





C^2N^2 : Complex-Valued Contourlet Neural Network

Mengkun Liu , Licheng Jiao , *Fellow, IEEE*, Xu Liu , *Member, IEEE*, Lingling Li , *Senior Member, IEEE*, Fang Liu , *Senior Member, IEEE*, Shuyuan Yang , *Senior Member, IEEE*, Yuwei Guo , *Senior Member, IEEE*, and Puhua Chen , *Senior Member, IEEE*

Abstract—Complex-valued convolutional neural networks (CV-CNN) have recently gained recognition in feature representation learning. It implements the repeated application of the operations in convolution, local average pooling, and the absolute value of the resulting vectors. However, it is only conducted in the complex spatial domain, and lacks effective representation of directionality, singularity, and regularity in the complex spectral domain for anomaly detection of images. This is the key to feature learning representation of high-order singularity. To solve this problem, a complex-valued contourlet neural network (C^2N^2) is proposed in this article. It is novel in this sense that, different from the CV-CNN in the spatial domain, the spectral stream of C^2N^2 can enhance the multiresolution sparse representation of nonsubsampling contourlet (NSCT) with multiscales and multidirections for images. Furthermore, the spectral feature integration module is proposed to capture the statistical properties of the NSCT coefficients. It is shown that the proposed network can improve the distinguishability of feature learning and classification ability in theoretical analysis and experiments on three benchmark datasets (Flevoland, Xi'an, and Germany) compared with developed methods. Polarimetric synthetic aperture radar image classification is widely used in the fields of agriculture, forestry, and military. It must be emphasized that there is potential in effective feature learning representation and the generalization capability of C^2N^2 in deep learning, recognition, and interpretation.

Index Terms—Classification, complex-valued contourlet neural network (C^2N^2), nonsubsampling contourlet (NSCT) transform, polarimetric synthetic aperture radar (PolSAR).

Manuscript received 24 August 2023; revised 27 November 2023; accepted 16 January 2024. Date of publication 26 January 2024; date of current version 13 February 2024. This work was supported in part by the Key Scientific Technological Innovation Research Project by Ministry of Education, in part by the Joint Funds of the National Natural Science Foundation of China under Grant U22B2054, in part by the National Natural Science Foundation of China under Grant 62076192, Grant 61902298, Grant 61573267, Grant 61906150, Grant 62276199, and Grant 62276201, in part by the 111 Project, in part by the Program for Cheung Kong Scholars and Innovative Research Team in University under Grant IRT 15R53, in part by the ST Innovation Project from the Chinese Ministry of Education, in part by the Key Research and Development Program in Shaanxi Province of China under Grant 2019ZDLGY03-06, in part by the China Postdoctoral fund under Grant 2022T150506, and in part by the Shaanxi Provincial Department of Education 2020 Key Scientific Research Program through New Think Tank Project under Grant 20JT021. (*Corresponding author: Licheng Jiao.*)

The authors are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Joint International Research Laboratory of Intelligent Perception and Computation, School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: mengkunliu31@163.com; lchjiao@mail.xidian.edu.cn; xuliu361@163.com; llli@xidian.edu.cn; f63liu@163.com; syyang@xidian.edu.cn; yuweiguo18@126.com; phchen@xidian.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3358846

I. INTRODUCTION

CONVOLUTIONAL neural networks (CNNs), one of the representative methods in deep learning, are widely used in the image processing field of classification [1], [2], [3], [4], segmentation [5], [6], [7], [8], [9], and detection [10], [11], [12], [13]. By optimizing spatial filters in the neural network, models are provided with effective feature expression and good performance on conventional natural images. Although CNNs have been commonly studied in most fields, the CNN-based research on remote sensing, such as the polarimetric synthetic aperture radar (PolSAR) image interpretation, is relatively few.

As an active microwave imaging system, PolSAR can work under all-time and all-weather conditions. Various polarization combination modes make it enjoy a richer ability for land feature acquisition [14]. PolSAR image classification, one of the most fundamental issues in interpretation, is a pixel-level classification task that predicts the class a given sample belongs to. It is widely used in terrain classification [15], environment monitoring [16], and agriculture assessment [17]. The unsupervised terrain and land-use classification algorithm is proposed [15] using the PolSAR data, which combines the scattering model-based decomposition and the maximum-likelihood classifier. The pixel-level classification model MB-U2-ACNet [16] was proposed by dual-PolSAR images integrated with environmental information for dynamic intertidal zone land cover classification. The detection of grass-cutting practices is achieved by analyzing the parameters calculated from TerraSAR-X dual-polarimetric HH/VV and RADARSAT-2 fully PolSAR data [17].

The following is a brief review of several common methods for PolSAR image classification, which can be divided into the following three types.

Models based on polarimetric scattering characteristics: The paradigms of physics-based feature extraction and classifier selection are generally applied [18], [19], [20], [21], [22]. The polarimetric target decomposition is commonly considered for feature extraction, such as Pauli decomposition [18], Freeman–Durden decomposition [19], and the combination of different features [23], [24], [25]. The polarimetric scattering coding way [22] is designed for the feature representation of the polarimetric scattering matrix. The obtained features are then sent to the classifiers for classification, such as fuzzy cluster [26], Wishart classifier [15], and support vector machine (SVM) [27].

Models based on the statistical property: The statistical analysis was reported by Goodman [28] according to a multivariate Gaussian distribution in the complex-valued domain. Based on the Wishart distribution, Lee et al. [29] introduced a maximum

likelihood classifier for PolSAR imagery segmentation in line with the terrain types. They also introduced the target decomposition theory for unsupervised classification [30]. Specifically, Cloude and Pottier's method [31] was used to initially classify the PolSAR image with the Wishart method [29]. In [32], three sources of statistical information (context, dimensionality, and model adequacy) were considered to compensate for a relatively poor classification model.

Models based on polarimetric scattering characteristics: It mainly covers deep belief networks (DBN) [33], multilayer autoencoders [34], generative adversarial networks [35], fully convolutional networks (FCN) [36], U-Net [37], and transformer with contrastive learning [5], [38]. The SF-CNN [39] developed a dual-branch CNN for the PolSAR image classification with sample expansion, which alleviates the problem of insufficient labeled training set. The PolMPCNN with a multipath structure [40] is built for adaptive learning of the polarization rotation angles, which are related to different types of land covers.

However, real-valued neural networks only focus on the amplitude information of PolSAR images and ignore phase one. In order to make full use of the phase information, complex-valued contourlet network (C^2N^2) is proposed in this article. The spectral stream of C^2N^2 can enhance the multiresolution sparse representation of nonsampled contourlet (NSCT), and the spectral feature integration module (SFIM) is proposed to capture the statistical properties of the NSCT coefficients.

It is important to mention that the proposed method is consistent with the biological characteristics of the brain: sparsity [41], learnability [42], selectivity [43], directionality [44], knowledge [45], and diversity [46]. Besides, the model also reflects physical interpretability. The "data-driven multiscale windowed spectra" is calculated by the complex-valued convolutional networks. It characterizes certain stochastic processes (including patterns and textures) common in the modeling of natural images [47]. We motivate the construction of such multiscale spectra via the nonsampled directional filter bank (NSDFB) of contourlet, which compacts the energy into a few coefficients with sparsity.

The following three aspects are mainly considered in this article.

- 1) *Weight Self-learning of Features in the Spectral Domain:* The multiresolution analysis (MRA) of NSCT is integrated into the CV-CNNs, thus the weights of the spectral features are learned adaptively in the forward propagation. Prior knowledge is beneficial to the feature perception ability of the network. Integrating spectral features into the spatial domain can improve the capability of approximation and optimization in the model.
- 2) *Enhancing Features Representation:* Affected by speckle, poor gray resolution but the rich texture information is preserved in the PolSAR image. In this article, the SFIM is proposed, which calculates the mean vector of a series of contourlet coefficient maps and integrates the feature vector into the network. In addition, the spectral features and spatial features are combined to enhance the distinguishability of features. Thus, the impact of speckles on the PolSAR image can be alleviated.

- 3) *Light Weight Structure:* Due to the overcomplete transform and tight frame of NSCT with sparsity, C^2N^2 achieves shallow network structure with few trainable parameters. In addition, the weights in the NSCT kernel increase the interpretability of the network.

C^2N^2 integrates the MRA of NSCT to CNNs in the complex domain. The main contributions of this article can be summarized as follows.

- 1) A novel complex-valued contourlet network C^2N^2 is proposed with a two-stream structure, which combines the information both in the spectral domain and the spatial domain, and improves the distinguishability of feature learning and classification ability.
- 2) To enhance the feature learning representation of high-order singularity in the spectral domain, the NSCT is integrated into complex-valued CNNs to enhance the feature perception ability of the network.
- 3) The SFIM is designed to capture the statistical properties of the NSCT coefficients, which aims to enhance the distinguishability of feature learning and alleviate the impact of speckle.
- 4) The comparative experiments with previously developed methods on three PolSAR databases are conducted quantitatively and qualitatively to verify the effectiveness of our model. Furthermore, the ablation study provided a deeper insight into the proposed method.

The rest of this article is organized as follows. Section II presents a brief review of related work. The proposed C^2N^2 is described in Section III. Section IV provides empirical studies and discussions on three public PolSAR datasets. Finally, Section V concludes this article.

II. RELATED WORK

The literature on PolSAR image classification with complex-valued models and MRA-based models are referenced in our work. For brevity, only the relevant works are discussed.

A. Complex-Valued Models

As reported by Hirose [48], the phase information of the PolSAR image is beneficial for scatterer identification with different types. Therefore, some works attempted to leverage the phase information for PolSAR image classification, such as complex-valued convolutional neural networks (CV-CNNs) [49], where all mathematical operations (data, parameters, and training algorithm) are designed under complex analysis theory. L-CV-DeepLabv3+ [50] is proposed, which consists of CV-Xception, CV-ASPP, and decoder in a lightweight complex-valued network. The complex-valued PDAS [51] first applied the neural architecture search on PolSAR image classification and customized for the data of PolSAR, thus the performance is improved. The triplet complex-valued network [52] classifies PolSAR images by learning and comparing the complex-valued distance, which is computed by the K nearest neighbor. The CV-EPLS [53] extracts nonredundant sparse features of the amplitude and phase information in different polarimetric channels, which is conducive to the subsequent classification. The

CV-SDFCN [54] is a multiscale FCN network based on the U-Net structure, which is defined in the complex-valued domain with stacked dilated convolutions. Qin et al. [55] introduced superpixel oriented into CV-CNN to reduce the computational cost, thus the image details are preserved.

Nevertheless, the mathematical properties of CNNs still remain mysterious due to their accumulated nonlinearities, and the relationship between the network performance and the optimal configuration is not yet clear.

B. MRA-Based Models

With the recent studies on MRA, multiscale filter banks have been integrated into CNNs to explore a wider interpretable network [56], [57], [58], [59], [60]. By exploiting the spatial features and the corresponding spectral features of contourlet, the multiscale joint spatial–spectral feature maps are used for classification [56]. MSCCN uses the multiscale curvelet-scattering module to extract deep features of different scales for remote sensing scene classification [57]. In [58], 3D-DWT features and MRF priors were simultaneously utilized to integrate contextual information for accurate segmentation. Besides, the low-frequency subband and contour subband (LC subbands) were used to feature mining and selection with channel modeling, which strengthens the LC-PSENet [59] with efficiency and flexibility. DSTFN [60] employed the geometric ridgelet filters, which developed from the least squares SVM for feature representation, and then, the sparse features are cascaded to CNNs for SAR image classification. However, the extracted multiscale features are directly cascaded with the corresponding CNN features [56] or scattering features [57], thus the spectral information is not fully mined. In terms of the input in real-valued CNNs, the elements' intensities and combination of complex coherency polarimetric matrix T are used in [58], and the imaginary part and the real part in the T matrix are cascaded in [59], thus the phase information of the PolSAR image is ignored.

III. COMPLEX-VALUED CONTOURLET NEURAL NETWORK

In the following, Section III-A shows the motivation of the proposed method. Section III-B introduces the NSCT sparse representation and explains the advantages of NSCT over contourlet and wavelet. Section III-C presents the details of the components in our model.

A. Motivation

PolSAR image classification is a pixel-level classification task and usually the CNN-based classifiers are adopted. Given a complex-valued signal $x \in \mathbb{C}^2$ for the sample patch of complex coherency polarimetric matrix T in PolSAR image. The complex convolution operation in layer i can be described as $W_i * x + b_i$. It can be seen that the W_i is covariant to translations. The geometric transformations could alleviate this problem by data augmentation and deep architecture. In addition, the pooling operation is robust to a small translation by local average operation, but CNNs still lack the capability of geometric transformation.

Moreover, although neural networks have the ability to approximate any deterministic nonlinear process with little knowledge and no assumptions, sigmoid neural networks have several drawbacks. Typically, the transfer function is of sigmoidal type always accompanied by the local minimums of the loss function in the training procedure. In addition, the values of the weights will vary each time, although minimization of the loss function can be replicated with random initialization [61].

In contrast to classical “sigmoid neural networks,” contourlet has compact support as the basis and approximates the “contour-like” components, which significantly outperforms the wavelet and sigmoid function of CNNs by its higher degree of directional selectivity. Contourlet can represent edges effectively and sparsely, which is the key to PolSAR image classification. The introduction of contourlet into neural networks can enhance the ability of geometric transformation, and avoid the local minimums of the loss function in the early training stage. Through the feature extraction by the multiscale and multidirectional filters of contourlet, the initial values of the network parameters can be constructed efficiently. Since it is approximate to the same weight vector to minimize the loss function each time, efficient initialization can reduce the number of iterations of the network in the training phase.

Therefore, the $\mathbb{C}^2\mathbb{N}^2$ is proposed with the inspiration of CV-CNNs [49]. Since the PolSAR image is represented by the coherency matrix T with complex form. To make full use of the phase information, $\mathbb{C}^2\mathbb{N}^2$ utilizes both amplitude and phase information and integrates the MRA of NSCT in the complex domain. NSCT can construct a shift-invariant image decomposition with flexible multiscales and multidirections. It uses an overcomplete transform through the nonseparable two-channel nonsubsampling filter bank (NSFB) [62]. Unlike wavelet basis with point singularities, the NSCT enjoys singularities distributed along curves and curve surfaces in high-dimensional space. Therefore, the combination of NSCT characteristics and CNNs not only greatly improve the ability of feature representation and geometric transformations, but also increase the mathematical interpretability of the network.

B. Sparsity of NSCT

Research on the human visual system (HVS) shows that the “optimal” image representation method should have the properties of multiresolution, locality, and directionality [63]. Sparse representation of signals can be realized by MRA, such as Fourier [64], wavelet [65], curvelet [66], and contourlet [67]. An effective transform characterizes a given signal f with few sets of basis ϕ_i

$$f = \sum_{i=1}^{\infty} \alpha_i \phi_i \quad (1)$$

for the M -largest $|\alpha_i|$, the approximation order of Fourier is $O(M^{-1/2})$ and wavelet is $O(M^{-1})$. Since the contourlet is a new extension of the curvelet, they have the same approximation errors of $O(M^{-2})$. The contourlet expansion is considered the optimal approximation rate M^{-2} [64] for sparse representation with few coefficients.

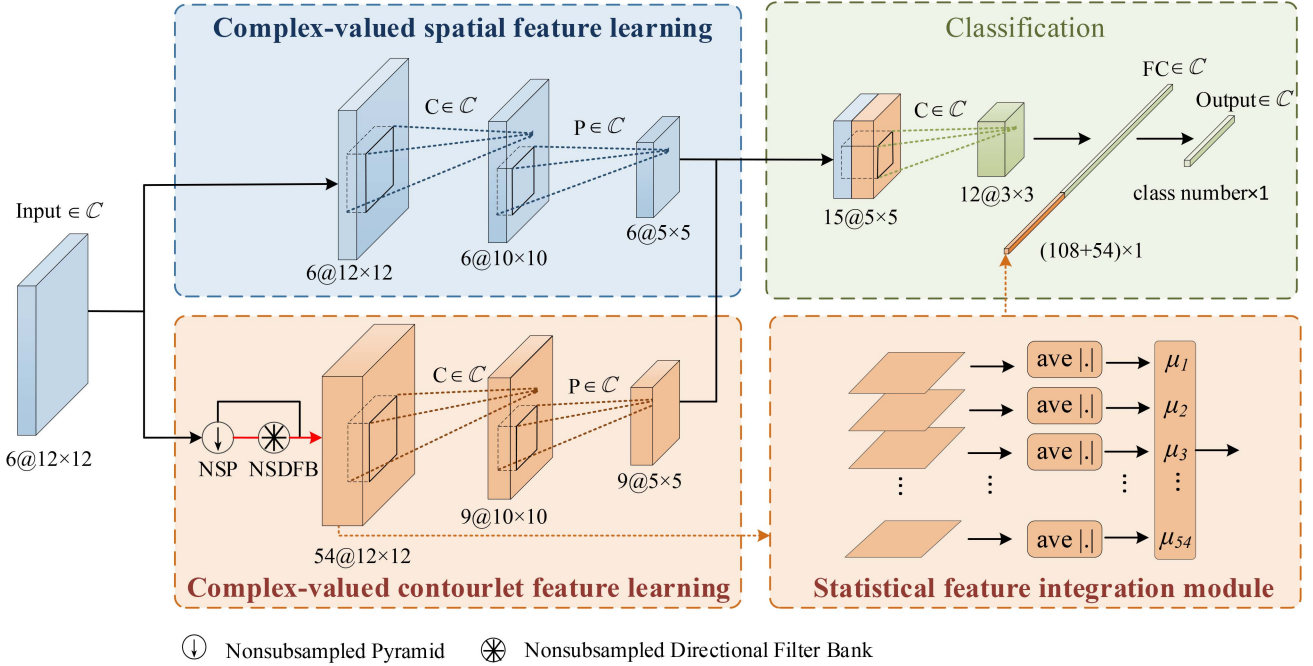


Fig. 1. Overview of the C^2N^2 . It is a two-stream architecture, one is the spatial feature learning branch, and the other is the spectral feature learning branch, where all operations (convolution, activation function, pooling, and BP algorithm) are carried out in the complex domain \mathbb{C} . SFIM is the statistical feature integration module. P and C denote pooling and convolution operations, respectively. The format below feature maps denotes [channels number @ width \times height] of the corresponding feature maps.

Contourlet overcomes the defects of the wavelet with its flexible directional filter bank. However, due to the existence of the downsampling process, both wavelet and contourlet lack shift invariance and result in pseudo-Gibbs phenomena.

Thanks to the NSCT [65], the distortion can be suppressed to a certain extent with its overcomplete transform. Compared with contourlet, NSCT has the advantages of shift-invariant, easy implementation, and better frequency selectivity.

Most notably, the shift-invariant of NSCT makes the transformed subbands consistent with the size of the original image, and the corresponding pixel position is unchanged. Therefore, NSCT is suitable for pixel-based classification with the geometrical information pixel by pixel from the coefficient maps.

C. Complex-Valued Contourlet Neural Network

As an efficient image representation of the geometrical structure, the NSCT can efficiently describe images with a flexible multiscale, multidirectional, and shift-invariant basis in the spectral domain. Compared with contourlet, NSCT has better frequency selectivity and regularity. Therefore, we integrated the NSCT into CV-CNNs named C^2N^2 [see Fig. 1], which can utilize amplitude and phase information for complex imagery and enhance feature representation in both spatial and spectral domains. Thus, it is beneficial to the approximation ability and further optimization of the network.

1) *Complex-Valued Contourlet Feature Learning*: The essence of NSCT is that the signal is convoluted iteratively with the filter banks of the nonsubsampled pyramid (NSP) and NSDFB. For input X , the corresponding NSCT coefficients

$C(x, y) = \{X_{l,i}, X_{h,i}\}_{0 \leq i \leq I}$, where I is the decomposition level. The $X_{l,i}$ and $X_{h,i}$ are the low-pass component and high-pass component in level i , respectively, and $X_{l,0} = X$. The decomposition of NSCT in level i is defined as

$$\begin{aligned} X_{l,i+1}, X_{h,i+1} &= (X_{l,i} * F_{\text{NSP}}) \\ X_{h_bds,i+1} &= (X_{h,i+1} * F_{\text{NSDFB}}) \end{aligned} \quad (2)$$

where F_{NSP} denotes the NSP and F_{NSDFB} is the NSDFB. The subscript h_bds represents the bandpass directional subbands in the spectral domain, which is exactly ignored by CV-CNNs [see (7)]. Thus, the spectral branch is integrated into CV-CNNs, as shown in Fig. 1, and the spectral features flow generated by the NSCT is highlighted with a red arrow.

Fig. 2(a) shows the NSCT transform with two shift-invariant filters of NSP and NSDFB, the multiscale is ensured by an NSP structure of NSP, and the directionality is given by an NSDFB structure of NSDFB. As shown in Fig. 2(b), the combination of these two filters splits the 2-D frequency plane into the subbands. In terms of space decomposition of features, given an input image in V_0 , the NSP is first applied to generate a low-pass subband in V_1 and high-pass subbands in W_1 , then the NSDFB is employed to decompose the high-pass subbands into 2^i directional subspaces. The relationship between these subspaces satisfies

$$V_j = V_{j+1} \oplus W_{j+1} \quad \text{s.t. } V_j \supset V_{j+1} \quad (3)$$

then the multiscale and multidirection features are obtained with space decomposition.

The details of NSP and NSDFB are as follows.

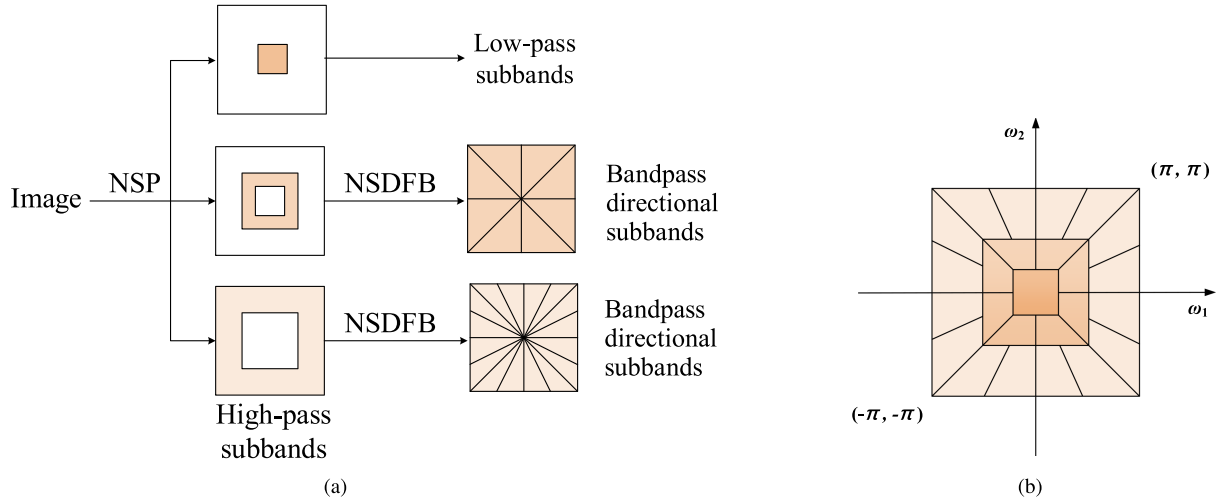


Fig. 2. NSCT transform. (a) NSCT achieved by the NSFB structure. (b) Idealized frequency partitioning obtained with the structure of (a).

NSP: The NSP is an improved version of the Laplace pyramid, where the low-pass subband is not downsampled, but upsampled for the filter accordingly. Therefore, there is no need to design additional filters, and the same size features as the original image can be obtained. Similar to the 1-D nonsubsampled wavelet transform (NSWT), NSP can be achieved by the *à trous* algorithm efficiently [68]. In terms of redundancy for one bandpass image at each decomposition level J , NSWT has redundancy of $J + 3$ and NSCT only has $J + 1$.

The ideal passband support of the low-pass filter is $[-\pi/2^j, \pi/2^j]^2$ at j th level. Similarly, the support region of the high-pass filter is $[-\pi/2^{j-1}, \pi/2^{j-1}]^2 \setminus [-\pi/2^j, \pi/2^j]^2$, which is the complement of the low pass.

NSDFB: An oriented 2-D filter bank [69] is introduced with the l -level binary tree in the 2-D frequency domain, thus the 2^l wedge-shaped subbands are obtained. To increase the shift-invariant of the transformation, NSDFB is introduced by eliminating the downsampled and upsampled operations in DFB. The number of directions is flexible on different scales of NSDFB. As shown in Fig. 2, the number of directions are $2^3 = 8$ and $2^4 = 16$ from coarse to fine scales, respectively.

The relationship between the low-pass decomposition filter $H_0(\mathbf{z})$ and the high-pass decomposition filter $H_1(\mathbf{z})$ satisfies: $H_0(\mathbf{z}) + H_1(\mathbf{z}) = 1$. Essentially, convolution is the basic unit of the filtering process in NSP and NSDFB.

Theorem 1: Given a discrete signal $x[n] \in \mathbb{C}^2$ and the discrete filter $h[n] \in \mathbb{R}^2$ of NSCT, the convolution operation is distributive on the addition term. Specifically, the result is the sum of the convolutions corresponding to the real part and the imaginary part, respectively. Therefore, the processing unit of the filter $h[n]$ in NSCT can be expressed as

$$x[n] * h[n] = \Re(x[n]) * h[n] + j\Im(x[n]) * h[n]. \quad (4)$$

Proof:

$$x[n] * h[n] = \sum_k x[k] * h[n - k]$$

$$\begin{aligned} &= \sum_k (\Re(x[k]) + j\Im(x[k])) * h[n - k] \\ &= \sum_k \Re(x[k]) * h[n - k] \\ &\quad + j \sum_k \Im(x[k]) * h[n - k] \\ &= \Re(x[n]) * h[n] + j\Im(x[n]) * h[n]. \end{aligned} \quad (5)$$

Besides, Cunha et al. [62] proved that NSCT is a frame operator. If the NSP constitutes a frame with frame bounds A_p and B_p , and the NSDFB is A_q and B_q . Therefore, the NSCT with bounds A and B satisfies

$$A_p^J A_q^{\min\{l_j\}} \leq A \leq B \leq B_p^J B_q^{\max\{l_j\}} \quad (6)$$

when $A_p = B_p = A_q = B_q = 1$, the NSCT is a tight frame with bound 1 ($A = B = 1$).

2) *Complex-Valued Spatial Feature Learning*: As shown in Fig. 1, the $\mathbb{C}^2\mathbb{N}^2$ is a two-stream architecture trained in an end-to-end supervised way, one is the spatial feature learning branch, and the other is the spectral domain feature learning branch. It consists of the input layer, the NSCT transform layer, the convolutional and pooling layers, the fully connected (FC) layer, and the classified output layer. Significantly, the operations of all layers are carried out in the complex domain \mathbb{C} . In terms of the complex convolution operation, any one of the output feature maps in layer i can be described as

$$\begin{aligned} x_i &= \rho(W_i x_{i-1} + b_i) \\ &= \rho \sum_k w_{i,k_i}(n) * x_{i-1}(n) + b_i \\ &= \rho \left[\sum_k \Re(w_{i,k_i}(n)) + j\Im(w_{i,k_i}(n)) \right] \\ &\quad * [\Re(x_{i-1}(n)) + j\Im(x_{i-1}(n))] + b_i \end{aligned}$$

$$\begin{aligned}
&= \rho \left[\sum_k \Re(w_{i,k_i}(n))\Re(x_{i-1}(n)) - \Im(w_{i,k_i}(n))\Im(x_{i-1}(n)) \right] \\
&+ j\rho \left[\sum_k \Im(w_{i,k_i}(n))\Re(x_{i-1}(n)) + \Re(w_{i,k_i}(n))\Im(x_{i-1}(n)) \right] + b_i
\end{aligned} \tag{7}$$

where i ($0 < i \leq I$) is the total number of layers in the network, and W_i and b_i are the weights and bias respectively. ρ denotes the activation function with nonlinearity transforms. k_i represents a channel index in layer i . $x_{i-1}(n)$ is the unit of the n th input feature map of layer $i-1$. The symbol $*$ is the convolution operator. $\Re(\cdot)$ and $\Im(\cdot)$ are the real and imaginary parts of complex-valued $x = a + jb$, then $a = \Re(x)$, $b = \Im(x)$.

A complex average pooling operation in the specific layer x_i is calculated by

$$x_i = \text{AVE}(x_{i-1}(m \times s + u, n \times s + v))_{u,v=0,1,\dots,p-1 \in \mathbb{Z}} \tag{8}$$

where (m, n) is the location of the input feature map, p is the pooling size, and s denotes the pooling stride. AVE is defined as the average pooling.

The filtering operation of the NSCT transform layer is given in (4). The details of the FC layer and classified output layer are described as follows.

The FC layer is considered a special convolution, including two situations. 1) *The connection between the feature map and the FC layer*: Given the feature maps of size $w*h*c$ (width*height*channel number), and connected to n neurons in the FC layer. This process is equivalent to the n kernels of size $w*h*c$ convolved with the feature maps of size $w*h*c$. 2) *The connection between the FC layer and the FC layer*: Suppose the number of neurons in the previous layer and the current layer is m and n , respectively. This process can be seen as the m kernels of size $1*1*n$ convolved with the feature maps of size $1*1*n$. Therefore, the FC operation can be expressed as

$$\begin{aligned}
x_i &= \rho(W_i x_{i-1} + b_i) \\
&= \rho \sum_k \Re(w_{i,k_i}(n) \cdot x_{i-1}(n) + b_i) \\
&\quad + j \Im(w_{i,k_i}(n) \cdot x_{i-1}(n) + b_i)
\end{aligned} \tag{9}$$

where i ($0 < i \leq I$) is the network depth, and k_i represents a channel index in layer i .

Different from the Softmax adopted in the output layer of the real-valued CNNs, the least-squares loss function is applied in CV-CNNs, since the output is not a probability when Softmax is used for the complex-valued input. The loss can be written as

$$\begin{aligned}
L &= \frac{1}{2} \frac{1}{M} \sum_{m=1}^M \sum_{c=1}^C [\hat{y}_c(m) - y_c(m)]^2 \\
&= \frac{1}{2} \frac{1}{M} \sum_{m=1}^M \sum_{c=1}^C [\Re(\hat{y}_c(m)) - \Re(y_c(m))]^2 \\
&\quad + [\Im(\hat{y}_c(m)) - \Im(y_c(m))]^2
\end{aligned} \tag{10}$$

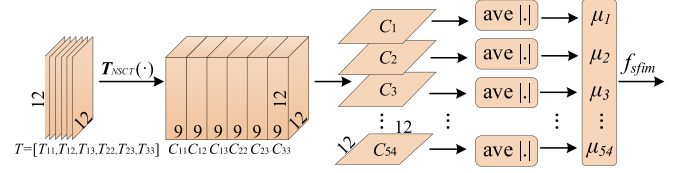


Fig. 3. Statistical feature integration module.

where M and C are the numbers of samples and categories, respectively. $\hat{y}_c(m)$ and $y_c(m)$ are the labels and the prediction in the complex domain, respectively. The label is a one-hot vector, where the value of the correct position is $1+1*j$, and the others are 0.

The proposed model is trained by the SGD to minimize the loss function, and then, the weights and biases are iteratively updated in the complex backpropagation stage. Similar to the backpropagation stage in the real-valued domain, the error propagation is calculated according to the complex chain rule in the complex-valued domain. For more details, refer to [49].

3) *Statistical Feature Integration Module*: As shown in Fig. 3, given a patch of the PolSAR image $T(x, y) = [T_{11}, T_{12}, T_{13}, T_{22}, T_{23}, T_{33}]$ in the complex domain, the corresponding contourlet coefficients $C(x, y) = [C_{11}, C_{12}, C_{13}, C_{22}, C_{23}, C_{33}]$ are obtained after the multiresolution sparse representation $T_{\text{NSCT}}(\cdot)$ of NSCT, and each element in $C(x, y)$ consists of one low-pass subband and eight high-pass subbands. Therefore, there are $6*(1+8) = 54$ subband coefficients. Calculate the mean μ_n of each coefficient and form the statistical vector f_{sfim} , which can be expressed as

$$\begin{aligned}
f_{\text{sfim}} &= (\mu_1, \mu_2, \dots, \mu_{54}) \\
\mu_n &= \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H |C_n(x, y)|
\end{aligned} \tag{11}$$

where W and H are the width and height of the C_n , respectively.

IV. VALIDATION AND GENERALIZATION

In this section, our model is evaluated on three PolSAR datasets, including images of Flevoland, Germany, and Xi'an. Section III-A and III-B provide the datasets description and implementation details of the proposed method, respectively. Section III-C introduces the evaluation metrics of overall accuracy (OA) and the kappa coefficient. Section III-D presents the experimental results on different datasets. Finally, the detailed ablation studies are conducted in Section III-E.

A. Datasets Description

Three benchmark PolSAR datasets are used in our experiments, and 1% of labeled samples per class are randomly chosen for training. Details of these datasets are described as follows.

The Flevoland : The PolSAR image of Flevoland is captured by NASA/Jet Propulsion Laboratory AIRSAR platform in 1989 [70]. The azimuth resolution is 12.1 m and the slant angle resolution is 6.6 m. Flevoland is an agricultural area in the

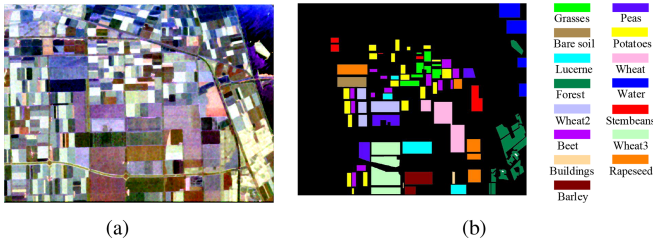


Fig. 4. PauliRGB image and ground truth of Flevoland. (a) PauliRGB image. (b) Ground truth.

TABLE I
NUMBER AND RATIO OF SAMPLES OF FLEVOLAND

Class	Name	Number of pixels	Ratio
1	Stem beans	6338	3.78%
2	Peas	9582	5.71%
3	Forest	18044	10.76%
4	Lucerne	10181	6.07%
5	Wheat	16386	9.77%
6	Beet	10033	5.98%
7	Potatoes	16156	9.63%
8	Bare soil	5109	3.05%
9	Grasses	7058	4.21%
10	Rapeseed	13863	8.27%
11	Barley	7595	4.53%
12	Wheat2	11159	6.65%
13	Wheat3	22241	13.26%
14	Water	13232	7.89%
15	Buildings	735	0.44%

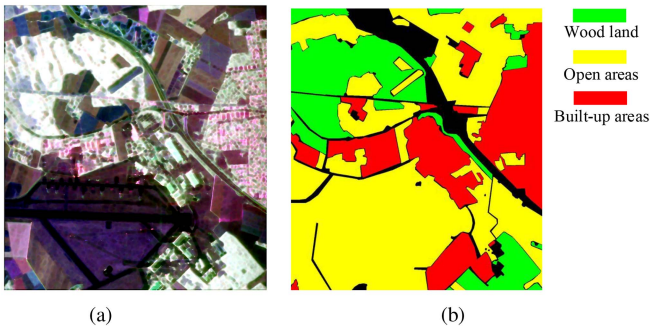


Fig. 5. PauliRGB image and ground truth of Germany. (a) PauliRGB image. (b) Ground truth.

Netherlands, including stem beans, rapeseed, bare soil, forest, buildings, and so on. Fig. 4 shows the Pauli RGB image and the corresponding ground truth of Flevoland with a size of 1024×750 , and the details number and ratio of samples are given in Table I.

The Germany : The L-band PolSAR image of Germany is obtained by DLR's experimental SAR in Oberpfaffenhofen. The size of the image is 1300×1200 pixels with a resolution of 3×2.2 . There are three land cover classes including wood land, built-up areas, and open areas. Fig. 5 shows the Pauli RGB image and the corresponding ground truth of Germany with a size of

TABLE II
NUMBER AND RATIO OF SAMPLES OF GERMANY

Class	Name	Number of pixels	Ratio
1	Built-up areas	339784	24.72%
2	Wood land	268010	19.50%
3	Open areas	766504	55.77%

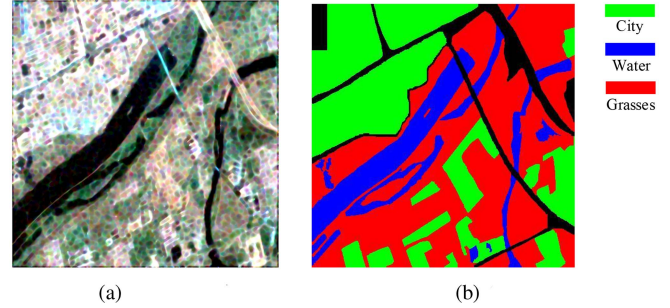


Fig. 6. Pauli RGB image and ground truth of Xi'an. (a) Pauli RGB image. (b) Ground truth.

TABLE III
NUMBER AND RATIO OF SAMPLES OF XI'AN

Class	Name	Number of pixels	Ratio
1	Grass	119584	50.37%
2	City	81071	34.15%
3	Water	36761	15.48%

1300×1200 , and the details number and ratio of samples are given in Table II.

The Xi'an : The C-band PolSAR image of Xi'an is produced by a RADARSAT-2 sensor, which covers western Xi'an, Shaanxi, China. The size of the image is 512×512 pixels with a resolution of 8×8 . There are three land cover classes including grass, city, and water. Fig. 6 shows the Pauli RGB image and the corresponding ground truth of Xi'an. The details number and ratio of samples are shown in Table III.

B. Implementation Strategy

The proposed C^2N^2 is implemented in MATLAB. It consists of two branches, as shown in Fig. 1. Both the spatial and spectral branches are composed of one convolution layer and one pooling layer. Then, two branches are combined into one branch where the feature maps are cascaded accordingly, followed by one convolution layer and one FC layer. Complex-valued patches of $12 \times 12 \times 6$ are used as input, and the output format is one-hot encoding. Kernels of 3×3 and 2×2 are used in convolution and pooling layers, respectively. More details of the network structure are listed in Table IV.

The training procedure is described in Algorithm 1, where C and M are the numbers of categories and samples, respectively. f_d^{spa} and f_d^{spe} , f_{c1}^{spa} and f_{c1}^{spe} , and f_p^{spa} and f_p^{spe} are data layers, convolution layers, and pooling layers of the spatial branch and spectral branch in the complex domain, respectively. f_{stack} , f_{c2} , and f_c are the stacking layer, convolution layer, and

TABLE IV
LAYERS OF C^2N^2

Layer		C@W×H	Kernel Size/ Stride/ Pad		
data		6@12×12	- / - / -		
data_spa	data_spe	6@12×12	54@12×12	- / - / -	- / - / -
conv1_spa	conv1_spe	6@10×10	9@10×10	3×3×6/1/0	3×3×9/1/0
pooling1_spa	pooling1_spe	6@5×5	9@5×5	2×2/1/0	2×2/1/0
stacking		15@5×5	- / - / -		
conv2		12@3×3	3×3×6×12		
FC		108+54@1×1	- / - / -		

Algorithm 1: Training Procedure.

-
- Input:** Training dataset: $\mathbf{X} = \{x_m | m = 1, 2, \dots, M\}$, the patch $x_m \in \mathbb{C}^{6 \times 12 \times 12}$, and their corresponding labels $\mathbf{G}_i = \{g_m^i | m = 1, 2, \dots, M\}$, and $i = 1, \dots, C$.
- Output:** The learned C^2N^2 .
- 1: **Preparation:** The model can be expressed as two branches: $[f_d^{spa}, f_{c1}^{spa}, f_p^{spa}]$ and $[f_d^{spe}, f_{c1}^{spe}, f_p^{spe}]$, then followed by $[f_{stack}, f_{c2}, f_c]$. $F_{NSP} = \text{'maxflat'}$, $F_{NSDFB} = \text{'dmaxflat7'}$.
 - 2: **Begin**
 - 3: **for** $n = 1$ to M **do**
 - 4: Using Eq. (2) to calculate spectral coefficients f_d^{spe} of the input x_m .
 - 5: Resize the f_d^{spe} to the size of N^2 , where $N = \max(\text{height}(f_d^{spe}), \text{width}(f_d^{spe}))$.
 - 6: Using Eq. (11) to calculate statistical features f_{sfim} of NSCT coefficients in statistical feature integration module, and then cascade f_{sfim} to the fully connected layer f_c .
 - 7: Stacking outputs of two branches as $f_{stack} = f_p^{spa} \oplus f_p^{spe}$ and fed f_{stack} to f_{c2} and f_c , sequentially;
 - 8: Minimize Eq. (10) and trained by the SGD until convergence.
 - 9: **end for**
 - 10: **End**
-

full connection layer in the complex domain, respectively. The decomposition level of the contourlet is 1, and the number of directional subbands is $2^3 = 8$.

C. Evaluation Metrics

In order to evaluate our model in a general view, accuracy is applied as the basic metric for any classification task.

Accuracy is the proportion of correctly classified samples in the total samples in the test dataset. The value of accuracy ranges from 0 to 1. The higher the value, the better the classification, and vice versa. It can be formulated as

$$\text{Accuracy} = \frac{\sum_{k=1}^K \text{TP}(k) + \text{TN}(k)}{\sum_{k=1}^K \text{TP}(k) + \text{TN}(k) + \text{FP}(k) + \text{FN}(k)} \quad (12)$$

where $\text{TP}(k)$ and $\text{TN}(k)$ are the numbers of true positive and true negative, respectively, and the $\text{FN}(k)$ and $\text{FP}(k)$ is the

number of false negative and false positive associated with label k , respectively.

The Kappa coefficient is a multivariate statistical metric, which describes the unbalanced confusion matrix

$$\text{Kappa} = \frac{N \sum_{k=1}^K x_{kk} - \sum_{k=1}^K (x_{k:} x_{:k})}{N^2 - \sum_{k=1}^K (x_{k:} x_{:k})} \quad (13)$$

where N denotes the total number of test samples, and $x_{k:}$ and $x_{:k}$ are the sum of row k and column k of the confusion matrix, respectively.

D. Results Comparison on Datasets

1) *Result on the Flevoland Dataset:* Table V and Fig. 7 provide the quantitative and qualitative results on the Flevoland dataset. As can be seen obviously in Table V, the proposed C^2N^2 integrated statistical feature (C^2N^2 -sf) outperforms the other seven comparison methods. The improvements of C^2N^2 -sf are 16.57% (SVM), 59.49% (SAE), 55.37% (DBN), 31.27% (SF-CNN), 33.43% (C-CNN), 19.93% (complex contourlet-CNN), and 29.4% (CV-CNN) in terms of OA, respectively, and 0.1883 (SVM), 0.6390 (SAE), 0.6374 (DBN), 0.3357 (SF-CNN), 0.3563 (C-CNN), 0.2095 (complex contourlet-CNN), and 0.3165 (CV-CNN) in terms of the Kappa coefficient, respectively. Specifically, our method not only achieves the highest classification accuracy in large categories, such as forest, wheat, and potatoes, but also performs well in small categories, such as stem beans, bare soil, and barley. It verifies the contribution of phase information of the complexity theory and the edge information of MRA for classification. In addition, C^2N^2 -sf is further improved than the C^2N^2 that not integrated statistical feature (C^2N^2 -nsf) by 7.85% in terms of OA, which demonstrates the effectiveness of the statistical features.

As can be seen from Fig. 7(a)–(g), the methods involved in the comparison cannot distinguish well between wheat, wheat2, and wheat3. SAE misclassified most pixels into rapeseed and beet. DBN mistook a large number of pixels for wheat3. In addition, SVM, SF-CNN, C-CNN, complex contourlet-CNN, and CV-CNN are not well classified in most land cover types. This is may be due to insufficient training samples (1%), which affect the classification performance of the compared method. However, our model achieves good classification results and local consistency. As the regions marked with numbers 1–5 in white ellipses or rectangles in Fig. 7, the number of isolated pixels is reduced in our methods with better visual continuity. There are two reasons for this encouraging result. First, due to the C^2N^2 with the multiscale and multidirection representations, the abundant details of the land cover can be captured and represented. Second, the separability between classes is increased with the help of the integration of the statistical features, which is beneficial to the classification task.

2) *Result on the Xi'an Dataset:* Table VI and Fig. 8 reported the quantitative and qualitative results of the Xi'an dataset. It can be observed [see Table VI] that the result

TABLE V
CLASSIFICATION RESULTS (ACCURACY %) FOR NETWORKS ON THE FLEVLAND DATASET

Class	SVM[27]	SAE[71]	DBN[33]	SF-CNN[28]	C-CNN[56]	Complex Contourlet-CNN[72]	CV-CNN[49]	C ² N ² -nsf	C ² N ² -sf
Stem beans	1.99±0.05	26.35±6.54	0.00±0.00	69.77±6.62	22.35±4.49	76.65±5.49	37.95±8.53	78.78±2.98	82.58 ±1.47
Peas	23.54±0.35	2.04±0.17	25.96±2.92	21.75±5.54	80.18±4.80	87.70±2.23	25.43±9.89	85.33±8.13	90.42 ±8.09
Forest	87.12±0.06	51.02±13.07	0.53±0.02	2.92±2.68	49.27±0.82	45.07±2.73	83.45±0.38	94.08 ±0.44	92.84±0.25
Lucerne	61.34±0.30	0.01±0.01	0.01±0.01	74.85±4.87	62.29±0.26	88.12 ±3.24	48.94±4.51	0.01±0.01	12.57±0.03
Wheat	48.29±0.13	17.42±2.14	2.80±0.20	49.27±2.59	84.12±2.29	67.22±5.59	5.82±1.08	96.30 ±1.12	93.07±0.95
Beet	27.96±0.14	20.97±1.39	1.41±0.16	18.31±2.49	35.04±1.83	44.18±3.40	39.71±3.14	51.63 ±2.52	47.83±1.22
Potatoes	70.18±0.05	6.56±0.48	0.07±0.07	13.63±1.37	3.51±2.87	58.35±1.65	55.42±1.72	61.07±1.37	90.59 ±1.21
Bare soil	79.86±0.14	0.01±0.01	0.01±0.01	98.18 ±1.09	79.14±0.01	92.19±7.32	24.42±7.62	90.64±4.03	75.83±0.38
Grasses	0.67±0.02	0.01±0.01	0.01±0.01	33.91±2.13	2.19±0.52	8.49±3.63	49.25 ±4.27	6.91±3.80	15.14±2.17
Rapeseed	33.24 ±0.09	18.99±0.82	8.71±2.32	11.62±4.07	28.20±0.01	23.42±4.81	24.79±10.31	0.01±0.01	5.95±3.91
Barley	0.01±0.01	0.01±0.01	0.01±0.01	71.76±0.41	0.13±0.08	50.47±3.50	77.53±0.66	98.48 ±0.57	98.44±0.42
Wheat2	69.13±0.28	1.84±0.16	5.06±0.62	44.01±5.69	75.06±2.25	89.49 ±2.07	9.00±8.01	75.38±8.23	84.46±6.22
Wheat3	82.11±0.08	0.42±0.06	99.70 ±0.23	40.82±6.41	0.23±0.21	25.33±0.14	50.96±7.78	86.42±6.39	97.85±1.92
Water	81.27±0.29	1.03±0.10	0.01±0.01	89.54 ±1.75	71.21±4.71	45.41±3.38	39.19±5.92	41.35±5.53	63.71±5.05
Buildings	57.41 ±0.21	20.53±6.91	2.33±0.54	8.82±1.34	1.34±1.26	3.78±2.95	3.56±0.77	0.63±0.27	7.56±0.58
OA	55.03±0.04	12.11±0.30	16.23±0.12	40.33±0.39	38.17±1.49	51.67±2.05	42.20±0.27	63.75±0.16	71.60 ±0.23
Kappa	50.45±0.09	5.38±0.30	5.54±0.06	35.71±0.41	33.65±0.85	48.33±0.93	37.63±0.33	60.70±0.27	69.28 ±0.25

The bold values denote the best result.

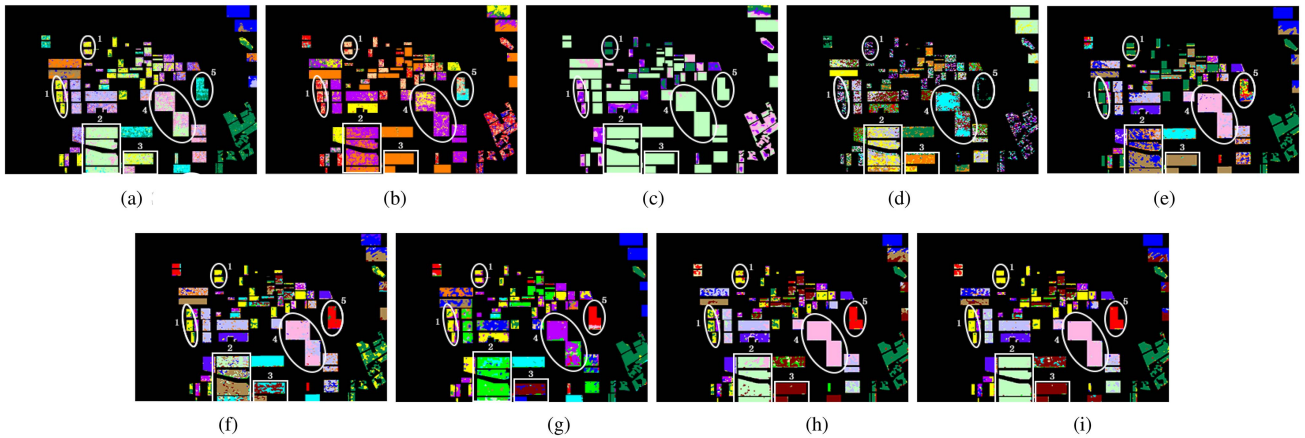


Fig. 7. Classification results of Flevoland with different methods. (a) SVM. (b) SAE. (c) DBN. (d) SF-CNN. (e) C-CNN. (f) Complex contourlet-CNN. (g) CV-CNN. (h) C²N²-nsf. (i) C²N²-sf.

TABLE VI
CLASSIFICATION RESULTS (ACCURACY %) FOR NETWORKS ON THE XI'AN DATASET

Class	SVM[27]	SAE[71]	DBN[33]	SF-CNN[28]	C-CNN[56]	Complex Contourlet-CNN[72]	CV-CNN[49]	C ² N ² -nsf	C ² N ² -sf
Grass	91.36 ±0.11	78.39±0.32	13.03±0.45	68.97±0.32	79.87±0.14	73.28±0.48	82.58±0.13	82.64±0.11	78.29±0.10
City	52.73±0.19	67.47±0.18	2.54±0.33	83.47±0.15	83.74±0.36	85.62±0.34	86.65±0.10	90.06±0.12	90.77 ±0.08
Water	17.78±0.12	30.23±1.18	99.00 ±0.14	81.02±0.60	50.54±0.42	84.49±0.57	65.95±0.37	88.78±0.24	90.03±0.27
OA	66.77±0.13	67.20±0.30	22.76±0.15	75.79±0.17	71.68±0.37	81.36±0.11	80.61±0.14	88.44±0.08	88.53 ±0.13
Kappa	39.43±0.36	42.19±0.13	0.87±0.44	61.73±0.25	57.76±0.23	73.25±0.29	68.48±0.16	82.66±0.10	82.80 ±0.08

The bold values denote the best result.

of our model is superior to other methods on the evaluation criteria of OA and Kappa coefficient. The improvements of our C²N²-sf are 21.76% (SVM), 21.33% (SAE), 65.77% (DBN), 12.74% (SF-CNN), 16.85% (C-CNN), 7.17% (complex contourlet-CNN), and 7.92% (CV-CNN) in terms of OA, respectively, and 0.4337 (SVM), 0.4061 (SAE), 0.8193

(DBN), 0.2107 (SF-CNN), 0.2504 (C-CNN), 0.0955 (complex contourlet-CNN), and 0.1432 (CV-CNN) in terms of the Kappa coefficient, respectively. Although the sampling rate of our training set is only 1%, the effectiveness of our model has been verified by the remarkable performance on the Xi'an dataset.

TABLE VII
CLASSIFICATION RESULTS (ACCURACY %) FOR NETWORKS ON THE GERMANY DATASET

Class	SVM[27]	SAE[71]	DBN[33]	SF-CNN[28]	C-CNN[56]	Complex Contourlet-CNN[72]	CV-CNN[49]	C^2N^2 -nsf	C^2N^2 -sf
Built-up areas	10.70 \pm 0.24	18.09 \pm 0.17	64.86 \pm 0.13	46.78 \pm 0.29	62.31 \pm 0.35	64.08 \pm 0.18	61.70 \pm 0.07	76.33 \pm 0.12	68.37 \pm 0.11
Wood land	89.32 \pm 1.19	62.72 \pm 0.18	36.31 \pm 0.27	93.43 \pm 0.43	73.85 \pm 1.05	79.93 \pm 0.41	77.25 \pm 0.11	67.92 \pm 0.14	82.58 \pm 0.13
Open areas	24.53 \pm 1.07	22.96 \pm 0.43	3.63 \pm 0.16	91.37 \pm 0.05	94.87 \pm 0.24	95.36 \pm 0.27	95.23 \pm 0.04	93.28 \pm 0.07	93.72 \pm 0.09
OA	33.25 \pm 0.14	29.22 \pm 0.15	25.09 \pm 0.13	80.61 \pm 0.36	77.98 \pm 0.43	84.49 \pm 0.23	83.56 \pm 0.10	84.31 \pm 0.09	85.36 \pm 0.06
Kappa	0.59 \pm 0.20	-10.53 \pm 0.04	-0.18 \pm 0.12	67.56 \pm 0.25	65.81 \pm 0.28	69.85 \pm 0.16	75.33 \pm 0.50	76.47 \pm 0.17	78.03 \pm 0.10

The bold values denote the best result.

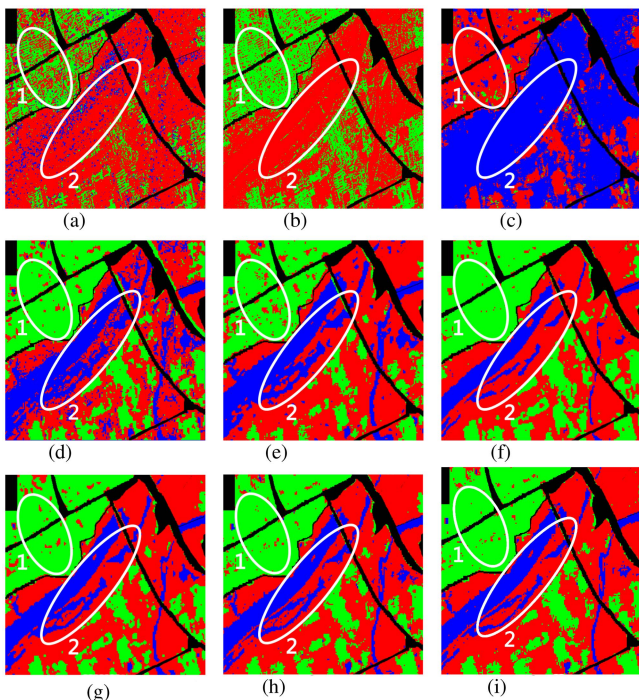


Fig. 8. Classification results of Xi'an with different methods. (a) SVM. (b) SAE. (c) DBN. (d) SF-CNN. (e) CV-CNN. (f) C-CNN. (g) Complex contourlet-CNN. (h) C^2N^2 -nsf. (i) C^2N^2 -sf.

Fig. 8(a)–(i) shows the classification results on different type of land cover. The performance of CV-CNN and complex contourlet-CNN is acceptable, followed by SF-CNN, C-CNN, SVM, and SAE, and DBN is the worst. Specifically, SVM and SAE have a low recognition rate of water and wrongly classified water as grass, which is shown as the white ellipses with the number 2 in Fig. 8(a) and (b). DBN is not doing well in the categories of the city and grass, which wrongly classified grass as water and city as grass. Compared with other methods, our model has a better classification result with clearer boundaries and regional consistency, where the outstanding areas are marked with white ellipses.

3) *Result on the Germany Dataset:* Table VII and Fig. 9 reported the quantitative and qualitative results of the Germany dataset. As is given in Table VII, the proposed C^2N^2 -sf performs best among the compared models. The OA of the C^2N^2 -sf is 52.11%, 56.14%, 60.27%, 4.75%, 7.38%, 0.87%, and 1.8% higher than SVM, SAE, DBN, SF-CNN, C-CNN, complex contourlet-CNN, and CV-CNN, respectively. In terms of the Kappa coefficient, C^2N^2 -sf is 0.7744 (SVM), 0.8856 (SAE), 0.7821(DBN), 0.1047(SF-CNN), 0.1222 (C-CNN),

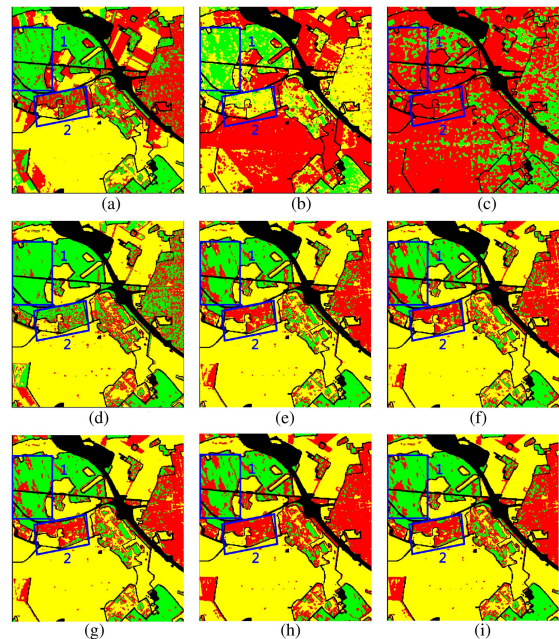


Fig. 9. Classification results of Germany with different methods. (a) SVM. (b) SAE. (c) DBN. (d) SF-CNN. (e) CV-CNN. (f) C-CNN. (g) Complex contourlet-CNN. (h) C^2N^2 -nsf. (i) C^2N^2 -sf.

0.0818 (complex contourlet-CNN), and 0.027 (CV-CNN) higher than the compared methods. For the accuracy of each category, C^2N^2 -sf has certain advantages in the classification results of built-up areas, which has increased by 6.67% on the basis of CV-CNN. This remarkable improvement comes from the fact that the high-frequency components of the contourlet have the potential to describe edge and texture information, and the built-up area has a snowflake-like texture with abundant details [see Fig. 5(a)], thus the representation and classification capabilities of our model are improved.

Fig. 9 presents the classification result of different methods for the Germany dataset. By comparing Fig. 9(a)–(c), the number of isolated pixels is reduced in Fig. 9(d)–(i). Our model performs better than the baseline CV-CNN in the wood land and built-up areas, which are indicated by the blue rectangles in Fig. 9. Although the classification result on wood land is slightly inferior to that of SVM. Overall, the experimental results prove the effectiveness of our method.

E. Ablation Study

To gain a deeper insight into the proposed method, the ablation experiment is given in Table VIII, where the “CV-CNN” denotes

TABLE VIII
ABLATION EXPERIMENTAL RESULTS (ACCURACY %)

Methods	Flevoland		Xi'an		Germany	
	OA	Kappa	OA	Kappa	OA	Kappa
CV-CNN	42.20 \pm 0.27	37.63 \pm 0.33	80.61 \pm 0.14	68.48 \pm 0.16	83.56 \pm 0.10	75.33 \pm 0.50
C ² N ² -nsf	63.75 \pm 0.16	60.70 \pm 0.27	88.44 \pm 0.08	82.66 \pm 0.10	84.31 \pm 0.09	76.47 \pm 0.17
C ² N ² -sf	71.60 \pm 0.23	69.28 \pm 0.25	88.53 \pm 0.13	82.80 \pm 0.08	85.36 \pm 0.06	78.03 \pm 0.10

The bold values denote the best result.

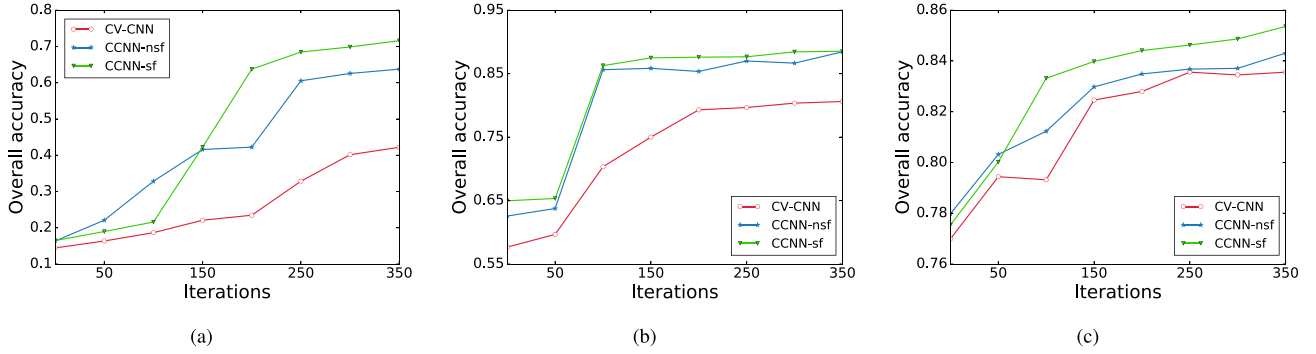


Fig. 10. OA curves of our methods on the Flevoland dataset, the Germany dataset, and the Xi'an dataset, respectively.

TABLE IX
CLASSIFICATION RESULTS (ACCURACY %) OF C²N²-SF WITH DIFFERENT NUMBERS OF NSDFB

Numbers of NSDFB	Flevoland	Xi'an	Germany
2 ² = 4	58.34 \pm 0.19	84.86 \pm 0.11	83.32 \pm 0.20
2 ³ = 8	71.60 \pm 0.23	88.53 \pm 0.13	85.36 \pm 0.06
2 ⁴ = 16	68.23 \pm 0.22	86.05 \pm 0.25	84.45 \pm 0.17

The bold values denote the best result.

the basic complex-valued convolutional network without the branch of the spectral domain. "C²N²-sf" and "C²N²-nsf" are the C²N² with the SFIM and without SFIM, respectively.

The improvement in Table VIII and Fig. 10 shows the effectiveness of the spectral branch of contourlet and SFIM on the Flevoland dataset, the Germany dataset, and the Xi'an dataset. Besides, columns 7–9 of Tables V–VII and subfigures (e)–(i) of Figs. 7 to 9 present the effectiveness of each module of the proposed method quantitatively and qualitatively.

As shown in Fig. 10, the result of "CV-CNN" is the weakest since only the spatial features are employed in CNNs. After adding the spectral branch and SFIM to the model, the proposed method converges to higher precisions than the basic CV-CNN after several iterations. The reason can be attributed to the integration of the contourlet and its statistical characteristics, which is consistent with the HVS [73] with biological interpretation. In addition, the geometrical and textural features of the PolSAR image are more suitable for MRA. Therefore, the approximation performance is enhanced by the contourlet with the efficient tight frame, and good classification results are further achieved.

Furthermore, our model parameters are analyzed from two aspects. On the one hand, the study on different direction numbers of NSDFB is provided in Table IX. We can see that

TABLE X
CLASSIFICATION RESULTS (ACCURACY %) OF DIFFERENT RELATIVE WEIGHTS ON THE XI'AN DATASET

Class	[0, 1]	[0.25, 0.75]	[0.5, 0.5]	[0.75, 0.25]	[1, 0]
Grass	76.97 \pm 0.13	75.23 \pm 0.17	77.72 \pm 0.21	79.56 \pm 0.19	65.60 \pm 0.12
City	88.84 \pm 0.25	93.27 \pm 0.11	93.98 \pm 0.16	91.15 \pm 0.10	89.58 \pm 0.38
Water	82.22 \pm 0.45	85.82 \pm 0.39	88.66 \pm 0.28	82.35 \pm 0.05	88.65 \pm 0.41
OA	82.68 \pm 0.19	85.11 \pm 0.18	86.45 \pm 0.14	84.42 \pm 0.13	81.94 \pm 0.11
Kappa	74.01 \pm 0.32	76.76 \pm 0.26	77.68 \pm 0.10	76.23 \pm 0.43	72.91 \pm 0.29

The bold values denote the best result.

direction number 8 of NSDFB shows the largest improvement among other direction numbers and achieves the best performance. Although direction number 16 has the most spectral feature representation, we find that it achieves lower accuracy compared with direction number 8. The reason may be that as the feature dimension increases, the difficulty of classification also increases. On the other hand, in order to explore the sensitivity of relative weights on two feature learning branches, we roughly divided them into five groups: [0, 1], [0.25, 0.75], [0.5, 0.5], [0.75, 0.25], and [1, 0]. The results of our model under different relative weights are provided in Table X. It can be seen that the result of [1, 0] is the weakest since only the spatial features are employed. After gradually increasing spectral information, the result of [0.75, 0.25] and [0.25, 0.75] achieved higher classification performance. The reason can be attributed to the integration of the NSCT with the high-frequency feature representation. However, the result of [0, 1] is slightly worse than that of [0.75, 0.25] and [0.25, 0.75], which indicates that the classification results of feature representations in both domains are better than those of a single domain. The best classification performance is achieved when the weight is [0.5, 0.5]. Note that the weight used in other experiments is [1, 1] in this article.

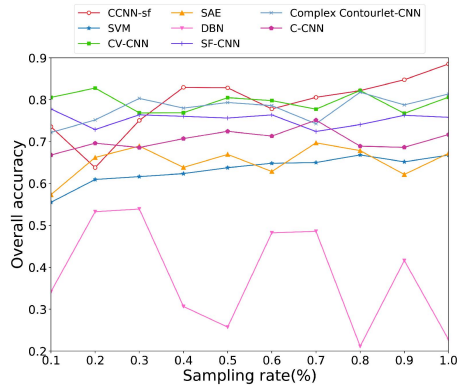


Fig. 11. Variation of OA with sampling rate for different methods on the Xi'an dataset.

TABLE XI
COMPARISON OF MODEL PARAMETERS AND FLOPS

Methods	Parameters	FLOPs
SAE	4.848 K	9.384 K
DBN	32.453 K	64.608 K
SF-CNN	102.752 K	0.205 M
C-CNN	905.147 K	5.398 M
Complex Contourlet-CNN	192.022 K	47.694 M
CV-CNN	35.898 K	0.968 M
C^2N^2-nsf	54.700 K	1.047 M
C^2N^2-sf	60.076 K	1.057 M

The bold values denote the best result.

In order to study the performance of compared methods on different numbers of training samples, the experiments were conducted with sampling rates ranging from 0.1% to 1% and intervals of 0.1%. Fig. 11 shows the variation of OA with sampling rate for different methods. It can be seen that our method achieves the best results at low sampling rates (0.1%–1%), and the OAs of our method are robust to changes in the sampling rate. The results show that our method can obtain ideal and stable classification results at a low sampling rate.

In Table XI, we report two different indices of the model efficiency, i.e., the model parameters and the floating point operations (FLOPs), and the Germany dataset is used to compute FLOPs. The structure of C-CNN and complex contourlet-CNN is slightly adjusted without affecting the core idea, so as to ensure the input size is consistent with the proposed method. Compared with CV-CNN, the parameters of the proposed model are slightly larger due to the existence of the spectral feature learning branch. In terms of the FLOPs, C^2N^2 and CV-CNN are comparable since the filters of the contourlet are “fixed,” and no need for additional learning. Thus, the spectral feature learning branch integrated into the CV-CNN is reasonable with the tradeoff on accuracy.

V. CONCLUSION

In this article, the C^2N^2 was proposed with the multiresolution sparse representation of the NSCT and the statistical feature integration module. The multiscale and multidirection contourlet

features provide effective representation of directionality, singularity, and regularity in the complex spectral domain, and the corresponding weights and biases can be adaptively learned in an end-to-end fashion. In addition, the statistical feature integration module is designed to further capture the statistical properties of the NSCT coefficients. Therefore, the amplitude and phase information of the complex imagery can be fully utilized, and the feature representation is enhanced in both spatial and spectral domains.

The extensive experiments on three PolSAR datasets showed the effectiveness of our model qualitatively and quantitatively, and verified the doughty capacity of approximate and robust with the combination of time–frequency localization and self-studying by the neural network. Furthermore, the interpretability of our model was increased with the NSCT kernel. Future work will build a complete multiscale contourlet network to explore the more effective feature representation and increase the perception capability of the network.

REFERENCES

- [1] C. Sun, X. Zhang, H. Meng, X. Cao, J. Zhang, and L. Jiao, “Dual-branch spectral-spatial adversarial representation learning for hyperspectral image classification with few labeled samples,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1–15, Jun. 2023.
- [2] G. Cheng, X. Xie, J. Han, L. Guo, and G.-S. Xia, “Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, Jun. 2020.
- [3] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, “Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6308–6325, Sep. 2020.
- [4] Z. Xue, Q. Xu, and M. Zhang, “Local transformer with spatial partition restore for hyperspectral image classification,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4307–4325, May 2022.
- [5] H. Dong, L. Zhang, and B. Zou, “Exploring vision transformers for polarimetric SAR image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–5, Dec. 2021.
- [6] L. Sun, S. Cheng, Y. Zheng, Z. Wu, and J. Zhang, “SPANet: Successive pooling attention network for semantic segmentation of remote sensing images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4045–4057, 2022.
- [7] Z. Zhong, Y. Li, L. Ma, J. Li, and W.-S. Zheng, “Spectral–spatial transformer network for hyperspectral image classification: A factorized architecture search framework,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–5, Oct. 2021.
- [8] R. Hang, Z. Li, Q. Liu, P. Ghamisi, and S. S. Bhattacharyya, “Hyperspectral image classification with attention-aided CNNs,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2281–2293, Mar. 2021.
- [9] S. K. Roy, S. Manna, T. Song, and L. Bruzzone, “Attention-based adaptive spectral–spatial kernel ResNet for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7831–7843, Sep. 2021.
- [10] W. Li, B. Zou, and L. Zhang, “A robust man-made target detection method based on relative spectral stationarity for high-resolution SAR images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–20, Nov. 2022.
- [11] H. Sun, Y. Chen, X. Lu, and S. Xiong, “Decoupled feature pyramid learning for multi-scale object detection in low-altitude remote sensing images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, Jul. 2023.
- [12] J. Wang, F. Li, and H. Bi, “Gaussian focal loss: Learning distribution polarized angle prediction for rotated object detection in aerial images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–3, May, 2022.
- [13] J. Zhao, G. Wang, B. Zhou, J. Ying, and J. Liu, “SRA-CEM: An improved CEM target detection algorithm for hyperspectral images based on sub-region analysis,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, Jun. 2023.

- [14] J.-S. Lee and E. Pottier, *Polarimetric Radar Imaging: from Basics to Applications*. Boca Raton, FL, USA: CRC Press, 2017.
- [15] J.-S. Lee, M. R. Grunes, E. Pottier, and L. Ferro-Famil, "Unsupervised terrain classification preserving polarimetric scattering characteristics," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 4, pp. 722–731, Apr. 2004.
- [16] G. Liu, B. Liu, G. Zheng, and X. Li, "Environment monitoring of Shanghai Nanhui intertidal zone with dual-polarimetric SAR data based on deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–8, Aug. 2022.
- [17] 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.
- [18] W. L. Cameron and L. K. Leung, "Feature motivated polarization scattering matrix decomposition," in *Proc. IEEE Int. Conf. Radar*, 1990, pp. 549–557.
- [19] A. Freeman and S. L. Durden, "Three-component scattering model to describe polarimetric SAR data," *Proc. SPIE*, vol. 1748, pp. 213–224, 1993.
- [20] G. Singh, Y. Yamaguchi, and S.-E. Park, "General four-component scattering power decomposition with unitary transformation of coherency matrix," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 3014–3022, May 2013.
- [21] W. Song, M. Li, P. Zhang, and Y. Wu, "Fuzziness modeling of polarized scattering mechanisms and PolSAR image classification using fuzzy triplet discriminative random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4980–4993, Jul. 2019.
- [22] X. Liu, L. Jiao, X. Tang, Q. Sun, and D. Zhang, "Polarimetric convolutional network for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 3040–3054, May 2019.
- [23] 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.
- [24] L.-K. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 780–795, Mar. 1999.
- [25] 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, pp. 1–9, 2009.
- [26] L. Du and J. Lee, "Fuzzy classification of earth terrain covers using complex polarimetric SAR data," *Int. J. Remote Sens.*, vol. 17, no. 4, pp. 809–826, 1996.
- [27] C. Lardeux et al., "Support vector machine for multifrequency SAR polarimetric data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 12, pp. 4143–4152, Dec. 2009.
- [28] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 152–177, 1963.
- [29] J.-S. Lee, M. R. Grunes, and R. Kwok, "Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution," *Int. J. Remote Sens.*, vol. 15, no. 11, pp. 2299–2311, 1994.
- [30] J.-S. Lee, M. R. Grunes, T. L. Ainsworth, L.-J. Du, D. L. Schuler, and S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex Wishart classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2249–2258, Sep. 1999.
- [31] 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.
- [32] A. C. Frery, A. H. Correia, and C. D. C. Freitas, "Classifying multi-frequency fully polarimetric imagery with multiple sources of statistical evidence and contextual information," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3098–3109, Oct. 2007.
- [33] B. Hou, X. Luo, S. Wang, L. Jiao, and X. Zhang, "Polarimetric SAR images classification using deep belief networks with learning features," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 2366–2369.
- [34] B. Hou, H. Kou, and L. Jiao, "Classification of polarimetric SAR images using multilayer autoencoders and superpixels," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 3072–3081, Jul. 2016.
- [35] F. Liu, L. Jiao, and X. Tang, "Task-oriented GAN for PolSAR image classification and clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2707–2719, Sep. 2019.
- [36] F. Zhao, M. Tian, W. Xie, and H. Liu, "A new parallel dual-channel fully convolutional network via semi-supervised FCM for PolSAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4493–4505, Aug. 2020.
- [37] R. Kotru, M. Shaikh, V. Turkar, S. Simu, S. Banerjee, and G. Singh, "Semantic segmentation of PolSAR images for various land cover features," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 351–354.
- [38] Y. Cui, F. Liu, X. Liu, L. Li, and X. Qian, "TCSPANET: Two-staged contrastive learning and sub-patch attention based network for PolSAR image classification," *Remote Sens.*, vol. 14, no. 10, 2022, Art. no. 2451.
- [39] R. Shang, J. Wang, L. Jiao, X. Yang, and Y. Li, "Spatial feature-based convolutional neural network for PolSAR image classification," *Appl. Soft Comput.*, vol. 123, 2022, Art. no. 108922.
- [40] Y. Cui et al., "Polarimetric multipath convolutional neural network for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–8, May, 2021.
- [41] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.
- [42] J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman, "How to grow a mind: Statistics, structure, and abstraction," *Science*, vol. 331, no. 6022, pp. 1279–1285, 2011.
- [43] X. Zhang, L. Zhaoping, T. Zhou, and F. Fang, "Neural activities in V1 create a bottom-up saliency map," *Neuron*, vol. 73, no. 1, pp. 183–192, 2012.
- [44] A. Finkelstein, D. Derdikman, A. Rubin, J. N. Foerster, L. Las, and N. Ulanovsky, "Three-dimensional head-direction coding in the bat brain," *Nature*, vol. 517, no. 7533, pp. 159–164, 2015.
- [45] S. A. Josselyn and S. Tonegawa, "Memory engrams: Recalling the past and imagining the future," *Science*, vol. 367, no. 6473, 2020, Art. no. eaaw4325.
- [46] T. E. Bakken et al., "Comparative cellular analysis of motor cortex in human, marmoset and mouse," *Nature*, vol. 598, no. 7879, pp. 111–119, 2021.
- [47] M. Tygert, J. Bruna, S. Chintala, Y. LeCun, S. Piantino, and A. Szlam, "A mathematical motivation for complex-valued convolutional networks," *Neural Computation*, vol. 28, no. 5, pp. 815–825, 2016.
- [48] A. Hirose, *Complex-Valued Neural Networks: Advances and Applications*. Hoboken, NJ, USA: Wiley, 2013.
- [49] 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.
- [50] L. Yu et al., "A lightweight complex-valued DeepLabv3 for semantic segmentation of PolSAR image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 930–943, Jan. 2022.
- [51] H. Dong, B. Zou, L. Zhang, and S. Zhang, "Automatic design of CNNs via differentiable neural architecture search for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6362–6375, Sep. 2020.
- [52] X. Tan, M. Li, P. Zhang, Y. Wu, and W. Song, "Deep triplet complex-valued network for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10179–10196, Dec. 2021.
- [53] Y. Jiang, M. Li, P. Zhang, X. Tan, and W. Song, "Unsupervised complex-valued sparse feature learning for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–6, Jun. 2022.
- [54] W. Xie, L. Jiao, and W. Hua, "Complex-valued multi-scale fully convolutional network with stacked-dilated convolution for PolSAR image classification," *Remote Sens.*, vol. 14, no. 15, 2022, Art. no. 3737.
- [55] 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.
- [56] M. Liu, L. Jiao, X. Liu, L. Li, F. Liu, and S. Yang, "C-CNN: Contourlet convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 6, pp. 2636–2649, Jun. 2021.
- [57] J. Gao, L. Jiao, F. Liu, S. Yang, B. Hou, and X. Liu, "Multiscale curvelet scattering network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 7, pp. 3665–3679, Jul. 2023.
- [58] H. Bi, L. Xu, X. Cao, Y. Xue, and Z. Xu, "Polarimetric SAR image semantic segmentation with 3D discrete wavelet transform and Markov random field," *IEEE Trans. Image Process.*, vol. 29, pp. 6601–6614, Jun. 2020.
- [59] R. Qin, X. Fu, and P. Lang, "PolSAR image classification based on low-frequency and contour subbands-driven polarimetric SENet," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4760–4773, Aug. 2020.

- [60] S. Yang, M. Wang, Z. Feng, Z. Liu, and R. Li, "Deep sparse tensor filtering network for synthetic aperture radar images classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3919–3924, Aug. 2018.
- [61] U. Anders and O. Korn, "Model selection in neural networks," *Neural Netw. Official J. Int. Neural Netw. Soc.*, vol. 12, pp. 309–323, 1999.
- [62] A. L. Da Cunha, J. Zhou, and M. N. Do, "The nonsubsampling contourlet transform: Theory, design, and applications," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3089–3101, Oct. 2006.
- [63] D. L. Donoho and A. G. Flesia, "Can recent innovations in harmonic analysis 'explain' key findings in natural image statistics?," *Netw.: Computation Neural Syst.*, vol. 12, no. 3, pp. 371–393, 2001.
- [64] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2435–2476, Oct. 1998.
- [65] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
- [66] E. J. Candés, "What is...a curvelet?," *Notices Amer. Math. Soc.*, vol. 50, no. 11, pp. 1402–1403, 2003.
- [67] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.
- [68] M. J. Shensa et al., "The discrete wavelet transform: Wedding the a Troun and Mallat algorithms," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2464–2482, Oct. 1992.
- [69] R. H. Bamberger and M. J. Smith, "A filter bank for the directional decomposition of images: Theory and design," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 882–893, Apr. 1992.
- [70] "Earth online," 2009. [Online]. Available: <http://envisat.esa.int/POLSARpro/datasets.html2>
- [71] H. Xie, S. Wang, K. Liu, S. Lin, and B. Hou, "Multilayer feature learning for polarimetric synthetic radar data classification," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2014, pp. 2818–2821.
- [72] L. Li, L. Ma, L. Jiao, F. Liu, Q. Sun, and J. Zhao, "Complex contourlet-CNN for polarimetric SAR image classification," *Pattern Recognit.*, vol. 100, 2020, Art. no. 107110.
- [73] L. F. Rossi, K. D. Harris, and M. Carandini, "Spatial connectivity matches direction selectivity in visual cortex," *Nature*, vol. 588, pp. 648–652, 2020, doi: [10.1038/s41586-020-2894-4](https://doi.org/10.1038/s41586-020-2894-4).



Mengkun Liu received the B.S. degree in digital media technology and M.S. degree in signal and information processing from the Xi'an University of Technology, Xi'an, China, in 2014 and 2017, respectively. She is currently working toward the Ph.D. degree in intelligent information processing with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence, Xidian University, Xi'an, China.

Her research interests include machine learning and image processing.



Licheng Jiao (Fellow, IEEE) received the B.S. degree in circuit and system from Shanghai Jiaotong University, Shanghai, China, in 1982, and the M.S. and Ph.D. degrees in circuit and system 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 Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China. His research

interests include image processing, natural computation, machine learning, and intelligent information processing.

Dr. Jiao is the Chairman of the Awards and Recognition Committee, the Vice Board Chairperson of the Chinese Association of Artificial Intelligence, the Foreign member of the Academia Europaea, the Foreign Member of the Russian Academy of Natural Sciences, the Fellow of IET/CAAI/CIE/CCF/CAA, a Councilor of the Chinese Institute of Electronics, a Committee Member of the Chinese Committee of Neural Networks, and an Expert of the Academic Degrees Committee of the State Council.



Xu Liu (Member, IEEE) received the B.S. degree in mathematics and applied mathematics from the North University of China, Taiyuan, China, in 2013, and the Ph.D. degree in circuit and system from Xidian University, Xi'an, China, in 2019.

He is currently an Associate Professor of Huashan Elite and Postdoctoral Researcher with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence, Xidian University, Xi'an. His current research interests include machine learning and image processing.

Dr. Liu is the Chair of IEEE Xidian University student branch (2015–2019).



Lingling Li (Senior Member, IEEE) received the B.S. and Ph.D. degrees in circuit and system from Xidian University, Xi'an, China, in 2011 and 2017, respectively.

Between 2013 and 2014, she was an Exchange Ph.D. Student with the Intelligent Systems Group, Department of Computer Science and Artificial Intelligence, University of the Basque Country UPV/EHU, Leioa, Spain. She is currently a Postdoctoral Researcher with the School of Artificial Intelligence, Xidian University. Her current research

interests include quantum evolutionary optimization, machine learning, and deep learning.



Fang Liu (Senior Member, IEEE) received the B.S. degree in computer science and technology from Xi'an Jiaotong University, Xi'an, China, in 1984, and the M.S. degree in computer science and technology from Xidian University, Xi'an, in 1995.

She is currently a Professor with the School of Computer Science, Xidian University. Her research interests include signal and image processing, synthetic aperture radar image processing, multiscale geometry analysis, learning theory and algorithms, optimization problems, and data mining.



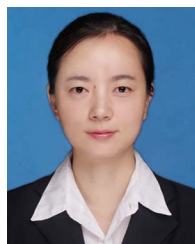
Shuyuan Yang (Senior Member, IEEE) received the B.A. degree in electrical engineering and the M.S. and Ph.D. degrees in circuit and system from Xidian University, Xi'an, China, in 2000, 2003, and 2005, respectively.

She has been a Professor with the School of Artificial Intelligence, Xidian University, Xi'an. Her research interests include machine learning and multiscale geometric analysis.



Yuwei Guo (Senior Member, IEEE) was born in Shaanxi, China. She is currently working toward the M.S. and Ph.D. degrees in circuit and system with Xidian University, Xi'an, China.

She is currently an Associate Professor with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University. Her interests include rough set theory, data mining, and image processing.



Puhua Chen (Senior Member, IEEE) received the B.S. degree in environmental engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2009, and the Ph.D. degree in circuit and system from Xidian University, Xi'an, China, in 2016.

She is currently an Associate Professor with the School of Artificial Intelligence, Xidian University. Her current research interests include machine learning, pattern recognition, and remote sensing image interpretation.