Random Region Matting for the High-Resolution PolSAR Image Semantic Segmentation

Jun Ni, *Student Member, IEEE*, Fan Zhang, *Senior Member, IEEE*, Fei Ma, Qiang Yin, *Senior Member, IEEE*, and Deliang Xiang, *Member, IEEE*

***Abstract*—Polarimetric synthetic aperture radar (PolSAR) imagery can provide more intuitive and detailed SAR polarization information, and it is widely used in the classification and semantic segmentation of remote sensing. To bridge the PolSAR data and application, the 2020 Gaofen Challenge on Automated High-Resolution Earth Observation Image Interpretation provides a set of high-quality PolSAR semantic segmentation dataset. A series of preprocessing methods is first used to analyze the PolSAR images to improve the semantic segmentation performance of the PolSAR imagery. A special polarimetric decomposition method is used to extract the features, and the filter and the data truncation are implemented to enhance local and global information of images. And the random region matting method is proposed to expand the training samples. Finally, the DeepLabV3+ method with the ResNet101-V2 is employed to achieve the semantic segmentation. A variety of comparison experiments verifies the effectiveness of our methods. Simultaneously, compared with the classification methods of other groups in the competition, our methods have obvious advantages in the inference time and semantic segmentation accuracy. The proposed method achieved a frequency weighted intersection over union of 75.29% in the contest.**

***Index Terms*—Data augmentation, DeepLabV3+, Gaofen-3, image classification, polarimetric synthetic aperture radar (PolSAR), semantic segmentation.**

I. Introduction

OVER the last decade, the synthetic aperture radar (SAR) imagery has been widely used in the geographical survey, climate change research, and other applications due to its high resolution, day-and-night, and weather-independence. Polarization characteristics can reflect targets' physical properties from different information provided by the magnitude, frequency, and phase. Thus, it is a crucial research part of the image analysis and interpretation of SAR [1]–[3]. As a fundamental step in polarimetric SAR (PolSAR) image processing and applications, the semantic segmentation of PolSAR imagery can realize segmentation and categorization simultaneously, thus obtains smooth and fine-grained classification to assign one land cover category to each pixel [4]–[6]. It is beneficial in disaster monitoring, land cover classification, water identification, and so on. How to accurately segment PolSAR imagery to solve practical problems has become a hot topic, as well as one of the main topics in the Gaofen Challenge Contest.

In earlier PolSAR image segmentation algorithm, some methods have been proposed for PolSAR imagery [7], such as the Markov random field [8], [9], the conditional random field [10], [11]. However, it is still challenging to get a reliable and consistent scene semantic segmentation. With the continuous development of deep learning networks and hardware devices, the precision and the computing efficiency of semantic segmentation have made significant progress [12]. The convolutional neural network (CNN) is especially good at learning features from raw data automatically. It has pushed the performance of semantic segmentation tasks to a soaring height on natural images [13], [14]. Based on two publicly available optical image datasets: Visual Object Classes Challenge 2012 (VOC2012) and Microsoft Common Objects in Context (MSCOCO), some semantic segmentation methods have been proposed, including the fully convolutional network (FCN) [15], U-net [16], SegNet [17], Dilated Convolutions [18], RefineNet [19], PSPNet [18], mask R-CNN [20], and their combination methods [21]. Due to the similarity of semantic segmentation between natural images and PolSAR images, these semantic segmentation methods are also suitable for PolSAR [22] images. Although some methods, such as FCN and Atrous-ResNet50, have been implemented in SAR image semantic segmentation [23]–[28], their methods are subject to the low-resolution dataset, the ignorance of polarization information, and imperfect selection of methods, resulting in unsatisfactory classification results.

Generally, the refined polarization scattering mechanism model and interpretation can extract features with intense discrimination for different feature categories from the fully polarized information. Given the mature polarization decomposition methods, the classification performance of the classifier with the polarization scattering mechanism model is superior. The effective use of the scattering mechanism and statistical

Manuscript received January 5, 2021; revised February 15, 2021; accepted February 19, 2021. Date of publication February 26, 2021; date of current version March 22, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61871413, Grant 41801236, and Grant 61801015, and in part by the Fundamental Research Funds for the Central Universities under Grant XK2020-03. *(Corresponding author: Fei Ma.)*

Jun Ni, Fei Ma, and Qiang Yin are with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China (e-mail: nijunbuct@163.com,; mafei@mail.buct.edu.cn; yinq@mail.buct.edu.cn,).

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

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

Digital Object Identifier 10.1109/JSTARS.2021.3062447

characteristics to preprocess the PolSAR images can improve the training efficiency of the neural network and enhance the generalization ability of the neural network. Since the GaoFen-3 high-resolution PolSAR imagery disclosed in the Gaofen Challenge Contest has not provided the phase information, the conventional PolSAR decomposition methods cannot be directly used to process the PolSAR imagery. Two kinds of pseudocolor images are synthesized to efficiently extract polarimetric features of the PolSAR images, so as to facilitate human observation of ground objects and improve the classification performance of the semantic segmentation network.

Nonetheless, two factors also affect the semantic segmentation performance of the PolSAR imagery: a large amount of speckle noise and the limited number of training samples. The former is formed by the coherent imaging mechanism of the PolSAR systems. Although the existing speckle suppression methods, such as the Refined-Lee filter, are widely implemented in PolSAR imagery, they are not the best in terms of the particular contest data and the processing efficiency. Although the semantic segmentation network has a certain ability to resist noise interference, the outlier caused by the multiplicative noise and the strong backscatter will vastly impact the classification performance of the neural network, and this negative effect is independent of the quality of the segmentation method. In this article, the statistical characteristics of the high-resolution PolSAR datasets have been analyzed to address this issue, and then some inessential extreme values are discarded to simplify the PolSAR imagery according to the statistical histogram to enhance the image representation, thereby improving the classification performance of the semantic segmentation method. In order to ensure the smoothness of the image and save more backscattering information, a simple filter is designed to reduce the strong scattering and speckle noise while ensuring the integrity of the weak scattering scene. The median filter is implemented to minimize the outlier noise, and a conditional constraint is designed to provide the complete information of surface scatting, e.g., waters and road, and unrecognized region in the PolSAR image.

Aiming at the limitation of training samples, the existing sample enhancement methods, such as rotation, flip, translation, color change, and so on, can improve the robustness of the semantic segmentation network. Still, these algorithms are only limited to transform the known image scene instead of constructing the unknown scene, thus restricting the classification performance. Inspired by the random neighbor pixel-block (RNPB) method [29] and background matting [30], an intelligent PolSAR image synthesis method, namely, random region matting (RRM), is devised to enrich the training samples. First, different categories can be cut into different components according to their labels in the training samples. Then, one PolSAR image can be randomly selected as the background image, and the category components are randomly selected from other images to replace the corresponding pixels of the background image, so as to build new training samples. Simultaneously, considering the scattering characteristics and resolution of different images, the components are adjusted according to the statistical information of the background image, and the final combination image can be used for the training samples. Moreover, the stitching of slices from different images will result in obvious boundaries between scenes in the combination image, which may weaken the training effect of the expanded sample on the neural network. Therefore, the previous conditional filter is also implemented to ensure the smoothing of the synthetic image.

However, an appropriate semantic segmentation network is essential to gain ideal classification results. Although multichannel networks can effectively improve the performance of image segmentation, the complex network structure reduces the reasoning efficiency of the network. Especially, the segmentation efficiency accounts for 20% of the score in the competition, and the limited resources restrict the execution of these complex networks, while the performance of a single-channel network is usually affected by the test sample with unfixed resolution. Dilated convolutions can effectively segment for images with different resolutions [31], and DeepLabV3+ designed by the multiscale feature is suitable for the unfixed resolution PolSAR image. Therefore, ResNet101-V2 is selected as the backbone of our DeepLabV3+ model. The main contributions of our work are as follows.

- The RRM is proposed to enhance the segmentation performance of the semantic segmentation network. The RRM aims to build the possibility of unseen scenarios and enhance the correlation among training samples.
- The statistical distribution of the PolSAR image in the contest is analyzed. According to its influence on the performance of the segmentation network, a data expansion method is given to improve the segmentation accuracy.
- Two pseudocolor images are synthesized to simplify the four-channel PolSAR images, so as to increase the speed of network reasoning and reduce computer resource usage. In addition, a simple filter is designed to minimize the strong scattering and speckle noise, and smooth the combination images of the RRM method.

The rest of this article is organized as following. The data processing and analysis, the RRS method, and the semantic segmentation method are presented in Section II; the comparative experiments are shown in Section III; conclusions and discussions are included in Section IV.

II. Methodology

As mentioned in Section I, a single-channel semantic segmentation method, namely DeepLabV3+, is developed to classify the high-resolution PolSAR images subtly, and a series of improved processes are implemented in data expansion and preprocessing to enhance the performance of the segmentation method. The overall classification framework is shown in Fig. 1.

First, based on the RRM method, each training samples can be expended into difference scenarios with slices from each category. According to the proportion of the image occupied by each slice, each slice is given a probability of being selected, and the small target has a higher probability to ensure the balance of training samples. Next, according to the statistical characteristics of each polarimetric channel, the PolSAR images are filtered, enhanced, and normalized, respectively. Then, the polarization

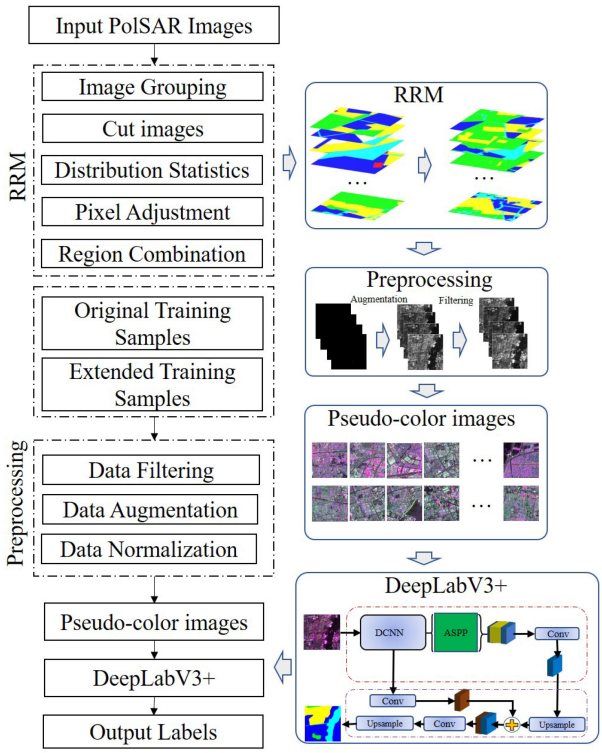

Fig. 1. Overall classification framework in this article.

decomposition method is used to extract the polarimetric feature and compose the pseudocolor map. Finally, based on the pseudocolor maps, the DeepLabV3+ is implemented to realize the fined classification of the PolSAR image.

Four groups of the polarimetric SAR images are randomly selected to illustrate the algorithm, and the labels and images are shown in Fig. 2. Due to the influence of strong backscatter data and speckles, the raw image cannot display the essential attributes of the samples, even if a very small amount of scattering information can be noted by artificial observation. After histogram equalization, the overall information is displayed, but most of the images are still fuzzy, as shown in Fig. 2(b)–(d). The appearance indicates that most pixel values are concentrated in one suitable interval. The strong backscattering weaken most of the reasonable pixels, leading to excessively enhance some pixels in the histogram equalization results. Similarly, these images cannot be directly normalized by the usual methods (e.g., all pixels divided by the largest pixel value) to the segmentation method. Otherwise, the effective pixel will be compressed to a minimal value, affecting the segmentation result. This negative effect has nothing to do with the performance of the segmentation network.

A. PolSAR Data Representations

The representation of PolSAR image is the basis of classification. A PolSAR image contains four channels of information, namely, HH, HV, VH, and VV, which represent four linear orthogonal polarization combinations. The complex backscattering matrix $[S]$ for each image pixel can be expressed as

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

where H and V indicate the horizontal and vertical polarization channels, respectively. Under the monostatic backscattering case, there is a reciprocity theorem, i.e., $S_{VH} = S_{HV}$. Therefore, the polarimetric information can be interpreted by the coherent matrix $[T]$ and covariance matrix $[C]$

$$[T] = \langle \vec{k} \cdot \vec{k}^{*T} \rangle \tag{2}$$

$$[C] = \langle \vec{\Omega} \cdot \vec{\Omega}^{*T} \rangle \tag{3}$$

where $\langle\dot{}\rangle$ denotes the ensemble average, $\dot{}^{*T}$ indicates the complex conjugation and transpose, $\vec{k} = \frac{\sqrt{2}}{2}[\, S_{HH} + S_{VV} \;\; S_{HH} - S_{VV} \;\; 2S_{HV} \,]^T$ is the Pauli-based scattering vector, and $\vec{\Omega} = [\, S_{HH} \;\; \sqrt{2}S_{HV} \;\; S_{VV} \,]^T$ is the lexicographic scattering vector. Accordingly, the coherency matrix $[T]$ and covariance matrix $[C]$ are 3×3 in the monostatic backscattering case, i.e., (4) and (5)

$$C_3 = \begin{bmatrix} \langle |S_{HH}|^2 \rangle & \sqrt{2}\langle S_{HH} S_{HV}^* \rangle & \langle S_{HH} S_{VV}^* \rangle \\ \sqrt{2}\langle S_{HV} S_{HH}^* \rangle & 2\langle |S_{HV}|^2 \rangle & \sqrt{2}\langle S_{HV} S_{VV}^* \rangle \\ \langle S_{VV} S_{HH}^* \rangle & \sqrt{2}\langle S_{VV} S_{HV}^* \rangle & \langle |S_{VV}|^2 \rangle \end{bmatrix}. \tag{5}$$

Due to the lack of phase information and the consideration of computational efficiency, it is difficult to proceed with more hierarchical decomposition models. The pseudocolor images can be represented with $|S_{HH} + S_{VV}|^2$, $|S_{HH} - S_{VV}|^2$ and $4|S_{HV}|^2$ in the coherent matrix $[T]$, and can also be represented with $|S_{HH}|^2$, $2|S_{HV}|^2$ and $|S_{VV}|^2$ in the covariance matrix $[C]$. The pseudocolor images can reduce the amount of data processed by the network, thus improving the processing efficiency of the network. Correspondingly, the effective data representation can improve the training efficiency and robustness of the upgraded network.

B. High-Resolution PolSAR Scattering Statistics

Inevitably, the speckle noise of the PolSAR image is caused by the coherent nature of the scattering phenomena. Even though the speckle noise carries itself information about the illuminated area, it degrades the appearance of images and affects the performance of scene analysis tasks carried out by computer programs (e.g., segmentation and classification) [32]. However, the intense speckle prevents the use of the PolSAR images in the automatic feature extraction. Even though this problem can be countered by resorting to some forms of the multilook technique, it is accompanied by the remarkable side effect of losing spatial resolution. Meanwhile, the complex classification algorithm reduces the efficiency of the semantic segmentation algorithm. Considering the antinoise ability of the neural network, it is more practical to find the problem that affects the classification accuracy than to design a complex filtering algorithm.

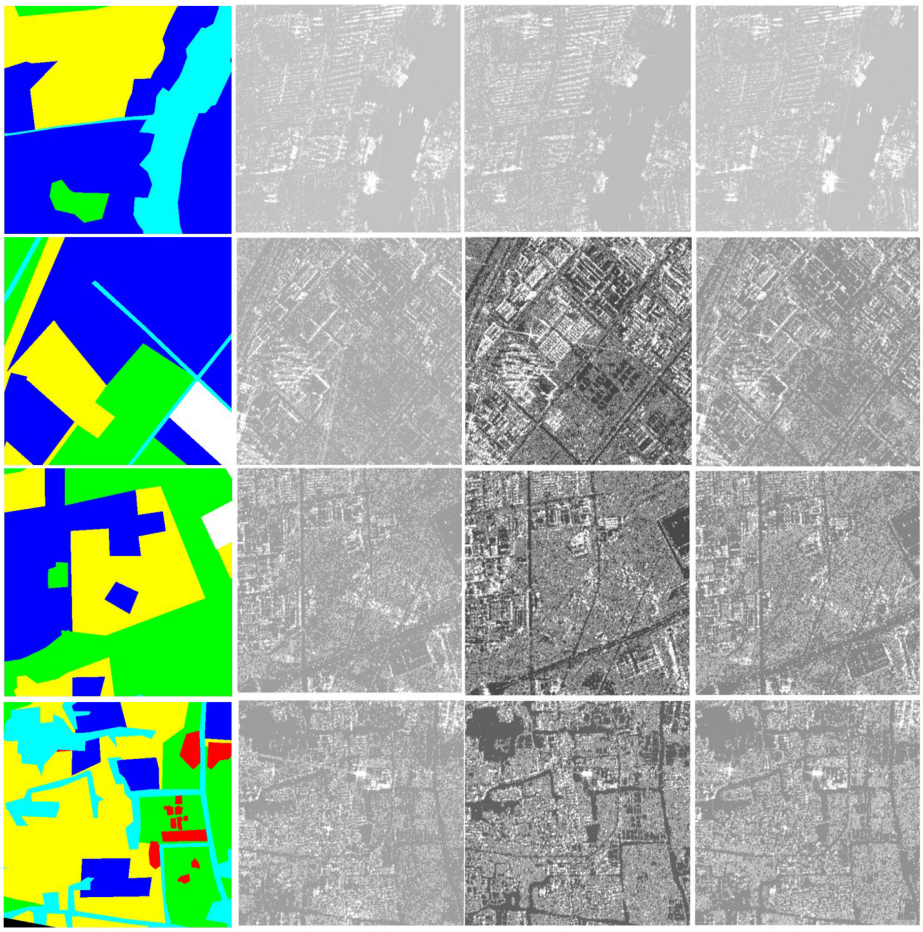

Fig. 2. Labels and images of the validation datasets. (a) Labels; (b) HH; (c) HV; and (d) VV.

Under the hypothesis of fully developed speckle, the observed backscattered signal $z(s)^3$ can be expressed as

$$z(s) = x(s)u(s) \tag{6}$$

where $x(s)$ is the noise-free reflectance and $u(s)$ is the speckle in intensity format. In the case of multiplicative noise, $z(s)$ will differ several times or even hundreds of times from $x(s)$.

Simultaneously, as the basis of the backbone network, CNN has been proved to extract the advanced features in images, thus realizing effective classification and segmentation. Irreplaceably, the convolution process is considered as the key of the network, which can be expressed as

$$y_i = \sum \mathbf{W}_{i,j} * \mathbf{z}'_{i,j} + b_i \tag{7}$$

where y_i is the output of the convolution layer corresponding to the ith feature map, $*$ is the convolution processing, $\mathbf{W}_{i,j}$ is the weight of the convolution kernel, b_i is the bias, and $\mathbf{z}'_{i,j}$ is the input block of convolution layer corresponding to $\mathbf{W}_{i,j}$. If there are speckles in $\mathbf{z}'_{i,j}$ in the input layer, the speckle causes a difference of several hundred times between other pixels in $\mathbf{z}'_{i,j}$, and the corresponding convolution weight is not 0, then the result of $W_{i,j} * x_{i,j}$ will be seriously distorted, thus weakening the network with a lot of speckles in $z(s)$. In addition to the speckle noise, a small number of the strong backscattering can also cause this negative situation, and it is necessary to find an effective numerical interval by the scattering statistics. For the four PolSAR images with the data type of the 16-b unsigned integer (UINT16), the statistical results of each channel are shown in Fig. 3.

Obviously, although the pixel value can ranges into the tens of thousands, more than 90% of the pixels are in the [0, 511]. In particular, in HH-channel and VV-channel, 91.05% and 89.11% of the pixels are less than 256, respectively. The statistical results means that the 99.6% pixel interval in [0, 65 535] is essentially

$$T_3 = \frac{1}{2}\begin{bmatrix} \langle |S_{HH}+S_{VV}|^2 \rangle & \langle (S_{HH}+S_{VV})(S_{HH}-S_{VV})^* \rangle & 2\langle (S_{HH}+S_{VV})S_{HV}^* \rangle \\ \langle (S_{HH}-S_{VV})(S_{HH}+S_{VV})^* \rangle & \langle |S_{HH}-S_{VV}|^2 \rangle & 2\langle (S_{HH}-S_{VV})S_{HV}^* \rangle \\ 2\langle S_{HV}(S_{HH}+S_{VV})^* \rangle & 2\langle S_{HV}(S_{HH}-S_{VV})^* \rangle & 4\langle |S_{HV}|^2 \rangle \end{bmatrix} \tag{4}$$

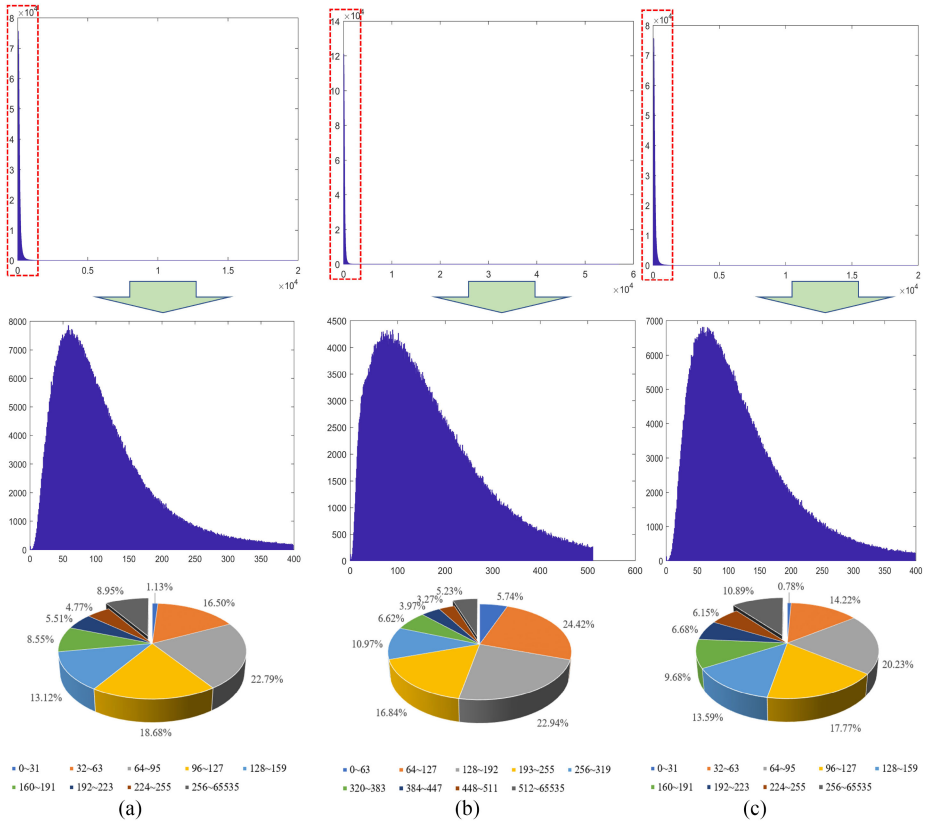

Fig. 3. Pixel statistics of the PolSAR images in different channels. (a) HH; (b) HV; and (c) VV.

invalid, resulting in low segmentation accuracy. To enhance the data without reducing the processing efficiency, a hard threshold N is set to deal with each pixel value x_i, that is

$$p(x_i) = \begin{cases} x_i & x_i \le N \\ N & x_i > N \end{cases} \tag{8}$$

where $p(x_i)$ is the pixel value after the processing, N is set to 255, 511, and 255 in HH-, HV-, and VV-channel. To represent more data and reduce the interference of weak of weak speckle noise, a sample conditional filter is designed as

$$f(x_i) = \begin{cases} filter(x_i) & x_i > M \\ M & x_i \le M \end{cases} \tag{9}$$

where $f(x_i)$ represents the pixel value after filtering, $filter(\cdot)$ is a filter, it is designed with a median filter in this article. To conserve the integrity of the surface scattering information in the image, such as rivers, roads and unknown areas, pixels with a value less than M are not filtered.

In view of the antinoise capability of the backbone network in the semantic segmentation network, the filter is not to eliminate the speckle completely, but to further expand the representation range of effective data, and save the image texture information as much as possible. Equation (9) and (8) are, respectively, aimed at reducing speckle and strong scattering in local area and global area. After filtering, the statistics of the pixels are shown in Fig. 4. In the three channels, 7.89%, 4.33%, and 9.52% of the pixels are regarded as the strong speckle noise and the strong backscattering, and are processed by (8).

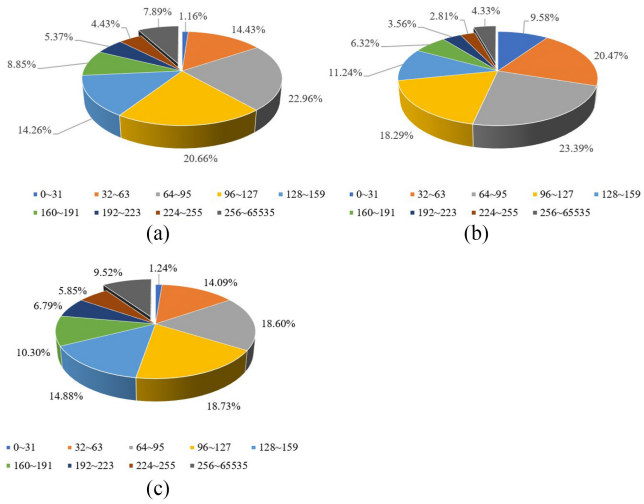

Fig. 4. Pixel statistics after filtering. (a) HH; (b) HV; and (c) VV.

Although the median filtering can change the statistical information of the PolSAR image, which may cause interference to the unsupervised learning task of the PolSAR images, the goal of the CNNs is to perform the high-level semantic interpretation (i.e., feature extraction) through given images, rather than to

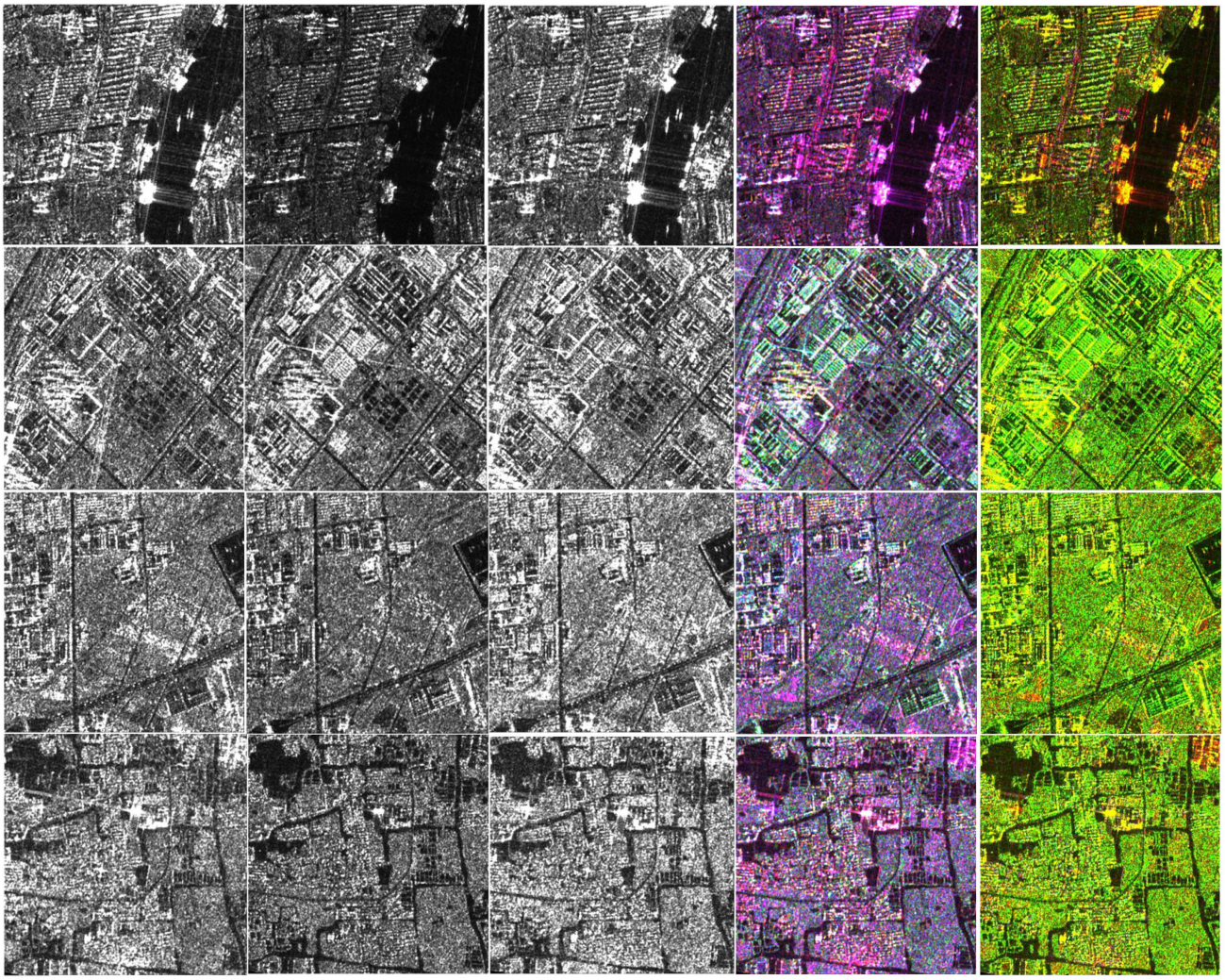

Fig. 5. PolSAR images processed by the algorithm in this article. (a) HH; (b) HV; (c) VV; (d) pseudo-color images ($|C_{11}|$, $|C_{22}|$, $|C_{33}|$); (e) pseudo-color images ($|T_{11}|$, $|T_{22}|$, $|T_{33}|$).

use the experience information directly [33]. Therefore, the negative effect of the changed statistical information caused by the median filtering is basically insignificant for the CNNs. On the contrary, the strong backscattering information will greatly weaken the feature learning and feedback adjustment effect of the effective pixels in the network training process, resulting in the degradation of the performance of the neural network. In this situation, the truncation and enhancement of the training dataset are workable and indispensable. In addition, based on the RRM method in Section II-C, the matting scene will have some boundaries between the background image, and this divergence will weaken the influence of the composite image on the segmentation network, so the filter is also implemented to smooth the combination image to suppress the negative effect.

After these processing, the PolSAR images can be represented in Fig. 5. Compared with the Fig. 2, the image quality of each channel in Fig. 5 is significantly enhanced. Finally, the PolSAR images are processed to produce the two pseudocolor images, and the data type is changed from UINT16 to the 8-b unsigned integer (UINT8). The data memory is compressed to $3/8$ of the original.

C. RRM Method

The deep learning method based on a large number of training samples performs better than the traditional machine learning algorithms in classification, target recognition, and so on. However, the training samples are often difficult to obtain in practical application, thereby limiting the performance of the network. Normally, the sample enhancement methods, such as rotate, flip, transform, etc., can improve the adaptability of the segmentation network to the target, but these methods basically do not change the target scene, resulting in limited performance improvement. The RNBP method has been proven to improve the classification accuracy of the CNN and the long-short term memory under limited samples. Inspired by the background matting method and the RNBP method, the RRM method is proposed to create unknown scenes for different categories.

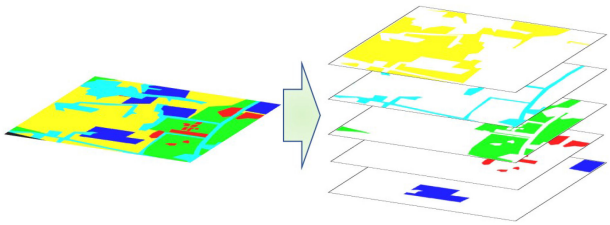

Fig. 6. Scene cutting schematic diagram of the SAR image.

Based on the labels on one training sample, the segmentation range of different scenes in the image can be roughly determined, so the image can be cut into different slices, as shown in Fig. 6. Accordingly, the four-channel PolSAR images can also be cut according to the labels. Then, the slices can be randomly transformed, shifted, rotated, resized and so on, to create different shapes. Based on one background selected from the training samples, the changed slices can be used to form a new training sample, hence the workflow of the RRM method can be shown in Fig. 7. The relationship between the background diagram and slices can be expressed as

$$I'(i,j) = u_0(i,j) \cdot I(i,j) + \sum_{k=1}^{K} u_k(i,j) \cdot f(s_k(\alpha_i, \theta_j)) \tag{10}$$

where $I'(i,j)$ is the combination image, $I(i,j)$ is a background, K is the number of slice, $u_*(i,j)$ is the pixel decision function, which determines whether the pixel is retained or not, $s_k(\cdot)$ is the inlaid slice, and $f(\cdot)$ is the transformation function and used for rotation, shrink, and other transformation processes. $u_k(i,j) \cdot f(s_k(\alpha_i, \theta_j))$ is noted as $S_k(\alpha_i, \theta_j)$. Considering the balance of the sample in the image, each slice is given a probability according to the size of slice, so that ensure the slice with the small target will be selected as the priority. In view of the different scattering information of the PolSAR images, data conversion should be carried out first to reduce the difference when the slices are synthesized into the background, that is

$$I'(i,j) = u_0(i,j) \cdot I(i,j) + \sum_{k=1}^{K} \Phi(S_k(\alpha_i, \theta_j)) \tag{11}$$

where $\Phi(\cdot)$ is the correction function. Notably, limited by computer resources, the image with a large scene is often cut into several small images, and then realize refined classification. Hence, $\Phi(\cdot)$ should be designed from two perspectives. 1) The slice and background are obtained from one large PolSAR image. 2) The slice is completely unrelated to the background image. Based on the pseudocolor images, the training samples are easily divided into different groups, such as Fig. 8.

If the original image of the slice and the background are grouped together, the function $\Phi(\cdot)$ does not need to transform the slice. If the original image of the slice and the background are in different groups, and $\Phi(\cdot)$ should be used to adjust the slice. Suppose that the background is in group-t, based on the Kullback–Leibler (KL) distance, the distance between the the background $I(i,j)$ and other images $I_t(i,j)$ in the group-t can be expressed as

$$dist(P||Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)} \tag{12}$$

where X is a set of pixels, $P(x)$ and $Q(x)$ corresponding to $I(i,j)$ and $I_t(i,j)$ are the distribution probability of X. Then, we can find the image that contains the slice n_k of the same category as the slice $s_k(\cdot)$ and the image is closest to the background. For the slice n_k and the slice $s_k(\cdot)$, their distributions can be calculated, respectively, and the pixels with a concentration of 80% in their distributions will be used to calculated their mean values, respectively. Finally, the pixels of the slice $s_k(\cdot)$ can be adjusted by the mean values.

One composite image is shown in Fig. 9. In this article, K is set to 5. From Fig. 9, the background image is inserted into five slices from five PolSAR images, and the resolution of each image is different. Still, the composite image as a whole has no obvious difference.

D. DeepLabV3+ Method

Since 2012, the image feature extraction methods based on the CNN architecture have been evolving, and greatly revolutionized pattern recognition [34]–[36]. Various structures are established benchmarks and used as backbone in the CNN-based semantic segmentation framework, such as VGGNet [37], ResNet [38], Xception [39], etc. Based on these backbone networks, a variety of semantic segmentation methods in the form of the encoder–decoder type architecture have been proposed. The encoder module is mainly composed of a combination of the backbone network and various convolution modules to obtain the necessary feature information from the image. It will gradually downsample the input image and reduce the resolution of the feature map to capture high-level details of the scene. The decoding module is used to complete the challenging job of reconstructing the segmentation map with a size equal to the original image from the feature map of the encoder module.

Based on the DeepLabV3+ method, this article designed the semantic segmentation network of the PolSAR image, which mainly composed of ResNet101-V2 and the atrous spatial pyramid pooling (ASPP) module, as shown in Fig. 10.

1) ResNet101-V2: Considering the complex speckles, the deep backbone network can effectively reduce the speckle interference and extract the useful information. Meanwhile, continuous research has shown that the depth of the neural network has a crucial impact on the performance of the segmentation method [37], [40]. As the network continues to deepen, the gradient will disappear, which hinders the convergence of the network, making the model easy to converge to a local minimum, resulting in the network performance not reaching the ideal effect [41]. This problem can be addressed by a residual learning framework, namely ResNet.

In a variety of image classification applications [42], [43], it is shown that the ResNet101 network with 101 layers has better segmentation performance than the shallow residual network, such as ResNet34, ResNet50. The ResNet101 consists of 33 three-layer building blocks of the residual learning. In

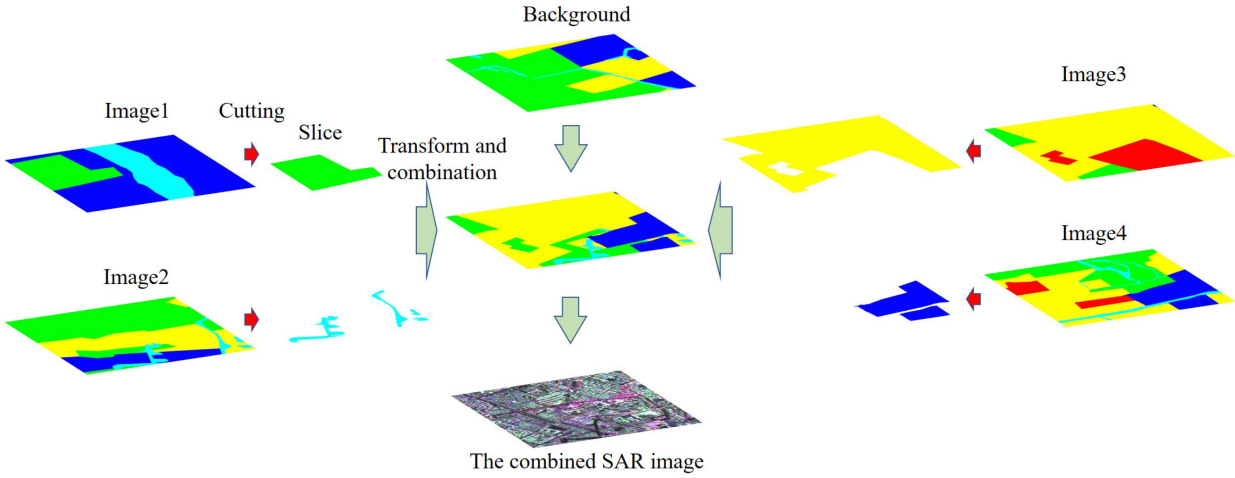

Fig. 7. Workflow of the RRM method.

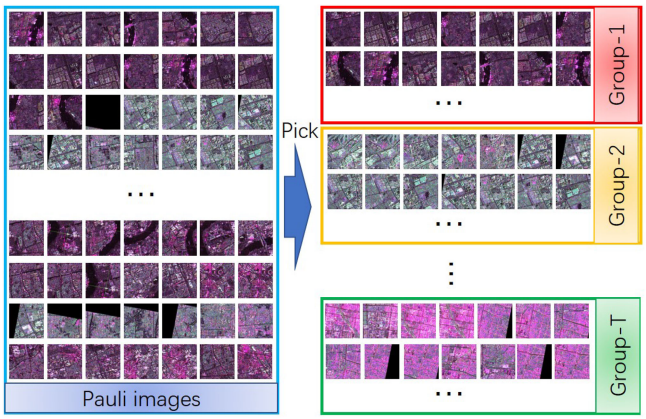

Fig. 8. Group the training samples.

PSPNet [18], the ResNet101 network structure adopts original Inception block, namely ResNet101-V1, while the ResNet-V2 adopts cheaper one [44]. In multiple applications [45], [46], the ResNet-V2 can achieve higher classification accuracy than the ResNet-V1. In our network architecture, the ResNet101-V2 is used as the backbone network.

2) Atrous Spatial Pyramid Pooling: The ASPP module derivers the concept of pyramid pooling from the SSPNet [13] and the atrous convolution. It is implemented to resist the limitation of the fixed size of the input image to extract the high-level multiscale information accurately and efficiently at a different atrous rate and the synthesize global information, as shown in Fig. 10. Given the varying resolution of the PolSAR image, the ASPP module can offer a better segmentation result. Our experiments show that the segmentation model has a poor segmentation effect on small scenes, such as the river, road (it may be the segmentation boundaries of different scenes), etc. In order to improve the segmentation ability, the atrous rates of three atrous convolution kernels are set as 4, 8, 12, respectively, and the number of all convolution kernels in the ASPP module is set to 256.

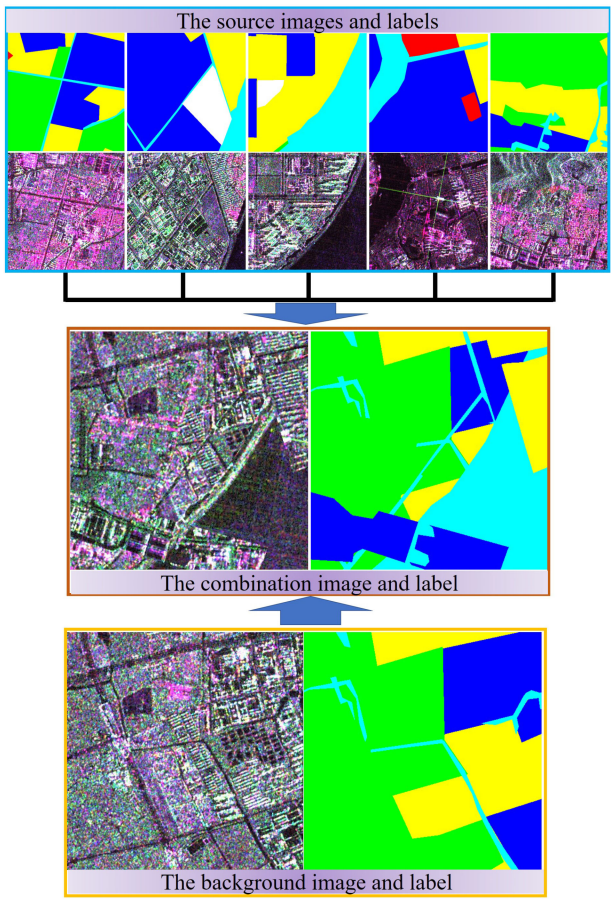

Fig. 9. Composite image using the RRM method.

III. Experiments and Results

In order to prove the effectiveness of the algorithm, we will present the experiment results from two aspects. 1) We designed a series of comparative experiments based on the classic segmentation method, and the results demonstrate the effectiveness of

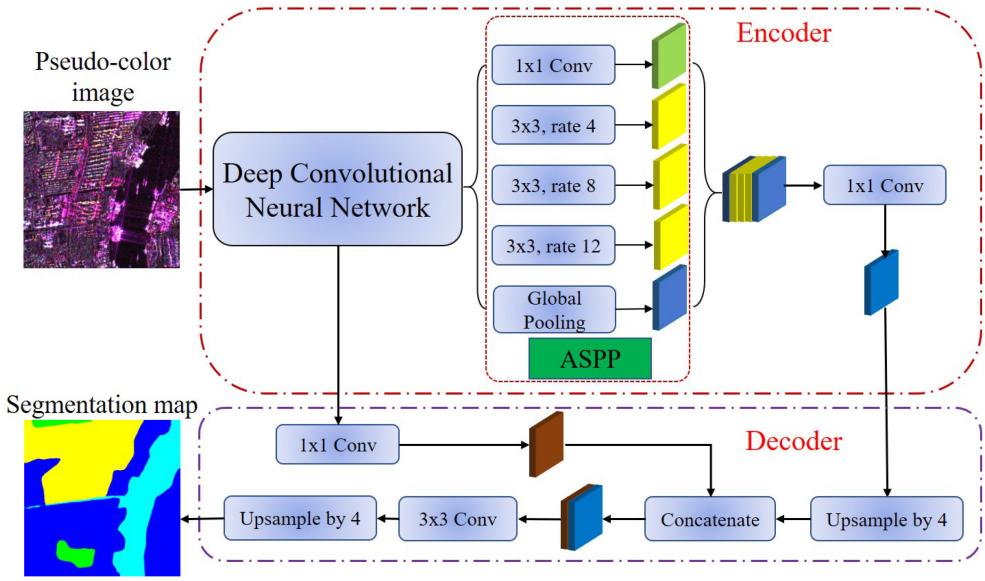

Fig. 10. Architecture of DeepLabV3+ for semantic segmentation in this article.

TABLE I
SAMPLE DETAILS IN THE EXPERIMENTS

Property	The Pre-training Data			The Contest Data		
	Training	RRM	Test	Training	RRM	Test
Number	100	400	400	500	2000	400
Resolution	1m-3m	-	1m-3m	1m-3m	-	1m-3m
Size	512×512	512×512	512×512	512×512	512×512	512×512 to 1500×1500
Memory	200 MB	802 MB	802 MB	0.97 GB	4 GB	1 GB
Pseudo-color Images (JPG)	37 MB	144 MB	144 MB	176 MB	694 MB	-

our method. 2) All the participating models are calculated on the same server and under the same computer resources at one time, and the convincing results obtained from each participating method in the final contest are used to illustrate the advantages of our method.

TABLE II
ENVIRONMENT OF OUR EXPERIMENTS

Environment	Specification
Keras	2.2.5
Tensorflow-gpu	1.14.0
CPU	Intel Core i7-7700k
Memory	16GB
GPU	GeForce RTX 2080Ti
Video memory	11GB
CUDA	10.0

A. Dataset

In the 2020 Gaofen Challenge on Automated High-Resolution Earth Observation Image Interpretation, 500 high-resolution full PolSAR images were released. The size of each image in the opened PolSAR images is 512 × 512, and the resolution range is 1 to 3 m. The ground truth maps are annotated with the six recognized land-cover types: water-body, buildings, industrial area, lawn, and two other unknown categories, as well as an unrecognized area. Due to the nondisclosure of the image tags, the type of the two unknown categories cannot be accurately determined and some special targets are difficult to identify, so this article uses the category label number to mark different categories to ensure the rigor of the content. The test data are composed of the 400 high-resolution fully PolSAR images, the image size ranges from 512 × 512 to 1500 × 1500, and the occupied memory of the test data is 1 GB. Since the test sample is not visible to users, a more detailed test sample introduction can no be provided. Therefore, 100 images are randomly selected from the public dataset as training samples, and the remaining 400 images are used as test samples to analyze the segmentation method. Considering the computational complexity, the pseudo-color images synthesized by $|C_{11}|$, $|C_{22}|$ and $|C_{33}|$ are used in the experiments. The sample details in the experiment are listed in Table I.

Three evaluation indicators are calculated to evaluate the quality of PolSAR image segmentation, namely overall accuracy (OA), frequency weighted intersection over union (FWIoU), and mean intersection over union (MIoU). MIoU is calculated as

$$MIoU = \frac{1}{C}\sum_{i=1}^{C}\frac{s_{ii}}{\sum_{j=1}^{C} s_{ij} + \sum_{j=1}^{C} s_{ji} - s_{ii}} \tag{13}$$

TABLE III
CLASSIFICATION RESULTS OF DIFFERENT METHODS

Methods	Category Accuracy (%)							OA(%)	MIoU(%)	FWIoU(%)
	1	2	3	4	5	6	7			
Unet	74.79	53.61	66.53	28.44	24.61	81.89	91.40	71.30	49.39	55.82
D-Link	72.35	66.69	72.38	39.62	25.71	79.53	91.65	72.71	53.63	57.71
DeepLabV3	73.43	74.10	70.07	47.88	63.52	81.88	75.41	74.93	59.56	60.34
DeepLabV3+	72.55	**78.18**	74.89	48.64	**66.75**	80.53	78.46	75.63	61.49	61.19
F-DeepLabV3+	74.42	71.27	**76.00**	49.64	65.82	**84.55**	78.88	76.25	61.83	62.02
Gen-DeepLabV3+	**79.63**	75.21	64.10	54.84	64.40	75.86	93.09	76.78	61.99	62.82
RRM-DeepLabV3+	79.25	73.02	65.95	**60.90**	66.61	78.89	**93.80**	**77.32**	**63.16**	**63.54**

The bold entities mean the optimal results.

where s_{ij} is the member of the confusion matrix S, i, and j are the corresponding coordinates, and C is the total number of categories. FWIoU is calculated as

$$FWIoU = \frac{1}{\sum_{i=1}^{C}\sum_{j=1}^{C} s_{ij}} \sum_{i=1}^{C} \frac{s_{ii}}{\sum_{j=1}^{C} s_{ij} + \sum_{j=1}^{C} s_{ji} - s_{ii}}. \tag{14}$$

In the competition, FWIoU is the only evaluation index for network segmentation performance.

B. Experimental Environment

The employed DeepLabV3+ segmentation method is implemented under Tensorflow-GPU and Keras architecture. Other comparison methods in our designed comparative experiments are also tested in this environment, and the basic configuration is shown in Table II.

C. Comparative Experiments in the Pretraining

In the preliminaries, the UNet, D-LinkNet [47] and DeepalavV3+ methods are implemented to test the accuracy of image segmentation before and after processing using (3), (4), and (8). In the UNet method, FWIoU accuracy is 58.7191 using the original PolSAR images of four channels. In contrast, the FWIoU accuracy is 62.6895 after processing with our method. Similarly, in the DeepLabV3+ method, the FWIoU accuracy before and after using these equations are 67.6806 and 70.2312, respectively. The preliminary experiment shows that the algorithm with (3), (4), and (8) can effectively improve the segmentation performance of the segmentation network, and the performance of the DeepLabV3+ method is better than that of the UNet method. In view of the long training time of the DeepLabV3+ method, we will not investigate this issue in detail but will add more comparative experiments to prove the effectiveness of our method.

In this article, based on the ResNet101-V2 network, UNet, D-LinkNet, DeepLabV3, and DeepLabV3+ are exploited to realize the refined segmentation of PolSAR images after the samples are processed by (3), (4), and (8). In addition, DeepLabV3+ combined with our filtering method is denoted as F-DeepLabV3+, DeepLabV3+ combined with the general sample enhancement method (i.e., rotate, flip, etc.) is denoted as Gen-DeepLabV3+, and DeepLabV3+ combined with the RRM method is denoted as RRM-DeepLabV3+. The results of the different segmentation methods are shown in Table III.

TABLE IV
CLASSIFICATION RESULTS OF DIFFERENT GROUPS IN THE FINAL CONTEST

No.	Method	FWIoU(%)	Inference Time(s)	Technology Maturity
1	The Proposed Method	**75.29**	166	**93.06**
2	MP-ResNet [48]	74.81	301	71.32
3	HRNetv2 [49]	74.48	1481	77.10
4	DeepLabV3+	72.73	208	80.37
5	-	72.48	**148**	85.36
6	-	65.76	721	-

The bold entities mean the optimal results.

From Table III, DeepLabV3 has dramatically improved the segmentation performance compared to UNet and DLink. Especially for the features of small targets and limited samples, such as categories 2, 4, and 5, the recognition ability of DeepLabV3 has significantly been improved. Compared with DeepLabV3, the accuracies of OA, MIoU, and FWIoU in DeepLabV3+ have increased from 74.926%, 59.561%, and 60.3358% to 75.6284%, 61.4945%, 61.1912%. After processing using the filtering algorithm proposed in this article, the network's segmentation accuracy has been improved slightly. Obviously, OA, MIoU, and FWIoU have been increased to 76.2482%, 61.8298%, and 62.0219%, respectively. Compared with training the segmentation network using original samples directly, the performance of the segmentation network trained by the sample amplification method has been greatly improved. In terms of the traditional sample enhancement methods, 400 samples are randomly synthesized to train the segmentation network. The segmentation results of OA, MIoU, and FWIoU have improved from 75.6284%, 61.4945%, 61.1912% to 76.7845%, 61.9856%, and 62.8219%. In contrast, 400 images obtained by the RRM method are added to the training samples, and OA, MIoU, and FWIoU have reached 77.3174%, 63.1606%, and 63.5427%.

D. Comparison of Segmentation Methods in the Final Contest

To fully illustrate the effectiveness of our method, the classification results of the different models in the finals are added to the comparison experiment. Since the specific methods of the contestants are non-transparent, their methods are temporarily named after the final ranking numbers. The segmentation accuracy, inference time, and the technology maturity of the segmentation method are used for the evaluation criteria of different segmentation algorithms, as shown in Table IV.

More than 100 teams submitted their methods in the preliminary round, of which eight teams with excellent segmentation

accuracy entered the final contest, and six teams obtained the classification results. From Table IV, our method has obvious advantages over the algorithms of other contestants. In addition to the highest segmentation accuracy, the inference time of our method is also ranked in the top 2 position. Besides, the official also proposed a technical maturity concept to measure the use-value of each segmentation method. The technical maturity is calculated by weighting inference time, resource occupation, segmentation accuracy, etc. The technical maturity of our segmentation method reached 93.0587, which is much higher than other contestants, and the result indicates our method is more practical than other participating models. It is worth noting that $No.4$ also uses DeepLabV3+ for the PolSAR segmentation, in comparison, our accuracy has been greatly improved. Simultaneously, due to the optimization of the data preprocessing, our algorithm also has an obvious advantage in inference time. In the end, our segmentation method won first place in the competition.

IV. Conclusion

This article proposed a novel method for the GaoFen-3 high-resolution PolSAR image. The decomposition algorithm is applied to extract image features of PolSAR to composite pseudocolour images, so as to compress the data volume of the sample. Meanwhile, a normalized method and a filtering method for reducing strong speckle noise and strong backscattering are designed to enhance local and global information of images, thus improving the segmentation performance of the network. On the other hand, in order to construct training samples for unknown scenes, a sample enhancement algorithm, namely RRM, is designed to improve the robustness of the segmentation network. Finally, the adjusted DeepLabV3+ method is implemented to achieve the semantic segmentation of PolSAR images and demonstrated its excellent segmentation performance. Whether it is our design of the comparative experiment results or comparing the results of other contestants in the finals, it shows that our algorithms have high accuracy and high practical value.

References

[1] F. Zhang *et al.*, "Nearest-regularized subspace classification for PolSAR imagery using polarimetric feature vector and spatial information," *Remote Sens.*, vol. 9, no. 11, 2017, Art. no. 1114.

[2] F. Zhang, Y. Wang, J. Ni, Y. Zhou, and W. Hu, "SAR target small sample recognition based on CNN cascaded features and AdaBoost rotation forest," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1008–1012, Jun. 2020.

[3] J. Ni, F. Zhang, Q. Yin, and H.-C. Li, "Robust weighting nearest regularized subspace classifier for polsar imagery," *IEEE Signal Process. Lett.*, vol. 26, no. 10, pp. 1496–1500, Oct. 2019.

[4] Y. Duan *et al.*, "Hierarchical multinomial latent model with G_0 distribution for synthetic aperture radar image semantic segmentation," *IEEE Access*, vol. 6, pp. 31783–31797, 2018.

[5] W. Wu, H. Li, X. Li, H. Guo, and L. Zhang, "PolSAR image semantic segmentation based on deep transfer learning-realizing smooth classification with small training sets," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 977–981, Jun. 2019.

[6] L. Jiao, M. Gong, S. Wang, B. Hou, Z. Zheng, and Q. Wu, "Natural and remote sensing image segmentation using memetic computing," *IEEE Comput. Intell. Mag.*, vol. 5, no. 2, pp. 78–91, May 2010.

[7] J. S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*. Boca Raton, FL, USA: CRC Press, 2017.

[8] X. Yang, X. Gao, D. Tao, X. Li, and J. Li, "An efficient MRF embedded level set method for image segmentation," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 9–21, Jan. 2015.

[9] F. Wang, Y. Wu, Q. Zhang, W. Zhao, M. Li, and G. Liao, "Unsupervised SAR image segmentation using higher order neighborhood-based triplet Markov fields model," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5193–5205, Aug. 2014.

[10] W. Yang, X. Zhang, L. Chen, and H. Sun, "Semantic segmentation of polarimetric SAR imagery using conditional random fields," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2010, pp. 1593–1596.

[11] P. Zhang, M. Li, Y. Wu, and H. Li, "Hierarchical conditional random fields model for semisupervised SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4933–4951, Sep. 2015.

[12] Z. Li, J. Ni, F. Zhang, W. Li, and Y. Zhou, "Multi-GPU implementation of nearest-regularized subspace classifier for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3534–3544, Jun. 2020.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[14] X. Sun, A. Shi, H. Huang, and H. Mayer, "BAS^4 NET: Boundary-aware semi-supervised semantic segmentation network for very high resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5398–5413, Sep. 2020.

[15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-NET: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.

[17] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNET: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[18] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2881–2890.

[19] G. Lin, A. Milan, C. Shen, and I. Reid, "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1925–1934.

[20] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2961–2969.

[21] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-SCNN: Gated shape CNNs for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 5229–5238.

[22] F. Mohammadimanesh, B. Salehi, M. Mahdianpari, E. Gill, and M. Molinier, "A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem," *ISPRS J. Photogrammetry Remote Sens.*, vol. 151, pp. 223–236, 2019.

[23] W. Yao, D. Marmanis, and M. Datcu, "Semantic segmentation using the fully convolutional networks for SAR and optical image pairs," in *Proc. Conf. Big Data Spcace*, 2017, pp. 289–292.

[24] 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.

[25] Y. Duan, F. Liu, L. Jiao, P. Zhao, and L. Zhang, "SAR image segmentation based on convolutional-wavelet neural network and Markov random field," *Pattern Recognit.*, vol. 64, pp. 255–267, 2017.

[26] J. Geng, W. Jiang, and X. Deng, "Multi-scale deep feature learning network with bilateral filtering for SAR image classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 167, pp. 201–213, 2020.

[27] 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.*, 2018, pp. 1–6.

[28] X. Wang, L. Cavigelli, M. Eggimann, M. Magno, and L. Benini, "Hr-SAR-NET: A deep neural network for urban scene segmentation from high-resolution SAR data," in *Proc. IEEE Sensors Appl. Symp.*, 2020, pp. 1–6.

[29] J. Ni, F. Zhang, Q. Yin, Y. Zhou, H. C. Li, and W. Hong, "Random neighbor pixel-block-based deep recurrent learning for polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3037209.

[30] S. Sengupta, V. Jayaram, B. Curless, S. M. Seitz, and I. Kemelmacher-Shlizerman, "Background matting: The world is your green screen," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2291–2300.

[31] P. Wang, X. Sun, W. Diao, and K. Fu, "FMSSD: Feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3377–3390, May 2020.
[32] S. Parrilli, M. Poderico, C. V. Angelino, and L. Verdoliva, "A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 606–616, Feb. 2012.
[33] 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.
[34] B. Baheti, S. Innani, S. Gajre, and S. Talbar, "Semantic scene segmentation in unstructured environment with modified DeepLabv3," *Pattern Recognit. Lett.*, vol. 138, pp. 223–229, 2020.
[35] X. Sun, P. Wang, C. Wang, Y. Liu, and K. Fu, "PBNet: Part-based convolutional neural network for complex composite object detection in remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 173, pp. 50–65, 2021.
[36] X. Sun, Y. Liu, Z. Yan, P. Wang, W. Diao, and K. Fu, "SRAF-Net: Shape robust anchor-free network for garbage dumps in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3023928.
[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[39] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1251–1258.
[40] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1026–1034.
[41] B. Yu, L. Yang, and F. Chen, "Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3252–3261, Sep. 2018.
[42] T. Rahman *et al.*, "Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization," *IEEE Access*, vol. 8, pp. 191586–191601, 2020.
[43] J. S. Velasco, M. V. C. Padilla, N. M. Arago, E. P. E. De Vera, F. E. M. Domingo, and R. E. R. Ramos, "Canine semen evaluation using transfer learning models," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 7, 2020, pp. 1–6.
[44] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, pp. 4278–4284.
[45] A. Kolesnikov, X. Zhai, and L. Beyer, "Revisiting self-supervised visual representation learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1920–1929.
[46] S. Peng, H. Huang, W. Chen, L. Zhang, and W. Fang, "More trainable inception-ResNet for face recognition," *Neurocomputing*, vol. 411, pp. 9–19, 2020.
[47] L. Zhou, C. Zhang, and M. Wu, "D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 182–186.
[48] L. Ding *et al.*, "MP-ResNet: Multi-path residual network for the semantic segmentation of high-resolution polSAR images," 2020, *arXiv:2011.05088*.
[49] K. Sun *et al.*, "High-resolution representations for labeling pixels and regions," 2019, *arXiv:1904.04514*.

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

His research interests include PolSAR classification, parallel computing, and imaging processing.

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

He is currently a Full Professor in electronic and information engineering at the Beijing University of Chemical Technology, Beijing, China. His research interests include remote sensing image processing, high performance computing, and artificial intelligence.

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

Fei Ma (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronic and information engineering from the Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China, in 2013, 2016, and 2020 respectively.

He is currently working with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China, as an Associate Professor. His research interests include radar signal processing, image processing, machine learning, and target detection.

Qiang Yin (Senior Member, IEEE) received the B.S. degrees in electronic and information engineering from the Beijing University of Chemical Technology, Beijing, China, in 2004, the M.S. and Ph.D. degrees in signal and information processing from the Institute of Electronics, Chinese Academy of Science, Beijing, China, in 2008 and 2016, respectively.

From 2008 to 2013, she was a Research Assistant with the Institute of Electronics, Chinese Academy of Sciences. From 2014 to 2015, she was a Research Fellow with European Space Agency, Roma, Italy. Currently, she is an Associate Professor with the College of Information Science and Technology, Beijing University of Chemical Technology. Her research interests include polarimetric/polarimetric interferometric SAR, and deep learning.

Deliang Xiang (Member, IEEE) received the B.S. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2010, the M.S. degree in photogrammetry and remote sensing from the National University of Defense Technology, Changsha, China, in 2012, and the Ph.D. degree in geoinformatics from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2016. In 2019, he was awarded a Humboldt Research Fellowship.

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

Dr. Xiang serves as a Reviewer for the *Remote Sensing of Environment*, the *ISPRS Journal of Photogrammetry and Remote Sensing*, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and several other international journals in the remote sensing field.