Semi-Supervised PolSAR Image Classification Based on Improved Tri-Training With a Minimum Spanning Tree

Shuang Wang, *Member, IEEE*, Yanhe Guo, Wenqiang Hua, *Member, IEEE*, Xinan Liu, *Member, IEEE*, Guoxin Song, Biao Hou, *Member, IEEE*, and Licheng Jiao, *Fellow, IEEE*

***Abstract*— In this article, the terrain classifications of polarimetric synthetic aperture radar (PolSAR) images are studied. A novel semi-supervised method based on improved Tri-training combined with a neighborhood minimum spanning tree (NMST) is proposed. Several strategies are included in the method: 1) a high-dimensional vector of polarimetric features that are obtained from the coherency matrix and diverse target decompositions is constructed; 2) this vector is divided into three subvectors and each subvector consists of one-third of the polarimetric features, randomly selected. The three subvectors are used to separately train the three different base classifiers in the Tri-training algorithm to increase the diversity of classification; and 3) a help-training sample selection with the improved NMST that uses both the coherency matrix and the spatial information is adopted to select highly reliable unlabeled samples to increase the training sets. Thus, the proposed method can effectively take advantage of unlabeled samples to improve the classification. Experimental results show that with a small number of labeled samples, the proposed method achieves a much better performance than existing classification methods.**

***Index Terms*— Polarimetric synthetic aperture radar (PolSAR), semi-supervised learning, terrain classification, Tri-training.**

I. Introduction

In RECENT years, polarimetric synthetic aperture radar (PolSAR) has attracted a great deal of attention because of its wide application for Earth remote sensing in many fields such as ocean dynamics, geoscience, agriculture, and environmental monitoring [1]–[3]. A large number of PolSAR systems (AIRSAR, ESAR, TerraSAR-X, Radarsat-2, and ALOS2) have been developed in the past few decades and some PolSAR images have been made publicly available.

Manuscript received February 22, 2020; accepted March 25, 2020. This work was supported by the National Natural Science Foundation of China under Grant 61771379 and Grant 61173092, and in part by the Project supported the Foundation for Innovative Research Groups of the National Natural Science Foundation of China under Grant 61621005. *(Corresponding author: Shuang Wang.)*

Shuang Wang, Yanhe Guo, Biao Hou, Guoxin Song, and Licheng Jiao are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an 710071, China.

Wenqiang Hua is 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.

Xinan Liu is with the Department of Mechanical Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: shwang@mail.xidian.edu.cn).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2020.2988982

Amid the utilizations of PolSAR data, land cover classification is much in demand. A number of PolSAR classification methods, including target decomposition [4], [5], multiple statistical distributions [6], [7], and sparse representation [8], have been developed in recent years. In general, these methods can be divided into unsupervised and supervised classifications [9]. An unsupervised classification is mainly based on polarimetric target decomposition, for instance, Pauli decomposition [4], H/A/Alpha decomposition [5], Freeman three-component decomposition [6], Krogager decomposition [10], Huynen decomposition [11], Touzi decomposition [12], Yamaguchi four-component decomposition [13], and B&F decomposition [14]. The parameters of the polarimetric target decomposition are related to the physical properties of terrain objects and can be used to identify terrain classes. In contrast, a supervised classification is mainly based on machine learning. Several well-recognized classifications include k-nearest neighbor (KNN) [15], decision tree (DT) [16], and support vector machine (SVM) [17], where labeled samples are used to train a classifier.

Recently, researchers have proposed a number of semi-supervised classification (SSC) methods that use both labeled and unlabeled samples in the training sets [18], [19]. These methods include graph-based SSC [20], semi-supervised SVM (S3VM) [21], generative adversarial networks-based SSC [22], self-training-based SSC [23], and co-training-based SSC [24]. As a prominent SSC method, the co-training algorithm [25] uses two independent sets of views to separately train two different classifiers and enlarges the training set of one classifier with the unlabeled samples predicted by the other classifier. This semi-supervised method has been employed in certain tasks, such as clustering [26] and hyperspectral image classification [27]. In terms of that the two independent views in the co-training method are difficult to obtain in many applications, Zhou and Li [28] proposed a Tri-training algorithm that only requires one sufficient view but with three different base classifiers. Their method can not only easily deal with the prediction problems of the unlabeled samples but can also apply the idea of ensemble learning to improve generalization ability. Ideas similar to the Tri-training method have been used to solve certain machine-learning problems, for instance, hyperspectral image classification [29].

It should be pointed out that the base classifiers in both co-training and Tri-training methods receive further positive training if proper unlabeled samples are selected. However, if improper unlabeled samples are selected, the base classifiers receive certain noisy labels and these noisy labels decrease the accuracy of classification. Especially, when available labeled samples are few, it is difficult to simulate the distribution of the entire data and is a challenge to the selection of proper unlabeled samples in both co-training and Tri-training. The two independent views in the co-training and three base classifiers in the tri-training are used for enough diversity as a prerequisite for the effective operation of these two methods. This diversity directly affects the reliability of selected unlabeled samples. The quality of the base classifiers increases with increasing the reliability of selected unlabeled samples. The diversity of polarimetric features of the PolSAR data can also enhance the reliability of the selected samples. In addition, the spatial information in PolSAR images can further improve the reliability of the selected samples, which has not been considered in the methods.

In this article, we propose a novel SSC method based on the Tri-training scheme with two improvements on the diversity of classification and the reliability of the selected unlabeled samples. To increase the diversity, a high-dimensional vector of polarimetric features obtained from the coherency matrix and various target decompositions is constructed. This vector is divided into three subvectors and each subvector contains one-third of the polarimetric features, randomly selected. The three subvectors are used to separately train the three different base classifiers in the Tri-training with different labeled samples. The reliable unlabeled samples are selected with an improved neighborhood minimum spanning tree (NMST) [30]. This selection scheme is similar to the minimum spanning tree (MST) that exploits local spatial information [31].

The main contributions of this article are summarized as follows: 1) the SSC method based on the Tri-training proposed herein can effectively classify various PolSAR images with a small number of labeled samples; 2) the partition of the high-dimensional vector into three subvectors each containing one-third of the polarimetric features increases the diversity of the Tri-training, as the three different base classifiers in the Tri-training are separately trained by the three different subsets of randomly selected polarimetric features; and 3) the reliability of selected unlabeled samples is increased with the NMST method that utilizes the spatial neighborhood information between pixels.

The remainder of the article is organized as follows. Section II explains the proposed method. Section III shows the experimental design with real PolSAR data, and the experimental results and discussion are provided in Section IV. This is followed by the conclusions in Section V.

II. Proposed Method

Fig. 1 shows the framework of the complete process of the proposed method. As can be seen from the figure, the proposed method consists of two processes: training and classification. In the training process, there exist three steps. First, the polarimetric features are divided into three subsets with the same dimension. Each subset contains one-third of the polarimetric features, randomly selected. These three subsets are used to separately train the three different base classifiers that are depicted by three different colors in Fig. 1. Second, the trained classifiers are combined with the NMST method to test unlabeled samples in each data set. Finally, when the tested results of unlabeled samples obtained from one classifier are verified by the other two classifiers, these tested samples are labeled and added into the training set. This step is marked by two colors in one cell, as shown in Fig. 1. In this article, the number of iterations t is used to control the training process. In the classification process, the three different trained classifiers are employed to analyze all unlabeled samples and determine their final labels.

A. Polarimetric Features

Two sets of polarimetric features are used in this article. The first set of the polarimetric features is obtained from the coherency matrix. In general, each pixel in PolSAR data can be expressed in the form of the coherency matrix T as follows:

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

Since T is a complex conjugate matrix and the diagonal elements are real numbers, the first set of the polarimetric features is a 9-D vector as shown in the following:

$$\upsilon 1 = [T_{11}, T_{22}, T_{33}, \mathrm{Re}(T_{12}), \mathrm{Im}(T_{12}), \mathrm{Re}(T_{13}), \mathrm{Im}(T_{13}), \mathrm{Re}(T_{23}), \mathrm{Im}(T_{23})] \tag{2}$$

where Re( ) and Im( ) represent the real and imaginary parts of a complex number, respectively. Here, we chose the coherency matrix rather than the scattering matrix because the coherence matrix is a derivative of the scattering matrix and contains both the fundamental scattering characteristics and the second-order statistics of the scattering matrix [32].

The second set of polarimetric features is extracted from various target decompositions, including the Pauli [4], H/A/Alpha [5], Freeman [6], Krogager [10], Huynen [11], Touzi decomposition [12], Yamaguchi four-component decomposition [13], and B&F decomposition [14]. Thirty-six polarimetric features depicting different scattering mechanisms are listed in Table I and represented by a 36-D vector $\upsilon 2$ as follows:

$$\upsilon 2 = [|a|^2, |b|^2, |c|^2, H, a, A, \lambda_1, \lambda_2, \lambda_3, P_s, P_d, P_v, |K_s|^2, |K_d|^2, |K_h|^2, A_0, B_0 - B, B_0 + B, C, D, E, F, G, H_0, \alpha_s, \phi_s, \tau_s, \Psi, P_{y_s}, P_{y_d}, P_{y_v}, P_{y_h}, P_{bf_s}, P_{bf_d}, P_{bf_v}, P_{bfh}]. \tag{3}$$

Combining v1 and v2 yields a 45-D PolSAR feature vector $\upsilon = [\upsilon 1\ \upsilon 2]$ that is used in this article. It should be mentioned that the vector of the 45 polarimetric features may be extended as new polarimetric features are introduced from target decompositions. Also, some of those features mentioned above may be exchanged with other features that are not selected in this article, as discussed in Section III-B.

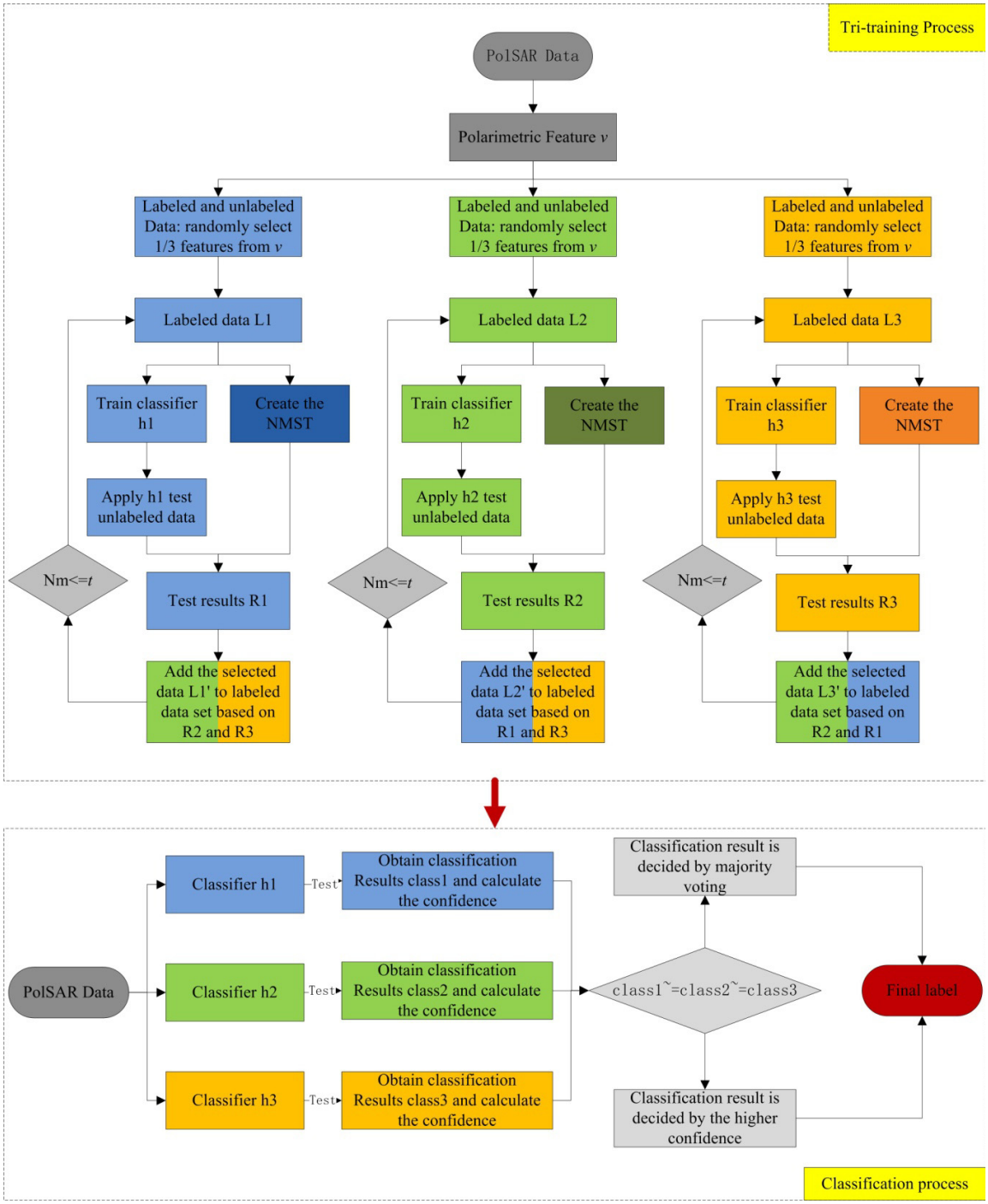


Fig. 1. Framework of the proposed method.

B. Neighborhood Minimum Spanning Tree

It is commonly assumed that pixels with similar characteristics can be assigned to the same category in image classification. In order to select highly reliable unlabeled samples, the NMST that is based on the Prim algorithm is adopted in this article. The NMST is an unsupervised clustering algorithm and uses both the coherency matrix and the spatial information, which is different from other existing clustering methods. The NMST used in this article is to select data with the features like a labeled sample. Herein, only a brief description of the NMST is given; readers interested in a more detailed account are referred to [30] and [31].

Let $G = (V, B, W)G = (V, B, W)$ be a weight-connected undirected graph, where V is the vertices in the graphic, B represents all the possible paths, and W is the collection of the weight of each path in G. A spanning tree, $g = (V, b, W)$, $g = (V, b, W)$ is a subset that contains all the vertices in G and only has one single path b between any two vertices. There may exist several spanning trees in one graph because of different paths of traversing vertices. As for a weight-connected graph, the MST refers to the spanning tree with a minimum total weight of paths.

When applying the MST to a PolSAR image, one needs to build an undirected weighted graph that contains all the

TABLE I
POLARIMETRIC FEATURES

Method	Dimensions	Polarimetric features
Pauli[4]	3	$\|a\|^2$: Single scattering coefficient, $\|b\|^2$: double scattering coefficient, and $\|c\|^2$: $\pi/4$ double scattering coefficient
H/A/Alpha[5]	6	H: Entropy, a: Alpha, A: Anisotropy and λ_1, λ_2, λ_3: three eigenvalues
Freeman[6]	3	P_s: Surface scattering power, P_d: double-bounce scattering power, and P_v: volume scattering power
Krogager[10]	3	$\|K_s\|^2$: Sphere scattering, $\|K_d\|^2$: Diplane scattering, and $\|K_h\|^2$: Helix scattering
Huynen[11]	9	A_0: scattering power of target symmetry; B_0: scattering power of target structure B: scattering power of other types B_0-B: scattering power of target non-symmetry; B_0+B: scattering power of target irregularity; C: global shape; D: local shape; E: local twist or surface torsion; F: global twist or target helicity; G: local coupling; H_0: global coupling due to target orientation;
Touzi[12]	4	α_s: Symmetric scattering type magnitude, ϕ_s: Symmetric scattering type phase, τ_s: helicity, and ψ:target orientation angle
Yamaguchi[13]	4	P_{ys} : Surface scattering power, P_{yd} : double-bounce scattering power, P_{yv} : volume scattering power, and P_{yh} : Helix scattering power
B&F[14]	4	P_{bfs} : Surface scattering power, P_{bfd} : double-bounce scattering power, P_{bfv} : volume scattering power, and P_{bfh} : Helix scattering power

pixels in the PolSAR image. The weight W for each path, which is also called the feature distance between two pixels, is computed as follows:

$$W_{i,j} = \frac{1}{2}\left\{ \ln(|T_i|) + \ln(|T_j|) + \mathrm{Tr}\left(T_i^{-1}T_j + T_j^{-1}T_i\right)\right\} \quad (4)$$

where T is the coherency matrix and $Tr(\ )$ is the trace of a matrix. It should be mentioned that this construction process requires a large amount of computation. Fortunately, Wu *et al.* [33] showed that for the PolSAR image, adjacent pixels are more likely to have the same label as the central pixel. Therefore, to reduce the computation and increase the reliability of selected vertices, we adopt an improved NMST method based on the Prim algorithm [30]. Fig. 2 is a schematic example showing the flowchart of the NMST. The blue point in Fig. 2(a) is the initial vertex V_0, while the green points represent the neighboring vertices (paths) of V_0. The numeric value at each neighboring point (path) denotes the distance (the weight W) between V_0 and that point. Select the point whose distance is the shortest, like "1," for example, in Fig. 2(a). This point is added to the vertex set V, and a new vertex is named V_1 [see Fig. 2(b)]. Next, select another point, marked V_2, nearest the new vertex V_1 in its neighborhood, as shown in Fig. 2(c). Repeat this process, until enough vertices are selected. A brief outline of the NMST procedure is given as follows.

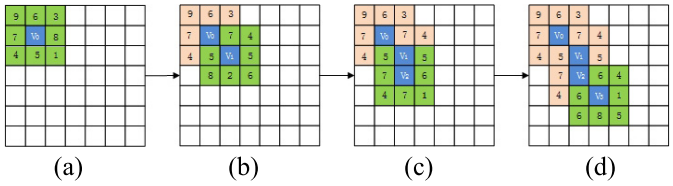


Fig. 2. Schematic example showing the flowchart of the NMST. (a) Initial vertex V_0, and (b), (c), and (d), respectively, correspond to the first, second, and third selected vertices (V_1, V_2, and V_3) with the NMST. The numeric value at each cell denotes the distance (the weight W) between the selected vertices (blue) and their neighboring vertices (green).

Step 1: Select the training samples as vertices V, and compute the set W for each vertex with its eight neighborhoods B.

Step 2: Select the nearest neighbor b of the eight neighborhoods of each training sample, and add it to the vertex set V.

Repeat steps 1 and 2 until enough vertices are selected.

C. Sample Selection and Training Process

To increase the reliability of selected unlabeled samples, a help-learning algorithm that is based on the NMST is adopted. The detailed algorithm is described as follows.

1) Define three different base classifiers (h1: SVM, h2: KNN, and h3: DT) and set their initial labeled samples $L = L_1 = L_2 = L_3$.
2) Use L_1, L_2, and L_3 to separately train the classifiers $h1$, $h2$, and $h3$.
3) Use the trained classifiers $h1$, $h2$, and $h3$ to separately test unlabeled samples and obtain three different sets of test results Test_1, Test_2, and Test_3.
4) Set the vertices of the labeled samples L as the initial vertex sets V_i $(i = 1, 2, 3)$ for each classifier. Use the NMST to label certain unlabeled samples and obtain the results $\text{Tree}_i (i = 1, 2, 3)$, respectively.
5) Select sample sets $R_i (i = 1, 2, 3)$, which have the same labels between Test_i and Tree_i, that is, $R_i = \{\text{Test}_i = \text{Tree}_i\}$.
6) Select a sample set $L1'$, which has the same label between R_2 and R_3, that is, $L1' = \{R_2 = R_3\}$. In the same way, select two sample sets $L2' = \{R_1 = R_3\}$ and $L3' = \{R_1 = R_2\}$. Then, update the labeled sample sets $L1 = L1 \cup L1'$, $L2 = L2 \cup L2'$, and $L3 = L3 \cup L3'$.
7) Update the vertex sets with $V1 = L1$, $V2 = L2$, and $V3 = L3$.
8) Repeat (from step 2 to step 7) t times until a satisfactory result is obtained.

D. Classification Process

$L1$ is employed to train the classifier $h1$. Then, the trained $h1$ is used to test all unlabeled samples, and class1 is the test result. The confidence $P_1(m) = u_1(m)/8$ is calculated for each pixel m in class1. Here, $u_1(m)$ is the number of neighboring pixels that have the same label as the central pixel m.

Similarly, $L2$ and $L3$ are employed to train the classifiers $h2$ and $h3$, respectively. Then, the trained $h2$ and $h3$ are used to separately test all unlabeled samples, and their test results are given in class2 and class3, respectively. The confidences $P_2(m) = u_2(m)/8$ and $P_3(m) = u_3(m)/8$ are calculated for each pixel m in class2 and class3.

The test results are compared for each pixel among class1, class2, and class3. If class1 $\neq$ class2 $\neq$ class3, the final class label of this pixel is decided by the highest confidence; otherwise, the final class label of the pixel is decided by majority of votes.

Table II shows the pseudocode describing the proposed method. In addition, it should be mentioned that as the spread of landcover type across the image is changed, the NMST method prevents to pick points across the boundary of the type spread because the method finds those points of the minimum feature distance with the spanning tree at each iteration. Also, there exists the extreme case that when all the base classifiers have misclassified, the final classification result will be wrong. To reduce the occurrence of such a situation, the base classifiers should be well-trained with reliable training sets. Thus, selecting reliable unlabeled samples to expand the training samples becomes essential.

TABLE II
PSEUDOCODE DESCRIBING THE PROPOSED METHOD

The proposed method
Input: L: Original Labeled sample set
U: Unlabeled sample set
v: 33-dimensional features
$h_i \quad i \in \{1..3\}$: three different classifiers
For $i \in \{1..3\}$ **do**
$v_i \leftarrow$ randomly selected from v
$L_i = L \leftarrow$ Original labeled samples
$V_i = L \leftarrow$ Original vertices
End of for
Repeat ***t*** **times**
For $i \in \{1..3\}$ **do**
$h_i \leftarrow$trained by v_i and L_i
$Test_i \leftarrow$results of U tested by h_i
$Tree_i \leftarrow$produced by V_i with NMST
If $(Test_i = Tree_i)$
$R_i \leftarrow Test_i \cap Tree_i$
End of if
For $j,k,m \in \{1..3\}$ **do**
$L_j^{'} \leftarrow R_k \cap R_m \quad (j \neq k \neq m)$
$L_j \leftarrow L_j \cup L_j^{'}$
$V_j \leftarrow L_j$
End of for
End of repeat

Output: Label : final label
For $i \in \{1..3\}$ **do**
$class_i \leftarrow$ Classified by h_i with U
$P_i \leftarrow u(m)/8$ % $u(m)$ is the number of neighboring pixels that have the same % label as the central pixel m.
If $(class_1 \neq class_2 = class_3)$ **do**
Label$\leftarrow \arg\max(P_i)$
else
Label$\leftarrow$majority voting ($class_i$)
End of if
End of for

III. EXPERIMENTAL DESIGN

In this article, experiments are carried out with the MATLAB 2014a on a desktop computer that runs Windows 7 operating system with an Intel(R) Core(TM) CPU processor (3.20 GHz) and 4 G memory. The three different base classifiers of the improved Tri-training in the proposed method are the DT, SVM, and KNN, respectively. The overall accuracy (OA) and kappa statistics (Kappa) of the classification

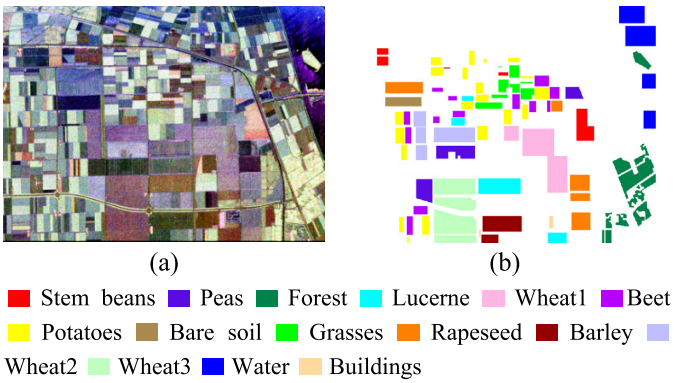


Fig. 3. Flevoland I AIRSAR data set acquired in Flevoland, August 1989 [32]. (a) Pauli RGB image. (b) Ground truth image.

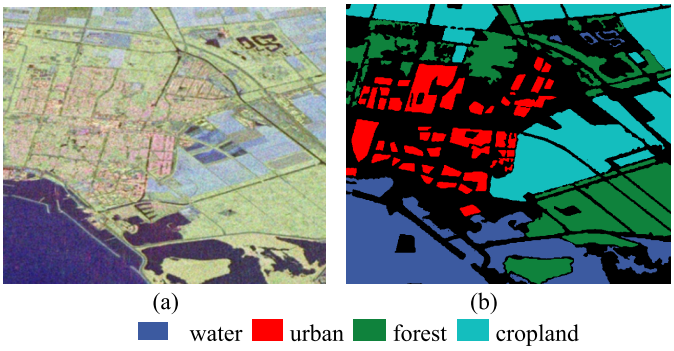


Fig. 4. Flevoland II RADARSAT-2 data acquired in Flevoland, April 2008 [32]. (a) Pauli RGB image. (b) Ground truth image.

results are adopted to characterize the effectiveness of the proposed classification method. Two real PolSAR data sets are used as shown below.

A. Data Sets

The first data set (Flevoland I) is a L-band four-look PolSAR data acquired by the NASA/JPL AIRSAR system in Flevoland, the Netherlands, August 1989 [32]. Fig. 3 shows both the Pauli RGB image and ground truth image of this data. The resolution of the Pauli RGB image is 750×1024 pixels, while the size of the ground truth data is around 167 712 pixels. The pixel size of the Pauli RGB image is 6.6 m in the slant range direction and 12.1 m in the azimuth direction. There are 15 terrain types, also called categories, marked in the ground truth, and most of which are related to agriculture, including water, rapeseed, grasses, bare soil, potatoes, beet, wheat, lucerne, forest, peas, and stem beans.

The second data set (Flevoland II) is a C-band single-look fully PolSAR data acquired by the RADARSAT-2 system in Flevoland, the Netherlands, April 2008 [32]. Fig. 4 shows both the Pauli RGB image and ground truth image of the data. The resolution of the Pauli RGB image is 1400×1200 pixels, while the size of the ground truth data is around 947 262 pixels. The pixel size of the Pauli RGB image is 10 m in the slant range direction and 5 m in the azimuth direction. As shown in the ground truth image [Fig. 4(b)], there are four different terrain types, including water, urban, forest, and cropland.

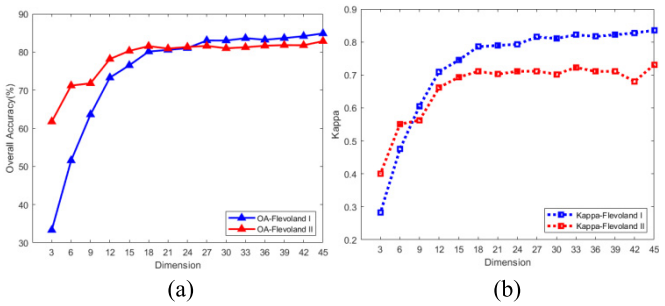


Fig. 5. (a) OA and (b) Kappa of two PolSAR data sets varying with the dimension of polarimetric feature vector.

B. Key Parameters in the Proposed Method

The performance of the proposed method is dominated by four key parameters: the dimension of polarimetric feature vector, the number of vertices in the NMST, and the number of iterations, and the number of labeled samples in the training process. The effects of those four parameters on the classification results of the proposed method are investigated experimentally with the two abovementioned real PolSAR data sets. In these experiments, if not specified, the dimension of the polarimetric feature vector is set to 33 and 10 labeled samples from each terrain type, i.e., 150 labeled samples for Flevoland I (15 types) and 40 for Flevoland II (four types), are used in the initial training data sets. The number of the initial vertices in NMST is 10 and 8 iterations are used in the training process.

Fig. 5 shows the OA and Kappa of the classification results given by the proposed method varying with the dimension of the polarimetric feature vector. As can be seen from Fig. 5(a), the OA values for both data sets rapidly increase as the vector dimension increases from 3 to 18. Fig. 5 also shows that both the OA and Kappa of Flevoland I are lower than that of Flevoland II when the vector dimension of the polarimetric features is less than 9. As the dimension of the polarimetric feature vector increases, the increase rates of the OA and Kappa of Flevoland I are greater than that of Flevoland II. These differences might be caused by various factors, including sensor type, wavelength of radar waves, image resolution, and data acquisition time as well. Also, the number of categories used in the classification may have markedly effects. It is conjectured that the different rates of the increase of both OA and Kappa between two data sets shown in Fig. 5 are mainly due to different radar waves and the different amount of terrain types in the ground truth images used in these two data sets. Flevoland I data set contains 15 types, while Flevoland II data set has only 4 types. As a rule of thumb, the more categories there are, the more difficult a classification will be in machine learning [33]. This is especially perceivable for a low-dimension vector of polarimetric features because of lack of enough information in the training process. However, when the dimension of the polarimetric feature vector exceeds 24, the OA and Kappa of the Flevoland I data set are greater than that of the Flevoland II data set because enough features provide the

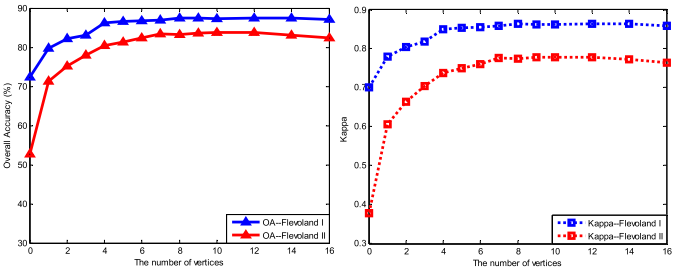


Fig. 6. Effects of the number of vertices on the (a) OA and (b) Kappa of two data sets given by the proposed method.

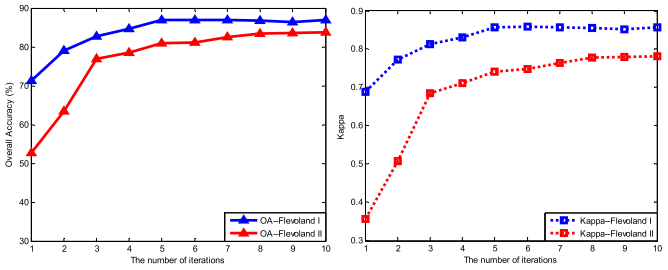


Fig. 7. (a) OA and (b) Kappa of the two PolSAR data sets vary with the number of iterations in the training process.

possibility of a fine classification, while the contribution to coarse-grained classification is indistinct. When the vector dimension is greater than 27, both OA curves become nearly flat. Similar trends for the Kappa are shown in Fig. 5(b), indicating that the improvement of the classification results becomes insignificant when the dimension of the polarimetric feature vector is beyond 27 because of the redundancy of selected polarimetric features [34]. It should be mentioned that the values of OA and Kappa for each data point in Fig. 5 are computed with certain polarimetric features that are selected from the pool of the 45 polarimetric features. This indicates that the classification results are affected by the dimension of the feature vector and that those polarimetric features are of equal importance to the classification. Thus, a 33-dimension vector is used in the late sections of this article.

Fig. 6 shows the OA and Kappa of the classification results of two real PolSAR data sets given by the proposed method as a function of the number of vertices used in the NMST. In the figure, the left data point (the number of vertices is equal to zero) of each curve corresponds to the case that the NMST is not used in the training process. The OA and Kappa at this point have the lowest values for each curve. Data points with a vertex number greater than zero refer to cases where the proposed NMST is applied. As can be seen from Fig. 6, the OA and Kappa of the classification results increase rapidly with increasing vertex number until 8. As the number of vertices is greater than 8, both curves become nearly flat. When the number of vertices exceeds 14, the OA and Kappa slightly decrease with increasing vertex number. These slight decreases are likely due to the fact that the similarity between the root nodes and their subnodes might be weakened as the number of selected vertices increases. This results in certain improper unlabeled samples selected. Thus, the optimum number of vertices per category is between 8 and 10 in the proposed method. Herein, we would like to point out that the optimum number of vertices is only for the two PolarSAR data mentioned above. It might vary a little bit when the proposed method is applied to another data set because there will be problematic if the initial labeled pixels fail to include all the polarimetric features for each category. Therefore, a pretest like Fig. 6 should be given for new data.

On the other hand, as the number of labeled samples increases the label error increases as well, which greatly reduces the accuracy of classification and may even cause

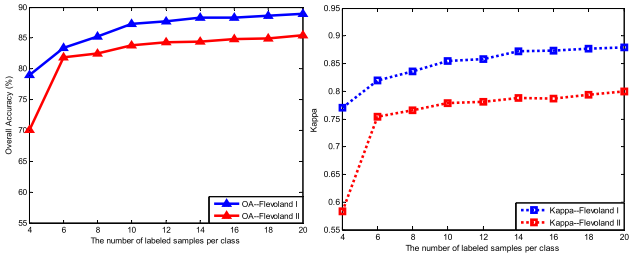


Fig. 8. (a) OA and (b) Kappa statistics of the classification results of the two PolSAR data sets vary with the number of labeled samples for each terrain type used in the proposed method.

the entire classification to collapse. Thus, selecting the most similar samples and improving the confidence of the selected unlabeled samples are important to improve the classification. Since the proposed method considers both the classification accuracy and the diversities of sample training and selection diversity, the strict sample expansion strategy like the NMST can be used.

Fig. 7 shows the OA and Kappa of the classification results of two PolSAR data sets varying with the number of iterations in the training process. As can be seen from Fig. 7, both OA and Kappa increase rapidly as the number of iterations increases until the iteration number reaches 4. This indicates that the performance of the classifiers is improved significantly within the first several iterations of the training process. When the number of iterations exceeds 4, slight increases of both the OA and Kappa occur. These slight increases are likely due to the fact that both proper and improper unlabeled samples might be added into the training process, resulting in a slow improvement in the classification accuracy. Therefore, eight iterations are suggested in the proposed method for efficiency. It should be mentioned after 10 iterations with 10 labeled samples in the initial training sets, the final number of the training samples can reach more 100 000 for each classifier.

Fig. 8 shows the OA and Kappa of the classification results of two PolSAR data sets as a function of the number of labeled samples (N_l) per category used in the proposed method. The total number of labeled samples for Flevoland I is $15N_l$ (15 categories) and $4N_l$ for Flevoland II (four categories). As can be seen from Fig. 8, the OA and Kappa are 78.96%, and 0.7709, respectively, when four labeled samples per category

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

Data / region	With filter	Original
Stembeans	96.40	85.14
Rapeseed	81.95	37.53
Bare soil	99.31	60.23
Potatoes	65.31	56.07
Beet	93.45	10.98
Wheat 2	72.48	9.75
Peas	92.29	65.96
Wheat 3	90.50	0.44
Lucerne	95.07	37.80
Barley	95.64	81.38
Wheat1	87.09	1.53
Grasses	72.13	0.07
Forest	90.32	7.30
Water	96.30	78.75
Building	76.87	73.06
OA	87.01	38.47
Kappa	0.8542	0.3424

(the left point in the figure) are used for the data set of the Flevoland I. The values of the OA and Kappa become 87.31% and 0.8577, respectively, as the number of labeled samples is ten per category. When the number of labeled samples is 20 per category, the OA is 88.78% and the kappa statistic is 0.8788. The OA increases 8.35% as the number of labeled samples per category used increases from 4 to 10, but only increases 1.47% as the number increases from 10 to 20. This is mainly because the fewer the labeled samples used in the training process, the greater the role of the semi-supervised method, resulting in greater changes in the classification results. Thus, the proposed method is effective for the PolSAR data where labeled samples are few. Similar conclusions are obtained from the Flevoland II data set. For efficiency, the optimum number of labeled samples per category is between 10 and 12 in the training process of the proposal method.

C. Effects of Speckle Filtering

To investigate the effects of speckle noise suppression on the classification results, the method proposed in this article is applied to the original PolSAR data of Flevoland I, without the speckle noise suppression. The results are given in Fig. 9 and Table III. For comparison, the classification results of the same data despeckled by the Lee filter [36] with a window size of 7×7 are also shown in the figure and table. These results are obtained with ten labeled samples per terrain type, i.e., a total of 150 labeled samples, in the initial training process.

As can be seen from Fig. 9 and Table III, the classification results are very poor when the algorithm is directly applied to the PolSAR data without the speckled noise suppression, indicating that the proposed method is extremely sensitive to noise.

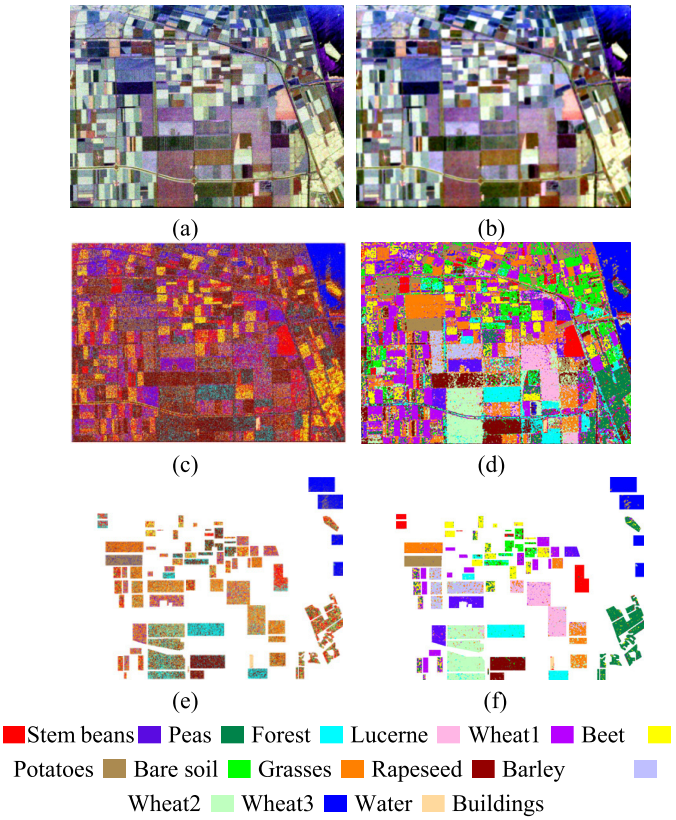


Fig. 9. Classification results of the Flevoland I data acquired by AIRSAR. (a) Pauli RGB image without filter [Fig. 3(a)]. (b) Denoised Pauli RGB image. (c) Classification results from the image data without filter. (d) Classification results of the denoised data. (e) Masked results corresponding to the ground truth of (c). (f) Masked results according to the ground truth of (d).

This poor performance is mainly due to the reason that the proposed method is based on 3-week classification training strategies to achieve the semi-supervised learning, a very important part of which is how to choose highly reliable unlabeled samples and add the correct pseudo labels into the training set. When this method is applied to an image without despeckling, due to the influence of speckle noise, certain deviation might occur in the selection of samples and the determination of pseudo labels, thereby introducing incorrect labels in the learning process. The introduction of more incorrect labels will reduce the effectiveness of this learning mechanism, especially when the number of labeled samples is small. Similar conclusions are obtained from the Flevoland II data set, but not shown herein.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the classification results of the two above-mentioned real PolSAR data sets given by the proposed method are presented and verified by the ground truth measurements of those PolSAR data sets. Before executing the proposed method, the Lee filter with a window size of 7×7 is applied to reduce the effects of speckle noise on the polarimetric features from the PolSAR images [35]. Thirty-three polarimetric features selected from the 45-D feature vector are used in the present experiment. These 33 polarimetric

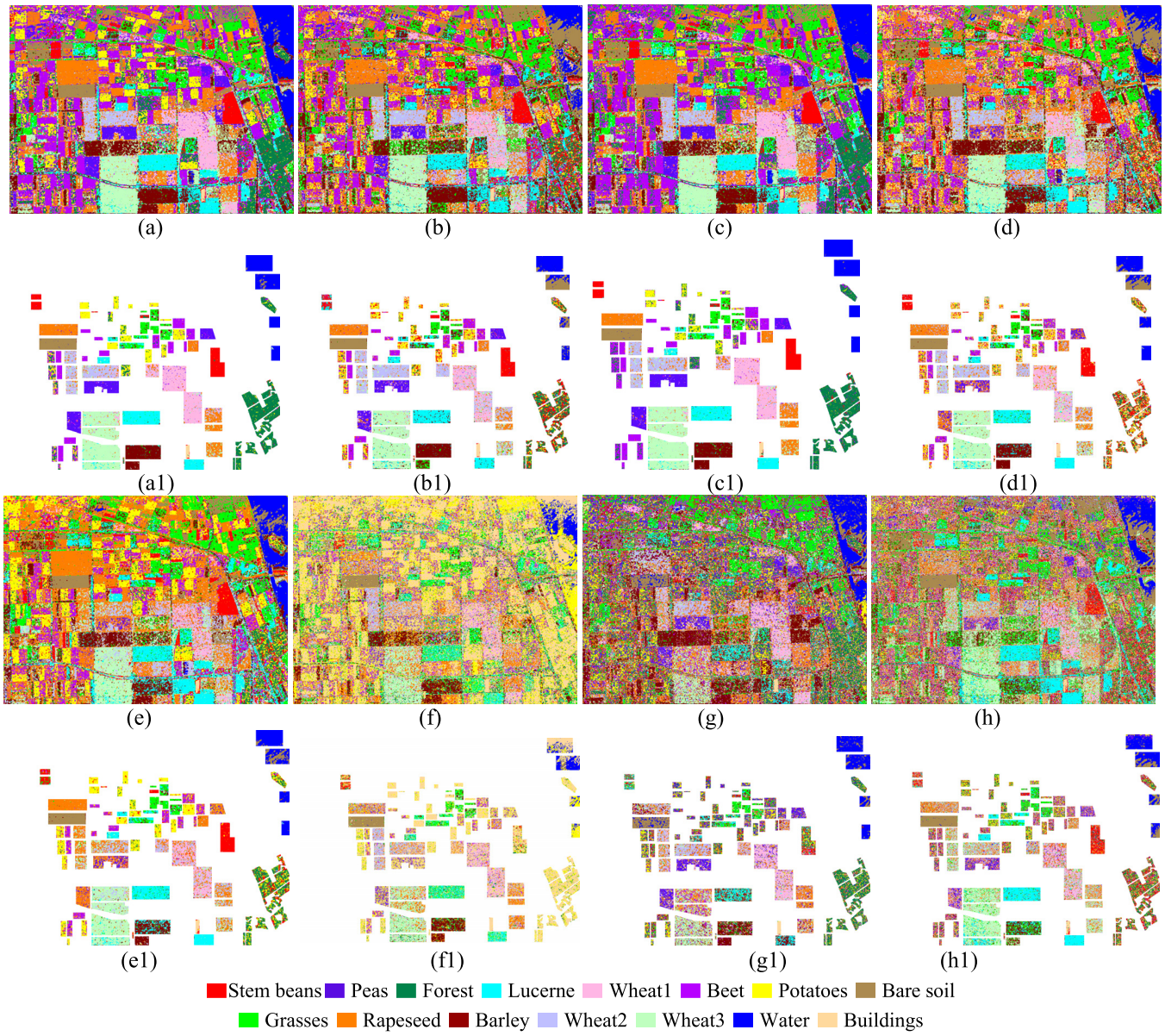


Fig. 10. Classification results of the Flevoland I data acquired by AIRSAR. (a) Proposed method. (b) Proposed + Neighbor. (c) Tri-training + NMST. (d) Tri-training. (e) Tri-training + Neighbor. (f) SVM. (g) DT. (h) KNN. (a1), (b1), (c1), (d1), (e1) (f1), (g1), and (h1) Masked results according to the ground truth of (a), (b), (c), (d), (e), (f), (g), and (h), respectively.

features are randomly divided into three subsets with 11 polarimetric features for each subset. The three subsets are used to separately train the three different base classifiers (DT, SVM, and KNN) of the improved Tri-training. We use ten labeled samples per category. There are 15 categories for Flevoland I data set; therefore, the total number of the labeled samples used for Flevoland I data set is 150. In contrast, 40 labeled samples (four categories) are used for Flevoland II data set. The experimental results are also compared with that obtained with a number of existing classification methods. For a fair comparison, the same labeled samples are used as the initial training set for all methods. To eliminate the instability effects on the results of certain methods, the experiments are repeated ten times and the average values are used as the final classification results.

A. Classification Results of the Flevoland I Acquired by AIRSAR in 1989

Fig. 10 shows the classification maps of the Flevoland I data set with different classification methods. The results of the proposed method are given in Fig. 10(a), while Fig. 10(b) shows the classification results of a procedure similar to the proposed method, but with the neighbor sample expanding strategy (a certain amount of pixels in the surrounding local area of a labeled pixel are added into the training data set), denoted as Proposed plus Neighbor (Proposed + Neighbor). Fig. 10(c) shows the classification results of the method of the Tri-training plus the NMST (Tri-training + NMST), while Fig. 10(d) shows the classification results of the Tri-training. Fig. 10(e) shows the results of the method of the Tri-training

TABLE IV
CLASSIFICATION ACCURACY (%) OF THE FLEVOLAND I DATA SET ACQUIRED BY AIRSAR

Method / Region	Proposed Method	Proposed+ Neighbor	Tri-training +NMST	Tri-training + Neighbor	Tri-training	SVM	DT	KNN
Stem beans	96.40	90.44	**97.14**	91.56	90.47	16.96	30.47	52.49
Rapeseed	81.95	49.29	**83.85**	60.61	70.17	45.42	40.14	43.39
Bare soil	**99.31**	99.45	99.20	99.22	98.38	93.81	79.86	98.02
Potatoes	65.31	61.96	33.94	33.95	**85.21**	26.81	30.07	34.59
Beet	93.45	66.46	**94.28**	64.41	65.73	29.11	5.42	16.93
Wheat 2	72.48	81.12	**75.66**	71.59	26.60	49.20	52.24	37.76
Peas	**92.29**	74.24	91.15	40.04	22.80	24.19	64.52	37.18
Wheat 3	90.50	84.09	**92.76**	88.69	80.54	60.95	51.35	67.46
Lucerne	**95.07**	80.39	93.17	89.62	91.53	45.78	45.34	89.97
Barley	**95.64**	90.13	94.73	87.00	69.22	72.47	67.58	41.28
Wheat1	**87.09**	60.99	81.62	49.86	44.64	45.52	52.24	36.79
Grasses	72.13	53.30	**72.51**	57.48	63.49	52.00	50.77	51.80
Forest	**90.32**	40.06	88.28	24.28	49.27	5.19	56.26	33.93
Water	96.30	74.09	**99.63**	70.56	81.54	34.56	84.46	64.72
Building	76.87	81.36	70.88	88.44	67.48	**93.88**	86.26	67.21
OA	**87.01**	72.82	84.59	67.82	67.14	46.93	53.13	51.57
Kappa	**0.8542**	0.7056	0.8323	0.6525	0.6426	0.4294	0.4923	0.4754

The bold number indicates the best result in each row.

plus the Neighbor (Tri-training + Neighbor). Fig. 10(f)–(h) shows the classification results of the SVM, DT, and KNN, respectively. Table IV shows the classification accuracies of each terrain type, also called category, with different classification methods and the OAs and Kappa for this entire data set. The results shown herein are obtained with ten randomly selected labeled samples per terrain type (150 labeled samples for 15 categories) in the training process.

As can be seen from Table IV, the classification accuracies of nine categories in the proposed method are higher than 90%, and the lowest classification accuracy of one category is 65.31%. In contrast, only eight categories in the Tri-training + NMST method can achieve classification accuracies above 90%, while the lowest classification accuracy of one category is only 33.94%. In the Proposed + Neighbor method, the classification accuracies of only three categories are higher than 90%, and the classification accuracies of two categories are below 50%. In the Tri-training method, the classification accuracies of three categories are higher than 90%, and the classification accuracies of four categories are lower than 50%. In the Tri-training + NMST, there are two categories that reach the accuracy of 90%, and four categories are below 50%. In the SVM method, the classification accuracies of two categories are higher than 90%, while the classification accuracies of ten categories are below 50%. In the DT method, there is no category with accuracy higher than 90%, and the classification accuracies of five categories are less than 50%. In the KNN method, the classification accuracy of only one category is higher than 90%, and the classification accuracies of eight categories are under 50%. In addition, the Kappa of the proposed method is 0.8542, which is higher than other methods.

By comparing the classification accuracies of all the categories shown in Fig. 10 and Table IV, one can see that the classification accuracies of the proposed method in peas, lucerne, barley, wheat, and forest are much higher than other classification methods. The data set is an L-band PolSAR data and the penetrability of L-band radar waves is stronger than that of C-band radar waves [37]. Thus, the backscattering signal from a crop canopy field may consist of different types of contributions, such as volume scattering from plants, surface scattering from the underlying soil, and plant–soil multiple-bounce scattering, depending on the growth period of the plants [38], [39]. The differences in scattering mechanisms of different plants are often ultimately reflected in the radar wavelength, and the overall roughness formed by the plants and others. The high classification accuracies of peas, lucerne, barley, wheat, and forest with this AIRSAR data indicate that the proposed method can significantly improve the classification of the abovementioned regions where a complex backscattering occurs.

On the other hand, the L-band AIRSAR data has a low resolution and is unable to accurately capture the subtle effects of plant size and features. For example, the accuracies of both potato and beet in the comparative methods vary considerably because they are misclassified in a number of methods. Both potatoes and beets are root plants and their leaves are very similar when the data was acquired in August. The leaves of potato in this growth period are oval (10–20 cm long, and about 3 cm wide), while the leaves of beet are also oval (10–20 cm long and 10–15 cm wide). As a result, these two fields have very similar roughness under the observation of AIRSAR, making them difficult to distinguish. From Fig.10(d), one can see that many beet areas are

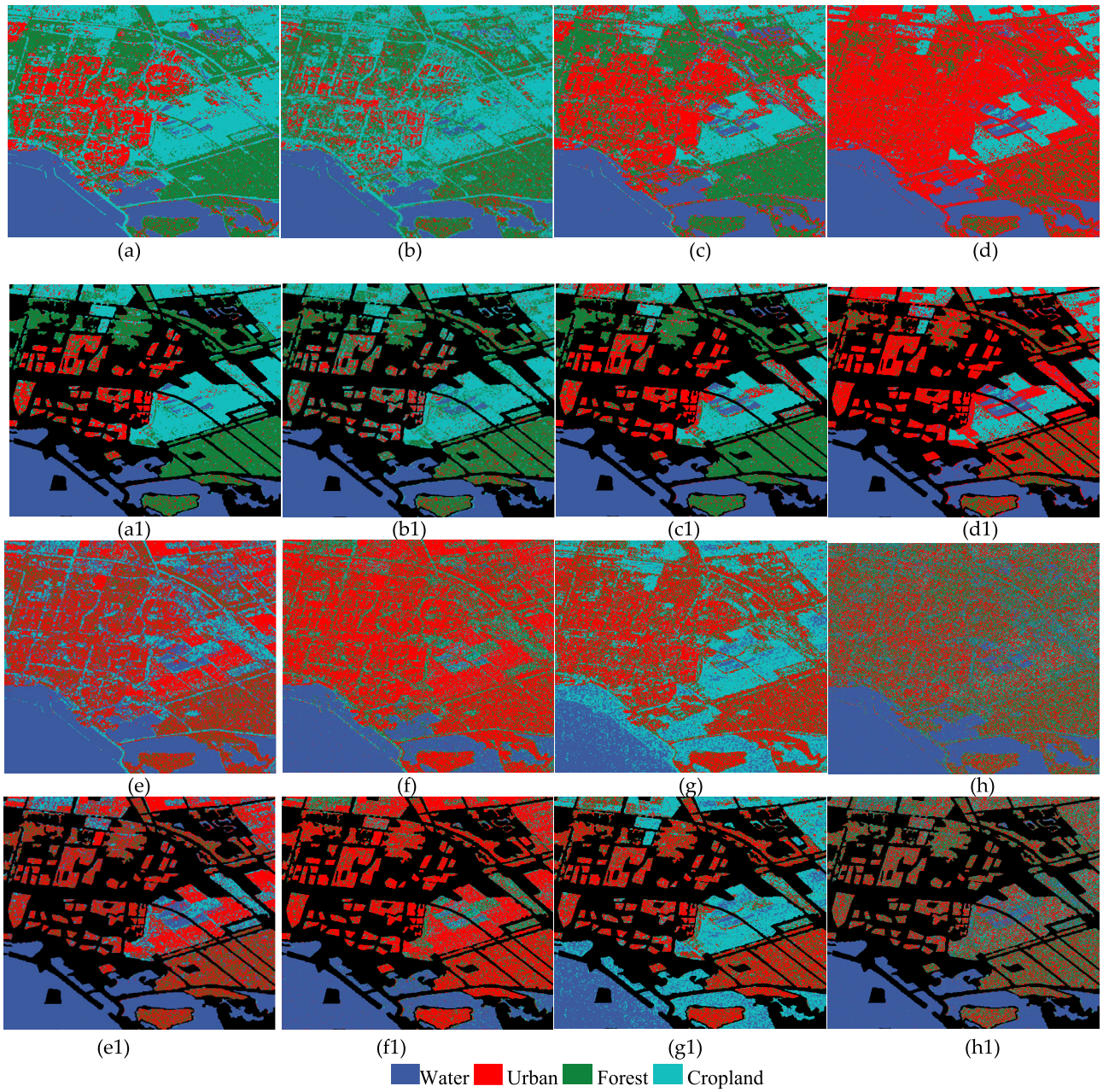


Fig. 11. Classification results of the Flevoland II data set acquired by RADARSAT-2. (a) Proposed method. (b) Proposed + Neighbor. (c) Tri-training + NMST. (d) Tri-training. (e) Tri-training + Neighbor. (f) SVM. (g) DT. (h) KNN. (a1), (b1), (c1), (d1), (e1) (f1), (g1), and (h1) Masked results according to the ground truth of (a), (b), (c), (d), (e), (f), (g), and (h), respectively.

misclassified as potato areas with the traditional Tri-training. However, in Fig. 10(a), the beet areas and potato areas have been divided well. The classification accuracy of beet with the proposed method is 93.45%, which is 27.72% higher than that of the traditional Tri-training method. Also, the sum of the accuracies of potatoes and beet in the proposed method is 7.82% higher than that given by the traditional Tri-training. In addition, the Kappa of the proposed method is also higher than the Tri-training + NMST and traditional Tri-training methods, respectively.

From Table IV and Fig. 10(b)–(d), one can see that the OA of the proposed method is 14.19% higher than the proposed plus neighbor, while the OA of the Tri-training plus NMST is 16.77% higher than the Tri-training + Neighbor. Those substantial improvements in the OA of the classification show the effectiveness of the selection strategy with the NMST.

Table IV also shows that the OA of the proposed method is 40.08%, 33.88%, and 35.44% higher than SVM, DT, and KNN, respectively. As for Kappa, the proposed method is 0.4248, 0.3619, and 0.3788 greater than SVM, DT, and KNN, respectively. Through analysis of the classification accuracies of different crops, one can see that the proposed method improves the representation of the model and performs well in distinguishing categories with similar characteristics.

TABLE V
CLASSIFICATION ACCURACY (%) AND KAPPA OF THE FLEVOLAND II DATA SET ACQUIRED BY RADARSAT-2

Method / Region	Proposed Method	Proposed+ Neighbor	Tri-training +NMST	Tri-training + Neighbor	Tri-training	SVM	DT	KNN
Urban	68.00	39.52	87.68	65.68	**88.85**	79.47	67.31	48.50
Water	**98.44**	97.25	98.38	98.39	97.30	91.57	62.62	96.19
Forest	**85.53**	75.08	77.42	32.66	39.48	24.34	32.64	44.52
Cropland	**83.17**	73.30	65.24	23.86	49.79	18.70	66.21	24.27
OA	**83.79**	67.90	82.18	55.15	68.86	53.52	57.20	53.37
Kappa	**0.7784**	0.4807	0.7631	0.2619	0.5591	0.4123	0.4329	0.3854

The bold number indicates the best result in each row.

B. Classification Results of the Flevoland II Acquired by RADASAT-2 in 2008

Fig. 11 shows the classification maps of the Flevoland II data set with different classification methods. The results of the proposed method are given in Fig. 11(a), while Fig. 11(b) shows the classification results of the procedure similar to the proposed method, but with the Neighbor sample expanding strategy, denoted as Proposed plus Neighbor (Proposed + Neighbor). Fig. 11(c) shows the classification of the method of the Tri-training plus the NMST (Tri-training + NMST), while Fig. 11(d) shows the classification of the Tri-training. Fig. 11(e) shows the classification of the Tri-training plus the Neighbor (Tri-training + Neighbor). Fig. 11(f)–(h) show the classifications of the SVM, DT, and KNN, respectively. Table V shows the classification accuracies of each category with different classification methods and the OA and Kappa for this data set. The results shown herein are obtained with 10 randomly selected labeled samples per category (40 labeled samples total for four categories) in the training process.

As can be seen from Table V and Fig. 11(a)–(d), the OA of the proposed method is 15.89% higher than the Proposed + Neighbor, while the OA of the Tri-training + NMST is 27.03% higher than the Tri-training + Neighbor. These substantial improvements in the OA of the classification show the effectiveness of the NMST for C-band data set as well.

Comparing Fig. 11(a) with Fig. 11(f)–(h), one can see that the classification results of the proposed method are much better than other methods. Table V shows that the classification accuracies of the water areas are exceptionally well (more than 90%) for all the methods, except DT. This is likely due to that the scattering type of water is mainly surface scattering, which is distinctly different from the scattering types of other areas. As for the forest areas and cropland areas, the classification accuracies of the proposed method are much higher than other methods. In contrast, the class accuracy of urban (red color) of the proposed method is only 68%, which is 20.85% lower than the classification accuracy of the traditional Tri-training method. This is mainly because the urban areas usually contain some trees and vegetation, which have similar scattering types to the forest and cropland areas; as a result, urban, forest, and cropland are difficult to differentiate from each other with the traditional methods. For instance, Fig. 11(d) shows that the forest areas are misclassified as urban areas with the Tri-training method, resulting in a high classification accuracy of the urban areas, as shown in Table V. However, the classification accuracy of the forest areas with the Tri-training method is only 39.48%, which is 49.05% lower than that of the proposed method because these three areas (forest, cropland, and urban) are separated well with the proposed method, as shown in Fig. 11(a) and Table V.

In summary, the proposed method is very effective for those regions where multiple scattering mechanisms are involved. One evidence is that the classification accuracies of cereal crops (Wheat1, Wheat2, Wheat3, Barley, and Grasses) for the Flevoland I data set with the proposed method are significantly improved in comparison with other methods. The Flevoland I data set is a L-band four-look PolSAR data acquired by the NASA/JPL system in August 1989. The L-band radar has stronger penetrability than C-band radar, so cereals in August in this region are generally considered to include a hybrid scattering mechanism composed of a two-layer medium (soil + stalks) before heading and a three-layer medium (soil + stalks + heads) after heading [39]. For the Flevoland II data set, the proposed method significantly improves the classification of Forest and Cropland. These two categories are generally considered to have certain mixed scattering mechanisms [40].

C. In Comparison With the Existing Methods in Terms of the Number of Labeled Samples

To test the proposed method and compare it with other methods, we experimentally study the effects of the number of labeled samples (N_l) on the classification results. Herein, only the results for the data set of Flevoland I are presented, while similar conclusions are obtained from the data set of Flevoland II. Fig. 12 shows the classification accuracies of different methods varying with the initial number of labeled samples used in the training process. In order to better show the effect of the number of labeled samples on the classification, the abscissa of Fig. 12 is expressed in logarithmic form. In Fig. 12, the color notation (N_l, OA) on each curve, corresponding to each method, represents a point that the

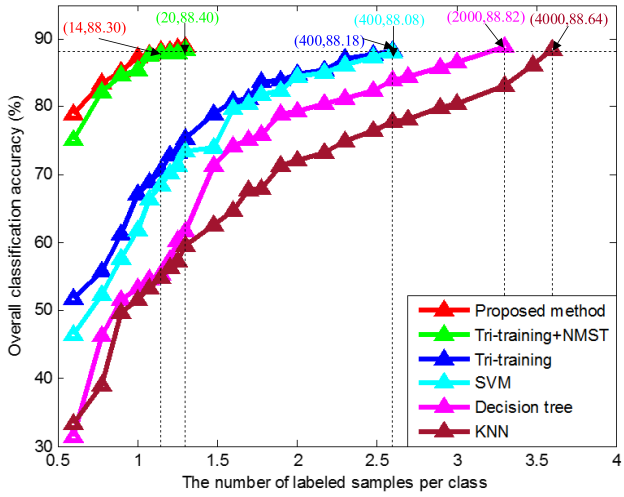


Fig. 12. Overall accuracies of different classification methods varying with the number of labeled samples for the Flevoland I data set. In (N_l, OA), N_l refers to the number of labeled samples and OA refers to classification accuracy.

number of labeled samples N_l per category (the total number of labeled samples of $15N_l$ for 15 categories) is required to reach the OA of about 88% for the method.

As can be seen from Fig. 12, in order to reach the classification accuracy of about 88%, the proposed method only needs 14 ($14 \approx 10^{1.146}$) labeled samples per category, while the Tri-training + NMST method needs 20 ($20 \approx 10^{1.301}$) labeled samples per category. In contrast, the traditional Tri-training method needs 400 ($400 \approx 10^{2.602}$) labeled samples per category, the SVM method needs 400, the DT method needs 2000 ($2000 \approx 10^{3.302}$), and the KNN method needs 4000 ($4000 \approx 10^{3.602}$). Note that when labeled samples are smaller than 20 per category, the classification accuracy of the Tri-training + NMST method is slightly lower than the proposed method, but is much higher than other methods. This indicates that the proposed sample selection method in the NMST is very effective. In other words, the selected unlabeled samples with the NMST improve the performances of both the base classifiers and the final classification processes.

Fig. 12 also shows that the classification accuracies of the proposed method and Tri-training + NMST method are nearly the same, as the number of the labeled samples per category exceeds 12. Even though the classification accuracies of the traditional methods (Tri-training, SVM, DT, and KNN) increase with increasing the number of the labeled samples, the rates of their increases with the number of labeled samples are much slower than that of the proposed method and the Tri-training plus NMST. This is mainly because the Tri-training, SVM, DT, and KNN require sufficient labeled samples as training samples. When labeled samples are few, the classification accuracies of these four methods are poor. For the proposed method and the Tri-training + NMST method, training samples gradually increase during the iteration of the training process, and the proposed selection method for unlabeled samples ensures the reliability of those samples added into the training sample set. The enlarged training samples provide sufficient information for the proposed method and Tri-training + NMST method.

V. Conclusion

This article presents an improved SSC method for PolSAR image, which combines the advantages of the co-training and Tri-training methods. For the co-training and Tri-training method, the diversities of both classifiers and features are important, and these diversities determine whether the co-training and tri-training methods can work effectively. Herein, we propose a new sample selection method to produce three different feature subsets to increase the diversity among the base classifiers. Then, an improved tri-training process is used to choose reliable samples from the selected sample set and improve the three base classifiers. In this process, in order to increase the reliability of the selected samples, we adopt the improved NMST method to select the unlabeled samples based on spatial neighborhood information between the pixels of PolSAR image. Finally, we apply these three base classifiers to classify the whole PolSAR image, and use major voting and confidence to decide the final class label.

The experiments with two real PolSAR data sets show that the proposed method is able to achieve satisfactory classification accuracy, especially when the training labeled samples are few. The unlabeled samples selected as training samples with our proposed method are reliable, and the performance of the classifiers is gradually improved by selecting unlabeled samples to expand the label samples set. As for these two data sets, not only is the overall classification accuracy of the proposed method higher than other existing methods, but also the accuracy on most individual classes. In addition, we analyze the influence of the initial number of labeled samples in the training process on the results of the classification.

Acknowledgment

The authors would like to thank anonymous reviewers for their constructive comments which were very helpful in improving the manuscript.

References

[1] F. Nunziata, M. Migliaccio, X. Li, and X. Ding, "Coastline extraction using dual-polarimetric COSMO-SkyMed PingPong mode SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 104–108, Jan. 2014.

[2] X. Ding and X. Li, "Shoreline movement monitoring based on SAR images in Shanghai, China," *Int. J. Remote Sens.*, vol. 35, nos. 11–12, pp. 3994–4008, Jun. 2014.

[3] A. Buono, F. Nunziata, M. Migliaccio, X. Yang, and X. Li, "Classification of the Yellow River Delta area using fully polarimetric SAR measurements," *Int. J. Remote Sens.*, vol. 38, no. 23, pp. 6714–6734, Dec. 2017.

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

[5] 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.

[6] 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.

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

[8] B. Ren, B. Hou, J. Zhao, and L. Jiao, "Sparse subspace clustering-based feature extraction for PolSAR imagery classification," *Remote Sens.*, vol. 10, no. 3, p. 391, 2018.

[9] S.-E. Park and W. M. Moon, "Unsupervised classification of scattering mechanisms in polarimetric SAR data using fuzzy logic in entropy and alpha plane," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 8, pp. 2652–2664, Aug. 2007.

[10] E. Krogager, "New decomposition of the radar target scattering matrix," *Electron. Lett.*, vol. 26, no. 18, pp. 1525–1526, 1990.

[11] J. R. Huynen, "Phenomenological theory of radar targets," Ph.D. Dissertation, Dept. Elect. Eng., Math Comput. Sci., Univ. Technol., Delft, The Netherlands, 1970.

[12] R. Touzi, "Target scattering decomposition in terms of roll-invariant target parameters," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 73–84, Jan. 2007.

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

[14] A. Bhattacharya, A. Muhuri, S. De, S. Manickam, and A. C. Frery, "Modifying the Yamaguchi four-component decomposition scattering powers using a stochastic distance," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3497–3506, Jul. 2015.

[15] L. 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, p. 782, 2018.

[16] T. Berhane *et al.*, "Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory," *Remote Sens.*, vol. 10, no. 4, p. 580, 2018.

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

[18] S. Uhlmann, S. Kiranyaz, and M. Gabbouj, "Semi-supervised learning for ill-posed polarimetric SAR classification," *Remote Sens.*, vol. 6, no. 6, pp. 4801–4830, 2014.

[19] Z. Xue, P. Du, H. Su, and S. Zhou, "Discriminative sparse representation for hyperspectral image classification: A semi-supervised perspective," *Remote Sens.*, vol. 9, no. 4, p. 386, 2017.

[20] L. Ma, A. Ma, C. Ju, and X. Li, "Graph-based semi-supervised learning for spectral-spatial hyperspectral image classification," *Pattern Recognit. Lett.*, vol. 83, pp. 133–142, Nov. 2016.

[21] L. Yang, S. Yang, P. Jin, and R. Zhang, "Semi-supervised hyperspectral image classification using spatio–spectral Laplacian support vector machine," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 3, pp. 651–655, Mar. 2014.

[22] Z. He, H. Liu, Y. Wang, and J. Hu, "Generative adversarial networks-based semi-supervised learning for hyperspectral image classification," *Remote Sens.*, vol. 9, no. 10, p. 1042, 2017.

[23] X. Lu, J. Zhang, T. Li, and Y. Zhang, "Incorporating diversity into self-learning for synergetic classification of hyperspectral and panchromatic images," *Remote Sens.*, vol. 8, no. 10, p. 804, 2016.

[24] W. Hua, S. Wang, H. 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.

[25] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. 11th Annu. Conf. Comput. Learn. Theory*, 1998, pp. 92–100.

[26] A. Appice and D. Malerba, "A co-training strategy for multiple view clustering in process mining," *IEEE Trans. Services Comput.*, vol. 9, no. 6, pp. 832–845, Nov. 2016.

[27] X. Zhang, Q. Song, R. Liu, W. Wang, and L. Jiao, "Modified co-training with spectral and spatial views for semisupervised hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2044–2055, Jun. 2014.

[28] Z.-H. Zhou and M. Li, "Tri-training: Exploiting unlabeled data using three classifiers," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 11, pp. 1529–1541, Nov. 2005.

[29] J. Zhu, K. Tan, and Q. Du, "Tri_training for remote sensing classification based on multiscale homogeneity," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 3055–3058.

[30] W. Hua, S. Wang, Y. Guo, and W. Xie, "Semi-supervised PolSAR image classification based on the neighborhood minimum spanning tree," *J. Radars.*, vol. 8, no. 4, pp. 458–470, Aug. 2019.

[31] M. Laszlo and S. Mukherjee, "Minimum spanning tree partitioning algorithm for micro aggregation," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 7, pp. 902–911, Jul. 2005.

[32] 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.

[33] X. Wu, U. Jang, J. Chen, L. Chen, and S. Jha, "Reinforcing adversarial robustness using model confidence induced by adversarial training," 2017, *arXiv:1711.08001*. [Online]. Available: http://arxiv.org/abs/1711.08001

[34] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Jan. 2003.

[35] R. Kohavi and G. John, "Wrapper for feature subset selection," *Artif. Intell.*, vol. 97, no. 1, pp. 273–324, 1997.

[36] J. S. Lee, M. R. Grunes, and G. De Grandi, "Polarimetric SAR speckle filtering and its impact on terrain classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2363–2373, Sep. 1999.

[37] A. Muhuri, S. Manickam, and A. Bhattacharya, "Scattering mechanism based snow cover mapping using RADARSAT-2 C-band polarimetric SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 7, pp. 3213–3224, Jul. 2017.

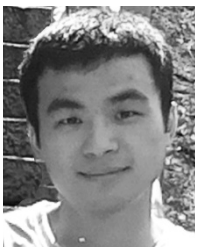
[38] I. Heine, T. Jagdhuber, and S. Itzerott, "Classification and monitoring of reed belts using dual-polarimetric TerraSAR-X time series," *Remote Sens.*, vol. 8, no. 7, p. 552, 2016.

[39] 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.

[40] M. Neumann, L. Ferro-Famil, and A. Reigber, "Estimation of forest structure, ground, and canopy layer characteristics from multibaseline polarimetric interferometric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1086–1104, Mar. 2010.

Shuang Wang (Member, IEEE) was born in Xi'an, Shaanxi, China, in 1978. She received the B.S. degree in electronic engineering, the M.S. degree in computer applications technology, and the Ph.D. degree in circuits and systems from Xidian University, Xi'an, China, in 2000, 2003, and 2007, respectively.

In 2017, she joined the Department of Mechanical Engineering, University of Maryland, College Park, MD, USA, as a Visiting Scholar. She is a Professor with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an, China. Her research interests include sparse representation, image processing, synthetic aperture radar (SAR) automatic target recognition, and polarimetric SAR data analysis and interpretation.



Yanhe Guo received the B.S. degree in intelligent science and technology and the M.S. degree in electronic engineering from Xidian University, Xi'an, China, in 2013 and 2016, respectively. He is pursuing the Ph.D. degree with the School of Artificial Intelligence, Xidian University, Xi'an, China.

His research interests include machine learning and polarimetric symmetric radar data classification.

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 of Ministry of Education, Xidian University, Xi'an, China, in 2018.

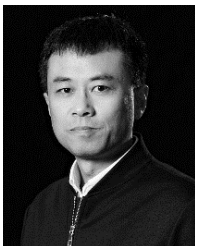
Since 2018, he has been with the School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an, where he is a Lecturer. His research interests include machine learning, deep learning, and PolSAR image processing.

Xinan Liu (Member, IEEE) received the B.S. degree in mathematics from Jilin University, Changchun, China, in 1986, the M.S. degree in physical oceanography from the Ocean University of China, Qingdao, China, in 1989, and the Ph.D. degree in mechanical engineering from the University of Maryland at College Park, College Park, MD, USA, in 2002.

From 1989 to 1997, he was a Research Scientist with the First Institute of Oceanography, State Oceanic Administration, Qingdao, China, where he was involved in the field of measurements of coastal dynamics on sand ocean beach, interaction between sea waves and vertical breakwaters, and shoaling of nonlinear random waves. Since 2002, he has been with the University of Maryland at College Park, where he is a Research Associate Professor with the Department of Mechanical Engineering and teaches undergraduate courses in electronics and instrumentation, and fluid mechanics. He has been involved in using radar and SAR for understanding the fundamental physics of a number of flow phenomena on the ocean surface. His research interests include radar backscattering of rough water surface, the response of the ocean surface to rainfall, water wave breaking, surfactant dynamics, and low-frequency waves in coastal areas.

Guoxin Song received the B.S. degree from Xiamen University, Xiamen, China, in 2017. He is pursuing the M.S. degree with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an, China.

His research interests include deep learning and PolSAR terrain classification.



Biao Hou (Member, IEEE) received the B.S. and M.S. degrees in mathematics from Northwest University, Xi'an, China, in 1996 and 1999, respectively, and the Ph.D. degree in circuits and systems from Xidian University, Xi'an, in 2003.

Since 2003, he has been with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University, where he is a Professor. His research interests include compressive sensing and synthetic aperture radar image interpretation.

Licheng Jiao (Fellow, IEEE) received the B.S. degree in high voltage from Shanghai Jiao Tong University, Shanghai, China, in 1982, and the M.S. and Ph.D. degrees in electronic engineering from Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively.

From 1984 to 1986, he was an Assistant Professor with the Civil Aviation Institute of China, Tianjin, China. From 1990 to 1991, he was a Post-Doctoral Fellow with the Key Laboratory for Radar Signal Processing, Xidian University, Xi'an, where he is the Director of the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China. He has authored or coauthored more than 200 scientific articles. His research interests include signal and image processing, nonlinear circuits and systems theory, wavelet theory, natural computation, and intelligent information processing.

Dr. Jiao is a member of the IEEE Xian Section Executive Committee and an Executive Committee Member of the Chinese Association of Artificial Intelligence. He is the Chairman of the Awards and Recognition Committee.