, labeling work is time-consuming and laborious. As a result, an insufficient number of labeled samples will lead to a limited ability of the network to recognize different terrains. To alleviate this problem, we take advantage of labeled and unlabeled samples simultaneously to train the deep learning model and, thus, propose a semisupervised complex network with spatial statistics fusion (SCN-SSF) for PolSAR image classification. First, the semisupervised complex network (SCN) continuously updates the pseudolabels of unlabeled samples during the training of complex-valued CNN, and their errors constitute the regularization term of the objective function, which improves the generalization of the network. As a result, SCN can recognize different terrains more accurately, and the classification has a higher belief. Then, a parameter-free spatial statistics module is built to model neighborhood label interactions based on the product of experts (POEs), thus enhancing contextual smoothness and correcting some misclassifications. Finally, based on the Dempster–Shafer theory, the contextual label information of POE and pixel-level information obtained by SCN are integrated to preserve image structure. Overall, with only a small number of labeled samples, SCN-SSF can accurately identify each terrain and obtain smooth classification while preserving edge information. The effectiveness of SCN-SSF is demonstrated by classifying PolSAR images with a small number of labeled samples.**

***Index Terms*—Semisupervised complex network (SCN), polarimetric synthetic aperture radar (PolSAR) image classification, contextual label information, product of experts (POEs), Dempster–Shafer theory (DST).**

I. Introduction

POLARIMETRIC synthetic aperture radar (PolSAR) transmits and receives electromagnetic signals with various polarimetric states [1], [2], [3]. As a result, PolSAR can provide much richer terrain information than single-polarization systems and has been developed rapidly recently. Terrain classification plays an important role in the analysis and interpretation of PolSAR data. A lot of methods have been proposed to mine effective information for PolSAR image classification, where statistical modeling and deep learning play important roles.

The statistics-based methods are classical techniques for PolSAR image classification. Lee et al. [4] model the covariance matrix of PolSAR data using the complex Wishart distribution, and the final classification is obtained by maximum likelihood (ML). Furthermore, the scalar product model is proposed for PolSAR data analysis. It models the backscattering signal by the product of a random scalar texture component and a complex Gaussian speckle variable [5] and has been widely applied in non-Gaussian modeling of PolSAR images, such as the Gamma distribution [6], $\mathcal{G}$ distribution [7], and Fisher distribution [8]. However, these pixel-based methods have poor performance against speckles. Considering the spatial relationships, Markov random field (MRF) and conditional random fields (CRF) methods are introduced to process PolSAR data [9], [10], [11], [12], [13], [14], [15]. In [14], Wishart MRF (WMRF) model combines the Wishart distribution with MRF to explore the spatial relation of neighboring pixels for initial oversegmentation. In [13], mixture Wishart-generalized Gamma-MRF (MWG Γ-MRF) utilizes Wishart-generalized Gamma distribution mixture model (WGΓMM) to maintain statistical correlations and introduce MRF to capture spatial–contextual information. In [15], CRF is combined with SVM to describe the spectral and spatial information in the max-margin frame. These MRF/CRF-based models have achieved satisfactory results. However, it is necessary to choose an appropriate statistical distribution for different scenes, which requires expert knowledge. Besides, inference is a relatively difficult task [16].

Recently, with the development of computer hardware, deep learning models have achieved great success in many fields, such as image processing [17], [18], [19], [20], speech recognition [21], [22], [23], natural language processing [24], [25], [26], etc. Deep learning models can automatically extract representative deep features through multiple cascaded convolution layers, which are also applied to PolSAR image processing. CNN is first applied to PolSAR image classification by Zhou et al. [27]. The complex coherence matrix of PolSAR is converted into a six-dimensional (6-D) real vector fed into the network. Complex-valued CNN (CVCNN) [28] extends real networks to the complex domain, including complex convolution and

Manuscript received 20 May 2023; revised 18 August 2023 and 26 September 2023; accepted 13 October 2023. Date of publication 17 October 2023; date of current version 27 October 2023. This work was supported in part by the Basic Research Program for National Defense under Grant KGJ202304, in part by the Natural Science Foundation of China under Grant 62172321, and in part by Civil Space Thirteen Five Years Pre-Research Project under Grant D040114. *(Corresponding author: Ming Li.)*

Yinyin Jiang is with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China, and also with the School of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an 710054, China (e-mail: jiang_yinyin@foxmail.com).

Ming Li and Peng Zhang are with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China (e-mail: liming@xidian.edu.cn; pzhang@xidian.edu.cn).

Wanying Song is with the School of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an 710054, China (e-mail: wanyingsong@hotmail.com).

Digital Object Identifier 10.1109/JSTARS.2023.3325365

complex pooling, etc. As a result, CVCNN can directly extract effective deep complex features for classification. The amplitude and phase information of complex data is fully utilized. Furthermore, the complex-valued 3-dimensional CNN (CV3DCNN) [29] performs 3-D complex-valued convolution to extract hierarchical complex features in spatial and scattering dimensions. These models consider the characteristics of PolSAR data and achieve satisfactory results. All the above deep models rely on sufficient labeled samples to ensure good classification performance while annotations are time-consuming and require expert knowledge. In practice, labeled samples are often insufficient, which can lead to overfitting of the network. As a result, the network has a limited ability to identify different terrains.

To alleviate this problem, semisupervised networks have been studied and applied to PolSAR image classification [10], [30], [31], [32], [33]. Bi et al. [10] propose a graph-based semisupervised model. The model extracts representative polarimetric features through a deep network and efficiently propagates class labels through a graph model. Finally, these two models are integrated into a principled framework. In [30], the K-means algorithm is used to generate pseudolabels for unlabeled samples. Unlabeled samples with pseudolabels and labeled samples are fed into 3DCNN to obtain discriminative deep features. Guo et al. [32] propose a memory CNN (MCNN), which introduces a memory module and designs an additional loss for unlabeled samples. These semisupervised networks effectively introduce unlabeled samples into training, thus expanding the training set and alleviating overfitting.

In addition, neighboring pixels within PolSAR images are highly correlated. To explore spatial relations, Zhang et al. [34] and Hu et al. [35] utilize the spatial distance of the neighborhood to measure the weights and feed the weighted inputs into a stacked sparse autoencoder (SSAE) for further feature extraction and classification. These two models exploit the relationship between neighboring pixels and are robust against speckles. However, SSAE extracts features by reconstructing the inputs while discriminative features useful for the classification task are ignored. In [36], minigraph convolutional networks (miniGCNs) are applied to model middle- or long-range spatial relationships of samples and are fused with spatial–spectral features extracted by CNN to achieve better classification results. In [37], multimodal deep learning-remote sensing (MDL-RS) combines spatial–spectral joint classification guided by CNN and pixel-level labeling provided by fully connected in a unified MDL framework. These models design different network structures to capture spatial information and integrate it with pixel-level information, whose superiority has been demonstrated experimentally. Inspired by the above models, we propose a semisupervised complex network with spatial statistics fusion (SCN-SSF) model for PolSAR image classification. The model has the following contributions.

1) The semisupervised complex network (SCN) part of SCN-SSF is an off-the-shelf CVCNN whose training involves unlabeled samples. Concretely, SCN can generate and update pseudolabels for unlabeled samples during training, and their errors constitute the regularization term of the objective function, which improves the generalization of the network. In this way, SCN is able to recognize different terrains more accurately and improve the belief of classification.
2) To capture spatial information, a parameter-free spatial statistics module is designed to compute neighborhood label dependencies based on the product of experts (POEs) to enforce label consistency. As a result, SCN-SSF can obtain smooth homogeneous regions and correct some misclassifications.
3) To preserve the image structure, Dempster–Shafer theory (DST) is applied to integrate pixel-level information provided by SCN and contextual label information of POE. In this way, with a small number of labeled samples, the SCN-SSF model can preserve the edge structure and achieve smoother classification within the consistent class region.

The rest of the article is organized as follows. Section II briefly introduces the PolSAR data. Section III describes SCN-SSF model in detail. Section IV shows the experimental results and discussion. Finally, Section V concludes this article.

II. Preliminary

Each resolution cell of PolSAR data can be represented as a complex scattering matrix $\mathbf{S}$ [1]

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

where S_{HV} denotes the complex scattering coefficient of vertical transmitting and horizontal receiving polarization, and the other coefficients have similar definitions. In the reciprocity theorem, S_{HV} is assumed to be equal to S_{VH}. Under the Pauli basis, the scattering matrix $\mathbf{S}$ can be expressed by the scattering vector $\mathbf{k}$

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

To reduce the effect of speckle, multilook processing is applied to polarimetric data [33]. An n-look coherence matrix can be given by

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

where the superscript T in (2) and H in (3) indicate transpose and conjugate transpose operations, respectively. T is a Hermitian matrix, and $T_{ij} = \bar{T}_{ji}$. The elements of the upper triangular, i.e., $\{T_{11}, T_{12}, T_{13}, T_{22}, T_{23}, T_{33}\}$, are used to represent the input data. Finally, a sliding window is used to crop the PolSAR image into a set of patches with the size being $m \times m \times 6$, where m is the window size and 6 denotes the number of channels.

III. Methodology

Deep networks trained by insufficient number of labeled samples suffer from overfitting and have a poor ability to identify different terrains. To alleviate this problem, we propose the SCN-SSF model, as shown in Fig. 1. First, SCN is a special

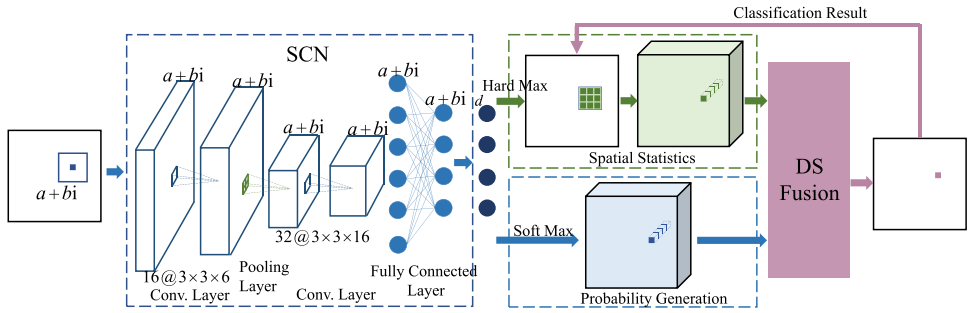

Fig. 1. Backbone of the SCN-SSF model. The model has three main components. The first SCN module learns the network parameters in a semisupervised way. The second part has two branches, including spatial statistics and probability generation. The third part fuses the information from these two branches by DST.

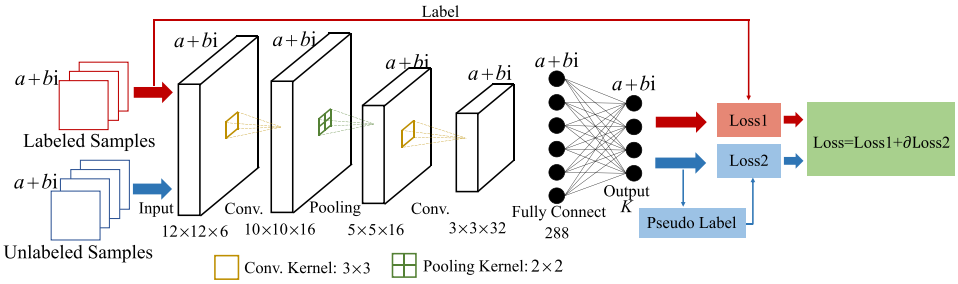

Fig. 2. Training process of SCN.

CVCNN whose training involves unlabeled samples. Specifically, SCN continuously generates pseudolabels for unlabeled samples during training, and their corresponding errors play the role of regularization in the objective function. Thus, the generalization of the network is improved, and the network can recognize different terrains more accurately.

Then, the output of SCN is fed to the spatial statistics and probability generation branches to explore the pixel-level information and contextual label information. Finally, these two different kinds of information are combined for classification. Consequently, the SCN-SSF model can achieve smoother homogeneous regions and preserve edge information.

A. Training of SCN

SCN is a CVCNN jointly trained by labeled and unlabeled samples. The main structure and training process of SCN are shown in Fig. 2, where the input patches are a set of complex patches extracted from the PolSAR images. SCN consists of two complex convolutional layers, a complex pooling layer, and a complex fully connected layer [28]. The complex network can integrate amplitude and phase information directly in the complex domain, thus effectively extracting more complete deep complex features.

In this part, the training process using labeled and unlabeled samples is described in detail. The labeled dataset is represented as $\{X_L, T_L\}$, where X_L denotes the labeled N_1 samples, and T_L is its corresponding complex labels consisting of $0+0\cdot \mathrm{i}$ and $1+\mathrm{i}$. The unlabeled dataset is denoted as X_U with N_2 unlabeled samples. Then, the training set $X=\{X_L, X_U\}$ is fed into SCN, and the corresponding output set is $O_T=\{O_L, O_U\}$. Thus, the loss function of the labeled data is given by

$$E_L=\frac{1}{N_1}\sum_{i=1}^{N_1}\sum_{k=1}^{K}\left[\left(\Re\left(T_i^k\right)-\Re\left(O_i^k\right)\right)^2+\left(\Im\left(T_i^k\right)-\Im\left(O_i^k\right)\right)^2\right] \quad (4)$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts, respectively, and K represents the number of classes. Furthermore, based on the maximum modulus of the output O_U, the pseudolabels $\tilde{T}_U$ of unlabeled samples are generated. Thus, the loss function of unlabeled data can be given by

$$E_U=\frac{1}{N_2}\sum_{i=1}^{N_2}\sum_{k=1}^{K}\left[\left(\Re\left(\tilde{T}_i^k\right)-\Re\left(O_i^k\right)\right)^2+\left(\Im\left(\tilde{T}_i^k\right)-\Im\left(O_i^k\right)\right)^2\right]. \quad (5)$$

The pseudolabels of unlabeled samples are iteratively updated during training. Finally, the loss function of the SCN can be expressed as

$$E=E_L+\alpha E_U \quad (6)$$

where α is a coefficient varying with time [38]. In this way, the error of unlabeled samples is the regularization term of the final objective function, which can prevent overfitting and improve the generalization performance of SCN. Thus, SCN can recognize different terrains more accurately.

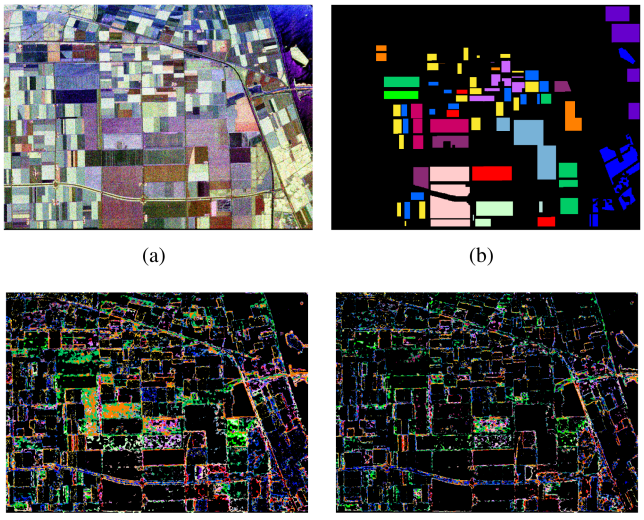

Fig. 3. Uncertainty of CVCNN and SCN on the Flevoland1 dataset. The black areas in (c) and (d) indicate that the probabilities of the classified class are greater than 0.1 while the colorful areas are the opposite. (a) Pauli image. (b) Ground truth. (c) Uncertainty of CVCNN. The colorful uncertain regions account for 29.74% of the whole image. (d) Uncertain of SCN. The colorful uncertain regions account for 18.33% of the whole image.

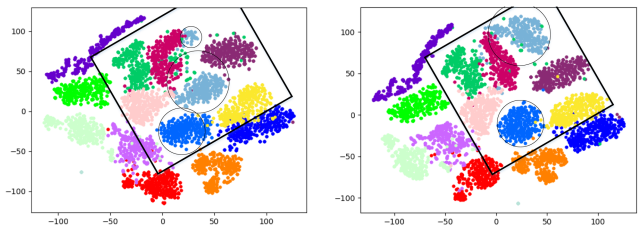

Fig. 4. T-SNE visualization of the features learned by different methods on the Flevoland1 dataset. (a) CVCNN. (b) SCN. The corresponding boxes in (a) and (b) have the same size.

With the introduction of unlabeled samples, classified labels have a higher class probability. An experiment is conducted here to visualize this. As shown in Fig. 3, it is obvious that the colorful uncertainty regions in Fig. 3(d) are smaller than those in Fig. 3(c). This indicates that the introduction of unlabeled data reduces the uncertainty of the classification. The classification has higher beliefs, which brings two benefits. On the one hand, CVCNN with unlabeled data can learn more effective and discriminative deep features. As shown in Fig. 4, the features of SCN have larger interclass distances (see box) and smaller intraclass distances (see circle), thus improving the classification accuracy. On the other hand, the higher beliefs facilitate the preservation of structural information in the subsequent SSF.

B. Pixel-Level and Contextual Label Information Capturing

Moreover, the spatial statistics and probability generation branches are designed to explore pixel-level and contextual label information from the output of SCN. The probability generation

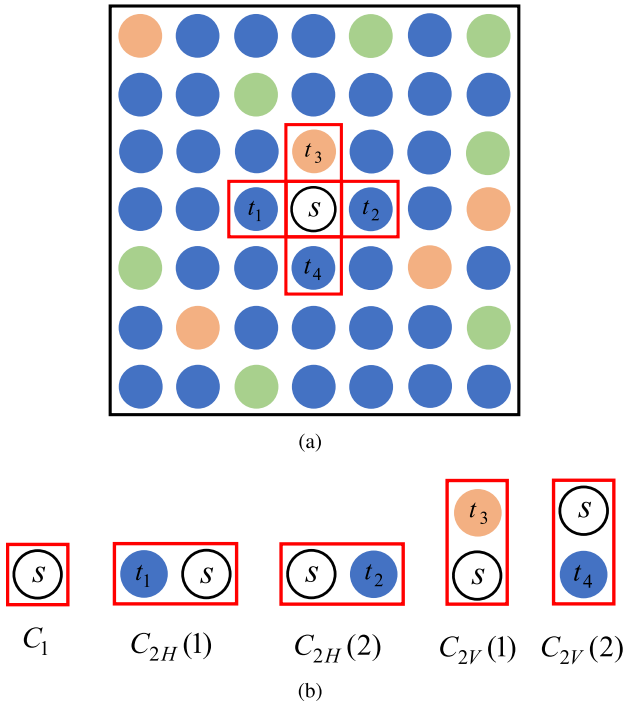

Fig. 5. Neighborhood system of pixel s. (a) High-order neighborhood window H_s is centered on pixel s, where the red box is its low-order neighborhood. (b) Five cliques decomposed from the red box in (a) are the single clique C_1, the two horizontal pairwise cliques C_{2H}, and the two vertical pairwise cliques C_{2V}. The blue, green, and yellow represent three different categories, and the center pixel has no color, indicating that it can take any category and calculate the corresponding probability.

branch assigns a probability value to each SCN output through the softmax function, which is expressed as

$$p(k) = \text{softmax}(z(k)) = \frac{e^{z(k)}}{\sum_{k\in\Omega} e^{z(k)}} \tag{7}$$

where $p(k)$ is the probability that pixel belongs to class k, and $z(k)$ is the modulus of $O(k)$, where $\Omega = \{1, 2, ..., K\}$ denotes a set of K labels. In this branch, pixel-level information is obtained, preserving structural information and preventing oversmoothing in subsequent SSF.

The spatial statistics branch first obtains the labels $Y = \{Y_s \in \Omega\}_{s\in S}$ based on the output of SCN, where $S = \{s\}_{s=1,...,N}$ represents a set of N pixels. Then, to correct some misclassifications, contextual label information is utilized. Specifically, a high-order neighborhood window H_s centered on pixel s is shown in Fig. 5(a). Here, the red box is the low-order neighborhood of s, which consists of a single clique C_1, two horizontal pairwise cliques C_{2H}, and two vertical pairwise cliques C_{2V}, as shown in Fig. 5(b). Then, the label of the center pixel is set to the label k, and k takes the labels in Ω. And the other labels remain unchanged. At this time, the label frequency of each clique is counted, including single cliques $Y_s = k$ and pairwise cliques $(Y_{t_1}, Y_s) = (k, q_1)$, $(Y_s, Y_{t_2}) = (k, q_2)$, $(Y_{t_3}, Y_s) = (k, q_3)$, and $(Y_s, Y_{t_4}) = (k, q_4)$ in the neighborhood window H_s, where $t_{1,2} \in C_{2H}$, $t_{3,4} \in C_{2V}$, and $q_j \in \Omega$. The frequency of each clique in H_s is then considered as an expert.

In this way, the high-order label dependencies of pixel s are decomposed into low-order ones. Based on POE, the spatial label interactions of s are modeled, expressed as

$$f_s(k) = \frac{g_s(k)\prod_{t=1}^{4} g_s(k,q_t)}{\sum_{k=1}^{K} g_s(k)\prod_{t=1}^{4} g_s(k,q_t)} \tag{8}$$

where $f_s(k)$ denotes the probability that Y_s takes k. $g_s(k) = M_k/M_{C_1}$ and M_k denote the frequency and number of single cliques with the label being k in H_s, where M_{C_1} indicates the total number of single cliques within H_s. $g_s(q_1,k) = M_{(q_1,k)}/M_{C_{2H}}$ and $M_{(q_1,k)}$ represent the frequency and number of pairwise cliques with the label being (q_1,k) in H_s, which is counted by the normalized joint histogram (NJH) [39]. $M_{C_{2H}}$ indicates the total number of horizontal pairwise cliques within H_s. $g_s(k,q_2)$, $g_s(k,q_3)$, and $g_s(k,q_4)$ have similar definitions.

The second branch measures the label dependencies based on the classification of SCN and obtains the POE-based class probabilities. First, the label frequencies of single cliques are calculated in high-order neighborhoods. Thus, for homogeneous regions, the label with the largest proportion can correct the misclassification caused by speckle. In addition, the model considers the labels of pairwise cliques. Therefore, pairwise cliques with the same labels [e.g., (k,k)] can further suppress speckle. In contrast, pairwise cliques with different labels [e.g., $(k,q)_{k\neq q}$] take into account the edge information, which prevents the loss of structure information due to overcorrection.

C. Information Fusion

Finally, to preserve more edge structure, DST is applied to introduce the pixel-level information obtained by SCN into the fusion iteration. Specifically, in evidential theory, the label set Ω is considered as the frame of discernment. The class probability p and the spatial statistics class probability f are Bayesian basic belief assignment (BBA) functions m, satisfying two constraints

$$\begin{cases} m(\varnothing) = 0 \\ \sum_{A\subseteq\Omega} m(A) = 1 \end{cases} \tag{9}$$

where $\varnothing$ is an empty set, and proposition A denotes a subset of Ω. The value of $m(A)$ can be interpreted as a part of the unit belief mass, which is assigned to A based on a corpus of evidence.

Based on the evidential theory, the class probability p and spatial statistics class probability f are fused by

$$m(k) = p \oplus f(k) = \frac{p(k)f(k)}{\sum_{k=1}^{K} p(k)f(k)} \tag{10}$$

where $\oplus$ represents the fusion operation. Next, the classification result of the fusion probability m is fed to the second branch to calculate its spatial statistics class probability, which is fused with the class probability of SCN to obtain a new fusion probability. After several iterations, the final classification results are obtained. In this way, the SCN-SSF model effectively preserves the edge structure while enhancing label consistency.

IV. Experiment Results and Discussion

A. Implementation

SCN-SSF model is tested against other models: 1) CNN [17]; 2) CVCNN [28]; 3) SCN; 4) the semisupervised PL3DCNN [30]; 5) CVCNN-SSF. These models have similar parameter settings. Concretely, 100 samples with the size being $12 \times 12 \times 6$ per class [40] are fed into the networks. For the semisupervised model, some unlabeled samples are randomly selected and participate in the training. The networks are optimized by the Adam algorithm with the learning rate being 0.001, and the number of iterations is set to 300. The batch sizes of the labeled set and the unlabeled set are 50 and 120, respectively. The experiments are performed on Pytorch-CPU (4.2 GHz).

B. Hyperparameter Analysis

Besides, there are two hyperparameters that need to be determined through experiments. The experiments are conducted on two representative datasets, including the San Francisco and Flevoland1 dataset, and the results are shown in Fig. 6.

The first experiment is to choose a suitable size for the neighborhood window. Based on the classification results of SCN, the size of the neighborhood window in the spatial statistics branch varies from 3 to 23, and the corresponding classification results of SCN-SSF are shown in Fig. 6(a). For the Flevoland1 dataset, as the window size increases to 11, the classification accuracy gradually converges. For the San Francisco dataset, the classification accuracy gradually improves as the window size increases. In addition, larger window sizes mean longer running times. Overall, when the window size is 15, the accuracies of the Flevoland1 and San Francisco datasets are relatively high. Consequently, in our experiments, the window size is set to 15.

Another experiment is performed to determine the number of iterations. In the SSF part, the number of iterations is increased from 0 to 15, and the change in classification accuracy is presented in Fig. 6(b). For the Flevoland1 and San Francisco datasets, the classification accuracies gradually converge as the number of iterations increases to 6. As a result, the number of iterations in our experiment is set to 6.

C. Ablation Experiment

To clearly illustrate the roles of SCN, POE, and DST in SCN-SSF, three ablation experiments are designed, including SCN, CVCNN-SSF, and SCN-POE. The comparison of SCN and CVCNN can validate the effectiveness of the semisupervised network. In addition, the comparison of CVCNN-SSF and SCN-SSF can clarify the role of the SCN module in SCN-SSF. Similarly, the role of POE is illustrated by comparing SCN and SCN-POE, and the role of DST is illustrated by comparing SCN-POE and SCN-SSF.

SCN and CVCNN-SSF are used as reference models, and their experiment results are shown in Figs. 9, 11, and 12. The corresponding quantitative evaluations are listed in Tables I–III. Comparing the classification results of SCN and CVCNN, it can be found that Wheat2 regions in Fig. 11(j) are well classified, and Wheat2 in Fig. 11(f) is not identified, which is also confirmed

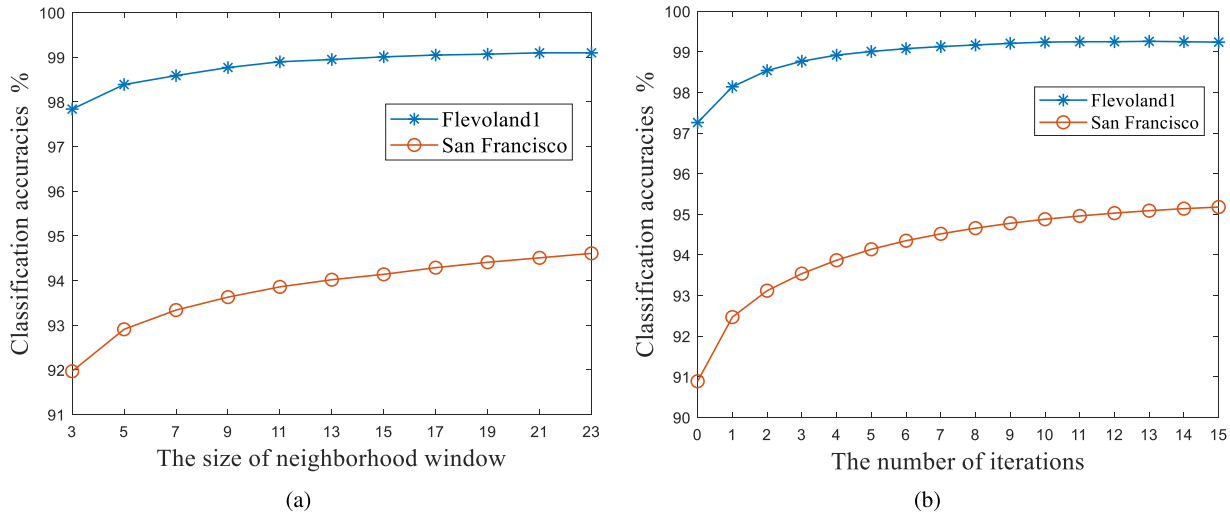

Fig. 6. Effect of different hyperparameters on classification accuracy. (a) Effect of the neighborhood window size on classification accuracy. (b) Effect of iteration number on classification accuracy.

TABLE I
QUANTITATIVE EVALUATION (OA/KAPPA) OF THE CLASSIFICATION FOR THE SAN FRANCISCO DATASET

Class	CNN	CVCNN	PL3DCNN	SCN	CVCNN-SSF	SCN-SSF
High-density urban	84.58%	84.13%	81.82%	87.97%	90.27%	93.92%
Vegetation	71.46%	77.91%	74.59%	80.51%	84.19%	87.96%
Water	98.16%	98.84%	99.02%	99.13%	99.17%	99.46%
Developed urban	91.99%	88.15%	89.29%	87.23%	94.45%	92.97%
Low-density urban	84.84%	84.61%	78.45%	83.33%	95.53%	95.72%
OA	88.42%	89.53%	88.43%	90.90%	93.33%	94.88%
Kappa	0.8390	0.8539	0.8386	0.8728	0.9067	0.9281
running time(s)	31	68	643	106	68+254	106+254

by the quantitative evaluation listed in Table II. Similarly, the Cropland regions framed by the black box in Fig. 12(f) are well classified while some Cropland regions in Fig. 12(d) are misclassified as Water. As a result, with a small number of labeled samples, SCN possesses a stronger ability to identify different terrains compared to CVCNN. In addition, comparing the classification results of CVCNN-SSF and SCN-SSF, it can be found that SSF as a simple postprocessing method can effectively enhance contextual smoothness and correct some misclassifications. However, in Figs. 11(k) and 12(g), the misclassifications of Wheat2 and Cropland framed by the black boxes still exist and cannot be corrected by SSF. Therefore, to fully utilize the ability of SSF, deep networks trained with a small number of samples need to have stronger recognition ability. SCN introduces unlabeled samples to participate in the training of CVCNN and designs the objective function to induce a more focused network output. In this way, SCN improves the generalization of the model and alleviates overfitting, thus improving the ability to recognize different terrains.

To illustrate the effectiveness of DST, we conduct experiments on the Flevoland1 dataset, and the classification accuracies of SCN-POE and SCN-SSF with different numbers of iterations are shown in Fig. 7. As the number of iterations increases, the classification accuracy of SCN-SSF gradually increases and converges while that of SCN-POE gradually increases and then decreases. Compared with SCN, SCN-POE can efficiently smooth

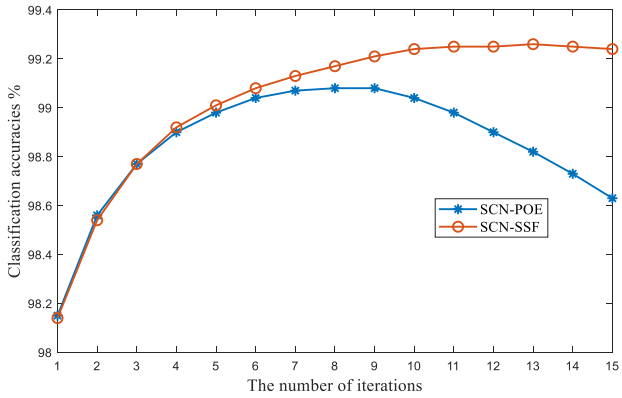

Fig. 7. Classification accuracies of the Flevoland1 dataset with different numbers of iterations.

homogeneous regions and correct some misclassifications, but it can also lead to the loss or even distortion of edge information due to oversegmentation. To visualize this, a simulated PolSAR image with clear edges is generated [13] and classified. In detail, four homogeneous regions in the Flevoland2 dataset are selected and then fill the ground truth map in Fig. 8(c). The obtained simulated image is shown in Fig. 8(b). The experiment results of SCN, SCN-POE, and SCN-SSF are shown in Fig. 8(d)–8(f). Obviously, as shown in Fig. 8(e), POE has a strong ability to enhance label consistency, and the misclassifications in Fig. 8(d) are corrected. However, at the same time, the smoothing ability of POE leads to the loss or even distortion of edge information. Fortunately, based on DST, SCN-SSF continuously introduces pixel-level information in the fusion iterations to prevent the loss of edge information. As a result, compared to Fig. 8(d), the classification of Fig. 8(f) possesses smoother homogeneous regions while preserving structural information. Quantitatively, the classification accuracy of SCN-SSF is 94.27%, which is 1.09% higher than that of SCN-POE. Overall, POE has a strong

TABLE II
QUANTITATIVE EVALUATION (OA/KAPPA) OF THE CLASSIFICATION FOR THE FLEVOLAND1 DATASET

Class	CNN	CVCNN	PL3DCNN	SCN	CVCNN-SSF	SCN-SSF
Stembeans	99.61%	99.75%	99.57%	99.69%	99.74%	99.52%
Peas	97.18%	97.48%	94.52%	97.21%	99.63%	99.48%
Forest	96.77%	94.23%	89.69%	97.35%	99.65%	99.50%
Lucerne	95.25%	96.44%	96.58%	95.52%	97.71%	96.01%
Wheat	94.37%	97.63%	93.69%	96.57%	99.11%	98.77%
Beet	98.82%	98.70%	94.35%	98.53%	99.50%	99.37%
Potatoes	95.83%	94.28%	89.64%	98.84%	99.55%	99.81%
Bare soil	99.16%	99.29%	99.58%	1	1	1
Grasses	93.19%	96.00%	92.23%	95.69%	99.03%	98.82%
Rapeseed	90.32%	92.81%	78.83%	94.29%	98.29%	98.94%
Barley	97.07%	96.30%	96.95%	98.94%	99.34%	1
Wheat2	91.00%	0	89.30%	91.95%	0	96.64%
Wheat3	98.13%	98.83%	97.52%	99.13%	99.96%	99.81%
Water	0	1	99.86%	98.66%	1	99.09%
Buildings	93.70%	95.59%	1	98.53%	91.18%	1
OA	87.48%	90.36%	93.19%	97.26%	92.65%	99.01%
Kappa	0.8639	0.8947	0.9256	0.9701	0.9198	0.9892

TABLE III
QUANTITATIVE EVALUATION (OA/KAPPA) OF THE CLASSIFICATION FOR THE FLEVOLAND2 DATASET

Class	CNN	CVCNN	PL3DCNN	SCN	CVCNN-SSF	SCN-SSF
Urban area	72.71%	69.19%	71.54%	72.89%	75.77%	78.60%
Cropland	79.63%	80.68%	74.24%	82.27%	83.91%	85.91%
Water	93.36%	94.59%	92.08%	95.04%	94.78%	94.93%
Woodland	84.32%	85.87%	74.59%	86.80%	91.63%	91.64%
OA	81.95%	82.23%	76.89%	83.86%	86.35%	87.62%
Kappa	0.7544	0.7578	0.6848	0.7801	0.8140	0.8313

ability to smooth homogeneous regions, but it can also lead to the loss and distortion of edge information. In SCN-SSF, DST is applied to continuously introduce pixel-level information into the fusion iterations, thus preventing edge distortion and preserving structural information.

D. Experiment on the San Francisco Dataset

The San Francisco dataset (900×1024) is acquired by NASA/JPL AIRSAR (*L*-band). It has five terrain classes including High-density urban, Vegetation, Water, Developed urban, and Low-density urban, as shown in Fig. 9.

As shown in Fig. 9(c), CNN achieves good results in the Developed urban areas (yellow), where fewer pixels are misclassified as High-density urban (red). Compared to CNN, CVCNN identifies the Vegetation regions better (green), as shown in Fig. 9(d). However, they both misclassify the Water regions in the upper right corner of Fig. 9(c)–(d) (marked by the white box) due to the lack of sufficient labeled samples. As shown in Fig. 9(e), the semisupervised PL3DCNN is severely affected by speckle. Fortunately, our SCN is more robust against speckles, as shown in Fig. 9(f). Furthermore, SCN-SSF utilizes the contextual information of samples to model the spatial label dependencies. As shown in Fig. 9(h), SCN-SSF achieves smoother homogeneous areas and identifies the Vegetation areas more accurately.

Quantitative evaluation including overall accuracy (OA) and Kappa parameter [40] further validates the superiority of the proposed model. The objective evaluations and Running time are listed in Table I. Concretely, CNN with short running times can identify the Developed urban and the Low-density urban well but performs poorly in the Vegetation regions. CVCNN fully utilizes the amplitude and phase information of the polarimetric data, thus achieving higher accuracy. The semisupervised PL3DCNN achieves good results in the Water regions, but unsatisfactory results in the Low-density urban. Moreover, the training of PL3DCNN is time-consuming. Different from the fixed pseudolabels of PL3DCNN, our SCN continuously generates the pseudolabels of unlabeled data in iterations, thus capturing more useful information. Consequently, the classification accuracy of SCN reaches 90.90%. Moreover, the Running time of SCN is relatively low compared to the semisupervised PL3DCNN. Besides, SCN-SSF also extracts contextual label information through SSF, which further improves classification accuracy by 3.98%. Finally, the classification accuracy of SCN-SSF enhances by 6.46%, 5.35%, and 6.45%, compared to that of CNN, CVCNN, and PL3DCNN, respectively, with the Kappa increasing by 0.0891, 0.0742, and 0.0895, respectively.

Besides, these models with different numbers of labeled samples are experimented on the San Francisco dataset. Fig. 10 shows the classification accuracy when the number of labeled samples for each category is 100, 300, 500, 700, 900, 1100, and 1300. Comparing the results of SCN-SSF, SCN, CVCNN-SSF, and CVCNN, it can be found that SSF can effectively improve classification accuracy. Fortunately, the classification performance of our SCN-SSF is significantly superior to that of CNN, CVCNN, PL3DCNN, and CVCNN-SSF for different numbers of samples. And the experiment shown in Fig. 10 can also be used to determine the number of labeled samples. In this figure, as the number of labeled samples increases, it can be found that the classification accuracies of CNN, CVCNN, PL3DCNN, CVCNN-SSF, and our SCN part all improve significantly. In contrast, the number of labeled samples has less effect on SCN-SSF, and a small number of labeled samples can still achieve good performance. Therefore, in our experiments, only 100 labeled samples per class are selected.

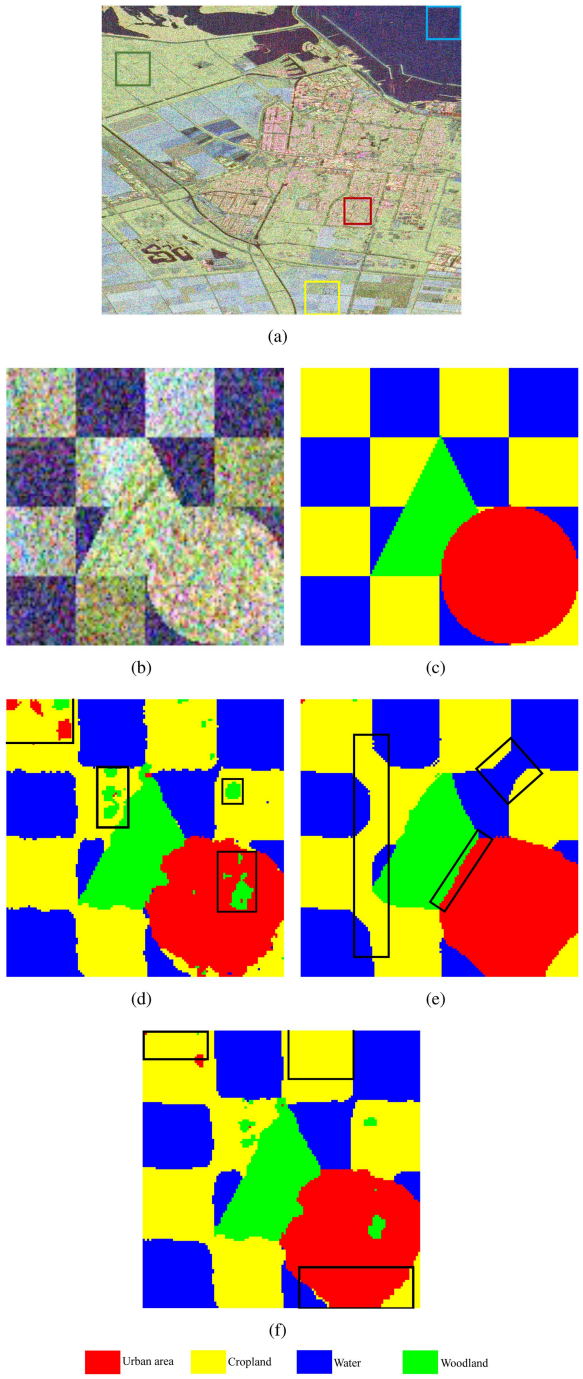

Fig. 8. Classification results of the simulated dataset. (a) Pauli RGB image of Flevoland2 dataset. (b) Pauli RGB image of the simulated dataset. (c) Ground truth. (d) SCN. (e) SCN-POE. (f) SCN-SSF. The classification accuracies of SCN, SCN-POE, and SCN-SSF are 92.49%, 93.18%, and 94.27%, respectively.

E. Experiment on the Flevoland1 Dataset

The Flevoland1 dataset (750×1024) is acquired by NASA/JPL AIRSAR (*L*-band) and has 15 different categories, including the Stembeans, Peas, Forest, Lucerne, Wheat, Beet, Potatoes, Bare soil, Grasses, Rapeseed, Barley, Wheat2, Wheat3, Water, and Buildings. Its pseudocolor image is shown in Fig. 11(a), and the corresponding ground truth is shown in Fig. 11(b).

CNN, as the basic deep learning model, obtains good classification results, as shown in Fig. 11(c). Considering the characteristics of polarimetric data, CVCNN extends the real network to the complex domain and exploits the amplitude and phase information of polarimetric data. Thus, compared to Fig. 11(c), the Water regions in the upper right of Fig. 11(e) (marked by the white box) are correctly classified. However, the two deep models rely on a large number of labeled samples to ensure excellent performance while labeled samples are limited in our experiment. The semisupervised PL3DCNN assigns pseudolabels to unlabeled samples by the K-means algorithm before network training, thus expanding the training set. As shown in Fig. 11(g)–(h), PL3DCNN can accurately identify each category but suffers greatly from speckle. SCN generates and updates pseudolabels for unlabeled samples in training, which captures more useful information to improve the performance, as shown in Fig. 11(i). Besides, the SSF part is designed to explore contextual label information. Obviously, in Fig. 11(n), some misclassifications of SCN are corrected. In addition, comparing the white box regions in Fig. 11(i) and (m), we can find that the SCN-SSF model not only corrects the misclassification using spatial information but also preserves the clear edges. In summary, SCN-SSF obtains smooth homogeneous regions and preserves the edge structure well.

Quantitatively, Table II lists its objective evaluation. For the results of the Flevoland1 dataset, CNN obtains high accuracy in all categories except the Water regions. Similarly, CVCNN has poor classification performance in the Wheat2 regions. Classifying images containing many terrains using a small number of labeled samples is a challenging task, and some terrains may not be identified. By introducing unlabeled samples into the training, the semisupervised PL3DCNN can identify all the categories and improve the classification accuracy to 93.19%. As a semisupervised model, our SCN improves the accuracy to 97.26%, which is attributed to two aspects. On the one hand, the pseudolabels of unlabeled samples are constantly updated during training, which is more favorable for classification. On the other hand, SCN makes full use of the amplitude and phase information of the polarimetric data. The extracted features contain more useful information. Furthermore, SCN-SSF is designed with a simple postprocessing module that only needs to count the labels in the neighborhood, which increases the accuracy by another 1.75%. Finally, the OA of SCN-SSF is higher than that of CNN, CVCNN, and PL3DCNN by 11.53%, 8.65%, and 5.82%, respectively, with the Kappa increasing by 0.1253, 0.0945, and 0.0636, respectively.

F. Experiment on the Flevoland2 Dataset

The Flevoland2 dataset with the size being 1400×1200 is acquired by RADARSAT-2 (*C*-band) in April 2009. The pseudocolor and ground truth image are shown in Fig. 12(a) and (b), respectively. The dataset contains four different terrains, involving the Urban area, Cropland, Water, and Woodland.

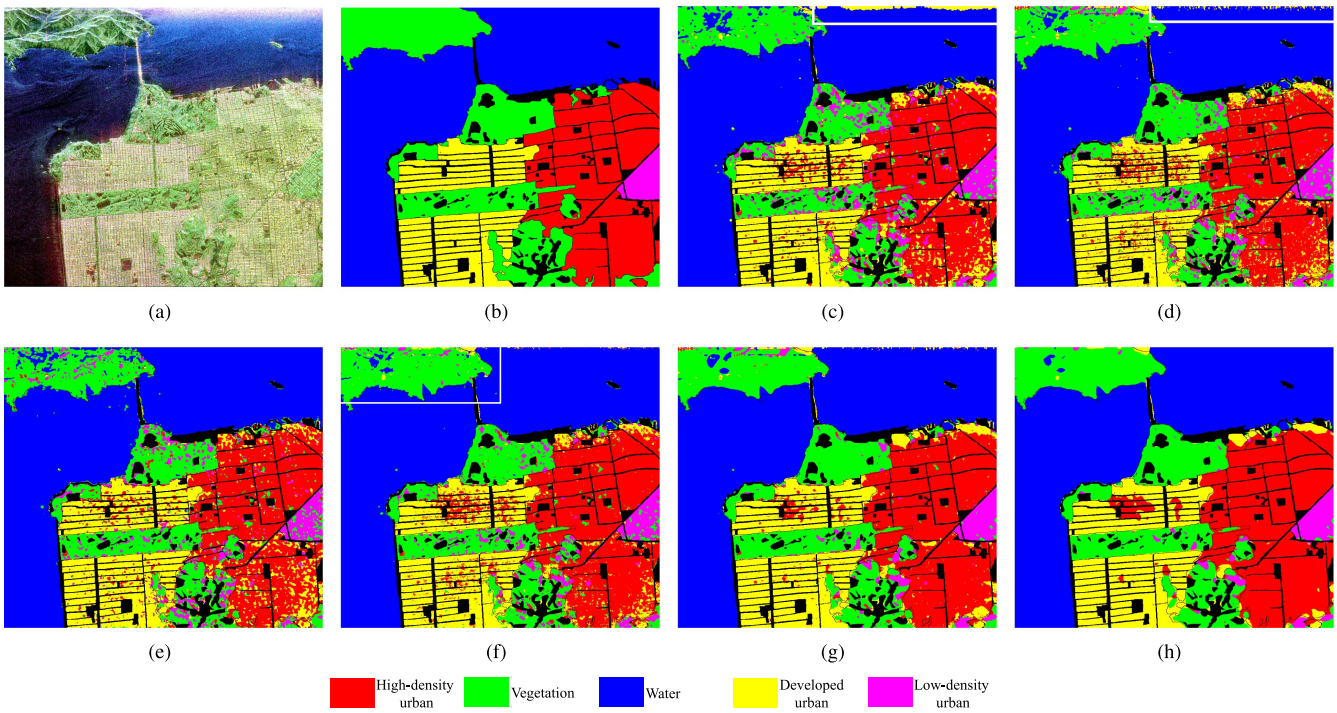

Fig. 9. Classification results of the San Francisco dataset. (a) Pauli RGB. (b) Ground truth. (c) CNN result overlaid with the ground truth. (d) CVCNN result overlaid with the ground truth. (e) PL3DCNN result overlaid with the ground truth. (f) SCN result overlaid with the ground truth. (g) CVCNN-SSF result overlaid with the ground truth. (h) SCN-SSF result overlaid with the ground truth.

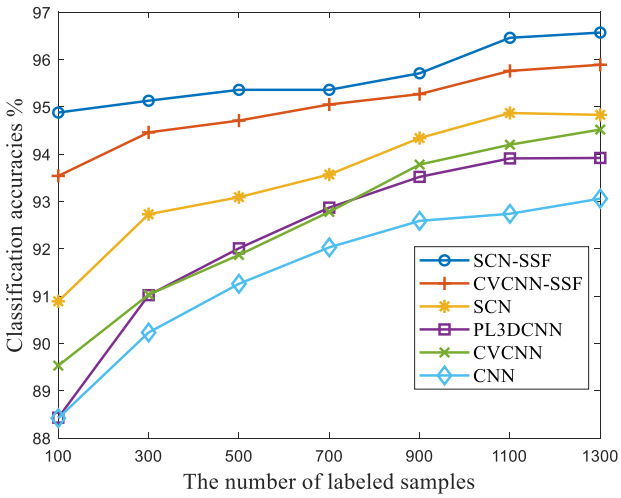

Fig. 10. Effect of the number of labeled samples on the classification accuracy.

CNN and CVCNN achieve good results in general, as shown in Fig. 12(c) and (d). However, these two supervised networks lack sufficient labeled samples for training, and the networks are not discriminative enough. Consequently, there are many obvious misclassifications, such as the regions framed by the black box. By contrast, PL3DCNN achieves better results in this region, as shown in Fig. 12(e). As a semisupervised model, PL3DCNN generates pseudolabels for unlabeled samples, thus expanding the training dataset. However, there are many misclassifications caused by speckle. Different from PL3DCNN, pseudolabels of SCN are generated iteratively during training. Our SCN makes better use of unlabeled samples. The Cropland areas within the black box are correctly classified. And there are fewer misclassifications caused by speckle, such as the regions framed by the purple box in Fig. 12(f). In addition, SSF is introduced to improve the performance. As shown in Fig. 12(h), SCN-SSF has smoother homogeneous regions and corrects some misclassifications in SCN.

Furthermore, their corresponding quantitative evaluations are presented in Table III. CNN has a good performance in the Urban areas while CVCNN performs well in the Cropland, Water, and Woodland areas. The OA of CNN and CVCNN is 5.06% and 5.34% higher than that of PL3DCNN, respectively. Fortunately, our SCN achieves good results in all categories, especially in the Cropland regions, which is consistent with the previous visual analysis. The OA of SCN enhances by 1.91% and 1.63%, respectively, compared to that of CNN and CVCNN. Besides, SSF is introduced in SCN-SSF, which improves the classification accuracy to 87.62%. Overall, the OA of SCN-SSF is 5.67%, 5.39%, and 10.73% higher than that of CNN, CVCNN, and PL3DCNN, respectively, and the Kappa is improved by 0.0769, 0.0735, and 0.1465, respectively.

In addition, our proposed SCN-SSF is compared with the other three state-of-the-art semisupervised models to illustrate its advantages. Improved Tritraining with a minimum spanning

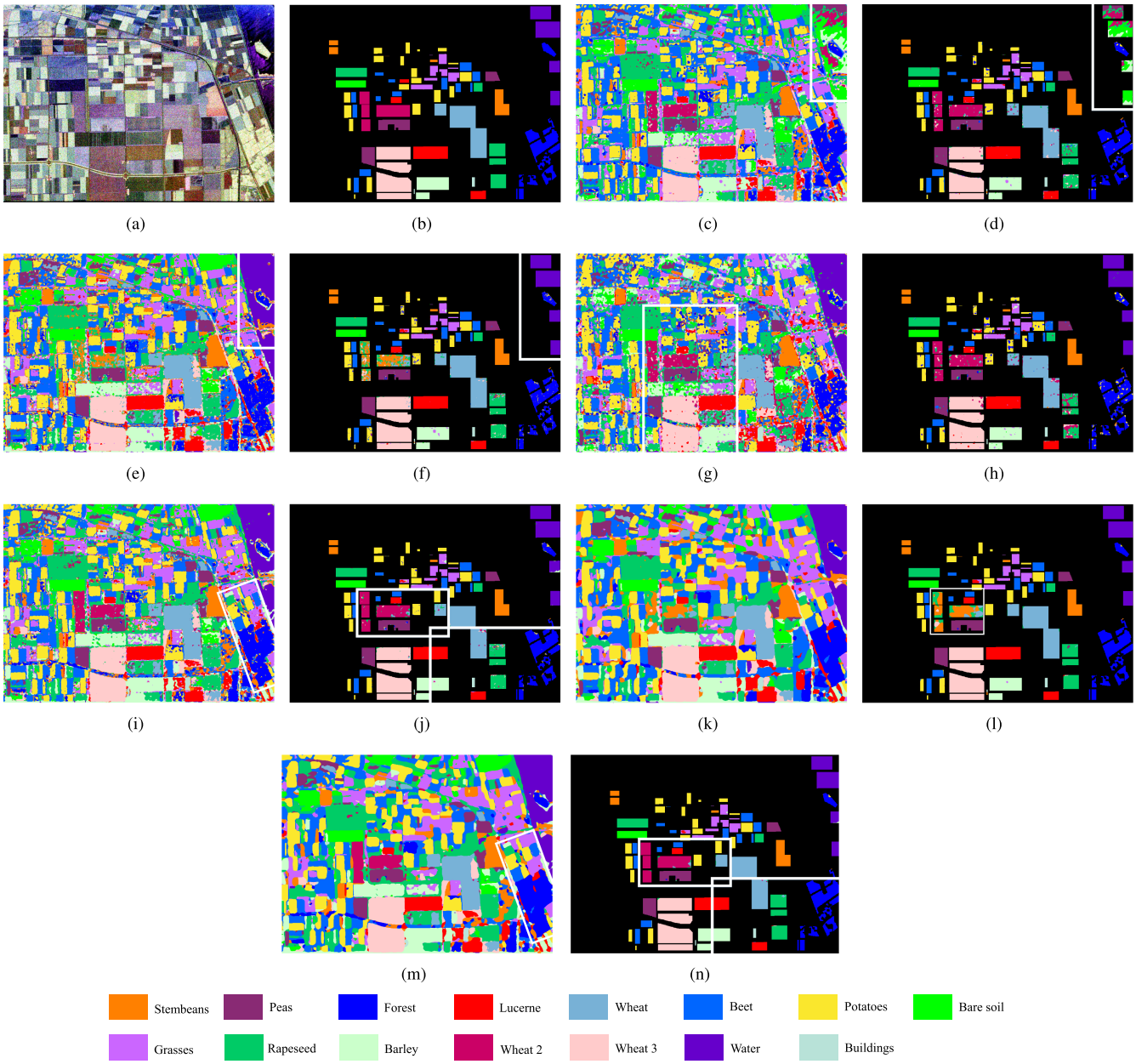

Fig. 11. Classification results of the Flevoland1 dataset. (a) Pauli RGB. (b) Ground truth. (c) CNN. (d) CNN result overlaid with the ground truth. (e) CVCNN. (f) CVCNN result overlaid with the ground truth. (g) PL3DCNN. (h) PL3DCNN result overlaid with the ground truth. (i) SCN. (j) SCN result overlaid with the ground truth. (k) CVCNN-SSF. (l) CVCNN-SSF result overlaid with the ground truth. (m) SCN-SSF. (n) SCN-SSF result overlaid with the ground truth.

tree (ITT-MST) [41] is a semisupervised method that recognizes different terrains well, as shown in Fig. 13(b). However, ITT-MST is not robust enough against speckles, and Fig. 13(b) has many misclassifications. Compared to ITT-MST, cross label-information network (CLIN) [42] obtains smooth homogeneous regions as shown in Fig. 13(c). CLIN finds similar sample pairs by the scattering mechanism of PolSAR and achieves consistency regularization. However, the Cropland regions in the black box are misclassified as Water regions, and the edge locations of the Water regions in the purple box are inaccurate. PolNet [43] efficiently obtains the polar-spatial fusion features and a smooth classification map by imposing a Gaussian random field prior to the fusion features and a CRF posterior on the label configuration. Thus, as shown in Fig. 13(d), PolNet achieves a better classification result, with the accuracy being 85.25%. Our proposed SCN-SSF utilizes unlabeled samples to improve the recognition ability of CVCNN. Besides, a parameter-free SSF module is designed to enhance the classification smoothness and correct some misclassifications. In this way, with a small number of labeled samples, our SCN-SSF recognizes different terrains more accurately and obtains smooth homogeneous regions, as shown in Fig. 13(e).

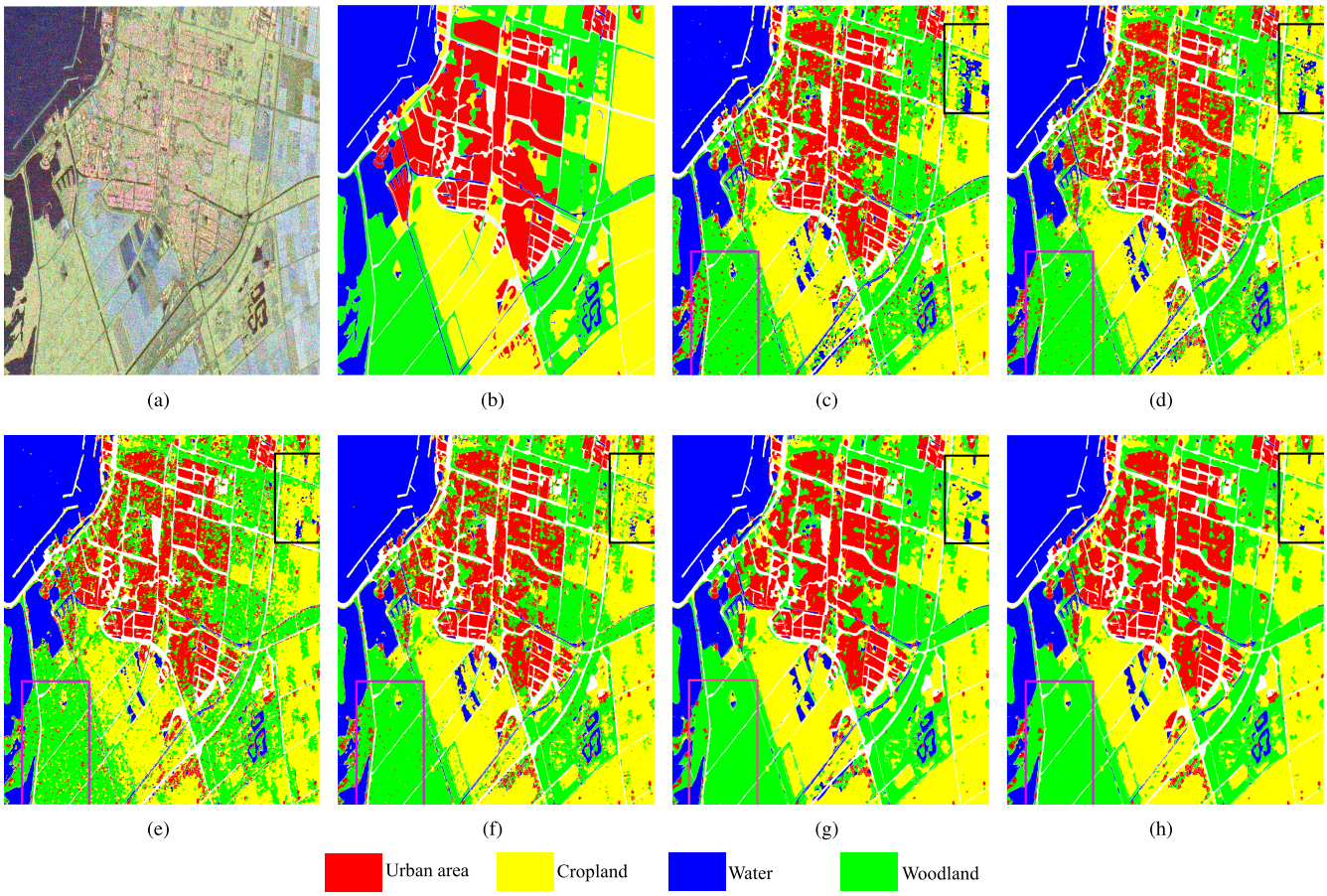

Fig. 12. Classification results of the Flevoland2 dataset. (a) Pauli RGB. (b) Ground Truth. (c) CNN result overlaid with the ground truth. (d) CVCNN result overlaid with the ground truth. (e) PL3DCNN result overlaid with the ground truth. (f) SCN result overlaid with the ground truth. (g) CVCNN-SSF result overlaid with the ground truth. (h) SCN-SSF result overlaid with the ground truth.

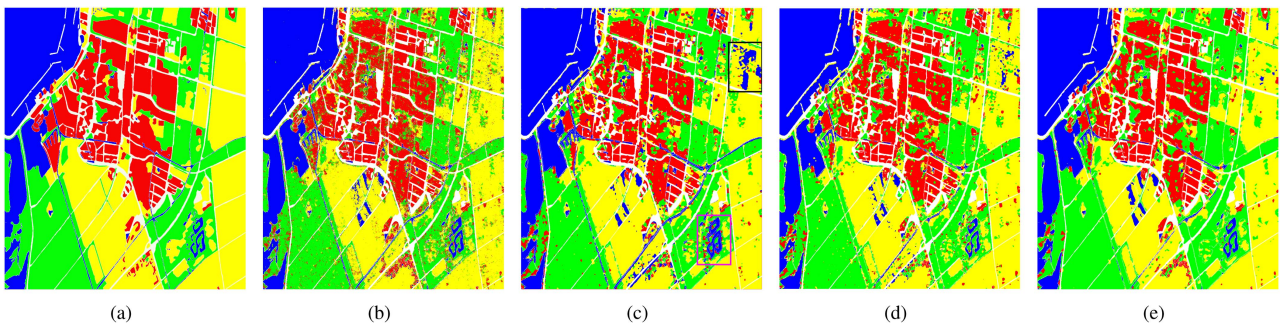

Fig. 13. Classification results for different semisupervised models. (a) Ground truth. (b) 84.60% ITT-MST. (c) 84.92% CLIN. (d) 85.25% PolNet. (e) 87.62% SCN-SSF.

V. Conclusion

In this article, an SCN-SSF model is proposed for PolSAR image classification, which achieves satisfactory classification results with a small number of labeled samples. First, the SCN part of SCN-SSF iteratively updates the pseudolabels of unlabeled samples during the training of CVCNN. Their errors constitute the regularization term of the objective function, thus improving the recognition ability of the network. And the classification has higher accuracy and belief. Then, a parameter-free spatial statistics module is designed to model the spatial label dependencies by POE, which effectively enhances contextual smoothness and corrects some misclassifications. Finally, DST is used to integrate contextual label information and pixel-level information to preserve structural information and prevent oversegmentation. Experimental results have demonstrated that the SCN-SSF model can accurately identify different terrains using only a small number of labeled samples and obtain smoother homogeneous regions while preserving edge information.

In the future, there are two directions to try. First, SSF can be used as a postprocessing module for other networks to enhance classification smoothness. Besides, we will investigate incorporating the spatial statistics module into an appropriate network framework to construct an end-to-end classification model, rather than as a postprocessing component. In this way, the network will also be able to take spatial information into account during training, fully integrating pixel-level and spatial information for classification.

REFERENCES

[1] J.-S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*. New York. NY, USA: CRC Press, 2009.

[2] L. Deng, Y.-N. Yan, and C. Sun, "Use of sub-aperture decomposition for supervised PolSAR classification in urban area," *Remote Sens.*, vol. 7, no. 2, pp. 1380–1396, 2015.

[3] Y. Jiang, M. Li, P. Zhang, X. Tan, and W. Song, "Unsupervised complex-valued sparse feature learning for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 1, Jun. 2022, Art. no. 5230516.

[4] 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.

[5] T. Eltoft, S. N. Anfinsen, and A. P. Doulgeris, "A multitexture model for multilook polarimetric synthetic aperture radar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2910–2919, May 2014.

[6] A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Classification with a non-Gaussian model for PolSAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2999–3009, Oct. 2008.

[7] C. C. Freitas, A. C. Frery, and A. H. Correia, "The polarimetric G distribution for SAR data analysis," *Environmetrics*, vol. 16, no. 1, pp. 13–31, 2005.

[8] L. Bombrun and J.-M. Beaulieu, "Fisher distribution for texture modeling of polarimetric SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 512–516, Jul. 2008.

[9] A. Masjedi, M. J. V. Zoej, and Y. Maghsoudi, "Classification of polarimetric SAR images based on modeling contextual information and using texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 932–943, Feb. 2016.

[10] H. Bi, J. Sun, and Z. Xu, "A graph-based semisupervised deep learning model for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2116–2132, Apr. 2019.

[11] H. Bi, L. Xu, X. Cao, Y. Xue, and Z. Xu, "Polarimetric SAR image semantic segmentation with 3D discrete wavelet transform and Markov random field," *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 6601–6614, Jun. 2020.

[12] H. Bi, J. Yao, Z. Wei, D. Hong, and J. Chanussot, "PolSAR image classification based on robust low-rank feature extraction and Markov random field," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, no. 1, Nov. 2022, Art. no. 4005205.

[13] W. Song, M. Li, P. Zhang, Y. Wu, X. Tan, and L. An, "Mixture WG γ -MRF model for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 905–920, Feb. 2018.

[14] Y. Wu, K. Ji, W. Yu, and Y. Su, "Region-based classification of polarimetric SAR images using Wishart MRF," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 668–672, Oct. 2008.

[15] S. Zhang, B. Hou, L. Jiao, Q. Wu, C. Sun, and W. Xie, "Context-based max-margin for PolSAR image classification," *IEEE Access*, vol. 5, pp. 24070–24077, 2017.

[16] Y. Jiang, M. Li, P. Zhang, and Z. Wang, "Unsupervised deep sparse features extraction for SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 1, Aug. 2022, Art. no. 5234315.

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

[18] W. Li, C. Yang, Y. Peng, and J. Du, "A pseudo-siamese deep convolutional neural network for spatiotemporal satellite image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, no. 1, pp. 1205–1220, Jan. 2022.

[19] X. Wu, D. Hong, and J. Chanussot, "UIU-Net: U-net in U-net for infrared small object detection," *IEEE Trans. Image Process.*, vol. 32, no. 1, pp. 364–376, Dec. 2022.

[20] C. Li, B. Zhang, D. Hong, J. Yao, and J. Chanussot, "LRR-Net: An interpretable deep unfolding network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, no. 1, May 2023, Art. no. 5513412.

[21] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 6645–6649.

[22] M. Chen, X. He, J. Yang, and H. Zhang, "3-D convolutional recurrent neural networks with attention model for speech emotion recognition," *IEEE Signal Process. Lett.*, vol. 25, no. 10, pp. 1440–1444, Oct. 2018.

[23] L. Yi and M.-W. Mak, "Improving speech emotion recognition with adversarial data augmentation network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 1, pp. 172–184, Jan. 2022.

[24] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3104–3112.

[25] M. Lippi, M. A. Montemurro, M. D. Esposti, and G. Cristadoro, "Natural language statistical features of LSTM-generated texts," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3326–3337, Nov. 2019.

[26] R. Xiao et al., "Towards energy-preserving natural language understanding with spiking neural networks," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 31, no. 1, pp. 439–447, Nov. 2023.

[27] 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.

[28] 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.

[29] X. 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.

[30] Z. Fang, G. Zhang, Q. Dai, Y. Kong, and P. Wang, "Semisupervised deep convolutional neural networks using pseudo labels for PolSAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, no. 1, Nov. 2020, Art. no. 4005605.

[31] Y. Li et al., "Semi-supervised PolSAR image classification based on self-training and superpixels," *Remote Sens.*, vol. 11, 2019, Art. no. 1933.

[32] J. Guo, L. Wang, D. Zhu, and G. Zhang, "Semisupervised classification of PolSAR images using a novel memory convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, no. 1, Jan. 2021, Art. no. 4007605.

[33] F. Zhao, M. Tian, W. Xie, and H. Liu, "A new parallel dual-channel fully convolutional network via semi-supervised FCM for PolSAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, no. 6, pp. 4493–4505, Aug. 2020.

[34] 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.

[35] Y. Hu, J. Fan, and J. Wang, "Classification of PolSAR images based on adaptive nonlocal stacked sparse autoencoder," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 7, pp. 1050–1054, Jul. 2018.

[36] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[37] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

[38] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Int. Conf. Mach. Learn.*, vol. 3, p. 2, 2013.

[39] P. Zhang, M. Li, Y. Wu, L. An, and L. Jia, "Unsupervised SAR image segmentation using high-order conditional random fields model based on product-of-experts," *Pattern Recogn. Lett.*, vol. 78, Jul. 15, pp. 48–55, 2016.

[40] X. Tan, M. Li, P. Zhang, Y. Wu, and W. Song, "Deep triplet complex-valued network for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10179–10196, Dec. 2021.

[41] 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.

[42] X. Xin et al., "Semi-supervised classification of dual-frequency PolSAR image using joint feature learning and cross label-information network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 1, Sep. 2022, Art. no. 5235716.

[43] Z. Wen, Q. Wu, Z. 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.

Yinyin Jiang was born in 1994. She received the B.S. degree in electrical science and technology from AnHui Normal University, Wuhu, China, in 2017, and the Ph.D. degree in signal and information processing from the National Key Laboratory of Radar Signal Processing, Xidian University, Xi'an, China, in 2023.

Her primary research interests include synthetic aperture radar image processing and deep learning.

Ming Li (Member, IEEE) was born in 1965. He received the B.S. degree in electrical engineering and the M.S. and Ph.D. degrees in signal processing from Xidian University, Xi'an, China, in 1987, 1990, and 2007, respectively.

In 1987, he joined the Department of Electronic Engineering, Xidian University. He is currently a Professor with the National Key Lab of Radar Signal Processing, Xidian University. His research interests include adaptive signal processing, detection theory, ultrawideband, and SAR image processing.

Peng Zhang (Member, IEEE) received the B.S. degree in electronic and information engineering and the M.S. and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 2006, 2009, and 2012, respectively.

He is currently an Associate Professor with the National Key Lab of Radar Signal Processing, Xidian University. His main research interests include SAR image interpretation and statistical learning theory.

Wanying Song received the B.S. degree in electrical science and technology from the Shandong University of Science and Technology, Qingdao, China, in 2012, and the Ph.D. degree in signal and information processing from the National Laboratory of Radar Signal Processing, Xidian University, Xi'an, China, in 2018.

Her main research interest includes polarimetric synthetic aperture radar image classification.