




A Lightweight Complex-Valued DeepLabv3+ for Semantic Segmentation of PolSAR Image

Lingjuan Yu , Zhaoxin Zeng, Ao Liu, Xiaochun Xie, Haipeng Wang , *Senior Member, IEEE*,
Feng Xu , *Senior Member, IEEE*, and Wen Hong, *Senior Member, IEEE*

Abstract—Semantic image segmentation is one kind of end-to-end segmentation method which can classify the target region pixel by pixel. As a classic semantic segmentation network in optical images, DeepLabv3+ can achieve a good segmentation performance. However, when this network is directly used in the semantic segmentation of polarimetric synthetic aperture radar (PolSAR) image, it is hard to obtain the ideal segmentation results. The reason is that it is easy to yield overfitting due to the small PolSAR dataset. In this article, a lightweight complex-valued DeepLabv3+ (L-CV-DeepLabv3+) is proposed for semantic segmentation of PolSAR data. It has two significant advantages when compared with the original DeepLabv3+. First, the proposed network with the simplified structure and parameters can be suitable for the small PolSAR data, and thus, it can effectively avoid the overfitting. Second, the proposed complex-valued (CV) network can make full use of both amplitude and phase information of PolSAR data, which brings better segmentation performance than the real-valued (RV) network, and the related CV operations are strictly true in the mathematical sense. Experimental results about two Flevoland datasets and one San Francisco dataset show that the proposed network can obtain better overall average, mean intersection over union, and mean pixel accuracy than the original DeepLabv3+ and some other RV semantic segmentation networks.

Index Terms—Lightweight complex-valued DeepLabv3+ (L-CV-DeepLabv3+), polarimetric synthetic aperture radar (SAR), segmentation performance, semantic image segmentation.

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) has all-day and all-weather imaging capability, which is very important in military and civilian fields. As one of the research hotspots

Manuscript received August 1, 2021; revised October 26, 2021 and December 12, 2021; accepted December 29, 2021. Date of publication January 4, 2022; date of current version January 17, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61501210 and in part by the Science and Technology Project of Jiangxi Provincial Education Department under Grant 190459. (*Corresponding authors: Lingjuan Yu; Zhaoxin Zeng.*)

Lingjuan Yu, Zhaoxin Zeng, and Ao Liu are with the School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China (e-mail: yljsmile@163.com; zzx951128@gmail.com; liuao0510@gmail.com).

Xiaochun Xie is with the School of Physics and Electronic Information, Gannan Normal University, Ganzhou 341000, China (e-mail: xiexiaochun@gnnu.cn).

Haipeng Wang and Feng Xu are with the Key Laboratory for Information Science of Electromagnetic Waves (MoE), Fudan University, Shanghai 200433, China (e-mail: hpwang@fudan.edu.cn; fengxu@fudan.edu.cn).

Wen Hong is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China (e-mail: whong@mail.ie.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3140101

in the polarimetric synthetic aperture radar (PolSAR) field, the classification of land cover is of great significance. In the early days, the classification methods were mainly based on statistical distribution and physical scattering mechanism, respectively. In terms of statistical distribution, the single-look PolSAR data are subject to complex Gaussian distribution [1], and the multilook PolSAR data are subject to complex Wishart distribution [2]. In terms of physical scattering mechanism, the scattering types of land cover constituted by odd scattering, even scattering, and diffuse scattering [3] were used, and the scattering types characterized by polarization entropy and scattering angle [4] were also adopted. Moreover, hybrid algorithms, based on statistical distribution and physical scattering mechanism [5], and the improved hybrid algorithms [6] were proposed. However, the accuracy of these methods was not high enough because only some shallow features about targets were exploited.

In order to improve the classification accuracy of land cover, some methods based on machine learning were used in PolSAR image classification. In these methods, features were extracted, and then classification was carried out based on these features. The commonly used polarization features included polarimetric decomposition features (H/A/ α decomposition [4], Freeman decomposition [7], etc.), and polarimetric parameters (the intensities of the copolarized channel in linear and circular polarization, the span, the ration between different intensity channels in linear and circular polarization bases, the modulus and phase of polarimetric degree of coherence, the minimum and maximum of the degree of polarization [8], etc.). The commonly used classifiers were support vector machine [9]–[12], neural network [13]–[15], k-nearest neighbor [16], decision tree [17], boosting [18], and so on.

In recent years, PolSAR image classification based on deep learning has achieved remarkable results. The classic deep stacked networks were multilayer perceptrons [19], convolutional long short term memory [20], deep belief network (DBN) [21], stacked autoencoder [22], [23], convolutional neural network (CNN) [24]–[29], and so on. The classification processes of these deep learning networks also included the feature extraction and classification. First, the PolSAR data or polarimetric decomposed features were used as the input of network. Then, the deep features were extracted from the input data by layers in the front of the network. Finally, the extracted features were classified by layers in the back of the network. In addition, there were some other deep networks used in the classification of PolSAR image, such as

3-D-CNN combined with the conditional random field [30], CNN combined with graph [31], polarimetric convolutional network [32], Wishart deep stacking network [33], Wishart autoencoder (Wishart-AE) [34], Wishart convolutional autoencoder (Wishart-CAE) [34], and the improved deep stacked network [35], [36]. Because some superpixel segmentation methods were proved to be effective in preserving spatial structure information of PolSAR images [37]–[39], the superpixel-based graph convolutional network was also proposed for PolSAR image classification [40].

Among the above classification methods based on deep learning, some of them were pixel-by-pixel classification methods. In these methods, the sliding window was used in dividing the whole PolSAR image into many overlapping patches, and then the class of each pixel in the whole image was obtained by classifying the patch centered on this pixel as an image. The disadvantages of these methods were that the classification results were easily affected by the speckle noise, and the amount of computation was large. Besides, there were also a few region-based classification methods which could effectively retain the whole region of the target. In these methods, regions were generated by superpixel segmentation [23], [40] or other methods [12], and then the obtained target regions were classified.

With the development of deep learning, semantic image segmentation has also developed rapidly. It achieved not only pixel-by-pixel classification but also region-level segmentation. So far, this technology has also been applied in the single-polarization SAR and PolSAR images. For the semantic segmentation of single-polarization SAR image, it mainly focused on one kind of specific target, such as road [41], oil spill [42], and building [43]–[46]. Besides, the semantic segmentation of land cover is also studied [47]. For the semantic segmentation of PolSAR image, it mainly focused on the land cover [48]–[52]. Since the original semantic segmentation networks were proposed for optical images, the input data of these networks were real-valued (RV). However, both single-polarization SAR and PolSAR data are complex valued (CV). In order to make full use of both the amplitude and phase information of SAR data, some CV semantic segmentation networks were also proposed [53], [54]. Although these CV networks achieved good segmentation results, some mathematical operations involved in these networks were not strictly true in the mathematical sense. In this article, we propose a lightweight complex-valued DeepLabv3+ (L-CV-DeepLabv3+) for semantic segmentation of PolSAR image in order to obtain better segmentation performance than classic RV segmentation networks based on deep learning. The structure and parameters of the proposed network are simplified based on the original DeepLabv3+ [55], and all the CV operations involved in this network are mathematically strict. Two Flevoland datasets and one San Francisco dataset are used in verifying the effectiveness of the proposed network.

The rest of this article is organized as follows. Section II presents the detailed structure, parameters, and the related CV operations of L-CV-DeepLabv3+. Section III introduces three PolSAR datasets and the corresponding preprocessing. The experimental results are shown in Section IV. Finally, Section V concludes this article.

II. THEORY FOR L-CV-DEEPLABV3+

The architecture of L-CV-DeepLabv3+ is shown in Fig. 1. It includes two modules: 1) encoder and 2) decoder. In the encoder, there are backbone network, complex-valued atrous spatial pyramid pooling (CV-ASPP), and CV convolution (CV-Conv) operation with size 1×1 . In the decoder, there are CV-Conv operations with size 1×1 and 3×3 , and two upsampling operations with ratio 4. At the end of the decoder, there is a magnitude operation which converts the CV output to RV before the final softmax operation. It is worth noting that each convolution operation in the proposed network is followed by a CV activation function and CV batch normalization, although they are not shown in Fig. 1.

In order to further analyze the proposed network, both the structures of backbone network, CV-ASPP, decoder, and the involved mathematical operations in these subnetworks are presented in this section. In addition, the loss function is also given, and the CV batch normalization and CV weight initialization are simply introduced.

A. Backbone Network

In the original DeepLabv3+ [55], ResNet, Xception, MobileNet are always selected as the backbone network in semantic segmentation of optical images. However, when these deep structures are directly used in the semantic segmentation of PolSAR images, it is hard to obtain good segmentation results. The reason is that it is easy to yield overfitting due to the small PolSAR dataset. To solve this problem, we propose a lightweight complex-valued Xception (L-CV-Xception) as the backbone network. It also contains three parts: 1) entry flow, 2) middle flow, and 3) exit flow, which are shown in Fig. 2(a)–(c), respectively. As shown in Fig. 2(a), the entry flow includes CV-Conv operations, shortcut connections, and complex-valued separation convolution (CV-SepConv) operations, which is similar to the original entry flow. However, the number of convolution kernels in each layer is reduced to be about 1/2 or 1/4 of the original. In Fig. 2(b), the structure of middle flow is similar to the original. But the number of convolution kernels in each layer is reduced to be 1/4 of the original, and the number of repeat times of the structure is reduced to be 10, which is obtained by experiments in Section IV. In Fig. 2(c), the structure of exit flow is also similar to the original, but the number of convolution kernels in each layer is reduced to be 1/4 of the original. Overall, the proposed L-CV-Xception is lighter than the original RV Xception. The segmentation performance of L-CV-Xception obtained by experiments in Section IV is better than the original.

All the convolution operations involved in L-CV-Xception are CV. As shown in Fig. 2, it mainly includes CV-Conv operation with stride s ($s \geq 1$) and CV-SepConv operation. For the latter, it can also be divided into two steps just like the SepConv operation in the original Xception. The first step is complex-valued depthwise convolution (CV-DWConv) operation, and the second step is complex-valued pointwise convolution (CV-PWConv) operation. The schematic diagrams of these two steps are shown in Fig. 3(a)–(b), respectively. In Fig. 3(a),

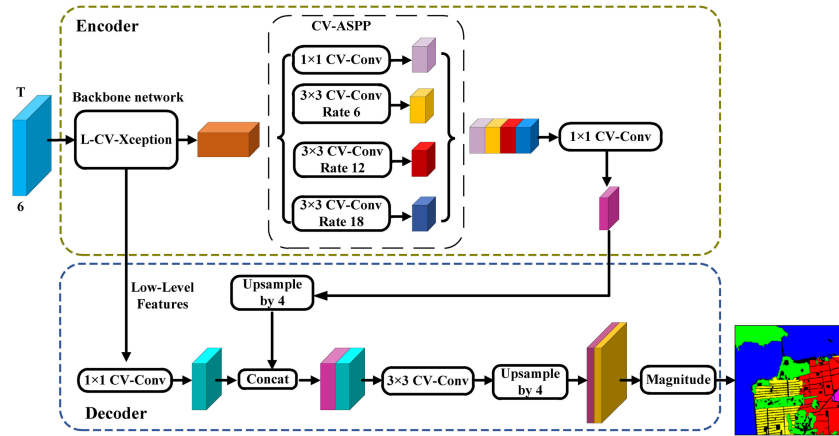


Fig. 1. Architecture of L-CV-DeepLabv3+. It includes two modules: Encoder and decoder. All the convolution operations in two modules are complex-valued (CV). Lightweight complex-valued Xception (L-CV-Xception) is used as the backbone network of the encoder.

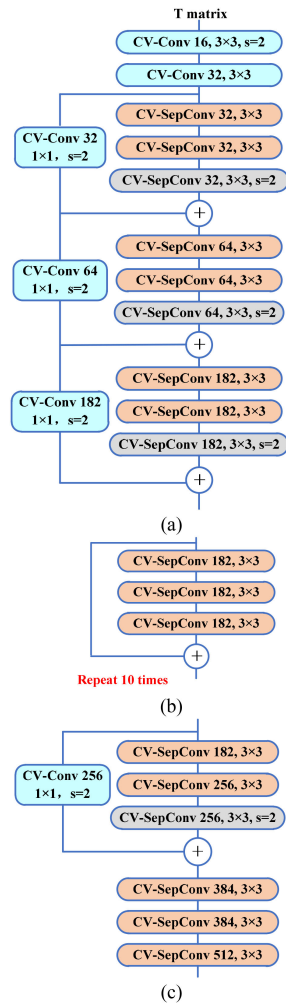


Fig. 2. Architecture of L-CV-Xception. (a) Entry flow. (b) Middle flow. (c) Exit flow.

one convolution kernel is convolved with one input channel, and the number of output channels is the same as that of input channels. In Fig. 3(b), each input feature map is convolved

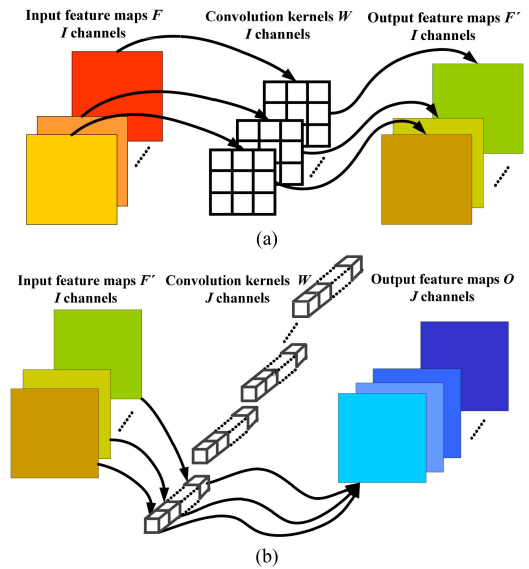


Fig. 3. Schematic diagram of CV-SepConv operation. (a) CV-DWConv operation. (b) CV-PWConv operation.

with one convolution kernel with size 1×1 , and then all the convolutional results of input feature maps are summed together to obtain one final feature map. In order to clearly illustrate all these convolution operations, their mathematical formulas are considered.

At first, some common symbols are defined. The i th ($i = 1, 2, \dots, I$) input feature map is denoted as F_i , where I is the number of CV input feature channels, and the j th ($j = 1, 2, \dots, J$) output feature map is denoted as O_j , where J is the number of CV convolution kernels. The pixel coordinates in each input feature map are denoted as (x, y) , and the pixel coordinates in each convolution kernel are denoted as (u, v) . Then, the mathematical formulas of CV-Conv, CV-DWConv, and CV-PWConv operations are presented as follows.

1) *CV-Conv operation with any stride*: Suppose the size of any one convolution kernel W_{ji} is $K_1 \times K_2$, then the CV-Conv operation with stride s ($s \geq 1$) can be written as follows [56]:

$$O_j(x, y) = \sum_{i=1}^I \sum_{u=0}^{K_1-1} \sum_{v=0}^{K_2-1} W_{ji}(u, v) F_i(xs + u, ys + v) \quad (1)$$

2) *CV-DWConv operation*: As shown in Fig. 3(a), the number of CV output feature channels is the same as that of input feature channels, which is equal to I . Denote the i th output feature maps as F_i' , and suppose the size of any one convolution kernel W_i is also $K_1 \times K_2$, then the CV-DWConv operation can be written by

$$F_i'(x, y) = \sum_{u=0}^{K_1-1} \sum_{v=0}^{K_2-1} W_i(u, v) F_i(x + u, y + v). \quad (2)$$

3) *CV-PWConv Operation*: As shown in Fig. 3(b), the input feature maps are also the output results of CV-DWConv operation. The number of input feature maps is I , and the number of output feature maps is J . Suppose the size of any one convolution kernel W_{ji} is 1×1 , then the CV-PWConv operation can be written by

$$O_j(x, y) = \sum_{i=1}^I W_{ji} F_i'(x, y). \quad (3)$$

B. Complex-Valued Atrous Spatial Pyramid Pooling

The structure of CV-ASPP is shown in Fig. 1. It includes one CV-Conv operation with size 1×1 , and three complex-valued dilated convolution (CV-DConv) operations with size 3×3 whose rates are 6, 12, and 18, respectively. The former is used in compressing the channels, while the latter is used in obtaining multiscale features. Obviously, there are two main differences between CV-ASPP and the original ASPP. The first one is that all the operations in CV-ASPP are CV, while they are RV operations in the original ASPP. The second one is that the global pooling in the original ASPP is removed in CV-ASPP. There are two reasons for the second difference. One reason is that the output feature size of the backbone network is too small to be used in obtaining the global feature by the pooling layer, and the other reason is that the CV pooling operation which is complicated in the mathematical sense can be avoided.

If the symbols defined in the CV-Conv operation with stride s are also used in CV-ASPP, then the CV-DConv operation can be written by

$$O_j(x, y) = \sum_{i=1}^I \sum_{u=0}^{K_1-1} \sum_{v=0}^{K_2-1} W_{ji}(u, v) F_i(x + ru, y + rv) \quad (4)$$

where r is the dilated ratio.

C. Decoder

As shown in Fig. 1, both feature maps obtained from L-CV-Xception and CV-ASPP are used in the decoder. The feature maps from L-CV-Xception represent low-level features, while the feature maps from CV-ASPP represents high-level features.

At the beginning of the decoder, the low-level features are convoluted by 1×1 convolution kernels in order to reduce the proportion of low-level features, and the high-level features are upsampled to be with the same size as the low-level features. Considering that CV nearest neighbor interpolation is easier than CV bilinear interpolation, the former is used in the CV upsampling operation. Then, the concatenation is performed to obtain rich feature maps containing both low-level and high-level features. After that, CV-Conv operation with size 3×3 is used to refine feature maps, and then the refined feature maps are upsampled four times again. Specifically, all the above operations are CV. Since the CV softmax operation is complicated in the mathematical sense [57], a magnitude operation is used in converting the extracted feature maps from CV into RV, which will not lead to information loss in the backward propagation of the whole network [56]. At the end of the decoder, the softmax classifier is used in obtaining the final semantic segmentation results.

If the symbols defined in the backbone network are also used in the decoder, then the mathematical formulas of CV upsampling, magnitude, and the softmax operations involved in the decoder can be given as follows.

1) *CV Upsampling Operation*: Since the CV upsampling operation is only used in enlarging the input feature maps, the number of output feature channels is the same as that of the input feature channels. Suppose each pixel in one input feature map is upsampled into a block with size $K \times K$, then the i th output feature map can be calculated by

$$O_i(xK + k, yK + m) = F_i(x, y) \quad (5)$$

where $k \in [0, K - 1]$ and $m \in [0, K - 1]$.

2) *Magnitude Operation*: According to the framework of L-CV-DeepLabv3+, the number of the input channels of softmax operation is the same as that of the output channels, which is also equal to the number of target classes. Suppose the number of target classes is N , then the n th ($n = 1, 2, \dots, N$) output of magnitude operation can be calculated by

$$O_n(x, y) = \sqrt{(\Re(F_n(x, y)))^2 + (\Im(F_n(x, y)))^2} \quad (6)$$

where $\Re(\bullet)$ and $\Im(\bullet)$ denote real and imaginary parts of a complex number, respectively.

3) *Softmax Operation*: After the magnitude operation, the softmax operation can be directly used in the final semantic segmentation of images. The probability of the pixel located at (x, y) in the feature map belonging to the n th class can be written by

$$p_n(x, y) = \frac{\exp(O_n(x, y))}{\sum_{n'=1}^N \exp(O_{n'}(x, y))}. \quad (7)$$

D. Loss Function

The cross-entropy [58] is chosen as the loss function of the proposed L-CV-DeepLabv3+, which is the same as the original DeepLabv3+. Taking one sample as an example, the

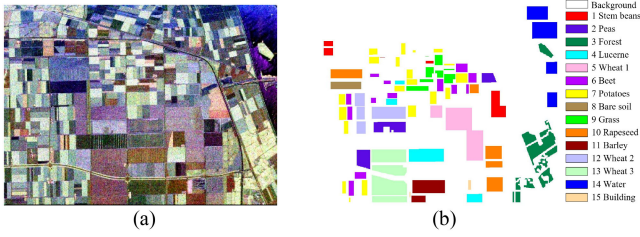


Fig. 4. Flevoland dataset 1. (a) Pauli RGB. (b) Ground truth and the legend of ground truth.

loss function can be expressed by

$$\text{Loss} = - \sum_{(x,y)} \sum_{n=1}^N q_n(x,y) \ln p_n(x,y) \quad (8)$$

where $q_n(x,y)$ represents the true classification result of one pixel located at (x,y) in the final segmentation image. If the label of this pixel is n , then $q_n(x,y)$ is equal to 1; otherwise, $q_n(x,y)$ is equal to 0.

E. CV Batch Normalization and Weight Initialization

Batch normalization is very important in the training of the deep network. It can not only accelerate the convergence speed of the network model but also alleviate the gradient dispersion problem in deep network to a certain extent. In the framework of L-CV-DeepLabv3+, the CV batch normalization in [59] is used. The CV convolution results are standardized to obey standard complex distribution with mean 0 and covariance 1.

Weight initialization also has a crucial impact on the convergence speed and performance of the network model. It can reduce the risk of gradient explosion and gradient dispersion. The CV weight initialization in [59] is also used in the proposed L-CV-DeepLabv3+. The magnitude of a CV weight obeys Rayleigh distribution, and the phase obeys the Uniform distribution between $-\pi$ and π .

III. DATASETS AND DATA PREPROCESSING

In this section, three fully PolSAR datasets (i.e., two Flevoland datasets and one San Francisco dataset) used in verifying the effectiveness of the proposed L-CV-DeepLabv3+ are introduced. Then, the data preprocessing of these datasets is presented. Finally, the mathematical formulas of four metrics for evaluating segmentation performance of the proposed network are given.

A. Datasets

The first L-band full polarimetric Flevoland dataset is acquired by AIRSAR airborne platform in August 1989. There are 15 types of land covers, namely, stem beans, peas, forest, lucerne, three types of wheat, beet, potatoes, bare soil, grass, rapeseed, barley, water, and a small amount of buildings. The RGB image with size 1024×750 obtained by Pauli decomposition is shown in Fig. 4(a). The ground truth and the legend of ground truth

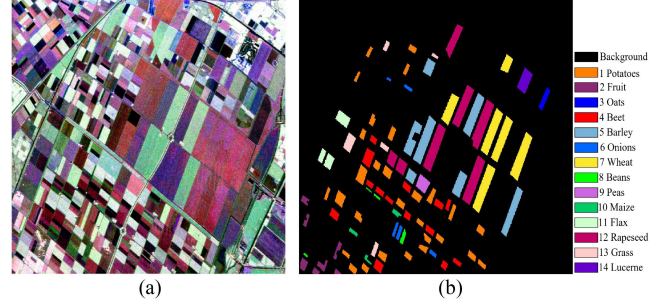


Fig. 5. Flevoland dataset 2. (a) Pauli RGB. (b) Ground truth and the legend of ground truth.

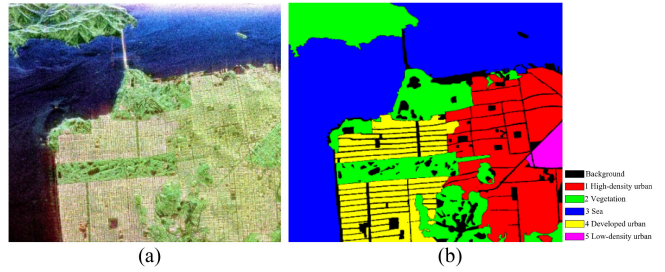


Fig. 6. San Francisco dataset. (a) Pauli RGB. (b) Ground truth and the legend of ground truth.

which are generated by LabelMe toolkit are shown in Fig. 4(b), where the white regions are regarded as the background.

The second L-band full polarimetric Flevoland dataset is acquired by AIRSAR airborne platform in 1991. There are 14 types of land covers, namely, potatoes, fruit, oats, beet, barley, onions, wheat, beans, peas, maize, flax, rapeseed, grass, and lucerne. The RGB image with size 1024×1020 obtained by Pauli decomposition is shown in Fig. 5(a), and the ground truth and the legend of ground truth are shown in Fig. 5(b), where the black regions are regarded as the background.

The third L-band full polarimetric San Francisco dataset is acquired by AIRSAR airborne platform in 2008. There are five types of land covers, namely, high-density urban, water, vegetation, developed urban, and low-density urban. The RGB image with size 1024×900 obtained by Pauli decomposition is shown in Fig. 6(a), and the ground truth and the legend of ground truth are shown in Fig. 6(b), where the black regions are regarded as the background.

B. Data Preprocessing

The proposed L-CV-DeepLabv3+ allows its input to be CV, so the CV input data is considered. For PolSAR datasets, in monostatic mode, the backscattering coefficient S_{HV} is equal to S_{VH} according to the reciprocity theorem, where the subscripts H and V represent the horizontal and vertical polarization bases, respectively. Thus, the scattering vector obtained by Pauli decomposition can be simplified, and then coherency matrix

TABLE I
TRAIN AND TEST SET

Dataset	Train	Test
Flevoland dataset 1	1836	75
Flevoland dataset 2	2160	85
San Francisco dataset	3456	144

obtained from multilook data processing can be expressed by

$$\mathbf{T} = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix} \quad (9)$$

where T_{11}, T_{22}, T_{33} are RV, and T_{12}, T_{13}, T_{23} are CV.

Because the coherency matrix \mathbf{T} is a Hermitian symmetric matrix, we choose the upper triangular $\{T_{11}, T_{22}, T_{33}, T_{12}, T_{13}, T_{23}\}$ as the 6-channel input of L-CV-DeepLabv3+.

In Section IV, five RV semantic segmentation networks are used for comparisons. For these RV networks, the input data must be RV. Thus, another 6-channel RV input data is calculated as follows [24]:

$$\begin{cases} A = 10\log_{10}(\text{SPAN}) \\ B = T_{22}/\text{SPAN} \\ C = T_{33}/\text{SPAN} \\ D = |T_{12}|/\sqrt{T_{11} \cdot T_{22}} \\ E = |T_{13}|/\sqrt{T_{11} \cdot T_{33}} \\ F = |T_{23}|/\sqrt{T_{22} \cdot T_{33}} \end{cases} \quad (10)$$

where $\text{SPAN} = T_{11} + T_{22} + T_{33}$.

From (9) and (10), both the CV and RV inputs of the network are derived from the matrix \mathbf{T} . As a result, the following data preprocessing is for the matrix \mathbf{T} . At first, two Flevoland datasets and one San Francisco dataset are expanded in mirror mode to be with size 1024×832 , 1024×1024 , and 1024×960 , respectively. Then three datasets are cut into many nonoverlapping blocks with size 64×64 through the sliding window, respectively. Subsequently, for each dataset, 40% of blocks are randomly chosen as the training samples, and the remaining 60% are chosen as the testing samples. Finally, each dataset is expanded by scaling and rotation. The numbers of training and testing samples of three datasets are shown in Table I, respectively. It is obvious that the number of training samples is greatly increased after the data expansion.

C. Metrics

Four metrics are used in evaluating the segmentation performance of the proposed network. They are intersection over union (IOU), mean intersection over union (MIOU), overall accuracy (OA), and mean pixel accuracy (MPA) [60]. IOU is used to evaluate the segmentation performance of each class, while other three metrics are used to evaluate the average segmentation performance of all classes. Suppose there are N classes in total. Denote P_{ii} ($i = 1, 2, \dots, N$) as the number of pixels of class i predicted to belong to class i , and denote P_{ij} ($j = 1, 2, \dots, N$) as the number of pixels predicted to belong to class j . Then mathematical formulas of four metrics

can be respectively written by

$$\text{IOU} = \frac{P_{ii}}{\sum_{j=0}^N P_{ij} + \sum_{j=0}^N P_{ji} - P_{ii}} \quad (11)$$

$$\text{MIOU} = \frac{1}{N+1} \sum_{i=0}^N \frac{P_{ii}}{\sum_{j=0}^N P_{ij} + \sum_{j=0}^N P_{ji} - P_{ii}} \quad (12)$$

$$\text{OA} = \frac{\sum_{i=0}^N P_{ii}}{\sum_{i=0}^N \sum_{j=0}^N P_{ij}} \quad (13)$$

$$\text{MPA} = \frac{1}{N+1} \sum_{i=0}^N \frac{P_{ii}}{\sum_{j=0}^N P_{ij}} \quad (14)$$

IV. EXPERIMENTAL RESULTS

In this section, the selection of structure and parameters of the backbone network is analyzed at first. Then, semantic segmentation experiments on three PolSAR datasets are implemented. In order to verify the segmentation performance of L-CV-DeepLabv3+, five classic RV networks (i.e., FCN [60], U-Net [61], SegNet [62], PSPNet [63], and DeepLabv3+ [55]) are used for comparisons. Among these five networks, FCN-8s is selected; softmax classifier is used in the multiclassification for U-Net; ResNet50 is used as the backbone network of PSPNet; and Xception is used as the backbone network of DeepLabv3+. The purpose of these options is to obtain good segmentation results as much as possible.

A. Selection of Structure and Parameters of the Backbone Network

The backbone network of the proposed L-CV-DeepLabv3+ is shown in Fig. 2. It is obvious that the number of convolution kernel channels in each layer of entry flow, middle flow and exit flow is reduced to be about 1/2 or 1/4 of the original DeepLabv3+. In order to clearly explain the influence of structure and parameters of the backbone network on segmentation performance, experiments on the number of repetitions of the middle flow structure are implemented firstly.

When the number of repetitions of the middle flow given in Fig. 2(b) is changed from 2 to 16, and the other structure and parameters shown in Fig. 2 are unchanged, the OA, MPA, and MIOU obtained by L-CV-DeepLabv3+ for three datasets are shown in Fig. 7(a)–(c), respectively. In Fig. 7(a), the OA is almost unchanged with the number of repetitions, while the MPA and MIOU are changed with the number of repetitions. When the number of repetitions is 10, two metrics get the maximum. In Fig. 7(b) and (c), the same conclusions as Fig. 7(a) can also be obtained. As a result, we choose 10 as the number of repetitions of the middle flow structure for three datasets. In addition, we can also find that the OA is higher than the MPA and MIOU for each dataset.

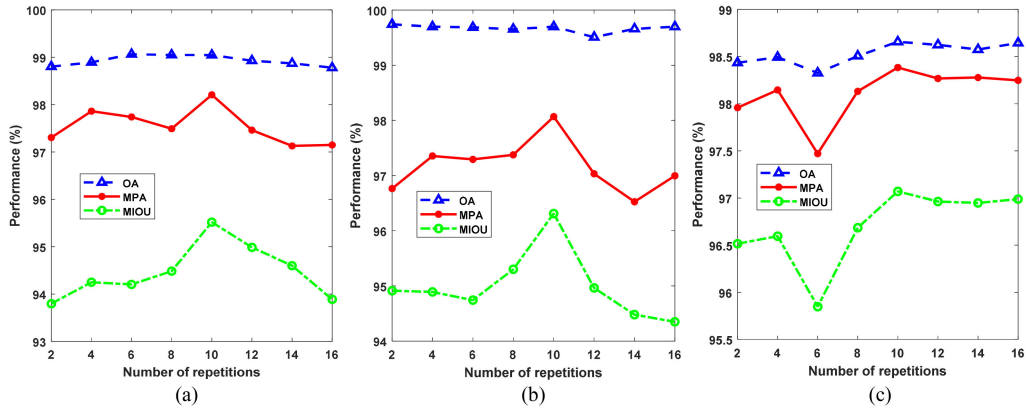


Fig. 7. Segmentation performance changing with the number of repetitions of the middle flow structure. (a) Flevoland dataset 1. (b) Flevoland dataset 2. (c) San Francisco dataset.

TABLE II
COMPARISON OF SEGMENTATION PERFORMANCE (%)

Backbone networks	Flevoland dataset 1			Flevoland dataset 2			San Francisco dataset		
	MIOU	OA	MPA	MIOU	OA	MPA	MIOU	OA	MPA
RV Xception	67.16	94.32	80.22	74.09	97.47	91.02	77.07	89.89	85.45
CV Xception	93.08	98.84	94.48	91.75	99.49	94.06	94.98	97.67	97.26
L-CV- Xception	95.52	99.05	97.74	96.31	99.7	98.07	97.07	98.66	98.38

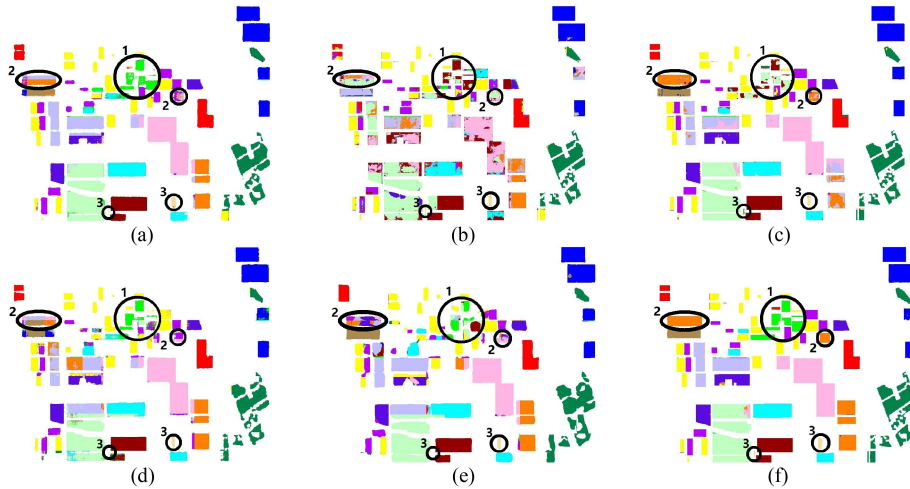


Fig. 8. Segmentation results of Flevoland dataset 1. (a) FCN. (b) U-Net. (c) SegNet. (d) PSPNet. (e) DeepLabv3+. (f) L-CV-DeepLabv3+.

Then ablation experiments are also implemented on three datasets to analyze the influence of structure and parameters of the backbone network on segmentation performance. The original RV Xception is chosen as the baseline. The CV Xception is an extension of the original RV Xception in the complex domain. Although its structure and parameters of CV Xception are the same as the original RV Xception, they are CV. The proposed L-CV-Xception is a lightweight CV Xception. The comparison of segmentation performance obtained by these three networks for three datasets is shown in Table II. It is obvious that the MIOU, OA, and MPA obtained by CV Xception are higher than those obtained by the original RV Xception. The MIOU, OA,

and MPA obtained by L-CV-Xception are the highest among the three networks.

B. Experiment on Flevoland Dataset 1

For Flevoland dataset 1, the semantic segmentation results using five classical RV networks and the proposed CV network are shown in Fig. 8(a)–(f), respectively. In these figures, there are some regions in black circles or ellipses. For these regions, segmentation results obtained by the proposed CV network are better than those obtained by the other five RV networks. In the region marked with number 1, grass is partially lost by

TABLE III
 SEGMENTATION PERFORMANCE OF FLEVOLAND DATASET 1 (%)

Networks	FCN	U-Net	SegNet	PSPNet	DeepLabv3+	L-CV-DeepLabv3+
Stem beans	85.84	63.18	94.15	87.69	83.23	96.72
Peas	73.45	6.42	93.80	74.95	59.72	93.87
Forest	75.22	87.78	95.60	76.40	71.28	94.82
Lucerne	75.10	54.24	91.89	71.68	77.82	98.86
Wheat1	77.85	50.52	78.66	76.86	79.31	95.03
Beet	76.58	54.59	91.84	47.51	49.24	98.24
Potatoes	67.02	80.96	95.85	68.67	66.28	96.20
Bare soil	43.70	4.81	95.03	39.87	73.11	99.41
Grass	57.91	6.41	0.81	47.48	26.02	97.23
Rapeseed	43.15	40.02	58.03	38.52	44.46	91.40
Barley	97.44	39.19	80.93	91.26	83.90	99.98
Wheat2	46.85	22.03	51.82	35.06	51.80	89.80
Wheat3	86.40	59.36	84.64	70.58	71.50	97.62
Water	91.94	83.42	98.10	88.06	93.57	94.19
Building	55.19	78.22	67.32	63.00	47.16	85.65
MIOU	71.80	51.95	79.90	67.21	67.16	95.52
OA	93.36	93.31	97.76	94.85	94.32	99.05
MPA	82.12	63.76	85.37	76.74	80.22	97.74

FCN; misclassified as barley and wheat 3 by U-Net and SegNet; partially misclassified as wheat 3 and beet by PSPNet; partially misclassified as barley and wheat 3 by DeepLabv3+. The segmentation boundaries of grass obtained by five RV networks are not clear enough. In the regions marked with number 2, rapeseed is misclassified as beet, wheat 1, and wheat 2 by FCN and PSPNet; misclassified as beet, wheat 1, and wheat 3 by U-Net; partially misclassified as wheat 3 by SegNet; misclassified as beet, wheat 1, wheat 2, and peas by DeepLabv3+. In the regions marked with number 3, building is partially misclassified as forest by SegNet; partially lost by FCN, PSPNet, and DeepLabv3+. However, these regions are correctly classified by the proposed L-CV-DeepLabv3+. There are two reasons. One reason is that when the amplitude information of one class is similar to others, these classes are difficult to be distinguished by their amplitude information, but they can be distinguished by phase information. The other reason is that the regions of some land covers in the blocks are so small that it is difficult to extract their features by using RV networks and RV input data without the help of phase information.

Furthermore, the IOU of each class, MIOU, OA, and MPA are, respectively, shown in Table III. It is obvious that the IOU of grass is very low by using five RV networks. Except that the IOU of grass obtained by FCN is 57.91%, the IOU obtained by any of the other four RV networks is less than 50%. However, the IOU of grass obtained by the proposed L-CV-DeepLabv3+ can achieve 97.23%. Besides, the IOU of rapeseed is also very low by using five RV networks. Except for SegNet, the IOU of rapeseed obtained by any of the other four RV networks is also less than 50%. But the IOU of rapeseed obtained by the proposed network can reach 91.4%. Because the number of building samples is very small and the structure of DeepLabv3+ is very deep, the IOU of building obtained by DeepLabv3+ is less than 50%, while the IOU of building obtained by the other four RV networks can be in the range of 50% to 80%. Although the IOU of building obtained by the proposed network can achieve 85.65%, it is lower than that of any other land cover obtained by the same

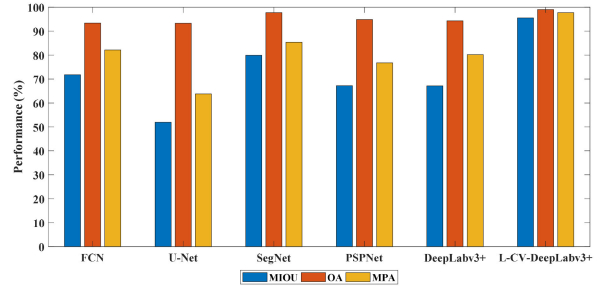


Fig. 9. Segmentation performance of Flevoland dataset 1.

network because of the very small number of samples. All these results are consistent with those shown in the black circles or ellipses in Fig. 8. In addition, the MIOU, OA, and MPA obtained by the proposed L-CV-DeepLabv3+ are more than 95%, and they are much higher than those obtained by the other five RV networks.

It is easy for a bar chart to display the differences of metrics obtained by different networks. Therefore, the MIOU, OA, and MPA listed in Table III are shown in Fig. 9. Obviously, for the proposed L-CV-DeepLabv3+, the MIOU, OA, and MPA are very high, and the difference between any two metrics is small. However, for five RV networks, the OA is higher than 90%, the MPA is in the range of 60–85%, and the MIOU is in the range of 50–80%. The MIOU and MPA are much smaller than the OA. Furtherly, for each RV network, we can obtain the difference between the OA and MIOU, and the difference between the OA and MPA from Table III. These two differences obtained by FCN are 21.56% and 11.24%, respectively. It means that two differences are greater than 10%. The similar results can also be obtained by other four RV networks. It can be explained by the formulas of three metrics. According to (16) and (18), MIOU and MPA are obtained by averaging the IOU and accuracy of all classes, respectively. According to (17), OA is the ratio of the total number of correctly classified pixels to the total number

TABLE IV
COMPARISON OF CLASSIFICATION ACCURACY (%)

Networks	W-DBN[21]	WCAE[34]	CV-CNN[25]	DMCNN [36]	3DDW-CNN[28]	L-CV-Deeplabv3+
OA	97.57	93.31	97.7	98.77	97.74	99.05

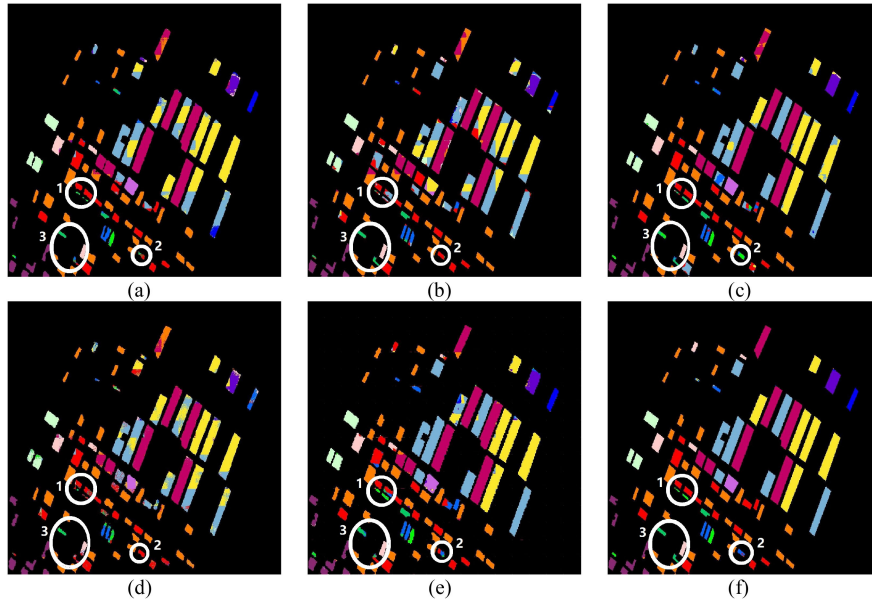


Fig. 10. Segmentation results of Flevoland dataset 2. (a) FCN. (b) U-Net. (c) SegNet. (d) PSPNet. (e) DeepLabv3+. (f) L-CV-DeepLabv3+.

of pixels for all classes. Suppose there is such a class of land cover, its total number of pixels is very small, and the number of correctly classified pixels is much less than its total number of pixels. Then, the IOU and accuracy of this class will be very small, which leads to the small MIOU and MPA. But the small number of pixels of this class has little impact on the OA. Take the segmentation results obtained by SegNet for example. The IOU of grass is only 0.81%. After the averaging operation of all classes, the MIOU is 79.9%. However, the OA can reach 97.76%.

In addition, the OA obtained by the proposed network is also compared with those obtained by some deep learning networks used in the classification of land cover. The results obtained by Wishart DBN (W-DBN) and local spatial information [21], Wishart CAE (WCAE) [34], complex-valued CNN (CV-CNN) [25], depthwise separable convolution based multitask CNN (DMCNN) [36], and 3-D depthwise separable convolution based CNN (3DDW-CNN) [28] are shown in Table IV. It is obvious that the proposed network can achieve the higher OA than other networks.

C. Experiment on Flevoland Dataset 2

For Flevoland dataset 2, the semantic segmentation results using five RV networks and the proposed CV network are shown in Fig. 10(a)–(f), respectively. In these figures, for regions in white circles or ellipses, segmentation results obtained by the proposed CV network are better than those obtained by the other five RV networks. In the region marked with number 1, beans

are partially or completely misclassified by five RV networks. In the region marked with number 2, onions are misclassified as beet or beans by five RV networks. In the region marked with number 3, maize is misclassified as beans, onions, or beet by five RV networks. However, these regions are correctly classified by the proposed L-CV-DeepLabv3+. The reason is the same as that obtained from Flevoland dataset 1.

The IOU of each class, MIOU, OA, and MPA are, respectively, shown in Table V. It is obvious that the IOU of onions obtained by any of the five networks is much lower than that obtained by the proposed network. A similar case holds for most of the other land covers. Although the IOU of beans obtained by the proposed network is larger than that obtained by the other five RV networks, it is lower than the IOU of any other land cover obtained by the same proposed network. The reason is also that the sample number of beans is smaller than that of the other land covers. All these results are consistent with those shown in the white circles or ellipses in Fig. 10. In addition, the MIOU, OA, and MPA obtained by the proposed L-CV-DeepLabv3+ are more than 96%, which are much higher than those obtained by the other five RV networks.

A bar chart is also used in representing the MIOU, OA, and MPA listed in Table V, which is shown in Fig. 11. It is easy to find that the proposed L-CV-DeepLabv3+ obtains the highest MIOU, OA, and MPA among all these networks, and the difference between any two metrics for this network is small. However, for any of the other five RV networks, the MIOU and MPA are much smaller than the OA. The difference between the OA and MIOU, and the difference between the OA and MPA

TABLE V
SEGMENTATION PERFORMANCE OF FLEVOLAND DATASET 2 (%)

Networks	FCN	U-Net	SegNet	PSPNet	DeepLabv3+	L-CV-DeepLabv3+
Potatoes	79.69	86.73	96.47	76.66	80.83	98.35
Fruit	80.63	99.69	99.84	80.49	80.92	97.47
Oats	54.80	62.12	87.10	0.13	72.29	95.48
Beet	68.21	45.57	67.81	58.49	68.40	98.33
Barley	53.20	34.99	65.82	49.65	83.02	99.31
Onions	35.42	34.42	5.47	22.45	35.69	90.26
Wheat	63.53	40.08	74.65	58.93	80.65	98.47
Beans	50.52	0.45	15.86	17.88	58.49	87.77
Peas	88.09	22.90	97.28	72.61	82.01	94.83
Maize	40.83	73.55	43.43	26.57	50.28	95.44
Flax	87.95	78.03	98.96	78.52	86.82	98.81
Rapeseed	85.59	87.50	96.47	83.07	87.31	99.28
Grass	60.62	69.21	58.56	53.65	71.00	96.27
Lucerne	73.64	78.21	88.74	67.79	75.76	94.97
MIOU	68.07	60.90	73.10	56.37	74.09	96.31
OA	96.92	96.59	97.83	96.55	97.47	99.70
MPA	80.47	71.44	80.28	65.40	91.02	98.07

TABLE VI
SEGMENTATION PERFORMANCE OF SAN FRANCISCO DATASET (%)

Networks	FCN	U-Net	SegNet	PSPNet	DeepLabv3+	L-CV-DeepLabv3+
High-density urban	80.82	73.83	66.72	82.76	73.26	96.00
Vegetation	82.16	74.21	80.37	84.24	80.39	94.91
Sea	96.86	90.12	96.28	97.80	97.86	99.10
Developed urban	63.99	70.24	66.49	78.21	68.09	96.53
Low-density urban	71.97	52.85	52.90	61.34	77.97	97.92
MIOU	75.39	76.69	76.93	81.18	77.07	97.07
OA	89.57	90.15	91.23	93.57	89.89	98.66
MPA	84.24	86.64	88.07	87.85	85.45	98.38

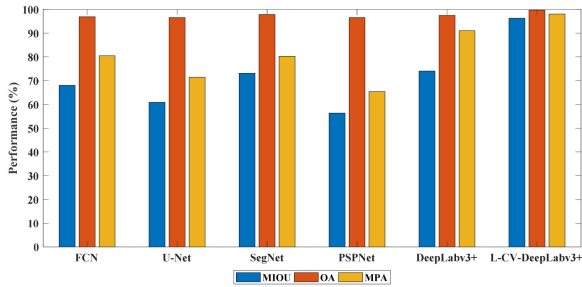


Fig. 11. Segmentation performance of Flevoland dataset 2.

are very large. The reason is the same as that obtained from Flevoland dataset 1.

D. Experimental Results on San Francisco Dataset

For San Francisco dataset, the semantic segmentation results using five RV networks and the proposed CV network are shown in Fig. 12(a)–(f) respectively. There are some regions in white circles in these figures. In the region marked with number 1, low-density urban is partially or completely misclassified as vegetation or high-density urban by five RV networks. In the region marked with number 2, high-density urban is partially misclassified as vegetation, and developed urban is partially misclassified as high-density urban. However, these regions are correctly classified by the proposed L-CV-DeepLabv3+.

The reason is the same as that obtained from the previous two datasets.

The IOU of each class, MIOU, OA, and MPA are respectively shown in Table VI. Unlike the results obtained from the previous two datasets, the IOU of each class obtained from this dataset by five RV networks is higher than 50%. The reason is that the number of training samples for each class is larger than those of the previous two datasets. Among all the land covers, the low-density urban is with poor segmentation performance by using five RV networks, which is consistent with those shown in white circles in Fig. 12. In addition, the MIOU, OA, and MPA obtained by the proposed network are more than 97%, which are much higher than those obtained by the other five RV networks.

A bar chart is also used in representing the MIOU, OA, and MPA listed in Table VI, which is shown in Fig. 13. The same conclusions as those obtained from the previous two datasets are still valid.

E. Discussion on Three Datasets

In the above experiments, the test loss curves obtained by the proposed L-CV-DeepLabv3+ for three datasets are shown in Fig. 14. They are all convergent, which verifies the effectiveness of the proposed network.

From the above experimental results about three datasets, we can find some similarities among them. The similarities include that the proposed L-CV-DeepLabv3+ can obtain the

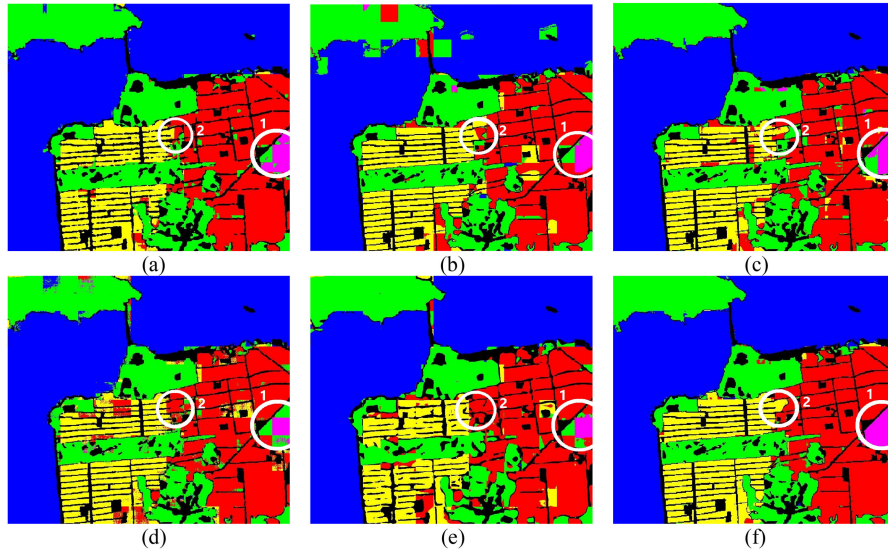


Fig. 12. Segmentation results of San Francisco dataset. (a) FCN. (b) U-Net. (c) SegNet. (d) PSPNet. (e) DeepLabv3+. (f) L-CV-DeepLabv3+.

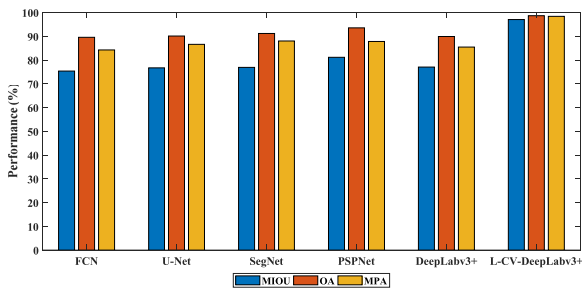


Fig. 13. Segmentation performance of San Francisco dataset.

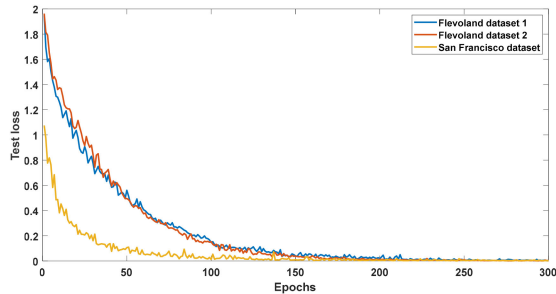


Fig. 14. Test loss curves obtained by L-CV-DeepLabv3+ for three datasets.

best segmentation performance among all the networks, and the OA is larger than the MIOU and MPA for each network. At the same time, there is one difference between the results of San Francisco dataset and two Flevoland datasets. The MIOU and MPA obtained by five RV networks from San Francisco dataset are higher than those from two Flevoland datasets in general. Besides, among all the RV networks, the network which can obtain the best segmentation performance is also different for three datasets. Based on these similarities and differences, we can draw some conclusions as follows.

- 1) The phase information of input data plays a very important role in the semantic segmentation of PolSAR images, which can greatly improve the segmentation performance.
- 2) The design of the structure and parameters of networks should be combined with the size of datasets. For some small datasets, deep networks may not achieve good segmentation results.
- 3) Generally, the more samples the dataset has, the better the segmentation performance is.
- 4) If there is one class with a small number of samples and the segmentation performance is poor by using one network, then the OA obtained by this network is always higher than the MIOU and MPA.

V. CONCLUSION

A lightweight L-CV-DeepLabv3+ was proposed for semantic segmentation of PolSAR image in this article. The structures of backbone network, CV-ASPP, and decoder were presented in detail. They were simplified based on the original DeepLabv3+, and all the operations involved in L-CV-DeepLabv3+ were mathematically strict. Since the proposed network was CV, the CV input data was also introduced. In addition, considering that three PolSAR datasets are very small, the data pre-processing including data expansion was also given. Finally, semantic segmentation experiments were implemented on three PolSAR datasets. Experimental results about the selection of structure and parameters of the backbone network show that the appropriate structure and parameters can effectively improve the segmentation performance. Experimental results about the semantic segmentation of three datasets show that the proposed lightweight network can avoid overfitting, and the phase information of PolSAR data can be very helpful in improving the segmentation performance. Because the structure and parameters of the proposed network are obtained by some experiments, it takes much time and may not achieve the best segmentation

performance. In future, we will use the searching strategy to obtain the optimal structure and parameters of the CV network.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments to improve the quality of this article.

REFERENCES

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

[2] 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.

[3] J. J. Van Zyl, "Unsupervised classification of scattering behavior using radar polarimetry data," *IEEE Trans. Geosci. Remote Sens.*, vol. 27, no. 1, pp. 36–45, Jan. 1989.

[4] 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.

[5] 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.

[6] A. Dargahi, Y. Maghsoudi, and A. A. Abkar, "Supervised classification of polarimetric SAR imagery using temporal and contextual information," in *Proc. Int. Arch. Photogram. Remote Sens. Spatial Inf. Sci.*, Oct. 2013, pp. 107–110.

[7] A. Freeman and S. L. Durden, "A three-component scattering model for polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 963–973, May 1998.

[8] 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.

[9] S. Fukuda and H. Hirose, "Support vector machine classification of land cover: Application to polarimetric SAR data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 1, Jul. 2001, pp. 187–189.

[10] L. Zhang, B. Zou, J. Zhang, and Y. Zhang, "Classification of polarimetric SAR image based on support vector machine using multiple-component scattering model and texture features," *EURASIP J. Adv. Signal Process.*, vol. 2010, pp. 1–9, May 2009.

[11] Q. Yin, W. Hong, F. Zhang, and E. Pottier, "Optimal combination of polarimetric features for vegetation classification in PolSAR image," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 10, pp. 3919–3931, Oct. 2019.

[12] H. Li, H. Gu, Y. Han, and J. Yang, "Object-oriented classification of polarimetric SAR imagery based on statistical region merging and support vector machine," in *Proc. Int. Workshop Earth Observ. Remote Sens. Appl.*, Beijing, China, Sep. 2008, pp. 147–152.

[13] K. S. Chen, W. Huang, D. Tsay, and F. Amar, "Classification of multifrequency polarimetric SAR imagery using a dynamic learning neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 3, pp. 814–820, May 1996.

[14] M. Hellmann, G. Jager, E. Kratzschmar, and M. Habermeyer, "Classification of full polarimetric SAR-data using artificial neural networks and fuzzy algorithms," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 1999, vol. 4, pp. 1995–1997.

[15] C. T. Chen, K. S. Chen, and J. S. Lee, "The use of fully polarimetric information for the fuzzy neural classification of SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 2089–2100, Sep. 2003.

[16] A. Richardson, D. G. Goodenough, H. Chen, B. Moa, G. Hobart, and W. Myrvold, "Unsupervised nonparametric classification of polarimetric SAR data using the k-nearest neighbor graph," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Honolulu, HI, USA, Jul. 2010, pp. 1867–1870.

[17] P. Mishra and D. Singh, "A statistical-measure-based adaptive land cover classification algorithm by efficient utilization of polarimetric SAR observables," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2889–2900, May 2014.

[18] X. She, J. Yang, and W. Zhang, "The boosting algorithm with application to polarimetric SAR image classification," in *Proc. 1st Asian Pacific Conf. Synth. Aperture Radar*, Huangshan, China, Jan. 2007, pp. 779–783.

[19] R. Hansch, "Complex-valued multi-layer perceptrons—An application to polarimetric SAR data," *Photogram. Eng. Remote Sens.*, vol. 76, no. 9, pp. 1081–1088, Sep. 2010.

[20] L. Wang, X. Xu, H. Dong, R. Gui, R. Yang, and F. Pu, "Exploring convolutional LSTM for PolSAR image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Valencia, Spain, Jul. 2018, pp. 8452–8455.

[21] F. Liu, L. Jiao, B. Hou, and S. Yang, "POL-SAR image classification based on Wishart DBN and local spatial information," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3292–3308, Jun. 2016.

[22] L. Zhang, W. Ma, and D. Zhang, "Stacked sparse autoencoder in PolSAR data classification using local spatial information," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1359–1363, Sep. 2016.

[23] B. Hou, H. Kou, and L. Jiao, "Classification of polarimetric SAR images using multilayer autoencoders and superpixels," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 3072–3081, Jul. 2016.

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

[25] 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.

[26] S. W. Chen and C. S. Tao, "PolSAR image classification using polarimetric-feature-driven deep convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 627–631, Apr. 2018.

[27] F. Gao, T. Huang, J. Wang, J. Sun, A. Hussain, and E. Yang, "Dual-branch deep convolution neural network for polarimetric SAR image classification," *Appl. Sci.*, vol. 7, no. 5, Apr. 2017, Art. no. 447.

[28] H. Dong, L. Zhang, and B. Zou, "PolSAR image classification with lightweight 3D convolutional networks," *Remote Sens.*, vol. 12, no. 3, Jan. 2020, Art. no. 396.

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

[30] P. Zhang *et al.*, "PolSAR image classification using hybrid conditional random fields model based on complex-valued 3-D CNN," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 3, pp. 1713–1730, Jun. 2021.

[31] H. Bi, J. Sun, and Z. Xu, "A graph-based semisupervised deep learning model for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2116–2132, Apr. 2019.

[32] 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.

[33] L. Jiao and F. Liu, "Wishart deep stacking network for fast POLSAR image classification," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3273–3286, Jul. 2016.

[34] W. Xie *et al.*, "PolSAR image classification via Wishart-AE model or Wishart-CAE model," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3604–3615, Aug. 2017.

[35] H. Liu, S. Yang, S. Gou, D. Zhu, R. Wang, and L. Jiao, "Polarimetric SAR feature extraction with neighborhood preservation-based deep learning," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1456–1466, Apr. 2017.

[36] L. Zhang, H. Dong, and B. Zou, "Efficiently utilizing complex-valued PolSAR image data via a multi-task deep learning framework," *ISPRS J. Photogramm.*, vol. 157, no. 8, pp. 59–72, Sep. 2019.

[37] D. Xiang *et al.*, "Adaptive statistical superpixel merging with edge penalty for PolSAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2412–2429, Apr. 2020.

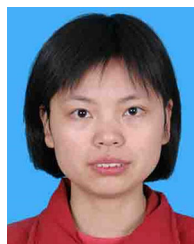
[38] S. Quan, D. Xiang, W. Wang, B. Xiong, and G. Kuang, "Scattering feature-driven superpixel segmentation for polarimetric SAR images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2173–2183, Jan. 2021.

[39] H. Gao, C. Wang, D. Xiang, J. Ye, and G. Wang, "TSPol-ASLIC: Adaptive superpixel generation with local iterative clustering for time-series quad- and dual-polarization SAR data," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2021.3126669](https://doi.org/10.1109/TGRS.2021.3126669).

[40] J. Cheng, F. Zhang, D. Xiang, Q. Yin, and Y. Zhou, "PolSAR image classification with multiscale superpixel-based graph convolutional network," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–14, doi: [10.1109/TGRS.2021.3079438](https://doi.org/10.1109/TGRS.2021.3079438).

[41] C. Henry, S. M. Azimi, and N. Merkle, "Road segmentation in SAR satellite images with deep fully convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1867–1871, Dec. 2018.

- [42] D. Cantorna, C. Dafonte, A. Iglesias, and B. Arcay, "Oil spill segmentation in SAR images using convolutional neural networks. A comparative analysis with clustering and logistic regression algorithms," *Appl. Soft Comput.*, vol. 84, Nov. 2019, Art. no. 105716.
- [43] M. Shahzad, M. Maurer, F. Fraundorfer, Y. Wang, and X. X. Zhu, "Buildings detection in VHR SAR images using fully convolution neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1100–1116, Feb. 2019.
- [44] J. Chen, X. Qiu, C. Ding, and Y. Wu, "CVCMMF net complex-valued convolutional and multifeature fusion network for building semantic segmentation of InSAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Apr. 2021, Art. no. 5205714.
- [45] J. Jing, X. Sun, Z. Wang, K. Chen, W. Diao, and K. Fu, "Fine building segmentation in high-resolution SAR images via selective pyramid dilated network," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6608–6623, Apr. 2021.
- [46] J. Xia, N. Yokoya, B. Adriano, L. Zhang, G. Li, and Z. Wang, "A benchmark high-resolution Gaofen-3 SAR dataset for building semantic segmentation," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5950–5963, May 2021.
- [47] Y. Duan, X. Tao, C. Han, X. Qin, and J. Lu, "Multi-scale convolutional neural network for SAR image semantic segmentation," in *Proc. IEEE Glob. Commun. Conf.*, Dec. 2018, pp. 1–6.
- [48] Y. Wang, C. He, X. Liu, and M. Liao, "A hierarchical fully convolutional network integrated with sparse and low-rank subspace representations for PolSAR imagery classification," *Remote Sens.*, vol. 10, no. 2, Feb. 2018, Art. no. 342.
- [49] Y. Li, Y. Chen, G. Liu, and L. Jiao, "A novel deep fully convolutional network for PolSAR image classification," *Remote Sens.*, vol. 10, no. 12, Dec. 2018, Art. no. 1984.
- [50] C. He, M. Tu, D. Xiong, and M. Liao, "Nonlinear manifold learning integrated with fully convolutional networks for PolSAR image classification," *Remote Sens.*, vol. 12, no. 4, Feb. 2020, Art. no. 655.
- [51] F. Zhao, M. Tian, W. Xie, and H. Liu, "A new parallel dual-channel fully convolutional network via semisupervised FCM for PolSAR image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4493–4505, Aug. 2020.
- [52] H. Jing, Z. Wang, X. Sun, D. Xiao, and K. Fu, "PSRN: Polarimetric space reconstruction network for PolSAR image semantic segmentation," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10716–10732, Sep. 2021.
- [53] J. Chen, L. Peng, X. Qiu, C. Ding, and Y. Wu, "A 3D building reconstruction method for SAR images based on deep neural network," *Sci. Sin. Informat.*, vol. 49, no. 12, pp. 1606–1625, Dec. 2019.
- [54] Y. Cao, Y. Wu, P. Zhang, W. Liang, and M. Li, "Pixel-wise PolSAR image classification via a novel complex-valued deep fully convolutional network," *Remote Sens.*, vol. 11, no. 22, Nov. 2019, Art. no. 2653.
- [55] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comp. Vis.*, Munich, Germany, Sep. 2018, pp. 801–818.
- [56] L. Yu, Y. Hu, X. Xie, Y. Lin, and W. Hong, "Complex-valued full convolutional neural network for SAR target classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1752–1756, Oct. 2020.
- [57] N. Guberman, "On complex valued convolutional neural networks," Feb. 2016, [Online]. Available: <https://arxiv.org/pdf/1602.09046.pdf>
- [58] M. A. Nielsen, *Neural Networks and Deep Learning*. San Francisco, CA, USA: Determination Press USA, 2015.
- [59] C. Trabelsi *et al.*, "Deep complex networks," May 2017, [Online]. Available: <https://arxiv.org/abs/1705.09792>
- [60] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Santiago, Chile, 2015, pp. 3431–3440.
- [61] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Munich, Germany, Oct. 2015, pp. 234–241.
- [62] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [63] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Venice, Italy, Jul. 2017, pp. 2881–2890.



Lingjuan Yu received the M.S. degree in signal and information processing from the South China University of Technology, Guangzhou, China, in 2009, and the Ph.D. degree in electromagnetic field and microwave technology from the National Space Science Center, Chinese Academy of Sciences, Beijing, China, in 2012.

From 2015 to 2019, she was a Postdoctoral Fellow with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. From 2020 to 2021, she was a Visiting Fellow with the School of Information Science and Technology, Fudan University, Shanghai, China. She is currently an Associate Professor with the School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, China. Her research interests include synthetic aperture radar data processing and application.



Zhaoxin Zeng received the B.E. degree in communication engineering in 2018 from the Jiangxi University of Science and Technology, Ganzhou, China, where he is currently working toward the M.S. degree with the School of Information Engineering.

His research interests include the application of machine learning and deep learning algorithms to synthetic aperture radar.



Ao Liu received the B.E. degree in electronic and information engineering from Hubei Engineering University, Xiaogan, China, in 2019. He is currently working toward the M.S. degree with the School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, China.

His research interests include the application of machine learning and deep learning algorithms to synthetic aperture radar.



Xiaochun Xie received the M.S. degree in signal and information processing from the Huazhong University of Science and Technology, Wuhan, China, in 2003, and the Ph.D. degree in computer application technology from the National Space Science Center, Chinese Academy of Sciences, Beijing, China, in 2010.

He is currently an Associate Professor with the School of Physics and Electronic Information, Gannan Normal University, Ganzhou, China. His research interests include synthetic aperture radar data processing and application.



Haipeng Wang (Senior Member, IEEE) received the B.S. and M.S. degrees in mechanical and electronic engineering from the Harbin Institute of Technology, Harbin, China, in 2001 and 2003, respectively, and the Ph.D. degree in environmental systems engineering from the Kochi University of Technology, Kochi, Japan, in 2006.

He was a Visiting Researcher with the Graduate School of Information, Production and Systems, Waseda University, Fukuoka, Japan, in 2008. He is currently a Professor with the Key Laboratory of Electromagnetic Wave Information Science (MoE), Department of Communication Science and Engineering, School of Information Science and Engineering, Fudan University, Shanghai, China. His research interests include signal processing, SAR image processing and analysis, speckle statistics, and applications to forestry and oceanography, and machine learning and its applications to SAR images.

Dr. Wang is a member of technical program committee of IEEE Geoscience and Remote Sensing Symposium (IGARSS) since 2011. He was the recipient of the Dean Prize of the School of Information Science and Engineering, Fudan University in 2009 and 2017, respectively.



Feng Xu (Senior Member, IEEE) received the B.E. degree (Hons.) in information engineering from Southeast University, Nanjing, China, in 2003, and the Ph.D. degree (Hons.) in electronic engineering from Fudan University, Shanghai, China, in 2008.

From 2008 to 2010, he was a Postdoctoral Fellow with the NOAA Center for Satellite Application and Research, Camp Springs, MD, USA. From 2010 to 2013, he was with Intelligent Automation Inc., Rockville, MD, USA, while partly involved as a Research Scientist with the NASA Goddard Space Flight

Center, Greenbelt, MD, USA. In 2012, he was selected into the China's Global Experts Recruitment Program, and in 2013 subsequently returned to Fudan University, where he is currently a Professor with the School of Information Science and Technology and the Vice Director of the Key Laboratory for Information Science of Electromagnetic Waves (MoE). He has authored more than 40 papers in peer-reviewed journals, coauthored three books, and holds two patents, among many conference papers. His research interests include fast electromagnetic modeling for complicated target and environments, intelligent interpretation of synthetic aperture radar (SAR) images, inverse scattering tomography, and SAR remote-sensing applications in earth observation.

Dr. Xu was awarded the second-class National Nature Science Award of China, among other honors, in 2011. He was a recipient of the Early Career Award of the IEEE Geoscience and Remote Sensing Society in 2014 and the SUMMA Graduate Fellowship in the advanced electromagnetics area in 2007. He is currently an Associate Editor for *IEEE Geoscience and Remote Sensing Letters*. He is the Founding Chair of the IEEE GRSS Shanghai Chapter.



Wen Hong (Senior Member, IEEE) received the M.S. degree in electronic engineering from Northwestern Polytechnical University, Xi'an, China, in 1993, and the Ph.D. degree from Beihang University, Beijing, China, in 1997.

From 1997 to 2002, she was a Faculty Member in signal and information processing with the Department of Electrical Engineering, Beihang University. In between, she was a Guest Scientist with DLR-HF, Wessling, Germany, from 1998 to 1999, for one year.

Since 2002, she has been a Scientist with the Science and Technology on Microwave Imaging Laboratory and a Supervisor of the Graduate Student Program with the Chinese Academy of Sciences, Beijing. Her main research interests include polarimetric/polarimetric interferometric synthetic aperture radar (SAR) data processing and application, 3-D SAR signal processing, circular SAR signal processing, SAR polarimetry application, and sparse microwave imaging with compressed sensing.