Reg-Superpixel Guided Convolutional Neural Network of PolSAR Image Classification Based on Feature Selection and Receptive Field Reconstruction

Ronghua Shang, *Senior Member, IEEE*, Keyao Zhu, Jie Feng, *Member, IEEE*, Chao Wang, Licheng Jiao, *Fellow, IEEE*, and Songhua Xu

***Abstract*—The convolutional neural network (CNN) has a poor performance in nonuniform and edge regions due to the limitation of fixed receptive field. At the same time, feature stacking of input data can bring burden and overfitting to the network. To solve these problems, this article proposes a reg-superpixel guided CNN based on feature selection and receptive field reconstruction. First, a feature selection method is designed, which uses polarimetric SAR statistical distribution features to calculate distance and similarity, and selects features that are easier to identify to avoid the negative impact of low distinguishing features on classification. Second, the reg-superpixel, which means regular superpixel, is used to reconstruct the receptive field and represent the features of the central pixel. The classification result of the central pixel is extended to the whole superpixel during the test. This method can extend the pixel-level CNN network to superpixel-level. Finally, by using edge information of the small-scale superpixel and spatial information of the large-scale superpixel to adjust receptive field of the central pixel, the classification results with uniform smooth region and high edge fitting can be generated. Experimental results with four state-of-the-art methods on four datasets show that feature selection and multiscale reg-superpixel network is effective for polarimetric SAR classification problems.**

***Index Terms*—Feature selection, image classification, polarimetric decomposition, receptive field remodeling.**

I. Introduction

POLARIMETRIC synthetic aperture radar (PolSAR), as an advanced active microwave imaging technology, has been widely used in disaster monitoring, military reconnaissance, agricultural evaluation, and other fields [1]. PolSAR can obtain multiple channel scattering information, and extract important features, such as volume scattering, even scattering, and surface scattering [2]. These scattering features are also used to improve the accuracy of image classification.

Remote sensing image interpretation methods include image denoising [3], change detection [4], image segmentation [5], and image classification. The semisupervised classification methods for PolSAR images can be divided into the following two steps: feature extraction and classifier design. Recently, many research works used the polarimetric decomposition components as the polarimetric features, which were generated by Pauli [6], Krogager [7], Cloud [8], Yamaguchi [9], and other methods. But using all these features affected the efficiency of the network, and some of them also contained noise and redundant information. Therefore, some methods used feature extraction or feature fusion for dimension reduction. Ren et al. [10] proposed a feature extraction method based on sparse subspace clustering, which constructed the projection matrix through subspace clustering, and then used it to project features. Song et al. [11] proposed a feature fusion method based on the CK-HDRF model, which could well maintain the position of classification edges. However, a drawback still existed in these methods that the extracted features lose the original data distribution and even some crucial information [12]. Therefore, other methods tended to use feature selection for dimension reduction. Dong et al. [13] proposed an attention-based polarimetric feature selection network, which obtained the relationship between the polarimetric features and ensure the effectiveness of classification. Liu et al. [14] proposed a neural architecture search method for polarimetric feature selection, which adaptively found the importance of polarimetric features during network training by using the new stem layer. Yang et al. [12] proposed a polarimetric feature selection method based on convolutional neural network (CNN), taking KLD distance as an evaluation to train CNN network. However, this method requires more training samples, which are usually limited in PolSAR data. In addition, Bai et al. [15] used an improved sparse support vector machine to select polarimetric features, and could automatically select a subset of polarimetric features suitable for all classes. Banerjee et al. [16] proposed a mutual information (MI) based polarimetric feature selection method, which used the power of MI to select the

Manuscript received 28 December 2022; revised 4 March 2023; accepted 8 April 2023. Date of publication 27 April 2023; date of current version 15 May 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62176200 and Grant 62271374, in part by the Natural Science Basic Research Program of Shaanxi under Grant 2022JC-45 and Grant 2022JQ-616, in part by the Open Research Projects of Zhejiang Lab under Grant 2021KG0AB03, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515110686, in part by the Research Project of SongShan Laboratory under Grant YYJC052022004, and in part by the Fundamental Research Funds for the Central Universities under Grant XJS221903. *(Corresponding author: Keyao Zhu.)*

Ronghua Shang, Keyao Zhu, Jie Feng, and Licheng Jiao are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: rhshang@mail.xidian.edu.cn; keyao_zhu@stu.xidian.edu.cn; jiefeng@xidian.edu.cn; lchjiao@mail.xidian.edu.cn).

Chao Wang is with the Research Center for Big Data Intelligence, Zhejiang Laboratory, Hangzhou 311100, China (e-mail: wangc@zhejianglab.com).

Songhua Xu is with the Institute of Medical Artiffcial Intelligence, The Second Afffliated Hospital of Xi'an Jiaotong University, Xi'an 710004, China (e-mail: songhuaxu@mail.xjtu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3268177

best set of classification parameters. Hariharan et al. [17] proposed a random forest (RF) based polarimetric feature selection technique to identify features that vary significantly with crop phenology. Huang et al. [18] proposed a multiview polarimetric feature selection method, carried out sparse regularization of the feature matrix, and built an optimization model to achieve feature selection. Yang et al. [19] proposed a polarimetric feature selection method, which used sparse variational autoencoder reconstruction errors to evaluate feature representation ability. However, none of these methods take into account the statistical characteristics of PolSAR data. Therefore, in order to make full use of limited samples and polarimetric information in network learning, a feature selection method for PolSAR images is needed.

In terms of classifiers design for PolSAR images, first, the traditional algorithms represented by Wishart distribution [20] and polarimetric K distribution [21] use polarimetric statistical distribution to model PolSAR images. These methods have poor fitting ability for more complex PolSAR images. Second, deep learning methods represented by CNN also achieved good results. Zhou et al. [22] first applied CNN to PolSAR classification task and achieved good classification results. On this basis, Zhang et al. [23] proposed complex-valued convolutional neural network (CV-CNN) according to the complex data features of the coherence matrix. In addition, various CNN-derived networks are also constantly proposed. Howard et al. [24] proposed a deep separable convolution to reduce the number of parameters of deep neural networks and construct lightweight deep neural networks. Huang et al. [25] proposed a densely connected structure to reduce the loss of shallow features by directly connecting the front and back layers. However, the network of CNN structure is classified pixel by pixel and usually takes rectangular window as input data. Therefore, classification errors are often generated in the edge area and the homogeneous area with large noise.

The methods based on superpixel can solve this problem well. The method based on superpixel [26] can take into account the background information and local structure, and also reduce the demand for training samples. Cheng et al. [27] proposed a method based on multiscale superpixel graph convolution network (MSSP-GCN), which used graph convolution network to classify small scale superpixels and used large scale superpixels to modify the classification results. In order to reduce classification errors, Guo et al. [28] proposed the concept of fuzzy superpixels, which were divided into mixed superpixels and pure superpixels, and the generation of mixed superpixels was reduced as much as possible during clustering. On this basis, Guo et al. [29] proposed an adaptive fuzzy superpixel generation method in order to reduce the influence of the proportion of undetermined pixels on the generation effect of fuzzy superpixels. Meanwhile, Guo et al. [30] applied fuzzy superpixels to network learning to achieve more accurate classification. In addition, Qin et al. [31] used a regularized superpixel input into the trained CV-CNN network to reduce the running time, but the improvement of classification accuracy is very limited because the training data is still pixel-level data.

The abovementioned classification networks based on superpixel have achieved good results on PolSAR images. However, due to its poor structural flexibility and scalability, the superpixel-based GCN network is difficult to train for the features of PolSAR data. Although CNN structure network has high flexibility, it is difficult to classify superpixels without fixed shape. To solve these problems, this article proposes a method based on feature selection and multiscale reg-superpixel network (FS-MSRSnet). First, three PolSAR decomposition methods are used to extract the decomposition features. And the supervised feature selection method Pol-ReliefF combining polarization distance and ratio dissimilarity is designed. According to the special statistical features of PolSAR, the polarimetric statistical distance is used to calculate the distance, and the average ratio is used to calculate the dissimilarity between samples. In this article, a small number of samples are selected as training samples of Pol-ReliefF. Second, in order to make the receptive field of CNN not contain heterogeneous regions, the reg-superpixel guided dense connection subnetwork (RSDSnet) is implemented. The reg-superpixel is a superpixel with the consistent shape and size generated by the regularization method. It is used to input the CNN structure network RSDSnet that only accepts Euclidean structure data, and adjust the receptive field. In RSDSnet, dense connections are also realized through regularized feature map, and then the classification results are extended to the whole superpixel. Finally, in order to make full use of different scale superpixel information, a branch is added to RSDSnet to form multiscale reg-superpixel network (MSRSnet). MSRSnet is fused in the output layer to obtain the central pixel classification result that integrates the information of the two scales. The main contributions of the proposed FS-MSRSnet are as follows.

1) The Pol-ReliefF method for feature selection is proposed. This method makes full use of the limited training set, reduces the influence of weak discriminative features on network learning, and avoids the overfitting phenomenon to improve the network efficiency.
2) Reg-superpixel is used to reshape the receptive field and realize the superpixel-level CNN structure network RSDSnet. The classification accuracy is greatly improved by using the features of high edge fit and homogenous region of the superpixel.
3) MSRSnet is formed by adding superpixels of two scales on the basis of 2). It can combine the spatial context information of large-scale superpixel with the details and edge information of small-scale superpixel at the same time to improve the classification accuracy of superpixels.

The rest of this article is organized as follows. The Section II introduces the proposed FS-MSRSnet method in detail. In Section III, parameter analysis, experimental results analysis, and ablation experiment are carried out. Finally, Section IV concludes this article.

II. Methodology

In order to effectively use the decomposition features of PolSAR images and the advantages of homogeneity in the superpixel region to improve the classification accuracy, this article proposes a reg-superpixel guided CNN based on feature selection and receptive field reconstruction, as shown in Fig. 1.

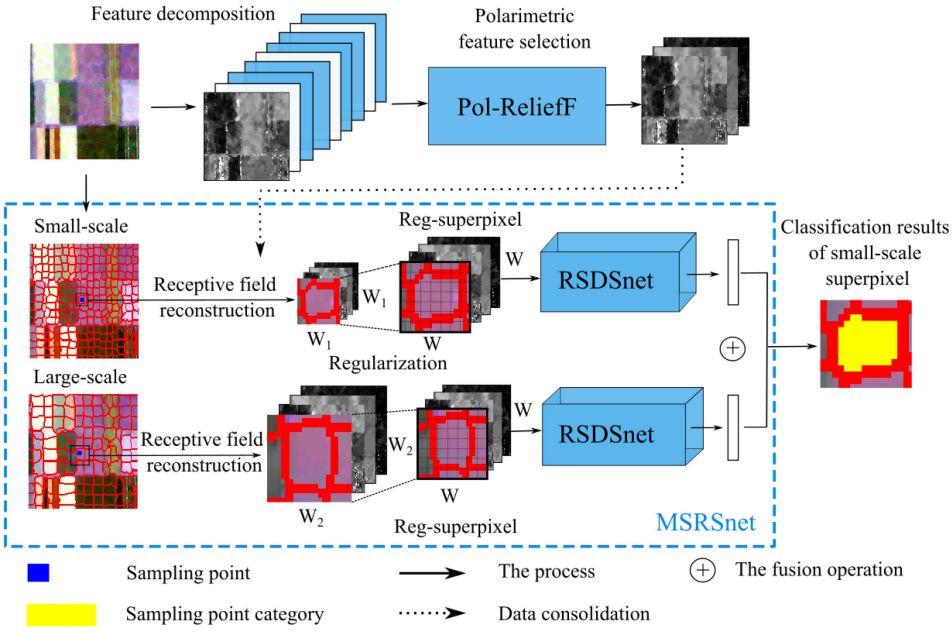

Fig. 1. Structure of PolSAR classification network FS-MSRSnet.

The overall structure of the proposed FS-MSRSnet consists of three parts. Polarimetric feature selection method Pol-ReliefF, reg-superpixel dense connected subnetwork based on receptive field reconstruction (RSDSnet), and multiscale reg-superpixel network (MSRSnet). The specific structure is shown in the blue dashed box. First, the traditional incoherent decomposition method is used to decompose the features of PolSAR images, and then the Pol-ReliefF method is used to select features by combining the features of the coherence matrix. Second, the superpixels of two scales are used to segment the image, and the superpixels of the sampling points are extracted and combined with the selected features. Then, regularization method is used to generate reg-superpixel, which is used to reconstruct the receptive field and input into RSDSnet network. Finally, the regularized superpixels of both sizes are input into RSDSnet network and fused on the output layer to obtain the classification results of small-scale superpixels.

A. *Pol-ReliefF for Polarimetric Feature Selection*

PolSAR is different from optical image and ordinary SAR image because of its unique radar signal receiving and transmitting mode. Polarimetric scattering matrix $\mathbf{S}$ is needed to describe the polarization information of different channels. However, due to the imaging features of PolSAR, one pixel contains information of multiple scattering centers, so only using the polarimetric scattering matrix $\mathbf{S}$ is not enough to support the network for learning. In order to analyze this kind of statistical scattering in more detail, the polarimetric covariance matrix $\mathbf{C}$ [27], or polarimetric coherence matrix $\mathbf{T}$ [22] is usually used instead of the original data. On this basis, some scholars also added some polarimetric decomposition features to make better use of the physical scattering features of polarization data.

However, although the abovementioned features can make full use of the limited PolSAR data and extract polarimetric scattering information more comprehensively, high-dimensional data imposes a burden on network training, reduces the learning efficiency, and is prone to overfitting. Therefore, a feature selection method is needed to select features with strong separateness to reduce the negative impact of weak distinguishing features on network learning. Based on this, a Pol-ReliefF method is proposed.

Before that, the feature construction process is as follows. The original data of fully polarized SAR are the polarization scattering matrix $\mathbf{S}$, which is a complex matrix as follows:

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

where S_{HH} represents that both transmitting and receiving are horizontally polarized electromagnetic waves, S_{HV} represents transmitting horizontally polarized electromagnetic waves and receiving vertically polarized electromagnetic waves, the S_{VH}, S_{VV} are similar. When the single station SAR satisfies the reciprocity theorem, $S_{\mathrm{HV}} = S_{\mathrm{VH}}$, at this moment $\mathbf{S}$ is a symmetric matrix. In this article, the polarization coherence matrix $\mathbf{T}$ is used to analyze the image, which can be expressed as

$$\mathbf{T} = \mathbf{K}_p \mathbf{K}_p^{*T} = \begin{bmatrix} T_{11} & T_{12}^* & T_{13}^* \\ T_{21}^* & T_{22} & T_{23}^* \\ T_{31}^* & T_{32}^* & T_{33} \end{bmatrix} \tag{2}$$

where $\mathbf{K}_p = 1/\sqrt{2}[S_{\mathrm{HH}} + S_{\mathrm{VV}} S_{\mathrm{HH}} - S_{\mathrm{VV}} 2 S_{\mathrm{HV}}]^T$ is the vectorization of scattering matrix $\mathbf{S}$, and $\langle * \rangle$ is the complex number. In the polarization coherence matrix, T_{11}, T_{22}, and T_{33} are real numbers and the remaining terms are complex numbers. Since the real number network is used in this article, the complex terms of the original polarization coherence matrix can be divided into real and imaginary parts as different input features. For

TABLE I
FOUR TYPES OF FEATURES USED IN FS-MSRSNET

Features	Types of features
T_{11}, $T_{12}{}^*$, $T_{13}{}^*$, T_{22}, $T_{23}{}^*$, T_{33}	Original coherence matrix scattering features
H, A, α	Cloude decomposition features
P_c, P_s, P_d, P_v	Yamaguchi decomposition features
SPAN	Roll-invariant feature

example, T_{12}^* can be split into T_{12}^{real} and T_{12}^{imag}. And according to the reciprocity theorem $T_{ij} = T_{ji}, i, j = 1, 2, 3$. Therefore, after deleting some identical features, the final features of each sample can be written as the following vector:

$$\mathbf{F}_0 = \left[T_{11}\ T_{12}^{\text{real}}\ T_{13}^{\text{real}}\ T_{22}\ T_{23}^{\text{real}}\ T_{33}\ T_{12}^{\text{imag}}\ T_{13}^{\text{imag}}\ T_{23}^{\text{imag}}\right]. \tag{3}$$

The polarization coherence matrix $\mathbf{T}$ can well describe the statistical scattering features of PolSAR images. However, for PolSAR data with limited labeled samples, only using $\mathbf{T}$ input network for learning is equivalent to ignoring part of the physical scattering features implied by fully polarized data. These physical information need to be obtained by polarization feature composition of $\mathbf{T}$. The four types of features used in this article are shown in Table I.

The polarization scattering matrix $\mathbf{S}$ of the original PolSAR image is decomposed into coherence matrix $\mathbf{T}$, and three different decomposition features are added. In the second row of the Table I, Cloude decomposition is used, which is a classical incoherent polarization feature composition method. It can extract the scattering entropy H, scattering anisotropy A, and average angle α of the target from the coherence matrix. Yamaguchi decomposition in the third row of the Table I provides a four-component scattering model, which can well distinguish different ground object types, such as volume scattering P_v, double-bounce scattering P_d, surface scattering P_s, and cross-polarized scattering P_c. Finally, SPAN describes a roll-invariant feature, and the abovementioned contents are shown in formula (4) to form the final original 17-dimensional feature

$$\mathbf{F} = \left[\mathbf{F}_0\ \ H\ \ A\ \ \text{alpha}\ \ P_c\ \ P_s\ \ P_d\ \ P_v\ \ \text{SPAN}\right]. \tag{4}$$

In order to make full use of the limited training samples of PolSAR images to select more separated features, this article proposes a Pol-ReliefF method for polarimetric data, and the details are as follows.

Suppose that some samples are selected from all samples of PolSAR data as the training set Q. In order to ensure the balance of training samples, m samples are randomly sampled from each class of the training set $Q = \{\mathbf{X}_i^c\}$ as the central sample set $\text{Cen} = \{\mathbf{X}_j^c\}$ for feature selection, $c = 1,..., C$. Where c is the number of classes, i is the number of samples of each class in training set, $i = 1,..., n$, j is the number of samples in each class of central samples, $j = 1,..., m$.

Then, for each central sample $\mathbf{X}_j^c$, find K-neighbor samples $H = \{\mathbf{X}_k^c\}$ of the same class as the central sample, $k = 1, ..., K$, and in $C - 1$ sample sets of different classes, find K-neighbor samples m of each class. $\mathbf{X}_k^{\overline{c}}$ represents the kth sample that does not belong to class C. Due to the different imaging modes between PolSAR data and ordinary optical images, Wishart distance is usually used to calculate the statistical distance of polarimetric data. The ordinary Wishart distance is an approximate distance and may have a negative value. In order to measure the distance between samples more accurately, the modified Wishart distance can be used for calculation. Among them, since the central sample set $\text{Cen} = \{\mathbf{X}_j^c\}$ is obtained by sampling from the training set Q, each sample $\mathbf{X}_j^c$ in the central sample set is a rectangular matrix. So for the Ith sample in the sample set Cen, $R_{\text{Cen}(I)}$ represents the average value of the sample matrix

$$\begin{aligned} R_{\text{Cen}(I)} &= \text{avg}(\mathbf{X}_I), \quad \mathbf{X}_I \in \text{Cen} \\ R_{H(I)} &= \text{avg}(\mathbf{X}_I), \quad \mathbf{X}_I \in H \\ R_{M_c(I)} &= \text{avg}(\mathbf{X}_I), \quad \mathbf{X}_I \in M_c \end{aligned} \tag{5}$$

where $\text{avg}(\cdot)$ is the average value of the sample matrix, that is, the sum of all elements of the matrix and then divided by the size of the matrix. Therefore, the distance between the Ith center sample $R_{\text{Cen}(I)}$ and the Jth similar sample $R_{H(J)}$ can be calculated as follows:

$$\begin{aligned} D(R_{\text{Cen}(I)}, R_{H(J)}) = &\left(\text{tr}\left(T\left[R_{\text{Cen}(I)}\right]^{-1} T\left[R_{H(J)}\right]\right)\right. \\ &\left. + \text{tr}(T\left[R_{H(J)}\right]^{-1} T\left[R_{\text{Cen}(I)}\right])\right) - n. \end{aligned} \tag{6}$$

It is the same with the calculation of the distance between samples of different classes, where $T[R_{\text{Cen}(I)}]$ is the coherence matrix for calculating the central sample $R_{\text{Cen}(I)}$. Only the features $\mathbf{F}_0$ of the coherence matrix $\mathbf{T}$ are used in the distance calculation, while all features $\mathbf{F}$ are needed in the later calculation of the similarity. After finding K-neighbor samples of the same and different classes, first, for each feature f in feature $\mathbf{F}$, the weight is initialized to 0. Then, the central sample set Cen is taken as the training set, and the weight of each feature f is trained and updated using the weight update formula (7):

$$\begin{aligned} w_f = w_f &- \frac{\sum_{J=1}^{K} \text{diff}\left(f, R_{\text{Cen}(I)}, R_{H(J)}\right)}{K} \\ &+ \sum_{c=1}^{C} \frac{\frac{P(c)}{1-P(\text{Cen})} \sum_{J=1}^{K} \text{diff}\left(f, R_{\text{Cen}(I)}, R_{M_c(J)}\right)}{K} \end{aligned} \tag{7}$$

$$\begin{aligned} &\text{diff}\left(f, R_{\text{Cen}(I)}, R_{H(J)}\right) = 1 \\ &- \min\left\{\frac{N\left(R_{\text{Cen}(I)}\right)}{N\left(R_{H(J)}\right)}, \frac{N\left(R_{H(J)}\right)}{N\left(R_{\text{Cen}(I)}\right)}\right\} \end{aligned} \tag{8}$$

where $P(c)$ and $P(\text{Cen})$ are the proportions of class C sample M_c and center sample Cen in the whole training set Q. $\text{diff}\left(f, R_{\text{Cen}(I)}, R_{H(J)}\right)$ is the dissimilarity between the Ith central sample $R_{\text{Cen}(I)}$ and the Jth similar sample $R_{H(J)}$, as shown in formula (8). The effect of PolSAR multiplicative speckle noise can be reduced by calculating the ratio of the normalized mean values of the two sample areas. As can be seen from formula (8), when two samples are more similar, the value of $\text{diff}\left(f, R_{\text{Cen}(I)}, R_{H(J)}\right)$ is also smaller. Since there are negative values in the data of polarimetric decomposition features, $N(\cdot)$

Algorithm 1: Pol-ReliefF.

Input: The training set Q divided by the original data $\mathbf{X}_{N\times f}$, the number of features c to be selected, the k-nearest neighbor algorithm parameter k, and the number of central samples m of each class.
Output: The training set Q' after dimensionality reduction.

1. Training set Q is randomly sampled and m center samples are selected from each category to form center sample set $\mathrm{Cen} = \{\mathbf{X}_i\}$, in which the total number of center samples is n and the feature number is f.
2. Initialize the weight matrix $\mathbf{W}_{n\times f} = 0$.
3. **for** $i \leftarrow 1$ to n **do**
4. Samples of the same and different classes of $\mathbf{X}_i$ are selected, and formula (5) is used to calculate $R_{\mathrm{Cen}(i)}$ and its corresponding R_H, R_{M_c}.
5. $H \leftarrow neighbors(R_{\mathrm{Cen}(i)}, R_H, k)$ /*Function *neighbors uses (6)*/*
6. $M_c \leftarrow neighbors(R_{\mathrm{Cen}(i)}, R_{M_c}, k)$
7. **for** $j \leftarrow 1$ to f **do**
8. $\mathbf{W}(i,j) \leftarrow w_j$ /*w_j*can be calculated from (7)*/*
9. **end**
10. **end**
11. Calculate the average of each row of the weight matrix $\mathbf{W}_{n\times f}$ to generate the final weight vector, then sort it in ascending order, and select c features with the largest weight from $\mathbf{F}$ to form $\mathbf{F}'$. The new training set Q' is formed from these $\mathbf{F}'$ features.

is to normalize the features to eliminate the influence of negative values on the ratio results.

As can be seen from formula (7), the greater the similarity between the central sample and the nearest K similar samples, the faster the weight increases. On the contrary, the greater the similarity with samples of different categories, the slower the weight increases. At the end of the training, the features with the largest weight of c are the most separable features. The process is shown in Algorithm 1.

B. RSDSnet Network Based on Reg-Superpixel Generation and Receptive Field Reconstruction

In order to use the homogeneity of superpixel to improve the classification accuracy of the network, and solve the problem that ordinary superpixel cannot input CNN structure network that require Euclidean structure data, a reg-superpixel method is proposed in this article. Unlike the traditional superpixel, this method converts non-Euclidian structured data into Euclidian structured data by using a regularization process to transform originally irregular superpixel into uniform shape and size. Reg-superpixel uses its homogeneity to enlarge and adjust the receiving field, and at the same time realizes the superpixel-level classification method of CNN structure network. Therefore, this method based on reg-superpixel has higher generalization ability and classification accuracy than the ordinary superpixel

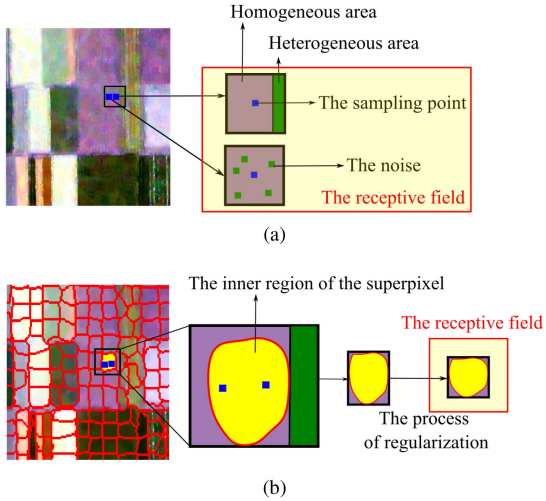

Fig. 2. Two methods of receptive field generation. (a) Use the rectangular area around the sampling point. (b) Use the superpixel where the sampling point is located.

classification based on traditional methods. And compared with the superpixel classification method based on GCN, this method has more flexible network construction and higher accuracy of superpixel classification.

Due to the difficulty of PolSAR image labeling, the whole image cannot be used to train the deep learning network [32]. Zhou et al. [33] regarded the field in the rectangular area around each pixel as an image, and constructed the training set by sampling pixels of PolSAR images. However, the receptive field used in this method treats all regions around the sampling point as homogeneous. Fig. 2 shows the process of generating receptive field by two methods. As shown in Fig. 2(a), the surrounding region near the sampling point (blue) of the edge region is partially homogeneous (purple) and partially heterogeneous (green). This sampling method takes both purple and green areas as receptive fields of sampling points, so classification errors are easy to occur in the edge areas. At the same time, the noise around the sampling point will also affect the classification. In order to reduce the influence of heterogeneous regions and noise, a regular superpixel training and testing method is designed in this article.

First, the improved SLIC method [34] is used to segment the PolSAR image into superpixels. Among them, in order to make the superpixel fit the edge better, modified Wishart distance is added on the basis of the traditional Euclidean distance for clustering

$$D_{\mathrm{srw}}(i,j) = \frac{1}{2}\left(\mathrm{Tr}\left(\hat{\mathbf{\Sigma}}_j^{-1}\mathbf{T}_i\right) + \mathrm{Tr}\left(\mathbf{T}_i^{-1}\hat{\mathbf{\Sigma}}_j\right)\right) - q \tag{9}$$

where $\hat{\mathbf{\Sigma}}_j$ is the average coherence matrix of class j, $\mathbf{T}_i$ is the coherence matrix of pixel i, and q is a constant, usually set as 3. The distance after adding D_{srw} is

$$D = \frac{D_{\mathrm{srw}}}{m} + \frac{D_{\mathrm{euc}}}{S} \tag{10}$$

where D_{euc} is Euclidean distance, S is the average distance between two adjacent superpixels, and m is the scale parameter.

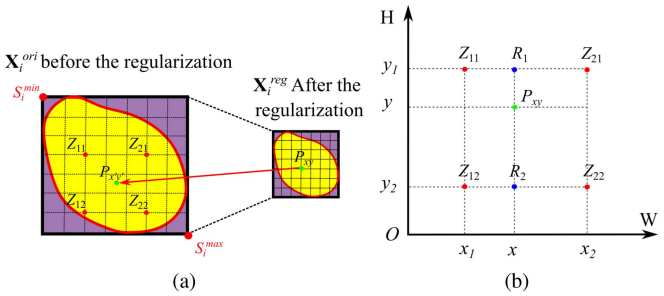

Fig. 3. Illustration of the reg-superpixel generation method. (a) Regularization process of superpixels. $Z_{ij}, i, j = 1, 2$ and P_{xy} are gray values. (b) Coordinate representation of regularization method.

In order to make the superpixel closer to the edge, m is set to be small in this article.

The method of using superpixel to generate receptive field is shown in Fig. 2(b). If the superpixel of the sampling point is taken as its receptive field, it can be guaranteed that for each sampling point, no matter where it is located, the receptive field can be limited by the shape of the superpixel. Meanwhile, the impact of noise can be suppressed because the sampling point category is mapped to the whole superpixel.

In addition, since the CNN-structured network requires input matrix data of a fixed size, a method is needed to map superpixels to another size, which called reg-superpixels. The specific process for reg-superpixel generation is shown in Fig. 3, where S_i is the superpixel to which the ith sampling point belongs in the training set Q selected by random sampling. As shown in Fig. 3(a), the maximum and minimum values of coordinates within S_i are selected, and two points $S_i^{\min}$ and $S_i^{\max}$ on the upper left and lower right are generated, through which the initial rectangle feature $\mathbf{X}_i^{\text{ori}}$ of superpixel is generated. Then, the following regularization method is used to adjust the size of this region, and the regularized $\mathbf{X}_i^{\text{reg}}$ is generated, so that the superpixels of different shapes can be input into the CNN structure network for training

$$\mathbf{X}_i^{\text{reg}} = \text{Reg}(\mathbf{X}_i^{\text{ori}}, L) \tag{11}$$

where $\text{Reg}(\cdot)$ is the regularized function and L is the regularized scale. The specific process of $\text{Reg}(\cdot)$ is shown in Fig. 3(a). P_{xy} represents the gray value of the point (x, y) in $\mathbf{X}_i^{\text{reg}}$, $P_{x'y'}$ represents the gray value of the point (x', y') projected by P_{xy} in $\mathbf{X}_i^{\text{ori}}$, four points Z_{11}, Z_{12}, Z_{21}, and Z_{22} represent the four gray values closest to $P_{x'y'}$, and the gray value of the target point P_{xy} can be obtained through the interpolation of the four points. To describe this process in more detail, Fig. 3(b) shows the coordinate system corresponding to the image, where the abscissa is the image width W and the ordinate is the image height H. First, the gray value Z_{ij} corresponding to the point (x_i, y_j) is used to calculate R_1 and R_2. Then, P_{xy} is calculated from R_1 and R_2. Formulas (12) and (13) can describe this process

$$P_{xy} = \frac{1}{k}\sum_{i=1}^{2}\sum_{j=1}^{2} Z_{ij}k_{ij} \tag{12}$$

$$k = (y_2 - y_1)(x_2 - x_1)$$

TABLE II
PARAMETER SETTINGS OF RSDSNET

Layer	Type/stride	Filter shape	Input size
0	Input	/	15×15×c
1	RSConv-1/s1	6×6×c	15×15×c
	RSConv-2/s1	1×1×c×W_1	15×15×c
2	Max Pool/s2	Pool 2×2	10×10×[c+W_1]
3	RSConv-1/s1	3×3×[c+W_1+(c+W_1)]	5×5×[c+W_1+(c+W_1)]
	RSConv-2/s1	1×1×[c+W_1+(c+W_1)]×W_2	5×5×[c+W_1+(c+W_1)]
4	RSConv-1/s1	3×3×[c+W_1+(c+W_1)+W_2]	3×3×[c+W_1+(c+W_1)+W_2]
	RSConv-2/s1	1×1×[c+W_1+(c+W_1)+W_2]×W_3	1×1×[c+W_1+(c+W_1)+W_2]
5	dropout-FC/s1	W_3×15	1×1×W_3
6	Softmax/s1	Classifier	1×1×15

$$k_{22} = (y - y_1)(x - x_1)\, k_{11} = (y_2 - y)(x_2 - x)$$
$$k_{21} = (y_2 - y)(x - x_1)\, k_{12} = (y - y_1)(x_2 - x) \tag{13}$$

where k, k_{ij} is the scaling ratio, $i, j = 1, 2$.

The main function of the $\mathbf{X}_i^{\text{reg}}$ is to use its edge to correct the receptive field of the sampling point. In addition, since the interior of the superpixel is regarded as a homogeneous region, the deformation generated by regularization can not affect the network training. Regularization is also required in densely connected networks to take advantage of shallow features. The input matrix $\mathbf{X}^{\text{reg}}$ extracts the deep features in the network through the formula (14)

$$\mathbf{X}^l = \text{RSConv}\left(\mathbf{W}, \mathbf{X}^{l-1}\right) + \sum_{k=0}^{l-1} \text{Reg}\left(\mathbf{X}^k, L\right) l = 1, 3$$
$$\mathbf{X}^l = \text{Pooling}\left(\mathbf{W}, \mathbf{X}^{l-1}\right) + \sum_{k=0}^{l-1} \text{Reg}\left(\mathbf{X}^k, L\right) l = 2 \tag{14}$$

where l is the number of layers, $\mathbf{X}^0 = \mathbf{X}^{\text{reg}}$, $\text{Reg}(\cdot)$ is the regularization function introduced above, $\mathbf{W}$ is the parameter matrix, $\text{RSConv}(\cdot)$ represents the separation convolution of regularized features to reduce the number of parameters. At the same time, for each convolution layer, the shallow features can be retained in the training through $\text{Reg}(\cdot)$ regularization. Finally, the extracted features pass through a layer of separation convolution $\text{RSConv}(\cdot)$ and full connection $\text{FC}(\cdot)$, and the softmax function is used to output the predicted results

$$\boldsymbol{Y} = \text{Softmax}\left(\text{FC}\left(\text{RSConv}\left(\mathbf{W}, \mathbf{X}^l\right)\right)\right) l = 4. \tag{15}$$

In addition, because the CNN network processes pixel-level data, the network trained with regular superpixel can still only predict the category of one pixel. In order to make full use of the homogeneity and edge information in the superpixel region, the predicted pixel category of network output is extended to the entire superpixel, that is, $y_S = y_i$, where $i \in S$. y_i is the category predicted by pixel point i through RSDSnet network, and y_S is the category of superpixel S. During the test, each superpixel is regularized and input into the network, and the abovementioned method is used to predict the category of superpixel, making RSDSnet a superpixel-level CNN structure network.

Table II shows the parameter settings of RSDSnet, and also presents the calculation process of the number of channels in

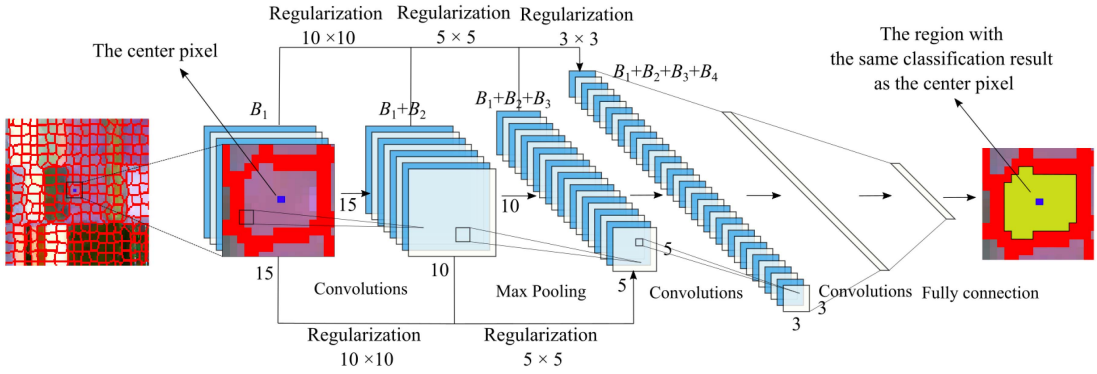

Fig. 4. Structure of RSDSnet based on reg-superpixel.

Algorithm 2: RSDSnet.

Input: The training set Q' after dimensionality reduction, the superpixel map $\mathbf{S}$, and the coordinate vector $\mathbf{E}_{1\times s}$ of superpixels, where s is the total number of superpixels.
Output: Superpixel classification result label $\mathbf{L}_{1\times s}$.

1. For each sample $\mathbf{X}_i^{ori}$ in Q', find its superpixel according to the superpixel map $\mathbf{S}$ and use formula (12) to generate the regularized sample $\mathbf{X}_i^{reg}$.
2. Input $\mathbf{X}_i^{reg}$ into RSDSnet network as shown in Fig. 4 for training.
3. **for** $i \leftarrow 1$ to s **do**
4. Find the superpixel i corresponding to $\mathbf{E}_i$ and regularize it.
5. The superpixel i is input into RSDSnet network for testing, and the predicted category is regarded as the classification result of the whole superpixel i.
7. **end**
8. Output the final classification results of superpixel $\mathbf{L}_{1\times s}$.

each layer. In Table II, RSConv-1 is depthwise convolution, and RSConv-2 is pointwise convolution. $c = 12$ is the number of input image channels after feature selection, $W_1 = 36$, $W_2 = 192$, and $W_3 = 288$ are the number of convolution kernels at the first, third, and fourth layers, respectively. Set the input size $L = 15$ based on the convolution kernel size. The deep learning network based on reg-superpixel is shown in Fig. 4.

In Fig. 4, $B_1 = c$ is feature number selected by Pol-ReliefF. $B_2 = W_1 = 36$, $B_3 = (c + W_1) = 53$, and $B_4 = W_2 = 192$ are calculated from Table II. The network structure consists of three separation convolution layers, one max-pooling layer, and one fully connection layer. The RSDSnet uses sigmoid as the activation function, the cross entropy loss function is used to calculate the loss value, and dropout operation is added to prevent the overfitting phenomenon. The RSDSnet based on reg-superpixel is shown in Algorithm 2.

C. MSRSnet Based on Multiscale Reg-Superpixel

For the same pixel in the image, superpixel information of multiple scales can be used for training at the same time. Small

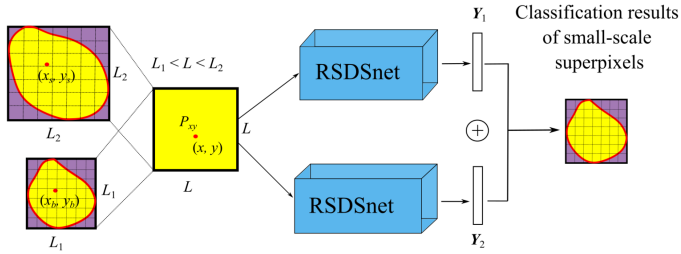

Fig. 5. Generated size of the small-scale superpixel and the large-scale superpixel.

scale superpixels segment the image finer and contain more edge information, while large scale superpixels segment the image wider and contain more spatial context information. Therefore, on the basis of the second section, a branch is added to the RSDSnet network to form the multiscale reg-superpixel network (MSRSnet), as shown in Fig. 5.

In Fig. 5, $L = 15$ is the size of the input rectangle of the network, L_1 and L_2 are sizes of the rectangle corresponding to the small-scale superpixel and large-scale superpixel. However, reg-superpixel has the possibility of oversmoothing the original image in the process of generating reg-superpixel. In the process of generating reg-superpixel, formula (12) calculated the weighted average of the four points Z_{ij}, essentially equivalent to smoothing the four points to generate the target point P_{xy}. This smoothing operation has a negative effect on network feature extraction, but it can be reduced through the selection of the relative size of L_1 and L_2.

We regard the coordinate of a point P_{xy} in the generated reg-superpixel as (x, y), where $x \geq 0$ and $y \geq 0$. Then, its corresponding reconstructed coordinate point in the original image is (x', y'). The reconstruction formula of coordinates is as follows:

$$x' = (x + 0.5) * \frac{W_{x'}}{W_x} - 0.5$$
$$y' = (y + 0.5) * \frac{H_{y'}}{H_y} - 0.5 \tag{16}$$

where $W_{x'}$ and $H_{y'}$ are the width and height of the original image, W_x and H_y are the width and height of the reconstructed reg-superpixel image. As shown in Fig. 5, the reconstruction point of the large size superpixel in the reduction operation is

(x_s, y_s), so the reconstruction formula can be rewritten as

$$x_s = (x + 0.5) * a - 0.5 = x * a + 0.5 * (a - 1)$$
$$y_s = (y + 0.5) * a - 0.5 = y * a + 0.5 * (a - 1) \quad (17)$$

where $a = L_2/L$. Because in the reduction operation $L_2 > L$, $L \in N^*$, so $a - 1 > 0$. Thus, $x_s > x$ can be derived from formula (17), and in the same way $y_s > y$ can be derived. The amplification operation is also shown in Fig. 5. The reconstructed point is (x_b, y_b). So $x_b < x$ and $y_b < y$ can also be deduced from $L_1 < L$.

Take the x-coordinates x^1 and x^2 of two adjacent points on reg-superpixel ($x^2 - x^1 = 1$) as an example, from the abovementioned inequality derivation process, $0 \le x_b^i < x^i$ ($i = 1, 2$) can be obtained. The distance between two corresponding reconstruction points x_b^1 and x_b^2 of the amplification operation is $x_b^2 - x_b^1 < x^2 - x^1$, so $x_b^2 - x_b^1 < 1$. The y-coordinate is in the same way. The adjacent reconstruction points required by the amplification operation often appear within 1 unit coordinate, and multiple reconstruction points are calculated by four same points Z_{ij}. Therefore, the generated reg-superpixel often results in image degradation due to excessive smoothness.

Although small-size superpixels can provide detailed information, amplification operation can reduce the image quality. On the contrary, the reduction operation on large-size superpixels loses the detail information, but at the same time can reduce the impact of image degradation. In order to reduce the negative impact of image degradation on feature extraction of the network as much as possible while enlarging the receptive field of the network, the L_1 and L_2 are set as $L_1 < L < L_2$ in this article to obtain the optimal classification effect. Among them, since $L = 15$ as shown in Table II of Section II-B is determined by the size of the convolution kernel, the range of L_1 and L_2 is $L_1 < 15 < L_2$.

MSRSnet first regularizes the large and small superpixels of the same sampling point, and then inputs them as the training features of the sampling points into the RSDSnet network described in Section B. After three convolution layers, one pooling layer and one fully connected layer, the network outputs the prediction results, as shown in formula (18)

$$\mathbf{Y}_1 = \text{softmax}(\text{FC}(\mathbf{W}_{\text{small}}, \mathbf{X}_{\text{small}}^{\text{reg}}))$$
$$\mathbf{Y}_2 = \text{softmax}(\text{FC}(\mathbf{W}_{\text{large}}, \mathbf{X}_{\text{large}}^{\text{reg}})) \quad (18)$$

where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are the $1 \times C$ dimensional output layers of small scale and large scale, and C is the number of categories. $\mathbf{W}_{\text{small}}$ is the weight matrix of small-scale superpixels, and $\mathbf{X}_{\text{small}}^{\text{reg}}$ is the input feature map of the convolution layer. $\mathbf{W}_{\text{large}}$ and $\mathbf{X}_{\text{large}}^{\text{reg}}$ are in the same way. FC() is the full connection layer. Then, the network output layer is fused, and the fusion method is shown in formula

$$\mathbf{Y} = \text{avg}(\mathbf{Y}_1, \mathbf{Y}_2) = \frac{1}{2}(\mathbf{Y}_1 + \mathbf{Y}_2) \quad (19)$$

where $\text{avg}(x, y)$ represents the average of x and y. The $\mathbf{Y}$ after averaging synthesizes the classification results of the same sampling point under the two scale superpixels. Finally, in order not to lose the edges and other details of the small-scale superpixel, the final classification results are expanded on the small-scale superpixel.

Algorithm 3: MSRSnet.

Input: The training set Q' after dimensionality reduction, the extracted su–perpixel map $\mathbf{S}_{large}$ and $\mathbf{S}_{small}$, and the coordinate vector $\mathbf{E}_{1\times s}$ of small-scale superpixel, where s is the total number of small-scale superpixel.
Output: Classification result labels for small-scale superpixel $\mathbf{L}_{1\times s}$.

1. For each sample $\mathbf{X}_i^{ori}$ in Q', find the two scale superpixels $\mathbf{S}_{large}$ and $\mathbf{S}_{small}$ to which it belongs.
2. $\mathbf{X}_{large}^{reg}$ and $\mathbf{X}_{small}^{reg}$ is generated by two scale superpixels according to Algorithm 2, and is fed into two-branch RSDSnet network. Formula (19) is used to fuse the output layer during training.
3. **for** $i \leftarrow 1$ to s **do**
4. Find out the large-scale superpixel corresponding to the small-scale superpixel i of $\mathbf{E}_i$, and regularize them respectively.
5. Algorithm 2 is used to input the two scale regular superpixels into the two-branch RSDSnet network for testing, and the output layers $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are obtained.
6. Formula (19) is used to fuse $\mathbf{Y}_1$ and $\mathbf{Y}_2$, and the prediction category is regarded as the classification result of small-size superpixel i.
8. **end**
9. Output the final classification results of small-scale superpixel $\mathbf{L}_{1\times s}$.

The multiscale reg-superpixel classification network MSRSnet is constructed as shown in Algorithm 3.

D. Proposed FS-MSRSnet Method

FS-MSRSnet uses the separation feature selection method considering the polarimetric decomposition features, and its network structure is based on RSDSnet by adding two scale branches to form MSRSnet. MSRSnet classifies small-scale superpixels by synthesizing the superpixel information of the two scales, and the final segmentation map is represented by the superpixel classification result. The overall process of FS-MSRSnet is shown in Algorithm 4. Among them, the training set Q is used for both feature selection and network learning, which makes full use of the limited training data of PolSAR to generate better classification results.

III. Experiments

In this section, to verify the effectiveness of the FS-MSRSnet, we conduct experiments with four different sizes of PolSAR datasets. First, four kinds of datasets and four compared algorithms are introduced briefly, and the parameter setting of the experiment is analyzed and discussed. Second, three indexes of average accuracy (AA), overall accuracy (OA), and Kappa coefficient are used as evaluation indexes, and the classification result graph is shown. Finally, the ablation experiment is carried out.

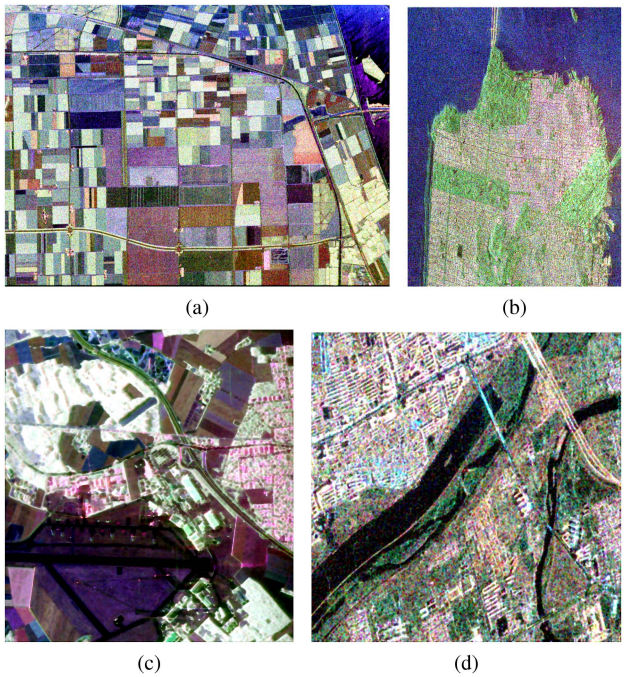

Fig. 6. Pauli RGB images for four datasets. (a) Flevoland. (b) San Francisco. (c) Oberpfaffenhofen. (d) Xi'an.

Algorithm 4: FS-MSRSnet.

Input: Original PolSAR coherence matrix $\mathbf{T}$. Selected feature number c, superpixel scale L_1 and L_2 $(L_1 < 15 < L_2)$.

Output: Classification result labels for small-scale superpixel $\mathbf{L}_{1\times s}$.

1. The improved SLIC method is used to extract superpixel maps $\mathbf{S}_{large}$ and $\mathbf{S}_{small}$ of sizes L_1 and L_2 from the coherence matrix $\mathbf{T}$.
2. The $H, A, \alpha, P_c, P_s, P_d, P_v$ and SPAN features are extracted from the coherence matrix $\mathbf{T}$, and the 17-dimensional feature $\mathbf{F}$ is formed together with the 9-dimensional data $\mathbf{F}_0$ to generate the original data $\mathbf{X}_{N\times f}$, where N is the total number of pixels.
3. Generate a training set Q by randomly sampling $\mathbf{X}_{N\times f}$, use Algorithm 1 to extract c-dimensional features $\mathbf{F}'$ that are useful for classification, and generate the new training set Q' after dimensionality reduction.
4. Combined with $\mathbf{S}_{large}$, $\mathbf{S}_{small}$ and Q', Algorithm 3 is used to input data into MSRSnet, and output the classification result $\mathbf{L}_{1\times s}$ of small-scale superpixel.

A. Dataset Introduction

In order to test the performance of the FS-MSRSnet more objectively, experiments are carried out on four datasets with different sizes. The details of the four datasets are as follows, and their Pauli decomposition images are given in Fig. 6.

1) Flevoland Dataset: This dataset is the four-polarization SAR dataset obtained by NASA/JPL AIRSAR in Flevoland, the Netherlands, which is widely used as the benchmark dataset for PolSAR. The image size is 750×1024 pixels, as shown in Fig. 6(a).

2) San Francisco Dataset: This dataset is the PolSAR dataset in San Francisco, USA, acquired by RADARSAT-2 system, and annotated data are provided by Liu et al. [35]. The image size is 1800×1380 pixels, as shown in Fig. 6(b).

3) Oberpfaffenhofen Dataset: This dataset is the PolSAR data collected by ESAR in Oberpfaffenhofen, Germany. The image size is 1300×1200 pixels, as shown in Fig. 6(c).

4) Xi'an Dataset: This dataset is also the Xi'an region of China acquired by RADARSAT-2 system, which is composed of 512×512 pixels, as shown in Fig. 6(d).

TABLE III
EXPERIMENTAL ENVIRONMENT SETTING

Name	The experimental setup
Platform	Windows 10
Python/Tensorflow	V 3.6.5 / V 1.5.0
MATLAB	R2019a
CPU	Intel(R) Core(TM) i7-10750H
Memory	16G
GPU	NVIDIA GeForce RTX 2060

B. Compared Algorithms

The four compared algorithms used in this article are PolSAR classification method based on CV-CNN [23], PolSAR classification method based on real number field deep CNN with the same structure [23], dense connection and deep divisible convolution PolSAR image classification method based on CNN (DSnet) [36], and a PolSAR image classification method based on MSSP-GCN [27].

Among them, the first three compared algorithms are pixel-level methods. CNN network and CV-CNN have the same structure as shown in [23], including two convolutional layers, one pooling layer, and one fully connected layer. DSnet network structure consists of three convolutional layers, a pooling layer and a fully connected layer, and is also classified pixel by pixel. The fourth is a superpixel level multiscale graph convolutional network classification method (MSSP-GCN). In order to control the experimental variables, the size scale superpixel size set by the MSSP-GCN method and the proposed method is the same.

C. Experimental Setup and Parameter Analysis

All experiments in this article are carried out on the same platform, as shown in Table III.

The experimental settings are as follows: the parameter m of SLIC segmentation is set to 0.7, the proposed network uses a learning rate of 0.001, the dropout value is 0.5, the input image matrix size is $L = 15$, the feature selection number is set as $c = 12$ through parameter analysis, and the average side length of superpixels of size is $L_1 = 10/L_2 = 20$ or $L_1 = 10/L_2 = 25$, respectively. Due to the poor performance of CNN and CV-CNN in the low sampling rate on Flevoland dataset, both of

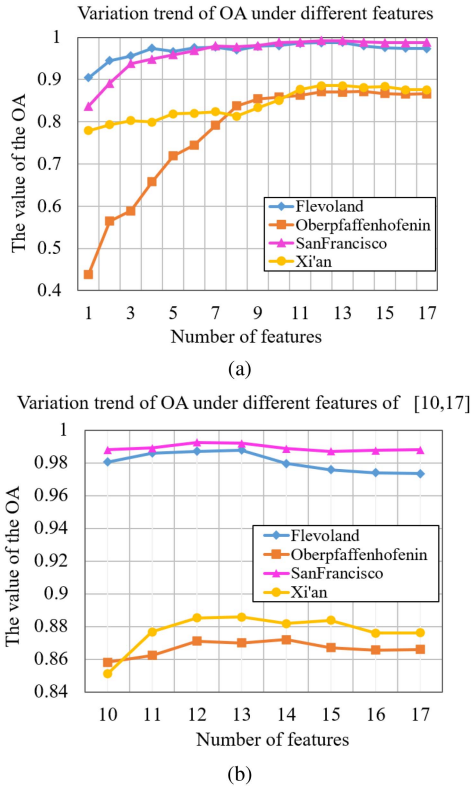

Fig. 7. OA values of four datasets under different feature numbers. (a) OA value at $c \in [1, 17]$. (b) OA value at $c \in [10, 17]$, which is the part intercepted from the black dashed box in (a).

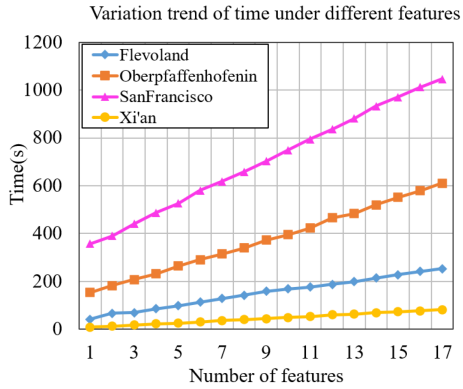

Fig. 8. Computation times of the algorithms of four datasets with different number of features.

TABLE IV
OA VALUES OF SUPERPIXEL AT DIFFERENT SCALES ON FOUR DATASETS

Dataset	Small superpixel scale	Large superpixel scale (+5)	Large superpixel scale (+10)	Large superpixel scale (+15)
Flevoland	W_1=10	0.9834	0.9885	**0.9907**
	W_1=15	0.9795	0.9787	0.9816
Oberpfaffenhofen	W_1=10	0.8518	**0.8701**	0.8695
	W_1=15	0.8506	0.8666	0.8701
Xi'an	W_1=10	0.8822	**0.8865**	0.8829
	W_1=15	0.8791	0.8708	0.8803
San Francisco	W_1=10	0.9906	0.9913	**0.9924**
	W_1=15	0.9910	0.9908	0.9912

them are set as the best 10% sampling rate in this experiment, and the other compared algorithms and the proposed method all have 1% sampling rate. On other datasets, the training data of all compared algorithms and proposed methods are set at 1% sampling rate.

In Section II-A, a Pol-ReliefF method is proposed to select separation features for classification, which reduces the original 17-dimensional polarization feature $\mathbf{F}$ to c dimension to generate $\mathbf{F}'$. However, when the number of c is too small, it is difficult for the network to learn an effective classification model, while if the value of c is too large, it cannot play the role of feature selection. Therefore, the classification accuracy of FS-MSRSnet network with different feature numbers c =[1, 17] is calculated. The OA is used as the classification accuracy index to observe the influence of different feature numbers c on the classification results. The experimental results are shown in Fig. 7, where Fig. 7(a) is the change of OA value of the four datasets when the number of features increases from 1 to 17, and Fig. 7(b) is the result of intercepting the black dashed box in Fig. 7(a).

As can be seen from Fig. 7(a), when the selected feature number c is small, the OA value of the four data sets is low. With the increase of the number of features, the network can learn more features, and the OA value also increases. When c increases to more than 13, it can be seen from Fig. 7(b) that the OA value of Flevoland dataset and Xi'an dataset decreases greatly with the increase of the number of features. This is because some poorly performing decomposition features have a negative impact on classification. The decomposition features of the other two datasets have less negative impact on the classification. On the four PolSAR datasets, although the variation amplitude of parameter c is different, the variation regularity is consistent. Flevoland dataset and San Francisco dataset have high OA value, and the range of change is not obvious, but it can still be seen that OA reaches the maximum value in the range c = [8, 13], and fluctuates in a small range. Fig. 8 shows the change of computation time under different feature numbers.

It can be seen that with the increase of the number of features, the computation times of the algorithms increases almost linearly. Therefore, based on the variation trend of running time and OA values, $c = 12$ can be chosen to select the most beneficial features for classification and improve the classification efficiency of the network.

The multiscale superpixel network uses two different scales of superpixels, namely the small-scale superpixel L_1, and the large-scale superpixel L_2. For reference to the experimental settings of superpixels of two scales in paper [27], we conducted experiments, respectively, under several conditions: $L_1 = L - 5n$, $L_2 = L + 5\ m$ (n, $m \in N$), where $L = 15$. The experimental results are shown in Table IV, where bold entities means the maximum accuracy.

As can be seen from Table IV, the optimal value of classification accuracy is concentrated in the region where $L_1 < 15$ and $L_2 > 15$. When $L_1 = 10$ and $L_2 = 20$, Oberpfaffenhofen dataset and Xi'an dataset obtain the optimal classification result.

TABLE V
CLASSIFICATION ACCURACY OF THE FIVE ALGORITHMS ON FLEVOLAND DATASET

Class	CNN (10%)	CV-CNN (10%)	DSnet (1%)	MSSP-GCN (1%)	FS-MSRSnet (1%)
Stembeas	0.9518	0.9813	**0.9920**	0.9536	0.9887
Peas	0.9890	0.9832	0.9780	0.9871	**0.9946**
Forest	0.9737	0.9841	0.9840	0.9899	**0.9963**
Lucerne	0.9882	0.9804	0.9712	0.9601	**0.9944**
Wheat	0.9692	0.9785	0.9810	0.9957	**0.9964**
Beet	0.9792	**0.9881**	0.9671	0.9682	0.9683
Potatoes	0.9631	0.9798	0.9668	0.9668	**0.9836**
Bare soil	0.9867	**0.9991**	0.9974	0.9902	0.9961
Grass	0.9534	0.9399	0.9461	0.9721	**0.9873**
Rapeseed	0.9190	0.9441	0.9397	0.9801	**0.9968**
Barely	0.9781	0.9789	0.9787	**0.9992**	0.9885
Wheat2	0.9489	0.9608	0.9574	0.9900	**0.9982**
Wheat3	0.9750	0.9914	0.9853	**0.9916**	0.9900
Water	0.9985	0.9968	**0.9994**	0.9989	0.9987
Buildings	0.8171	0.8201	**0.9626**	0.7932	0.8194
AA	0.9594	0.9671	0.9738	0.9722	**0.9798**
OA	0.9622	0.9749	0.9809	0.9855	**0.9907**
Kappa	0.9694	0.9751	0.9721	0.9846	**0.9899**

When $L_1 = 10$ and $L_2 = 25$, Flevoland dataset and San Francisco dataset obtain the optimal classification result. Therefore, the selection of parameter L_1 and L_2 in this article is $L_1 = 10$ and $L_2 = 20/25$.

D. Classification Results on the Flevoland Dataset

The experimental results on Flevoland dataset are shown in Table V. The percentage in parentheses represents the sampling rate of the training set. In order to maintain the experimental variables, the FS-MSRSnet and MSSP-GCN methods proposed in this article use the same size of two scale superpixels, namely, the average side length is $L_1 = 10$ and $L_2 = 25$, respectively. The CNN and CV-CNN networks are tested at their best sampling rate of 10%.

As can be seen from Table V, the FS-MSRSnet has high classification accuracy in eight categories of pea, forest, lucerne, two kinds of wheat, potato, grass, and rapeseed. Compared with CNN and CV-CNN, AA value increases by about 1%–2%, OA value, and Kappa value significantly increase by about 2%–3%. Compared with DSnet and MSSP-GCN, OA value, and Kappa value increase by about 1%, and AA value is also improved. Taking the OA value as an example, the OA value between CNN and CV-CNN is about 96%–97% at 10% sampling rate. The OA value of DSnet using separate convolution and dense connection network and superpixel-level classification MSSP-GCN network can reach more than 98% at 1% sampling rate. Under the same sampling rate, more than half of the subcategories of FS-MSRSnet proposed in this article are improved compared with the four algorithms, and the OA value can also reach about 99%. The AA and Kappa value are also slightly improved. Although there is not much difference between indicators in Table V, the classification effect of edge and homogeneous regions is different in the classification result map. The classification results on Flevoland dataset are shown in Fig. 9.

In Fig. 9, the first three pixel-level network classification results based on CNN structure all have noise points, and there are many pixels classified incorrectly in the edge area. Taking the black box in the figure as an example, there are a large number of noises in the two wheat and rapeseed regions of CNN and CV-CNN. DSnet has a good result, but there are still some noises, indicating that the network has a poor discrimination of these regions. From Fig. 9(e) and (f), both superpixel-level methods can fit the image edges well. Similarly, as shown in the black rectangle area in the figure, some misclassified superpixels appear in MSSP-GCN, while the proposed FS-MSRSnet has fewer misclassified superpixels in the same area, indicating that the proposed method achieves the best in terms of uniformity, edge fit, and superpixel classification effect.

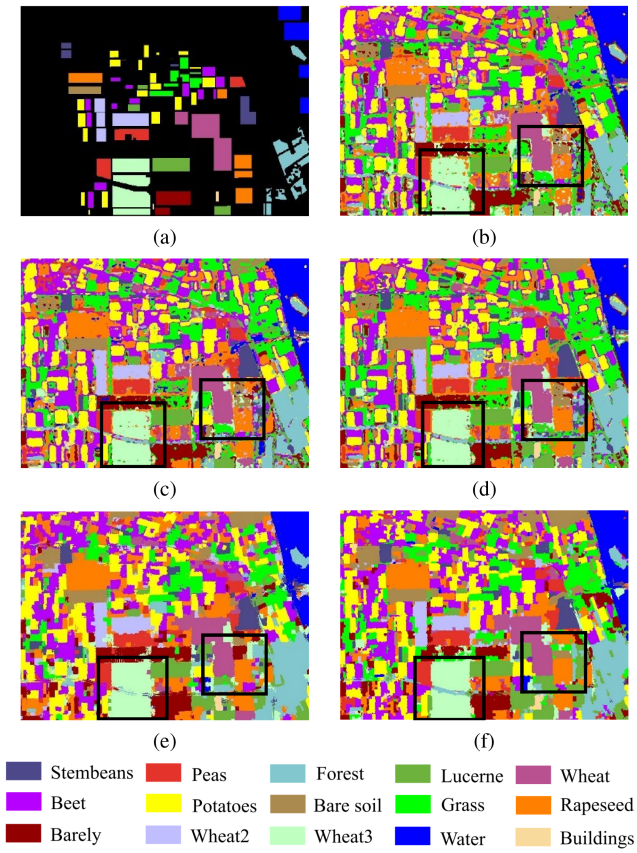

Fig. 9. Classification results of five algorithms on Flevoland dataset. (a) Ground truth map. (b) CNN. (c) CV-CNN. (d) DSnet. (e) MSSP-GCN. (f) FS-MSRSnet.

E. Classification Results on the San Francisco Dataset

The experimental results on the San Francisco dataset are shown in Table VI. On this dataset, 1% sampling rate is used for all methods, and the superpixel size is unchanged. As can be seen from Table VI, the OA, AA, and Kappa on the two superpixel-level networks are all higher than those on the three pixel-level networks, indicating that the superpixel-level classification method can better learn the local features of PolSAR data. The proposed FS-MSRSnet has achieved the highest classification accuracy in the five categories.

In Table VI, compared with CNN, CV-CNN, and DSnet, the AA value of FS-MSRSnet can be significantly improved by about 3%–4%, and the OA value and Kappa value can be improved by about 2%–3%. It shows that the introduction of

TABLE VI
CLASSIFICATION ACCURACY OF THE FIVE ALGORITHMS ON SAN FRANCISO DATASET

Class	CNN (1%)	CV-CNN (1%)	DSnet (1%)	MSSP-GCN (1%)	FS-MSRSnet (1%)
Water	0.9994	0.9993	0.9995	0.9671	**0.9995**
Vegetation	0.9460	0.9515	0.9518	0.9641	**0.9750**
High-Density	0.9269	0.9288	0.9533	0.9718	**0.9892**
Low-Density	0.9312	0.9498	0.9521	0.9857	**0.9895**
Developed	0.9457	0.9725	0.9487	0.9931	**0.9966**
AA	0.9499	0.9605	0.9611	0.9763	**0.9900**
OA	0.9646	0.9711	0.9741	0.9854	**0.9924**
Kappa	0.9529	0.9613	0.9630	0.9774	**0.9892**

TABLE VII
CLASSIFICATION ACCURACY OF THE FIVE ALGORITHMS ON OBERPFAFFENHOFEN DATASET

Class	CNN (1%)	CV-CNN (1%)	DSnet (1%)	MSSP-GCN (1%)	FS-MSRSnet (1%)
Other	0.8231	0.8327	**0.8961**	0.8901	0.8949
Woodland	0.8290	0.8674	0.8543	0.8615	**0.9049**
Farmland	0.8104	0.8332	0.8234	0.8656	**0.8999**
Suburban	0.7506	0.7353	0.7389	0.8289	**0.8747**
Road	0.5381	0.6974	0.7152	0.7383	**0.7704**
AA	0.7502	0.7932	0.8056	0.8369	**0.8650**
OA	0.8127	0.8352	0.8501	0.8609	**0.8701**
Kappa	0.7249	0.7425	0.7759	0.7998	**0.8195**

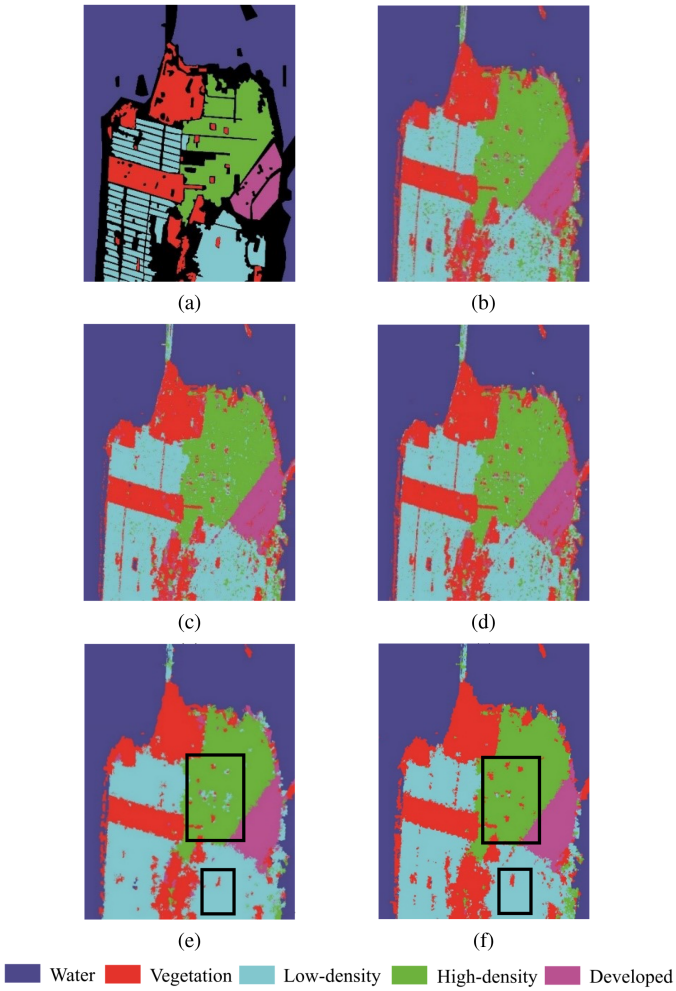

Fig. 10. Classification results of five algorithms on San Francisco dataset. (a) Ground truth map. (b) CNN. (c) CV-CNN. (d) DSnet. (e) MSSP-GCN. (f) FS-MSRSnet.

superpixel information can improve the classification accuracy. Compared with the superpixel-level MSSP-GCN network, the three indicators can also be improved by about 1%–2%, the OA value of FS-MSRSnet can reach about 99%, indicating that the proposed CNN structure network based on regular superpixel can classify superpixel more effectively. The classification result map on this dataset is shown in Fig. 10.

It can be seen from Fig. 10 that the results of CNN have great noise, especially the discrimination between high-density cities and low-density cities is obviously poor. The classification results of CV-CNN and DSnet are similar. Compared with CNN, the misclassification noise is less, but there are still a lot of noise and holes in the homogeneous region. In Fig. 10(e), the MSSP-GCN method based on GCN network can well reduce the misclassification phenomenon in the low-density area, but there are still some errors in the green high-density area. In addition, as can be seen from the black box in the figure, for some small target point in the red area, due to the superpixel classification error of the MSSP-GCN method, all pixels in the same area also have classification errors, resulting in the loss of most of some red targets. The proposed FS-MSRSnet has a higher accuracy rate of superpixel classification, so it can better retain the red small target point in the same region. Compared with the three pixel-level networks CNN, CV-CNN, and DSnet, it can also obtain better results in the homogeneous region and the edge region.

F. Classification Results on the Oberpfaffenhofen Dataset

In order to verify the effectiveness of the proposed method on more complex PolSAR images, the five algorithms are tested on the German Oberpfaffenhofen dataset with the same five categories. In this dataset, there are fine road regions, and because the scattering characteristics of homogeneous regions are complex, it is easy to produce misclassified noise and holes. On this dataset, all methods are tested with 1% sampling rate, and the superpixel scale is $L_1 = 10/L_2 = 20$, while other indicators are unchanged. The experimental results are shown in Table VII.

In Table VII, the FS-MSRSnet has achieved a significant improvement of about 2%–9% in Kappa. The classification accuracy of CNN and CV-CNN is low, DSnet, and MSSP-GCN are partially improved, and FS-MSRSnet has the largest improvement in this category. It can be seen from the three indicators that the OA values of CNN and CV-CNN are relatively low, and AA is below 80%, indicating that the effects of these two methods are limited for more complex PolSAR images at a certain sampling rate. In contrast, DSnet's OA value can reach about 85%, and the superpixel-level MSSP-GCN network can reach about 86%, and also has about 2% and 4% improvement. From the subcategories, it can be seen that the accuracy of road and other two categories is mainly improved greatly, indicating that DSnet and MSSP-GCN can better distinguish these two categories. The OA value of the proposed FS-MSRSnet can be increased by about 1%, and the AA and Kappa can be increased by about 3%, indicating that the proposed method can better

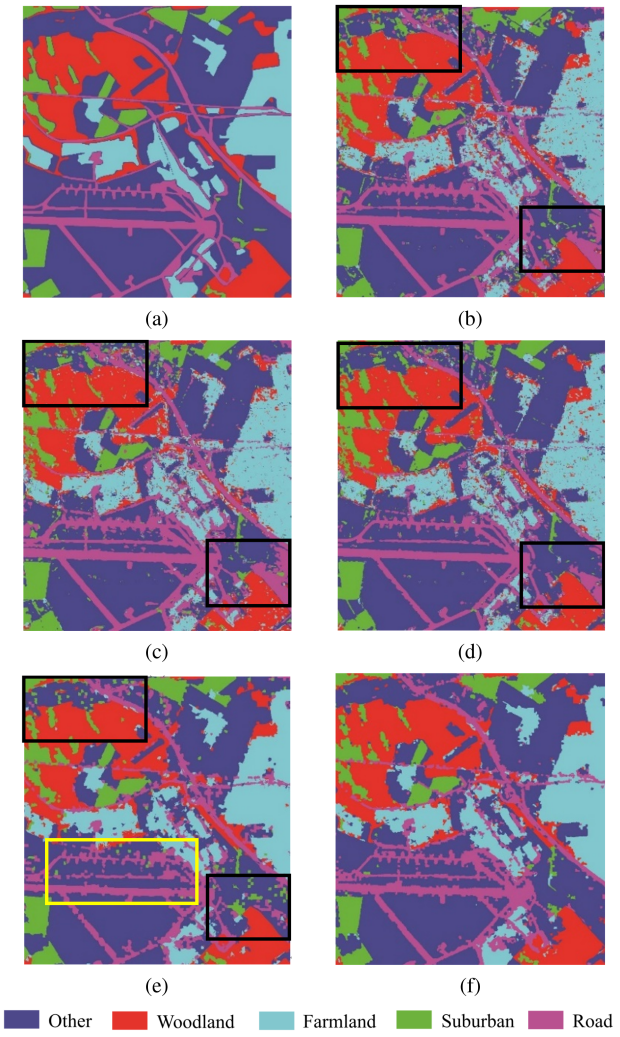

Fig. 11. Classification results of five algorithms on Oberpfaffenhofen dataset. (a) Ground truth map. (b) CNN. (c) CV-CNN. (d) DSnet. (e) MSSP-GCN. (f) FS-MSRSnet.

distinguish the four categories of woodland, farmland, suburban, and road. The experimental results are shown in Fig. 11.

It can be seen from Fig. 11(b)–(d) that there are a lot of noise and voids in both the farmland area and the suburban area. As can be seen from the black box area in the figure, CNN and CV-CNN misclassified pixels of many other areas into road areas and suburban areas. While DSnet can better distinguish roads and other categories, there are still noise points caused by misclassification in the homogeneous area. The superpixel method in Fig. 11(e) and (f) can effectively reduce noise and voids. However, it can be seen from the black rectangular area that although it performs well in the homogeneous and edge areas, there are still many misclassified superpixels in this dataset. As can be seen from the yellow box, for the relatively elongated road area, the MSSP-GCN method is partially broken and missing. Compared with the other four methods, FS-MSRSnet has uniform classification results in farmland, suburban, woodland, and other areas, and the edge fit is also good. Although there are still some superpixel classification errors in the road area due to the influence of the superpixel size, it has been greatly improved compared with the other algorithms.

TABLE VIII
CLASSIFICATION ACCURACY OF THE FIVE ALGORITHMS ON XI'AN DATASET

Class	CNN (1%)	CV-CNN (1%)	DSnet (1%)	MSSP-GCN (1%)	FS-MSRSnet (1%)
City	0.9001	0.8970	0.8964	0.8996	**0.9149**
Grass	0.7014	0.7381	0.7972	0.8289	**0.8816**
Water	0.9324	**0.9359**	0.8876	0.8810	0.8443
Crop	0.9084	0.8394	0.7618	0.8005	**0.9485**
AA	0.8606	0.8526	0.8357	0.8525	**0.8973**
OA	0.8181	0.8275	0.8211	0.8678	**0.8865**
Kappa	0.7348	0.7413	0.7356	0.8178	**0.8250**

G. Classification Results on the Xi'an Dataset

The experimental results on another relatively complex Chinese Xi'an dataset are shown in Table VIII. In this dataset, the sampling rate is also set as 1%, and the superpixel scale is $L_1 = 10/L_2 = 20$.

From Table VIII, it can be seen that three pixel-level networks, CNN, CV-CNN, and DSnet, perform poorly on this dataset, and it is difficult to distinguish between grass and crop areas. Especially in grass area, large classification errors occur, and the classification accuracy is below 80%. In contrast, better classification results can be obtained in cities and water areas, among which CV-CNN achieves the best classification accuracy in water areas. The MSSP-GCN method can distinguish grass and crop areas well, and the classification accuracy of these two categories is more than 80%, but the improvement is very limited. However, the proposed FS-MSRSnet method can well distinguish city, grass and crop areas, and has been significantly improved in these three categories. It can be seen from the three evaluation indicators that the OA value of CNN, CV-CNN, and DSnet only reached about 81%–82%. The AA value of CNN and CV-CNN is high, but the Kappa value is low, indicating that the classification results are unbalanced. From the classification accuracy of each category, it can be seen that these two networks can classify the other categories well, but the classification effect of grass is poor. Compared with MSSP-GCN, FS-MSRSnet can improve the accuracy by about 1%–2% in all three indicators, especially for crop area, which can increase the classification accuracy from 80% to about 94%, indicating that the proposed method can better distinguish crop area and improve the accuracy of superpixel classification.

The experimental results are shown in Fig. 12. It can be seen that the three pixel-level networks in Fig. 12(b)–(d) all classified part of grass as crops, and the red and green areas are mixed and difficult to distinguish. However, the three pixel-level networks can get good classification effect in the water area. The blue region in the figure has some noise points in the edge area, but the whole can well fit the blue river region in the ground truth.

In Fig. 12(e), the MSSP-GCN network can identify the grass and crop areas well, but there are still superpixel classification errors at some points (as shown in the black rectangular box). At the same time, due to the limitation of superpixel size, the water area is broken and missing to different degrees. Since this article uses the deep separated convolution method, which is more suitable for PolSAR data, compared with the MSSP-GCN based on GCN network, the superpixel classification error is

TABLE IX
OA OF RS AND MS-RS ON FOUR DATASETS

Indicator	Flevoland		San Francisco		Oberpfaffenhofen		Xi'an	
Network	RS	MS-RS	RS	MS-RS	RS	MS-RS	RS	MS-RS
AA	0.9781	**0.9845**	0.9855	**0.9900**	0.8383	**0.8650**	0.8660	**0.8973**
OA	0.9817	**0.9907**	0.9856	**0.9924**	0.8644	**0.8701**	0.8756	**0.8865**
Kappa	0.9779	**0.9899**	0.9852	**0.9892**	0.7968	**0.8195**	0.8107	**0.8250**

TABLE X
OA OF MS AND POLRF-MS ON FOUR DATASETS

Indicator	Flevoland		San Francisco		Oberpfaffenhofen		Xi'an	
network	MS	PolRF-MS	MS	PolRF-MS	MS	PolRF-MS	MS	PolRF-MS
AA	0.9761	**0.9845**	0.9812	**0.9900**	0.8571	**0.8650**	0.8700	**0.8973**
OA	0.9768	**0.9907**	0.9901	**0.9924**	0.8656	**0.8701**	0.8737	**0.8865**
Kappa	0.9739	**0.9899**	**0.9899**	0.9892	**0.8270**	0.8195	0.8168	**0.8250**

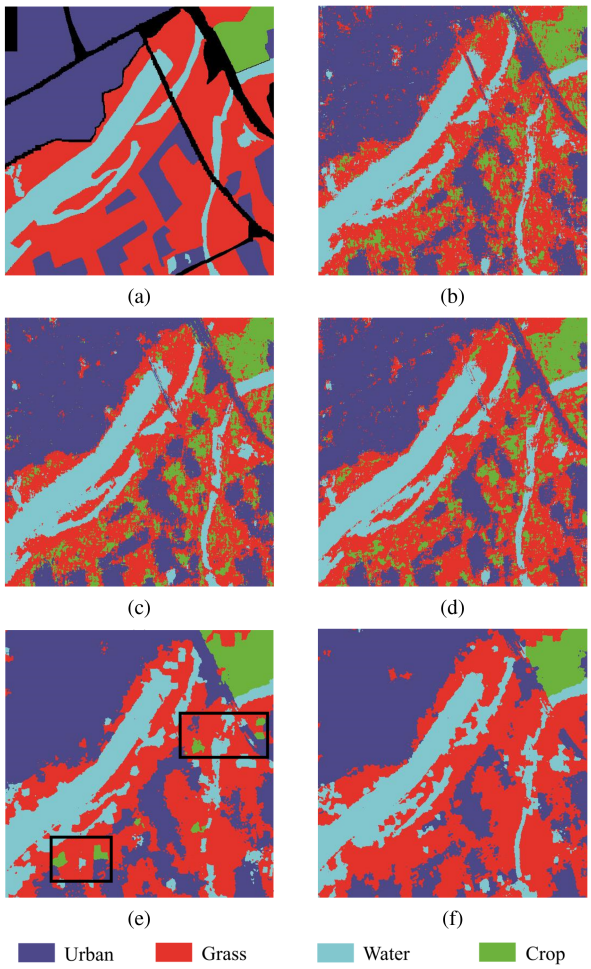

Fig. 12. Classification results of five algorithms on Xi'an dataset. (a) Ground truth map. (b) CNN. (c) CV-CNN. (d) DSnet. (e) MSSP-GCN. (f) FS-MSRSnet.

lower. It can be seen from Fig. 12(f) that there is almost no mixing of red and green. Compared with the previous pixel-level methods, the classification results are more uniform and the edge fit is better.

H. Ablation Experiments

In order to verify the effectiveness of each module of the proposed method, in addition to the compared experiment with the advanced method, the following two ablation experiments are also set up. The ablation experiments between FS-MSRSnet with multiscale idea (MS-RS) and RSDSnet with only small-scale superpixels (RS) are, respectively, conducted under the condition of both using feature selection. And ablation experiments between Pol-ReliefF feature selection method (PolRF-MS) and multiscale (MS) method without feature selection.

The experimental setup between MS-RS and RS is as follows: RSDSnet, which only uses small-scale superpixels, and FS-MSRSnet, which adds MS idea, are compared on four datasets. RSDSnet is a single-branch network using only small-scale superpixels, while FS-MSRSnet is a two-branch network with two scales. Both methods classify small-scale superpixels, and other modules are the same.

The RSDSnet and FS-MSRSnet classification results with the MS method are shown in Table IX. Compared with RSDSnet, FS-MSRSnet has about 1% improvement in each dataset, indicating that the addition of large-scale superpixel information in network learning can well correct the classification results of small-scale superpixels. In Flevoland dataset, since the ground truth area of each class is small, the addition of large-scale superpixel information does not improve the classification effect much. However, the classification indexes of the other three datasets have been greatly improved, which can fully prove the effectiveness of the MS superpixel algorithm.

The experimental setup between PolRF-MS and MS are as follows: an ablation experiment is conducted between the MS method using Pol-ReliefF feature selection and the MS method without feature selection. All modules are the same between PolRF-MS and MS except feature selection.

The experimental results are shown in Table X. On Oberpfaffenhofen and San Francisco, the AA value of PolRF-MS using Pol-ReliefF feature selection method has been increased by about 1% compared with the MS method without feature selection, and the OA value and Kappa value have little difference. In Flevoland and Xi'an, PolRF-MS improves the accuracy by 1%–2% compared with MS. It can be seen that the Pol-ReliefF feature selection method can well select effective features, and

delete some features with weak discrimination between classes. At the same time, it can reduce the running time and improve the efficiency of network learning.

IV. Conclusion

Due to the limitation of receptive field, the classification results of PolSAR data in pixel-level CNN structure network are not uniform. At the same time, the accumulation of polarization features is easy to cause overfitting in network learning. To solve these problems, this article proposes a reg-superpixel guided convolutional neural network based on feature selection and receptive field reconstruction (FS-MSRSnet). First, Cloude and Yamaguchi decomposition methods are used to extract seven different decomposition features for the original PolSAR image, and then the original 17-dimensional features are composed with the original coherence matrix and features. The Pol-ReliefF method is then used to select features. According to the results of PolRF-MS and MS ablation experiments, the Pol-ReliefF method can improve the classification accuracy by about 1%–2% on Flevoland and Xi'an datasets, and also improve the other two datasets to different degrees. Second, reg-superpixel is used to adjust the receptive field, which is regularized and input into the network for training. The class of sampling points is designated as the class of the superpixels in the test, so as to realize the superpixel classification of CNN network. RSDSnet using this method has a significant improvement compared with DSnet network. The OA values of DSnet on the four datasets are 98.09%, 97.41%, 85.01%, and 82.11%, respectively, while RSDSnet can improve 1%–5% on this basis, reaching 98.17%, 98.56%, 86.44%, and 87.56%, respectively. Finally, the two scale superpixels are simultaneously input into RSDSnet and fused in the output layer to adjust the classification results of small scale superpixels by using the spatial context information of large-scale superpixel. It can be seen from the ablation experiment results of MS-RS and RS that the method combining MS superpixel can improve the classification accuracy by about 1%. Compared with the pixel-level network, FS-MSRSnet can improve the classification accuracy by about 1%–6%, and compared with the latest superpixel-level network MSSP-GCN, it can improve the classification accuracy by about 1%–2%. At the same time, it can be seen from the classification result map that FS-MSRSnet method can distinguish several regions with complex structure and regions with low discrimination well, and can obtain more complete and accurate classification results compared with other methods. Compared with the GCN structure network, FS-MSRSnet has significantly fewer superpixel classification errors. It shows that the classification accuracy of PolSAR superpixel can be better improved by inputting superpixel into the CNN structure network that is more targeted to PolSAR images.

FS-MSRSnet has achieved good results in both classification accuracy and classification result maps. However, some small items with scales smaller than superpixels are easy to be ignored. In the future, we can consider introducing pixel-level information on the basis of this method to improve the classification accuracy of small objects. The regularized superpixel method can also be applied to deep networks with other CNN structures.

References

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

[2] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Boston, MA, USA: Artech House, 1998.

[3] S. Baraha, A. K. Sahoo, and S. Modalavalasa, "A systematic review on recent developments in nonlocal and variational methods for SAR image despeckling," *Signal Process.*, vol. 196, 2022, Art. no. 108521.

[4] R. Shang et al., "A self-paced learning algorithm for change detection in synthetic aperture radar images," *Signal Process.*, vol. 142 pp 375–387, 2018.

[5] R. Shang et al., "A thumbnail-based hierarchical fuzzy clustering algorithm for SAR image segmentation," *Signal Process.*, vol. 171, 2020, Art. no. 107518.

[6] E. Pottier, "Dr. J. R. Huynen's main contributions in the development of polarimetric radar techniques and how the 'radar targets phenomenological concept' becomes a theory," in *Proc. SPIE Radar Polarimetry*, 1993, pp. 72–85.

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

[8] 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.

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

[10] B. Ren et al., "Sparse subspace clustering-based feature extraction for PolSAR imagery classification," *Remote Sens.*, vol. 10, no. 3, 2018, Art. no. 391.

[11] W. Song, Y. Wu, and P. Guo, "Composite kernel and hybrid discriminative random field model based on feature fusion for PolSAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 6, pp. 1069–1073, Jun. 2021.

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

[13] H. Dong, L. Zhang, D. Lu, and B. Zou, "Attention-based polarimetric feature selection convolutional network for PolSAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Sep. 2022, Art. no. 4001705.

[14] G. Liu, Y. Li, Y. Chen, R. Shang, and L. Jiao, "Pol-NAS: A. neural architecture search method with feature selection for PolSAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9339–9354, Oct. 2022.

[15] Y. Bai, D. Peng, X. Yang, L. Chen, and W. Yang, "Supervised feature selection for polarimetric SAR classification," in *Proc. 12th Int. Conf. Signal Process.*, 2014, pp. 1006–1010.

[16] B. Banerjee, A. Bhattacharya, and K. M. Buddhiraju, "A generic land-cover classification framework for polarimetric SAR images using the optimum touzi decomposition parameter subset—An insight on mutual information-based feature selection techniques," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1167–1176, Apr. 2014.

[17] S. Hariharan et al., "A novel phenology based feature subset selection technique using random forest for multitemporal PolSAR crop classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4244–4258, Nov. 2018.

[18] X. Huang and X. Nie, "Multi-view feature selection for PolSAR image classification via $l_{2,1}$ sparsity regularization and manifold regularization," *IEEE Trans. Image Process.*, vol. 30, pp. 8607–8618, Oct. 2021.

[19] C. Yang et al., "Reconstruction error-based decomposition feature selection for PolSAR image," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Nov. 2022, Art. no. 5216719.

[20] N. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (An introduction)," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 152–177, Mar. 1963.

[21] S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak, "K-distribution and polarimetric terrain radar clutter," *J. Electromagn. Waves Appl.*, vol. 3, no. 8, pp. 747–768, 1989.

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

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

[24] A. G. Howard et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017. [Online]. Available: https://arxiv.org/abs/1704.04861
[25] G. Huang, Z. Liu, K. Q. Weinberger, and L. Van Der Maaten, "Densely connected convolutional networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit*, 2017, pp. 2261–2269.
[26] D. Li, F. Kong, and Q. Wang, "Hyperspectral image classification via nonlocal joint kernel sparse representation based on local covariance," *Signal Process.*, vol. 180, 2021, Art. no. 107865.
[27] 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, May 2022, Art. no. 5209314.
[28] Y. Guo, L. Jiao, S. Wang, S. Wang, F. Liu, and W. Hua, "Fuzzy superpixels for polarimetric SAR images classification," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 2846–2860, Oct. 2018.
[29] Y. Guo et al., "Adaptive fuzzy learning superpixel representation for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Nov. 2022, Art. no. 5217818.
[30] Y. Guo et al., "Fuzzy superpixels based semi-supervised similarity-constrained CNN for PolSAR image classification," *Remote Sens.*, vol. 12 no. 10, 2020, Art. no. 1694.
[31] X. Qin, H. Zou, W. Yu, and P. Wang, "Superpixel-oriented classification of PolSAR images using complex-valued convolutional neural network driven by hybrid data," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10094–10111, Dec. 2021.
[32] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press. 2016.
[33] Y. Zhou, H. Wang, F. Xu, and Y.-Q. Jin, "Polarimetric SAR image classification using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1935–1939, Dec. 2016.
[34] F. Qin, J. Guo, and F. Lang, "Superpixel segmentation for polarimetric SAR imagery using local iterative clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 13–17, Jan. 2015.
[35] X. Liu et al., "PolSF: PolSAR image dataset on San Francisco," 2019. [Online]. Available: https://arxiv.org/abs/1912.07259
[36] R. Shang et al., "Dense connection and depthwise separable convolution based CNN for polarimetric SAR image classification," *Knowl.-Based Syst.*, vol. 194, 2020, Art. no. 105542.

Ronghua Shang (Senior Member, IEEE) received the B.S. degree in information and computation science and the Ph.D. degree in pattern recognition and intelligent systems from Xidian University, Xi'an, China, in 2003 and 2008, respectively.

She is currently a Professor with Xidian University. Her current research interests include machine learning, pattern recognition evolutionary computation, image processing, and data mining.

Keyao Zhu received the B.E. degree in intelligent science and technology from the School of Artificial Intelligence, Xidian University, Xi'an, China, in 2020, where she is currently working toward the M.S. degree in computer science and technology with the School of Artificial Intelligence in Xidian University, Xi'an, China, in 2022.

Her current research interests include image processing and deep learning.

Jie Feng (Member, IEEE) received the B.S. degree from Chang'an University, Xi'an, China, in 2008 and the Ph.D. degree from Xidian University, Xi'an, China, in 2014.

She is currently an Associate Professor with the Laboratory of Intelligent Perception and Image Understanding, Xidian University. Her current interests include remote sensing image processing, deep learning, and machine learning.

Chao Wang received the B.S. degree from Lanzhou University, Lanzhou, China, in 2016 and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2021.

She is currently an Assistant Research Scientist with the Research Center for Big Data Intelligence, Zhejiang Laboratory, Zhejiang, China. Her research interests include spatial data mining and geographic information science.

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

From 1990 to 1991, he was a Postdoctoral Fellow with the National Key Laboratory for Radar Signal Processing, Xidian University, Xi'an, China. Since 1992, he has been a Professor with the School of Electronic Engineering, Xidian University. He is currently the Director of the Key Lab of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University. He has charged of about 40 important scientific research projects, and authored or coauthored more than 20 monographs and a hundred papers in international journals and conferences. His research interests include image processing, natural computation, machine learning, and intelligent information processing.

Dr. Jiao is a member of IEEE Xi'an Section Execution Committee and the Chairman of Awards and Recognition Committee, Vice Board Chairperson of Chinese Association of Artificial Intelligence, Councilor of Chinese Institute of Electronics, Committee Member of Chinese Committee of Neural Networks, and an Expert of Academic Degrees Committee of the State Council.

Songhua Xu, photograph and biography not available at the time of publication.