Polarimetric Convolutional Network for PolSAR Image Classification

Xu Liu, Student Member, IEEE, Licheng Jiao, Fellow, IEEE, Xu Tang, Member, IEEE, Qigong Sun, Student Member, IEEE, and Dan Zhang

Abstract—The approaches for analyzing the polarimetric scattering matrix of polarimetric synthetic aperture radar (PolSAR) data have always been the focus of PolSAR image classification. Generally, the polarization coherent matrix and the covariance matrix obtained by the polarimetric scattering matrix are used as the main research object to extract features. In this paper, we focus on the original polarimetric scattering matrix and propose a polarimetric scattering coding way to deal with polarimetric scattering matrix and obtain a close complete feature. This encoding mode can also maintain polarimetric information of scattering matrix completely. At the same time, in view of this encoding way, we design a corresponding classification algorithm based on the convolution network to combine this feature. Based on the polarimetric scattering coding and convolution neural network, the polarimetric convolutional network is proposed to classify PolSAR images by making full use of polarimetric information. We perform the experiments on the PolSAR images acquired by AIRSAR and RADARSAT-2 to verify the proposed method. The experimental results demonstrate that the proposed method get better results and has huge potential for PolSAR data classification. Source code for polarimetric scattering coding is available at https://github.com/liuxuvip/Polarimetric-Scattering-Coding.

Index Terms—Classification, convolution network, polarimetric scattering matrix, polarimetric synthetic aperture radar (PolSAR).

I. INTRODUCTION

POLARIMETRIC synthetic aperture radar (PolSAR) images collected with airborne and satellite sensors are a wealthy source of information concerning the Earth’s surface and have been widely used in urban planning, agriculture assessment, and environment monitoring [1], [2]. These applications require the fully understanding and interpretation of PolSAR images.

Hence, PolSAR image interpretation is of much significance in theory and application. Land use classification of PolSAR images is an important and indispensable research topic, since these images contain rich character of the target (e.g., scattering properties, geometric shapes, and the direction of arrival). The land use classification is arranging the pixels to the different categories according to the certain rule. The common objects within the PolSAR images include land, buildings, water, sand, urban areas, vegetation, road, bridge, and so on [3]. In order to distinguish them, the features of the pixels should be fully extracted and mined. With the development of the PolSAR image classification, many feature extraction algorithms based on physical scattering mechanisms have been introduced. The feature extraction techniques based on polarimetric characteristics can be divided into two kinds: coherent target decomposition and incoherent target decomposition. The former acts on the scattering matrix to characterize completely polarized scattered waves, which contains the fully polarimetric information. The latter acts only on the Mueller matrix, covariance matrix, or coherency matrix in order to characterize partially polarized waves [4].

The coherent target decomposition algorithms include the Pauli decomposition, the sphere–diplane–helix decomposition [5], the symmetric scattering characterization method [6], Cameron decomposition [7], Yamaguchi four-component scattering power decomposition [8], general polarimetric model-based decomposition [9], [10], and some advances [11], [12]. The incoherent target decomposition algorithms include Huynen decomposition [13], Freeman–Durden decomposition [14], Yamaguchi four-component decomposition [15], and Cloude–Pottier decomposition [16], [17], and a number of approaches have been reported [18]–[20]. In addition to the feature based on the polarization mechanism [21]–[24], there are some traditional features of natural images, which have been utilized to analyze PolSAR image, such as color features [25], texture features [26], spatial relations [27], and so on. Based on the above-mentioned basic features, some multiple features of PolSAR data have been constructed to improve the classification performance [25], [28], [29].

For classification tasks, besides the feature extraction, classifier design is also a key point. According to the degree of data mark, the classification methods can be broadly divided into
three groups, including unsupervised classification (without any labeled training data), semisupervised classification (SSC) (with a small amount of labeled data and a large amount of unlabeled data), and supervised classification (with completely labeled training data) [30].

The unsupervised classification approaches design a function to describe hidden structure from unlabeled data. The traditional methods always make a decision rule to cluster PolSAR data into different groups, and the number of groups is also a hyperparameter. There are a lot of unsupervised classification methods for PolSAR data, such as the H/α complex Wishart classifier [31], polarimetric scattering characteristic-preserved method [32], fuzzy k-means cluster classifier [33], [34], classification based on deep learning [35]–[37], and so on. The SSC is a class of supervised learning tasks and techniques that also make use of a small amount of labeled data and a large amount of unlabeled data for training. Compared with the unsupervised method, SSC can improve classification performance so long as the target is to make full use of a smaller number of labeled samples [38]. There are many semisupervised classification for PolSAR data, such as the classification based on hypergraph learning [39], the method based on parallel auction graph [40], spatial-anchor graph [38], and so on. Unlike unsupervised approaches and semisupervised approaches, the supervised classifications use enough labeled samples to train the classifiers, which can be applied to determine the class of other samples. Lots of methods have been introduced, including maximum likelihood (MLD) [41], support vector machines (SVMS) [42]–[44], sparse representation [45], and deep learning [46]–[48].

Recently, deep learning has attracted considerable attention in the computer vision community [49]–[52], as it provides an efficient way to learn image features and to represent certain function classes far more efficiently than shallow ones [53]–[55]. Deep learning has also been introduced into the geoscience and remote sensing community [35], [56]–[64]. Especially in the direction of PolSAR image classification, in [36], a specific deep model for PolSAR image classification is proposed, which is named Wishart deep stacking network. A fast implementation of Wishart distance is achieved by a special linear transformation, which speeds up the classification of PolSAR image. In [35], a new type of restricted Boltzmann machine (RBM) is specially defined, which we name the Wishart–Bernoulli RBM, and is used to form a deep network named Wishart deep belief networks. It is hard for researchers to consider all kinds of features. What is more, a fully convolutional network (FCN) is successfully used for natural image semantic segmentation [68]–[74] and remote sense image classification based on one-by-one pixel [75]–[78]. In [75], a fully CNN is trained to segment water on Landsat imagery. In [76], a novel deep model, i.e., a cascaded end-to-end convolutional neural network (CasNet), was proposed to simultaneously cope with the road detection and centerline extraction tasks. Specifically, CasNet consists of two networks. One aims at the road detection task. The other is cascaded to the former one, making full use of the feature maps produced formerly, to obtain the good centerline extraction. In [77], a novel hyperspectral image classification framework, named deep multiscale spatial–spectral feature extraction algorithm, was proposed based on a fully CNN. Volpi and Tuia presented a CNN-based system relying on a downsampling-then-upsampling architecture. Specifically, it first learns a rough spatial map of high-level representations by means of convolutions and then learns to upsample them back to the original resolution by deconvolution. By doing so, the CNN learns to densely label every pixel at the original resolution of the image.

Inspired by the previous research works, a supervised PolSAR image classification method based on polarimetric scattering coding and convolution network is proposed in this paper. Our goal is to solve the problems of PolSAR data coding, feature extraction, and land cover classification. This paper can be summarized into three main parts as follows.

1) First, a new encoding mode of polarimetric scattering matrix is proposed, which is called polarimetric scattering coding. It not only completely preserves the polarization information of data but also facilitates to extract high-level features by deep learning, especially convolutional networks.

2) Second, a novel PolSAR image classification algorithm based on polarimetric scattering coding and convolutional network is proposed, which is called a polarimetric convolutional network and also an end-to-end learning framework.

3) Third, feature aggregation is designed to fuse the two kinds of feature and mine more advanced features.

This paper is organized as follows. In Section II, the representation of PolSAR images is described. In Section III, the proposed method named polarimetric convolutional network is given. The experimental setting is presented in Section IV. The results are presented in Section V, and the conclusions and discussions are presented in Section VI.

II. PROPOSED METHOD

In the PolSAR image classification task, the land use classes are determined by different analyses, including polarization of the target responses, scattering heterogeneity determination, and determination of the polarization state for target discrimination, which need to be decided by different features. It is hard for researchers to consider all kinds of features. PolSAR data is a 2-D complex matrix. The traditional feature extraction method represents PolSAR data in a 1-D vector, which destroys data space structures. In order to solve this
problem, the intuitive way is to express the original data directly. In this paper, the polarimetric scattering coding is proposed to express the original data directly, which can maintain structure information completely. Next, the polarimetric scattering matrix obtained by the encoding is fed into a classifier based on a fully convolutional network. This section consists of three parts. First, representation of PolSAR images is given. Second, the polarimetric scattering coding for complex scattering matrix $S$ is explained. Third, the proposed method called polarimetric convolutional network is presented.

### A. Representation of PolSAR Images

The fully PolSAR measures the amplitudes and phases of backscattering signals in four combinations: 1) HH; 2) HV; 3) VH; and 4) VV. Here $H$ means horizontal mode and $V$ means vertical mode. These signals form a $2 \times 2$ complex scattering matrix $S$ to represent the information for one pixel, which relates the incident and the scattered electric fields. Scattering matrix $S$ can be expressed as

$$
S = \begin{bmatrix}
S_{HH} & S_{HV} \\
S_{VH} & S_{VV}
\end{bmatrix} = \begin{bmatrix}
|S_{HH}|e^{i\phi_{HH}} & |S_{HV}|e^{i\phi_{HV}} \\
|S_{VH}|e^{i\phi_{VH}} & |S_{VV}|e^{i\phi_{VV}}
\end{bmatrix} \tag{1}
$$

where $S_{HH}, S_{HV}, S_{VH}$, and $S_{VV}$ are the complex scattering coefficients, and $S_{HV}$ is the scattering coefficient of the horizontal (H) transmitting and vertical (V) receiving polarization. Other elements have similar definitions. $|S_{HH}|$, $|S_{HV}|$, $|S_{VH}|$, and $|S_{VV}|$ denote the amplitudes of the measured complex scattering coefficients, and $\phi_{HH}, \phi_{HV}, \phi_{VH}$, and $\phi_{VV}$ are the value of phases. $i$ is the complex unity.

The characteristics of the target can be specified by vectorizing the scattering matrix. Based on the two important basis sets, lexicographic basis and Pauli spin matrix set, in the case of monostatic backscattering with reciprocal medium, the lexicographic scattering vector $\vec{k}_L$ and the Pauli scattering vector $\vec{k}_p$ are defined as

$$
\vec{k}_L = [S_{HH}, \sqrt{2}S_{HV}, S_{VV}]^T \tag{2}
$$

$$
\vec{k}_p = 1/\sqrt{2}[S_{HH} + S_{VV}, S_{HH} - S_{VV}, 2S_{HV}]^T \tag{3}
$$

where superscript $T$ denotes the transpose of vector.

The scattering characteristics of a complex target are determined by different independent subscatterers and their interaction. The scattering characteristics are described by a statistic method due to the randomness and depolarization. Moreover, the inherent speckle in the SAR data is reduced by spatial averaging at the expense of losing spatial resolution. Therefore, for the complex target, the scattering characteristics should be described by a statistic coherence matrix or covariance matrix. Covariance and coherence matrices can be generated from the outer product of $\vec{k}_L$ and $\vec{k}_p$, respectively, with its conjugate transpose

$$
C = \langle \vec{k}_L \vec{k}_L^* \rangle \tag{4}
$$

$$
T = \langle \vec{k}_p \vec{k}_p^* \rangle \tag{5}
$$

where $\langle \cdot \rangle$ denotes the average value in the data processing stage, and the superscript $H$ stands for the complex conjugate and transpose of vector and matrix.

The covariance matrix $C$ has been proved to follow a complex Wishart distribution [79]. Moreover, the coherence matrix $T$ is used to express PolSAR data, which has a linear relation with covariance matrix $C$. The PolSAR features are always extracted indirectly from the PolSAR data, such as color features, texture features, and the decomposition features. The color and texture features are extracted from the pseudocolor image, which is comprised of the decomposition components. The spatial relation of the pixels could be obtained from such a PolSAR pseudocolor image. The decomposition features are made up of matrix $C$ or $T$ by polarimetric target decompositions, e.g., Pauli decomposition, Cloude decomposition, Freeman–Durden decomposition, and so on. A number of works have included the computations of these features, as shown in [25] and [80].

### B. Polarimetric Scattering Coding

Polarimetric data returned by PolSAR is stored in the polarimetric scattering matrix. Polarimetric scattering matrix is used to show polarimetric information, in which the element is complex value. Inspired by the one-hot coding [81] and hash coding [82], we learned from the idea of position encoding and mapping relationship and proposed the polarimetric scattering coding for complex matrix encoding. We assume that $z = (x + yi)$ is a complex value, and $x$ and $y$ are the real and imaginary parts of $z$, respectively. Considering the sign of $x$ and $y$, there are four possibilities for $z$, and we give a complete encoding as follows, which is called sparse scattering coding $\varphi$ by us:

$$
\varphi(x + yi) = \begin{bmatrix}
x & 0 \\
0 & y
\end{bmatrix}, \quad \text{if } x \geq 0 \text{ and } y \leq 0
$$

$$
\begin{bmatrix}
x & y \\
0 & 0
\end{bmatrix}, \quad \text{if } x \geq 0 \text{ and } y > 0
$$

$$
\begin{bmatrix}
0 & y \\
x & 0
\end{bmatrix}, \quad \text{if } x < 0 \text{ and } y > 0
$$

$$
\begin{bmatrix}
0 & 0 \\
x & y
\end{bmatrix}, \quad \text{if } x < 0 \text{ and } y \leq 0
$$

Fig. 1 and (6) show the details of polarimetric scattering coding, the first column represents the real element, the second column represents the imaginary part of the element, the first line represents the positive element, and negative elements are expressed in the second row. $|\cdot|$ is the absolute value operation.
An example in (6) is given as follows:

\[
\varphi(x + yi) = \begin{bmatrix} x & 0 \\ 0 & |y| \end{bmatrix}, \quad \text{if } x \geq 0 \text{ and } y < 0. \quad (7)
\]

\(\varphi\) represents the function of polarimetric scattering coding, when \(x > 0\) and \(y < 0\). From (1), scattering matrix \(S\) can be written as

\[
S = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix}. \quad (8)
\]

Because \(S\) is a complex matrix, we can write its elements as follows:

\[
S_{HH} = a + bi \quad (9)
\]
\[
S_{HV} = c + di \quad (10)
\]
\[
S_{VH} = e + fi \quad (11)
\]
\[
S_{VV} = g + hi. \quad (12)
\]

In order to facilitate the understanding and explanation, we give a general assumption, where \(a, b, c, d, e, f, g > 0\). This assumption can consider the characteristics of the PolSAR data. For instance, some PolSAR data format is int16 (−32,768 to +32,767). Polarimetric scattering coding of the scattering matrix \(S\) can be given, which is called a polarimetric scattering coding matrix

\[
\varphi(S) = \varphi\left( \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \right)
\]

\[
\varphi\left( \begin{bmatrix} a + bi & c + di \\ e + fi & g + hi \end{bmatrix} \right)
\]

\[
\begin{bmatrix} a & b & 0 & 0 \\ 0 & 0 & |c| & |d| \\ 0 & 0 & |e| & |f| \\ 0 & |g| & 0 \end{bmatrix}. \quad (13)
\]

Based on this new coding way, we can get the polarimetric scattering coding matrix, which is a 2-D sparse matrix. We also avoid transforming a complex matrix into a 1-D vector, as shown in Fig. 2. This will bring great convenience to the processing of PolSAR data. The proposed polarimetric coding could show more information and is easy to generate and restore polarimetric covariance matrix. Next to it, we proposed a new corresponding classification method named polarimetric convolutional network.
of $s1 \times s2$. Fourth, the feature maps generated from the two-layer convolutional networks can be fed into a FCN whose input size is the same as the output size. Fifth, another pathway is also a FCN, and the input data is the raw polariometric data $I$. Sixth, the part of feature aggregation is designed to fuse the two kinds of the feature. The last-layer feature maps of the two pathway are stacked together as new feature maps, the size of the new feature maps is $s1 \times s2 \times (2 \times M)$, and $M$ is the number of feature maps in each pathway. At last, in order to use softmax to classify data into $C$ classes, we add a convolutional layer to reduce the dimension of the new feature maps from $(2 \times M)$ to $C$.

Above all, the whole network is an end-to-end network. For the entire network architecture, the loss function is a sum over the spatial dimensions of the final layer. We can use stochastic gradient descent to optimize it.

In the proposed method, the model mainly includes a polarimetric scattering coding layer, a convolutional layer, and a deconvolutional layer. The details of the polarimetric scattering coding layer are in Section II-B. The convolutional layer can be defined as follows:

$$O_{ij} = F \left[ \sum_{m_i,m_j=0}^{k} G(X_{s_i+m_i,s_j+m_j}) \right]$$

where $O_{ij}$ is the output map of the convolutional layer in $i$ row and $j$ column, $F$ denotes the normalized function, and rectified linear unit is a good choice and used in this paper. $G$ is the convolutional function, $k$ is the size of the kernel, $X$ denotes the input pixels from the form layer, and $s$ denotes the sampling stride.

The deconvolution layer is composed of upsampling and convolutional layers, and an upsampling layer corresponds to a max-pooling one in the downsampling stage. Those layers upsample feature maps using the max-pooling indexes from their corresponding feature maps. The upsampled maps are then convolved with a set of trainable convolutional kernels to produce dense feature maps.

At last, we compute the energy function by a pixelwise softmax over the final feature maps and the cross entropy loss function. The softmax is defined as

$$p_c(x) = \frac{\exp(a_c(x))}{\sum_{c=1}^{C} \exp(a_c(x))}$$

where $a_c(x)$ denotes the activation in the $c$'th feature channel at the pixel position $x \in \Omega$ with $\Omega \subset \mathbb{Z}^2$, $C$ is the number of land cover categories, and $p_c(x)$ is the approximated maximum function, which represents the probability of each pixel for each category. The cross entropy calculates the deviation of $p_{i}(x)$ at each position

$$E = \sum_{x \in \Omega} \log(p_{i}(x))$$

where $i: \Omega \rightarrow \{1, \ldots, C\}$ is the ground truth of each pixel, and $i(x)$ is the label in the location of $x$.

III. EXPERIMENTS

In this section, four PolSAR images are described in detail, which are used to verify the performance of the proposed algorithm. The four images have significant representativeness, obtained from two airborne systems and two cities, and the details are listed in Tables I–V. The paramenter settings of the proposed method are discussed. Meanwhile, evaluation metrics are given.

A. Data Set Description

For our experiments and evaluations, we select four PolSAR images from an airborne system (NASA/JPL-Caltech AIRSAR) and a spaceborne system (Canadian Space Agency RADARSAT-2). The AIRSAR supports full polarimetric modes for C, L, and P-bands, where we focus on the L and C-bands. RADARSAT-2 works in C-band with the
support of full polarimetric mode. The four selected PolSAR pseudoimages are from two different areas, including Flevoland, The Netherlands, and the San Francisco bay area, CA, USA. In the four PolSAR images, the first two images come from Canadian Space Agency RADARSAT-2 [83], as shown in Figs. 4 and 5, and the latter two are from NASA/JPL-Caltech AIRSAR [84]. We consider that this setup can fully test the performance of the algorithm over a variety of PolSAR images in terms of the system (AIRSAR and RADARSAT-2), the underlying classification problem, and the operative band (C and L). The information about the four images is shown in the following.

1) San Francisco, RADARSAT-2, C-Band: The area around the bay of San Francisco with the golden gate bridge is probably one of the most used scenes in PolSAR image classification over the past decades. It provides a good coverage of both natural (e.g., water and vegetation) and man-made targets (e.g., high density, low density, and developed). This RADARSAT-2 fully PolSAR image at fine quad-pol mode (8-m spatial resolution) was taken in April 2, 2008. The selected scene is a 1380 × 1800-pixel subregion. The Pauli-coded pseudocolor image, the used ground truth data, and the color code are shown in Fig. 6. In the ground truth map in Fig. 6(b), there are five kinds of objects, including developed, high density, low density, water, and vegetation, which can be simply written as c1–c5. The size of original image is 2820 × 14416, which is a relatively large scene with great research significance. Subarea coordinate is (7326:9125, 661:2040), which was shown in Fig. 4. This coordinate can help researchers easily find the area and know the source of the data. Table I shows the numbers of the train and test samples.

2) Flevoland, Radarsat-2, C-Band: This RADARSAT-2 fully PolSAR image at one quad-pol mode (8-m spatial resolution) of Flevoland, The Netherlands, was taken in April 2, 2008. The selected scene is a 1635 × 2375-pixel subregion, which mainly contains four terrain classes: 1) woodland/forest; 2) cropland; 3) water; and 4) urban area. The Pauli

![Fig. 4. Pauli RGB image of the San Francisco big figure. The whole data of this image is downloaded from the RADARSAT-2 web [83], and the area in box is often used. The area coordinate is (7326:9125, 661:2040).](image1)

![Fig. 5. Pauli RGB image of the Flevoland big figure. The whole data of this image is downloaded from the RADARSAT-2 web [83], and the area in box is often used. The area coordinate is (4061:6435, 97:1731).](image2)

![Fig. 6. San Francisco pseudoimage and ground truth, Radarsat-2.](image3)

(a) SanFrancisco image. (b) Ground truth image. (c) Color code.

| LAND CLASSES AND NUMBERS OF PIXELS IN THE FIRST DATA SET |
|--------------------------------------------------------|
| class code | name          | No. of training samples | No. of testing samples |
|------------|---------------|-------------------------|------------------------|
| 1          | Water         | 1000                    | 852078                 |
| 2          | Vegetation    | 1000                    | 237237                 |
| 3          | High-Density Urban | 1000   | 351181                 |
| 4          | Low-Density Urban | 1000  | 282975                 |
| 5          | Developed     | 1000                    | 80616                  |
color-coded image, the ground truth data, and the color code are shown in Fig. 7. The size of original image is 2820 × 12944. Subarea coordinate is (4061:6435, 97:1731), as shown in Fig. 5. Table II shows the number of the train and test samples.

3) San Francisco, AIRSAR, L-Band: This PolSAR image of San Francisco bay has been used in many literature works, and Fig. 8 shows the Pauli RGB image, the ground truth map, and the color code. The size of this image is 900 × 1024. The spatial resolution is 10 m for 20 MHz. Pixels in this image can be classified into five categories, and the abbreviated letters c1–c5 indicate the categories of mountain, ocean, urban, vegetation, and bare soil, respectively. Table III shows the number of the train and test samples.

4) Flevoland, AIRSAR, L-Band: The PolSAR image of Flevoland is shown in Fig. 9(a), there are 15 categories in the ground truth map in Fig. 9(b), and the color code is shown in Fig. 9(c). It is the picture with the most kinds of objects in the public PolSAR data collection at present. The spatial resolution is 10 m for 20 MHz. The size of this PolSAR data is 750 × 1024. There are 15 kinds of objects to be identified, including stem beans, rapeseed, bare soil, potatoes, beet, wheat2, peas, wheat3, lucerne, barley, wheat, grasses, forest, water, and building. These objects are simply written as c1–c15 in this paper. The numbers of the train and test samples are shown in Table IV.

B. Experimental Setup

Traditional polarimetric features are extracted to compare with the polarimetric scattering coding in the proposed method and contrast algorithms. In this paper, we used a 22-D feature
vector as the traditional polarimetric feature, including the upper right element’s absolute value of the $3 \times 3$ polarimetric coherency matrix $\mathbf{T}$, the upper right element’s absolute value of the $3 \times 3$ polarimetric covariance matrix $\mathbf{C}$, three components of Pauli decomposition, three components of Freeman decomposition [14], and four components of Yamaguchi decomposition [8], expressed as PF22. The feature extracted by polarimetric scattering coding can be written as SSCF. In the experiment, we consider two aspects for comparison experiments: feature extraction and classifier design. For feature extraction, the common feature PF22 and the proposed SSCF are adopted. For classifier design, the traditional algorithms’ MLD [85] and SVM [86] are used to classify the image data, and the deep learning algorithms also be adopted, such as a CNN [49], PFDCN [67], and FCN [87]. Based on these considerations, the comparison methods include PF22-MLD, PF22-CNN, PF22-FCN, and PFDCN. The proposed method can be represented as PCN.

In contrast algorithms, in order to make as fair a comparison as possible, the standard deviation has been added in the result through 10 random experiments, and the parameters of CNN and FCN are set to the same as possible. The CNN is structured as follows: the first convolutional layer filters the input image patch with 64 kernels of size $5 \times 5$; the second convolutional layer takes the output of the first convolutional layer as the input and filters it with 128 kernels of size $5 \times 5$. The third, fourth, and fifth convolutional layers are connected to one another without any intervening pooling or normalization layers. The third convolutional layer has 256 kernels of size $5 \times 5$ connected to the (normalized and pooled) outputs of the second convolutional layer. The fourth convolutional layer has 256 kernels of size $5 \times 5$, and the fifth convolutional layer has 128 kernels of size $5 \times 5$. At last, the fully connected layer has 100 neurons.

The structure of the FCN is as follows: the relevant setting of the first five layers is the same as that of the CNN, the latter four layers is corresponding to the first five layers and the original image size. A skip layer has also been adopted at the FCN’s sixth to eighth layers to keep image edge information. In Fig. 3, the last feature maps of the two FCNs are stacked. And then, the output layer then performs a $1 \times 1$ convolution to produce the same number of feature maps as the number of classes in each data set. In the case that the input is polarimetric scattering coding matrix, we added a specially designed two-layer network in front of the CNN and FCN: the first layer is 32 kernels of size $4 \times 4$ with a stride of 1 pixel and the second layer is 64 kernels of size $4 \times 4$ with a stride of 2 pixels. Through this network, we can get the classification map with the same size as the original image. We train the model in a single step of optimization, and the weight is initialized by Glorot and Bengio [88]. The stochastic gradient descent with momentum 0.9 is used to train the weights of the model. The initial learning rate is set to 0.01, and the train batch size is 1. In the experiments, we randomly selected 1000 points per class for training, and the remaining labeled samples were used for testing. In addition to training data, other samples do not participate in training and learning. In SVM, the kernel function is the radial basis function. The other samples do not participate in training and learning.

All the experiments are running on an HP Z840 workstation with an Intel Xeon(R) CPU, a GeForce GTX TITAN X GPU, and 64G RAM under Ubuntu 16.04 LTS. All of these methods are implemented using the deep learning framework of TensorFlow.

### C. Evaluation Metrics

For evaluating the classification performance, the experiment results were assessed by single class recall rate, overall accuracy (OA), average accuracy (AA), and kappa coefficient (Kappa). Overall accuracy can be defined as follows:

$$OA = \frac{M}{N}$$  \hspace{1cm} (17)

where $M$ is the number of classified correctly, and $N$ is the total number of samples. AA can also be written as follows:

$$AA = \frac{1}{C} \sum_{i=1}^{C} \frac{M_i}{N_i}$$  \hspace{1cm} (18)

where $C$ is the number of categories, $i$ is the category index, $M_i$ is the number of correct samples for the $i$th categories, and $N_i$ is the number of samples in the $i$th class. $(M_i/N_i)$ is the $i$th class recall rate. The formula of kappa coefficient can be written as follows:

$$Kappa = \frac{OA - P}{1 - P} = \frac{1}{N^2} \sum_{i=1}^{C} \bar{Z}(i, :) \ast \bar{Z}(, i)$$  \hspace{1cm} (19)

where $OA$ is the overall accuracy, $Z$ is the confusion matrix, $\bar{Z}(i, :)$ is the sum of the $i$th row elements, and $\bar{Z}(, i)$ is the

### TABLE IV

| Class Code | Name | No. of Training Samples | No. of Testing Samples |
|------------|------|-------------------------|------------------------|
| 1          | Water| 1000                    | 12252                  |
| 2          | Barely| 1000                    | 6955                   |
| 3          | Peas  | 1000                    | 8582                   |
| 4          | Stem beans| 1000                  | 5338                   |
| 5          | Beet  | 1000                    | 9033                   |
| 6          | Forest| 1000                    | 17044                  |
| 7          | Bare soil| 1000                  | 4109                   |
| 8          | Grapes| 1000                    | 6058                   |
| 9          | Rapeseed| 1000                 | 12863                  |
| 10         | Locorene| 1000                   | 9181                   |
| 11         | Wheat2| 1000                    | 10159                  |
| 12         | Wheat1| 1000                    | 15386                  |
| 13         | Buildings| 200                    | 555                    |
| 14         | Potatoes| 1000                  | 15156                  |
| 15         | Wheat3| 1000                    | 21241                  |

### TABLE V

| Name          | Experimental ID | Abv | Date       | Spatial Resolution | Dimensions | Label |
|---------------|-----------------|-----|------------|--------------------|-------------|-------|
| San Francisco | RADARSAT-2 C    | SP-L| Apr 2003   | 100x100            | 3            | 3     |
| Pfaffiand     | RADARSAT-2 C    | PIV-E| Apr 2003   | 100x100            | 3            | 3     |
| San Francisco | AIRSAR C        | SP-L| Apr 1999   | 100x100            | 3            | 3     |
| Pfaffiand     | AIRSAR C        | PIV-E| Apr 1999   | 100x100            | 3            | 3     |
| Pfaffiand     | AIRSAR C        | PIV-E| Aug 1999   | 100x100            | 3            | 3     |
sum of the $i$th column elements. $N$ is the total number of samples.

IV. RESULTS AND DISCUSSION

Our experimentation will be separated into three parts. In the first part, we demonstrate the proposed algorithm using the two data sets from the RADARSAT-2. In the second part, we evaluate the proposed algorithm using the other two data sets from the AIRSAR. In the third part, we give a significant analysis of the results. Finally, we show the computation times of the proposed algorithm and comparison algorithms in part four. The results are shown in Figs. 10–15 and Tables VI–IX.

A. Experiment With the Data Set From the RADARSAT-2

The classification results from different algorithms are shown in Figs. 10(a)–(g) and 11(a)–(g), and the accuracies for each class are listed in Tables VI and VII, respectively. Fig. 14(a) and (b) shows a clear contrast of classification accuracy, and the trend line is generated by the data in Tables VI and VII. It can be seen that the performance of the proposed method is better than others. The classification accuracies are higher than the contrast algorithms. The classification maps are closer to the ground truth. Trend chart of accuracy in Fig. 14(a) and (b) shows that the PCN is better than others.

Fig. 10 shows that PF22-MLD and PF22-SVM cannot distinguish high-density urban well, and mistake much high-density urban for low-density urban, which is due to the fact that the two objects are similar, and the method is difficult to distinguish its characteristics. The result of the FCN is better than the CNNs, which can be seen from the contiguous area of two classes, such as coastal area. The classification map...
of PF22-CNN is worse than that of SSC-CNN in recognizing mixed terrain. Meanwhile, the result of PCN is better than that of PF22-FCN and PFDCN, and the proposed algorithms PFDCN and PCN almost detect all terrain. Especially, the low-density urban misclassified in high-density urban is almost completely corrected. Table VI shows that the accuracy of deep learning methods is higher than that of the conventional methods, with a difference of 2–6 percentage points. It can be clearly seen that the high-density urban is misclassified. The classification accuracy of high-density urban is 80.54%.

In Fig. 11, there are only four types of objects that need to be distinguished, namely, urban, water, forest, and cropland. Water is the most easily recognizable relatively, and inland water can also be judged right. It is difficult to judge fake forest in urban and cropland. Some urban and cropland are mistakenly divided into forests in the classification maps obtained by PF22-MLD and PF22-SVM. The classification methods based on convolution get the better classification maps, and some homogeneous areas can be correctly obtained. Furthermore, the performance of the proposed method is more outstanding. The urban and cropland can be seen completely, and there is almost no noise. Meanwhile, Table VII gives a quantitative result, PF22-MLD and PF22-SVM get a low accuracy, especially in urban and cropland areas, and other methods show an upward trend. PCN has the highest accuracy.

B. Experimental With the Data Set From the AIRSAR

In this section, Figs. 12(a)–(g) and 13(a)–(g) show the classification maps from the proposed method and the contrast algorithms. Tables VIII and IX list the statistical classification accuracy from the proposed method and the compared algorithms. A clear exhibition of classification accuracy is shown in Fig. 15(a) and (b).
From the above-mentioned results, we can get that the proposed method has the best performances. The misclassified pixels are much less in the classification maps. The classification accuracies are 95.82 and 95.59, respectively. It is 2–8 percentage points higher than other algorithms.

In Fig. 12, there are five kinds of objects to be determined, including mountain, ocean, urban, vegetation, and bare soil. There is an island on the upper right of the image, called Alcatraz Island, and it did not appear in Fig. 12(a). On contrast, the island was distinguished as a mountain in Fig. 12(b)–(e). Although the island has not been marked, it is still detected. Fig. 12(g) shows a clearer outline. What is more, in the bare soil detection task, the difference of the algorithm performance can also be shown. The shape and area of the soil can be described more accurately, which can be seen in Fig. 12(g). The accuracy is listed in Table VIII, the OAs of PF22-MLD and PF22-SVM are only 87.50% and 87.24%, other algorithms are larger than 0.90, and the proposed algorithm is as high as 95.82.

Fig. 13 shows the classification results of the fourth data set. This data set is shown in Fig. 9 and contains 15 kinds of objects, so many kinds of labeled data are extremely rare. Naturally, the problem will increase in difficulty. For instance, it is difficult to distinguish wheat1, wheat2, and wheat3,
some wheat3 was wrongly divided into wheat1 and wheat2 in Fig. 13(a)–(c), Fig. 13(a) is the largest, and Fig. 13(b) and (c) is relatively few. The proposed algorithm can judge almost all pixels correctly, which is shown in Fig. 13(g). Similarly, some potatoes are wrongly classified as peas, but the proposed method can solve this problem. In Table IX, the above-mentioned experimental phenomena can be seen accurately through numbers, such as the classification accuracy of wheat1, wheat2, and wheat3 get a 5 percentage point increase.

We can see that the proposed approach outperforms the compared methods. It indicates that the encoded data through polarimetric scattering coding is easier to be identified and distinguished. At the same time, we can find that the modified FCN has a better classification performance than the conventional convolutional network.

C. Significance Analysis

In this section, we give a significance analysis about the experiment results by T-test score. The analysis results are shown in Table X. In the table, low values are good and significant. If the value is less than 0.5, it means the result is significant. It can be seen that the proposed approach has a significant advantage.

D. Computation Times

In Table XI, we show the computation times achieved by different methods on the data sets. Factors that affect the computation times include the size of image, the complexity of the data sets, and the methods. The computation times then increase with the size of image and the complexity of the data sets. When compared with Figs. 6 and 8, the size of the image is a key factor that impacts the computational efficiency. From Figs. 7 and 9, it can be easily seen that the complexity of the data sets is an import factor for the computational efficiency, since there are 15 categories in Fig. 9. At last, the speed of PF22-FCN and PCN is faster than others, and the main reason is that this algorithm does not need sliding window to calculate the pixels.

V. CONCLUSION

In this paper, we proposed a new PolSAR image classification method named polarimetric convolutional network, which is based on the polarimetric scattering coding and FCN. Polarimetric scattering coding can keep the structure information of scattering matrix and avoid breaking the matrix into a 1-D vector.

Coincidentally, a convolution network needs a 2-D input, where polarimetric scattering coding matrix meets this condition. We design an improved FCN to classify data encoded...
by polarimetric scattering coding. In order to make the experiment more effective and complete, the experimental data sets consist of four data sets from two satellites, and the contrast algorithms include traditional methods and latest methods. Experimental results show that the proposed algorithm PCN is robust and get a better result, the classification maps of the proposed method are very close to the ground truth maps, and the classification accuracies are higher than contrast algorithms. These results are mainly due to the fact that the proposed algorithm can preserve the structural semantic information of the image in the raw data. In contrast algorithms, PF22-MLD, PF22-SVM, and PF22-CNN do not give full consideration to the structural semantic information, SSC-CNN only thinks about the structural semantic information of the raw data, and PF22-FCN only considers the structural semantic information of model. The results of PFDNCN outperform other comparison algorithms. Experimental results can also confirm the above inference. From comparative experiments, we also found that this polarimetric coding is effective. For this coding, the performance of the designed classification network is better.

ACKNOWLEDGMENT

The authors would like to thank the NASA/JPL-Caltech and the Canadian Space Agency for kindly providing the polarimetric AIRSAR/RADARSAT2 data used in this paper. They would also like to thank the anonymous reviewers for their helpful comments.

REFERENCES

[1] H. Wang, R. Magagi, and K. Goita, “Comparison of different polarimetric decompositions for soil moisture retrieval over vegetation covered agricultural area,” Remote Sens. Environ., vol. 199, pp. 120–136, Sep. 2017.
[2] K. Voormansik, T. Jagdhuber, K. Zalite, M. Noorma, and I. Hajnsek, “Observations of cutting practices in agricultural grasslands using polarimetric SAR,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 9, no. 4, pp. 1382–1396, Apr. 2016.
[3] F. Liu et al., “Hierarchical semantic model and scattering mechanism based PolSAR image classification,” Pattern Recognit., vol. 59, pp. 325–342, Nov. 2016.
[4] C. Dickinson, P. Siqueira, D. Clewley, and R. Lucas, “Classification of forest composition using polarimetric decomposition in multiple landscapes,” Remote Sens. Environ., vol. 131, pp. 206–214, Apr. 2013.
[5] E. Kroogager, “New decomposition of the radar target scattering matrix,” Electron. Lett., vol. 26, no. 18, pp. 1525–1527, Aug. 1990.
[6] R. Touzi and F. Charbonneau, “Characterization of target symmetric scattering using polarimetric SARs,” IEEE Trans. Geosci. Remote Sens., vol. 40, no. 11, pp. 2507–2516, Nov. 2002.
[7] W. L. Cameron and L. K. Leung, “Feature motivated polarization scattering matrix decomposition,” in Proc. IEEE Int. Radar Conf., May 1990, pp. 549–557.
[8] Y. Yamaguchi, A. Sato, W.-M. Boerner, R. Sato, and H. Yamada, “Four-component scattering power decomposition with rotation of coherency matrix,” IEEE Trans. Geosci. Remote Sens., vol. 49, no. 6, pp. 2251–2258, Jun. 2011.
[9] S. W. Chen, X. S. Wang, S. P. Xiao, and M. Sato, “General polarimetric model-based decomposition for coherency matrix,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 3, pp. 1843–1855, Mar. 2014.
[10] S.-W. Chen, X.-S. Wang, Y.-Z. Li, and M. Sato, “Adaptive model-based polarimetric decomposition using PolInSAR coherence,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 3, pp. 1705–1718, Mar. 2014.
[11] M. Mahdianpari et al., “Fisher Linear Discriminant Analysis of coherency matrix for wetland classification using PolSAR imagery,” Remote Sens. Environ., vol. 206, pp. 300–317, Mar. 2018.
[12] J. R. Huyven, Phenomenological Theory of Radar Targets. 1970.
[13] A. Freeman and S. L. Durden, “Three-component scattering model to describe polarimetric SAR data,” in Radar Polarimetry, vol. 1748, pp. 213–225, Feb. 1993.
[14] Y. Yamaguchi, T. Moriya, M. Ishido, and H. Yamada, “Four-component scattering model for polarimetric SAR image decomposition,” IEEE Trans. Geosci. Remote Sens., vol. 43, no. 8, pp. 1699–1706, Aug. 2005.
[15] S. R. Cloude and E. Pottier, “A review of target decomposition theorems in radar polarimetry,” IEEE Trans. Geosci. Remote Sens., vol. 34, no. 2, pp. 498–518, Mar. 1996.
[16] 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.
[17] J. S. Lee, T. L. Ainsworth, and Y. Wang, “Generalized polarimetric model-based decompositions using incoherent scattering models,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 5, pp. 2474–2491, May 2014.
[18] N. Besic, G. Vasilje, J. Chanussot, and S. Stankovic, “Polarimetric incoherent target decomposition by means of independent component analysis,” IEEE Trans. Geosci. Remote Sens., vol. 53, no. 3, pp. 1236–1247, Mar. 2014.
[19] H. Aghababae and M. R. Sahebi, “Incoherent target scattering decomposition of polarimetric SAR data based on vector model roll-invariant parameters,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 8, pp. 4392–4401, Aug. 2016.
[20] S.-W. Chen, Y.-Z. Li, X.-S. Wang, S.-P. Xiao, and M. Sato, “Modeling and interpretation of scattering mechanisms in polarimetric synthetic aperture radar: Advances and perspectives,” IEEE Signal Process. Mag., vol. 31, no. 4, pp. 79–89, Jul. 2014.
[21] S.-W. Chen, X.-S. Wang, and M. Sato, “Uniform polarimetric matrix rotation theory and its applications,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 8, pp. 4756–4770, Aug. 2014.
[22] F. Xu, Q. Song, and Y. Q. Jin, “Polarimetric SAR image factorization,” IEEE Trans. Geosci. Remote Sens., vol. 55, no. 9, pp. 5026–5041, Sep. 2017.
[23] C. Tao, S. Chen, Y. Li, and S. Xiao, “PolSAR land cover classification based on roll-invariant and selected hidden polarimetric features in the rotation domain,” Remote Sens., vol. 9, no. 7, pp. 660, 2017.
[24] S. Uhlmann and S. Kiranyaz, “Integrating color features in polarimetric SAR image classification,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 4, pp. 2197–2216, Apr. 2014.
[25] G. D. Grandi, J. S. Lee, and D. L. Schuler, “Target detection and texture segmentation in polarimetric SAR images using a wavelet frame: Theoretical aspects,” IEEE Trans. Geosci. Remote Sens., vol. 45, no. 11, pp. 3437–3453, Nov. 2007.
[26] X. Ma, H. Shen, J. Yang, L. Zhang, and P. Li, “Polarimetric-spatial classification of SAR images based on the fusion of multiple classifiers,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 7, no. 3, pp. 961–971, Mar. 2014.
[27] T. Zou, W. Yang, D. Dai, and H. Sun, “Polarimetric SAR image classification using multifeatures combination and extremely randomized clustering forests,” EURASIP J. Adv. Signal Process., vol. 2010, Jan. 2010, Art. no. 4.
[28] B. Ren, B. Hou, J. Zhao, and L. Jiao, “Unsupervised classification of polarimetric SAR image via improved manifold regularized low-rank representation with multiple features,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 10, no. 2, pp. 580–595, Feb. 2017.
[29] H. Liu et al., “Semisupervised feature extraction with neighborhood constraints for polarimetric SAR classification,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 9, no. 7, pp. 3001–3015, Jul. 2016.
Xu Tang (M’17) received the B.S., M.S., and Ph.D. degrees from Xidian University, Xi’an, China, in 2007, 2010, and 2017, respectively, where he is currently pursuing the Ph.D. degree in circuit and systems.

He is currently a member of the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education, Xi’ an University. His current research interests include remote sensing image processing, remote sensing image content-based retrieval, and reranking.

Qigong Sun (S’15) received the B.Eng degree in intelligence science and technology from Xidian University, Xi’an, China, in 2015, where he is currently pursuing the Ph.D. degree in circuit and system.

He is currently a member of the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education, and the International Research Center for Intelligent Perception and Computation, Xidian University. His research interests include deep learning and image processing.

Xu Liu (S’15) received the B.Sc. degree in mathematics and applied mathematics from the North University of China, Taiyuan, China, in 2013. He is currently pursuing the Ph.D. degree in circuit and system with Xidian University, Xi’an, China.

He is currently a member of the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education, the International Research Center for Intelligent Perception and Computation, and the Joint International Research Laboratory of Intelligent Perception and Computation. He is the Chair of the IEEE Xidian University Student Branch, Xidian University. His research interests include machine learning, deep learning, and image processing.

Licheng Jiao (F’17) 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.

Since 1992, he has been a Professor with the School of Electronic Engineering, Xidian University, Xi’an, where he is currently the Director of the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education of China. His research interests include image processing, natural computation, machine learning, and intelligent information processing.

Dan Zhang received the B.S. degree in intelligent science and technology and the M.S. degree in circuits and systems from Xidian University, Xi’an, China, in 2011 and 2016, respectively.

She was a Counselor with Xidian University from 2011 to 2013, where she is currently with the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education. Her research interests include image processing and machine learning.