DNN-Based PolSAR Image Classification on Noisy Labels

Jun Ni, *Student Member, IEEE*, Deliang Xiang, *Member, IEEE*, Zhiyuan Lin, Carlos López-Martínez, *Senior Member, IEEE*, Wei Hu, and Fan Zhang, *Senior Member, IEEE*

***Abstract*—Deep neural networks (DNNs) appear to be a solution for the classification of polarimetric synthetic aperture radar (PolSAR) data in that they outperform classical supervised classifiers under the condition of sufficient training samples. The design of a classifier is challenging because DNNs can easily overfit due to limited remote sensing training samples and unavoidable noisy labels. In this article, a softmax loss strategy with antinoise capability, namely, the probability-aware sample grading strategy (PASGS), is developed to overcome this limitation. Combined with the proposed softmax loss strategy, two classical DNN-based classifiers are implemented to perform PolSAR image classification to demonstrate its effectiveness. In this framework, the difference distribution implicitly reflects the probability that a training sample is clean, and clean labels can be distinguished from noisy labels according to the method of probability statistics. Then, this probability is employed to reweight the corresponding loss of each training sample during the training process to locate the noisy data and to prevent participation in the loss calculation of the neural network. As the number of training iterations increases, the condition of the probability statistics of the noisy labels will be constantly adjusted without supervision, and the clean labels will eventually be identified to train the neural network. Experiments on three PolSAR datasets with two DNN-based methods also demonstrate that the proposed method is superior to state-of-the-art methods.**

***Index Terms*—Deep neural network (DNN), image classification, noisy label, polarimetric synthetic aperture radar (SAR).**

Manuscript received January 17, 2022; revised March 3, 2022 and March 29, 2022; accepted April 15, 2022. Date of publication April 22, 2022; date of current version May 18, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61871413 and Grant 61801015, in part by the Fundamental Research Funds for the Central Universities under Grant XK2020-03, in part by China Scholarship Council under Grant 2020006880033, and in part by Grant PID2020-114623RB-C32 funded by MCIN/AEI/10.13039/501100011033. (*Corresponding author: Fan Zhang.*)

Jun Ni is with the College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China, and also with the Signal Theory and Communication Department, Polytechnic University of Catalonia, 08034 Barcelona, Spain (e-mail: nijunbuct@163.com).

Deliang Xiang is with the Advanced Innovation Center for Soft Matter Science and Engineering, Beijing 100029, China, and also with the Interdisciplinary Research Center for Artificial Intelligence, Beijing University of Chemical Technology, Beijing 100029, China (e-mail: xiangdeliang@gmail.com).

Zhiyuan Lin and Wei Hu are with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China (e-mail: lam13986916961@126.com; huwei@mail.buct.edu.cn).

Carlos López-Martínez is with the Signal Theory and Communication Department, Polytechnic University of Catalonia, 08034 Barcelona, Spain (e-mail: carlos.lopezmartinez@upc.edu).

Fan Zhang is with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China, and also with the Interdisciplinary Research Center for Artificial Intelligence, Beijing University of Chemical Technology, Beijing 100029, China (e-mail: zhangf@mail.buct.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3168799

I. Introduction

POLARIMETRIC synthetic aperture radar (PolSAR) imagery classification is important due to the unique physical properties that can be retrieved from electromagnetic wave scattering. Inspired by the image processing methods [1]–[3], deep neural network (DNN) seems to be a promising approach to perform classification tasks because of its precise classification effect and independence from complex models [4]–[6]. The convolutional neural network (CNN) has been employed as an effective deep learning method for image feature extraction, and it has been widely developed for PolSAR imagery classification [7]–[10].

These DNN-based methods aim to mine more reliable features by improving the network structure and loss functions and by expanding or enhancing samples with ideal training samples and labels. The current research progress is relatively slow in actual applications due to the limitations of transportability and sufficient training samples. Simultaneously, ground reference data play a fundamental role in the classification methods. Although the ground dataset is generally considered perfect, there may still be mislabeled samples, even though the data may be obtained from reliable sources [11]. In SAR/PolSAR images, training samples with noisy labels are almost inevitable because of many reasons such as mislabeling caused by human error, indistinguishable edge mixing, inaccurate scene image, and label registration, and even labels obtained from unsupervised automatic labeling technology. Under the condition of training samples with noisy labels, the DNN-based classification method can easily learn the characteristics of the noise samples and disturb the original classification rules, which is susceptible to overfitting [12]–[16].

Actually, learning with noisy datasets has been studied extensively in classification. In terms of computer vision, DNN-based methods for evaluating noise transfer matrices have been proposed successively [17]. To prevent overfitting, some researchers try to isolate the noisy labels and update the DNN using only the separated clean labels. Two DNN-based models trained by a decoupling technique are designed to select samples that have different predictions [12]. The weights are implemented to mediate training samples to select clean samples [13]. Some methods try to select clean training samples using curriculum learning [14]. However, the discarded training samples may contain some vital information, and excessive discarding of training samples will result in a severe decline in the classification performance compared with the DNN updated by the same number

of clean samples. The noise-tolerant loss is designed to correct noisy labels [15], and Ghosh *et al.* [18] constructed a knowledge graph to guide the learning process. A simple joint optimization framework is implemented to learn the probable correct labels of training samples, and then the corrected labels are used to train models [19]. Han *et al.* [20] proposed an extra-label correction phase to correct the wrong labels. In addition, label regression is proposed for noisy supervision [21]. To avoid relying on too much auxiliary information, the probability difference distributions are designed to evaluate the noisy labels [22].

Unfortunately, there are few studies on PolSAR imagery classification with noisy labels, and many researchers of PolSAR classification design their classifiers under an ideal condition, except those carried out by some researchers based on target recognition and scene segmentation of SAR images. In target recognition, a transition matrix with noise rate is proposed to change the probability vector of softmax using the noise rate, but they idealized that noise labels were evenly distributed in each category [23]. Wang *et al.* [24] attempted to build a noise label corrected model by fitting the loss curve. In scene segmentation, A class label correction method is designed to correct the misclassified pixels in SAR image segmentation [25]. A top-2 smooth loss function with cost-sensitive parameters is designed to tackle the noise label and alleviate the overfitting problem caused by the noise label [26]. In PolSAR image classification, Hou *et al.* [27] designed a robust classification loss function to solve the influence of outliers in the PolSAR pixels, and a probability graph framework combined support vector machine (SVM) with Markov fields is designed to smooth images. Indeed, most of the methods in hyperspectral imagery classification regard the noisy labels as outliers [28]–[34]. Practically, additional errors may be introduced for some clean samples because of high interclass similarity compared with low intraclass similarity [35]. Simultaneously, the noise labels are not limited by the outliers, for example, large areas of error labeling caused by human mistakes limit the outlier-detection methods.

In addition, labeling is not only the category attribute problem in PolSAR imagery, but also crossed samples between different scenarios (pixels near the boundary or mix-pixels), and these neutral and hard understand samples are not suitable for classification. For the training processing of deep learning, the most simple and effective softmax-loss strategy is to establish a statistical model to guide neural networks to learn simple samples with common characteristics, then to learn difficult samples, and finally to discard the samples with unrecognizable features.

In this article, we design general softmax-loss strategies, namely PASGS and Auto-PASGS, for a DNN-based PolSAR image classification method to tolerate noisy labels. First, the influence of noisy labels on the DNN classifier in PolSAR image classification is analyzed. Given the influence of noisy labels, a probability-aware distribution diagram is implemented to learn DNN classifiers. Then, the probability-aware strategy is designed to estimate the probability of a sample to be noisy according to the probability-aware distribution diagram. Two kinds of softmax-loss strategies are developed for two situations: the noise ratio can be artificially evaluated or not. For the former, the noise ratio can be directly used for the dynamic weight adjustment algorithm. For the latter, the noise rate needs to be automatically evaluated according to the training and convergence status of the neural network, therefore, an evaluation indicator of sample differentiation is defined to estimate the noise profile. Finally, the dynamic weight adjustment algorithm is used to suppress the influence of noisy labels on the training process, and the softmax-loss strategy does not rely on additional auxiliary information, such as datasets, stages, and models. Two PolSAR images and one temporal PolSAR dataset are employed to verify the validity of our method. Considering the existing PolSAR classification methods, we have made the following contributions:

1) A novel noisy label identification method, namely a probability-aware distribution diagram, is implemented to estimate the probability of a PolSAR sample to be noisy. Based on the probability-aware distribution diagram, the sample grading strategy is designed to select clean samples and suppress noisy labels.
2) Two sample grading strategies according to the noise ratio can be artificially evaluated or not. And an evaluation indicator of sample differentiation is defined to estimate the noise profile. Without any prior knowledge, the antilabel noise method can be easily applied to the existing softmax loss to show its excellent performance.
3) Based on two DNN-based classification methods of PolSAR imagery, namely VGGNet and LeNet, hypothetical experiments with different noise conditions and five different forms of noise labels are designed to demonstrate the advantages of our method.

The rest of this article is organized as follows. Section II describes some preliminaries of the influence of noisy labels on DNN-based PolSAR image classification. Section III introduces the proposed noisy label detection method in detail, and the softmax loss strategy is formulated. In Section IV, the experimental results are analyzed. Finally, Section V concludes this article.

II. Background

Noise severely affects PolSAR image interpretation. The existing noise includes not only speckle noise of the PolSAR images, but also label noise in some special tasks (e.g., classification, target recognition, and semantic segmentation). The former is generated by the imaging mechanism of the SAR system and has been extensively studied [36]–[39], while the latter is generated by unavoidable human error or automatic labeling systems, which has a significant impact on the performance of data-driven classifier models, such as DNN-based classifiers.

To illustrate the influence of noisy labels on the PolSAR image classification, a VGGNet-based land cover classification method is designed with training samples in different noisy label ratios in Flevoland I (It will be introduced in Section IV-A1), and the results are shown in Fig. 1. Under the same samples, the labels of training samples are randomly changed, and the noise ratios are, respectively, 10%, 20%, and 30%. In the network, the "dropout" is set to 0.2 to prevent the overfitting caused by too many epochs, "Adam" is selected as the optimizer, and the noise label ratio of the training samples is the result obtained by random modification of the training labels. From Fig. 1,

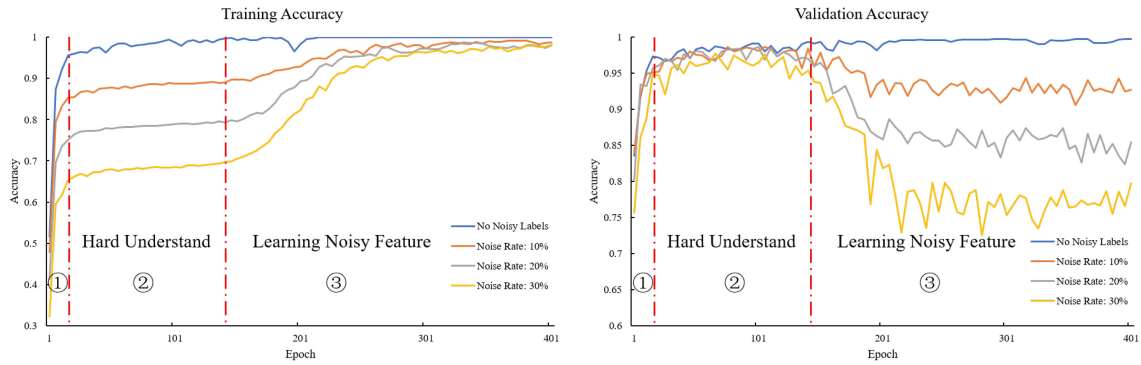


Fig. 1. Convergence curves of PolSAR image classification with different noise rates using VGGNet-8 on Flevoland I.

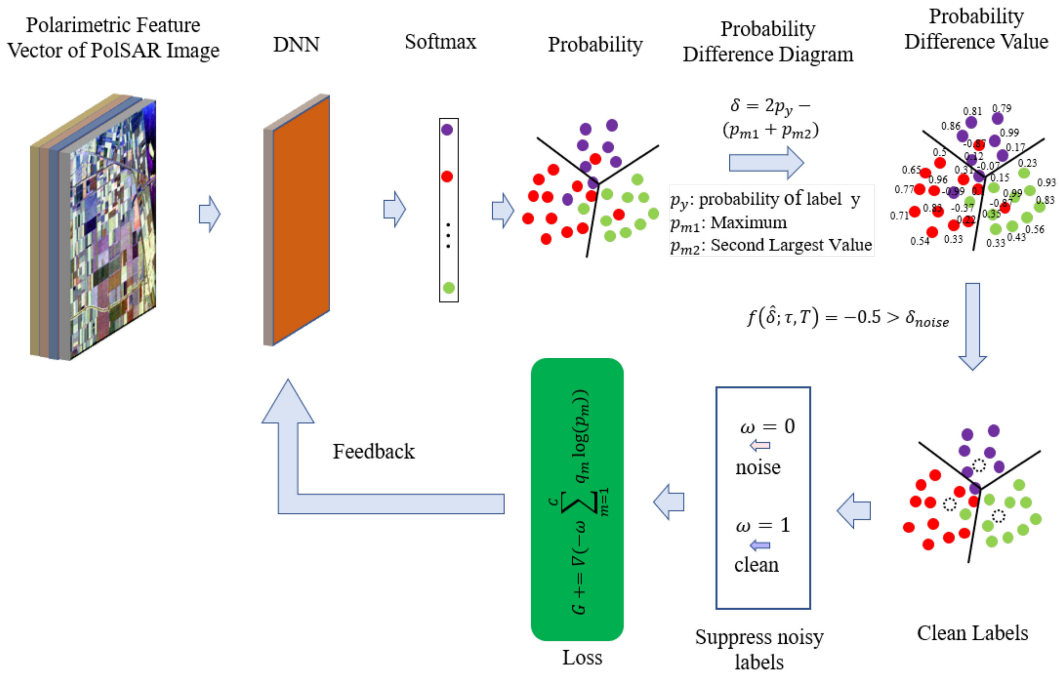


Fig. 2. Overall flowchart of the proposed noise-tolerant PolSAR classification method.

regardless of there are noise labels or not, the curve of training processing can be divided into three stages:

1) Rapid convergence stage, easy-to-understand samples dominate network training, so network convergence is fast, and training accuracy and test accuracy are constantly greatly improved. The noise samples have little influence on the network training process.
2) Convergence tends to moderate increase, the samples with noise labels and hard-to-understand samples dominate the training processing, so network convergence stagnate, and the improvement of training and test accuracy is slowing down. Due to the network learning noise samples, the accuracy of verification began to decline in the latter half, and the network appeared overfitting.
3) Stability stage, the networks are stable and all the training accuracy remains unchanged. In the training network without noise labels, the verification accuracy is stable and unchanged, but in the training networks with noisy labels, the verification accuracy fluctuates constantly.

Actually, the influence of noise samples gradually increases and eventually leads to overfitting. Although complete validation samples may be able to find the optimal network model, it is difficult to find such a complete validation sample set in practical applications, thus increasing the difficulty of selecting the optimal model. If the constraints of noise samples are gradually strengthened in the training processing, the overfitting problem will be alleviated.

III. Proposed Classification Framework

The softmax loss strategy has been proven to be an effective way to select clean samples [40]–[42]. This article also designs a softmax loss strategy in the DNN-based PolSAR image classification method, mainly including the probability-aware distribution and softmax-loss strategy formulation, see Fig. 2.

The proposed noise-tolerant classification framework mainly consists of four steps: pixelwise polarimetric feature vector extraction, spatial-polarimetric feature extraction using the

DNN-based method, calculation of the probability difference distribution diagram, and calculation of the loss using separated clean samples. First, a three-dimensional feature matrix is constructed for the PolSAR image by integrating various polarimetric feature extraction methods. Then, the polarimetric features are input into a CNN to extract spatial-polarimetric features and output the predicted probability using the softmax function. Next, the probability-aware distribution diagram can be calculated to separate clean samples from training samples. Finally, an algorithm is designed to dynamically constrain the noise proportion of the probability difference distribution diagram in one minibatch according to the number of iterations, and the loss can be calculated by the separated clean samples. In addition, two kinds of softmax loss strategies, namely, a given noise label ratio strategy and an unknown noise label ratio strategy, are developed for the PolSAR image classification.

A. Polarimetric Feature Vector Construction

As the basis of PolSAR imagery classification, polarimetric feature vector construction is an essential process in PolSAR image classification. The PolSAR complex backscattering matrix $[S]$ can be expressed as

$$[S] = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix} \tag{1}$$

where *h* and *v* represent the horizontal polarization channel and the vertical polarization channel, respectively. The scattering matrix [*S*] can be interpreted by the second-order polarimetric descriptors' multilook coherent matrix [T_3]or covariance matrix [C_3] in a single station backscattering mass.

The target decomposition (TD) theorems, divided into incoherent decomposition and coherent decomposition algorithms, are designed to interpret the scattering mechanisms of the targets physically. The incoherent TD theorems are usually used to process the coherent matrix [*T*], the covariance matrix [*C*], and the Kennaugh matrix [*K*]. The coherent TD algorithms are employed to process the scattering matrix [*S*]. The features obtained by TD have more vital image expression ability than the original matrices (i.e., [*K*], [*T*], [*C*], and [*S*]), and many studies have shown the effectiveness of these features [43], [44]. In summary, 107 features are selected for our PolSAR imagery classification, namely, 10 elements of the coherency and covariance matrixes, 18 polarimetric descriptors, and 79 TD parameters, as shown in Table I.

TABLE I
POLARIMETRIC FEATURES EMPLOYED IN THE PROPOSED METHOD

Category	Features (Dimension)	#parameters
Matrix Elements	Coherence Matrix (7), Covariance Matrix (3),	10
Polarimetric Descriptors	Span (1), Polarization Fraction(1), SERD/DERD(2), Shannon Entropy(1), Polarization Asymmetry (1), Radar Vegetation Index (1), Pedestal Height (1), Alpha Approximation (2), Entropy Approximation (2), Scattering Mechanism Entropy (2), Kozlov Anisotropy (1), Lueneburg Anisotropy (1), Conformity Coefficient (1), Scattering Predominance (1)	18
Decomposition Components	H/A/α (11), Holm (6), Huynen (3), Barnes (6), Cloude (3), Freeman (2), Freeman-Durden (3), Krogager (3), Vanzyl (3), Yang (7), Yamaguchi (7), Mcsm5 (6), Neuman (3), Tsvm (16)	79
Total		107

B. Disadvantages of the Cross-Entropy Loss

In PolSAR data classification, the category label of samples is treated as a probability vector $\vec{P} = [p_0, p_1, \ldots, p_C]^T$, where C is the number of categories. $p_c \in [0, 1]$ is the predicted probability of belonging to the cth class and can be output by the softmax function as

$$p_c(\vec{x}) = \frac{\exp(\vec{W}_c^T \vec{x} + \vec{b}_c)}{\sum_{j=1}^{C} \exp(\vec{W}_j^T \vec{x} + \vec{b}_j)} \tag{2}$$

where $\vec{x}$ is the input of the softmax function obtained by the DNN-based method, and $\vec{W}_*^T$ and $\vec{b}_*$ are the weight and bias of the full connection layer located before the softmax layer, respectively.

The cross-entropy loss fully believes in the training labels, and the predicted probability is directly calculated by the cross-entropy loss as

$$\mathcal{L} = -\sum_{c=1}^{C} q_c \log(p_c) \tag{3}$$

where q_c is the preset truth probability of the cth category in the training samples and is defined as

$$q_c = \begin{cases} 0 & c \neq y \\ 1 & c = y \end{cases} \tag{4}$$

Due to the belief in the authenticity of the sample label, in the update process of the DNN-based method, only when the prediction label c is the same as the sample label y (i.e., $c = y$) will the corresponding prediction probability contribute to the loss calculation, but the possibility of sample label error is ignored. In fact, the predicted probability p_y of the training sample labeled with y is a noisy label, and it will mislead the robustness of the DNN-based classifier. Therefore, the softmax loss strategy appears to be a fundamental algorithm for solving the problem of training samples with noisy labels and has been proven effective in selecting clean samples.

C. Probability-Aware Distribution

As shown in Fig. 1, the complexity of the training data can be reflected by the DNN-based classifier according to the training and validation curves, and the relevant network state in the update may be hinted in the probability vector $\vec{P}$. Therefore, this article attempts to design a probability distribution to distinguish noisy labels and clean labels.

Suppose there is a probability distribution mapping that can separate the noisy samples from clean samples, and the mapping relationship between the probability-aware distribution δ and

probability vector $\vec{P}$ is presented as

$$\delta = \Theta(\vec{P}) \tag{5}$$

where probability-aware function $\Theta(\cdot)$ is the mapping relationship function, and $\delta \in [\delta_{\min}, \delta_{\max}]$ should be a bounded distribution to make a reasonable analysis of the corresponding sample.

In probability vector $\vec{P}$ of one sample labeled the yth category, the special predicted probability p_y is developed to calculate the loss, and the other predicted probability hardly participates in any calculations. Considering y may not be the real ground truth because it may be a noisy label, the other predicted probability of this training sample should be considered in the calculation. The other predicted probability of this training sample is recorded as $p_{c\neq y}$. According to the probability relationship between p_y and $p_{c\neq y}$, various probability-aware distributions can be proposed.

1) Label Probability-Aware Distribution: In the training update of a DNN-based classifier, p_y is a process increasing from 0 to 1, while $p_{c\neq y}$ decreases and approaches 0. Finally, the predicted probability p_y of one training sample y in the trained DNN-based classifier is close to 1, and the probability $p_{c\neq y}$ of other labels in $\vec{P}$ is near 0. Therefore, the probability relationship between p_y and $p_{c\neq y}$ can be directly constructed by

$$\delta = p_y - \sum_{c=1, c\neq y}^{C} p_c \tag{6}$$

where $\delta \in [-1, 1]$. In the training processing, δ of clean samples will easily approach 1 because these samples have more similar characteristics. According to (2), $\sum_{c=1}^{C} p_c = 1$ and

$$\sum_{c=1, c\neq y}^{C} p_c = \sum_{c=1}^{C} p_c - p_y = 1 - p_y \tag{7}$$

Equation (6) can be revised as

$$\delta = 2 \cdot p_y - 1. \tag{8}$$

Although $p_{c\neq y}$ is implemented to calculate the probability distribution, (8) does not make detailed distinctions on other category attributes of the corresponding sample.

2) Distinguishable Probability-Aware Distribution: Actually, another predicted probability needs to be taken into account in the predicted probability vector $\vec{P}$, i.e., the maximum predicted probability p_m. Most clear samples have common characteristics that are easily extracted by the DNN-based classifier, so the label predicted probability p_y is usually the maximum predicted probability p_m, i.e., $y = m$. However, the characteristics of noisy samples are considered to be difficult to understand, and their predicted probabilities are easily mixed with p_m. To distinguish those predicted probabilities, p_m is defined as the maximum probability other than p_y in $\vec{P}$, and the distinguishable probability distribution is defined as

$$\delta = p_y - p_m \tag{9}$$

where $\delta \in [-1, 1]$. For example, the output probability vector of softmax is {0.01, 0.02, 0.3, 0.07, 0.6}, if the corresponding training label is $y = 5$, the distinguishable probability should be $\delta = 0.3$ because of $p_y = 0.6$ and $p_m = 0.3$, if the corresponding training label is $y = 3$, the distinguishable probability should be $\delta = -0.3$ because of $p_y = 0.3$ and $p_m = 0.6$. If y is a clear label, but the sample characteristics are not extracted in the initial update stage, it can easily be considered a noise sample. To avoid this issue, the predicted probabilities in $\vec{P}$ are sorted, and the largest probability p_{m1} and the second-largest probability p_{m2} are considered to be more easily confused with p_y. Then, (9) should be revised as

$$\delta = (p_y - p_{m1}) + (p_y - p_{m2}) = 2p_y - (p_{m1} + p_{m2}) \tag{10}$$

where the range of δ is still $[-1, 1]$ because of $p_{m1} + p_{m2} \leq 1$.

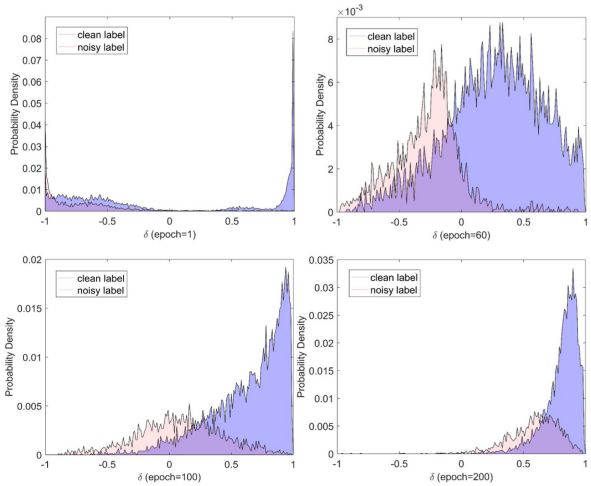


Fig. 3. Distinguishable probability-aware distribution of VGGNet-8 with a 30% noise label ratio with Flevoland I.

Based on the classification results of Fig. 1, the VGGNet-based PolSAR classification method with a 30% noisy label ratio of training samples can be expressed by the distinguishable probability-aware distribution, as shown in Fig. 3. The vertical axis represents the statistics of δ, which is a percentage of the total number of training samples. Obviously, the universal distinction of the sample features in each category is easily captured in the initial training stage of the network. Therefore, most of the samples with $\delta \approx 1$ are clean samples, while δ of the noisy label is generally close to -1. As the number of epochs increases, the features that are not easy to distinguish are gradually extracted by the neural network. However, the features of the noise samples are also learned by the neural network, resulting in overfitting of the DNN-based model.

D. Probability Statistic Distribution

Considering Sections III-B and III-C, a sample grading strategy is designed to clean the noisy labels. First, δ is considered an independent variable of the probability-aware distribution, and its probability density function $f(\delta)$ is expressed as

$$F(\delta) = \int_{-\infty}^{+\infty} f(\delta) d\delta = \int_{\delta_{\max}}^{\delta_{\min}} f(\delta) d\delta = 1 \tag{11}$$

and when $\delta = \delta_k$, the cumulative distribution function $F(\delta_k)$ is expressed by

$$F(\delta_k) = \int_{\delta_{\min}}^{\delta_k} f(\delta) d\delta, \qquad \delta_k \in [\delta_{\min}, \delta_{\max}]. \tag{12}$$

According to the analysis in Section III-C, the noise samples can be distinguished from the clean samples by δ_k. With the increase in the number of epochs in the training process, δ_k can slowly changes its constraint value (from $\delta_{\min}$ to $\delta_{\max}$) to suppress the interference of separated noise samples. Due to the limitation of the number of training samples, the probability-aware variable $f(\delta)$ can be equally discretized into N bins, and suppose that δ_n locates in nth bin

$$\delta_n \in (\frac{(n-1)(\delta_{\max}-\delta_{\min})}{N}+\delta_{\min}, \frac{n(\delta_{\max}-\delta_{\min})}{N} + \delta_{\min}], \quad n \in \{1,2,\ldots,N\}. \tag{13}$$

The proportion of training samples in the nth bin is

$$\begin{aligned}\triangle F(\delta_n) &= \int_{\frac{(n-1)(\delta_{\max}-\delta_{\min})}{N}+\delta_{\min}}^{\frac{(n)(\delta_{\max}-\delta_{\min})}{N}+\delta_{\min}} f(\delta)d\delta \\ &\approx \lim_{N\to+\infty} \sum_{i=\frac{n(\delta_{\max}-\delta_{\min})}{N}+\delta_{\min}}^{\frac{n(\delta_{\max}-\delta_{\min})}{N}+\delta_{\min}} \triangle f(i)\end{aligned} \tag{14}$$

where $\triangle f(i)$ is the proportion of training samples with $\delta = i$; thus, the cumulative distribution function $F(\delta_n)$ is obtained by

$$F(\delta_n) = \sum_{i=1}^{n} \triangle F(\delta_i) \tag{15}$$

To improve the computational efficiency, N is set to 100 in this article. According to (10), $\delta_{\max} = 1$ and $\delta_{\min} = -1$. Furthermore, one minibatch is considered to reduce the stabilization of sample selection because of the variability of the loss values in one epoch and the demand for calculation. To avoid increasing the computational burden of computing δ frequently, δ is set as a global variable and stored in memory. As the number of updates increases in one epoch, δ is constantly accumulated. According to the probability-aware distribution δ, the softmax-loss strategy can be formulated.

E. Sample Grading Strategy for a Given Noise Ratio

As shown in Fig. 3, the clean samples are more concentrated around $\delta_{\max}$ in the probability-aware distribution. This trend becomes more apparent with the increase in the number of epochs. If the probability-aware distribution δ close to $\delta_{\min}$ is considered noisy samples, the ideal sample grading strategy should concentrate the noise samples in the back part of the probability-aware distribution diagram, while the clean samples are distributed in the front of the probability-aware distribution diagram so that the noise samples and the clean samples are separated.

Suppose that $\widehat{\delta}$, namely, the probability-aware threshold value, is a threshold for distinguishing between noisy samples and clean samples, and corresponds to the zth bin. $\widehat{\delta}$ is calculated by

$$\widehat{\delta} = \frac{z}{N}(\delta_{\max}-\delta_{\min})+\delta_{\min}. \tag{16}$$

Algorithm 1: The Probability-Aware Sample Grading Strategy (PASGS) for a Given Noise Ratio.

Require: the PolSAR samples *D*, DNN-based network, the initialize wight $\vec{W}$ of networks, T_{max}, τ
Ensure: the updated wight $\vec{W}$
get mini-batch samples $\bar{D}$ from *D*;
calculate softmax and output the probability vector $\vec{P}$;
if end of one epoch **then**
  initialize probability-aware distribution δ;
else
  cumulative probability-aware distribution δ;
end if
if $T < T_{max}$ **then**
  calculate $\hat{\delta}$ using Function (18) and update ω;
else
  $F(\hat{\delta}) = \tau$;
  calculate $\hat{\delta}$ and update ω;
end if
calculate loss $\mathcal{L}$ and calculate the gradient $G+ = \nabla(-\omega \sum_{c=1}^{C} q_c \log(p_c))$
get the updated wight $\vec{W}$

Given the proportion τ of noisy labels, δ_{noise} of noisy labels should meet

$$\delta_{\text{noise}} \le \widehat{\delta} \le \delta_{\max} - \tau(\delta_{\max}-\delta_{\min}) \tag{17}$$

and $F(\delta_{\text{noise}}) \le \tau$. Since the DNN-based classifiers do not have the ability to classify samples correctly at the beginning of the training process, $F(\widehat{\delta})$ should be 0 and $\widehat{\delta} \approx \delta_{\min}$. As the number of epochs increases, the noise samples are gradually separated from training samples and finally reach the condition of $F(\widehat{\delta}) \approx \tau$. If the network can completely distinguish the noisy labels after $T_{\max}$ iterations, the dynamic drop rate $F(\widehat{\delta})$ can be set to distinguish noisy samples

$$F(\widehat{\delta}) = \tau \cdot \frac{T}{T_{\max}}, \quad T \in \{0,1,2,\ldots,T_{\max}\} \tag{18}$$

where T is the number of training epochs. Since N is not an infinite value in actual calculations, it is not easy to find an xth bin to satisfy (18). Therefore, the upper limit value $\widehat{\delta}'$ satisfying the condition $F(\widehat{\delta}) \le \tau \cdot \frac{T}{T_{\max}}$ is selected as the constraint threshold. Therefore, the softmax loss strategy can be formulated as

$$\mathcal{L} = -\omega \sum_{c=1}^{C} q_c \log(p_c) \tag{19}$$

where ω is a weight adjustment matrix for each training sample in one minibatch and is defined as

$$\omega = \begin{cases} 1 & \delta \ge \widehat{\delta}' \\ 0 & \delta < \widehat{\delta}' \end{cases} \tag{20}$$

In addition, $F(\widehat{\delta})$ will be identical to τ when the number of network updates is greater than $T_{\max}$. Therefore, the overall flowchart of sample grading with the softmax loss strategy for a given noise ratio, namely, the probability-aware sample grading strategy (PASGS), can be designed as shown in Algorithm 1.

F. Sample Grading Strategy Without Noise Ratio

Generally, the noisy label ratio is hard to obtain due to the lack of information on the training samples in actual applications. Although the noisy label ratio can be inferred by the validation set [45], the inferred rate cannot accurately reflect the noisy label ratio of the training samples. Therefore, a method independent of the noise sample proportion is more suitable in practical applications.

In the absence of a given noise ratio, two problems need to be addressed: 1) what is the reference standard of $\widehat{\delta}$; and 2) how to evaluate the noise ratio.

1) Reference Standard of $\widehat{\delta}$: In the ideal method, the noise samples should be separated from clean samples. The maximum predicted probability p_m of the noisy label is far greater than the predicted label probability p_y. This means the sample should be a noise sample when $\delta \to \delta_{\min}$, the sample should be a clean sample when $\delta \to \delta_{\min}$, and the sample should be a hard recognition sample when δ is near $\frac{(\delta_{\min}+\delta_{\max})}{2}$. Intuitively, from Fig. 3, the proportion of clean samples, in the case of $\delta < 0$, is continuously decreasing with the increase in the number of network updates. Therefore, the sample grading strategy without a noise ratio can be designed with the ideal threshold value $\widehat{\delta} = \frac{(\delta_{\min}+\delta_{\max})}{2}$.

Similarly, suppose that the network can completely distinguish the noisy label after $T_{\max}$ iterations; the dynamic drop rate $F(\widehat{\delta})$ can be set to distinguish noisy samples

$$F(\widehat{\delta}) = \frac{T}{T_{\max}} + \frac{2\delta_{\min}}{\delta_{\max} - \delta_{\min}}, \qquad T \in \{0, 1, \ldots, T_{\max}\} \quad (21)$$

However, the estimated threshold value $\widehat{\delta} = \frac{(\delta_{\min}+\delta_{\max})}{2}$ cannot be completely trusted since the DNN-based method can eventually memorize incorrect labels [46]. The ideal approach is to clean the training samples further using a noise ratio.

2) Evaluate the Noise Ratio: Theoretically, the probability-aware distribution δ of noisy samples and clean samples should be close to $\delta_{\min}$ and $\delta_{\max}$, but it cannot be completely proved that the samples are clean samples when $\widehat{\delta} > \frac{(\delta_{\min}+\delta_{\max})}{2}$ and the samples are noisy samples when $\widehat{\delta} < \frac{(\delta_{\min}+\delta_{\max})}{2}$ after $T = T_{\max}$ because $T_{\max}$ is also an inference value. Therefore, an appropriate method needs to be developed to evaluate the training level of the network.

Suppose that only δ_{noise} of the noisy labels is considered and the number of noise samples is K_{noise}; the average $\bar{\delta}_{\text{noise}}$ of noise samples can be calculated as

$$\bar{\delta}_{\text{noise}} = E[\delta_{\text{noise}}] = \frac{1}{K_{\text{noise}}} \sum_{i=1}^{K_{\text{noise}}} \delta_{\text{noise},i}. \quad (22)$$

Similarly, the average $\bar{\delta}_{\text{clear}}$ of the clear samples is calculated as

$$\bar{\delta}_{\text{clear}} = E[\delta_{\text{clear}}] = \frac{1}{K_{\text{clear}}} \sum_{i=1}^{K_{\text{clear}}} \delta_{\text{clear},i}. \quad (23)$$

Algorithm 2: The Automatic Probability-Aware Sample Grading Strategy (Auto-PASGS) without Noise Ratio.

Require: the PolSAR samples D, DNN-based network, the initialize wight $\vec{W}$ of networks, T_{max}
Ensure: the updated wight $\vec{W}$
get mini-batch samples $\bar{D}$ from D;
calculate softmax and output the probability vector $\vec{P}$;
if end of one epoch **then**
  initialize probability-aware distribution δ;
else
  cumulative probability-aware distribution δ;
end if
if $T < T_{max}$ **then**
  calculate $\hat{\delta}$ using Function (21) and update ω;
else
  calculate ξ using Function (15);
  if $\xi < \xi_p$ **then**
    $\hat{\delta} = 0$ and update ω;
  else
    calculate $F(\hat{\delta})$ and estimate $\hat{\tau}$;
    calculate $\hat{\delta}$ using Function (18) and update ω;
  end if
end if
calculate the loss $\mathcal{L}$ and gradient $G+ = \nabla(-\omega \sum_{c=1}^{C} q_c \log(p_c))$
get the updated wight $\vec{W}$;

The separation level of noise samples and clear samples can be expressed as

$$\xi = \frac{|\bar{\delta}_{\text{clear}} - \bar{\delta}_{\text{noise}}|}{\delta_{\max} - \delta_{\min}} \quad (24)$$

where $\xi \in [0, 1]$. However, ξ cannot be obtained because it is impossible to distinguish the respective distribution of clean samples and noise samples in the probability-aware distribution. Therefore, a hybrid evaluation index is defined as

$$\xi = \frac{2\sum_{i=1}^{N} |\delta_i - \frac{\delta_{\max}+\delta_{\min}}{2}| \cdot \Delta F(\delta_i)}{\delta_{\max} - \delta_{\min}} \quad (25)$$

where ξ indicates the concentration degree of the probability distribution diagram. The greater the value of ξ is, the more concentrated the probability distribution value δ of the training samples is at $\delta_{\max}$ or $\delta_{\min}$, indicating the higher the degree of discrimination between noise samples and clean samples. When the hybrid evaluation index ξ reaches a certain value ξ_p, the probability-aware distribution is considered to be able to distinguish noise samples from clean samples with $\delta = \frac{\delta_{\max}+\delta_{\min}}{2}$ as the boundary condition. Therefore, the noise ratio is evaluated as $\tau = F(\delta \leq \frac{\delta_{\max}+\delta_{\min}}{2})$. Finally, the noise-tolerant method can be improved by the sample grading strategy with the noise ratio.

The automatic sample grading strategy without the noise ratio, namely, Auto-PASGS, can be designed as in Algorithm 2.

Considering the noisy samples and clean samples have a large distinction in the probability difference distribution diagram after $\xi > \xi_p$, the loss and backpropagation of the network are

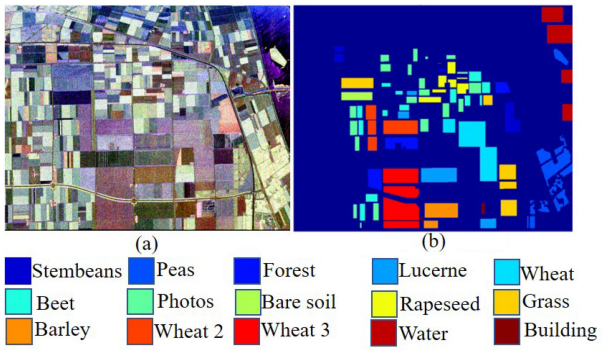


Fig. 4. Information of Pauli image and ground truth in Flevoland I. (a) Pauli image. (b) Ground truth.

TABLE II
INFORMATION OF TRAINING AND TEST SAMPLES IN FLEVOLAND I

No.	Name	Training Samples	Test Samples	No.	Name	Train Samples	Test Samples
1	Stembeans	271	6103	9	Rapeseed	316	6269
2	Peas	435	9111	10	Grass	555	12690
3	Forest	717	14944	11	Barley	337	7156
4	Lucerne	462	9477	12	Wheat 2	489	10591
5	Wheat	808	17283	13	Wheat 3	970	21300
6	Beet	428	10050	14	Water	620	13476
7	Photos	721	15292	15	Building	27	476
8	Bare soil	140	3078	Total	-	7296	157296

mainly affected by the clean samples. However, δ of some uncertain samples will still be near 0, and clean samples will increasingly be closer to 1 as the number of updates increases. Due to the differentiability of noise samples, the samples with $\delta < 0$ are roughly considered noise samples to evaluate noise ratio $\hat{\tau}$. Then, the dynamic threshold $\hat{\delta}$ can be inferred according to $F(\hat{\delta}) = \hat{\tau}$.

IV. Experimental Results and Analysis

In this section, two experiments are implemented to discuss the effectiveness of the designed classification method. Two airborne PolSAR datasets are employed to conduct the experiments and to validate the proposed classification scheme. In addition, two CNNs are implemented to prove the applicability of the proposed methodology.

A. Experimental Data

1) AIRSAR Data in Flevoland I: One L-band quad-polarimetric SAR dataset was obtained by the NASA/JPL AIRSAR system in Flevoland in August 1989. The size of these data is 750×1024 pixels, and 157 296 pixels are labeled into 15 different terrain types. The Pauli image and the ground truth are shown in Fig. 4. The detailed information of the training and test samples can be seen in Table II.

In the experiments, each sample is extracted into one cube centered on the label, and the block size is set to 14×14. Approximately 5% of pixels are randomly selected as training samples, and the rest are used as test samples. In the training samples, approximately 10%, 20%, and 30% of the training labels are randomly labeled as noisy labels to interfere with the DNN-based classifier, and 0% means that the training samples are clear.

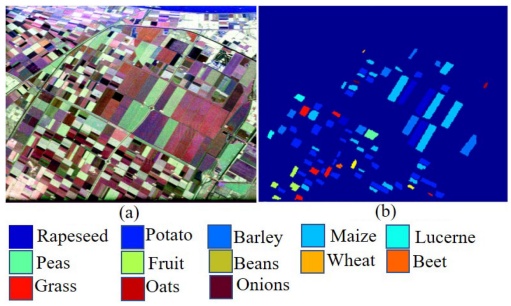


Fig. 5. Information of Pauli image and ground truth in Flevoland II. (a) Pauli image. (b) Ground truth.

TABLE III
INFORMATION OF TRAINING AND TEST SAMPLES IN FLEVOLAND II

No.	Name	Train Samples	Test Samples	No.	Name	Train Samples	Test Samples
1	Rapeseed	866	17621	8	Beans	63	1258
2	Potato	1691	32854	9	Wheat	5	120
3	Barley	823	16475	10	Beet	18	559
4	Maize	2002	40019	11	Grass	292	5800
5	Lucerne	114	2336	12	Oats	12	360
6	Peas	123	2409	13	Onions	9	247
7	Fruit	192	4156	Total	-	6210	124214

TABLE IV
INFORMATION OF TRAINING AND TEST SAMPLES IN NETHERLAND DATASET

No.	Name	Training Samples	Test Samples
1	Flower bulb	1208	156855
2	Fruit	1244	161879
3	Grass	5711	748886
4	Maize	2113	268156
5	Miscellanea	2720	350675
6	Onion	4108	542516
7	Pea	680	85962
8	Potato	7854	1025700
9	Spring-wheat	529	71536
10	Sugar-bee	5110	661697
11	Winter-wheat	8123	1116549
Total	-	40000	5190411

2) AIRSAR Data in Flevoland II: The AIRSAR system acquired another L-band quad-polarimetric SAR dataset in June 1991. The size of the image scene is 1024×1279 pixels with 13 different terrain types and 122 928 identified pixels. The Pauli image and the ground truth are shown in Fig. 5. The detailed information of training and test samples can be found in Table III.

In the experiment, approximately 5% of pixels are also randomly selected as training samples, and each sample is cut into one 14×14 cube centered on the label.

3) Temporal PolSAR Dataset in the Netherlands: Driven by solving the actual problem, a set of temporal PolSAR datasets are implemented to prove the practicability of our method. The C-band temporal quad-polarimetric SAR datasets were obtained by Radarsat-2 in the Netherlands from April 14, 2009, to September 29, 2009 in the framework of AgriSAR 2009. The size of these data is 5300×3100 pixels at 8 different times, and 5 190 411 pixels are labeled into 11 different terrain types are shown in Fig. 6. The detailed information of training and test samples can be found in Table IV. In the experiment,

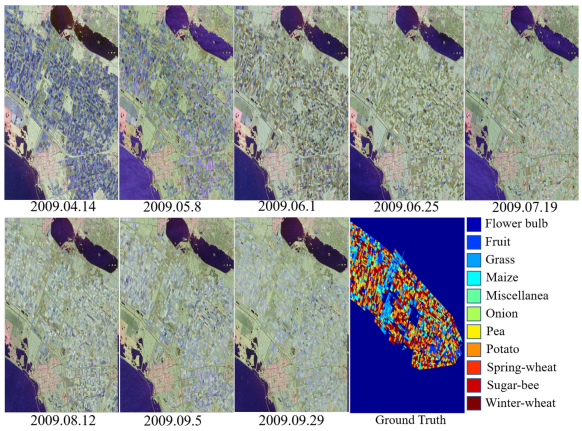


Fig. 6. Information of Pauli Image and Ground Truth about the Temporal PolSAR Datasets in the Netherlands.

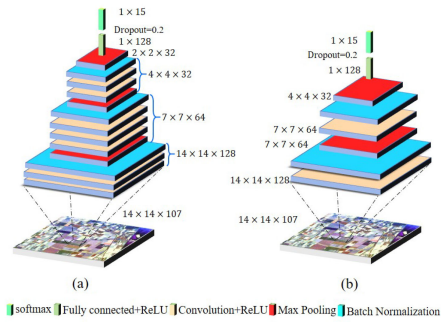


Fig. 7. Structure of VGGNet-8 and LeNet-5 for the PolSAR imagery classification in this article. (a) VGGNet-8. (b) LeNet-5.

approximately 0.77% of pixels are randomly selected as training samples, and each sample is cut into one 14×14 cube centered on the label.

B. DNN-Based Classification Method

To fully illustrate the effectiveness of our method, two DNN-based methods are designed to implement PolSAR image classification, including VGGNet-8 and LeNet-5. Their structures are shown in Fig. 7. In the experiments, Caffe [47] is implemented as the basic framework of our DNN. In each classification method, "SGD" is designed as the optimizer, the initialization learning rate is set to 0.0001, and the batch size is set to 128.

Two classical noise-tolerant classification methods have also been developed for comparative experiments, namely, Co-teaching [41] and JOCOR [48]. In addition, the traditional method without any noise-tolerant ability is referred to General. In all experiments, $T_{\max}$ is set to 30. In different classification methods, we separately combine VGGNet-8 and LeNet-5 with the general DNN-based method, Co-teaching, JOCOR, PASGS, and Auto-PASGS. In noise-tolerant methods of Co-teaching, JOCOR, and PASGS, the noise ratio in the training samples should be preset, while Auto-PASGS does not need to set the noise ratio in advance but rather automatically evaluates the noise ratio of the training samples through changes in the probability difference distribution diagram in the training process. In each method, the maximum epoch number is set to 400. Moreover, the average accuracy (AA), overall accuracy (OA), and kappa coefficient are calculated to evaluate the performance of different methods.

TABLE V
CLASSIFICATION RESULTS OF FLEVOLAND I IN DIFFERENT METHODS WITH DIFFERENT NOISE RATIOS (OA %, AA %, AND KAPPA)

Ratio	Method	LeNet-5			VGGNet-8		
		OA	AA	Kappa	OA	AA	Kappa
0%	General	99.53	98.7	0.9948	99.88	99.86	0.9986
	Co-teaching	**99.74**	**99.59**	**0.9972**	**99.97**	**99.97**	**0.9996**
	JOCOR	99.65	98.95	0.9962	99.86	99.7	0.9985
	PASGS	99.69	98.99	0.9966	99.94	99.9	0.9994
	Auto-PASGS	99.56	98.69	0.9952	99.92	99.68	0.9991
10%	General	98.74	97.29	0.9863	97.06	96.27	0.9679
	Co-teaching	96.89	96.1	0.966	96.69	96.02	0.9639
	JOCOR	99.06	**98.46**	0.9897	98.69	98.26	0.9857
	PASGS	**99.65**	98.28	**0.9962**	**99.66**	**99.23**	**0.9963**
	Auto-PASGS	99.15	97.79	0.9907	99.63	98.95	0.996
20%	General	94.98	94.64	0.9453	93.31	92.69	0.9271
	Co-teaching	94.93	91.22	0.9447	94.9	94.84	0.9444
	JOCOR	96.2	94.71	0.9586	97.34	95.98	0.971
	PASGS	**99.54**	**98.4**	**0.995**	**99.54**	**98.42**	**0.995**
	Auto-PASGS	98.52	97.04	0.9839	98.73	98.25	0.9861
30%	General	92.29	91.46	0.9161	87	86.74	0.8585
	Co-teaching	93.25	87.48	0.9263	93.6	94.5	0.9301
	JOCOR	93.39	91.83	0.928	95.02	94.82	0.9457
	PASGS	**99.24**	**98.07**	**0.9917**	**99.42**	**99.13**	**0.9936**
	Auto-PASGS	98.17	96.53	0.9801	97.88	97.65	0.9769

C. Analysis of Antinoise Performance

Two datasets in Flevoland are designed to evaluate VGGNet-8 and LeNet-5 for demonstrating the effectiveness of our proposed method. In the training samples, approximately 10%, 20%, and 30% of the training labels are also randomly labeled as noisy labels.

1) Results and Analysis of Flevoland I Data: In the first dataset, the number of samples can be considered balanced. The sample size of each category is amenable to the classifier because the design of the classifier does not need to account for the problem of sample balance. The classification results are shown in Table V.

In the clear training samples, OA of different classifiers can reach more than 99%, AA can reach more than 98%, and the kappa coefficient can reach more than 0.98. Compared with the results of LeNet-5 with clear samples, VGGNet-8 has better classification performance in each classification method. However, with the continuous increase of the noise ratio, the classification accuracy of the conventional classifiers has decreased significantly with the General method in LeNet-5 and VGGNet-8. The classification performance of VGGNet-8 has decreased more obviously than that of LeNet-5 because the inclusion of noise labels in training samples makes the deep network model more challenging to be trained. Although both Co-teaching and JOCOR can distinguish the noise samples, their influence on the classification performance of the classifier is not ideal. The OA of the classifier combined with Co-teaching and JOCOR is better than that of general classifiers, but the downward trend of the classification accuracy does not seem to stop. In addition, the AA of Co-teaching and JOCOR with LeNet

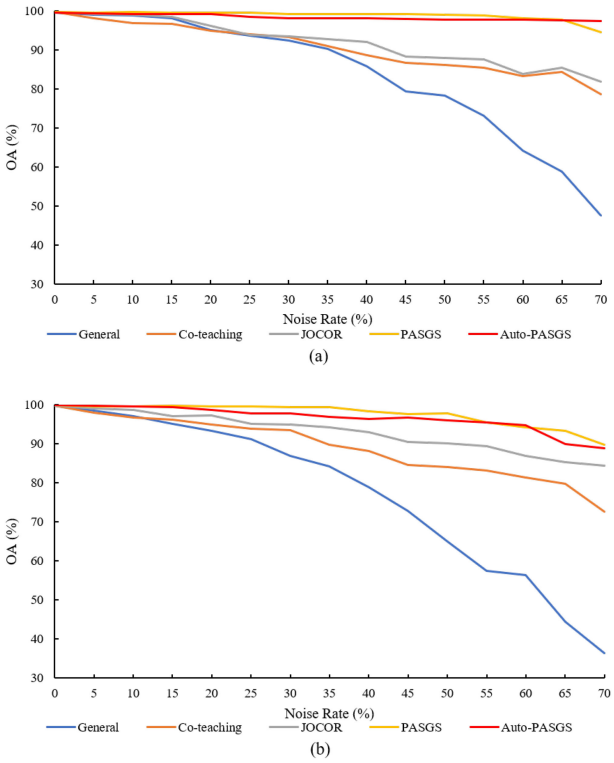


Fig. 8. Classification Accuracy with LeNet-5 and VGGNet-8 in Different Noise Rate. (a) Validation accuracy of LeNet-5 with different noise rate. (b) Validation accuracy of VGGNet-8 with different noise rate.

has no advantage compared to the general method, and even the AA of Co-teaching is lower than that of the general method. By contrast, the classification performance of the classifiers combined with PASGS and Auto-PASGS does not decrease significantly. Although the proportion of the noise ratio is constantly increasing, the OA of the combination of PASGS can always remain above 99%. Although the classification performance of the classifiers combined with Auto-PASGS is slightly lower than PASGS, this method still has significant advantages compared with other methods.

To track the antinoise performance of different models, more experiments with different noise rates are implemented, ranging from 0% to 70% with every 5% interval. The validation accuracy of different methods can be plotted, as shown in Fig. 8. With the increase of the noise rate, the uncertainty of OA in verification samples increases in the general method, which can be concluded from Figs. 1 and 8. At the same time, the classification performance of Co-teaching and JOCOR also gradually decreased. Nevertheless, the accuracy of PASGS and Auto-PASGS is relatively stable. Especially, the overall accuracy of Auto-PASGS in LeNet-5 is over 97%.

In detail, the probability difference diagrams of the combination methods in LeNet-5, VGGNet-8, and PASGS, and Auto-PASGS in the training process with the training samples that contained 30% noisy labels are presented in Fig. 9. Compared with the probability difference diagram of the VGGNet-based general classifier in Fig. 3, the methods combined with PASGS and Auto-PASGS can clearly distinguish noise samples and

TABLE VI
CLASSIFICATION RESULTS OF FLEVOLAND II IN DIFFERENT METHODS WITH DIFFERENT NOISE RATIOS (OA %, AA %, AND KAPPA)

Ratio	Method	LeNet-5			VGGNet-8		
		OA	AA	Kappa	OA	AA	Kappa
0%	General	99.79	98.18	0.9974	99.77	99.63	0.9971
	Co-teaching	99.8	98.87	0.9974	**99.93**	**99.9**	**0.9991**
	JOCOR	**99.83**	**99.77**	**0.9978**	99.91	99.85	0.9989
	PASGS	99.75	94.93	0.9968	99.77	99.63	0.9971
	Auo-PASGS	99.6	95.96	0.9949	99.72	99.49	0.9965
10%	General	98.18	80.31	0.9768	98.18	80.31	0.9768
	Co-teaching	98.39	94.12	0.9795	98.16	95.68	0.9765
	JOCOR	97.68	95.24	0.9705	97.55	94.08	0.9689
	PASGS	99.48	93.76	0.9934	99.66	95.2	0.9957
	Auo-PASGS	**99.59**	**95.39**	**0.9948**	**99.61**	**95.82**	**0.995**
20%	General	96.58	79.11	0.9565	98.18	80.31	0.9768
	Co-teaching	97.07	87.74	0.9628	97.18	86.59	0.9642
	JOCOR	95.73	87.04	0.9458	95.24	85.65	0.9397
	PASGS	99.2	82.2	0.9898	99.54	**93.54**	0.9941
	Auo-PASGS	**99.51**	**89.46**	**0.9937**	**99.61**	90.89	**0.9951**
30%	General	95.17	69.09	0.9387	80.66	69.05	0.7615
	Co-teaching	95.09	76.59	0.9375	94.51	79.05	0.9305
	JOCOR	93.71	80.74	0.9206	93.39	78.51	0.9166
	PASGS	**98.89**	**85.11**	**0.9858**	99.11	89.8	0.9887
	Auto-PASGS	98.53	81.02	0.9812	**99.53**	**93.68**	**0.994**

clear samples. This trend becomes increasingly evident with the increase in updating times. The noise samples and the clear samples are wholly separated in PASGS when epoch=100, and some samples are still mixed in Auto-PASGS, but the effect of this phenomenon on classification accuracy is weak. As the network continues to be trained, the probability difference diagrams of clear samples are closer to 1. In comparison, the probability difference diagrams of noise samples are closer to −1, which we expect to obtain in Sections III-E and III-F.

2) Results and Analysis of Flevoland II Data: Different with respect to the first dataset, Flevoland II dataset intensifies the imbalance of training samples, especially the *Wheat* labeled as 9th category only has five training samples, while *Maize* labeled as 4th category has 2002 training samples. The unbalance of the training samples will affect the discriminating ability of classifier. The classification results with 10 different classification methods can be shown in Table VI.

In the clear training samples, i.e., the noise ratio is 0%, the classification performance of VGGNet-8 is better than that of LeNet-5. After adding the noisy labels, the overall accuracy of LeNet-5 is higher than that of VGGNet-8, but this does not mean that LeNet-5 is better than VGGNet-8 in the classification performance because VGGNet-8 has higher AA accuracy than LeNet, which indicates that VGGNet-8 is more suitable than LeNet-5 to deal with the classification problem of unbalanced training samples. On the other hand, the classification accuracy of cooperation networks with two-path branches in Co-teaching and JOCOR is higher than that of the one-path network in General, PASGS, and Auto-PASGS.

After adding the noisy labels to the training samples, the overall classification results of Co-teaching are better than those of JOCOR but not as good as the proposed methods, i.e., PASGS and Auto-PASGS. In addition, the classification performance of Auto-PASGS is better than other methods, which indicates that the given noise ratio is not better than the autonoise ratio method in the imbalance of training samples.

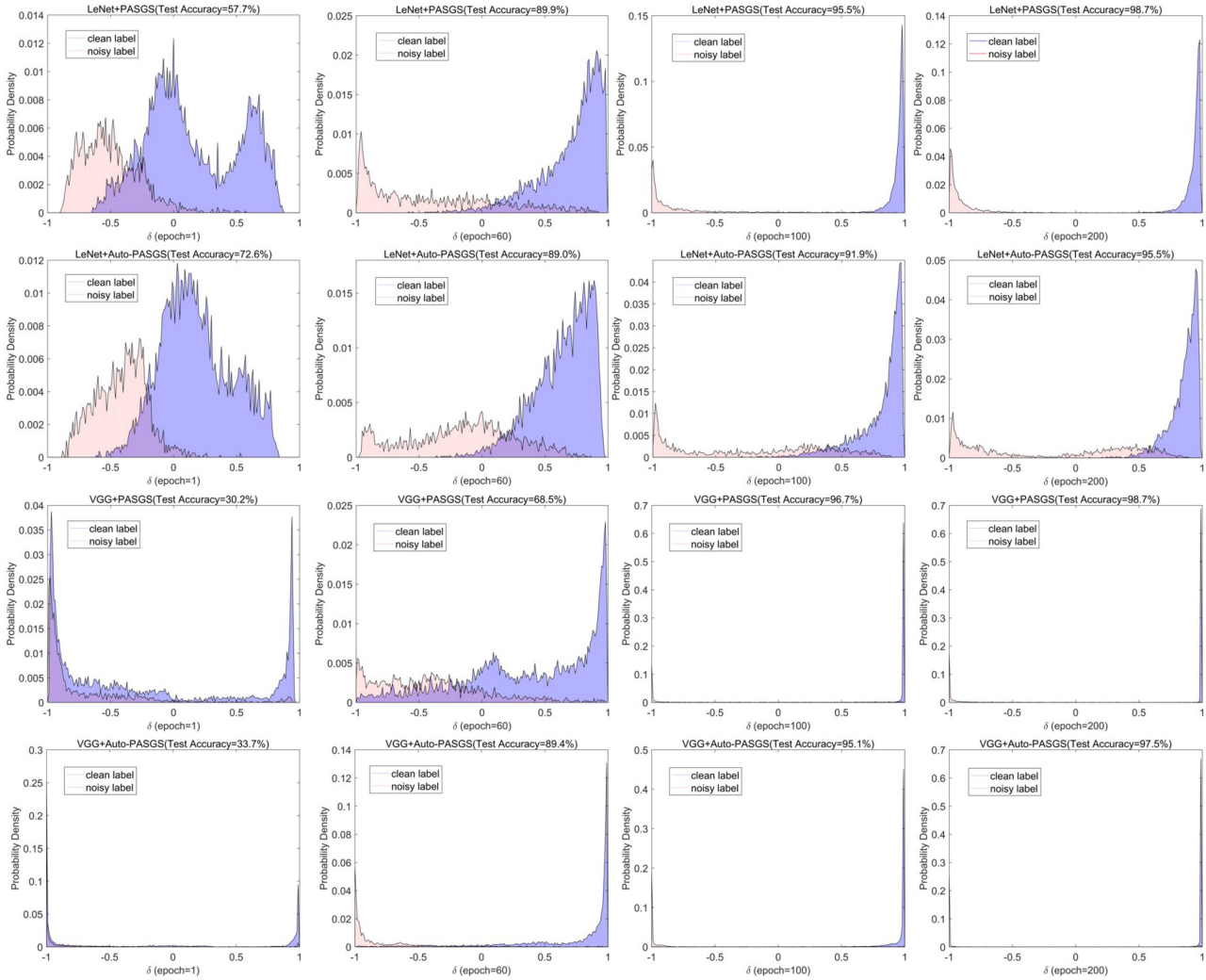


Fig. 9. Probability-aware distribution diagram by the proposed Softmax-Loss strategies with training samples contained 30% noisy labels.

D. Hypothetical Experiments in Different Noise Label Types

Considering the diversity of noise labels in PolSAR image classification, three hypothetical experiments of different noise distribution are implemented to prove the effectiveness of our methods.

1) Asymmetric Noisy Labels in Similar Areas: The first hypothetical experiment is that in the fine-grained PolSAR classification, two similar categories are confused by human errors. Based on Flevoland I, one area of Wheat-2 is mislabeled into Wheat-3, and the mislabeled labels accounted for around 14.77% of Wheat-2 samples and 0.99% of total samples, as shown in Fig. 10. In the asymmetric noise labels, 5% samples of the mislabeled map are selected to train network, and the real ground truth is used to test the performance. The classification results are shown in Table VII.

From the classification results, the classification accuracy of Wheat-2 is reduced in the general method, Co-teaching, and JOCOR. Their classification accuracy of OA dropped from 99% in Table V to around 98% in Table VII, mostly because the classification ability of Wheat-2 declined. The noise labels have

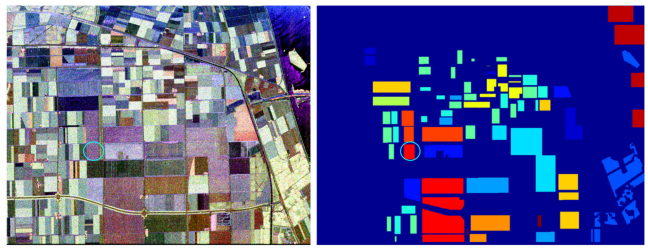


Fig. 10. One area of Wheat-2 mislabeled as Wheat-3. (a) Pauli image. (b) Noisy labels.

TABLE VII
CLASSIFICATION RESULTS OF ONE AREA OF WHEAT-2 MISLABELED TO WHEAT-3 (OA %, AA %, AND KAPPA)

Method	LeNet-5			VGGNet-8		
	OA	AA	Kappa	OA	AA	Kappa
General	98.88	98.08	0.9878	98.81	98.54	0.987
Co-teaching	98.59	98.03	0.9846	98.94	98.83	0.9884
JOCOR	98.58	98.54	0.9845	98.94	98.73	0.9884
PASGS	**99.47**	**98.76**	**0.9942**	**99.19**	**98.99**	**0.9912**
Auto-PASGS	99.27	98.57	0.9921	99.14	98.93	0.9906

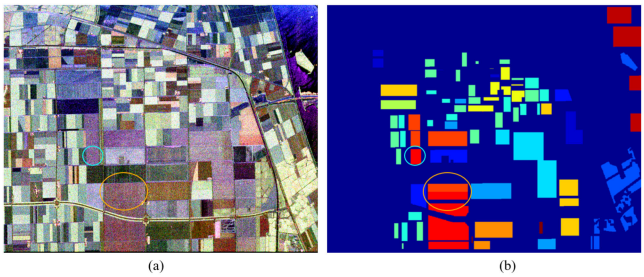


Fig. 11. Symmetric noise labels of Wheat-2 and Wheat-3. (a) Pauli image. (b) Noisy labels.

TABLE VIII
CLASSIFICATION RESULTS OF THE SYMMETRIC NOISE LABELS OF WHEAT-2 AND WHEAT-3 (OA %, AA %, AND KAPPA)

Method	LeNet-5			VGGNet-8		
	OA	AA	Kappa	OA	AA	Kappa
General	97.34	97.08	0.9709	96.87	97.6	0.9658
Co-teaching	97.05	96.69	0.9678	97.07	97.46	0.968
JOCOR	96.72	97.16	0.9642	96.63	91.25	0.9632
PASGS	98.05	**97.58**	0.9787	**97.69**	**97.86**	**0.9748**
Auto-PASGS	**98.11**	97.44	**0.9793**	97.37	97.77	0.9713

TABLE IX
CLASSIFICATION RESULTS OF ONE AREA OF WHEAT-2 IS MISLABELED TO WHEAT-3 (OA %, AA %, AND KAPPA)

Method	LeNet-5			VGGNet-8		
	OA	AA	Kappa	OA	AA	Kappa
General	98.75	97.31	0.9864	98.44	98.57	0.983
Co-teaching	98.2	97.67	0.9804	98.55	98.42	0.9841
JOCOR	98.2	98.06	0.9804	98.51	98.2	0.9838
PASGS	99	98.04	0.9891	99.36	**99.26**	0.993
Auto-PASGS	**99.12**	**98.37**	**0.9904**	**99.38**	99.24	**0.9932**

less effect on Wheat-3 because the absence of noise labels causes some Wheat-3 to be classified as Wheat-2. In LeNet-5, AA of Co-teaching and JOCOR is lower than that of General, but OA of JOCOR is better than that of General. In VGGNet-8, the classification accuracy of Co-teaching and JOCOR is better than that of General in the three classification performance indicators. In terms of OA, AA, and Kappa in LeNet-5 and VGGNet-8, the classification accuracy of PASGS and Auto-PASGS are superior to that of other compared methods.

2) Symmetric Noisy Labels in Similar Areas: The symmetric noisy labels in the similar area mean that some labels of Wheat-2 are mislabeled as Wheat-3, and some labels of Wheat-3 are also mislabeled as Wheat-2, which is shown in Fig. 11. In the noisy ground truth, 14.77% samples of Wheat-2 are labeled as Wheat-3, while 14.65% samples of Wheat-3 are labeled as Wheat-2, therefore, the mislabeled labels accounted for about 2.98% of total samples. 5% training samples of the mislabeled map are randomly selected to realize the classification, and the results are shown in Table VIII.

Compared with the results in Table IX, the classification accuracy of different methods in Table VIII is decreased, but our proposed method still has advantages. Specifically, the classification accuracy of PASGS and Auto-PASGS is better than that of General, Co-teaching, and JOCOR in both LeNet-5 and VGGNet-8 methods, and even the classification accuracy of

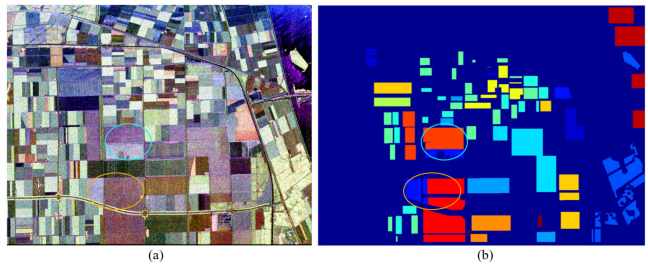


Fig. 12. Noisy labels at edge area. (a) Pauli image. (b) Noisy labels.

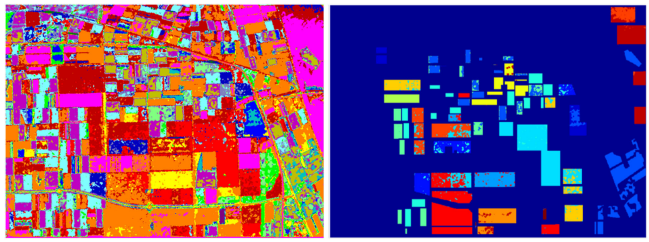


Fig. 13. Unsupervised classification results from $H/A/\alpha$-Wishart and the matched noisy labels. (a) Unsupervised classification labels. (b) The matched noisy labels.

Co-teaching and JOCOR is lower than that of the conventional method. In addition, the classification accuracy of Auto-PASGS in OA, AA, and Kappa is similar to that of PASGS, which means the proposed noise-tolerant network without artificial setting of noise ratio is feasible.

3) Noisy Labels at Edge Areas: The indistinguishable edge mixing label is another kind of noise label in scene classification and segmentation of PolSAR image. The noisy labels at the edge of one area can be divided into types: one is that the object is labeled into another class; The other is that the edge of one area does not belong to any labeled classes, as shown in Fig. 12. In the above-mixed edge scene, some irrelevant labels that may not be related to the classification types are also added to the ground truth, while the following mixed edge only considers the edge blends of marked types. The mislabeled labels accounted for 1.53% of total samples. Under 5% training samples from the mislabeled ground truth, the classification results are shown in Table IX.

Obviously, the classification accuracy of Co-teaching, JOCOR, PASGS, and Auto-PASGS in VGGNet-8 is more stable than that in LeNet-5, comparing the classification results in Tables V and IX. In LeNet-5, General has better OA performance disregarding PASGS and Auto-PASGS, but its AA is lower than that of Co-teaching and JOCOR. In VGGNet-8, OA, AA, and Kappa of Co-teaching and JOCOR have been improved compared with LeNet-5. Although OA and Kappa of General are decreased slightly, AA has been improved by 1.26%. In all classification results, Auto-PASGS for automatic recognition of noise proportions the optimal classification performance, PASGS is slightly lower than the former.

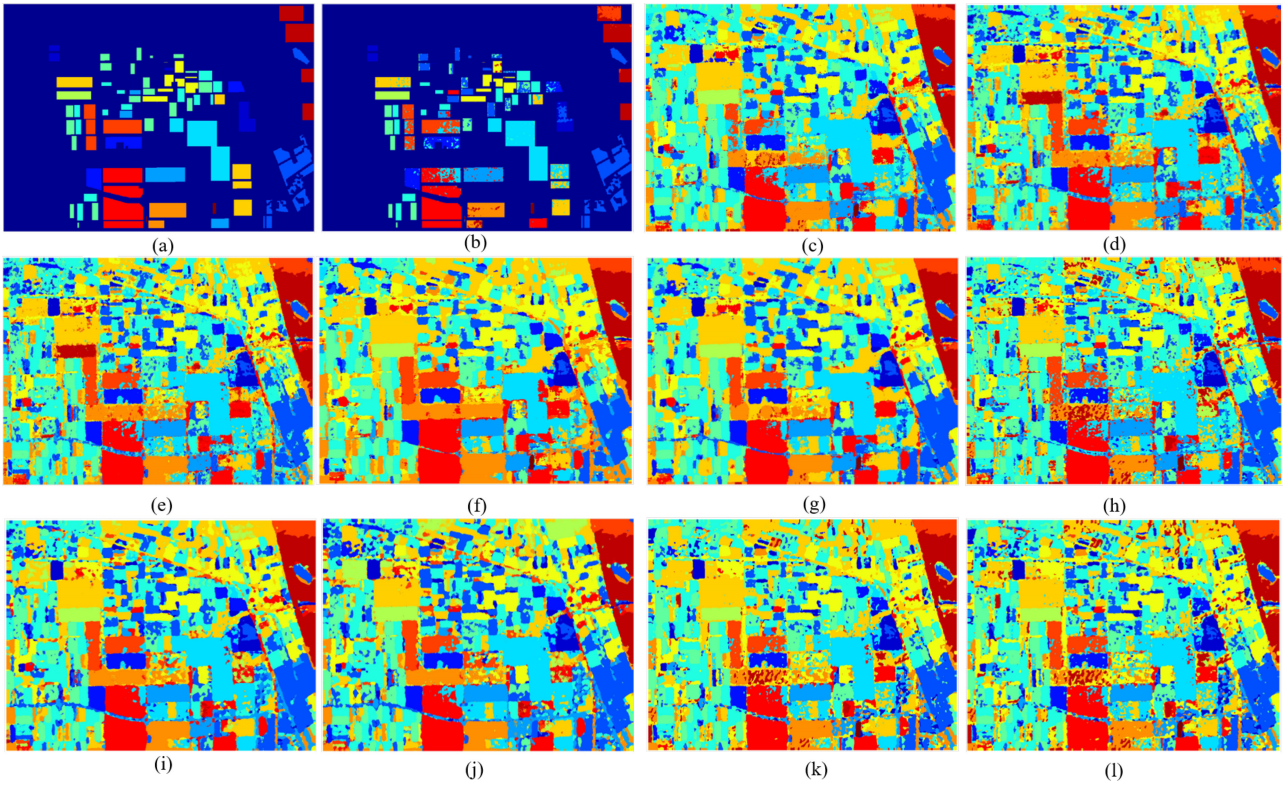


Fig. 14. Classification maps of different methods in the $H/A/\alpha$-Wishart result from Flevoland I. (a) Ground truth. (b) The corrected labels. (c) LeNet-5 general. (d) LeNet-5, co-teaching. (e) LeNet-5, JOCOR. (f) LeNet-5 PASGS. (g) LeNet-5 auto-PASGS. (h) VGGNet-8, general. (i) VGGNet-8, co-teaching. (j) VGGNet-8, JOCOR. (k) VGGNet-8 PASGS. (l) VGGNet-8, auto PASGS.

E. Applications in Different Noise Label Types

Considering the practicability in PolSAR image classification, two sets of practical applications are implemented to prove the effectiveness of our methods.

1) Noisy Labels With Unsupervised Classification Results: As one of the automatic labeling technologies, unsupervised classification is another important application in PolSAR image. More complicated situations of noise label will appear in the unsupervised classification results, limiting the performance of unsupervised classification. Although the deep learning method can reach a better-refined classification result, the limited training samples restrict its classifier performance. In this article, the unsupervised classification results can be regarded as a kind of the noisy ground truth for noise-tolerant classification methods. Based on the classical unsupervised classification method: $H/A/\alpha$-Wishart [49], the unsupervised classification labels can be obtained with 16 categories in Flevoland I dataset. To qualitatively evaluate the performance of each method, the 16 labels are matched with 15 refined categories in the manually labeled ground truth, as shown in Fig. 13.

The unsupervised classification result is obtained from PolSARPRO [50], the multilook size is set to 3×3, and the maximum of iteration with the Wishart method is set to 10. In the matching rule, the label with the maximum number of unsupervised labels in one category of the manual label is regarded as the modified label and others are regarded as noisy labels. After matching each category, the noisy labels account for 15.90% of whole labeled labels, and each class has a different proportion of noise labels. Finally, 5% samples of the noisy labels are randomly selected to realize the supervised classification, and there are around 84.10% clean labels in the training samples. The classification results are shown in Table X.

TABLE X
CLASSIFICATION RESULTS WITH THE MATCHED UNSUPERVISED LABELS (OA %, AA %, AND KAPPA)

Method	LeNet-5			VGGNet-8		
	OA	AA	Kappa	OA	AA	Kappa
General	83.77	84.44	0.8223	85.62	86.08	0.8427
Co-teaching	82.92	74.08	0.8304	85.78	81.33	0.8443
JOCOR	86.23	81.88	0.8128	84.49	80.79	0.8304
PASGS	**87.77**	**88.07**	**0.8662**	87.06	86.98	0.8585
Auto-PASGS	87.48	87.5	0.8631	**87.28**	**87.84**	**0.8609**

The overall classification process can be considered unsupervised except they need to preset noise rate in Co-teaching, JOCOR, and PASGS. In LeNet-5, the classification results of General and Co-teaching get worse because their OAs are lower than 84.10%. AA of JOCOR is improved by 2.13% compared with 84.10%, but its AA is worse than General. PASGS and Auto-PASGS are more superior to other methods, no matter in OA, AA, or Kappa. In VGGNet-8, the classification accuracy of all methods has been improved in terms of AA, but the classification performance of Co-teaching and JOCOR is not better than that General. Compared with General, AA, OA, and Kappa of PASGS and Auto-PASGS can reach more than 87.06%, 86.98%, and 0.8585, respectively, and there is little difference between PASGS and Auto-PASGS in classification

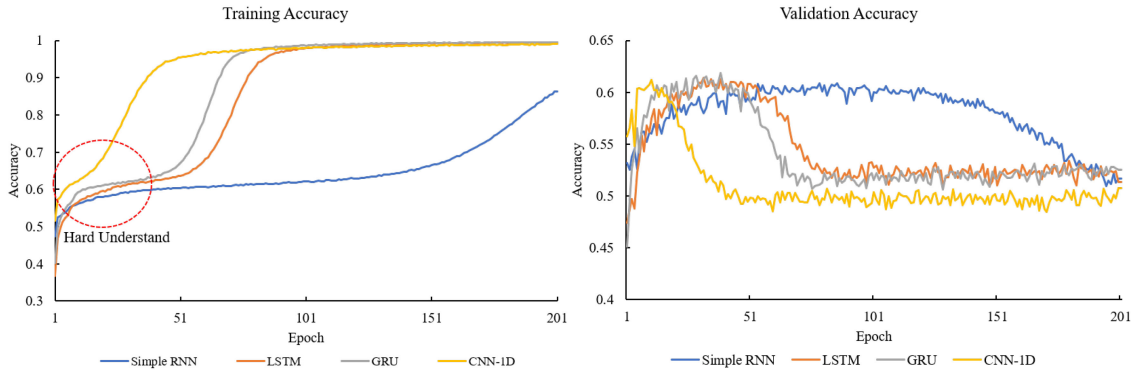


Fig. 15. Training and validation accuracy curves of different sequence methods on the temporal PolSAR images.

TABLE XI
CLASSIFICATION RESULTS OF TEMPORAL POLSAR IMAGES IN THE NETHERLAND (OA %, AA %, AND KAPPA)

Class	LeNet-5		VGGNet-8	
	General	Auto-PASGS	General	Auto-PASGS
1	85.6	89.27	91.63	92.4
2	89.64	88.32	92.19	93.23
3	89.57	90.04	93.01	93.62
4	85.14	81.69	91.08	92.15
5	75.17	77.26	84.34	85.62
6	89.74	90.65	93.74	94.32
7	71.5	78.42	86.87	88.35
8	90.78	94.68	95.93	96.83
9	31.02	21.12	69	64.23
10	95.25	94.27	95.32	95.39
11	96.6	96.54	94.89	95.65
OA	89.66	**90.47**	93.18	**93.84**
AA	81.82	**82.02**	89.82	**90.16**
kappa	0.8789	**0.8883**	0.9204	**0.928**

performance. The complete classification maps of the classification methods are shown in Fig. 14. The noise labels with unsupervised classification results are not evenly distributed in each category, and they are more complicated than those of the hypothesis. When there are a lot of noise labels, LeNet-5 in General is more stable than VGGNet-8 from Tables V and Table VI, which is similar as the classification maps in Fig. 14 of LeNet-5 of PASGS and Auto-PASGS are better than those of VGGNet-8.

2) Noisy Labels in Temporal PolSAR Data: Considering that PolSAR images at different times share one ground truth, there may be some errors after coregistration, resulting in edge mismatch of samples. Due to sequence-based deep learning classifiers being more sensitive to noise samples because of lack of spatial information, four network models, namely Simple RNN, LSTM, GRU, and CNN-1D, are used to test the impact of temporal PolSAR samples on the networks, as shown in Fig. 15. "SGD" is designed as the optimizer, the initialization learning rate is set to 0.0001, "dropout" is set to 0.2, and batch-size is set to 16. In addition, the multilook with 7×7 windows is implemented.

Obviously, the overfitting is easy to occur when there are noise samples, regardless of the network structure, and the training curve can be decomposed into three stages similar to Fig. 1. Even in the CNN-1D method, overfitting has already appeared when the epoch is less than 15. Since the noise rate in the training

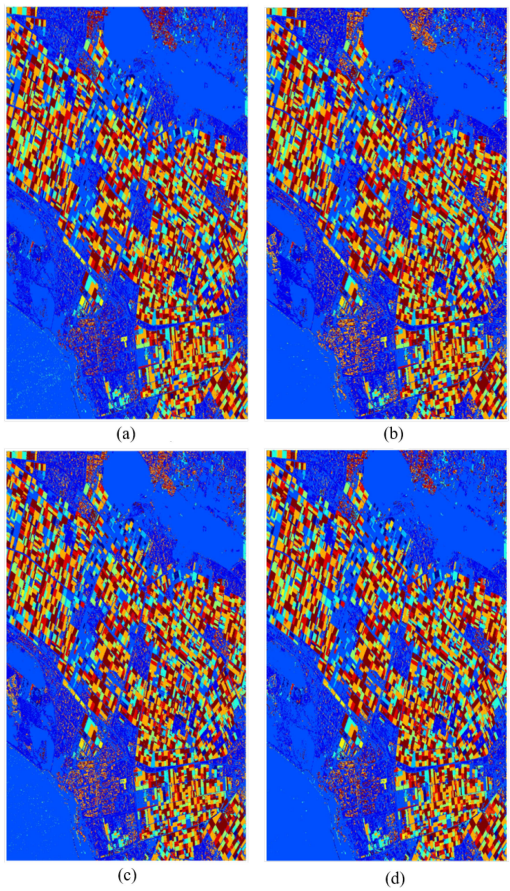


Fig. 16. Classification maps of different methods in the multi-temporal SAR dataset. (a) LeNet-5 General, 88.66% (b) LeNet-5 Auto-PASGS, 90.47% (c) VGGNet-8 general, 93.18%. (d) VGGNet-8 Auto-PASGS, 93.84%.

samples is not known, the noise-tolerant method based on artificial noise proportion (Co-teaching, JOCOR, and PASGS) is not suitable for actual temporal PolSAR classification. Based on the DNN-based methods (i.e., VGGNet-8 and LeNet-5), Auto-PASGS is combined with them to realize the temporal PolSAR image classification, as shown in Table XI. The complete classification maps of the four classification methods are shown in Fig. 16. In multitemporal datasets, the complexity of sample features is greatly increased, so the classification

maps of LeNet-5 with shallow networks are worse than those of VGGNet-8, but Auto-PASGS is superior to General.

Although the noise ratio can not be determined, the proposed Auto-PASGS does help improve the classification accuracy of OA, AA, and Kappa in LeNet-5 and VGGNet-8. In VGGNet-8, the classification accuracy of most categories in Auto-PASGS is better than that in General. Therefore, OA has been improved by 0.56%, AA has been improved by 0.34%, and Kappa has been improved by 0.0076 after using the Auto-PASGS method. By contrast, the classification accuracy of LeNet-5 is not better than that of VGGNet-8, and the classification accuracy of some categories in Auto-PASGS is better than that in General. Nevertheless, OA has been improved by 0.81%, AA has been improved by 0.2%, Kappa has been improved by 0.0094 after using the Auto-PASGS method.

V. Conclusion

In this article, we proposed a novel classification method for PolSAR classification with noisy labels. First, the probability-aware distribution is implemented to separate noisy labels from training samples. Then, a softmax loss strategy is designed to suppress the influence of the noise samples on DNN-based classifiers according to the probability-aware distribution diagram. Two sample grading strategies, namely, PASGS and Auto-PASGS, are developed to improve the classification performance when the training samples contain noisy labels. The former approach needs to set the noise ratio in advance and can reach a classification result similar to the classification accuracy with clear training samples. The latter approach does not need any prior information of training samples and can also obtain a superior classification accuracy. Finally, two PolSAR datasets are employed to prove the effectiveness of our proposed method with different noise ratios for labels. In addition, three hypothetical noise scenarios are implemented based on Flevoland I dataset, and two applications (i.e., unsupervised labels and temporal PolSAR images) are realized to prove the practicality of our proposed method. Two DNN-based classifiers and two noise-tolerant methods are exploited to demonstrate the effectiveness of our methods from different aspects.

Acknowledgment

The authors would like to thank the European Space Agency (ESA) for providing the temporal dataset through the AgriSAR 2009 campaign and PolSARPro software, and also like to thanks the NASA/JPL for providing the AIRSAR data.

References

[1] F. Ma, F. Zhang, Q. Yin, D. Xiang, and Y. Zhou, "Fast SAR image segmentation with deep task-specific superpixel sampling and soft graph convolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Sep. 2021.

[2] F. Zhang, Y. Wang, J. Ni, Y. Zhou, and W. Hu, "SAR target small sample recognition based on CNN cascaded features and adaboost rotation forest," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1008–1012, Jun. 2019.

[3] Y. Sun, X. Sun, Z. Wang, and K. Fu, "Oriented ship detection based on strong scattering points network in large-scale SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5218018, doi: 10.1109/TGRS.2021.3130117.

[4] G. Pajares, C. Lopez-Martinez, and S. Ma, "Improving wishart classification of polarimetric SAR data using the hopfield neural network optimization approach," *Remote Sens.*, vol. 4, no. 11, pp. 3571–3595, 2012.

[5] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.

[6] J. Ni, F. Zhang, Q. Yin, Y. Zhou, H. C. Li, and W. Hong, "Random neighbor pixel-block-based deep recurrent learning for polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 1–13, Sep. 2020.

[7] J. Ni, F. Zhang, F. Ma, Q. Yin, and D. Xiang, "Random region matting for the high-resolution polsar image semantic segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3040–3051, Feb. 2021.

[8] J. Cheng, F. Zhang, D. Xiang, Q. Yin, and Y. Zhou, "Polsar image classification with multiscale superpixel-based graph convolutional network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5209314, doi: 10.1109/TGRS.2021.3079438.

[9] H. Bi, F. Xu, Z. Wei, Y. Xue, and Z. Xu, "An active deep learning approach for minimally supervised polsar image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9378–9395, Nov. 2019.

[10] J. Tang, D. Xiang, F. Zhang, F. Ma, Y. Zhou, and H. Li, "Incremental SAR automatic target recognition with error correction and high plasticity," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1327–1339, Jan. 2022.

[11] B. Tu, W. Kuang, W. He, G. Zhang, and Y. Peng, "Robust learning of mislabeled training samples for remote sensing image scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5623–5639, Sep. 2020.

[12] E. Malach and S. Shalev-Shwartz, "Decoupling 'when to update' from 'how to update'," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 961–971.

[13] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.

[14] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 41–48.

[15] G. Patrini, A. Rozza, A. Krishna Menon, R. Nock, and L. Qu, "Making deep neural networks robust to label noise: A loss correction approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1944–1952.

[16] Y. Xu *et al.*, "Dual-channel residual network for hyperspectral image classification with noisy labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, Mar. 2021.

[17] A. Veit, N. Alldrin, G. Chechik, I. Krasin, A. Gupta, and S. Belongie, "Learning from noisy large-scale datasets with minimal supervision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 839–847.

[18] A. Ghosh, H. Kumar, and P. Sastry, "Robust loss functions under label noise for deep neural networks," in *Proc. AAAI Conf. Artif. Intell.*, pp. 1919–1925, 2017.

[19] D. Tanaka, D. Ikami, T. Yamasaki, and K. Aizawa, "Joint optimization framework for learning with noisy labels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 5552–5560.

[20] J. Han, P. Luo, and X. Wang, "Deep self-learning from noisy labels," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 5138–5147.

[21] J. Yao, H. Wu, Y. Zhang, I. W. Tsang, and J. Sun, "Safeguarded dynamic label regression for noisy supervision," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 9103–9110.

[22] W. Hu, Q. H. Zhao, Y. Huang, and F. Zhang, "P-Diff: Learning classifier with noisy labels based on probability difference distributions," in *Proc. 25th Int. Conf. Pattern Recognit.*, 2021, pp. 1882–1889.

[23] J. Zhao, W. Guo, B. Liu, Z. Zhang, W. Yu, and S. Cui, "Preliminary exploration of SAR image land cover classification with noisy labels," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 3274–3277.

[24] C. Wang *et al.*, "Label noise modeling and correction via loss curve fitting for SAR atr," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–10, Oct. 2021.

[25] R. Shang, J. Lin, L. Jiao, and Y. Li, "SAR image segmentation using region smoothing and label correction," *Remote Sens.*, vol. 12, no. 5, 2020, Art. no. 803.

[26] Z. Huang, C. O. Dumitru, Z. Pan, B. Lei, and M. Datcu, "Classification of large-scale high-resolution SAR images with deep transfer learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 1, pp. 107–111, Jan. 2020.

[27] B. Hou, Q. Wu, Z. Wen, and L. Jiao, "Robust semisupervised classification for polsar image with noisy labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6440–6455, Nov. 2017.
[28] X. Kang, P. Duan, X. Xiang, S. Li, and J. A. Benediktsson, "Detection and correction of mislabeled training samples for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5673–5686, Oct. 2018.
[29] J. Jiang, J. Ma, Z. Wang, C. Chen, and X. Liu, "Hyperspectral image classification in the presence of noisy labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 851–865, Feb. 2019.
[30] B. Tu, C. Zhou, W. Kuang, L. Guo, and X. Ou, "Hyperspectral imagery noisy label detection by spectral angle local outlier factor," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 9, pp. 1417–1421, Sep. 2018.
[31] B. Tu, X. Zhang, X. Kang, G. Zhang, and S. Li, "Density peak-based noisy label detection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1573–1584, Mar. 2019.
[32] B. Tu, X. Zhang, X. Kang, J. Wang, and J. A. Benediktsson, "Spatial density peak clustering for hyperspectral image classification with noisy labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5085–5097, Jul. 2019.
[33] B. Tu, C. Zhou, D. He, S. Huang, and A. Plaza, "Hyperspectral classification with noisy label detection via superpixel-to-pixel weighting distance," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4116–4131, Jun. 2020.
[34] J. Li, Q. Yuan, H. Shen, and L. Zhang, "Noise removal from hyperspectral image with joint spectral cspatial distributed sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5425–5439, Sep. 2016.
[35] Y. Xu *et al.*, "Dual-channel residual network for hyperspectral image classification with noisy labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, Mar. 2021.
[36] X. Nie, H. Qiao, B. Zhang, and X. Huang, "A nonlocal tv-based variational method for polsar data speckle reduction," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2620–2634, Jun. 2016.
[37] S. Foucher and C. Lopez-Martinez, "Analysis, evaluation, and comparison of polarimetric SAR speckle filtering techniques," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1751–1764, Apr. 2014.
[38] X. Nie, B. Zhang, Y. Chen, and H. Qiao, "A new algorithm for optimizing TV-based polsar despeckling model," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1409–1413, Oct. 2016.
[39] X. Nie, H. Qiao, and B. Zhang, "A variational model for polsar data speckle reduction based on the Wishart distribution," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1209–1222, Apr. 2015.
[40] L. Jiang, Z. Zhou, T. Leung, L.-J. Li, and L. Fei-Fei, "Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2304–2313.
[41] B. Han *et al.*, "Co-teaching: Robust training of deep neural networks with extremely noisy labels," in *Proc. 32nd Conf. Neural Inf. Process. Syst.*, 2018.
[42] W. Hu, Y. Huang, F. Zhang, and R. Li, "Noise-tolerant paradigm for training face recognition CNNS," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11879–11888.
[43] Y. Wang, J. Cheng, Y. Zhou, F. Zhang, and Q. Yin, "A multichannel fusion convolutional neural network based on scattering mechanism for polsar image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Jan. 2021.
[44] J. Ni, F. Zhang, Q. Yin, and H. Li, "Robust weighting nearest regularized subspace classifier for polsar imagery," *IEEE Signal Process. Lett.*, vol. 26, no. 10, pp. 1496–1500, Oct. 2019.
[45] T. Liu and D. Tao, "Classification with noisy labels by importance reweighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 447–461, Mar. 2016.
[46] D. Arpit *et al.*, "A closer look at memorization in deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 233–242.
[47] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.
[48] H. Wei, L. Feng, X. Chen, and B. An, "Combating noisy labels by agreement: A joint training method with co-regularization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 13726–13735.
[49] L. Ferro-Famil, E. Pottier, and J.-S. Lee, "Unsupervised classification of multifrequency and fully polarimetric SAR images based on the h/a/alpha-Wishart classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2332–2342, Nov. 2001.
[50] E. Pottier and L. Ferro-Famil, "Polsarpro v5. 0: An esa educational toolbox used for self-education in the field of polsar and pol-insar data analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, pp. 7377–7380.

Jun Ni (Student Member, IEEE) received the B.E. degree in electronic and information engineering in 2016 from the Beijing University of Chemical Technology, Beijing, China, where he is currently working toward the Ph.D. degree.

His research interests include the remote sensing image processing, deep learning, and parallel computing.

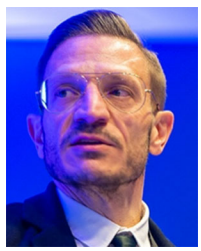
Deliang Xiang (Member, IEEE) received the B.S. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2010, the M.S. degree in photogrammetry and remote sensing from the National University of Defense Technology, Changsha, China, in 2012, and the Ph.D. degree in geoinformatics from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2016.

Since 2020, he has been a Full Professor with the Beijing University of Chemical Technology, Beijing, China. His research interests include urban remote sensing, synthetic aperture radar (SAR)/polarimetric SAR image processing, artificial intelligence, and pattern recognition.

Dr. Xiang is a Reviewer for *Remote Sensing of Environment, ISPRS Journal of Photogrammetry and Remote Sensing*, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATION AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and several other international journals in the remote sensing field. In 2019, he was awarded a Humboldt Research Fellowship.

Zhiyuan Lin received the B.E. degree in communication engineering in 2020 from the Beijing University of Chemical Technology, where he is currently working toward the M.S. degree.

His research focuses on regional classification of polarized time series SAR data.



Carlos López-Martínez (Senior Member, IEEE) received the M.Sc. degree in electrical engineering and the Ph.D. degree in remote sensing from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 1999 and 2003, respectively, and the Postgraduate Diploma in data science and big data from the Universitat de Barcelona, Barcelona, in 2021.

He is currently an Associate Professor in the area of remote sensing and microwave technology with UPC. He has a large professional international experience at DLR, Germany, at the University of Rennes 1 (France), and as a group leader of the Remote Sensing and Natural Resources Modelling Team at the Luxembourg Institute of Science and Technology (Luxembourg). He has broad academic teaching experience from bachelor, master, and Ph.D. levels to advanced technical tutorials presented at international conferences and space and research institutions worldwide. He has authored more than 200 articles in journals, books, and conference proceedings. His research interests include synthetic aperture radar (SAR) theory, statistics and applications, multidimensional SAR, radar polarimetry, physical parameter inversion, advanced digital signal processing, estimation theory, and harmonic analysis.

Dr. López-Martínez was the recipient of the EUSAR 2002 Conference Student Prize Paper Award, the EUSAR 2012 Conference First Place Student Paper Award, as a coauthor, and the IEEE-GRSS 2013 GOLD Early Career Award. He is an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING and *MDPI Remote Sensing*, acting also as invited Guest Editor for several special issues. He has collaborated in the Spanish PAZ and the ESA's SAOCOM-CS missions and in the proposal of the Parsifal mission. He is Member of the ESA's Sentinel ROSE-L Mission Advisory Group. He was appointed Vice-President of the IEEE-GRSS Spanish Chapter, and in 2016, he became its Secretary and Treasurer. Since 2011, he has been collaborating with the IEEE-GRSS Globalization initiative in Latin America, contributing to the creation of the IEEE-GRSS Chilean Chapter and the organization of the 2020 LAGIRSS conference, being appointed as Latin America liaison in 2019. He is also the Co-Chair of the Tutorial Technical Committee of the Indian 2020 and 2021 InGARSS conferences.

Wei Hu received the B.S. and M.S. degrees from the Dalian University of Science and Technology, Dalian, China, in 1999 and 2002, respectively, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2006, all in computer science.

He is currently an Associate Professor of Computer Science with the Beijing University of Chemical Technology, Beijing, China. His research interests include computer vision, computer graphics, and scientific visualization.

Dr. Hu is a Reviewer for the IEEE TRANSACTIONS ON VISUALIZATIONS AND COMPUTER GRAPHICS, *EuroGraphics*, and *Pacific Graphics*.

Fan Zhang (Senior Member, IEEE) received the B.E. degree in communication engineering from the Civil Aviation University of China, Tianjin, China, in 2002, the M.S. degree in signal and information processing from Beihang University, Beijing, in 2005, and the Ph.D. degree in signal and information processing from the Institute of Electronics, Chinese Academy of Sciences, Beijing, in 2008.

He is currently a Full Professor in Electronic and Information Engineering with the College of Information Science and Technology, and the Interdisciplinary Research Center for Artificial Intelligence, Beijing University of Chemical Technology, Beijing. His research interests include remote sensing image processing, high-performance computing, and artificial intelligence.

Dr. Zhang is an Associate Editor for IEEE ACCESS and a Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and *Journal of Radars*.