IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 January 2023, accepted 26 January 2023, date of publication 2 February 2023, date of current version 7 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3241630

RESEARCH ARTICLE

CRAS-YOLO: A Novel Multi-Category Vessel Detection and Classification Model Based on YOLOv5s Algorithm

WENXIAO ZHAO[1], **MUHAMMAD SYAFRUDIN**[2], **(Member, IEEE), AND NORMA LATIF FITRIYANI**[3], **(Member, IEEE)**

[1]School of Statistics, Beijing Normal University, Haidian, Beijing 100875, China
[2]Department of Artificial Intelligence, Sejong University, Seoul 05006, Republic of Korea
[3]Department of Data Science, Sejong University, Seoul 05006, Republic of Korea

Corresponding authors: Muhammad Syafrudin (udin@sejong.ac.kr) and Wenxiao Zhao (202122011050@mail.bnu.edu.cn)

This work was supported in part by the Beijing Normal University Graduate Student Innovation Foundation of China under Grant 202122011050.

ABSTRACT Multi-category vessel detection and classification based on satellite imagery attract a lot of attention due to their significant applications in the military and civilian domains. In this study, we generated a new Artificial-SAR-Vessel dataset based on the combination of the FUSAR-Ship dataset and the SimpleCopyPaste method. We further proposed a novel multi-category vessel detection called CRAS-YOLO which consisted of a convolutional block attention module (**C**BAM), receptive fields block (**R**FB), and adaptively spatial feature fusion (**AS**FF) based on YOLOv5s. The proposed CRAS-YOLO improved the feature pyramid network based on the path aggregation network (PANet), which integrates the **R**FB feature enhancement module and **A**SFF feature fusion strategy to obtain richer feature information and realize the adaptive fusion of multi-scale features (RA-PANet). At the same time, a CBAM is added to the backbone to accurately locate the vessel location and improve detection capability. The results confirmed that the proposed CRAS-YOLO model reached a precision, recall rate, and mean average precision (mAP) (0.5) of up to 90.4%, 88.6%, and 92.1% respectively. The proposed model also outperformed previous studies' results in another Sar Ship Detection (SSDD) dataset with precision, recall, and mAP scores of up to 97.3%, 95.5%, and 98.7% respectively.

INDEX TERMS Artificial-vessel dataset, feature fusion, multi-category vessel detection, satellite imagery, YOLOv5s.

I. INTRODUCTION

The shipping industry is quickly becoming more intelligent in the age of artificial intelligence. The port monitoring service has implemented the use of computer vision for multi-vessel/ship image detection and classification. A reliable detection and classification method based on remote synthetic aperture radar (SAR) images is receiving a lot of attention due to its considerable military and civilian applications. Synthetic aperture radar (SAR) satellites are active microwave imaging sensors that are not affected by weather, light, and other conditions, which cast an important role in monitoring ships, aircraft, vehicles, and bridges in the military and civilian fields [1], [2].

In the early days, ship detection methods for SAR images were mostly based on traditional object detection algorithms and were semi-automated. In terms of traditional methods, a constant false alarm rate (CFAR) algorithm was proposed [3], [4]. In addition, there are other methods such as entropy [5], wavelet transform [6], and template matching for ship target detection [7]. These traditional algorithms have a range of problems, including detection accuracy and model deployability.

Nowadays, deep learning has broken through the bottleneck of traditional object detection algorithms and is the mainstream algorithm of detection. The deep learning method

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Ayoub Khan.

does not need to separate sea and land in the SAR images and only needs to be trained by a labeled data set and has great advantages in target detection. The current popular object detection algorithms have two types. One is two-stage object detection algorithms based on region recommendation, of which the representative methods are region-based convolutional neural networks (R-CNN), Fast R-CNN, and Faster R-CNN [8], [9], [10]. The main idea is to utilize selective search methods to generate the suggested region, and regression classification is then made in the suggested area. Another type is one-stage object detection algorithms, which simplify detection problems to regression problems, requiring only convolutional neural networks to directly obtain class probability and position coordinates of targets. Representative algorithms include you only look once (YOLO) [11], single-shot multibox detector (SSD) [12], Retina-Net [13], and so on. YOLO series algorithms are generally faster than other algorithms and have a good effect on small object detection. They are classic one-stage detection methods, which generally have faster recognition speed than other algorithms, and show excellent detection capability in small object detection.

The detection method of deep learning is inseparable from the support of datasets, and at present, there are many SAR ship detection datasets. Li et al. [14] first proposed the SAR Ship Detection Dataset (SSDD) dataset. This dataset contains ships in various environments, such as image resolution, ship size, sea condition, and sensor type. Wang et al. [15] proposed the SAR-Ship-Dataset, which is derived from multimodal SAR images. Based on this dataset, they realize an integrated deep learning processing system for merchant ship detection and classification in the complex background and realize near-real-time automatic detection and classification of commercial ships without sea and land division. Sun et al. [16] proposed a high-resolution and large-scale SAR image ship detection dataset called AIR-SARShip. They also experimented with the dataset using a series of a deep learning algorithms. The backgrounds include near shore and the open sea. Su et al. published high-resolution SAR images called HRSID [17], which can be used for ship detection and instance segmentation. However, these datasets only contain ship location information and lack ship class information. Based on these datasets, deep learning methods have flourished in SAR image ship detection. Li et al. [14] proposed an improved Faster R-CNN model and achieved good results on SSDD datasets. Lei et al. [18] proposed a high-resolution SAR rotation ship detection dataset called SRSDD, which contains both vessel category information and vessel angle information. This dataset can be used for rotating frame target detection. However, the dataset has a category imbalance problem, which seriously affects the detection accuracy. Hu et al. [19] introduced the SENet channel attention mechanism in Faster R-CNN to strengthen feature extraction capabilities. Zhang et al. [20] proposed an improved YOLOv3 algorithm, which replaced DarkNet53 with DarkNet19 and had a fast detection speed on SSDD datasets. Hong et al. [21] input optical images and SAR images into Yolov3 with the aim of enhancing the generalization ability of the model, which introduces the k-means++ algorithm, and Gaussian parameter for ship detection and uses four anchor boxes in Yolov3.Gong and Wu [22] proposed an improved YOLOv4-tiny algorithm, which is based on an efficient channel attention mechanism to enhance channel feature expression. And the proposed cascade residual dilated fusion module can promote the algorithm to effectively improve the accuracy of object detection.

At the same time, considering that different models of SAR images have different resolutions, this leads to large differences in the area size of ship pixels in the same dataset. Feature pyramid network (FPN) [23] fuses multi-scale features to improve multi-scale object detection performance. However, FPN does not feed back the accurate positioning information existing in the low-level feature map to the high-level semantic feature map, and the feature transfer between the layers is limited to adjacent levels, resulting in the imbalance of feature fusion. Liu et al. [24] proposed the PANet network, introduced the bottom-up path augmentation structure and used the shallow features of the network to fuse the FPN features. Ghiasi et al. [25] proposed neural architecture search FPN (NAS-FPN) networks, which use neural network structure search methods to automatically design feature networks. The bi-directional feature pyramid network (BiFPN) [26] introduces contextual information and weight information based on PANet to balance different scale features and obtain larger receptive fields and richer semantic information. However, considering the large difference in ship scale, the existing feature fusion networks are difficult to meet the requirements of SAR ship detection in actual scenarios.

In multi-vessel/ship detection and classification, the ability to represent the point of interest more accurately is important, and one way to do it is by utilizing attention. A convolutional block attention module (CBAM) [27] is one of the attention networks that has widely been used to improve the detection capabilities in many applications such as fly species recognition [28], bamboo sticks counting [29], safety helmets wear-ing recognition [30], and human activity recognition [31]. Therefore, by leveraging attention mechanisms such as focusing on essential features and suppressing irrelevant ones, we hope to boost the power of representation. In addition, by applying CBAM, an accurate vessel location as well as an improved detection capability of our proposed model can be achieved.

Most of the aforementioned SAR ship detection datasets that have been publicly released so far only include ship position data and lack ship category data. At the same time, the only public multi-category ship detection dataset called SRSDD [18] has a serious category imbalance problem, which seriously affects the accuracy of ship detection. Thus, in this study, we generated a novel dataset called the ArtificialSAR-Vessel dataset based on the combination of the

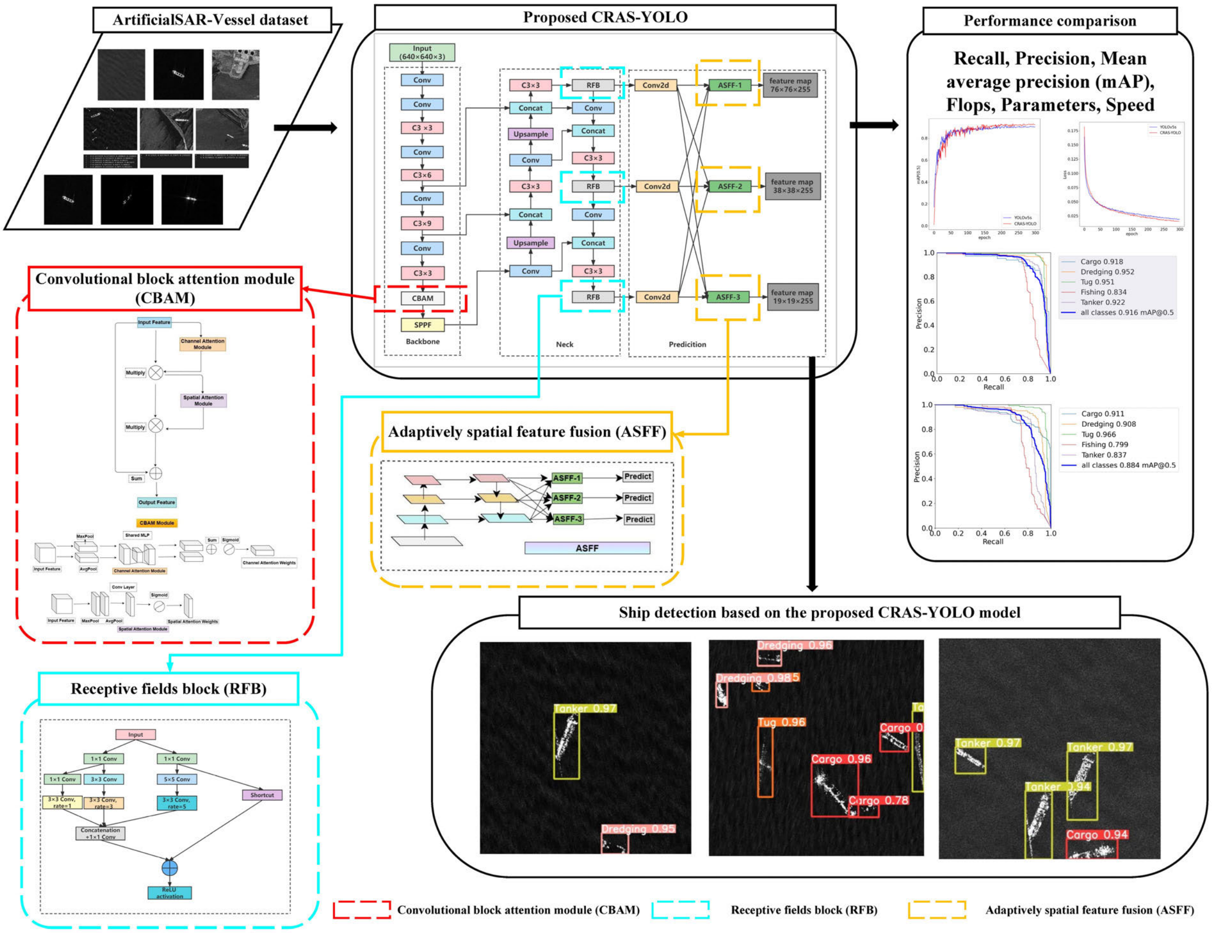


FIGURE 1. Our proposed CRAS-YOLO model which consisted of a convolutional block attention module (CBAM), receptive fields block (RFB), and adaptively spatial feature fusion (ASFF) based on the You Only Look Once version 5s (YOLOv5s) for satellite imagery multi-vessel detection and classification.

FusarShip dataset [32] and the sea surface remote synthetic aperture radar (SAR) images taken by the HAISI-1 satellite. we innovatively introduce the SimpleCopyPaste method into the dataset construction, hoping to supplement the SAR ship detection dataset and provide a new solution to the problem of lack of ship detection datasets.

And many studies based on ship detection only study ship position information, without further analysis of ship categories. Therefore, we proposed a novel model called CRAS-YOLO, which consisted of a convolutional block attention module (**C**BAM) [27], receptive fields block (**R**FB) [33], and adaptively spatial feature fusion (**AS**FF) [34] based on the You Only Look Once [11] version 5s (YOLOv5s) [35] algorithm. An important weakness of the current existing FPN-based models is that deep semantic features are used more thoroughly than shallow features, which makes the detection of small ships inaccurate. Therefore, the CRAS-YOLO model proposed an improved feature pyramid network called RA-PANet based on the path aggregation network (PANet) [24], which integrates the **R**FB feature enhancement module and **A**SFF feature fusion strategy to obtain richer feature information and realize the adaptive fusion of multi-scale features. At the same time, a CBAM is added to the backbone to accurately locate the vessel location and improve detection capability. Finally, the main contributions of this article are as follows:

- We generated a novel dataset called ArtificialSAR-Vessel which consists of multi-vessel SAR images and provides not only the vessel/ship location information but also its categories such as cargo, dredger, tug, fishing, and tanker.
- We also proposed a novel model called CRAS-YOLO which integrated CBAM, and RA-PANet into the YOLOv5s algorithm. The proposed model improved the feature pyramid structure called RA-PANet based on YOLOv5s, which integrates the RFB feature enhancement module and ASFF feature fusion strategy in the neck structure to enhance the model detection capability. CBAM was added to the backbone of the model to locate ship targets more accurately.
- We presented and performed a comprehensive analysis of our proposed CRAS-YOLO model on our novel ArtificialSAR-Vessel dataset as well as another publicly available SAR ship detection dataset (SSDD).

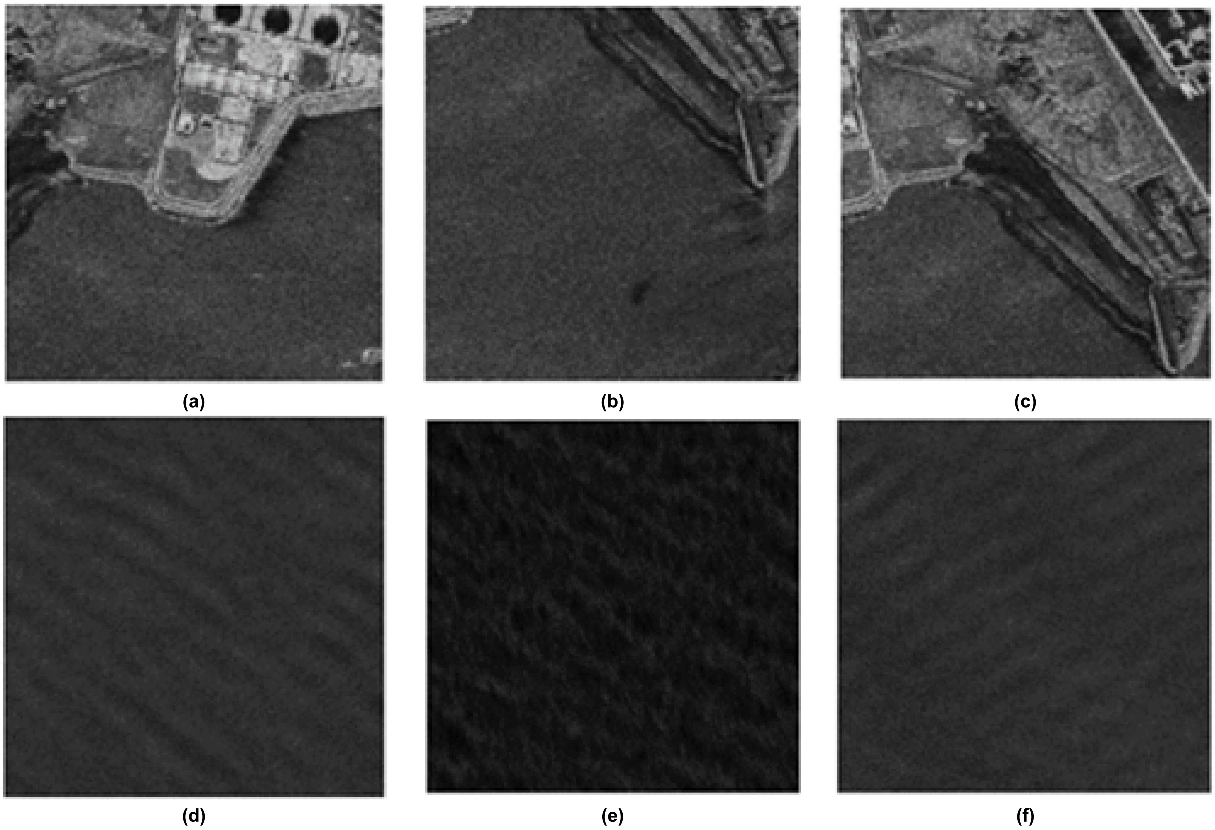


FIGURE 2. Image examples: (a), (b), and (c) are near-shore complexes, and (d), (e), and (f) are offshore sea surface background images, respectively.

Seven metrics were measured, such as precision (P), recall (R), FLOPs, mean average precision (mAP), number of parameters (Params), and speed. The results from previous studies were also compared with our proposed model. Finally, this study could be used by decision-makers to develop a ship detection model that can accurately and automatically detect and classify multiple vessels based on satellite imagery.

The remainder of our work is organized as follows. Section II presented the proposed CRAS-YOLO including datasets description, overall design, and modules of the proposed model as well as performance evaluation metrics. Section III discusses the performance evaluation of the proposed model. Finally, the concluding remarks and future works are presented in Section IV.

II. MATERIALS AND METHODS

The proposed CRAS-YOLO was formed to offer high performance in accurately locating the vessel location and improving detection capability given the SAR images. The flowchart in Fig. 1 shows how the proposed CRAS-YOLO is developed. The details of each step sequentially are presented and explained in the following subsections. First, the novel ArtificialSAR-Vessel dataset is generated. Second, the proposed CRAS-YOLO model is formed by adding the CBAM, RFB, and ASFF into the YOLOv5s network. Next, the performance metrics are presented to evaluate the performance of the proposed model as compared to other models. Finally, the developed CRAS-YOLO model is used in ship detection based on satellite imagery.

A. ARTIFICIALSAR-VESSEL DATASET

Based on the FusarShip [32] dataset and the sea surface remote sensing images taken by the HAISI-1 satellite, this paper constructs a novel dataset named ArtificialSAR-Vessel using the SimpleCopyPaste [36] method. SimpleCopyPaste [36] is a data enhancement method proposed by Google in 2021, the main idea is to directly copy and paste instances of an image to another image to obtain new sample data, which can create new data with more complex scenes, to enrich the dataset. SimpleCopyPaste is a hybrid data enhancement method that mixes information from different images while modifying the corresponding labels. This method copies only the pixels of the instance, not all the pixels of the image. First, randomly select two pictures, each for random scale shaking, and then randomly select some examples from one picture, paste them directly onto another picture, and update the detection box and category labels. Pasting some instances directly onto another image typically obscures the original image instance, and SimpleCopyPaste filters the occluded instances by detecting the box threshold and the mask pixel threshold.

In the sea surface background image, we constructed near-shore complex background images and offshore sea surface

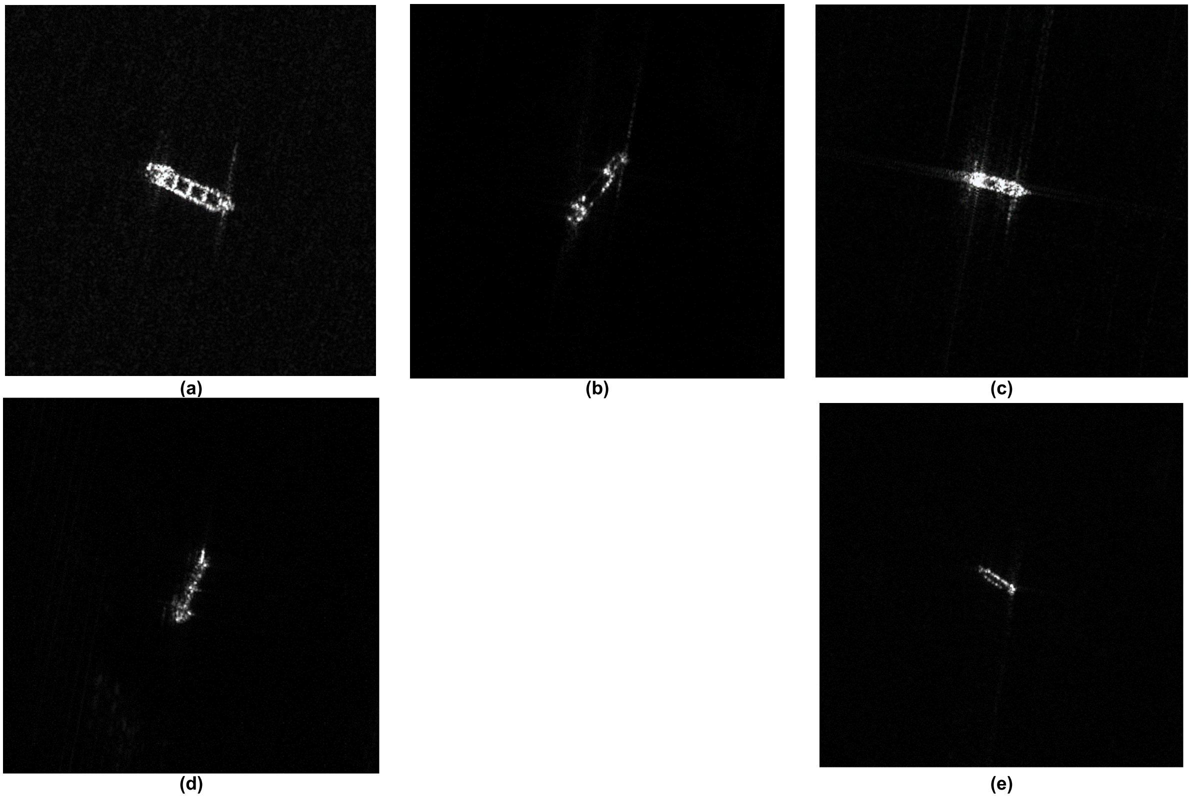


FIGURE 3. Representative images of different vessel/ship categories such as (a) cargo, (b) dredger, (c) fishing, (d) tank, and (e) tug, respectively.

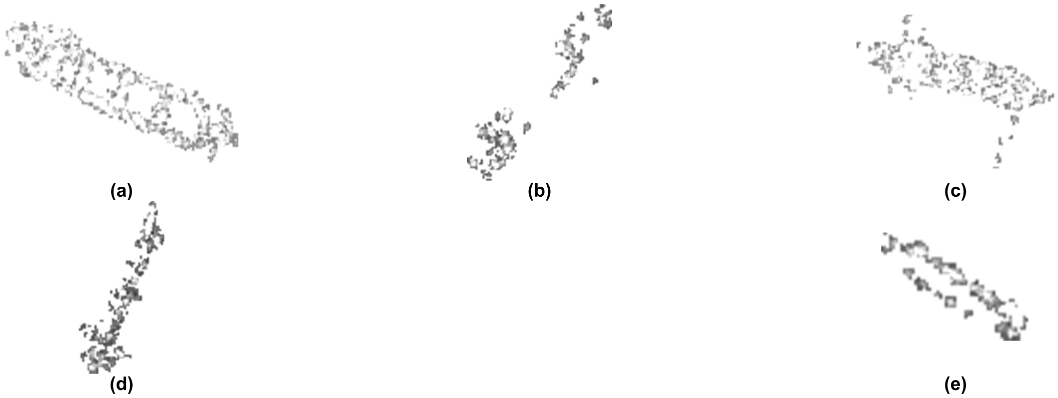


FIGURE 4. The segmented ship images with transparent backgrounds of different vessel/ship categories such as (a) cargo, (b) dredger, (c) fishing, (d) tank, and (e) tug, respectively.

background images with dimensions of 640 × 640. As shown in Fig. 2a, b, and c are near-shore complex background images, and d, e, and f are offshore sea surface background images.

Next, a total of 1550 vessel files in 5 categories are selected in the FusarShip [32] dataset, including 388 Cargo types, 368 Dredger types, 342 Fishing types, 228 Tanker types, and 224 Tug types. At the same time, the filtered vessel images are threshold-segmented, separating the vessel from the background. We selected the OTSU threshold segmentation method [37] to segment the vessels and backgrounds to obtain the vessel masks and used the obtained masks to multiply the original ship images by pixel to obtain ship images with transparent backgrounds. The OTSU method takes advantage of the difference in grayscale between the target and background and divides the pixel level into several classes by setting a certain threshold, to separate the target and background. The original image is $f(x, y)$, T is the threshold, and the formula for segmenting the image is as follows:

$$g(x, y) = \begin{cases} 1 & f(x, y) \geq T \\ 0 & f(x, y) < T. \end{cases} \tag{1}$$

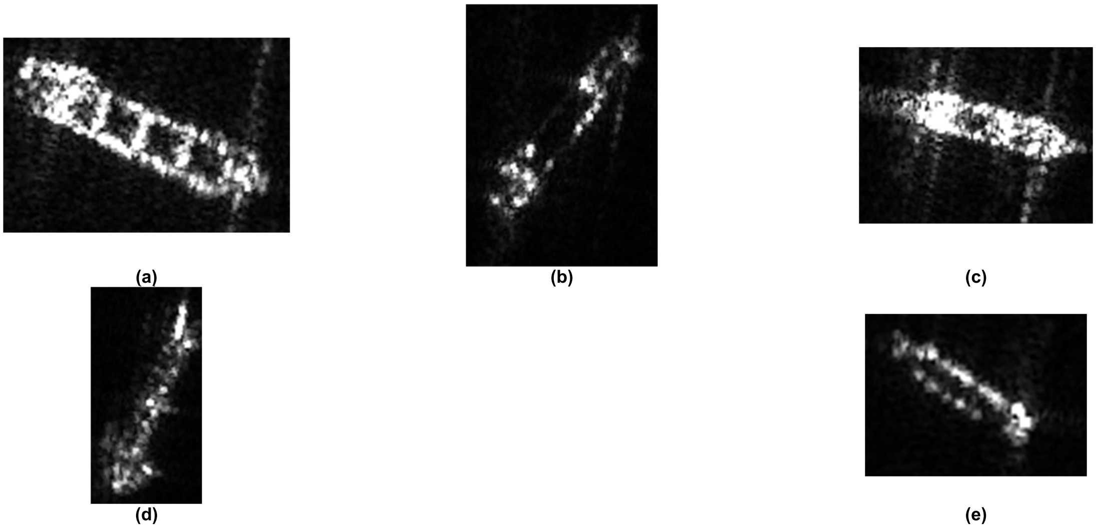
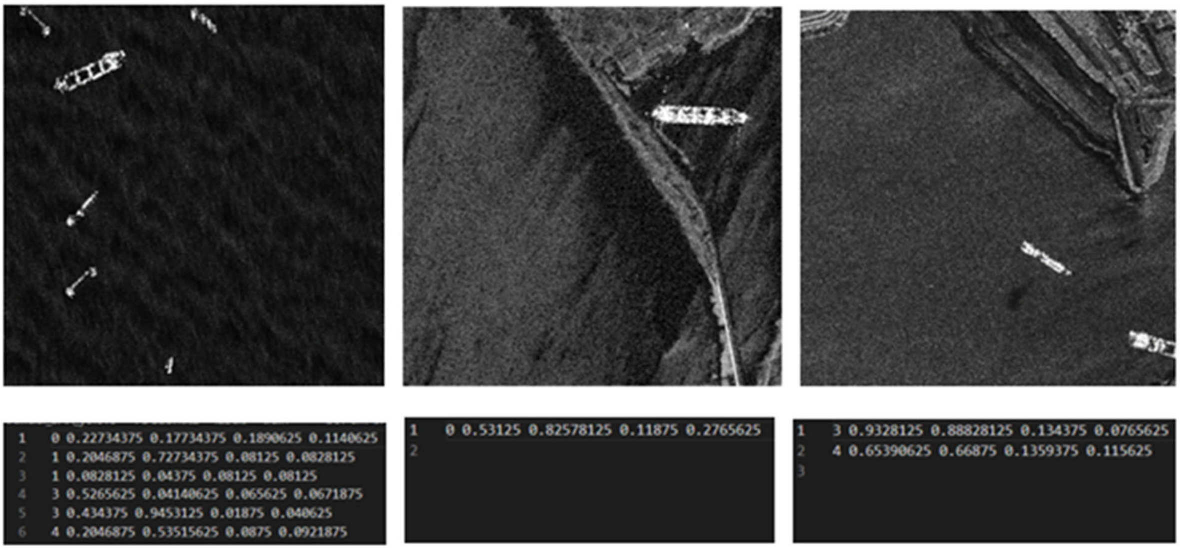


FIGURE 5. The segmented ship images with black backgrounds of different vessel/ship categories such as (a) cargo, (b) dredger, (c) fishing, (d) tank, and (e) tug, respectively.



```
0 0.22734375 0.17734375 0.1890625 0.1140625
1 0.2046875 0.72734375 0.08125 0.0828125
1 0.0828125 0.04375 0.08125 0.08125
3 0.5265625 0.04140625 0.065625 0.0671875
3 0.434375 0.9453125 0.01875 0.040625
4 0.2046875 0.53515625 0.0875 0.0921875
```

```
0 0.53125 0.82578125 0.11875 0.2765625
```

```
3 0.9328125 0.88828125 0.134375 0.0765625
4 0.65390625 0.66875 0.1359375 0.115625
```

FIGURE 6. ArtificialSAR-Vessel images and corresponding text files.

OTSU is a method of automatically determining thresholds using the maximum interclass variance, which is a global-based binary algorithm. When the threshold taken maximizes the variance between classes, the probability of misdivision is the smallest and the division effect is the best [31].

We set T as the threshold for segmentation, and the ratio of target pixels to the total pixels in the image is w_0, and the average grayscale of the target is u_0; the ratio of background pixels to the total pixels in the image is w_1, and the average grayscale of the background is u_1. The variance of the foreground and background in images is g. These variables satisfy the following formula:

$$\begin{cases} u = w_0 \times u_0 + w_1 \times u_1 \\ g = w_0 \times (u_0 - u)^2 + w_1 \times (u_1 - u)^2, \end{cases} \tag{2}$$

and g satisfies the following formula:

$$g = \frac{w_0}{1 - w_0} \times (u_0 - u)^2. \tag{3}$$

When g is the largest, the targets and background are the most different, and the grayscale T is the most optimal threshold.

Fig. 3 shows the representative images of the five types of vessels, and Fig. 4 shows vessel images with transparent backgrounds that have been segmented. Fig. 5 shows vessel images with black backgrounds that have been segmented. Judging from the separation results, the separation between ships and backgrounds is clean, and the separation effect is excellent. Next, we divided each class of ships in an 8:2 scale and get training and validation sets of each type, and mixed the vessel images and background images using

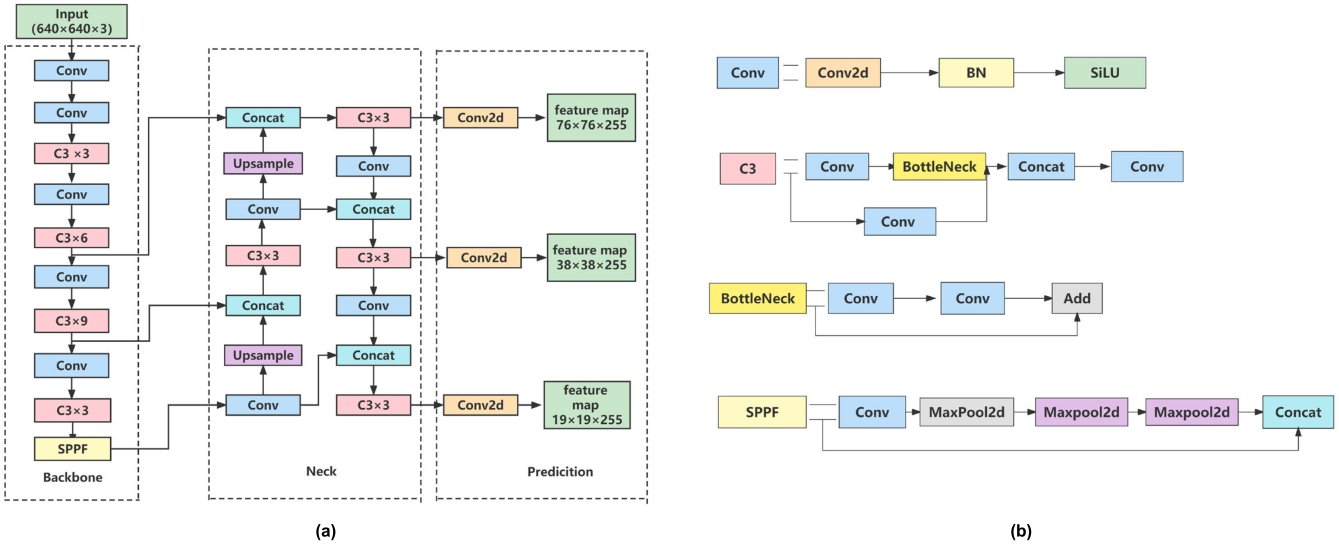


FIGURE 7. The original YOLOv5 (a) network and (b) module structure.

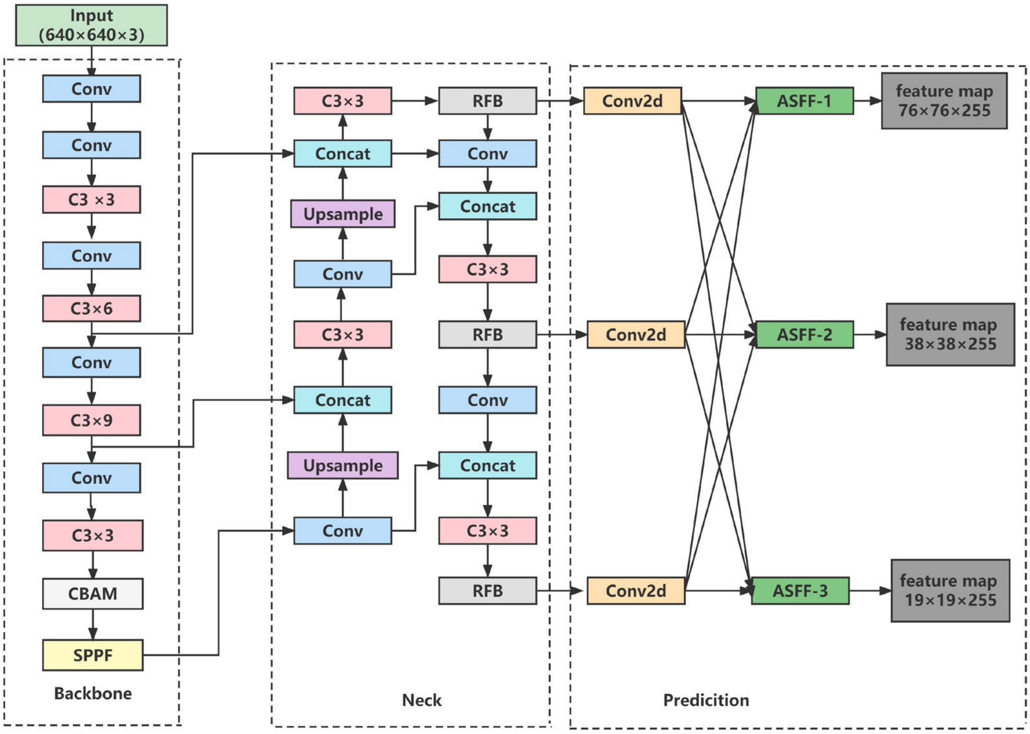


FIGURE 8. The proposed CRAS-YOLO network structure.

the SimpleCopyPaste [36] method. The purpose is to ensure that each vessel does not appear in both the training set and validation set, ensuring the rationality of the dataset. Delete the unreasonable generated images, and finally obtain a multi-classification ship dataset containing both the sea surface and ship targets of different categories, and generate the corresponding location information and category information files for ships in images. We finally generated 2073 images, of which 1658 (79.98%) were used as training sets and 415 (20.02%) as validation sets, and corresponding text files containing location and category information were generated for each image as visualized in Fig. 6.

B. PROPOSED CRAS-YOLO MODEL

The proposed CRAS-YOLO is based on the YOLOv5s [35] algorithm with the addition of CBAM [27], RFB [33], and ASFF [34].

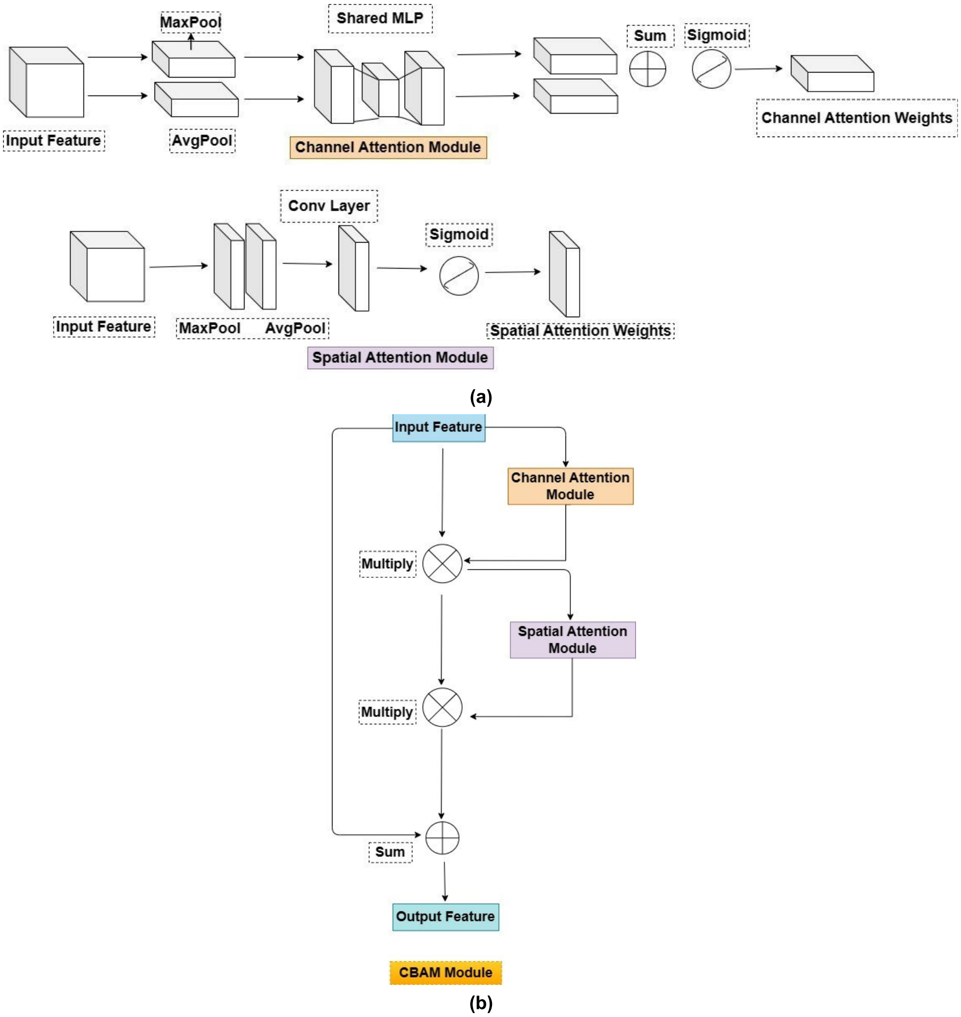


FIGURE 9. The structure of the (a) Channel attention module and spatial attention module, and (b) Convolutional block attention module (CBAM) network.

YOLOv5s [35] is the smallest model structure in the series of YOLOv5 models. The original YOLOv5 network and module structure are shown in Fig. 7a and b respectively. First, a three-channel RGB image is entered at the input with a feature size of 640 × 640 × 3. At the same time, Mosaic data enhancement and other methods are used to enrich the image background of the detection target and reduce the model's dependence on batch size. The backbone of YOLOv5 is CSP-Darknet53 [38], which aims to extract features from images, and consists mainly of Conv, C3, SPPF, and BottleNeck modules. The Conv module consists of convolution, batch normalization (BN), and activation functions (SiLU). The C3 module is constructed of the Conv module, BottleNeck module, and concat splicing, of which the BottleNeck module is composed of the Conv module and add operation. SPPF structure is an improvement of SPP, which is a spatial pyramid pooling network and can convert any size of the feature map into a fixed size of the feature vector.

YOLOv5s uses the feature pyramid structure of PANet in the neck. The structure uses top-down lateral connections to construct high-level semantic features on all scales, and at the same time, considering the fuzziness of the underlying target information, the bottom-up structure is added to compensate for and strengthen the positioning information. PANet better integrates shallow and deep feature information, so that the network fully extracts features at all levels and obtains richer feature information (strong semantic information and edge, texture, and other information). The output section has three Yolo Head detectors and can output three different dimensional feature maps.

In this study, our proposed CRAS-YOLO vessel detection and classification model is based on YOLOv5s with an improvement in the FPN by adding RFB and ASFF in the PANet (RA-PANet) to obtain richer feature information and achieve adaptive fusion of multi-scale features. The proposed CRAS-YOLO also integrated the CBAM into the neck of the

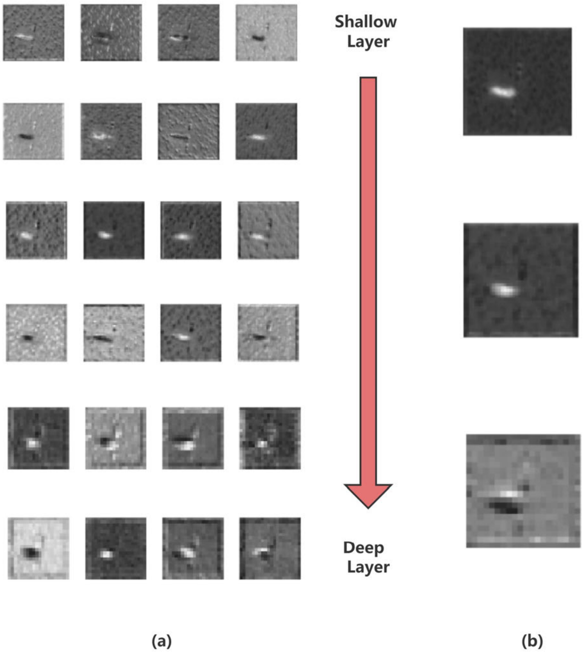


FIGURE 10. The visualization of feature maps of SAR images. (a) The visualization of feature maps extracted by filters from the backbone network of RA-PANet. (b) Visualization of the feature maps of small ships from shallow layers to deep layers.

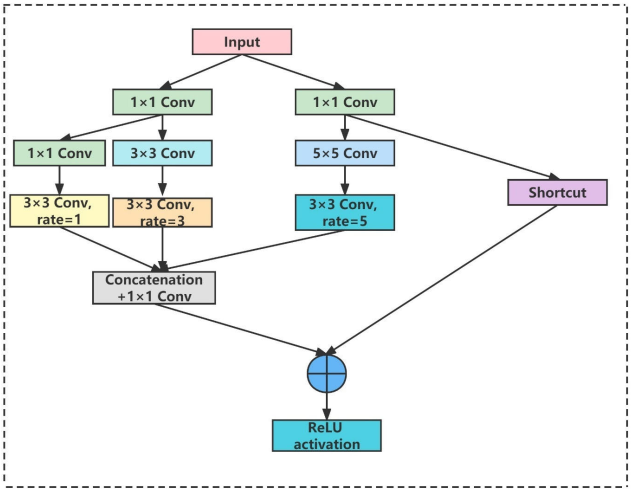


FIGURE 11. The detailed receptive fields block (RFB) feature enhancement module network structure.

network. The full network structure of CRAS-YOLO can be seen in Fig. 8.

Attention mechanisms mainly include spatial attention and channel attention. SE and ECA are representatives of channel attention mechanisms. CA is representative of the spatial attention mechanism. CBAM fuses channels and spatial attention mechanisms and has good performance. We inserted CBAM [27] into the neck structure for more efficient feature extraction. From Fig. 9a, CBAM extracts feature information through the channel module and spatial module and uses serial structure to fuse feature information. From Fig. 9b, first, the input feature map generates the channel weights through the channel module and multiplies the obtained weights with the input to generate the channel map. Next, import the channel feature map into the spatial module to generate spatial weights and multiply the weights with the imported feature map to generate the spatial map. Finally, the final weighted feature map and the original input are added element by element to obtain the final output result, and the detailed CBAM structure is shown in Fig. 9. The features extracted by the shallow filter contain more specific feature information. Therefore, we integrated CBAM in the shallow layer to learn and select important features from shallow feature maps and improve the quantitative performance of ships.

The proposed RA-PANet model is a variant model of FPN. The FPN model is designed in a top-down manner, acquiring fine-grained pyramids of features by using horizontal connections. An important weakness of existing FPN-based models is that deep semantic features are used more than shallow features, making the detection of small ships inaccurate. This is due to the fact that the characteristics of small objects have been smeared deep under the pooling operation. As can be seen from Fig. 10a, the features extracted by the shallow filter contain more specific feature information, such as edges, textures, and shapes, which are more useful than abstract features extracted by the deep filter. In addition, as shown in Fig. 10b, small ships have more pixels in the shallow feature map, and these pixels contain more features of small ships. Therefore, based on the FPN model, this paper combines shallow high-resolution feature maps with deep low-resolution feature maps to improve the detection effect of large and small ships in SAR images.

Another problem with FPN models is that it is not clear which features are more useful for inspecting multiscale vessels. Therefore, this paper proposed an improved RA-PANet weighted feature pyramid structure on the basis of PANet, integrates RFB and ASFF on the basis of PANet, learns and selects important features from multi-scale feature maps, and improves the detection performance of ships. This paper aims to combine shallow and deep feature maps to adaptively select important feature maps from multi-scale feature maps, and specially designed a multi-scale feature pyramid network (RA-PANet) for the accurate detection of multi-scale ships in SAR images.

We added RFB [33] feature enhancement modules at each output to enhance feature expression and improve the capability of multi-scale predictions. RFB is a feature extraction module that draws on the inception idea in structure and adds atrous convolution based on inception, thus effectively increasing the receptive field. RFB Block first forms a multi-branched structure through convolution-al layers of different sizes, and then atrous convolution is used to increase the receptive field, the specific structure is shown in Fig. 11.

One of the main drawbacks of the feature pyramid is inconsistencies between features at different scales, especially for one-stage detectors. To this end, we introduce an ASFF

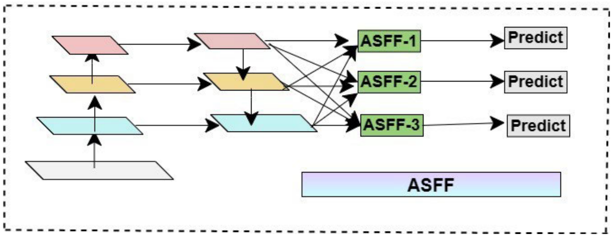


FIGURE 12. The detailed adaptively spatial feature fusion (ASFF) network structure.

feature fusion strategy to address inconsistencies within the feature pyramid in one-stage detectors. The ASFF strategy filters features and retains useful information for fusional features. For a feature of one level, first, adjust the features of the other levels to the same resolution and integrate simply, and then train to find the best way to blend. At each spatial location, different levels of features are adaptively blended together [34]. Fig. 12 shows the ASFF structure.

From Fig. 12, we can see that the fused vector is a weighted combination of the vectors of the first three feature maps, and the coefficients (the spatial importance weights of the feature map) are learned by the network adaptively, and they are shared across all channels.

We set the feature map to a different level l ($l \in 0, 1, 2$) based on the input map dimensions and the corresponding feature map is called x_l. This paper sets level l ($l \in 0, 1, 2$) to 256, 512, and 1024 depending on the characteristics of the input size and SAR vessel target. The output of the ASFF module is defined as follows:

$$y^l = \alpha^l * x^{1\rightarrow \text{æ}l} + \beta^l * x^{2\rightarrow \text{æ}l} + \gamma^l * x^{3\rightarrow \text{æ}l}, \quad 1 = 1, 2, 3, \tag{4}$$

where $x^{n\rightarrow l}$ represents the feature vector after resizing the feature of level n to level l. α, β, γ are the learning rate of level l, level $l+1$, and level $l+2$, and the definitions are as follows:

$$\alpha^l + \beta^l + \gamma^l = 1, \tag{5}$$

$$\alpha^l = \frac{e^{\lambda^l_\alpha}}{e^{\lambda^l_\alpha} + e^{\lambda^l_\beta} + e^{\lambda^l_\gamma}}, \tag{6}$$

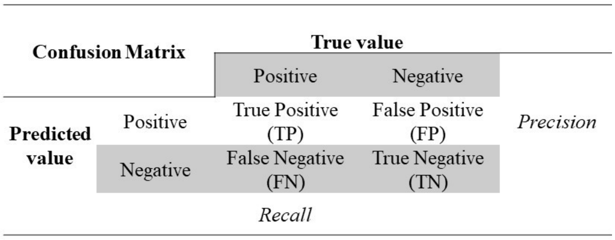
where λ_α, λ_β and λ_γ are obtained when backpropagating and updating the network and are compressed between [0,1] by softmax. In equation (4), y^l represents the output of ASFF-l. Taking ASFF-1 as an example, y^1 is obtained by the weighted sum of $x^{1\rightarrow \text{æ}1}$, $x^{2\rightarrow \text{æ}1}$, and $x^{3\rightarrow \text{æ}1}$. To be specific, $x^{1\rightarrow \text{æ}1}$ is the same as x^1. For $x^{2\rightarrow \text{æ}1}$, x^2 is twice the size of x^1, so 3*3 convolution with stride = 2 is used to convert x^2 to $x^{2\rightarrow \text{æ}1}$. We should use 3*3 convolution and the max pool operation to lower the size of x^3 and obtain $x^{3\rightarrow \text{æ}1}$.

We used the GIOU loss function to calculate the regression loss, where the union intersection (IOU) loss represents the difference in the intersection ratio between the prediction box and the real box. In equation 7, A and B represent the prediction box and the truth box. C is the smallest box containing A and B. GIOU losses are defined as follows:

$$L_{giou} = 1 - IOU + \frac{|C \backslash (A \cup \text{ú}B)|}{|C|}. \tag{7}$$

C. EXPERIMENTAL SETTINGS AND PERFORMANCE METRICS

The experiments are based on the Pytorch 1.9.1 framework and are computed using an NVIDIA RTX3090 (with 24GB of video memory) graphics processing unit (GPU) and CUDA11.1 environment. We used the lightest YOLOv5s [35], and network improvements are made on this basis. Our study also used the officially announced YOLOv5s [35] pre-training weights and uses the SGD optimizer to iteratively update the network parameters. In the training process, we set the momentum parameter to 0.937 and batch size to 16 and trained 300 epochs. We used a periodic learning rate and We used periodic learning rate and Warm-Up method to warm up the learning rate, where the initial lr0 was set to 0.01. In the Warm-Up phase, the learning rate of each iteration was updated to 0.1 using linear interpolation. After that, we updated the learning rate using the cosine annealing algorithm, and finally, the learning rate dropped to 0.002. For the performance metrics, we used the three evaluation metrics such as recall, precision, and mean average precision (mAP). We calculated various evaluation metrics based on the result of the confusion matrix as presented in Fig. 13.



Confusion Matrix		True value		
		Positive	Negative	
Predicted value	Positive	True Positive (TP)	False Positive (FP)	*Precision*
	Negative	False Negative (FN)	True Negative (TN)	
		Recall		

FIGURE 13. Confusion matrix.

The confusion matrix shown in Fig. 13 can be used to examine the accuracy of any classification task's predictions. The correct prediction is represented by the TP and TN boxes, while the incorrect prediction is represented by the FP and FN boxes. The proposed CRAS-YOLO model is thus more accurate as TP and TN values increase. The confusion matrix's elements (TP, FP, TN, and FN) were then used to calculate a more thorough evaluation of the proposed CRAS-YOLO model.

Recall (R), precision (P), mean average precision (mAP), parameters ($Params$), and $FLOPs$ are defined as follows:

$$R = \frac{TP}{TP + FN}, \tag{8}$$

$$P = \frac{TP}{TP + FP}, \tag{9}$$

$$mAP = \int_0^1 P(R)dR, \tag{10}$$

TABLE 1. The experimental results of CRAS-YOLO as compared to other models.

Model	CBAM	RA-PANet	P(%)	R(%)	mAP0.5(%)	mAP0.5-0.95(%)
Baseline	-	-	87.7	86.0	90.6	73.2
Baseline + CBAM	√		88.9	86.2	91.2	73.6
Baseline +RA-PANet		√	86.8	86.8	91.6	73.8
CRAS-YOLO	√	√	90.4	88.6	92.1	74.8

TABLE 2. Comparison results of different neck structures.

Model	P(%)	R(%)	mAP0.5(%)	mAP0.5-0.95(%)
FPN	87.7	85.8	89.8	69.8
PANet (baseline)	87.7	86.0	90.6	73.2
BiFPN	65.8	74.9	76.3	57.2
RA-PANet	86.8	86.8	91.6	73.8

$$Params = k_H \times k_W \times C_{in}/g \times C_{out}, \qquad (11)$$

$$FLOPs = (2 \times k_H \times k_W \times C_{in}/g - 1) \times C_{out} \times H_{out} \times W_{out}. \qquad (12)$$

In equation (10), $P(R)$ is the precision and recall curve. In this experiment, the detection threshold (intersection over union/IoU) is set to 0.5-0.95. In this study, we used the mAP value to measure the detection accuracy of the model.

We also introduced some other indexes, such as speed to evaluate the detection speed, parameters, and FLOPs to describe model complexity. Speed represents the time that takes to process an image. While the parameters and the flops can be used to describe the complexity of the model. Their calculation formulas are shown in Formula (11) and (12) respectively. In equations (11) and (12), $k_H \times k_W \times C$ is the kernel size, C_{out} is the output channels, g is the number of group convolutions, $C_{out} \times H_{out} \times W_{out}$ is the total number of units included in the output feature map.

III. RESULTS AND DISCUSSIONS

A. EXPERIMENTAL RESULTS

This section explains the results of the proposed CRAS-YOLO approach. In this study, the CRAS-YOLO approach which consisted of CBAM, RFB, and ASFF (RA-PANet) is compared with its baseline model including YOLOv5s, with/without CBAM as well as RA-PANet. Based on the YOLOv5s method, we selected CBAM and RA-PANet feature pyramid network as independent variable modules and adapt the control variable method to study the improvement of each module on the ArtificialSAR-Vessel dataset. Through several experiments, we verified the capacity of different modules in CRAS-YOLO through ablation experiments [42]. Table 1 shows the experimental results of CRAS-YOLO as compared to other models.

It can be seen that, compared with the original YOLOv5s (baseline) model, the precision (P) of the CRAS-YOLO model is improved by 2.7%, the recall (R) rate is increased by 2.6%, the mAP (0.5) value is increased by 1.5%, and the mAP (0.5-0.95) is improved by 0.6%, and the improved CRAS-YOLO performs better than the original YOLOv5s algorithm.

Both the CBAM attention mechanism and the RA-PANet feature pyramid network can enhance the capacity of the algorithm to some extent. Among them, CBAM has a certain degree of improvement in precision, recall, mAP (0.5), and mAP (0.5-0.95), which are increased by 1.2%, 0.2%, 0.6%, and 0.4% respectively. The RA-PANet feature pyramid network improved by 0.8%, 1.0%, and 0.6% in the recall, mAP (0.5), and mAP (0.5-0.95), respectively, and the precision decreased slightly, down 0.9%. Overall, the CBAM attention mechanism and the RA-PANet feature pyramid network have enhanced the capacity of the detection model.

We also investigated the impact of different neck structures in YOLOv5s, including FPN, PANet (baseline), BiFPN, and RA-PANet structures, to verify the performance of each neck structure. As presented in Table 2, the results revealed that the performance of the BiFPN structure is poor, and its precision, recall, mAP (0.5) and mAP (0.5-0.95) values are the lowest. The performance of the RA-PANet structure is the best, while the performance of the FPN and PANet structure is slightly lower than that of the RA-PANet. In short, considering the model detection capability, the RA-PANet structure is more suitable for multi-scale vessel detection.

Furthermore, we also conducted experiments on the effect of the attention mechanism on the RA-PANet network. Table 3 showed the experimental effect of adding different attention mechanisms to the backbone of the RA-PANet network, including not adding an attention mechanism (baseline) and adding SE, CA, ECA, and CBAM, to find the best-performing one. The results shown in Table 3 confirmed that SE and CA attention mechanisms do not play an effective role in improving detection accuracy, while ECA and CBAM can improve detection accuracy to a certain extent, of which CBAM performs better. Compared with the baseline, the detection precision is increased by 3.6%, the recall is increased by 1.8%, mAP (0.5) is increased by 0.5%, and mAP (0.5-0.95) is increased by 1.0%.

TABLE 3. Comparison results of different attention mechanisms in the backbone.

Model	P(%)	R(%)	mAP0.5(%)	mAP0.5-0.95(%)
No (Baseline)	86.8	86.8	91.6	73.8
SE	86.9	87.7	91.5	73.1
CA	88.4	84.9	89.9	72.4
ECA	89.3	85.0	91.9	73.6
CBAM	90.4	88.6	92.1	74.8

TABLE 4. Comparison results of different prediction models.

Model	P(%)	R(%)	mAP0.5(%)	mAP0.5-0.95(%)	Flops(G)	Params(M)	Speed(ms)
YOLOv3 [39]	91.3	86.0	91.4	69.7	23.1	9.3	8.8
YOLOv3-Tiny	85.0	77.8	85.5	55.8	12.9	7.7	4.3
YOLOv3-SPP	90.0	87.1	91.1	69.3	23.4	9.6	8.5
YOLOv4 [40]	89.1	88.1	91.2	72.1	20.5	10.2	10.1
YOLOv4-Tiny	87.6	82.1	89.3	69.4	11.3	8.7	5.6
YOLOv5s [35]	87.7	86.0	90.6	73.2	18.0	8.0	10.9
CRAS-YOLO	90.4	88.6	92.1	74.8	19.7	10.3	11.3

TABLE 5. Comparison between SAR ship detection methods on the SSDD dataset.

Model	P(%)	R(%)	mAP0.5(%)	mAP0.5-0.95(%)	Flops(G)	Params(M)	Speed(ms)
YOLOv3 [39]	96.8	93.8	96.7	68.5	23.1	9.3	4.9
YOLOv3-Tiny	96.1	92.1	96.1	66.8	12.9	7.7	2.9
YOLOv3-SPP	97.3	94.3	96.9	69.6	23.4	9.6	7.6
YOLOv4 [40]	97.1	94.6	97.1	70.9	20.5	10.2	7.3
YOLOv4-Tiny	96.7	93.5	95.9	68.1	11.3	8.7	3.6
YOLOv5s [35]	96.8	94.1	97.6	71.3	18.0	8.0	6.1
FBR-NET [41]	86.73	87.14	83.76	-	-	-	-
SSGE-NET [42]	88.12	88.97	85.63	-	-	-	-
ARP-NET [43]	89.36	89.73	88.42	-	-	-	-
MNE-NET [44]	94.77	82.72	91.7	-	-	-	-
CRAS-YOLO	97.3	95.5	98.7	72.0	19.7	10.3	7.9

In addition, to verify the reliability of the CRAS-YOLO model's performance, the performance of CRAS-YOLO and other models are compared on the ArtificialSAR-Vessel dataset, and the results are shown in Table 4.

Compared with other prediction models, our proposed CRAS-YOLO model has a good performance in terms of the model detection capability, with model detection precision reaching 90.4%, recall rate reaching 88.6%, mAP (0.5) reaching 92.1%, and mAP (0.5-0.95) reaching 74.8%. Overall, the proposed CRAS-YOLO model has a good detection performance to meet the requirements of ship detection. For each different Yolo model, they have different complexities and different detection accuracy. YOLOv3-tiny has the smallest number of parameters which is 7.7 M and the fastest calculation speed which is 4.3 ms, while CRAS-YOLO has the highest detection accuracy with an acceptable range of model complexity of which the model Flops and Parameters are 19.7 G and 10.3 M.

At the same time, to verify the effectiveness of the proposed CRAS-YOLO model, we performed additional experiments on another Sar ship detection publicly available dataset called SSDD [14]. The detailed experimental results are shown in Table 5. Table 5 revealed that, as compared with the original YOLOv5s, the mAP value of our proposed CRAS-YOLO is up to 98.7% which increased by 1.1%, the Flops is increased from 18.0 to 19.7 G, while the Parameters is increased from 8.0 to 10.3 M, the Speed is also increased from 6.1 to 7.9 ms. Additionally, as presented in Table 5, we compared the results of our proposed CRAS-YOLO model in the SSDD dataset with some previous studies such as feature balancing and refinement network (FBR-NET) [41], spatial shuffle-group enhance (SSE) attention module in

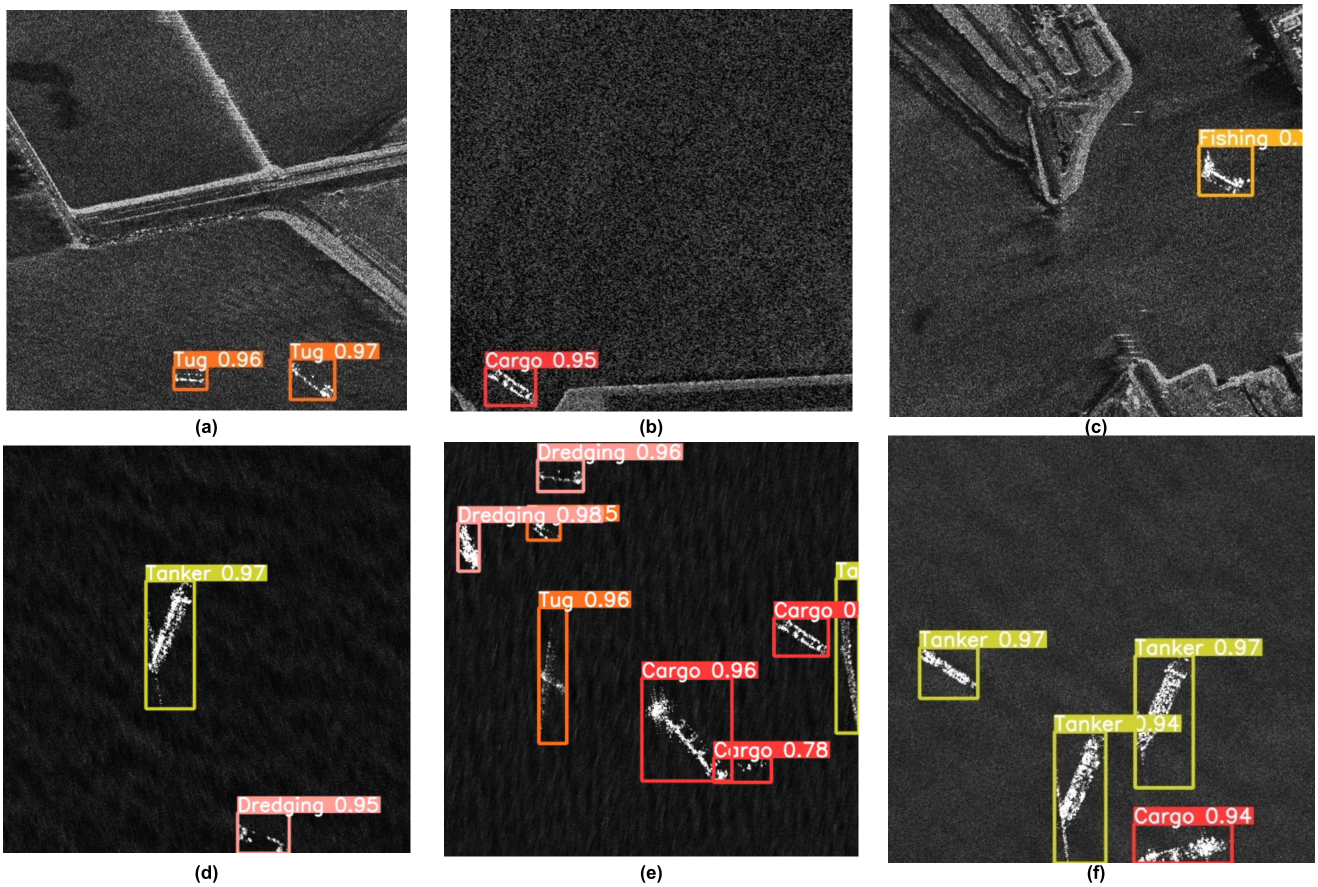


FIGURE 14. Ship detection results based on the proposed CRAS-YOLO model: (a), (b), and (c) are the ship test results in a complex coastal background, and (d), (e), and (f) are the ship test results with a sparse distribution in the deep sea, respectively.

CenterNet (SSGE-NET) [42], attention receptive pyramid network (ARP-NET) [43] and MobileNetV3 block module in YOLOv5 networks (MNE-NET) [44]. The comparison study revealed that our proposed CRAS-YOLO outperformed other previous studies in terms of precision, recall, and mAP by achieving scores of up to 97.3%, 95.5%, and 98.7% respectively. Overall, the proposed CRAS-YOLO can increase the ship detection accuracy to some extent with a little increase in the model complexity and size.

B. SHIP DETECTION RESULTS

Fig. 14 shows the ship detection results of the proposed CRAS-YOLO model on the ArtificialSAR-Vessel dataset, where a, b, and c are the ship test results in a complex coastal background, while d, e, and f are the ship test results with a sparse distribution in the deep sea respectively. We can see that the proposed model not only can effectively locate ships of different sizes on the shore but also can accurately locate deep-sea ships with sparse distribution and small sizes. This model has good detection performance and deployability. Fig. 15a and 15b show the precision-recall curves of CRAS-YOLO and YOLOv5s in each category respectively. As can be seen in Fig. 15a and b, the CRAS-YOLO model performed better than YOLOv5s in 4 categories, including cargo, dredging, fishing, and tanker. In particular, the mAP of CRAS-YOLO in the tanker category is 8.6% higher than that of YOLOv5s. The CRAS-YOLO's APs in the cargo, dredging, and fishing categories increase by 4.1%, 4.4%, and 3.5%, respectively. For the other category, CRAS-YOLO is only slightly down 1.5% in mAP than YOLOv5s. Thus, in the ArtificialSAR-Vessel dataset, CRAS-YOLO has better detection capability than YOLOv5s.

Fig. 16a shows the mAP (IOU = 0.5) curves for YOLOv5s and CRAS-YOLO on the ArtificialSAR-Vessel dataset. It can be seen that the mAP values of YOLOv5s grew faster than CRAS-YOLO at the beginning of training. As the training progressed, the mAP values of YOLOv5s and CRAS-YOLO tended to coincide. Finally, the mAP values of CRAS-YOLOv5s were higher than those of YOLOv5s. In summary, after a certain number of training, the improved CARS-YOLO model can achieve better detection capability.

Fig. 16b shows the training loss curves for YOLOv5s and CRAS-YOLO on the Arti-ficialSAR-Vessel dataset. It can be seen that the training loss of YOLOv5s declined faster than CRAS-YOLO at the beginning of training. As the training progressed, the training loss of CRAS-YOLOv5s descend

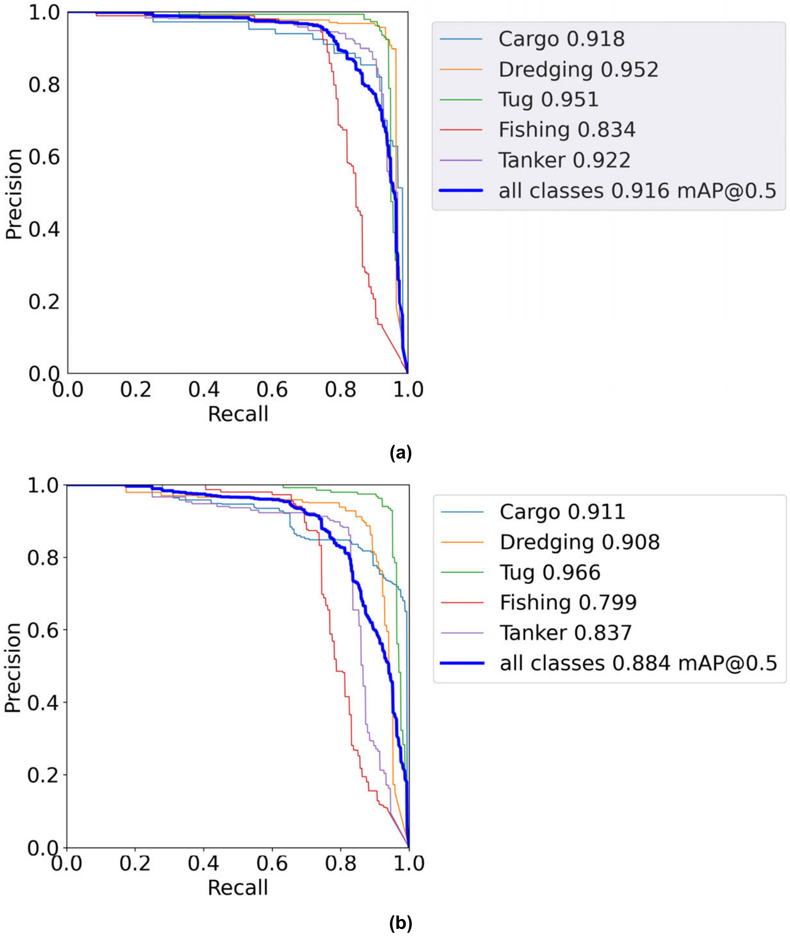


FIGURE 15. The precision-recall curves of (a) the proposed CRAS-YOLO and (b) the original YOLOv5s model.

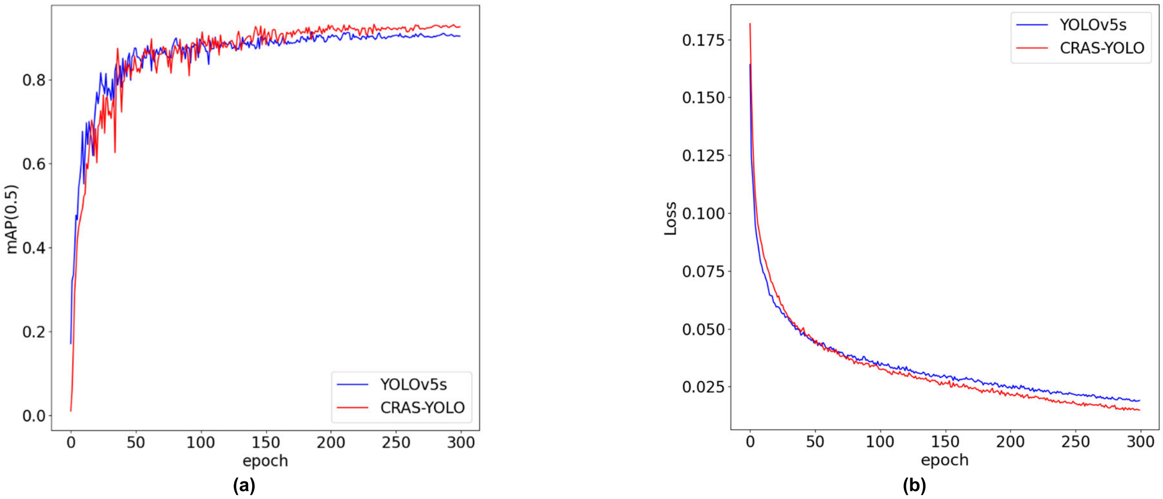


FIGURE 16. The training (a) mAP(0.5) values and (b) loss of different models.

faster than those of YOLOv5s. In summary, after a certain number of training, the improved CARS-YOLO model can better reduce detection loss.

IV. CONCLUSION AND FUTURE WORKS

In this study, we generated a novel dataset called the ArtificialSAR-Vessel dataset which consists not only of ship/vessel location information but also its ship/vessel category/type such as cargo, dredger, tug, fishing, and tanker. We also proposed a novel ship detection and classification model called CRAS-YOLO, which consisted of a CBAM, RFB, and ASFF based on the YOLOv5s algorithm. The experimental results confirmed that as compared to the original YOLOv5s and other models, the proposed CRAS-YOLO performed better in SAR ship detection, with an average improvement of up to 3.6% in detection precision, a 1.8% increase in recall, an 0.5% increase in mAP (0.5) and a 1.0% improvement in mAP (0.5-0.95). Additional experiments and comparison studies in another Sar Ship Detection (SSDD) dataset revealed that the proposed CRAS-YOLO model outperformed other models and previous studies' results with precision, recall, and mAP scores of up to 97.3%, 95.5%, and 98.7% respectively. Overall, the proposed model has strong generalization, excellent performance on ship target detection, as well as practical significance in the application field of multi-category ship detection. Finally, the results of our study could be used by decision-makers to develop a ship detection model that can accurately and automatically detect and classify multiple vessels based on satellite imagery.

Although CRAS-YOLO has some improvement over the original algorithm, our study only divided vessels/ships into five types/categories, lacking more types of ship data, and the detection capability, as well as the complexity of the network, needs to be further enhanced.

Currently, there are two challenges to ship inspection. Firstly, the image contour information of the SAR ship detection dataset is not clear and the background is complex, resulting in low ship detection capabilities. Second, some ship detection models are large, which results in poor model deployability. In the future, on the one hand, we will continue to expand our novel ArtificialSAR-Vessel dataset, by enriching it with different types and quantities of ships, and models that can detect more types of ships to meet practical applications. It is also necessary to further improve the accuracy of the model which can more accurately classify and detect various types of ships. On the other hand, the model complexity can be further lightened without sacrificing accuracy. Therefore, other lightweight model approaches, such as lightweight backbone networks, knowledge distillation, and network pruning can be further explored in the future to cope with the current model limitation.

REFERENCES

[1] Z. M. Li, L. Cheng, D. M. Zhu, Z. J. Yan, C. Ji, Z. X. Duan, M. Jing, N. Li, S. K. Dongye, Y. R. Song, and J. H. Liu, "Deep learning and spatial analysis based port detection," *Laster Optoelectron. Prog.*, vol. 58, no. 20, 2021, Art. no. 2028002, doi: 10.3788/LOP202158.2028002.

[2] T. Zhang and X. Zhang, "High-speed ship detection in SAR images based on a grid convolutional neural network," *Remote Sens.*, vol. 11, no. 10, p. 1206, May 2019, doi: 10.3390/rs11101206.

[3] C. C. Wackerman, K. S. Friedman, W. G. Pichel, P. Clemente-Colón, and X. Li, "Automatic detection of ships in RADARSAT-1 SAR imagery," *Can. J. Remote Sens.*, vol. 27, no. 5, pp. 568–577, Oct. 2001, doi: 10.1080/07038992.2001.10854896.

[4] P. W. Vachon, S. J. Thomas, J. Cranton, H. R. Edel, and M. D. Henschel, "Validation of ship detection by the RADARSAT synthetic aperture radar and the ocean monitoring workstation," *Can. J. Remote Sens.*, vol. 26, no. 3, pp. 200–212, Jun. 2000, doi: 10.1080/07038992.2000.10874770.

[5] J. N. Kapur, P. K. Sahoo, and A. K. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Comput. Vis. Graph. Image Process.*, vol. 29, no. 3, pp. 273–285, Mar. 1985, doi: 10.1016/0734-189X(85)90125-2.

[6] B. D. Wardlow, S. L. Egbert, and J. H. Kastens, "Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains," *Remote Sens. Environ.*, vol. 108, no. 3, pp. 290–310, Jun. 2007, doi: 10.1016/j.rse.2006.11.021.

[7] S. Wang, M. Wang, S. Yang, and L. Jiao, "New hierarchical saliency filtering for fast ship detection in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 351–362, Jan. 2017, doi: 10.1109/TGRS.2016.2606481.

[8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 580–587, doi: 10.1109/CVPR.2014.81.

[10] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1440–1448, doi: 10.1109/ICCV.2015.169.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Computer Vision—ECCV 2016*, vol. 9905, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.

[13] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2999–3007, doi: 10.1109/ICCV.2017.324.

[14] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proc. SAR Big Data Era, Models, Methods Appl. (BIGSARDATA)*, Beijing, China, Nov. 2017, pp. 1–6, doi: 10.1109/BIGSARDATA.2017.8124934.

[15] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, p. 765, Mar. 2019, doi: 10.3390/rs11070765.

[16] X. Sun, Z. R. Wang, Y. R. Sun, W. H. Diao, Y. Zhang, and K. Fu, "AIR-SARShip-1.0: High-resolution SAR ship detection dataset," *J. Radars*, vol. 8, no. 6, pp. 852–862, 2019, doi: 10.12000/JR19097.

[17] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234–120254, 2020, doi: 10.1109/ACCESS.2020.3005861.

[18] S. Lei, D. Lu, X. Qiu, and C. Ding, "SRSDD-v1.0: A high-resolution SAR rotation ship detection dataset," *Remote Sens.*, vol. 13, no. 24, p. 5104, Dec. 2021, doi: 10.3390/rs13245104.

[19] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.

[20] T. Zhang, X. Zhang, J. Shi, and S. Wei, "High-speed ship detection in SAR images by improved YOLOv3," in *Proc. 16th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process.*, Chengdu, China, Dec. 2019, pp. 149–152, doi: 10.1109/ICCWAMTIP47768.2019.9067695.

[21] Z. Hong, T. Yang, X. Tong, Y. Zhang, S. Jiang, R. Zhou, Y. Han, J. Wang, S. Yang, and S. Liu, "Multi-scale ship detection from SAR and optical imagery via a more accurate YOLOv3," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6083–6101, 2021, doi: 10.1109/JSTARS.2021.3087555.

[22] Q. Gong and Y. Wu, "Improved YOLOv4-tiny algorithm based on cascade residual dilated fusion," in *Proc. 20th Int. Symp. Distrib. Comput. Appl. Bus. Eng. Sci. (DCABES)*, Nanning, China, Dec. 2021, pp. 136–140, doi: 10.1109/DCABES52998.2021.00041.

[23] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 936–944, doi: 10.1109/CVPR.2017.106.

[24] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 8759–8768, doi: 10.1109/CVPR.2018.00913.

[25] G. Ghiasi, T.-Y. Lin, and Q. V. Le, "NAS-FPN: Learning scalable feature pyramid architecture for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 7029–7038, doi: 10.1109/CVPR.2019.00720.

[26] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 10778–10787, doi: 10.1109/CVPR42600.2020.01079.

[27] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Computer Vision—ECCV 2018*, vol. 11211, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.

[28] Y. Chen, X. Zhang, W. Chen, Y. Li, and J. Wang, "Research on recognition of fly species based on improved RetinaNet and CBAM," *IEEE Access*, vol. 8, pp. 102907–102919, 2020, doi: 10.1109/ACCESS.2020.2997466.

[29] L. Jia, Y. Wang, Y. Zang, Q. Li, H. Leng, Z. Xiao, W. Long, and L. Jiang, "MobileNetV3 with CBAM for bamboo stick counting," *IEEE Access*, vol. 10, pp. 53963–53971, 2022, doi: 10.1109/ACCESS.2022.3175818.

[30] L. Wang, Y. Cao, S. Wang, X. Song, S. Zhang, J. Zhang, and J. Niu, "Investigation into recognition algorithm of helmet violation based on YOLOv5-CBAM-DCN," *IEEE Access*, vol. 10, pp. 60622–60632, 2022, doi: 10.1109/ACCESS.2022.3180796.

[31] Y. Zhang, Y. Chen, Y. Wang, Q. Liu, and A. Cheng, "CSI-based human activity recognition with graph few-shot learning," *IEEE Internet Things J.*, vol. 9, no. 6, pp. 4139–4151, Mar. 2022, doi: 10.1109/JIOT.2021.3103073.

[32] X. Hou, W. Ao, Q. Song, J. Lai, H. Wang, and F. Xu, "FUSAR-ship: Building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition," *Sci. China Inf. Sci.*, vol. 63, no. 4, Apr. 2020, Art. no. 140303, doi: 10.1007/s11432-019-2772-5.

[33] S. Liu, D. Huang, and Y. Wang, "Receptive field block net for accurate and fast object detection," in *Computer Vision—ECCV 2018*, vol. 11215, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, pp. 404–419, doi: 10.1007/978-3-030-01252-6_24.

[34] S. Liu, D. Huang, and Y. Wang, "Learning spatial fusion for single-shot object detection," Nov. 2019, *arXiv:1911.09516*. Accessed: Oct. 24, 2022.

[35] Ultralytics. *YOLOv5*. Accessed: Nov. 1, 2021. [Online]. Available: https://github.com/ultralytics/yolov5

[36] G. Ghiasi, Y. Cui, A. Srinivas, R. Qian, T.-Y. Lin, E. D. Cubuk, Q. V. Le, and B. Zoph, "Simple copy-paste is a strong data augmentation method for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 2917–2927, doi: 10.1109/CVPR46437.2021.00294.

[37] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.

[38] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Seattle, WA, USA, Jun. 2020, pp. 1571–1580, doi: 10.1109/CVPRW50498.2020.00203.

[39] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," Apr. 2018, *arXiv:1804.02767*. Accessed: Oct. 24, 2022.

[40] J. Jiang, X. Fu, R. Qin, X. Wang, and Z. Ma, "High-speed lightweight ship detection algorithm based on YOLO-V4 for three-channels RGB SAR image," *Remote Sens.*, vol. 13, no. 10, p. 1909, May 2021, doi: 10.3390/rs13101909.

[41] J. Fu, X. Sun, Z. Wang, and K. Fu, "An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1331–1344, Feb. 2021, doi: 10.1109/TGRS.2020.3005151.

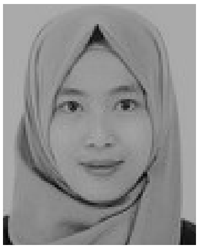
[42] Z. Cui, X. Wang, N. Liu, Z. Cao, and J. Yang, "Ship detection in large-scale SAR images via spatial shuffle-group enhance attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 379–391, Jan. 2021, doi: 10.1109/TGRS.2020.2997200.

[43] Y. Zhao, L. Zhao, B. Xiong, and G. Kuang, "Attention receptive pyramid network for ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2738–2756, 2020, doi: 10.1109/JSTARS.2020.2997081.

[44] L. Pang, B. Li, F. Zhang, X. Meng, and L. Zhang, "A lightweight YOLOv5-MNE algorithm for SAR ship detection," *Sensors*, vol. 22, no. 18, p. 7088, Sep. 2022, doi: 10.3390/s22187088.

WENXIAO ZHAO received the bachelor's degree from Chongqing University, Chongqing, China. She is currently pursuing the master's degree with Beijing Normal University. Her current research interests include deep learning, object detection, and data analysis.

MUHAMMAD SYAFRUDIN (Member, IEEE) received the bachelor's degree from Universitas Islam Negeri Sunan Kalijaga at Yogyakarta, Yogyakarta, Indonesia, in 2013, and the Ph.D. degree from Dongguk University, Seoul, South Korea, in 2019. He has been an Assistant Professor with the Department of Artificial Intelligence, Sejong University, Seoul, since March 2022. Previously, he worked as an Assistant Professor with the Department of Industrial and Systems Engineering, Dongguk University, from 2019 to 2022. He is also an Instructor in a practical course on undergraduate topics in linear algebra, programming languages, databases, and big data systems. In 2019, he was selected and invited to participate in the world class Scholar Symposium (SCKD) Event hosted by the Ministry of Research, Technology, and Higher Education, Indonesia, to make contributions on accelerating the Indonesian national development. He has published numerous research articles in several international peer-reviewed journals, including IEEE Access, *Food Control*, *Sensors*, and *Mathematics*. His research interests include industrial artificial intelligence (IAI), industrial analytics (IA), industrial informatics (II), the Industrial Internet of Things (IIoT), and industrial big data (IBD), ranging from theory to design and implementation. He is a Reviewer for well-known journals, such as *Expert Systems with Applications*, IEEE Transactions on Industrial Informatics, *Computers & Industrial Engineering*, and *Algorithms*. He has served as the guest editor for Scopus-and-Sciences Citation Index (SCI)-indexed journals.



NORMA LATIF FITRIYANI (Member, IEEE) received the bachelor's degree from Universitas Islam Negeri Sunan Kalijaga at Yogyakarta, Yogyakarta, Indonesia, in 2014, the master's degree from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 2016, and the Ph.D. degree from Dongguk University, Seoul, South Korea, in 2021. She has been an Assistant Professor with the Department of Data Science, Sejong University, Seoul, since March 2022. She is also an Instructor in undergraduate topics in decision-making and numerical analysis. She has published numerous research articles in several international peer-reviewed journals, including IEEE Access, *Food Control*, *Sensors*, *Mathematics*, *Applied Sciences*, *Asia Pacific Journal of Marketing and Logistics*, and *Sustainability*. Her research interests include health informatics, machine learning, the Internet of Things, sensors, and image processing. She has served as the guest editor and a reviewer for many peer-reviewed international journals.