A Novel Attention Fully Convolutional Network Method for Synthetic Aperture Radar Image Segmentation

Zhenyu Yue, Fei Gao, Qingxu Xiong, *Member, IEEE*, Jun Wang, Amir Hussain, and Huiyu Zhou

***Abstract*—As an important step of synthetic aperture radar image interpretation, synthetic aperture radar image segmentation aims at segmenting an image into different regions in terms of homogeneity. Because of the deficiency of the labeled samples and the existence of speckling noise, synthetic aperture radar image segmentation is a challenging task. We present a new method for synthetic aperture radar image segmentation in this article. Due to the large size of the original synthetic aperture radar image, we first divide the input image into small slices. Then the image slices are input to the attention-based fully convolutional network for obtaining the segmentation results. Finally, the fully connected conditional random field is adopted for improving the segmentation performance of the network. The innovations of our method are as follows: 1) The attention-based fully convolutional network is embedded with the multiscale attention network which is capable of enhancing the extraction of the image features through three strategies, namely, multiscale feature extraction, channel attention extraction, and spatial attention extraction. 2) We design a new loss function for the attention fully convolutional network by combining Lovasz-Softmax and cross-entropy losses. The new loss allows us to simultaneously optimize the intersection over union and the pixel classification accuracy of the segmentation results. The experiments are performed on two airborne synthetic aperture radar image databases. It has been proved that our method is superior to other state-of- the-art image segmentation approaches.**

***Index Terms*—Attention mechanism, conditional random field (CRF), fully convolutional network (FCN), image segmentation, synthetic aperture radar (SAR).**

Manuscript received May 9, 2020; revised June 21, 2020 and July 25, 2020; accepted August 9, 2020. Date of publication August 12, 2020; date of current version August 24, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61771027, Grant 61071139, Grant 61471019, Grant 61501011, and Grant 61171122. The work of Amir Hussain was supported in part by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/M026981/1. The work of Huiyu Zhou was supported in part by the Royal Society-Newton Advanced Fellowship under Grant NA160342. *(Corresponding author: Fei Gao.)*

Zhenyu Yue, Fei Gao, Qingxu Xiong, and Jun Wang are with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China (e-mail: yuezhenyu@buaa.edu.cn; 08060@buaa.edu.cn; qxxiong@buaa.edu.cn; wangj203@buaa.edu.cn).

Amir Hussain is with Cognitive Big Data and Cyber-Informatics (CogBID) Laboratory, the School of Computing, Edinburgh Napier University, EH10 5DT Edinburgh, U.K. (e-mail: ussain@napier.ac.uk).

Huiyu Zhou is with the School of Informatics, University of Leicester, LE1 7RH Leicester, U.K. (e-mail: hz143@leicester.ac.uk).

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2020.3016064

I. Introduction

Because of the penetrating capability of synthetic aperture radar (SAR), we can acquire high-resolution SAR images regardless of weather conditions [1], [2]. SAR image interpretation includes image segmentation, target detection, target recognition, and so on [3]–[5]. Image segmentation is a key step of SAR image interpretation. Its purpose is to categorize the SAR images into different regions [6]. Commonly used methods for SAR image segmentation include Markov random field (MRF), edge detection, and optimal thresholding-based methods [7], [8]. However, these methods heavily rely on hand-crafted features. Because of speckle noise in SAR images, it is difficult to extract the desired hand-crafted features [9].

Recently, convolutional neural network (CNN) has attracted wide attention because of its powerful feature extraction ability [10]. In contrast to the conventional methods [11], [12], CNN utilizes a multilayer structure for automatically extracting image features, which improves the efficiency of feature extraction. To employ CNN in the pixel-wise image segmentation tasks where pixels are predicted with labels, a fully convolutional network (FCN) is proposed in [13]. The FCN contains two processes, namely, downsampling and upsampling. In the downsampling process, the FCN extracts abstract image features using convolution layers. By contrast, the deconvolution layers are used in the upsampling process for improving the resolution of feature maps. Based on the FCN, the U-Net designed in [14] combines high-resolution features and the upsampling output for improving the segmentation accuracy. Badrinarayanan *et al*. propose the SegNet model which follows an encoder–decoder structure [15]. The SegNet first uses the encoder network for feature extraction, then the resolution of feature maps is improved by the decoder network. Zhao *et al*. present the pyramid scene parsing network (PSPNet) where the FCN is embedded with a pyramid pooling model [16]. Since the pyramid pooling features provide additional contextual information, the segmentation performance of PSPNet is superior to that of the traditional FCN. To simultaneously utilize the information at the full image resolution and robust features of the input images, the full-resolution residual network (FRRNet) is proposed in [17]. The FRRNet consists of the pooling and residual streams, wherein the pooling stream uses the pooling operation to obtain robust features while the residual stream carries the information of the full image resolution. Fariba *et al*. apply the FCN to SAR

image segmentation tasks. Comparative experiments against traditional SAR image segmentation methods show that FCN improves the segmentation accuracy of SAR images [18]. Based on FCN and transfer learning, Wu *et al*. propose a new SAR image segmentation algorithm where a pretrained network is adopted for improving the segmentation accuracy [19]. Corentin *et al*. apply the spatial tolerance rules to FCN and present a new loss function by modifying the mean square error (MSE) loss. This method is capable of achieving promising results despite of the imbalance of target categories in the SAR dataset [20]. In [21], the residual dense U-Net (RDU-Net) is proposed for pixel-wise sea-land segmentation. The RDU-Net includes the downsampling and upsampling paths where the densely connected residual network blocks are designed for aggregating multiscale contextual information.

FCN have effectively improved the segmentation accuracy of SAR images. To enhance the performance of FCN, many research studies focus on increasing the depth of the network [22]–[25]. Although deep structures can effectively enhance the representation of image features, the number of the parameters used in the network also increases. As a result, more labeled samples are needed to train the deep FCN. However, the sample annotation of SAR images is time consuming and the segmentation performance of the deep FCN decreases significantly if the labeled samples are insufficient [26]. When processing visual signals, the attention mechanism helps us to focus on important information while ignoring the rest [27]–[29]. Inspired by the attention mechanism, some researchers have designed attention modules embedded in neural networks. The attention modules calculate the attention values for assigning different weights to the image features, so that the neural networks focus on important features while ignoring the rest [30]. The residual attention network designed in [31] generates attention-aware features by the attention module. In [32], the squeeze-and-excitation network (SENet) is designed for improving the representation of image features. SENet calculates the channel attention of the feature map and use it to highlight useful features but restrain useless ones. The selective kernel network (SKNet) designed in [33] also utilizes the channel attention. Compared with the SENet, SKNet extracts image features with different scales by adopting a dynamic selection mechanism. The spatial and channel squeeze and excitation (SCSE) and CBAM modules utilize a pooling operation to calculate the channel and spatial attention maps, then the input feature map is multiplied with the two attention maps for feature optimization [34], [35]. The experiments have proved the capability of SCSE and CBAM modules for improving the accuracy of image classification. Besides, the numbers of the parameters used in SENet, SKNet, SCSE, and CBAM are much less than that in the deep FCN, which enables these attention modules to improve the performance of neural networks despite the deficiency of the labeled samples.

Because of the large size of the original SAR images, the segmentation methods first divide the images into small slices, then the segmentation results of the slices are combined for obtaining the final results [36], [37]. However, these methods neglect the spatial correlation of the image slices, which limits the segmentation accuracy [38], [39]. Recently, the fully connected conditional random field (CRF) is widely utilized for image segmentation [40], [41]. As a graph-based method, the fully connected CRF utilizes the spatial information by capturing the correlation of the image pixels. In [42], the DeepLab method first uses the CNN to obtain the segmentation results, then the fully connected CRF is adopted for improving the segmentation accuracy. Ma *et al*. present the hierarchically adversarial CRF (HACRF) method where the CRF is combined with the generative adversarial network [43]. The experiments prove that the CRF can effectively improve the segmentation accuracy of SAR images.

We present the attention fully convolutional network (AFCN) for SAR image segmentation. In our method, we first separate the input SAR images into small slices. Afterwards, the image slices are fed into the AFCN for obtaining the segmentation results of the input images. Finally, we utilize the fully connected CRF to improve the performance of AFCN. The innovations of our method are as follows.

1) We design the multiscale attention network (MANet) which is embedded within the AFCN model. The MANet, which is proposed to enhance the extraction of the SAR image features, contains three parts: Multiscale feature, channel attention, and spatial attention extraction. The multiscale feature extraction module extracts image features of different scales by convolution kernels of different sizes, which effectively enhances the feature representation. The channel and spatial attention extraction modules utilize attention values to reassign weights to the image features, thus the AFCN is capable of learning to focus on the important features.
2) The loss function of our AFCN consists of Lovasz-Softmax and cross-entropy losses, wherein the former component optimizes the Intersection over Union (IoU) of the segmentation results while the latter one optimizes the pixel classification accuracy. The segmentation performance of AFCN is further improved by utilizing the merits of the two loss functions.

The rest of this article is structured as follows. Section II describes our method in detail. Then the experiments are reported in the Section III. Finally, Section IV concludes the article.

II. Proposed Method

We propose a novel SAR image segmentation method in this article. As shown in Fig. 1, our method contains three parts: Image preprocessing, segmentation results acquisition, and CRF postprocessing. Due to the large size of the original SAR images, we first divide the input image into small slices in the image preprocessing. In the acquisition stage, the image slices are input to the AFCN to acquire the segmentation results. The AFCN model is embedded with the MANet which is proposed to enhance the extraction of the image features. Besides, the loss function for AFCN consists of Lovasz-Softmax and cross-entropy losses. This new loss function can simultaneously optimize the IoU and the pixel classification accuracy of the segmentation results. In the CRF stage, we adopt the fully connected CRF to improve

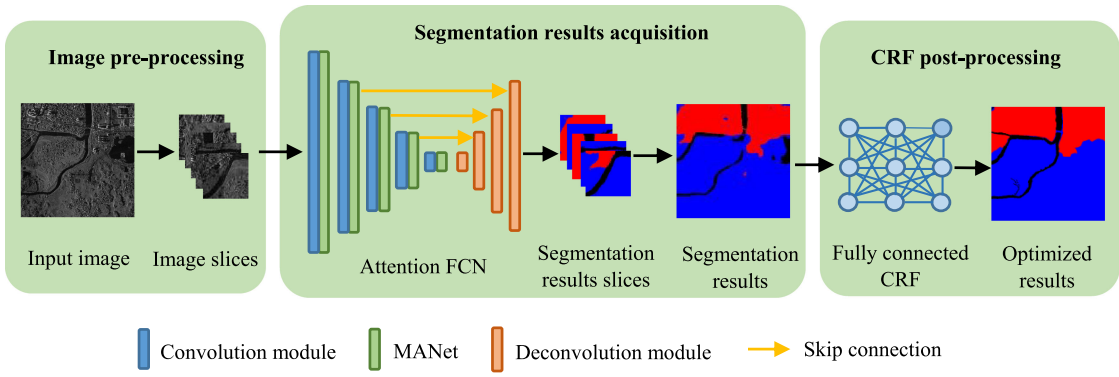

Fig. 1. Flowchart of our method.

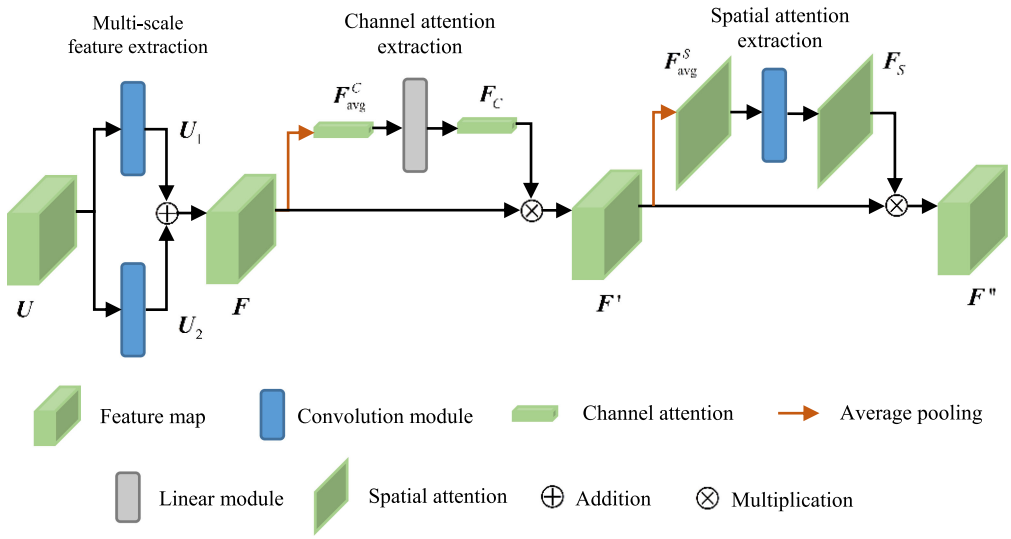

Fig. 2. Structure of MANet.

the segmentation performance of AFCN. The fully connected CRF captures the correlation of the image pixels, thus the spatial information contained in the SAR images is effectively utilized. Next, the AFCN, the loss function, and the fully connected CRF of our method are described, respectively.

A. AFCN Structure

In the proposed method, we design the AFCN for SAR image segmentation. As Fig. 1 shows, AFCN contains convolution modules, MANets, and deconvolution modules, wherein the skip connection method is utilized for fusing image features of different resolutions. Given the input image, the convolution modules extract image features using convolution and pooling layers. Each convolution module is followed by a MANet which is capable of enhancing the extraction of image features. The MANet first extracts the multiscale features, then the channel and spatial attention are calculated to reassign weights for the features. Hence, the AFCN is capable of learning to focus on the important features while ignoring the less important ones. The deconvolution modules utilize deconvolution layers for improving the resolution of image features, thereby obtaining the segmentation results.

In the structure of AFCN, we design the MANet to enhance the extraction of the SAR image features. As shown in Fig. 2, the MANet consists of three parts: Multiscale feature, channel attention, and spatial attention extraction. In the human visual system, the neurons in the same area adopt different receptive field sizes, so that the neurons are capable of collecting multiscale spatial information [33]. Inspired by this mechanism, recent CNNs have adopted convolutional kernels of different sizes to aggregate multiscale information [23]–[25]. Therefore, we utilize two convolution kernels of different sizes for extracting image features in the multiscale feature extraction part, thereby enhancing the representation of image features. Afterward, the attention maps are calculated in the channel and spatial attention extraction modules to further improve the feature representation. The detailed process of MANet is as follows.

Given the input feature map $\boldsymbol{U} \in \mathbb{R}^{C\times H\times W}$, where C, H, and W denote the number of channels, height, and width of the feature map, respectively. We first extract the multiscale feature using two convolution modules of different kernel sizes. Then

the output feature maps $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ are combined to generate the multiscale feature map $\boldsymbol{F} \in \mathbb{R}^{C\times H\times W}$:

$$\boldsymbol{F} = \boldsymbol{U}_1 \oplus \boldsymbol{U}_2 \tag{1}$$

where $\oplus$ denotes the element-wise addition.

Afterwards, we calculate the channel attention of the image features. Since the pooling operation is often used to obtain attention values, we adopt it in the channel attention extraction section. The brightness of different objects in SAR images varies greatly because of the imaging nature. Therefore, we utilize the averaging pooling operation for preserving the features of low brightness regions. Suppose $\boldsymbol{F}^C_{\text{avg}} \in \mathbb{R}^C$ represents the average pooling features in the channel dimension. $F^C_{\text{avg}}(n) \in \boldsymbol{F}^C_{\text{avg}}$ is calculated by

$$F^C_{\text{avg}}(n) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F(n,i,j) \tag{2}$$

where $F(n,i,j) \in \boldsymbol{F}$. Then $\boldsymbol{F}^C_{\text{avg}}$ is sent to a linear module and the channel attention $\boldsymbol{F}_C \in \mathbb{R}^C$ is obtained. To optimize $\boldsymbol{F}$, we reassign weights for the features in $\boldsymbol{F}$ by multiplying it with $\boldsymbol{F}_C$ in the C dimension, as shown in (3)

$$\boldsymbol{F}' = \boldsymbol{F} \otimes \boldsymbol{F}_C \tag{3}$$

where $\boldsymbol{F}' \in \mathbb{R}^{C\times H\times W}$ denotes the feature map optimized by channel attention. $\otimes$ denotes the multiplication operation.

Finally, the MANet uses spatial attention to optimize $\boldsymbol{F}'$, wherein the averaging pooling operation is also employed. Suppose $\boldsymbol{F}^S_{\text{avg}} \in \mathbb{R}^{H\times W}$ denotes the average pooling features in the spatial dimension. $F^S_{\text{avg}}(i,j) \in \boldsymbol{F}^S_{\text{avg}}$ is calculated by

$$F^S_{\text{avg}}(i,j) = \frac{1}{C}\sum_{n=1}^{C} F'(n,i,j) \tag{4}$$

where $F'(n,i,j) \in \boldsymbol{F}'$. Then $\boldsymbol{F}^S_{\text{avg}}$ is forwarded to a convolution module for obtaining the spatial attention $\boldsymbol{F}_S \in \mathbb{R}^{H\times W}$. To optimize $\boldsymbol{F}'$, we reassign weights to the features in $\boldsymbol{F}'$ by multiplying it with $\boldsymbol{F}_S$, as shown in (5)

$$\boldsymbol{F}'' = \boldsymbol{F}' \otimes \boldsymbol{F}_S \tag{5}$$

where $\boldsymbol{F}''$ represents the final optimized feature map.

B. Loss Function

Suppose $\boldsymbol{y}^* \in \mathbb{R}^N$, $\boldsymbol{y} \in \mathbb{R}^N$, and $\overline{\boldsymbol{y}} \in \mathbb{R}^{N\times K}$denote the ground-truth labels, the predicted labels and the label assignment probability of the image pixels, respectively. N represents the number of the image pixels. K is the number of the target categories. $\overline{\boldsymbol{y}}^* \in \mathbb{R}^{N\times K}$ denotes the one-hot encoding of $\boldsymbol{y}^*$. Most FCN-based image segmentation methods employ the cross-entropy loss which is designed for optimizing the pixel classification accuracy of segmentation results. The cross-entropy loss is expressed in (6)

$$\text{loss}_{\text{ce}} = -\frac{1}{N}\sum_{i=0}^{N-1}\sum_{k=0}^{K-1} \overline{y}^*_{i,k}\log \overline{y}_{i,k} \tag{6}$$

where $\overline{y}^*_{i,k} \in \overline{\boldsymbol{y}}^*$, $\overline{y}_{i,k} \in \overline{\boldsymbol{y}}$.

However, IoU is also an important evaluation method of image segmentation. Suppose $\boldsymbol{I}$ denotes the set of the image pixels. The IoU of category k can be calculated by

$$\text{IoU}_k = \frac{|\boldsymbol{P}_k \cap \boldsymbol{G}_k|}{|\boldsymbol{P}_k \cup \boldsymbol{G}_k|} \tag{7}$$

where $\boldsymbol{P}_k \in \boldsymbol{I}$ denotes the set of the image pixels whose predicted labels belong to category k, and $\boldsymbol{G}_k \in \boldsymbol{I}$ denotes the set of the image pixels whose ground-truth labels belong to category k. $\cap$ and $\cup$, respectively represent the intersection and union operations of two sets. During the training process, optimizing the IoU is an effective measure to promote the segmentation performance of our AFCN. The Lovasz-Softmax loss is designed to optimize IoU [44], and its derivation is as follows.

Based on (7), a corresponding loss function is expressed in (8)

$$\Delta_{\text{IoU}_k} = 1 - \text{IoU}_k. \tag{8}$$

Then the set of the mispredicted pixels for category k is defined

$$\boldsymbol{M}_k = (\boldsymbol{G}_k \cap (\boldsymbol{I} - \boldsymbol{P}_k)) \cup (\boldsymbol{P}_k \cap (\boldsymbol{I} - \boldsymbol{G}_k)). \tag{9}$$

We rewrite the loss function shown in (8) using $\boldsymbol{M}_k$

$$\Delta_{\text{IoU}_k}(\boldsymbol{M}_k) = \frac{|\boldsymbol{M}_k|}{|\boldsymbol{G}_k \cup \boldsymbol{M}_k|}. \tag{10}$$

To optimize (10) in a continuous setting, we utilize the Lovasz extension of $\boldsymbol{M}_k$, and the loss function is rewritten as

$$\overline{\Delta_{\text{IoU}_k}}(\boldsymbol{m}_k) = \sum_{i=0}^{N-1} m_k(i) g_i(\boldsymbol{m}_k) \tag{11}$$

where $\boldsymbol{m}_k \in \mathbb{R}^N$ represents the vector of the pixel errors. $m_k(i) \in \boldsymbol{m}_k$ is calculated as follows:

$$m_k(i) = \begin{cases} 1 - f_k(i) & \text{if } k = y^*_i \\ f_k(i) & \text{otherwise} \end{cases} \tag{12}$$

where $f_k(i) \in [0,1]$denotes the class probability that the ith pixel belongs to category k, $y^*_i \in \mathbf{y}^*$. $g_i(\boldsymbol{m}_k)$ is calculated by (13)

$$g_i(\boldsymbol{m}_k) = \Delta_{\text{IoU}_k}(\{\pi_1, \cdots, \pi_i\}) - \Delta_{\text{IoU}_k}(\{\pi_1, \cdots, \pi_{i-1}\}). \tag{13}$$

$\{\pi_1, \cdots, \pi_i\}$ denotes the permutation ordering the components of $\boldsymbol{m}_k$ in a decreasing order. Considering the mean IoU of all the categories, the Lovasz-Softmax loss is expressed as follows:

$$\text{loss}_{\text{ls}} = \frac{1}{K}\sum_{k=0}^{K-1} \overline{\Delta_{\text{IoU}_k}}(\boldsymbol{m}_k). \tag{14}$$

To simultaneously optimize the pixel classification accuracy and the IoU, a new loss function is designed for our AFCN by incorporating the Lovasz-Softmax and cross-entropy losses

$$\text{loss} = \text{loss}_{\text{ce}} + \alpha \text{loss}_{\text{ls}} \tag{15}$$

where α denotes the weight coefficient.

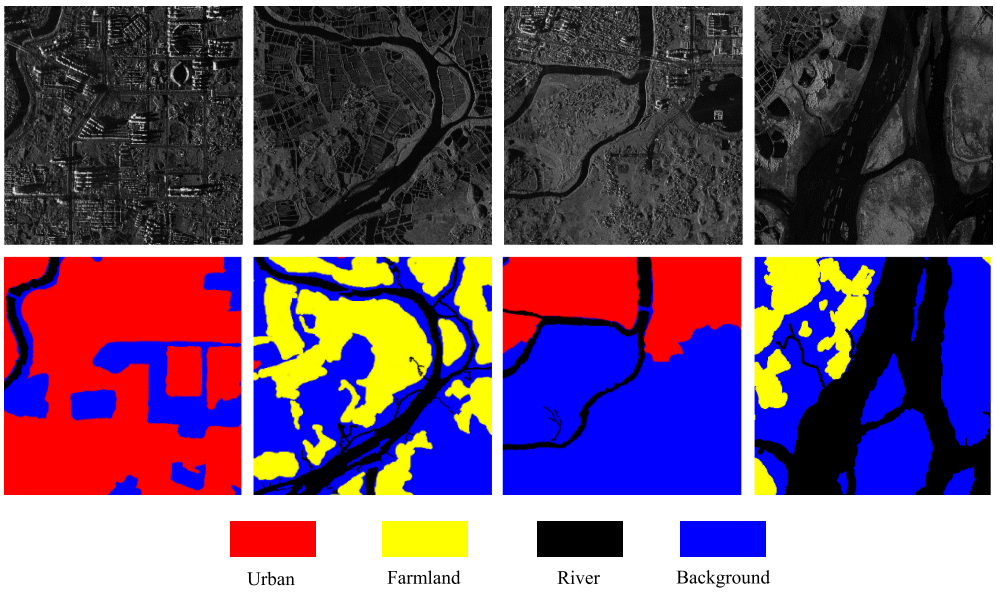

Fig. 3. SAR images and corresponding ground truth in the Fangchenggang dataset.

C. Fully Connected CRF

In the testing process of our method, we adopt the fully connected CRF to improve the segmentation performance of AFCN. We define the pixels as nodes. The energy function of the fully connected CRF is as follows:

$$E(\mathbf{y}) = \sum_i \psi_u(y_i) + \sum_{i,j} \psi_p(y_i, y_j). \tag{16}$$

$\psi_u(y_i)$ is the unary potential which is calculated by

$$\psi_u(y_i) = -\log P(y_i). \tag{17}$$

$P(y_i)$ denotes the label assignment probability of the ith pixel. $\psi_p(y_i, y_j)$ denotes the pairwise potential which captures the correlation of the image pixels. The expression of $\psi_p(y_i, y_j)$ consists of two weighted-Gaussian kernels functions

$$\psi_p(y_i, y_j) = \mu(y_i, y_j)\left(\omega_1 e^{\left(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|g_i - g_j|^2}{2\theta_\beta^2}\right)} + \omega_2 e^{\left(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}\right)}\right) \tag{18}$$

where ω_1 and ω_2 are the weight coefficients. p_i and p_j denote the pixel positions. g_i and g_j represent the grey values of the pixels. θ_α, θ_β, and θ_γ are the hyper parameters which control the scale of Gaussian kernels. $\mu(y_i, y_j)$ denotes the penalty function

$$\mu(y_i, y_j) = \begin{cases} 1 & \text{if} \quad y_i \neq y_j \\ 0 & \text{if} \quad y_i = y_j \end{cases}. \tag{19}$$

III. Experiments

In this section, the dataset, evaluation measures, and implementation details are first introduced. Then the performance of our method is compared with that of other segmentation methods. After that, we demonstrate the effectiveness of the MANet, the loss function, and the fully connected CRF. Finally, we discuss the computational efficiency of our method.

A. Preliminary

1) Dataset Description: We perform the experiments on the Fangchenggang and Pucheng datasets. The Fangchenggang dataset consists of the airborne SAR images obtained in Fangchenggang, China. The resolution of the SAR images is 2 m and the imaging area is about 30 × 30 km. There are 36 images in this dataset, and the image size is 875 × 883 pixels. In Fig. 3, we show some SAR images and the ground truth of this dataset. Different colors represent different categories of areas, where blue, black, yellow, and red represent background, river, farmland, and urban areas, respectively.

The Pucheng dataset contains 40 airborne SAR images collected from Pucheng, China. The image size is 850 × 850 pixels, and the resolution is 1 m. We show some SAR images and ground truth of the Pucheng dataset in Fig. 4. As can be seen, this dataset includes two categories of areas: Urban and farmland.

2) Evaluation Metrics: We utilize the mean intersection over union (MIoU), frequency weighted intersection over union (FWIoU), pixel accuracy (PA), and mean pixel accuracy (MPA) to estimate the performance of different methods. Suppose K and N represent the number of the target categories and the number of the image pixels, respectively. MIoU denotes the average IoU of all the categories

$$\text{MIoU} = \frac{1}{K}\sum_{i=0}^{K-1} \frac{p_{ii}}{\sum_{j=0}^{K-1} p_{ij} + \sum_{j=0}^{K-1} p_{ji} - p_{ii}} \tag{20}$$

where p_{ii} represents the element at coordinate (i, i) of the confusion matrix. $\sum_{j=0}^{K-1} p_{ij}$ and $\sum_{j=0}^{K-1} p_{ji}$ denote the element sums in the ith row and ith column in the confusion matrix, respectively. Compared with MIoU, FWIoU considers the weight

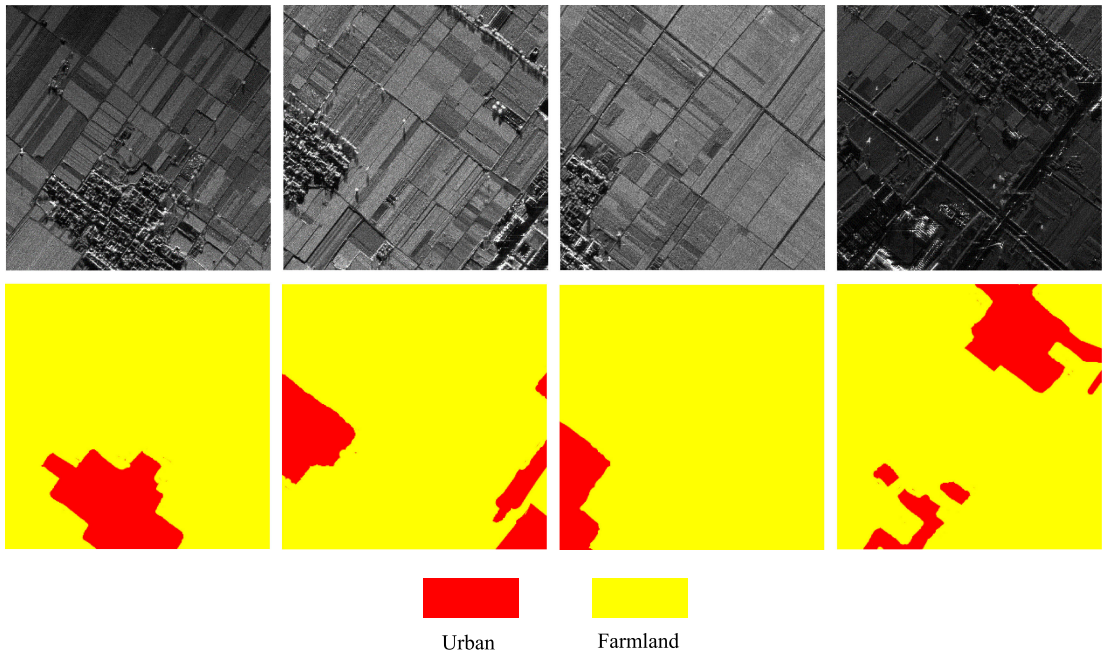

Fig. 4. SAR images and corresponding ground truth in the Pucheng dataset.

for each category in terms of the number of the pixels

$$\mathrm{FWIoU} = \frac{1}{N}\sum_{i=0}^{K-1}\frac{\left(\sum_{j=0}^{K-1} p_{ij}\right) p_{ii}}{\sum_{j=0}^{K-1} p_{ij} + \sum_{j=0}^{K-1} p_{ji} - p_{ii}}. \tag{21}$$

PA denotes the proportion of the correctly classified pixels to the total number of the pixels, and it is calculated by (22)

$$\mathrm{PA} = \frac{1}{N}\sum_{i=0}^{K-1} p_{ii}. \tag{22}$$

MPA represents the average PA of all the categories

$$\mathrm{MPA} = \frac{1}{K}\sum_{i=0}^{K-1}\frac{p_{ii}}{\sum_{j=0}^{K-1} p_{ij}}. \tag{23}$$

3) Implementation Details: During the image preprocessing, the input images are separated into small slices with a size of 224 × 224 and a step of 200. The network structure of AFCN includes five convolution modules, five MANets, and four deconvolution modules. The convolution modules consist of convolution, pooling, and batch normalization layers. The size of the convolution kernels is 3 × 3, and the pooling size adopted in pooling layers is 2 × 2. The numbers of the kernels contained in the five modules are 32, 64, 128, 256, and 512. The deconvolution module consists of a deconvolution layer which contains K kernels, and the kernel sizes of the 4 modules are 4 × 4, 4 × 4, 4 × 4, and 8 × 8, respectively.

In the multiscale feature extraction of MANet, the convolution modules consist of convolution and batch normalization layers. The kernel sizes of the two modules are 3 × 3 and 5 × 5, respectively. The number of the kernels in the two modules is C. The linear module in the channel attention extraction is composed of two linear layers. The number of the neurons in the two layers are C/r and C, where r denotes the reduction ratio which is set to 8. The convolution module in the spatial attention extraction section consists of convolution and batch normalization layers. The convolution layer contains one kernel, and the kernel size is 7 × 7.

TABLE I
NUMBERS OF PIXELS IN DIFFERENT CATEGORIES

	Urban	Farmland	River	Background
Training set	0.94M	0.63M	1.02M	2.23M
Testing set	2.40M	2.21M	3.75M	15.72M

M denotes the abbreviation of million.

B. Segmentation Performance Comparison

1) Experiments on the Fangchenggang Dataset: In this section, the experiments are performed on the Fangchenggang dataset. During the training process, we choose six images from this dataset as training set and the others are utilized as testing set. The numbers of the pixels contained in different categories are shown in Table I , where M is the abbreviation of million. As can be seen, the numbers of the pixels in different categories vary largely. For example, the background areas in the training set contain 2.23-M pixels, whereas the farmland areas contain 0.63-M pixels. This imbalance presents a challenge for the segmentation methods.

We compare our method with U-Net [14], SegNet [15], PSPNet [16], FRRNet [17], and DeepLab model [42]. The U-Net combines high-resolution features and upsampling output for improving the segmentation accuracy. The SegNet follows an encoder–decoder network where the encoder extracts image features whereas the decoder improves the resolution of the

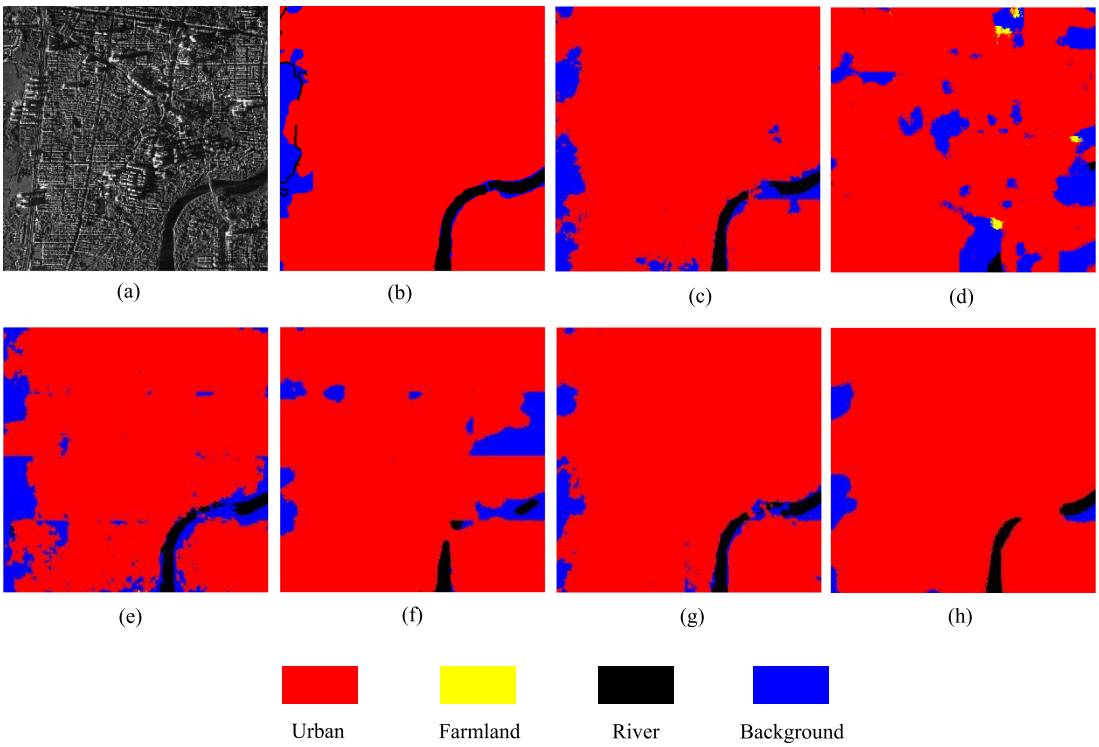

Fig. 5. Segmentation results of different methods. The fully connected CRF is utilized for optimizing the segmentation results of different methods (a) Test image. (b) Ground-truth. (c) U-Net. (d) DeepLab. (e) SegNet. (f) PSPNet. (g) FRRNet. (h) Our method.

TABLE II
PERFORMANCE OF DIFFERENT IMAGE SEGMENTATION METHODS

Methods	MIoU	FWIoU	PA	MPA
U-Net	68.03%	77.21%	87.22%	75.76%
DeepLab	54.44%	67.99%	81.10%	63.44%
SegNet	66.13%	76.23%	86.58%	74.25%
PSPNet	56.76%	70.40%	82.60%	65.12%
FRRNet	68.51%	77.66%	87.46%	76.14%
Ours	71.85%	79.65%	88.69%	79.49%

The fully connected CRF is adopted for improving the segmentation performance of different methods.

feature maps. By embedding the pyramid pooling module in FCN, the PSPNet extracts the global context information for improving the segmentation accuracy. The FRRNet is capable of simultaneously utilizing the information at full image resolutions and the robust features of the input images. DeepLab model first obtains the coarse segmentation results using the deep CNN, then the fully connected CRF is adopted for refinement. As a postprocessing algorithm, CRF can be embedded after all the algorithms. Thus we also perform CRF processing on the segmentation results of U-Net, SegNet, PSPNet, and FRRNet.

It is obvious that the MIoU, FWIoU, PA, and MPA of our method outperform the other methods. The MIoU of our method is 5.72% higher than the SegNet and 3.82% higher than the U-Net. This is because the SegNet and U-Net do not optimize image features in the feature extraction. In contrast, our AFCN is embedded with the MANet which optimizes the SAR image features through three strategies: Multiscale feature, channel attention, and spatial attention extraction. DeepLab model, PSPNet, and FRRNet are based on deep network structures. Although deep networks can improve the feature representation, the labeled samples in the SAR dataset is insufficient, which leads to the overfitting problem. By contrast, our method adopts a shallow structure which can effectively reduce the risk of overfitting.

Next, we choose four images in the testing set and draw the segmentation results in Figs. 5 to 8. As can be seen, the segmentation results of our method match well with the ground truth. Compared with SegNet and U-Net, our method effectively reduces the number of the misclassified pixels. This is because MANet embedded in our AFCN is capable of enhancing the extraction of the SAR image features. Due to the overfitting problem, the numbers of the misclassified pixels of the DeepLab model, PSPNet, and FRRNet are larger than that of our method, which is particularly obvious at the region boundaries in the segmentation results.

2) Experiments on the Pucheng Dataset: In this section, we validate the effectiveness of our method using the Pucheng dataset. Four images are chosen as training set and the others are used as testing set. The numbers of the pixels contained in the urban and farmland areas are shown in Table III. It is obvious that the number of the pixels in the farmland area is much larger than that in the urban area.

We compare the performance of our method with that of U-Net [14], SegNet [15], PSPNet [16], FRRNet [17], and DeepLab model [42]. The CRF processing is used to optimize the results

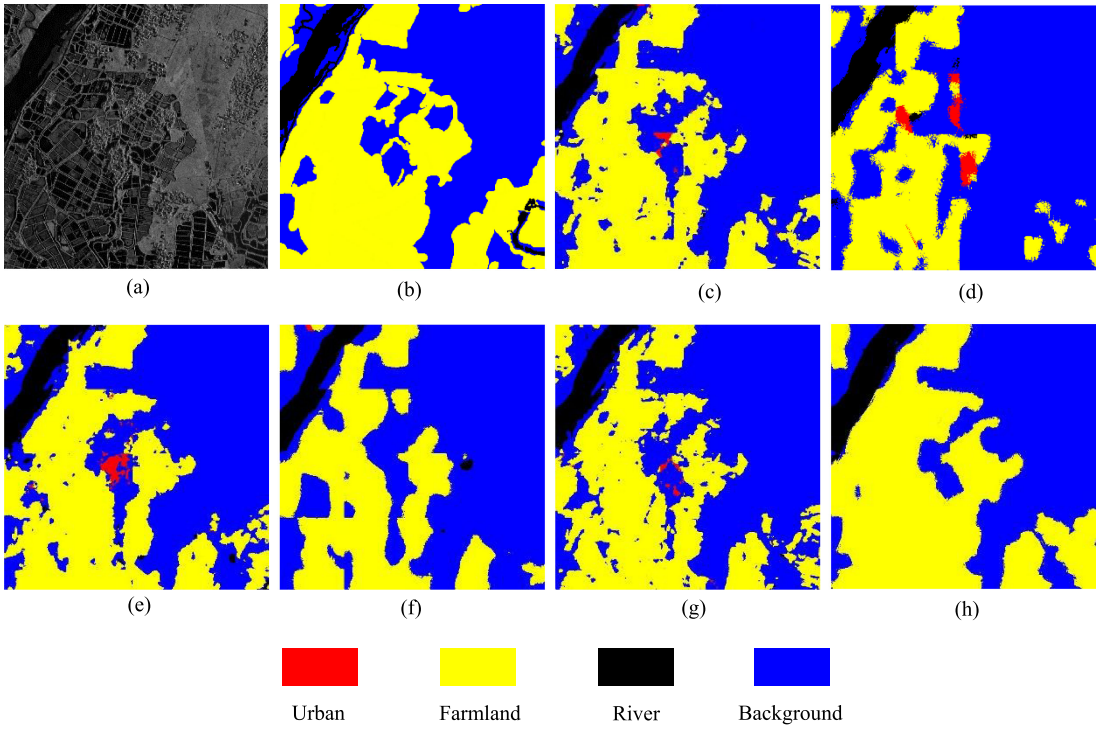

Fig. 6. Segmentation results of different methods. The fully connected CRF is utilized for optimizing the segmentation results of different methods (a) Test image. (b) Ground-truth. (c) U-Net. (d) DeepLab. (e) SegNet. (f) PSPNet. (g) FRRNet. (h) Our method.

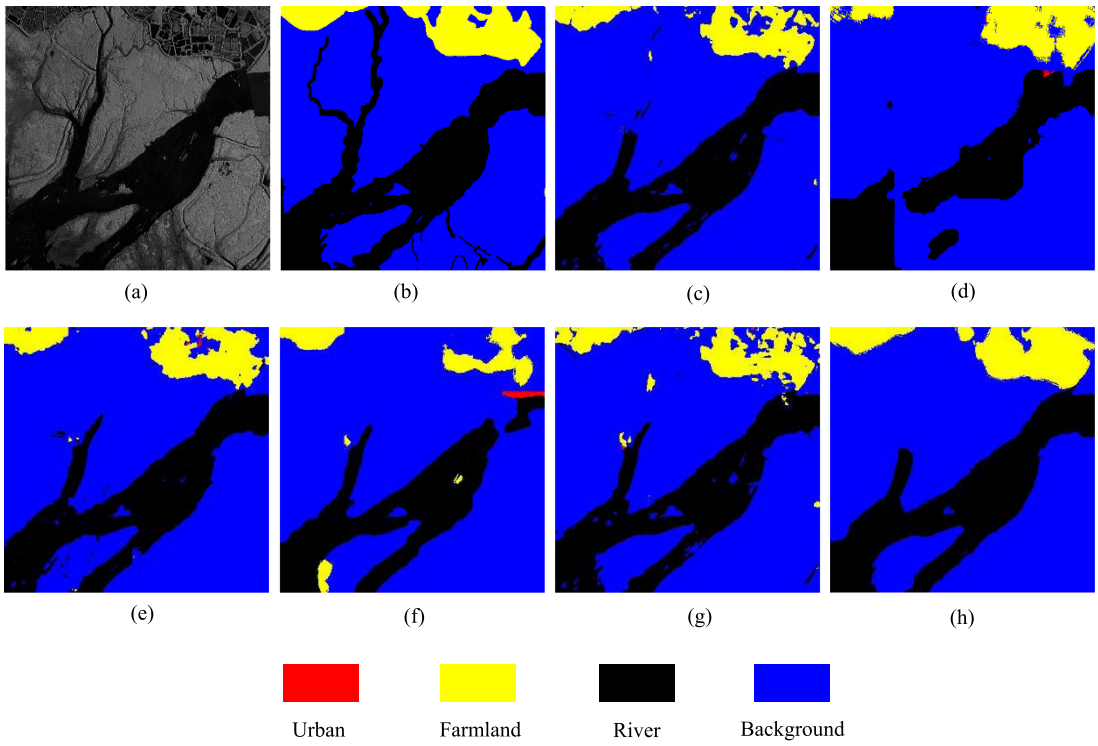

Fig. 7. Segmentation results of different methods. The fully connected CRF is utilized for optimizing the segmentation results of different methods (a) Test image. (b) Ground-truth. (c) U-Net. (d) DeepLab. (e) SegNet. (f) PSPNet. (g) FRRNet. (h) Our method.

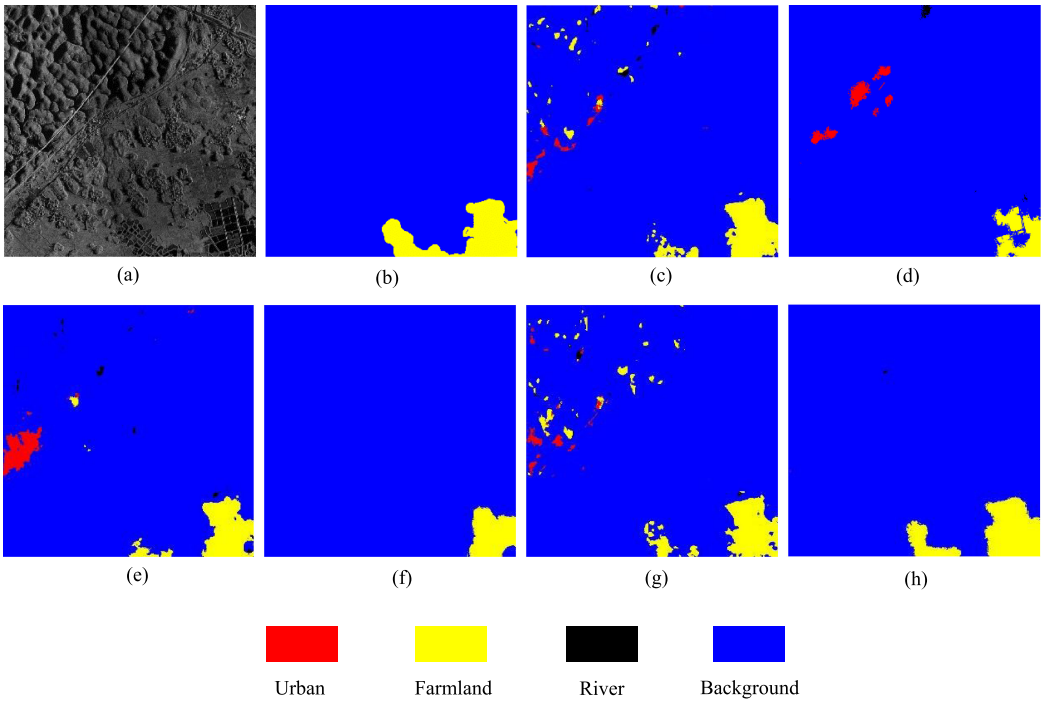

Fig. 8. Segmentation results of different methods. The fully connected CRF is utilized for optimizing the segmentation results of different methods (a) Test image. (b) Ground-truth. (c) U-Net. (d) DeepLab. (e) SegNet. (f) PSPNet. (g) FRRNet. (h) Our method.

TABLE III
NUMBERS OF PIXELS IN DIFFERENT CATEGORIES

	Urban	Farmland
Training set	0.27M	2.95M
Testing set	4.58M	24.32M

M denotes the abbreviation of million.

TABLE IV
PERFORMANCE OF DIFFERENT IMAGE SEGMENTATION METHODS

Methods	MIoU	FWIoU	PA	MPA
U-Net	84.74%	91.51%	95.39%	91.74%
DeepLab	72.62%	84.98%	91.65%	80.11%
SegNet	82.20%	90.14%	94.62%	89.26%
PSPNet	79.69%	88.42%	93.46%	89.38%
FRRNet	82.73%	90.63%	95.00%	88.01%
Ours	85.81%	92.18%	95.79%	91.99%

CRF processing is used to optimize the results of different methods.

of the different methods. As shown in Table IV, our method achieves the highest MIoU, FWIoU, PA, and MPA. The MIoU of our method (85.81%) is higher than that of U-Net (84.74%) and SegNet (82.20%), which demonstrates the effectiveness of feature optimization with MANet. Compared with the DeepLab model, PSPNet, and FRRNet, our method achieves the best segmentation performance despite the insufficiency of the labeled samples. These results prove that our method can obtain superior segmentation results on different SAR datasets.

We draw the segmentation results of the two testing images from the Pucheng dataset. In Figs. 9 and 10, our method can accurately distinguish the urban and farmland areas. In contrast, there are many misclassified pixels in the segmentation results of U-Net, SegNet, PSPNet, FRRNet, and the DeepLab model. Especially in Fig. 10 (c)–(g), a large number of pixels in the urban area are misclassified to the farmland.

C. *Evaluation of the MANet*

In our method, we design the attention module MANet and embed it in the AFCN for enhancing the extraction of the SAR image features. To demonstrate the effectiveness of MANet, we compare the segmentation performance of AFCN when it is embedded with different attention modules. The SENet [32], SKNet [33], and CBAM [35] are chosen as comparison attention modules. SENet and SKNet utilize the channel attention of the feature map to highlight useful features but restrain the others. Different from SENet, SKNet adopts the dynamic selection mechanism for extracting image features of different scales. CBAM utilizes the channel and spatial attention to optimize image features. Since the SENet, SKNet, and CBAM can be easily added to convolutional networks, we respectively embed them in the AFCN model to replace MANet. We adopt the cross-entropy loss and the experiments are performed on the Fangchenggang dataset. The segmentation performance of AFCN when it is embedded with different attention modules is shown in Table V.

As can be seen, the MIoU of MANet is 5.6% higher than that of CBAM. The CBAM uses average pooling and maximum

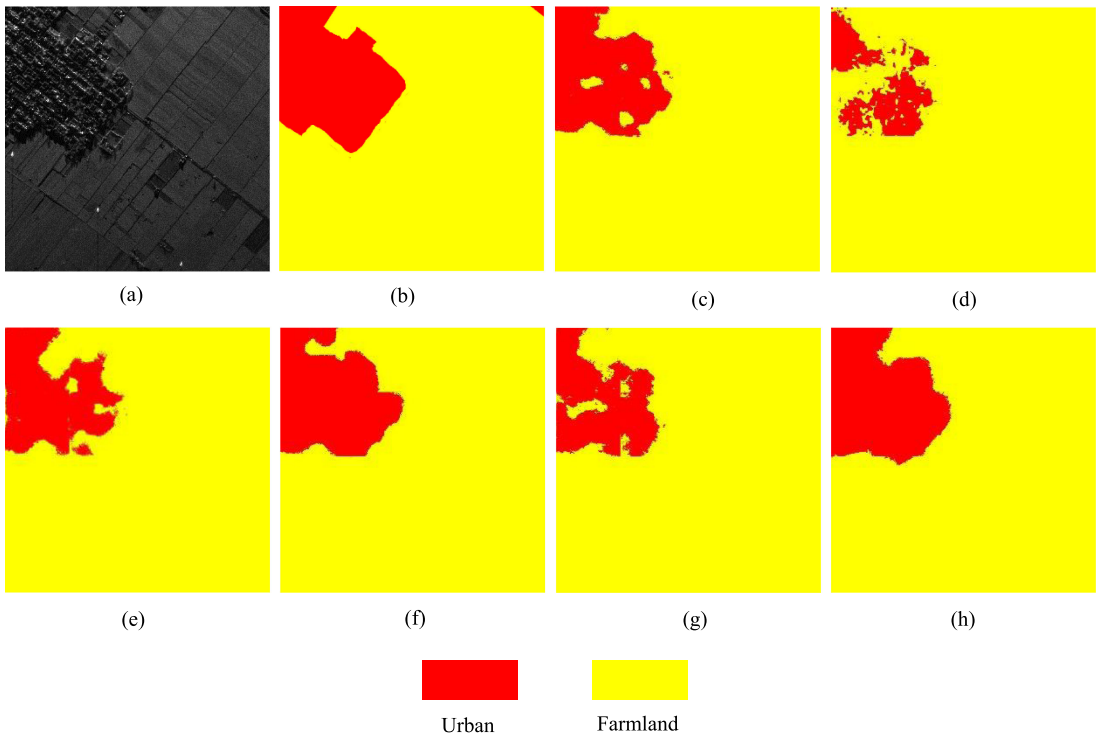

Fig. 9. Segmentation results of different methods. The fully connected CRF is utilized for optimizing the segmentation results of different methods (a) Test image. (b) Ground-truth. (c) U-Net. (d) DeepLab. (e) SegNet. (f) PSPNet. (g) FRRNet. (h) Our method.

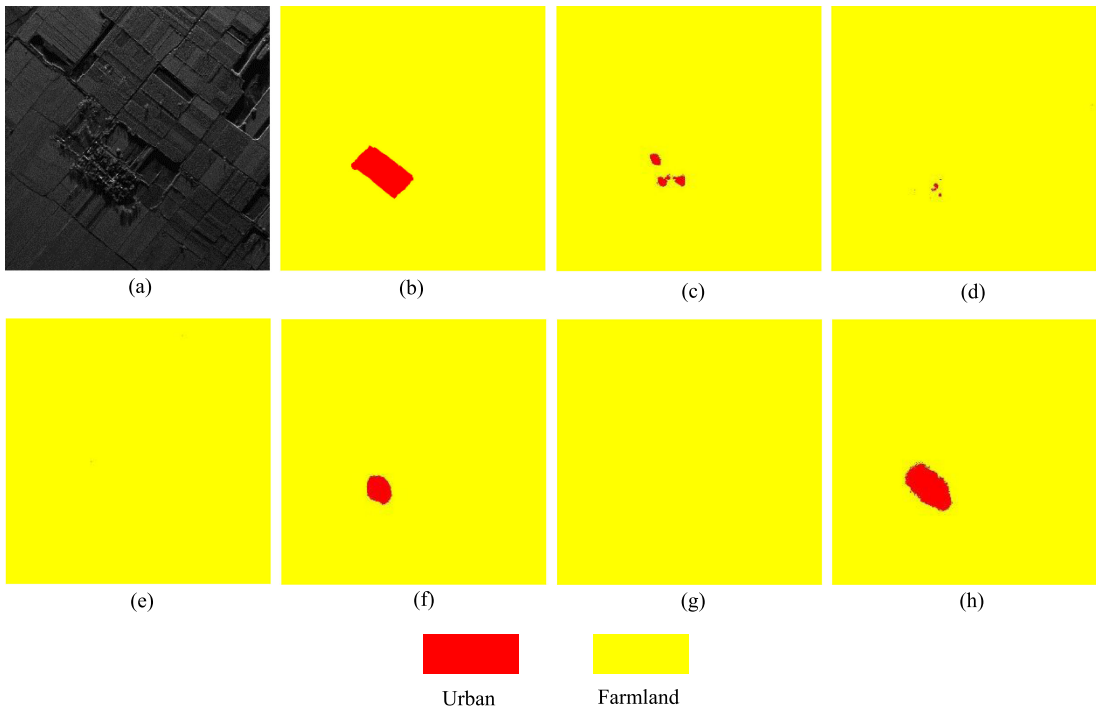

Fig. 10. Segmentation results of different methods. The fully connected CRF is utilized for optimizing the segmentation results of different methods (a) Test image. (b) Ground-truth. (c) U-Net. (d) DeepLab. (e) SegNet. (f) PSPNet. (g) FRRNet. (h) Our method.

TABLE V
SEGMENTATION PERFORMANCE OF AFCN WHEN IT IS EMBEDDED WITH DIFFERENT ATTENTION MODULES

Modules	MIoU	FWIoU	PA	MPA
CBAM	64.95%	75.01%	85.64%	73.78%
SENet	69.49%	78.09%	87.69%	77.83%
SKNet	69.88%	78.03%	87.49%	78.85%
MANet	70.61%	78.60%	87.99%	78.97%

TABLE VI
SEGMENTATION PERFORMANCE OF DIFFERENT FEATURE OPTIMIZATION STRATEGIES

Strategies	MIoU	FWIoU	PA	MPA
Baseline	69.71%	77.89%	87.49%	78.58%
Multi-scale	70.01%	78.04%	87.54%	78.86%
Channel	69.97%	78.37%	87.85%	78.51%
Spatial	70.35%	78.23%	87.63%	78.61%
Multi-scale + Channel + Spatial	70.61%	78.60%	87.99%	78.97%

"Baseline" denotes the AFCN model without feature optimization strategy. "Multiscale," "Channel," and "Spatial" denote three different feature optimization strategies.

TABLE VII
SEGMENTATION PERFORMANCE OF AFCN UNDER DIFFERENT LOSS FUNCTIONS

Loss	MIoU	FWIoU	PA	MPA
Cross-entropy	70.61%	78.60%	87.99%	78.97%
Focal	70.81%	78.50%	87.76%	79.05%
Dice	68.25%	76.66%	86.72%	77.67%
Lovasz-softmax	70.93%	78.77%	88.08%	79.04%
Ours	71.49%	79.17%	88.36%	79.12%

pooling operations to obtain attention values. However, the brightness of different regions in SAR images varies greatly. Thus maximum pooling operation will remove the features of low-brightness regions, thereby degrading the segmentation performance of CBAM. In contrast, MANet only utilizes average pooling operation to derive attention values. Hence the features of the low-brightness regions are well preserved, which improves the segmentation performance. Moreover, the performance of MANet is better than that of SENet. This is because MANet uses two convolution kernels of different sizes for extracting multiscale features, which effectively improves the representation of image features. Although SKNet adopts a dynamic selection mechanism to extract multiscale features, its performance is inferior to MANet. The reason is that SKNet only uses channel attention whereas MANet calculates both the spatial and channel attention for enhancing the feature extraction.

MANet is composed of three feature optimization strategies: Multiscale feature, spatial attention, and channel attention extraction. Next, we verify the effectiveness of the three strategies. In Table VI , the "Baseline" refers to the AFCN model without a feature optimization strategy. As can be seen, the three strategies outperform the "Baseline" in MIoU, FWIoU, PA, and MPA. In addition, we obtain a further performance improvement by combining the three strategies.

D. Evaluation of the Loss Function

In this article, a new loss function is designed for AFCN. We compare the new loss with cross-entropy loss, Lovasz-Softmax loss [44], focal loss [45], and dice loss [46]. The cross-entropy loss is used to optimize the pixel classification accuracy in the training process. The focal loss is capable of alleviating the category imbalance problem. Dice and Lovasz-Softmax losses are used for optimizing the IoU of the segmentation results. The difference between the two loss functions is that the former is a discrete loss whereas the latter is a continuous loss. We perform the experiments on the Fangchenggang dataset, and the weight coefficient α in our loss function is set as 7. The segmentation performance of AFCN under different loss functions is as follows.

It is obvious that the segmentation performance of our loss function outperforms that of the other loss functions. The MIoU of our loss function is 0.88% higher than the cross-entropy loss and 0.68% higher than the focal loss. The reason is that our loss function contains the Lovasz-softmax component that can effectively optimize the IoU of the segmentation results. Although dice loss is designed for optimizing IoU, it is a discrete loss which makes the training process unstable. In contrast, our loss function can optimize the IoU in a continuous manner, thereby improving the performance of the AFCN model. In addition, our loss function contains the cross-entropy component which can optimize the pixel classification accuracy. As a result, the segmentation results of our loss function outperform those of the Lovasz-Softmax loss.

In this article, the new loss function consists of Lovasz-Softmax and cross-entropy losses, wherein the former component is multiplied by the weight coefficient α. Next, the influence of α on the performance of AFCN is discussed. We, respectively, set α to 1, 3, 5, 7, 9, 11, and 13. Then we calculate MIoU and PA of AFCN with different α, as shown in Fig. 11. As α increases, the performance of AFCN improves, and we obtain the best MIoU (71.49%) and PA (88.36%) when α is set to 7. As α continues to increase, MIoU and PA of AFCN become worse. It is obvious that setting an appropriate α for the proposed loss function can further improve the performance of AFCN. Thus, the weight coefficient α is set to 7 in the experiments.

E. Evaluation of the Fully Connected CRF

In our method, the fully connected CRF is adopted for improving the segmentation performance of AFCN. Next, we compare the segmentation results before and after the CRF processing. In this section, the experiments are performed on the Fangchenggang dataset. As shown in Table VIII, MIoU, FWIoU, PA, and MPA of AFCN are improved after the CRF processing. This is because the fully connected CRF utilizes the spatial information in the input images by capturing the correlation of pixels, which improves the segmentation accuracy. We choose two images in the Fangchenggang testing set and draw the segmentation results before and after the CRF processing. As

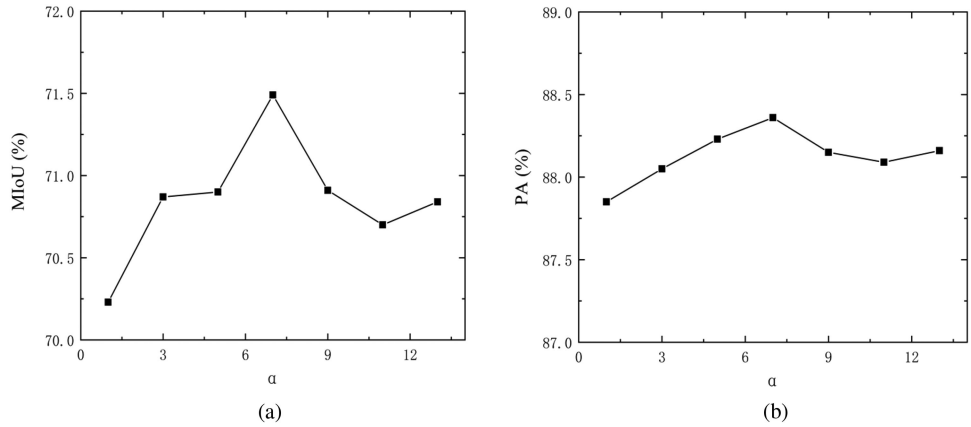

Fig. 11. MIoU and PA of AFCN with different weight coefficient in the proposed loss function (a) MIoU. (b) PA.

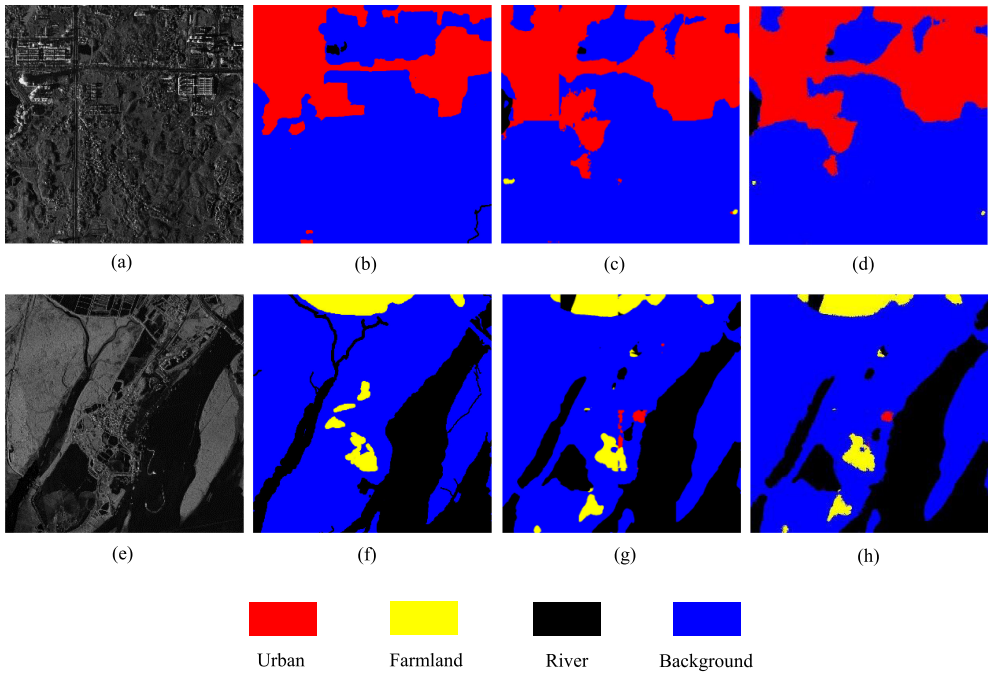

Fig. 12. Segmentation results of our method before and after CRF processing (a) Test image. (b) Ground-truth. (c) Before CRF. (d) After CRF. (e) Test image. (f) Ground-truth. (g) Before CRF. (h) After CRF.

TABLE VIII
SEGMENTATION PERFORMANCE OF OUR METHOD BEFORE AND AFTER CRF PROCESSING

	MIoU	FWIoU	PA	MPA
Before CRF	71.49%	79.17%	88.36%	79.12%
After CRF	71.85%	79.65%	88.69%	79.49%

Fig. 12 shows, the number of the misclassified pixels after CRF processing is less than that before the CRF processing. Besides, as can be seen in Fig. 12(d) and (h), the segmentation results at the slice boundaries become visually smoother after the CRF processing.

F. Computational Efficiency

To compare the computational efficiency of our method with that of the other methods, we calculate the running time required for different methods to segment all the images in the Pucheng testing set. The experiments are implemented with the Pytorch framework and the GPU of our computer is GeForce GTX 1070 with 8GB memory. The running time of different methods before and after the CRF processing is shown in TABLE IX. As can be seen, our method has the shortest running time among the six methods (about 8.9 s before CRF processing). Because of the deep structures, PSPNet, DeepLab model, and FRRNet require more running time. Besides, the running time required for CRF processing for different methods is almost the same (around 92 s).

TABLE IX
RUNNING TIME OF DIFFERENT METHODS TO SEGMENT ALL THE IMAGES IN THE PUCHENG TESTING SET

Methods	Before CRF	After CRF
PSPNet	28.4s	120.3s
DeepLab	26.4s	117.9s
FRRNet	19.8s	111.7s
SegNet	9.3s	102.7s
U-Net	9.1s	101.9s
Ours	8.9s	101.7s

IV. CONCLUSION

We proposed the AFCN model for SAR image segmentation. In the structure of AFCN, each convolution module is followed by a MANet which is utilized for enhancing the extraction of the SAR image features. MANet consists of three feature optimization strategies: Multiscale feature, channel attention, and spatial attention extraction. Besides, a new loss function was designed for AFCN by integrating Lovasz-Softmax and cross-entropy losses. This new loss can simultaneously optimize IoU and the pixel classification accuracy. To further improve the segmentation performance of AFCN, we adopted the fully connected CRF to capture the spatial information in the SAR images. The experiments which were performed on two airborne SAR image datasets prove that our method effectively improved the segmentation accuracy of the SAR images. For example, PA of our method achieves 95.79% in the Pucheng dataset, which is superior to that of the other methods such as SegNet, PSPNet, and FRRNet.

ACKNOWLEDGMENT

The Authors would like to thank the Beijing Institute of Radio Measurement for the SAR datasets used in the experiments.

REFERENCES

[1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci Remote Sens Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.

[2] H. Chen, F. Zhang, B. Tang, Q. Yin, and X. Sun, "Slim and efficienct neural network design for resource-constrained SAR target recognition," *Remote Sens.*, vol. 10, no. 10, Oct. 2018, Art. no. 1618.

[3] F. Gao, T. Huang, J. Sun, J. Wang, A. Hussain, and E. Yang, "A new algorithm of SAR image target recognition based on improved deep convolutional neural network," *Cogn. Computation*, vol. 11, no. 6, pp. 809–824, Dec. 2019.

[4] F. Gao, F. Ma, J. Wang, J. Sun, and H. Zhou, "Visual saliency modeling for river detection in high-resolution SAR imagery," *IEEE Access*, vol. 6, pp. 1000–1014, Nov. 2017.

[5] J. Geng, J. Fan, H. Wang, X. Ma, B. Li, and F. Chen, "High-resolution SAR image classification via deep convolutional autoencoders," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2351–2355, Nov. 2015.

[6] M. Ma, J. Liang, M. Guo, Y. Fan, and Y. Yin, "SAR image segmentation based on artificial bee colony algorithm," *Appl. Soft. Comput.*, vol. 11, no. 8, pp. 5205–5214, Dec. 2011.

[7] M. Li, Y. Wu, and Q. Zhang, "SAR image segmentation based on mixture context and wavelet hidden-class-label Markov random field," *Comput. Math. Appl.*, vol. 57, no. 6, pp. 961–969, Mar. 2009.

[8] H. B. Kekre and S. Gharge, "SAR image segmentation using co-occurrence matrix and slope magnitude," in *Proc. Int. Conf. Adv. Comput., Commun. Control*, Mumbai, India, 2009, pp. 368–372.

[9] J. Zhao, W. Guo, S. Cui, Z. Zhang, and W. Yu, "Convolutional neural network for SAR image classification at patch level," in *Proc. IEEE. Int. Geosci. Remote Sens. Symp.*, Beijing, China, 2016, pp. 945–948.

[10] Y. L. Huang, B. B. Xu, and S. Y. Ren, "Analysis and pinning control for passivity of coupled reaction-diffusion neural networks with nonlinear coupling," *Neurocomputing*, vol. 272, pp. 334–342, Jan. 2018.

[11] X. Hong *et al.*, "Component-based feature saliency for clustering," *IEEE Trans. Knowl. Data Eng.*, to be published, doi: 10.1109/TKDE.2019.2936847.

[12] J. Zabalza *et al.*, "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging," *Neurocomputing*, vol. 185, pp. 1–10, Apr. 2016.

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

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, Munich, Germany, 2015, pp. 234–241.

[15] 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.

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

[17] T. Pohlen, A. Hermans, M. Mathias, and B. Leibe, "Full-resolution residual networks for semantic segmentation in street scenes," in *Proc IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, USA, 2017, pp. 4151–4160.

[18] 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. Photogramm. Remote Sens.*, vol. 151, pp. 223–236, May 2019.

[19] 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.

[20] 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.

[21] P. Shamsolmoali, M. Zareapoor, R. Wang, H. Zhou, and J. Yang, "A novel deep structure U-Net for sea-land segmentation in remote sensing images," *IEEE J. Sel. Top. Appl. Earth Observ.*, vol. 12, no. 9, pp. 3219–3232, Sep. 2019.

[22] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, USA, 2017, pp. 1492–1500.

[23] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 1–9.

[24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, USA, 2016, pp. 2818–2826.

[25] 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.*, San Francisco, USA, 2017, pp. 4278–4284.

[26] Z. Yue *et al.*, "A novel semi-supervised convolutional neural network method for synthetic aperture radar image recognition," *Cogn. Comput.*, Mar. 2019. [Online]. Available: https://doi.org/10.1007/s12559-019-09639-x

[27] L. Itti and C. Koch, "Computational modelling of visual attention," *Nat. Rev. Neurosci.*, vol. 2, pp. 194–203, Mar. 2001.

[28] F. Gao, A. Liu, K. Liu, E. Yang, and A. Hussain, "A novel visual attention method for target detection from synthetic aperture radar (SAR) images," *Chin. J. Aeronaut.*, vol. 32, no. 8, pp. 1946–1958, Aug. 2019.

[29] M. Corbetta and G. L. Shulman, "Control of goal-directed and stimulus-driven attention in the brain," *Nat. Rev. Neurosci.*, vol. 3, pp. 201–215, Mar. 2002.

[30] J. Park, S. Woo, J. Y. Lee, and I. S. Kweon, "BAM: Bottleneck attention module," 2018.

[31] F. Wang *et al.*, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3156–3164.

[32] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[33] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 510–519.

[34] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, Granada, Spain, 2018, pp. 421–429.
[35] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Munich, Germany, 2018, pp. 3–19.
[36] L. Bombrun, G. Vasile, M. Gay, and F. Totir, "Hierarchical segmentation of polarimetric SAR images using heterogeneous clutter models," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 726–737, Feb. 2011.
[37] F. Ma *et al.*, "Attention graph convolution network for image segmentation in big SAR imagery data," *Remote Sens*., vol. 11, no. 21, Nov. 2019, Art. no. 2586.
[38] 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, Apr. 2017.
[39] L. Wan, T. Zhang, Y. Xiang, and H. You, "A robust fuzzy c-means algorithm based on Bayesian nonlocal spatial information for SAR image segmentation," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 896–906, Mar. 2018.
[40] F. Liu, G. Lin, and C. Shen, "CRF learning with CNN features for image segmentation," *Pattern Recognit.*, vol. 48, no. 10, pp. 2983–2992, Oct. 2015.
[41] H. Zhou, J. Zhang, J. Lei, S. Li, and D. Tu, "Image semantic segmentation based on FCN-CRF model," in *Proc. Int. Conf. Image, Vis. Comput.*, Portsmouth, U.K., 2016, pp. 9–14.
[42] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
[43] F. Ma, F. Gao, J. Sun, H. Zhou, and A. Hussain, "Weakly supervised segmentation of SAR imagery using superpixel and hierarchically adversarial CRF," *Remote Sens.*, vol. 11, no. 5, Mar. 2019, Art. no. 512.
[44] M. Berman, A. Rannen Triki, and M. B. Blaschko, "The Lovász-Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4413–4421.
[45] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Venice, Italy, 2017, pp. 2999–3007.
[46] T. A. Soomro, A. J. Afifi, J. Gao, O. Hellwich, M. Paul, and L. Zheng, "Strided U-Net model: Retinal vessels segmentation using dice loss," in *Proc. Int. Conf. Digit. Image Comput., Tech. Appl.*, Canberra, Australia, 2018, pp. 1–8.

Zhenyu Yue received the B.S. degree in electronic and electrical engineering for civil aviation from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2016. He is currently working toward the Ph.D. degree in circuits and systems at the Beijing University of Aeronautics and Astronautics.

His research interests include radar signal processing, machine learning, image segmentation, and image recognition.

Fei Gao received the B.S. degree in industrial electrical automation and the M.S. degree in electromagnetic measurement technology and instrument from the Xi'an Petroleum Institute, Xi'an, China, in 1996 and 1999, respectively, and the Ph.D. degree in signal and information processing from the Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China, in 2005.

He is currently an Associate Professor with the School of Electronic and Information Engineering, BUAA. His research interests include radar signal processing, moving target detection, and image processing.

Qingxu Xiong (Member, IEEE) received the Ph.D. degree in electrical engineering from Peking University, Beijing, China, in 1994.

From 1994 to 1997, he worked with the Information Engineering Department at the Beijing University of Posts and Telecommunications, Beijing, China, as a Postdoctoral Researcher. He is currently a Professor with the School of Electrical and Information Engineering at the Beijing University of Aeronautics and Astronautics (BUAA). His research interests include high performance switching, performance modeling of networks, and semantic communication networks.

Jun Wang received the B.S. degree in electronic engineering from the North Western Polytechnical University, Xi'an, China, in 1995 and the M.S. and Ph.D. degrees in information and communication engineering from the Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China, in 1998 and 2001, respectively.

He is currently a Professor with the School of Electronic and Information Engineering, BUAA. His research interests include signal processing, DSP/FPGA real-time architecture, target recognition and tracking, and so on.

Amir Hussain received the B.Eng. and the Ph.D. degrees in electronic and electrical engineering from University of Strathclyde, Scotland, U.K., in 1992 and 1997, respectively.

Following Postdoctoral and Academic positions with University of West of Scotland (1996–98), Paisley, U.K., University of Dundee (1998–2000), Dundee, U.K., and University of Stirling (2000–18), Stirling, U.K., respectively, he is currently a Professor and Founding Head of the Cognitive Big Data and Cybersecurity (CogBiD) Research Lab at Edinburgh Napier University, U.K. His research interests include cognitive computation, machine learning, and computer vision.

Huiyu Zhou received the B.Eng. degree in radio technology from Huazhong University of Science and Technology, Wuhan, China, the M.S. degree in biomedical engineering from University of Dundee, Dundee, U.K., and the Ph.D. degree in computer vision from Heriot-Watt University, Edinburgh, U.K.

He is currently a Professor with the School of Informatics, University of Leicester, Leicester, U.K. His research interests include medical image processing, computer vision, intelligent systems, and data mining.