

Reverse Attention Dual-Stream Network for Extracting Laver Aquaculture Areas From GF-1 Remote Sensing Images

Binge Cui, Yanli Zhao , Mingkai Yang , Ling Huang, and Yan Lu 

Abstract—Extracting laver aquaculture areas from remote sensing images is very important for laver aquaculture monitoring and scientific management. However, due to the large differences in spectral features of laver aquaculture areas caused by factors such as different growth stages and harvesting conditions, traditional machine learning and deep learning methods face great challenges in achieving accurate and complete extraction of raft laver aquaculture areas. In this article, a reverse attention dual-stream network (RADNet) is proposed for the extraction of laver aquaculture areas with weak spectral responses by comprehensively considering both the aquaculture boundary and surrounding sea background information. RADNet consists of a boundary stream and a segmentation stream. Considering the weaker spectral responses of certain laver aquaculture areas, we introduce a reverse attention module in the segmentation stream to amplify the weaker responses of inapparent laver aquaculture areas. To suppress the response of nonboundary details in the boundary stream, we design a boundary attention module, which is guided by high-level semantics from the segmentation stream. The structural information of the laver aquaculture area learned from the boundary stream will be fed back to the segmentation stream through a specially designed boundary guidance module. The study is conducted in Haizhou Bay, China, and is verified using a self-labeled GF-1 multispectral dataset. The experimental results show that RADNet model performs better in extracting inapparent laver aquaculture areas compared to SOTA models.

Index Terms—Dual-stream network, raft aquaculture areas, reverse attention.

I. INTRODUCTION

LAVER aquaculture is an important part of the coastal marine economy and is of great significance to farmers and fishermen in increasing their production and income. However, the rapid growth of raft aquaculture areas has also caused marine ecological environmental problems, such as deterioration of water quality due to inadequate water body exchange. In addition, the scattered distribution of large-scale aquaculture

floating ropes has created inconveniences for marine traffic and port transportation [1]. Therefore, the dynamic monitoring of raft aquaculture is important for the ecological environment protection of near-coastal areas and the sustainable development of the local aquaculture industry.

In recent years, scholars have performed much research on the use of remote sensing technology to monitor aquaculture areas. Wu et al. [1] proposed a constrained energy minimization method based on orthogonal subspace projection to enhance the aquaculture area features for accurate extraction of off-shore aquaculture areas in complex water color backgrounds. Cheng et al. [2] proposed a threshold segmentation method combined with a gray-level co-occurrence matrix, which fused spectral and texture features to achieve aquaculture area extraction from GF-2 images. However, these methods require manual turning of parameters, and thus, their generalization ability is weak, especially when applied to complex shallow marine environments [3].

Deep convolutional neural networks can avoid frequent parameter tuning by learning deep features of the target object [4]. Liu et al. [5] introduced a richer convolutional feature [6] network to efficiently extract the boundaries of raft aquaculture areas in Sanduao, China. Cui et al. [7] improved the decoder part of U-Net and proposed a pyramid upsampling and squeeze-excitation structure to capture the context and edge information of aquaculture areas, which effectively alleviated the adhesion problem in laver aquaculture area extraction. Shi et al. [3] proposed a homogeneous convolutional neural network (HCN) for extracting raft aquaculture areas from GF-1 images, in which a dual-scale structure (DS-HCN) was designed to integrate high-level contextual information. Lu et al. [8] improved U-Net by using an ASPP structure and introducing flow alignment modules, which can correct the semantic misalignment and reduce “adhesion” of aquaculture areas in the extraction results. However, Liu’s method is prone to influence by the complex shallow sea environment, and the boundaries of the extracted aquaculture areas are easily broken. The boundaries of some rafted laver aquaculture areas extracted by Cui’s method are excessively smooth. When extracting aquaculture areas that are not obvious in the images, Shi’s and Lu’s methods can easily miss the aquaculture areas or extract incomplete areas.

Thanks to the ability to effectively emphasize important features of an image and suppress useless information, attention mechanisms have now been combined with deep convolution

Manuscript received 22 December 2022; revised 6 March 2023 and 25 April 2023; accepted 29 May 2023. Date of publication 1 June 2023; date of current version 15 June 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 42276185 and in part by the Shandong Province Natural Science Foundation of China under Grant ZR2020MD096 and Grant ZR2020MD099. (Corresponding author: Yan Lu.)

The authors are with the College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China (e-mail: cuibinge@sdust.edu.cn; 2365855406@qq.com; 18865386835@163.com; sdythl@126.com; luyan@sdust.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3281823

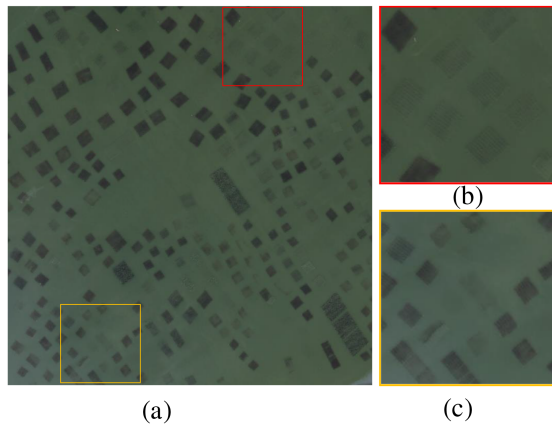


Fig. 1. Raft laver aquaculture areas in RSI. (a) GF-1 RSIs acquired near the coast of Lianyungang on February 17, 2017. The selected red and yellow box regions in the left image are enlarged in the right image. Region (b) shows raft laver aquaculture areas after harvesting. Region (c) shows raft laver aquaculture areas with high suspended sediment concentrations.

neural networks to solve a variety of deep learning tasks. Peng et al. [9] proposed a difference-enhanced dense-attention convolution neural network, which can be used for end-to-end change detection of bitemporal remote sensing images (RSIs). Li et al. [10] proposed a graph-feature-enhanced selective assignment network (GSANet) for hyperspectral and multispectral images fusion, in which an SFAM module was developed for adaptive fusion of hyperspectral and multispectral information. Shi et al. [11] proposed a centerness-aware network for object detection in RSIs, in which the center of objects with symmetrical shape is highlighted through the attention mechanism. Sun et al. [12] proposed a multistructure KELM algorithm with attention fusion strategy for HSI classification, in which a weighted self-attention fusion strategy was proposed for efficient fusion of multibranch KELM classification results. Moreover, they also proposed a successive pooling attention network for RSI segmentation, in which an SAPM module was proposed to extract salient features of images [13]. In these studies, the attention mechanism plays an important role in identifying and highlighting more discriminating features.

Raft laver aquaculture areas are characterized by large quantities, dense distribution, and complex spectra, as shown in Fig. 1. These characteristics render it challenging to accurately extract the aquaculture areas. In Fig. 1(a), most of the raft laver aquaculture areas appear black and can be easily distinguished from the seawater background. However, some raft laver aquaculture areas are not obvious in the RSIs, as shown in the red and yellow boxes of Fig. 1(a). Inapparent raft laver aquaculture areas have two main causes. First, affected by the harvest of laver, the chlorophyll content in the laver aquaculture area may be greatly reduced, narrowing the difference between the spectral characteristics of the raft laver aquaculture area and the surrounding seawater, as shown in Fig. 1(b). Second, in the nearshore area, the concentration of suspended sediment is high, which increases the reflectance of both the aquaculture area and the surrounding seawater and reduces the spectral difference between them, as shown in Fig. 1(c). The above factors hinder

the accurate extraction of inapparent rafted laver aquaculture areas.

In this article, a reverse attention dual-stream network (RAD-Net) was proposed for the extraction of rafted laver aquaculture areas in complex marine environments. The reverse attention mechanism proposed in [14] can amplify the weaker responses of the target objects. Inspired by this finding, we designed a reverse attention module (RAM) to learn features of both inapparent aquaculture areas and obvious aquaculture areas by suppressing features of seawater. In addition, considerable research [15], [16], [17], [18], [19], [20], [21] has demonstrated that dual-stream networks combining edge detection and semantic segmentation effectively utilize the boundary information of objects and significantly improve the boundaries of segmentation results. To obtain an accurate boundary of the aquaculture area, we designed a boundary attention module (BAM) using global semantic information to avoid the interference of nonboundary information on boundary extraction. Then, we designed a boundary guidance module (BGM) to enhance the boundary response of the semantic segmentation results of the raft laver aquaculture area. To evaluate the performance of RAD-Net for raft laver aquaculture area extraction, experiments on the created GF-1 dataset were carried out. Compared with other models, RADNet achieves higher F1-score and intersection over union (IoU) values for raft laver aquaculture area extraction.

The main contributions of this article are presented as follows.

- 1) We designed a novel RAM to enhance the spectral response of raft laver aquaculture areas, which was particularly beneficial for distinguishing aquaculture areas that were not obvious in RSIs from the surrounding seawater.
- 2) We designed a BAM to suppress the response of nonboundary details and designed a BGM to incorporate the refined boundary information into the segmentation stream to further restore the inherent shape of the laver raft aquaculture areas.
- 3) This article collects and releases a new GF-1 RSI dataset for aquaculture area extraction (https://github.com/cuibinge/RAA_dataset.git). Specifically, the dataset contains image blocks with pixel-level labels for laver aquaculture area and seawater, covering 264 651 750 pixels in Haizhou Bay, Lianyungang, Jiangsu province, China.

II. RELATED WORK

A. Multitask Learning

Multitask learning (MTL) improves model generalization and robustness by sharing representations across multiple tasks [19]. In the context of deep learning, MTL is typically performed via hard or soft parameter sharing [22]. Hard parameter sharing means that multiple tasks share the same few hidden layers of the network in the encoding part, and they start forking to perform different tasks near the decoding part of the network, as shown in Fig. 2(a). Soft parameter sharing means building independent models for different tasks, but using a loss function to constrain the distance between the individual model parameters, as shown in Fig. 2(b). MTL has been widely utilized in various computer

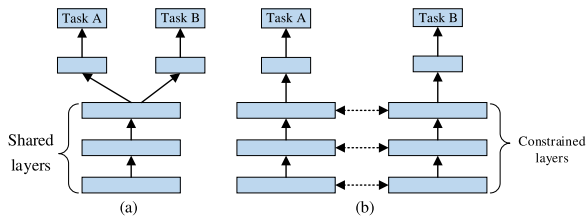


Fig. 2. Parameter sharing for MLT in deep neural networks. (a) Hard parameter sharing. (b) Soft parameter sharing.

vision applications, including semantic segmentation [15], [16], [17], [18], [19], [20], [21] and object detection [23].

Several approaches have been proposed to combine the semantic segmentation task with the edge detection task to refine the boundaries of the segmentation results [24], [25], [26], [27]. Yu et al. [24] designed a multitask network that shares a single encoder but uses different decoders to perform two independent tasks. Jing et al. [25] designed a boundary–semantic interaction module to achieve mutual guidance between the boundary detection task and the semantic segmentation task. To improve the detection performance for small and thin objects, Takikawa et al. [26] designed a dual-stream semantic segmentation network, where the shape stream was dedicated to processing boundary information and the learned boundaries were applied as intermediate representations to help the regular stream. Pan et al. [27] discovered that the pixels around the boundary are easily misclassified. To emphasize the error-prone boundary pixels, they designed an edge region detection module and incorporated the detected edge regions into the semantic segmentation task. Compared to single-task learning, MTL can obtain additional useful information by mining the relationships between two tasks, which often improves the performance of the model.

B. Attention Mechanism

The attention mechanism is a method of shifting attention to the most important regions of an image while ignoring irrelevant regions [28]. Attention modules have been widely utilized in various deep learning-based tasks, as they improve performance by introducing a small number of network parameters [18]. Typical attention mechanisms include channel attention, spatial attention, channel and spatial attention, self-attention, etc. SENet [29] is the earliest work on the channel attention mechanism, which employed the squeeze-and-excitation (SE) block to adaptively adjust the weight of the channels by modeling the relationships among them. ECA-Net [30] replaced the FC layer in the SE block using a one-dimensional convolution capable of adaptively resizing the kernel to reduce the number of network parameters. The spatial attention mechanism focuses on generating attention weights from spatial patches of the feature maps rather than the channels [37]. GENet [31] and the reverse attention network (RAN) [14] are representatives of spatial attention. Inspired by SENet, Hu et al. designed GENet to capture remote spatial contextual information by providing recalibration functions in the spatial domain. In the RAN, an attention mask is designed to highlight the prediction of the

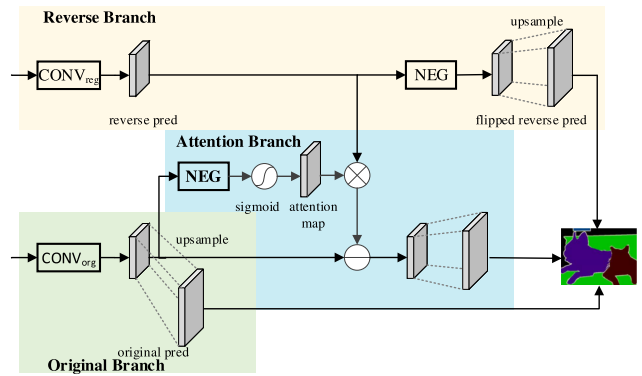


Fig. 3. System diagram of the RAN.

reverse object class, which is then subtracted from the original prediction to correct errors in the confusion region of the semantic segmentation [14]. The structure of reverse attention in the RAN is shown in Fig. 3. Channel and spatial attention mechanism combines the advantages of channel attention and spatial attention [28]. The CBAM [32] and BAM [33] are typical works that introduced channel and spatial attention to consider effective information among channels and within channels. The coordinate attention (CA) mechanism inherited the advantage of channel attention methods that model interchannel relationships and captured long-range dependencies with precise positional information [34]. Self-attention mechanism has shown great potential for capturing global context. DANet [36] captured global dependence based on self-attentive mechanisms in spatial and channel dimensions and achieved more accurate segmentation results.

III. PROPOSED METHOD

In this section, we describe the overall structure of RADNet and, then, describe in detail the functionality and structure of the three proposed modules.

A. Overall Structure of RADNet

RADNet consists of a segmentation stream and a boundary stream, as shown in Fig. 4.

The segmentation stream employs the classic U-Net structure of U-Net. It is worth noting that the double convolution layers in the encoder is replaced by a residual structure with batch normalization, as shown in Fig. 5. To enhance the model's ability to distinguish the inconspicuous raft laver aquaculture areas from the background seawater, a RAM is designed to replace the double convolution operation in each layer of the U-Net decoder. Moreover, the last two layers of the decoder use the BGM to incorporate the boundary map into the segmentation stream to enhance the boundary response of the raft laver aquaculture areas.

The boundary stream is aimed at obtaining a refined boundary map of the raft laver aquaculture areas. A 7×7 convolution operation is used to extract local features of the RSIs. Considering that the two of boundary detection and semantic segmentation tasks are closely related, the boundary stream and segmentation

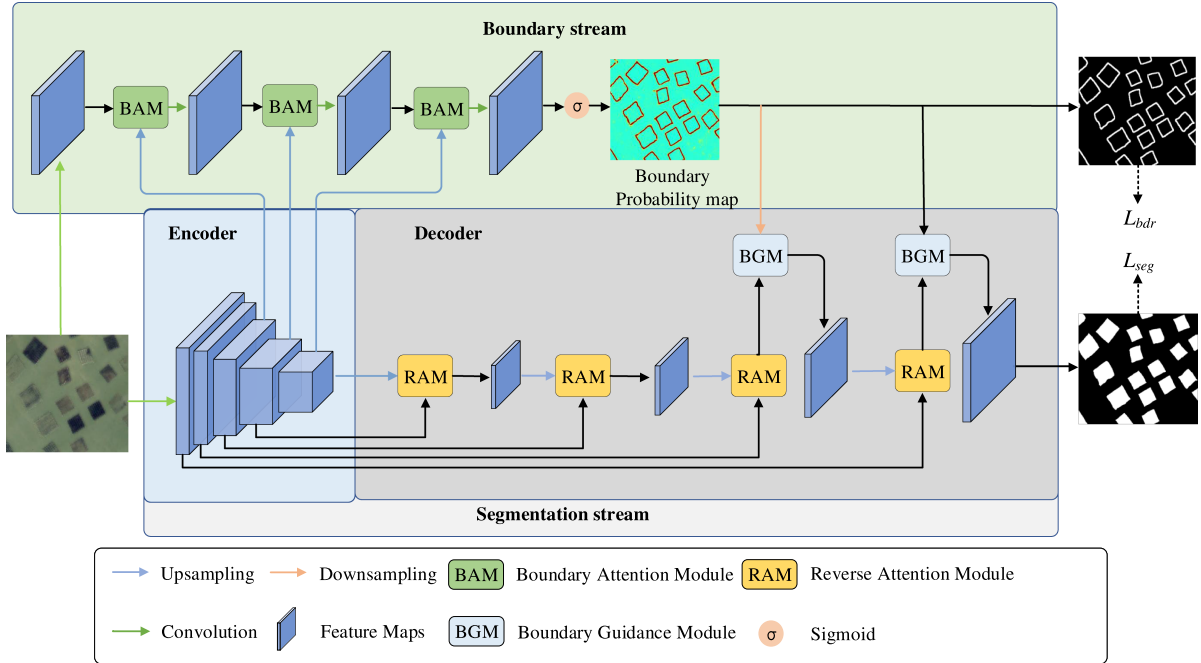


Fig. 4. Overview of the proposed RADNet. The input of the RADNet is a true color image after panchromatic multispectral fusion, and the output contains a boundary detection map and a semantic segmentation map.

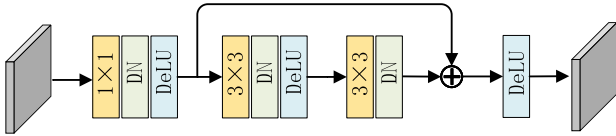


Fig. 5. Residual structure of RADNet. “ 1×1 ” and “ 3×3 ” denote convolution operations with convolution kernel sizes of 1 and 3, respectively. “BN” denotes the batch normalization layer.

stream of RADNet share the same encoder. The BAM is proposed to strengthen the features associated with the aquaculture area boundaries. By receiving high-level semantic information from the segmentation stream encoder, the BAM can effectively suppress the texture information inside the aquaculture area and the seawater. The output of the last BAM is fed to a sigmoid layer for the boundary probability prediction of the aquaculture area.

B. Reverse Attention Module

The structure of RAM is shown in Fig. 6. First, the feature maps F_{in}^l and F_{in}^{l+1} from $l+1$ and l layers are concatenated, and then, convolution operations are performed to fuse the information from different layers. The fused feature map M is fed into two separate branches.

The first branch is trained to learn explicitly the knowledge of seawater, which is the object to be excluded from the extraction results of the aquaculture area. Following the work in [14], we introduce a NEG operation for flipping the sign of the pixel values of the input feature map. The flipped feature map will undergo a 1×1 convolution and 3×3 convolution to learn the features of the seawater. Last, we performed another NEG

operation to obtain the features of the nonseawater area in the image. Mathematically, the feature map S of the nonseawater area can be written as

$$S = \text{NEG} (C_{3 \times 3} (C_{1 \times 1} (\text{NEG}(M)))) \quad (1)$$

where $C_{1 \times 1}(\cdot)$ and $C_{3 \times 3}(\cdot)$ denote convolution operations with kernel sizes of 1×1 and 3×3 , respectively.

The second branch focuses on learning the features of the inapparent part of the aquaculture area. First, a $1 - \sigma$ layer is used to obtain the reverse attention map. The pixel values in the reverse attention map are small in areas with obvious laver aquaculture features and large in areas with inapparent laver aquaculture features and in seawater. Second, the reverse attention map is multiplied by the nonseawater feature map S to obtain a feature map Q that contains only the inapparent part of the aquaculture area. Last, a residual connection is utilized to combine the feature maps M and Q to obtain a salient feature map of the intact aquaculture area, followed by a normalized 3×3 convolution layer. The module output F_{out} is calculated as follows:

$$Q = S \odot (1 - \sigma(M)) \quad (2)$$

$$F_{out} = C_{3 \times 3}(M + Q) \quad (3)$$

where $\sigma(\cdot)$ denotes the sigmoid function.

C. Boundary Attention Module

Fig. 7 shows the detailed structure of the BAM. By sharing semantic information from the segmentation stream encoder, the BAM is designed to enhance the boundary responses in the boundary feature map and refine the boundaries of the aquaculture area. The BAM has two inputs B_{in} and F_{in} , where B_{in} is the

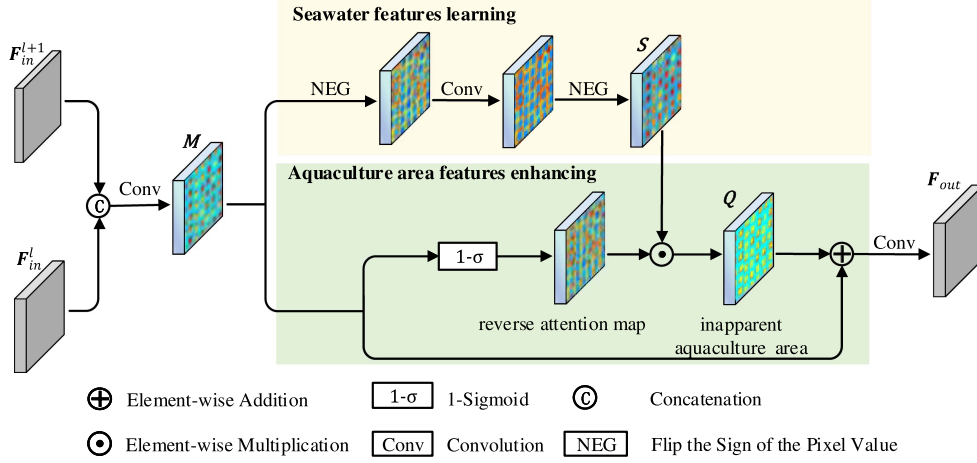


Fig. 6. Structure of the RAM.

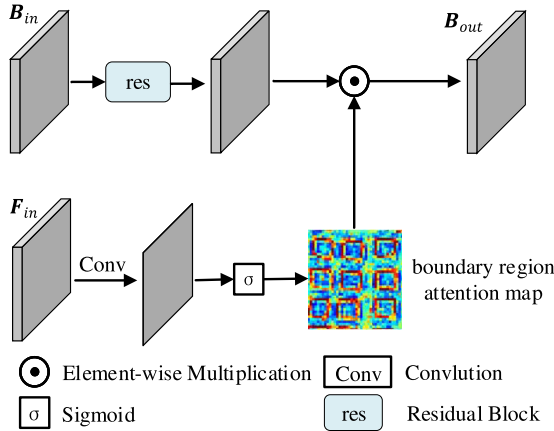


Fig. 7. Structure of the BAM.

boundary feature map from the previous layer of the boundary stream, and F_{in} is the high-level semantic feature map from the last three layers of the segmentation stream encoder. First, a residual block is applied to extract the rich edges in the feature map B_{in} . Second, a boundary region attention map is generated by feeding the feature map F_{in} into a 1×1 convolutional layer with a sigmoid activation function. Third, we combined the boundary region attention map with the edge response map using the elementwise multiplication to highlight the edges around the aquaculture area boundary. The module output B_{out} is calculated as follows:

$$B_{out} = \sigma(C_{1 \times 1}(F_{in})) \odot \text{res}(B_{in}) \quad (4)$$

where $\text{res}(\cdot)$ denotes the residual block [38].

D. Boundary Guidance Module

The specific structure of the BGM is shown in Fig. 8. The BGM is designed to incorporate the boundary probability information into the segmentation feature maps to enhance the boundaries of the aquaculture area. To selectively emphasize the feature maps with richer boundary information in the segmentation

stream, we introduce a feature map reweighting structure in the BGM module. First, we combine the minus 0.5 operation with the ReLU function to implement the $\max(0, x - 0.5)$ operation, where x is the pixel value in the boundary probability map P . Then, the nonboundary information in the input feature maps F_{in} is masked by an elementwise multiplication operation to obtain a set of feature maps F'_{in} containing mainly boundary information. Next, a global average pooling operation is performed in the spatial dimension to evaluate the richness of aquaculture area boundary details in each feature map of the segmentation stream, and then, a weight vector \mathbf{v} is obtained using a Sigmoid function. Finally, the aquaculture area feature maps are reweighted and fed into the 1×1 convolution layer to obtain a set of aquaculture area feature maps F''_{in} with overall enhanced boundary information. The above process can be expressed as

$$F'_{in} = \max(0, P - 0.5) \odot F_{in} \quad (5)$$

$$\mathbf{v} = \sigma(\text{GAP}(F'_{in})) \quad (6)$$

$$F''_{in} = C_{1 \times 1}(\mathbf{v} \odot F'_{in}) \quad (7)$$

where $\text{GAP}(\cdot)$ denotes global average pooling.

Unlike the traditional residual network, BGM is more concerned with learning detailed information around the boundary, so it multiplies the boundary probability map P with the feature maps F''_{in} in the residual branch. The module output F_{out} is calculated as follows:

$$F_{out} = C_{3 \times 3}(F''_{in} \odot P + F_{in}). \quad (8)$$

E. Loss Function

Our proposed RADNet consists of a boundary stream and a segmentation stream, and we apply different loss functions to train them. Since boundary pixels are in the minority in the aquaculture area image, boundary detection is a class imbalance problem; thus, we choose the focal loss as the loss function of the boundary stream. The loss function is expressed as follows:

$$L_{bdr} = \begin{cases} -\alpha(1-h(x))^\gamma \ln h(x), & y = 1 \\ -(1-\alpha)h(x)^\gamma \ln(1-h(x)), & y = 0 \end{cases} \quad (9)$$

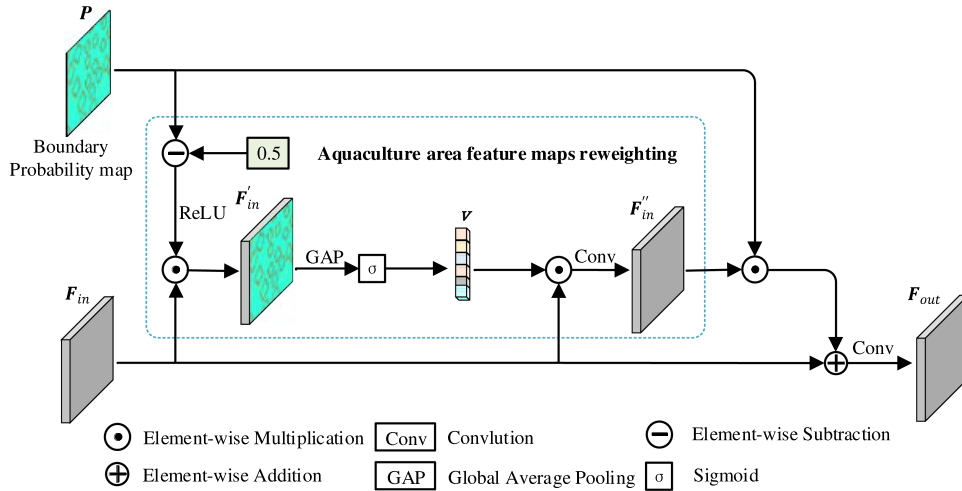


Fig. 8. Structure of the BGM.

where x is the input pixel, $h(x)$ is the value predicted by the network, and y is the value of the label. In this article, the label value of the boundary pixels of the laver aquaculture area is 0, and that of the other pixels is 1. α is a hyperparameter used to balance the importance of aquaculture area boundary and non-boundary samples, and γ is a hyperparameter used to smoothly adjusts the rate at which easy examples are down-weighted. The segmentation stream uses a binary cross-entropy function as the loss function. The loss function is expressed as follows:

$$L_{\text{seg}} = \begin{cases} -\ln h(x), & y = 1 \\ -\ln(1 - h(x)), & y = 0. \end{cases} \quad (10)$$

We set a parameter β to balance L_{bdr} and L_{seg} . Thus, the total loss of the network is expressed as follows:

$$L_{\text{total}} = L_{\text{seg}} + \beta L_{\text{bdr}}. \quad (11)$$

The procedure of our proposed method is summarized in Algorithm 1. By combining the boundary detection and semantic segmentation tasks, our proposed method effectively improves the extraction results of raft laver aquaculture areas.

IV. EXPERIMENTS AND RESULTS

A. Experimental Data Preparation

In this article, GF-1 RSIs were collected and a dataset was created to serve as the basis for the study. The study area was selected from Haizhou Bay, Lianyungang, China, where numerous raft laver aquaculture areas are distributed from November to April each year. The GF-1 RSI was shot on 17 February 2017, and the sensor is PMS2. We used the Pansharp algorithm to fuse the red, green, and blue bands of the multispectral images, which are more sensitive to the laver aquaculture area, with the panchromatic images to supplement the detail information (1.8 m spatial resolution after image fusion and resampling).

As shown in Fig. 9(a), the image contains more than 6000 raft aquaculture areas and has been labeled by visual interpretation. Fig. 9(b) and (c) shows a local zoom-in view of the selected area and the corresponding ground truth map, respectively. In

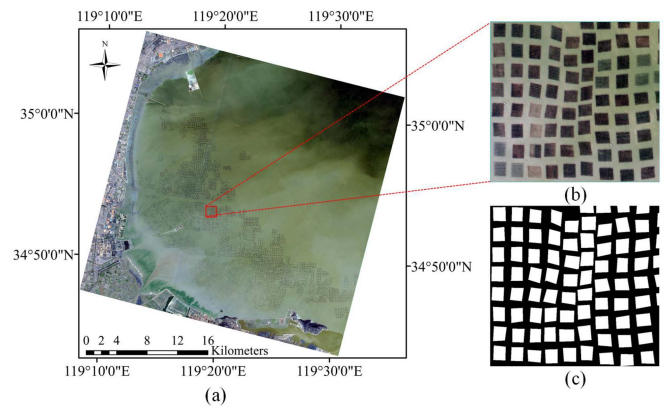


Fig. 9. Location of the study area. (a) GF-1 RSI. (b) Local magnification of the selected area in the red box. (c) Ground truth of the selected area in the red box.

the ground truth map, white pixels indicate aquaculture areas and black pixels indicate seawater. The fused RSI and ground truth map are cropped to 128×128 pixel patches, 30% of which are selected as training and validation sets.

B. Implementation Details

The experiments were conducted on a server equipped with an NVIDIA GeForce RTX 2080Ti GPU and Ubuntu 18.04.5 LTS operating system. All models in this article were implemented and trained based on the Keras framework. During training, Adam was chosen as the optimizer, the initial learning rate was set to $1e-4$, and the batch size was set to 4. In addition, the number of training epochs was set to 150. The hyperparameters α , γ , and β in the loss function were set to 0.1, 4, and 0.6, respectively.

C. Accuracy Evaluation

We evaluated the performance of the proposed RADNet, for which we chose four commonly employed semantic segmentation metrics: 1) precision (user accuracy), 2) recall (producer accuracy), 3) F1-score, and 4) IoU. The metrics are defined as

Algorithm 1: Detail of RADNet.

Input: Remote sensing image I and ground truth; The number of training epochs τ .

Output: Extraction results of aquaculture areas.

- 1: Normalize and clip the image I to obtain subimages with size of $128 \times 128 \times 3$;
- 2: **for** $i = 1$ to τ **do**
- 3: Encode each subimage to extract the multilevel feature maps;
- 4: Get full-resolution feature maps for each subimage using 7×7 convolution;
- 5: **for** $t = 1$ to 3 **do**
- 6: Feed the full-resolution feature maps and higher-level feature maps into BAM;
- 7: Obtain boundary semantics enhanced full-resolution feature maps via (4);
- 8: **end for**
- 9: Compute the boundary probability map P ;
- 10: **for** $t = 1$ to 4 **do**
- 11: Feed the higher-level feature maps and lower-level feature maps into RAM;
- 12: Obtain reverse attention enhanced feature maps via (1)–(3);
- 13: **if** $t > 2$ **then**
- 14: Feed the reverse attention enhanced feature maps and boundary probability map into BGM;
- 15: Obtain boundary-enhanced aquaculture area feature maps via (5)–(8);
- 16: **end if**
- 17: **end for**
- 18: Compute the boundary L_{bdr} via (9);
- 19: Compute the segmentation L_{seg} via (10);
- 20: Compute the loss L_{total} via (11) and update parameters of RADNet.
- 21: **end for**
- 22: Use the test dataset with the trained model to get predicted aquaculture area maps.

follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (12)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (15)$$

where TP, FP, and FN represent the number of true positives, false positives, and false negatives, respectively.

We use IoU as an example to analyze the scientific basis for choosing the above four metrics to assess the results of the regional segmentation of aquaculture areas (15). The more the aquaculture area is mistaken for seawater (FN) or seawater is

TABLE I
EVALUATION RESULTS OF DIFFERENT MODELS ON THE TEST SET

Attention	Precision (%)	Recall (%)	F1-score	IoU (%)
U-Net	93.02	84.19	0.884	79.23
Deeplabv3+	91.83	<u>88.07</u>	<u>0.899</u>	<u>81.68</u>
HRNet	<u>93.51</u>	85.72	0.894	80.90
DS-HCN	94.69	82.64	0.883	78.99
RaftNet	90.43	86.34	0.883	79.11
Improved U-Net	92.79	85.68	0.891	80.34
D-ResUnet	92.09	83.97	0.879	78.33
FRCNet	91.93	84.97	0.883	79.08
SAMALNet	92.70	86.23	0.890	80.75
RADNet	93.36	89.57	0.916	84.42

The best and second best values are marked in bold and underlined respectively.

mistaken for aquaculture (FP), the lower the IoU value, and vice versa. The scientific basis of the other three metrics can be analyzed in a similar way. Thus, precision, recall, F1-score, and IoU provide a visual representation of model performance.

D. Experimental Results

We performed aquaculture area extraction experiments on the test set. The aquaculture areas extracted by RADNet are shown in Fig. 10. The overall extraction results of the laver aquaculture areas are very good. Most of the aquaculture areas with inapparent spectral features can also be correctly identified. Nevertheless, the extraction results of some aquaculture areas are still slightly flawed. Specifically, several aquaculture areas are partially missing, and there is an overextraction phenomenon at the periphery of the aquaculture areas, as shown in blue and red in Fig. 10(c).

E. Comparison With the Other Models

We compare the proposed RADNet with seven other semantic segmentation networks, including U-Net [39], DeepLabv3+ [40], HRNet [41], DS-HCN [3], RaftNet [42], Improved U-Net [8], D-ResUnet [43], FRCNet [44], and SAMALNet [45]. Among them, the latter six methods were proposed for aquaculture area segmentation. Table I shows the quantitative results of the above models on the test set. The proposed model clearly outperforms the other models in terms of recall, F1-score, and IoU. Deeplabv3+ ranked second in terms of recall, F1-score, and IoU. The DS-HCN has the highest precision rate, but it has the lowest recall rate, which reduces its F1-score and IoU.

To visually illustrate the advantages of RADNet over other comparative models, some extraction results are shown in Figs. 11 and 12. The study area in Figs. 11 and 12 contains both inapparent aquaculture areas and obvious aquaculture areas. For the obvious aquaculture areas, each model shows good extraction results. However, for inapparent aquaculture areas, the extraction results of other models have different types of defects, including missing corners, adhesions, small fragments,

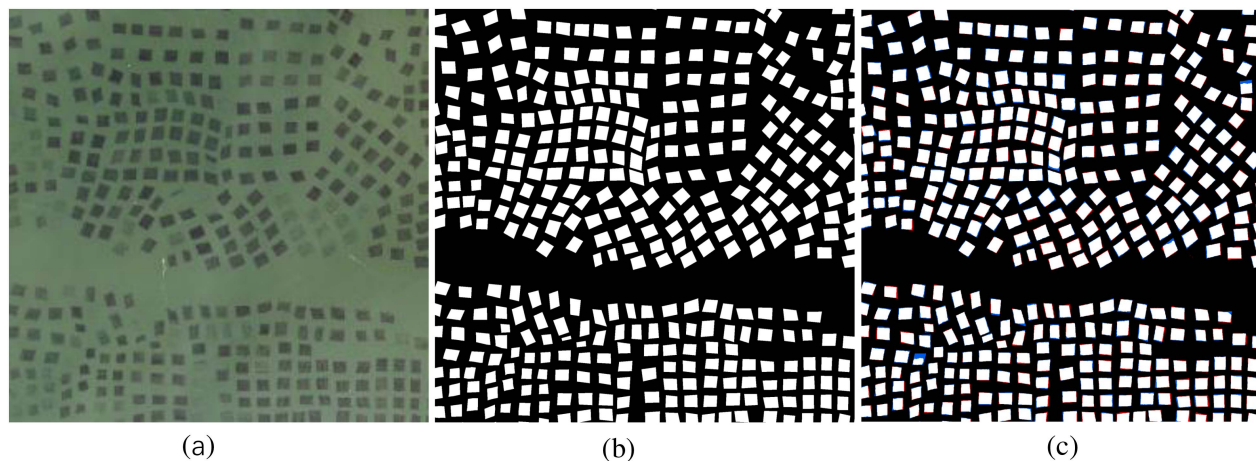


Fig. 10. Visualization of aquaculture area extraction results on the test set. (a) Test image. (b) Ground truth. (c) Aquaculture area extraction results of RADNet. The blue areas represent aquaculture area pixels that are missed by RADNet, and the red areas represent seawater pixels that are incorrectly extracted by RADNet.

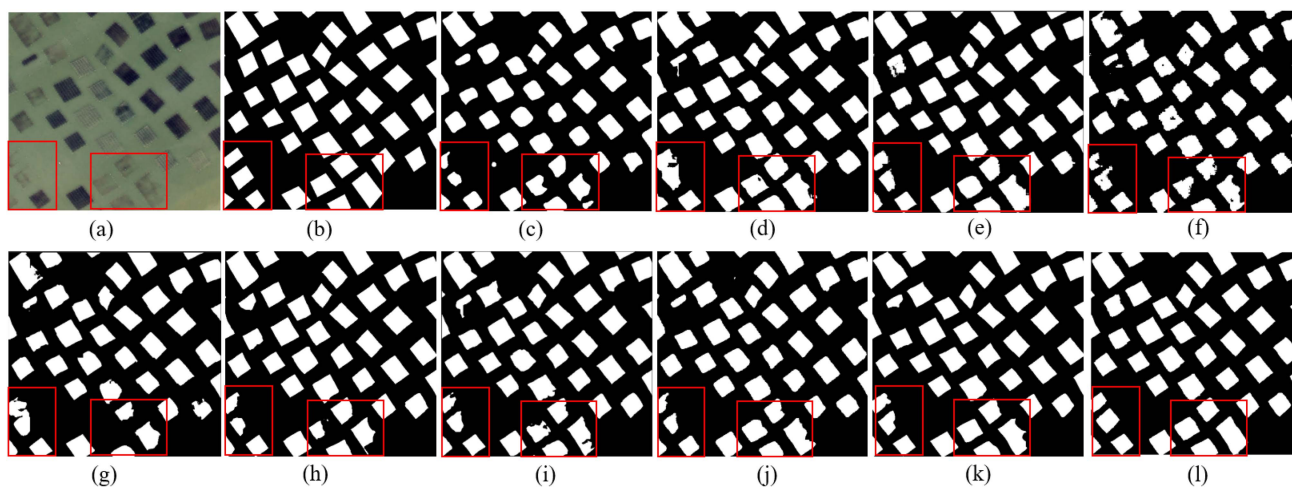


Fig. 11. Visualization of aquaculture area extraction results for each model on test image 1. (a) Test image 1. (b) Ground truth. (c) U-Net. (d) HRNet. (e) Deeplabv3+. (f) DS-HCN. (g) RaftNet. (h) Improved U-Net. (i) D-ResUnet. (j) FRCNet. (k) SAMALNet. (l) RADNet.

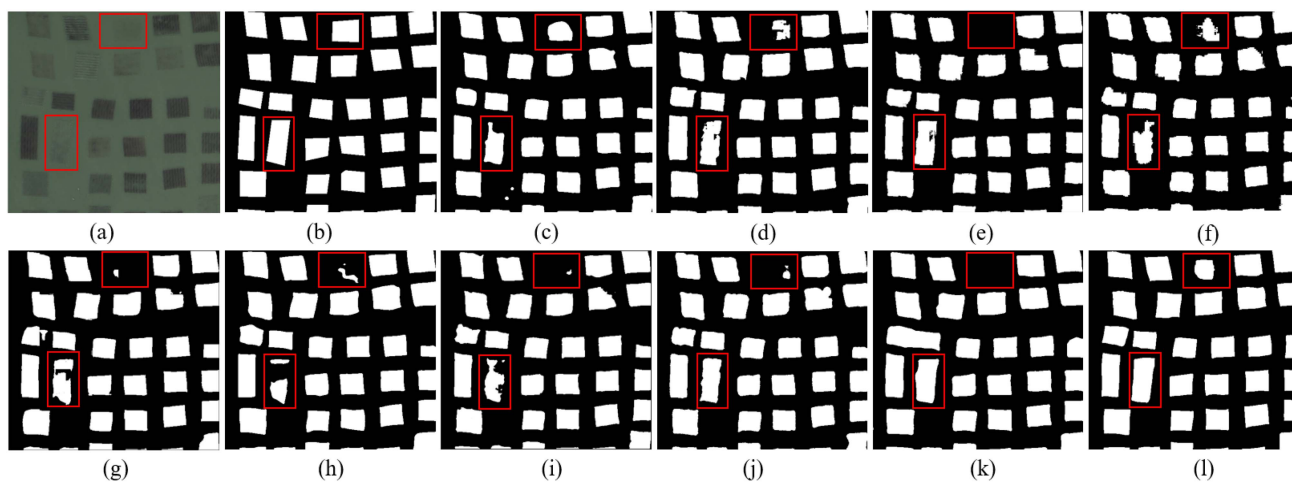


Fig. 12. Visualization of aquaculture area extraction results for each model on test image 2. (a) Test image 2. (b) Ground truth. (c) U-Net. (d) HRNet. (e) Deeplabv3+. (f) DS-HCN. (g) RaftNet. (h) Improved U-Net. (i) D-ResUnet. (j) FRCNet. (k) SAMALNet. (l) RADNet.

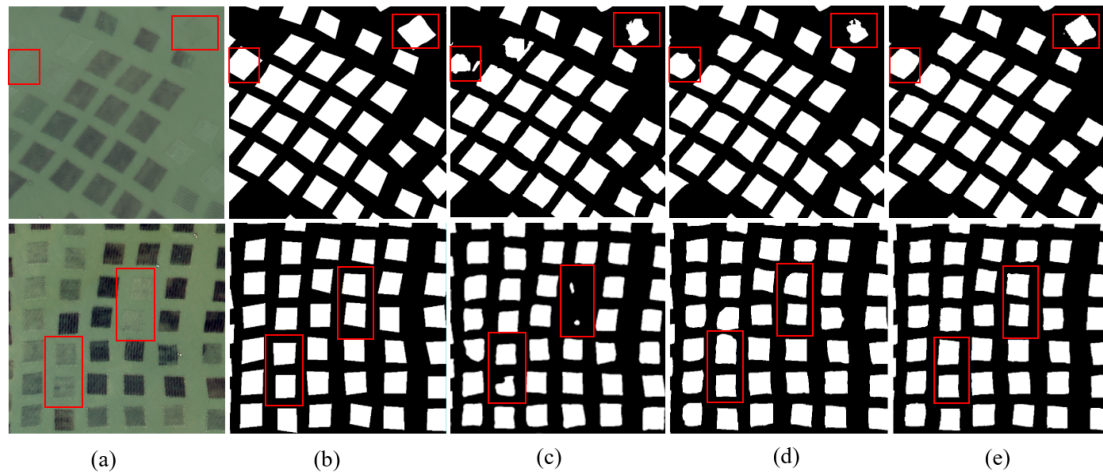


Fig. 13. Visualization of aquaculture area extraction results of the ablation experiment. (a) Test image. (b) Ground truth map. (c) Baseline. (d) With RAM. (e) With RAM, BAM, and BGM.

TABLE II
ABLATION EVALUATION RESULTS ON THE TEST SET

RAM	BAM	BGM	Precision (%)	Recall (%)	F1-score	IoU (%)
–	–	–	92.23	85.79	0.889	80.01
✓	–	–	93.35	88.72	0.908	83.12
–	✓	–	93.44	87.68	0.896	81.10
–	–	✓	93.34	86.10	0.896	81.17
✓	✓	–	93.50	89.29	0.913	84.07
✓	–	✓	93.53	88.61	0.910	83.49
–	✓	✓	93.43	88.26	0.908	83.11
✓	✓	✓	93.36	89.57	0.916	84.42

The best values are marked in bold.

holes, and complex boundary curves, as shown in the red boxes in Figs. 11 and 12. RADNet learns the high-level semantic features of seawater by introducing a RAM, which enables the network to identify and recognize inapparent aquaculture areas.

V. EXPERIMENTS AND ANALYSIS

A. Ablation Study

We validate the effectiveness of each module in our proposed model. The proposed RAM, BAM, and BGM are removed from our model as the baseline. The quantitative experimental results are shown in Table II. The introduction of RAM improved the model by 1.12%, 2.93%, 0.019, and 3.11% in terms of precision, recall, F1-score, and IoU, respectively, which indicates that RAM can effectively improve the performance. After adding the BAM and BGM, the precision, recall, F1 score, and IoU of the model improved by 1.20%, 2.47%, 0.019, and 3.10%, respectively, over the baseline. This finding indicates that optimizing the extracted aquaculture area boundaries and using them to guide segmentation can help improve the accuracy of aquaculture area extraction and identify inapparent aquaculture area components. By combining the RAM, BAM, and BGM, the performance of the model can be further improved.

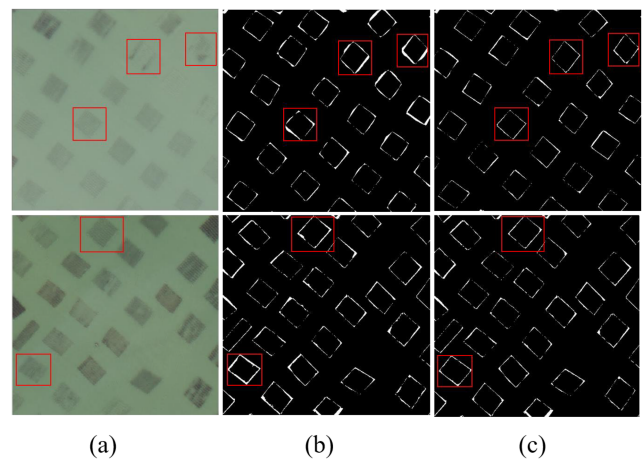


Fig. 14. Comparison of visualization of single-stream and dual-stream structures in terms of prediction errors. (a) Test image. (b) Single-stream. (c) dual-stream.

Fig. 13 shows the impact of the proposed three modules. The extraction results of the baseline network are shown in Fig. 13(c). The addition of the RAM greatly enhances the ability of the model to extract inapparent aquaculture areas, as shown in the red box in Fig. 13(d). However, there are still a few extracted aquaculture areas with anfractuosity or fragmented boundaries. With the addition of the BAM and BGM to the model, the integrity and boundaries of the extracted aquaculture areas become better, as shown in the red boxes in Fig. 13(e).

To further explore the advantages of dual-stream structures over single-stream structures, we conduct ablation experiments with the dual-stream structure and qualitatively compared the prediction errors between the single-stream and dual-stream structures, as shown in Fig. 14. As can be seen in the red boxes in Fig. 14, the dual-stream structure has fewer prediction errors near the boundary and the error boundary region looks narrower, thus proving the advantage of the dual-stream structure. In other words, although the dual-stream structure improves the accuracy of aquaculture area segmentation less (as shown in the 2nd and

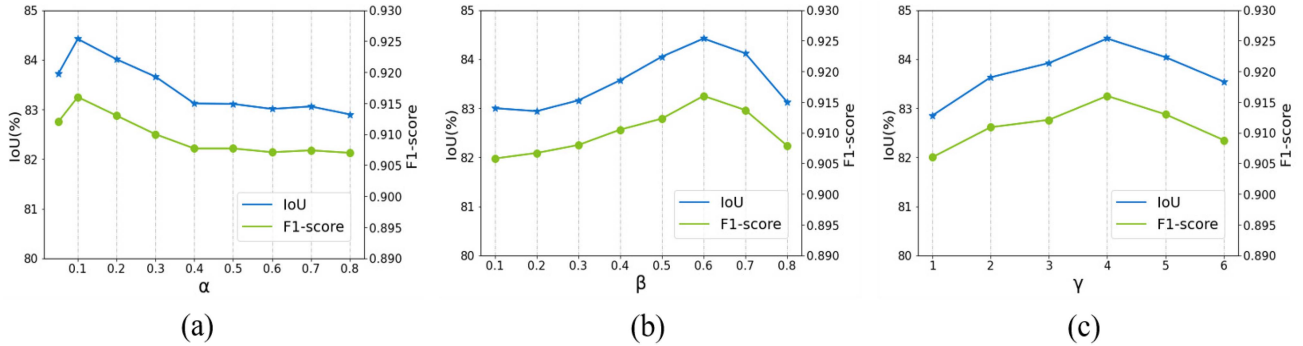


Fig. 15. IoU and F1-score of RADNet with different hyperparameter values. (a) With different α values. (b) With different β values. (c) With different γ values.

8th rows of Table II), it can significantly improve the accuracy of aquaculture area boundary extraction.

B. Hyperparameter Settings

We analyze the effects of parameters α , β , and γ in (9) and (11) on the final extraction results. Fig. 15 shows the segmentation performance of the proposed RADNet when the three parameters α , β , and γ vary. As shown in Fig. 15(a), when the value of α is greater than 0.1, the F1-score and IoU of the model gradually decrease with increasing values of α . Therefore, the value of the parameter α is set to 0.1 in this article. As shown in Fig. 15(b), there is an overall upward trend in F1-score and IoU as β increases from 0.1 to 0.6. When $\beta = 0.6$, the F1-score and IoU reach their peaks. Similarly, it is obvious from Fig. 15(c) that the performance of the proposed model increases and then tends to decrease as the value of γ increases. Therefore, β and γ were set to 0.6 and 4, respectively, in this article.

C. Model Visualization

To evaluate the effects of the three modules RAM, BAM, and BGM, we visualize the input–output feature maps of the relevant modules. We first map the corresponding feature maps to grayscale maps with pixel values between 0 and 255, and then use the `applyColorMap()` function in OpenCV to convert the grayscale maps to RGB maps to get the activation heatmap of the aquaculture area. The color from yellow to red in the heatmap indicates the activate value changing from low to high, and aquamarine indicating no activation. Fig. 16 shows the activation heatmaps of the RAM module for the two test images. The activation area of the aquaculture area in the heat map becomes larger and larger while the background noise of the seawater is effectively suppressed. Fig. 17 shows the activation heatmaps of the boundary of the aquaculture area. After applying the BAM module, the boundary of the aquaculture area becomes clearer and clearer, and the nonboundary texture activation is gradually suppressed. The activation heatmaps of the aquaculture area before and after the application of the BGM module is given in Fig. 18. The activation value near the boundary of the aquaculture area increases after applying the BGM module. This indicates that the BGM can effectively improve the boundary response of the aquaculture area.

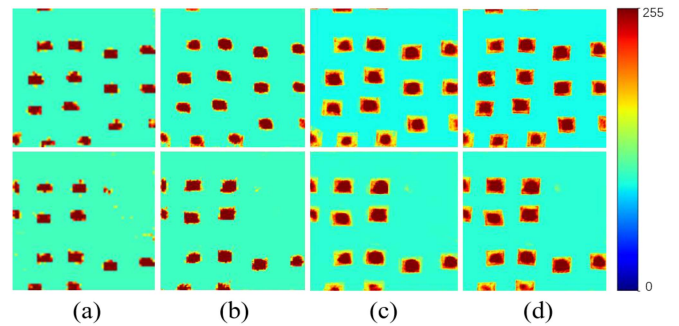


Fig. 16. Activation heatmaps of the aquaculture area of the RAM module for two test images. (a) Output feature map for the first RAM. (b) Output feature map for the second RAM. (c) Output feature map for the third RAM. (d) Output feature map for the fourth RAM.

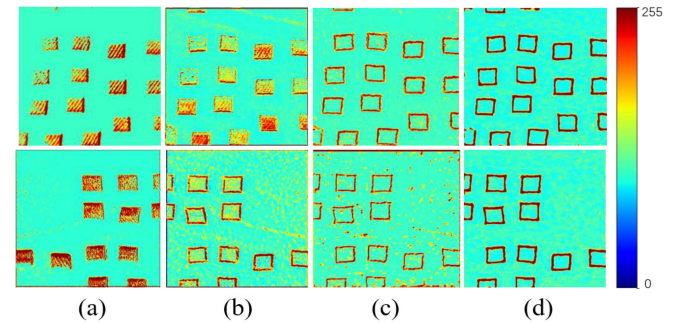


Fig. 17. Activation heatmaps of the boundary of the aquaculture area of the BAM module for two test images. (a) Input feature map of the first BAM. (b) Output feature map of the first BAM. (c) Output feature map of the second BAM. (d) Output feature map of the third BAM.

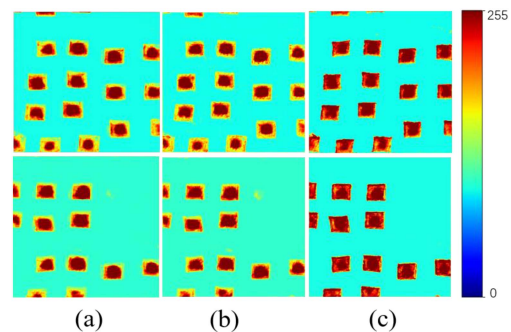


Fig. 18. Activation heatmaps of the aquaculture area of the BGM module for two test images. (a) Input feature map for the 1st BGM. (b) Output feature map for the first BGM. (c) Output feature map for the second BGM.

TABLE III
PARAMETERS AND INFERENCE TIME COMPARISON WITH DIFFERENT MODELS

Models	Parameters	Inference time(s)
U-Net	31.03M	1.24
DeepLabv3+	19.82M	1.81
HRNet	9.05M	13.57
DS-HCN	37.72M	2.09
RaftNet	45.12M	13.63
Improved U-Net	39.42M	2.48
D-ResUNet	26.00M	5.14
FRCNet	1.57M	2.51
SAMALNet	41.12M	8.47
RADNet	28.48M	6.13

D. Evaluation of Model Complexity

We compare the number of parameters and inference time of RADNet with the compared models. The results are listed in Table III. It can be seen that the number of parameters of our proposed RADNet is somewhat more than HRNet, DeepLabv3+, D-ResUNet, and FRCNet, and less than U-Net, DS-HCN, RaftNet, Improved U-Net, and SAMALNet. To obtain the inference time of each model, a test image of size 512×512 is selected and tested on the same computing platform. The inference time of RADNet is shorter than that of HRNet and RaftNet and longer than that of the other comparison models. Compared with U-Net, RADNet reduces the number of convolutional kernels from 1024 to 512 in the last layer of the encoder, which reduces the number of network parameters to a large extent. Therefore, the total number of parameters of RADNet is less than that of U-Net, although the RAM, BAM, and BGM modules bring some additional parameters. However, to avoid the “gradient disappearance” phenomenon, RADNet adds a batch normalization layer after each convolutional layer, which results in more computation and memory access. Therefore, how to reduce the inference time of the network as much as possible is the next focus of our research.

E. Comparison With Other Attention Modules

We compare the proposed RAM module with several classic attention methods, including SE block [29], efficient channel attention (ECA) [30], convolutional block attention module (CBAM) [32], and CA [34]. For a fair comparison, we replaced the RAM in RADNet with other attention modules and calculated the efficiency and performance of the model. F1-score, IoU, and Floating-point Operations (FLOPs) are used as metrics for the evaluation. The results are shown in Table IV. It can be seen that the performance and efficiency of SE, CBAM, and ECA are comparable. The IoU of the CA module is slightly higher than that of the first three attention modules due to the consideration of the spatial dependence over long distances. The RAM module significantly outperforms other comparative models in two metrics, F1-score and IoU, mainly because the features of seawater are fully exploited. However, the RAM is

TABLE IV
EVALUATION RESULTS OF DIFFERENT ATTENTION MODULES

Attention module	F1-score	IoU (%)	Parameters	FLOPs
SE	0.906	82.88	21.95M	43.86M
CBAM	0.907	83.00	21.95M	43.86M
ECA	0.907	82.99	21.95M	43.84M
CA	0.909	83.28	21.90M	44.02M
RAM	0.916	84.42	28.48M	57.62M

slightly less efficient than the other comparison models due to the inclusion of more operations.

VI. DISCUSSION

The spectral and textural features of different raft laver aquaculture areas vary significantly, posing a great challenge to the accurate recognition of raft aquaculture areas. Differences in laver growth stages, harvesting activities, suspended sediments, and detrital algae are important factors contributing to the variation in image features of raft laver aquaculture areas. In addition, during the growth of laver, the net curtain hanging laver needs to be lifted above the sea surface from time to time to bask in the sun to enhance photosynthesis and kill the attached algae. When the net curtain is located below the sea surface, the effect of waves or suspended matter may lead to inapparent spectral features in some raft laver aquaculture areas. In RADNet, to overcome the problem of large intraclass spectral variation of raft laver aquaculture areas, the RAM first learns seawater features and then enhances the features of inapparent raft laver aquaculture areas by seawater masks. The experimental results showed that RADNet was able to extract most of the raft laver aquaculture areas intact, including those with inapparent spectral features. This also leads to RAM having more parameters and increasing the FLOPs of the model.

The strong absorption of solar radiation by seawater usually results in a weak spectral feature of the target object. For such a problem, we propose to enhance the features of the target object indirectly by learning and suppressing the features of the background, which is also inspired by the reverse attention mechanism. Theoretically, this method can be applied to the recognition of objects with weak spectral features in other fields.

Although RADNet has achieved better extraction results for raft laver aquaculture areas, it still has two aspects that need further improvement. One is that the inference speed of RADNet is slower than classical deep learning models such as U-Net and Deeplabv3+, and the second is that there is a slight overextraction near the boundary of the aquaculture area.

In recent years, generative models have developed rapidly and have been successfully applied to tasks such as image enhancement and data augmentation. In our experiments, we also found that the model does not perform as well in shallow waters as in deep waters, which may be due to the higher sediment concentration and smaller number of samples in shallow waters. In the next work, we will try to use generative adversarial networks to generate more samples of raft aquaculture areas

in shallow water to improve the robustness and generalization ability of the model.

VII. CONCLUSION

In this article, we proposed a novel network RADNet based on reverse attention and boundary attention for the accurate extraction of laver aquaculture areas from RSIs. RADNet simultaneously learns useful features for aquaculture area extraction from both the target (aquaculture area) and background (seawater), significantly improving the detection rate and integrity of inapparent aquaculture areas. RADNet uses a dual-stream structure in which the segmentation stream uses RAM to obtain the activation of all aquaculture areas, whereas the boundary stream uses the BAM to enhance the boundary features of the aquaculture areas and deactivate the surrounding seawater. These two streams interact via the BGM to further expand the activation of aquaculture areas within the boundary. Experiments performed on the created dataset demonstrated that the proposed RADNet significantly outperformed other models.

REFERENCES

- [1] Y. Wu, F. Chen, Y. Ma, J. Liu, and X. Li, "Research on automatic extraction method for coastal aquaculture area using Landsat8 data," *Remote Sens. Land Resour.*, vol. 30, no. 3, pp. 96–105, Apr. 2018.
- [2] B. Cheng, Y. Liu, X. Liu, G. Wang, and X. Ma, "Research on extraction method of coastal aquaculture areas on high resolution remote sensing image based on multi-features fusion," *Remote Sens. Technol. Appl.*, vol. 33, no. 2, pp. 296–304, Sep. 2018.
- [3] T. Shi, Q. Xu, Z. Zou, and Z. Shi, "Automatic raft labeling for remote sensing images via dual-scale homogeneous convolutional neural network," *Remote Sens.*, vol. 10, no. 7, Jul. 2018, Art. no. 1130.
- [4] Y. Zhang et al., "Combining segmentation network and nonsubsampling contourlet transform for automatic marine raft aquaculture area extraction from Sentinel-1 images," *Remote Sens.*, vol. 12, no. 24, Dec. 2020, Art. no. 4128.
- [5] Y. Liu, X. Yang, Z. Wang, and C. Lu, "Extracting raft aquaculture areas in Sanduao from high resolution remote sensing images using RCF," *Haiyang Xuebao*, vol. 41, no. 4, pp. 119–130, 2019.
- [6] Y. Liu et al., "Richer convolutional features for edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1939–1946, Aug. 2019.
- [7] B. Cui, D. Fei, G. Shao, Y. Lu, and J. Chu, "Extracting Raft Aquaculture Areas from Remote Sensing Images via an Improved U-Net with aPSE structure," *Remote Sens.*, vol. 11, no. 17, Sep. 2019, Art. no. 2053.
- [8] Y. Lu, W. Shao, and J. Sun, "Extraction of offshore aquaculture areas from medium-resolution remote sensing images based on deep learning," *Remote Sens.*, vol. 13, no. 19, Sep. 2021, Art. no. 3854.
- [9] X. Peng, R. Zhong, Z. Li, and Q. Li, "Optical remote sensing image change detection based on attention mechanism and image difference," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7296–7307, Sep. 2021.
- [10] W. Li, J. Wang, Y. Gao, M. Zhang, R. Tao, and B. Zhang, "Graph-feature-Enhanced selective assignment network for hyperspectral and multispectral data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, Apr. 2022.
- [11] L. Shi, L. Kuang, X. Xu, B. Pan, and Z. Shi, "CANet: Centerness-aware network for object detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Apr. 2, 2021.
- [12] L. Sun, Y. Fang, Y. Chen, W. Huang, Z. Wu, and B. Jeon, "Multi-structure KELM with attention fusion strategy for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, Sep. 2022.
- [13] L. Sun, S. Cheng, Y. Zheng, Z. Wu, and J. Zhang, "SPANet: Successive pooling attention network for semantic segmentation of remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4045–4057, May 2022.
- [14] Q. Huang et al., "Semantic segmentation with reverse attention," in *Proc. Brit. Mach. Vis. Conf.*, 2017, pp. 1–13.
- [15] Y. Chong, X. Chen, and S. Pan, "Context union edge network for semantic segmentation of small-scale objects in very high-resolution remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Sep. 14, 2020.
- [16] D. Cheng, G. Meng, S. Xiang, and C. Pan, "FusionNet: Edge aware deep convolutional networks for semantic segmentation of remote sensing harbor images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5769–5783, Dec. 2017.
- [17] Y. Wang, W. Ding, R. Zhang, and H. Li, "Boundary-aware multitask learning for remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 951–963, Dec. 9, 2020.
- [18] Z. Nong, X. Su, Y. Liu, Z. Zhan, and Q. Yuan, "Boundary-aware dual-stream network for VHR remote sensing images semantic segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5260–5268, Apr. 2021.
- [19] A. Li, L. Jiao, H. Zhu, L. Li, and F. Liu, "Multitask semantic boundary awareness network for remote sensing image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, Feb. 2, 2021.
- [20] Z. Liu et al., "Edge guided context aggregation network for semantic segmentation of remote sensing imagery," *Remote Sens.*, vol. 14, no. 6, Mar. 2022, Art. no. 1353.
- [21] S. Liu, W. Ding, C. Liu, Y. Liu, Y. Wang, and H. Li, "ERN: Edge loss reinforced semantic segmentation network for remote sensing images," *Remote Sens.*, vol. 10, no. 9, Aug. 2018, Art. no. 1339.
- [22] S. Ruder, "An overview of multi-task learning in deep neural networks," 2017, *arXiv:1706.05098*.
- [23] Y. Zhou, F. Zhang, F. Ma, D. Xiang, and F. Zhang, "Small vessel detection based on adaptive dual-polarimetric feature fusion and sea-land segmentation in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2519–2534, Mar. 2022.
- [24] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Learning a discriminative feature network for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1857–1866.
- [25] W. Jing, B. Cui, Y. Lu, and L. Huang, "BS-Net: Using joint-learning boundary and segmentation network for coastline extraction from remote sensing images," *Remote Sens. Lett.*, vol. 12, no. 12, pp. 1260–1268, Dec. 2021.
- [26] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-SCNN: Gated shape CNNs for semantic segmentation," in *Proc. IEEE/CVF Int. Comput. Vis.*, 2019, pp. 5229–5238.
- [27] S. Pan, Y. Tao, C. Nie, and Y. Chong, "PEGNet: Progressive edge guidance network for semantic segmentation of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 4, pp. 637–641, Apr. 2021.
- [28] M.-H. Guo et al., "Attention mechanisms in computer vision: A survey," *Comput. Vis. Media.*, vol. 8, no. 3, pp. 331–368, Sep. 2022.
- [29] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
- [30] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Int. Comput. Vis. Pattern Recognit.*, 2020, pp. 11534–11542.
- [31] J. Hu, L. Shen, S. Albanie, G. Sun, and A. Vedaldi, "Gather-excite: Exploiting feature context in convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 9401–9411.
- [32] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.
- [33] J. Park, S. Woo, J.-Y. Lee, and I. S. Kweon, "BAM: Bottleneck attention module," in *Proc. Brit. Mach. Vis. Conf.*, 2018.
- [34] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13713–13722.
- [35] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7794–7803.
- [36] J. Fu et al., "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3146–3154.
- [37] M. Hassanin, S. Anwar, I. Radwan, F. S. Khan, and A. Mian, "Visual attention methods in deep learning: An in-depth survey," 2022, *arXiv:2204.07756*.
- [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. 770–780.
- [39] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.

- [40] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.
- [41] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5693–5703.
- [42] H. Su, S. Wei, J. Qiu, and W. Wu, "RaftNet: A new deep neural network for coastal raft aquaculture extraction from Landsat 8 OLI data," *Remote Sens.*, vol. 14, no. 18, Sep. 2022, Art. no. 4587.
- [43] L. Gao, C. Wang, K. Liu, S. Chen, G. Dong, and H. Su, "Extraction of floating raft aquaculture areas from Sentinel-1 SAR images by a dense residual U-net model with pre-trained Resnet34 as the encoder," *Remote Sens.*, vol. 14, no. 13, Jun. 2022, Art. no. 3003.
- [44] Y. Fu et al., "Marine aquaculture mapping using GF-1 WFV satellite images and full resolution cascade convolutional neural network," *Int. J. Digit. Earth*, vol. 15, no. 1, pp. 2047–2060, Jun. 2022.
- [45] B. Ai, H. Xiao, H. Xu, F. Yuan, and M. Ling, "Coastal aquaculture area extraction based on self-attention mechanism and auxiliary loss," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2250–2261, Dec. 21, 2022.



Binge Cui received the B.Sc., M.Sc., and Ph.D. degrees in computer science from Harbin Engineering University, Harbin, China, in 2000, 2003, and 2006, respectively.

In 2006, he joined the College of Computer Science and Engineering, Shandong University of Science and Technology. From 2010 to 2011, he was a Visiting Scholar with the Department of Information System, City University of Hong Kong. From 2012 to 2014, he was a Postdoctoral Researcher with the First Institute of Oceanography, State Oceanic Administration,

China. He is currently a Professor. His research interests include hyperspectral image classification and remote sensing images understanding with deep learning.



Yanli Zhao received the B.Sc. degree in computer science and technology from the Jiangxi University of Science and Technology, Jiangxi, China, in 2020. He is currently working toward the M.S. degree in remote sensing information extraction with the College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China.

His current research interests include deep learning and remote sensing image processing.



Mingkai Yang received the B.Sc. degree in computer science and technology from the Taishan College of Science and Technology, Taishan, China, in 2020. He is currently working toward the M.S. degree in remote sensing information extraction with the College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China.

His current research interests include deep learning and remote sensing image processing.



Ling Huang received the B.Sc. degree in computer science from Qufu Normal University, Qufu, China, in 2001, and the M.Sc. degree in software and theory of computer from Northwestern Polytechnical University, Xian, China, in 2004.

In 2004, she joined the College of Computer Science and Engineering, Shandong University of Science and Technology. She is currently a Lecturer. Her research interest includes hyperspectral image classification.



Yan Lu received the B.Sc. and M.Sc. degrees in computer science from Yanshan University, Qinhuangdao, China, in 1998 and 2000, respectively, and the Ph.D. degree in computer science from Fudan University, Shanghai, China, in 2003.

From 2003 to 2005, she was a Postdoctoral Researcher in computer science and technology with the Harbin Institute of Technology. In 2005, she joined the College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao, China. She is currently an Associate Professor.

Her research interests include hyperspectral image classification and object detection.