



OPEN ACCESS

EDITED BY

Hong Song,
Zhejiang University, China

REVIEWED BY

Islam Zada,
International Islamic University,
Pakistan
Mohammad Faisal,
University of Malakand, Pakistan
Yi Ma,
Ministry of Natural Resources, China

*CORRESPONDENCE

Wan Jianhua
19850014@upc.edu.cn

SPECIALTY SECTION

This article was submitted to
Ocean Observation,
a section of the journal
Frontiers in Marine Science

RECEIVED 01 November 2022

ACCEPTED 28 November 2022

PUBLISHED 13 January 2023

CITATION

Yasir M, Shanwei L, Mingming X, Hui S,
Hossain MS, Colak ATI, Wang D,
Jianhua W and Dang KB (2023) Multi-
scale ship target detection using SAR
images based on improved Yolov5.
Front. Mar. Sci. 9:1086140.
doi: 10.3389/fmars.2022.1086140

COPYRIGHT

© 2023 Yasir, Shanwei, Mingming, Hui,
Hossain, Colak, Wang, Jianhua and
Dang. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Multi-scale ship target detection using SAR images based on improved Yolov5

Muhammad Yasir¹, Liu Shanwei¹, Xu Mingming¹, Sheng Hui¹,
Md Sakaouth Hossain², Arife Tugsan Isiacik Colak³,
Dawei Wang¹, Wan Jianhua^{1*} and Kinh Bac Dang⁴

¹College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, China, ²Department of Geological Sciences, Jahangirnagar University, Dhaka, Bangladesh, ³National University International Maritime College Oman, Sahar, Oman, ⁴Faculty of Geography, VNU University of Science, Vietnam National University, Hanoi, Vietnam

Synthetic aperture radar (SAR) imaging is used to identify ships, which is a vital task in the maritime industry for managing maritime fisheries, marine transit, and rescue operations. However, some problems, like complex background interferences, various size ship feature variations, and indistinct tiny ship characteristics, continue to be challenges that tend to defy accuracy improvements in SAR ship detection. This research study for multiscale SAR ships detection has developed an upgraded YOLOv5s technique to address these issues. Using the C3 and FPN + PAN structures and attention mechanism, the generic YOLOv5 model has been enhanced in the backbone and neck section to achieve high identification rates. The SAR ship detection datasets and AirSARship datasets, along with two SAR large scene images acquired from the Chinese GF-3 satellite, are utilized to determine the experimental results. This model's applicability is assessed using a variety of validation metrics, including accuracy, different training and test sets, and TF values, as well as comparisons with other cutting-edge classification models (ARPN, DAPN, Quad-FPN, HR-SDNet, Grid R-CNN, Cascade R-CNN, Multi-Stage YOLOv4-LITE, EfficientDet, Free-Anchor, Lite-Yolov5). The performance values demonstrate that the suggested model performed superior to the benchmark model used in this study, with higher identification rates. Additionally, these excellent identification rates demonstrate the recommended model's applicability for maritime surveillance.

KEYWORDS

synthetic aperture radar (SAR), ship identification, artificial intelligence, deep learning (DL), YOLOv5S, SAR ship detection dataset (SSDD), AirSARship

1 Introduction

Detection of ships is a crucial task in the maritime industry for controlling maritime fisheries, marine transit, and rescue operations. However, some issues, such as intricate backdrop interferences, numerous ship size fluctuations, and imprecise little ship features, still pose difficulties and frequently thwart advancements in SAR ship recognition accuracy. Accurate position and trajectory determination of the target ship is essential for managing maritime traffic, recovering from maritime accidents, and the economy (Xiao et al., 2020).

According to the kinds of remote sensing technologies used, the two main categories of ship detection study are, respectively, the SAR image-based and optical satellite image-based methodologies. One of the major challenges of ship identification in optical remote sensing images is finding suitable areas from complex backgrounds fast and correctly (Wang et al., 2021). High-resolution capabilities, independence from the weather, and flight altitude independence are all attributes of SAR images. SAR's self-illumination capability ensures that they always produce high-quality images under any circumstance (Chang et al., 2019). SAR has been extensively employed in ship identification (Ma et al., 2018; Xu et al., 2021; Li et al., 2022; Yasir et al., 2022; Xiong et al., 2022), oil spill identification (Yekeen et al., 2020; Wang et al., 2022), change detection (Gao et al., 2019; Chen and Shi, 2020; Zhang et al., 2020b; Wang et al., 2022), and other fields (Niedermeier et al., 2000; Baseline and Ferraioli, 2013). Because of its broad observation range, brief observation duration, great data timeliness, and high spatial resolution (Ouchi, 2013), SAR performing a significant role in ship identification. The amount and quality of SAR data have been steadily improving recently due to the quick development of space-borne SAR-imaging technologies. As a result, many researchers are studying how to identify ships in HR SAR images (Wang et al., 2016; Li et al., 2017b; Salembier et al., 2018; Du et al., 2019; Lin et al., 2019; Wang et al., 2019; Wang et al., 2020b; Wang et al., 2020c; Wang et al., 2020d; Zhang et al., 2020c; Yasir et al., 2022). However, due to the complicated environment and other difficult issues, such as sidelobes and target defocusing (Chen et al., 2019; Han et al., 2019; Xiong et al., 2019; Yuan et al., 2020), identifying ship targets in HR SAR images is still challenging.

Deep learning (DL) technologies has enhanced quickly in recent years, enabling natural image identification. Convolutional neural networks (CNNs) were introduced into the target identification area by R-CNN (Girshick et al., 2014), and as a result, target identification has received new scientific research thoughts, and its use in SAR images has a wide range of potential applications. Currently, two stage identification approaches addressed by R-CNN, Fast R-CNN, and Faster R-CNN (Girshick, 2015; Ren et al., 2015) are the main convolutional neural network-based algorithms employed in

ship identification in SAR images. The complexity of their network topologies, the sheer number of parameters, and the slow recognition speed, however, prevent them from being able to complete ship detection tasks in the required amount of time. The target identification problem is also seen as a regression analysis task involving target location and category information by the single stage algorithms from the SSD (Liu et al., 2016) and YOLO (Redmon et al., 2016; Redmon and Farhadi, 2017; Redmon and Farhadi, 2018; Patel et al., 2022) series. They are more suited to ship identification applications that need virtually real-time identification since they output the identification results directly through a neural network model with high accuracy and speed (Willburger et al., 2020).

Although the aforementioned methods have strong detection performance, it is challenging to directly apply them to SAR ship identification. There are still a number of problem with the DL-based ship identification approaches in SAR images (Li et al., 2020; Zhang et al., 2021a). (i) Due to the unique imaging technique utilized by SAR, there is less contrast between the ocean and ship in the SAR images since there are more scattering noise and sea debris and less side flap. (ii) Different ships have various sizes and shapes which are reflected in SAR images as varying numbers of pixels, especially for tiny-scale ships. Smaller ships have less information about their whereabouts than large ships, and since they have fewer pixels, they are more susceptible to being deceived by the speckle noise in SAR images. While this is going on, the detection process becomes more complex, which lowers the accuracy of identification and recognition. (iii) SAR images cannot be directly supplied to the network for identification if the scene is large. It is anticipated that the network has now received the SAR image of the expansive landscape. The ship target will be resampled in this situation to a few or possibly only one pixel, which will significantly reduce the identification accuracy. The main goals of the current study are as follows:

- To identify optimum multi-size ship target in SAR images by modified YOLOv5 model.
- To offer the backbone extraction network a well-designed structure, a set of CSP framework and attention mechanisms have been upgraded, and the output layer has been expanded to four feature layers.
- To improve the overall performance throughout the recognition process, this improved version of the YOLOv5 model also produces effective results in a condensed amount of time with a relatively smaller database.
- To use the SSDD and AirSAR ship Datasets, two distinct and well-known datasets, in these simulations. The SSDD collection contains 1160 SAR images in total, collected by RadarSat-2, TerraSAR-X, and Sentinel-1, with resolutions ranging from 1m to 10m and

polarizations in HH, HV, VV, and VH. Gaofen-3 has collected 31 single-polarized SAR images, which are included in the AirSARship.

- To assess the suggested model's applicability utilizing cutting-edge benchmark convolutional neural network-based techniques.
- To employ several performance indicators for application evaluation reasons, including precision, accuracy, time consumption, and different training and test sets.
- To demonstrate the model's superiority the performance results would be demonstrated to the desired benchmark models (CNN-based SAR ship identification techniques).

The paper is organized as follows; Section 2 shows the study serves as an organizational framework for the remainder of the research, explaining the proposed methodology. The findings and analysis of the suggested research project are described in Section 3. Additionally, by contrasting it with other cutting-edge produced models, it has demonstrated the model's usefulness. Section 4 describes the ablation study and the paper is concluded in Section 5.

2 Proposed methodology

The target of this current study is to develop a ship detection model that could potentially function when there are inadequate hardware resources. Because of its reputation for speed and accuracy, the lightweight version of YOLO has received attention. Open source model YOLO was first presented by Joseph Redmon in 2016 (Redmon et al., 2016). It is suitable as a real-time system since it can identify things at extremely quick speeds. In this research work, the upgraded model lightweight version of YOLOv5 is used. This upgraded model resulted with higher accuracy and efficient identification capabilities (Caputo et al., 2022; Nepal and Eslamiat, 2022). Two datasets that are available in the literature, the AirSARship (Xian et al., 2019) dataset and the SSDD dataset (Li et al., 2017a), have both been considered.

2.1 Data augmentation

In order to train the model for deep learning, a lot of data is typically required. However, in practice, certain data sets are challenging to collect, leading to a small quantity of data in this category that falls short of the required data volume for deep learning. Experts have thus suggested data augmentation approaches to successfully address this issue (Najafabadi et al., 2015). The data augmentation techniques such as random rotation and mosaic was used. Given training data, mosaic randomly crops four images and stitch them together to create

one. It has the advantage of enriching the background of the image and enhancing the batch size discretely so that it can help to minimize the model dependence on a large batch size when training (Figure 1).

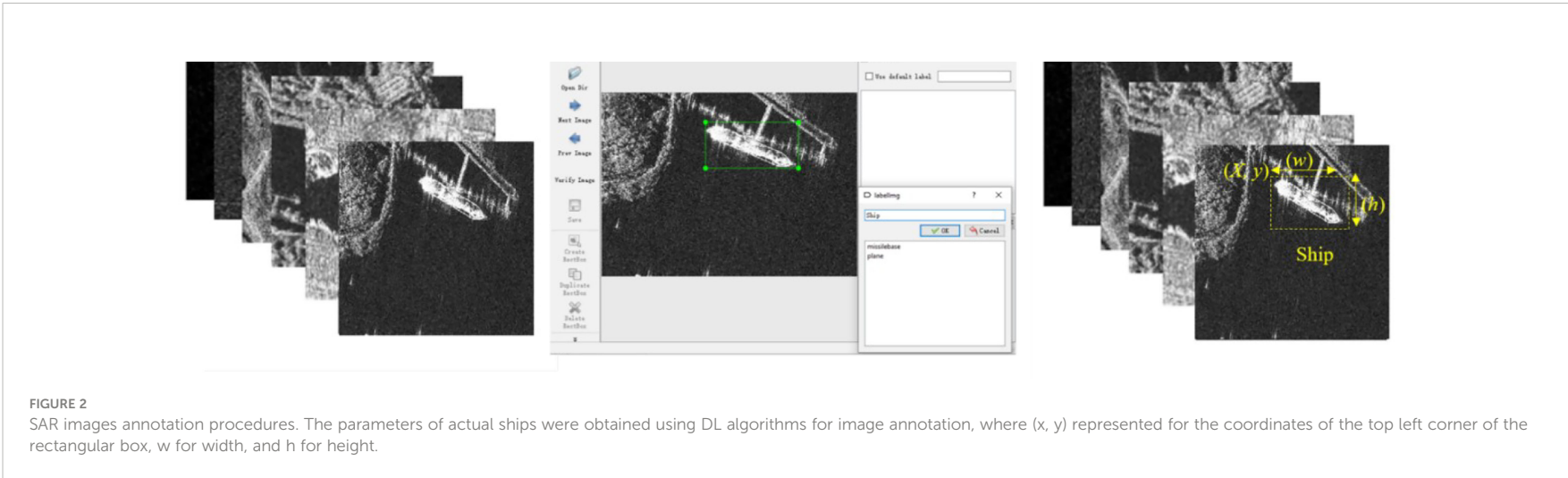
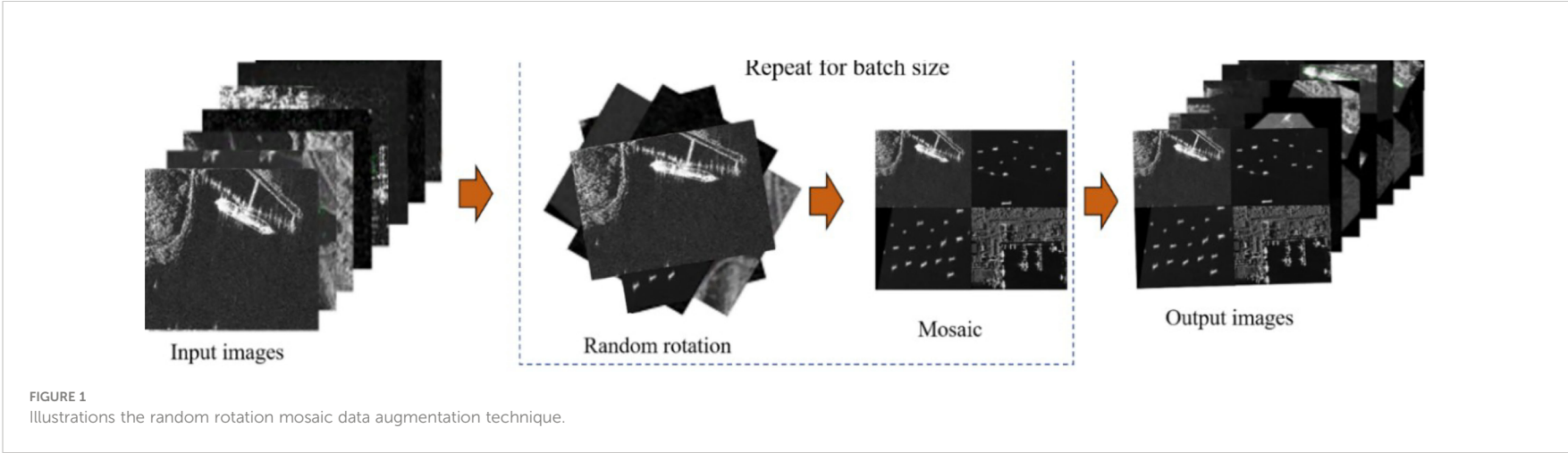
2.2 Data annotation

The images were annotated using the Labelling software, which generates a json annotation file and transforms it into a txt file. The type and number of the labeling target, the labeling image's standardized width and height, and the center point's coordinates are all information that can be found in the txt file. Figure 2 displays the labelling outcomes.

2.3 YOLOv5 network

YOLO is a regression-based technique and, despite being less accurate, is actually faster than region proposal-based methods like R-CNN (Girshick et al., 2014). The goal of YOLO is to achieve object identification by approaching it as a regression and classification issue. Identifying the bounding box coordinates for the objects in the images is the first step, and second step is to classifying the objects that are identified in a class. This is accomplished in a single step by first splitting the input images into a grid of cells, then determining the bounding box and relative confidence score for each cell's containing object.

The YOLOv5 network is one of the recent research advancements in the YOLO series of algorithms. Despite sharing a network structure with the YOLOv4 network, it is smaller, has a faster running speed and convergence speed, and uses a lightweight algorithm. Additionally, it improves precision at the same time. As a result, the YOLOv5s algorithm is used in the current study work to detect ships in SAR images. Four components make up the YOLOv5 network structure: input, backbone, neck, and prediction. The YOLOv5 framework architecture is displayed in Figure 3A. Networks can be categorized as YOLOv5l, YOLOv5m, YOLOv5x, and YOLOv5s. Their widths and depths may differ significantly, but their network structures are comparable. The network structure of YOLOv5s is the shortest, shallowest, runs the fastest, and has the least accuracy. As a result, the accuracy continues to rise, the speed of operation declines, and the other three network structures increasingly deepen and widen. Adaptive anchor box operation, mosaic data augmentation, image scaling, and CSP structure were used to process the input dataset, while focus and CSP framework were used to build the backbone. Focus increased network speed and cut down on floating-point operations (FLOPs) by clipping the input image. Figure 3A presented the focal structure. The two CSP (Wang et al., 2020a) framework that were used by YOLOv5 were CSP1_X and CSP2_X; CSP1_X was utilized for down



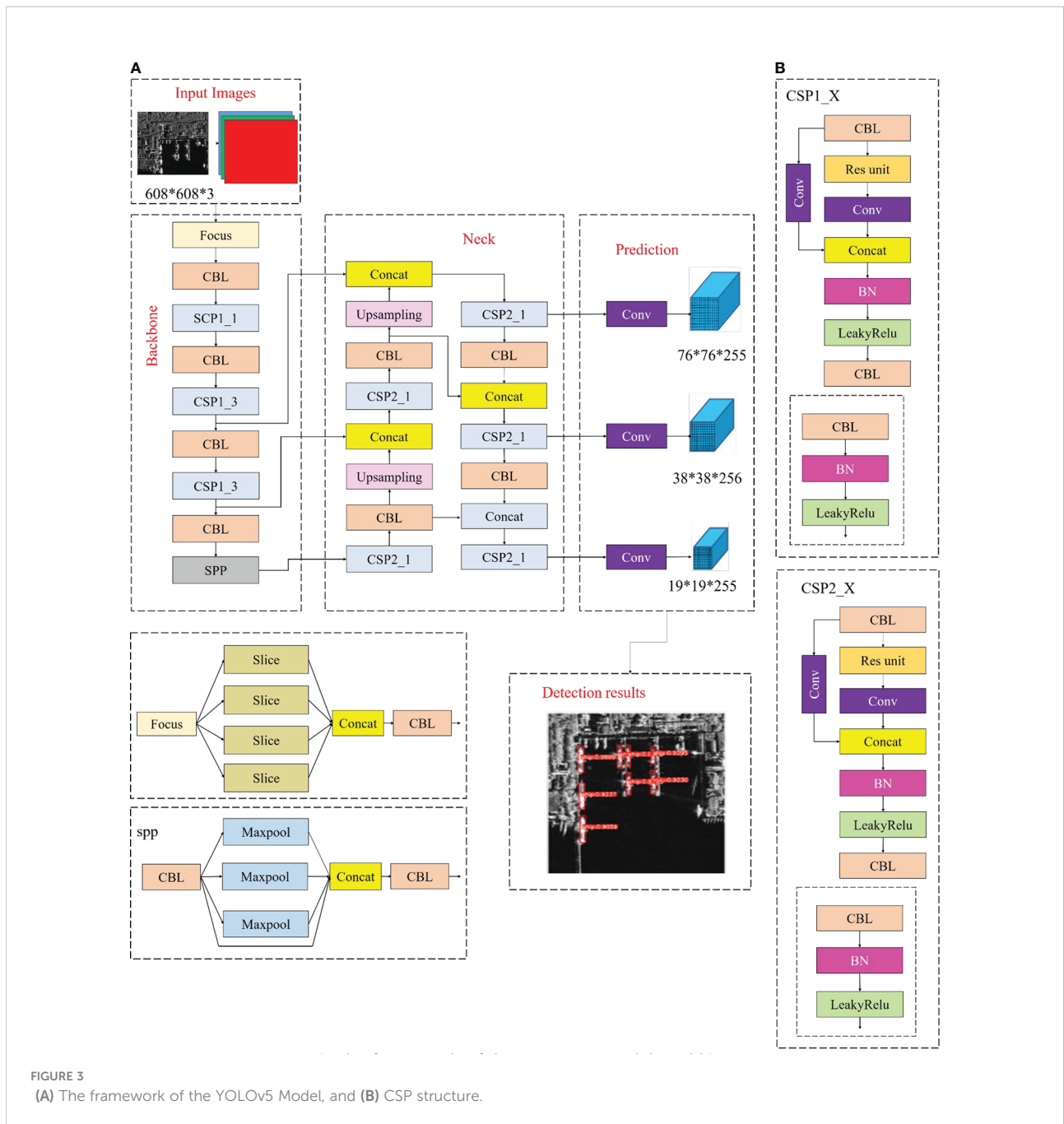


FIGURE 3 (A) The framework of the YOLOv5 Model, and (B) CSP structure.

sampling in the backbone while CSP2_ X was utilized in the neck. CSP can reduce operations while increasing the network’s capacity for learning and guaranteeing accuracy. Figure 3B depicted the two CSPs’ structures; the neck used the SPP-net and FPN + PAN framework to improve the network’s feature fusion effect, while the prediction employed the GIOU_ Loss (Rezatofighi et al., 2019), which did not only focus on the overlap between the prediction box and the ground truth but also on the non-overlapping areas. (Yu et al., 2016) found that GIOU maintains the benefits of IOU while solving its issues. The

computation for Equation (1-2) is as follows:

$$IOU = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

$$GLOU = IOU - \left| \frac{C}{\frac{A \cup B}{|C|}} \right| \tag{2}$$

Four separate networks were used, and YOLOv5 was continuously upgraded as well. Version 5.0 of YOLOv5s was used in this study project; in comparison to version 4.0, this

version modified all functional activation in the framework to SiLU (Elfving et al., 2018), eliminated the conv in the CSP, and designated it C3 as presented in Figure 4. Additionally, v5.0 has a smaller and faster network structure than v4.0.

2.4 YOLOv5 network improvement

In this section of the study, the improvements made to the YOLOv5 classifier have been described in accordance with the guidelines of the proposed research challenge. The neck and backbone parts are enhanced to produce the greatest identification outcomes.

2.4.1 Backbone improvement

It is frequently possible to combine features from various scales to obtain more meaningful object information. The high-level feature has lesser resolution and poor perception of object information, but the receptive field is bigger, which is suited for identifying huge objects. The low-level feature has higher resolution, a smaller receptive field, more texture information, and more noise. The complex background environment in the SSDD and AirSARship dataset results in some large ground objects having an inadequate detection effect. In the current research work, a set of C3 framework was used to construct the YOLOv5s backbone network. The original three sets of C3 were converted into four sets of C3 to further the network framework as a whole (Figure 4). In turn, the model’s detection accuracy may have improved as a result of the network’s increased ability to communicate and learn about larger ground objects.

2.4.2 Attentional mechanism

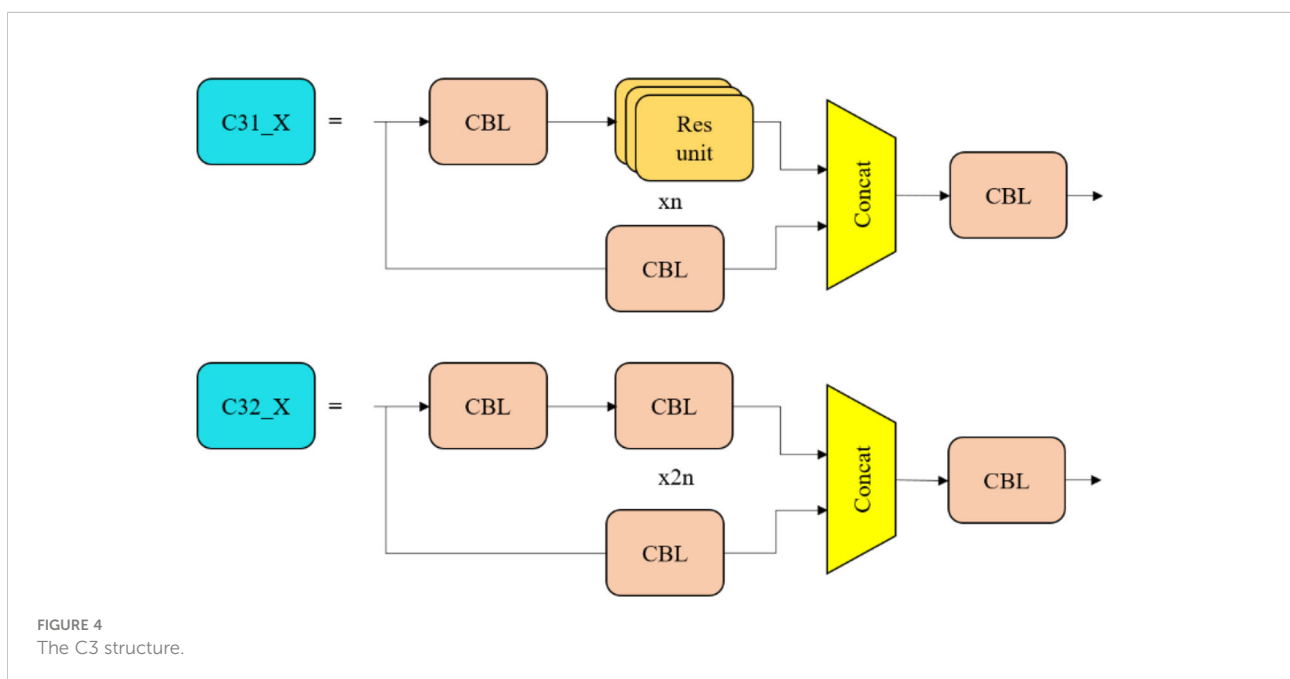
The human visual attention process is referred to as the “attention mechanism,” which concentrate on local details and blocks out redundant details. To put it another way, the network is able to identify critical information among a plethora of data due to the attention mechanism. The network performance is enhanced in this way by the addition of a small amount of computing. Figure 5 presented the increased backbone structure.

2.4.3 Neck enhancement

The neck was constructed using the FPN (Lin et al., 2017) + PAN (Liu et al., 2018) framework. This framework incorporated a bottom-up feature pyramid network after the FPN, which improved location information and semantic expression on various scales. The C32_X structure was incorporated into the neck of the YOLOv5s to enhanced the feature fusion impact of the network framework. Because of the development of a set of C3 structures in the current research work, an output layer was updated to the network’s neck to increase feature extraction. Figure 6 presented the increase structure of FPN + PAN.

2.4.4 Extending receptive field area

Each pixel in the output feature map must respond to an area in the image that is large enough for it to get information about the large object, which makes the size of the receptive field a major issue in many vision applications. Consequently, a maximum pooling layer has been chosen to be added to the space pyramid in order to improve multiple receptive field fusion and increase the accuracy of identification of tiny targets. The updated architecture is shown in Figure 7.



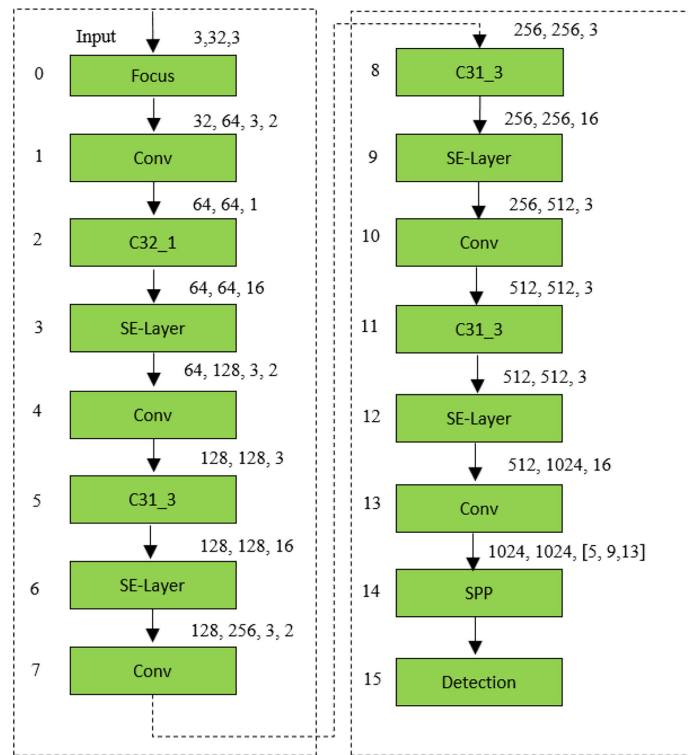


FIGURE 5
The improved backbone Network structure.

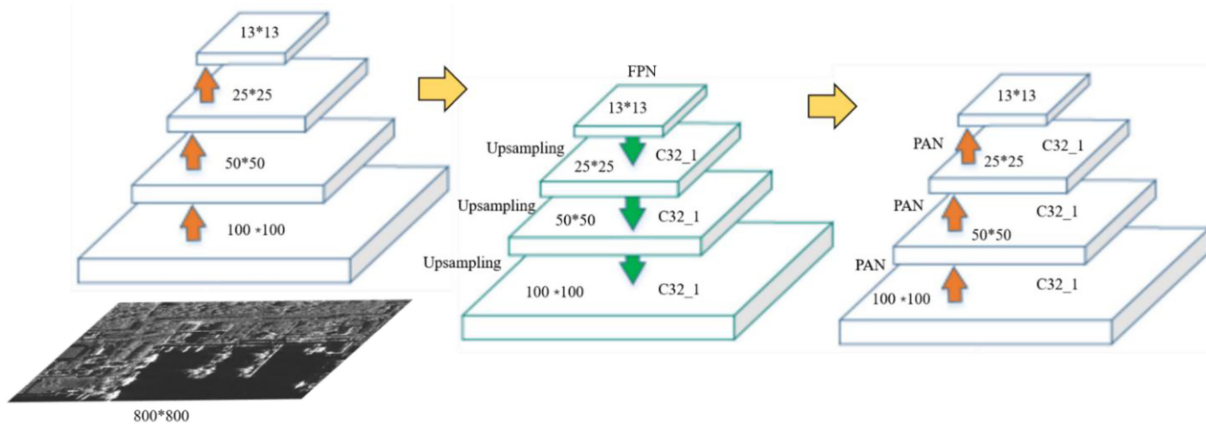
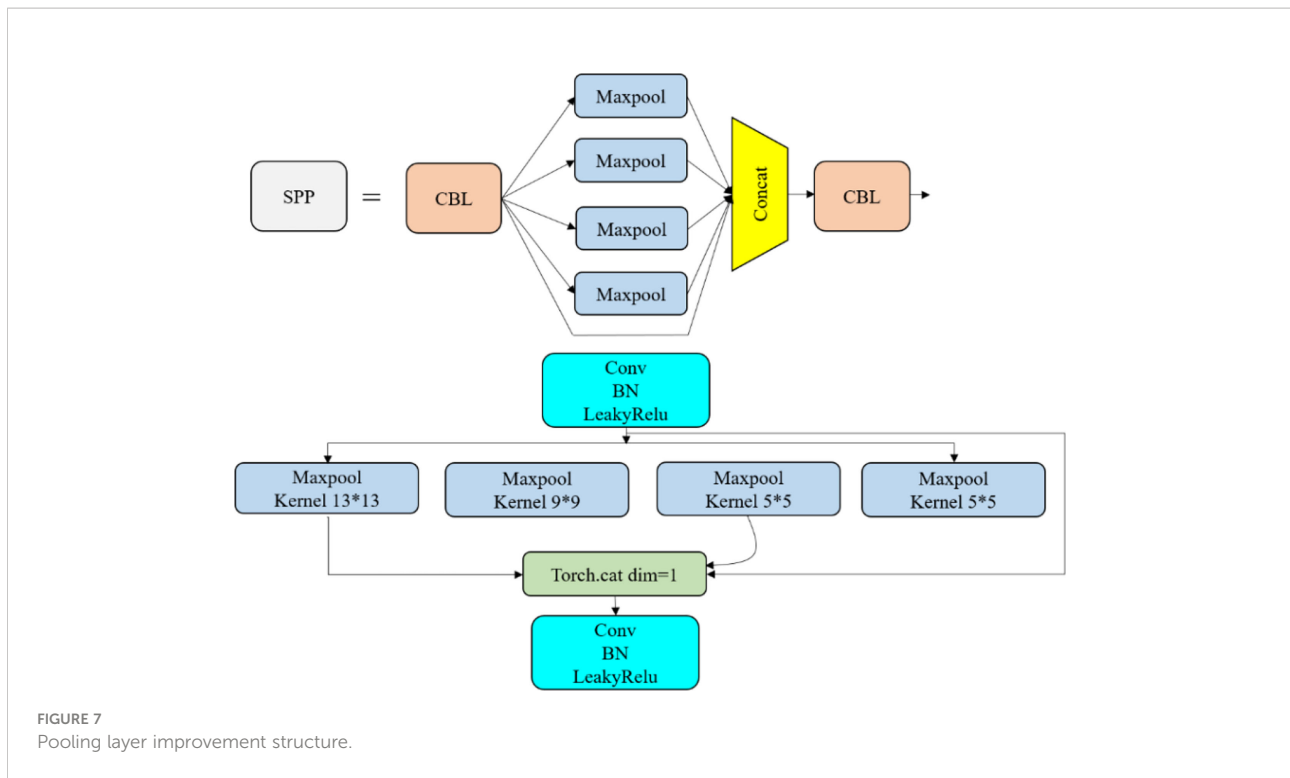


FIGURE 6
Improved FPN + PAN structure.

Figure 7 includes a graphic representation of the contribution of a maximum pooling layer. Figure 7 shows the spatial pyramid pooling module SPP and the combination module CBL, which combines convolutional layers, BN, and activation function layers. The addition of a 3*3 maximum pooling filter has increased the model's receptive field.

3 Results and discussion

This section has a detailed description of the SSDD, AirSARship datasets, experimental settings as well as evaluation metrics and assesses the performance of the technique. The testing set is then separated into two sets, one



is offshore ships and the other one is inshore ships, and each group has been used to assess the efficacy of the various strategies. The identification outcomes of the current model and a few unique CNN-based models are shown on the two SAR large scene images.

3.1 Dataset introduction

SSDD dataset: The first and most important stage in ship detection using deep learning techniques is the construction of a sizable and representative dataset. Therefore, the experiment utilize the SSDD (Li et al., 2017a) dataset, which have 1160 SAR images with resolutions ranging from 1m to 10m with polarizations in HH, HV, VV, and VH from RadarSat-2, TerraSAR-X, and Sentinel-1. Each sample image has a dimension of roughly 800 x 800, with a ratio of 7:1:2, where the SSDD dataset is divided into three sets for the experiment: a training set, a validation set, and a testing set (Table 1).

AirSARship Dataset: In the present study, experiments also use the AirSARShip-1.0 (Xian et al., 2019) dataset to assess the performance of proposed model utilize high-resolution SAR ship identification dataset. Gaofen-3 provided 31 single-polarized SAR images for AirSARShip-1.0. Most images have a size of 3000 x 3000 pixels, while one has a size of 4140 x 4140 pixels with resolutions ranging from 1 to 3 meters with HH polarization. The large scene image has been split into 1000 x 1000 slices with a ratio of 7:1:2, where the dataset is divided into three sets for the experiment: a training set, a validation set, and a testing set (Table 1).

Two SAR large scene images from the Chinese GF-3 satellite, as shown in Figure 13, further illustrate the efficacy of the suggested strategy for identifying different size ships in SAR large scene images with complicated sceneries. These images contain inshore and offshore scenery as well as ship targets at various scales. In Figure 8, some image slices are presented and offshore and inshore scenes as well as multiscale ship targets are primarily shown in Figure 8A. The dataset clearly shows that

TABLE 1 Reveal the dataset’s complete details.

| Dataset | Size (pixel) | Image (num) | Mode | Satellite | Resolution(m) |
|------------|----------------------|-------------|--------|--|---------------|
| SSDD | 390 ×205 500 ×500 | 1160 | SL | RadarSat-2 TerraSAR-X Sentinel-1 | 1-15 |
| AirSARship | 3000 × 3000 | 31 | SL/UFS | GF-3 | 1/3 |

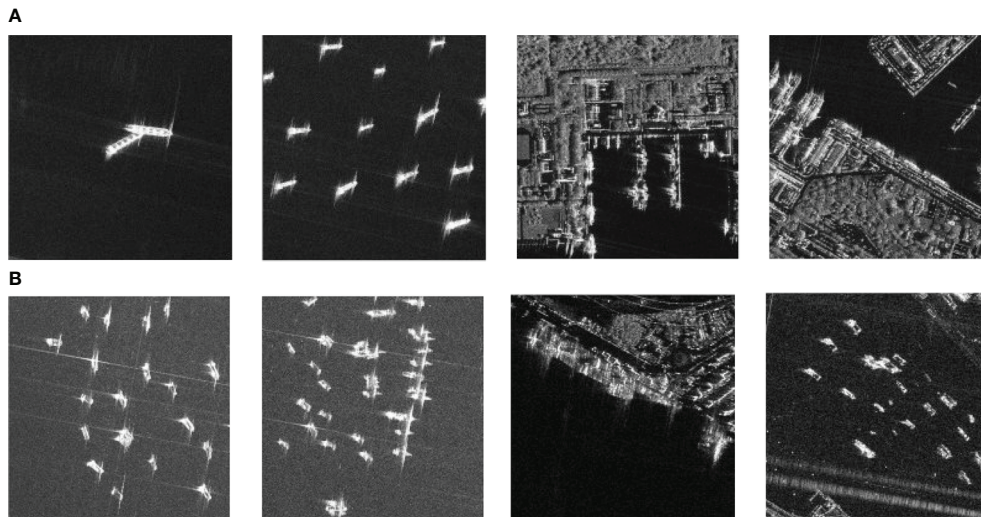


FIGURE 8
Inshore, offshore and different scale ship target on SAR images, (A) ships from the first SAR large scene images presented in Figure 13A, and (B) ship from the second SAR large images presented in Figure 13B.

both off-shore and inshore scenarios are included, and that the sizes of the ships fluctuate widely.

3.2 Experimental environment

The experiments are all carried out using PyTorch 1.7.0, CUDA 10.1, and CUDNN 7.6.5 on an NVIDIA Geforce GTX 2080Ti GPU and an Intel Core i9-9900KF CPU. The PC, which was equipped with a deep learning environment for our research and was running Windows 10, is depicted in Table 2. Additionally, each model was trained over 100000 iterations utilizing the Stochastic Gradient Descent (SGD) technique on a total of two images per minibatch. The initial learning rate was set at 0.001, while the weight decay was set to 0.00004. In every trial, the detection threshold IOU was set to 0.7. Table 2 displayed the experimental hardware and software configuration. During the experiment, the same platform was used for all comparison techniques.

3.3 Evaluation metrics

Since optical and SAR image object detection tasks are comparable, the effectiveness of various approaches is assessed using a variety of established indicators, such as average precision (AP), recall rate (r), precision rate (p), F score (F1), and these indications are specifically formulated in following equations (3-6):

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

The number of correctly recognized ships, false alarms, and missing ships are denoted by the acronyms FN (false negative), FP (false positives), and TP (true positives). The precision and recall are combined into the F1 score as follows:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

$$AP = \int_0^1 P(R) dR \quad (6)$$

The complete detection effectiveness of the various models is assessed using the AP and F1-score metrics, and a higher number indicates a superior detector performance.

The percentage of ground truth ships that networks correctly predict in all predictions is referred to as the precision rate. The percentage of ground truth ships that networks correctly predicted in

TABLE 2 Environment Configuration.

| Project | Model/Parameter |
|-----------|---------------------------------|
| System | windows 10 |
| RAM | 32GB |
| GPU | NVIDIA GTX Geforce 2080 Ti |
| CPU | Intel i9-9900KF |
| Framework | CUDA10.1/cudnn7.6.5/torch 1.7.0 |
| Code | python3.7 |

all ground truth ships is referred to as recall rate. F1 is a comprehensive statistic that combines precision rate and recall rate to assess the effectiveness of various framework. AP outlines the region beneath Precision-Recall (PR) curves and also shows the overall effectiveness of various approaches. Additionally, Frames-Per-Second (FPS), which is derived from Equation (7), is used to assess the detection speed of various approaches. A method achieves a higher speed the higher the FPS.

$$FPS = \frac{1}{T_{per-imag}} \quad (7)$$

When processing an image, the inference time ($T_{per-imag}$) is the cost of a method.

3.4 Detection performance of inshore and offshore ships

In this section, the proposed approach and alternative CNN-based approaches, such as Faster-RCNN (Lin et al., 2017), SSD (Liu et al., 2016), R2CNN (Jiang et al., 2017), ARPN (Zhao et al., 2020), DAPN (Cui et al., 2019), Quad-FPN (Zhang et al., 2021b), HR-SDNet (Wei et al., 2020), Grid R-CNN (Lu et al., 2019), Cascade R-CNN (Cai and Vasconcelos, 2018), YOLOv4-LITE (Liu et al., 2022), EfficientDet (Tan et al., 2020), Free-Anchor (Zhang et al., 2019), Lite-Yolov5 (Xu et al., 2022), and yolov5-X, are assessed using offshore and inshore ships of testing sets. Aside from these metrics, F1, AP, and FPS are also employed to investigate the applicability of various methodologies. The suggested model's identification performance against other CNN-based approaches tested on offshore ships and inshore

ships based on SSDD dataset and AirSAR ship dataset is presented in Tables 3, 4; Figures 9, 10. The current model provides the best accuracy for offshore SSDD (about 95.36% AP for the offshore scenes). The second-best result is 89.03% from the R2CNN approach, although it still performs better at detecting anomalies than the currently suggested model by 6.33% AP. The studied model also delivers the best accuracy for inshore on SSDD (about 92.27% AP for the inshore scenes). The second-best result is 83.53% from the R2CNN approach, although it still performs better in terms of detection than the currently suggested model by 8.74% AP. The researched model provides the best accuracy for offshore on AirSARship (about 94.57% AP for the offshore scenes). It has the best detection performance, as seen by the second-best result of 88.27% from the Quad-FPN approach, which is still 6.3% AP less than the proposed model. The proposed model also delivers the best accuracy for inshore on AirSARship (about 91.11% AP for the inshore scenes). It also has the best detection performance, with the second-best estimate coming from the Lite-Yolov5 approach at 84.94%, however it is still 6.17% AP lower than the current model.

The suggested model and the other state-of-the-art CNN based approaches include Faster-RCNN (Lin et al., 2017), SSD (Liu et al., 2016), R2CNN (Jiang et al., 2017), ARPN (Zhao et al., 2020), DAPN (Cui et al., 2019), Quad-FPN (Zhang et al., 2021b), HR-SDNet (Wei et al., 2020), Grid R-CNN (Lu et al., 2019), Cascade R-CNN (Cai and Vasconcelos, 2018), YOLOv4-LITE (Liu et al., 2022), EfficientDet (Tan et al., 2020), Free-Anchor (Zhang et al., 2019), Lite-Yolov5 (Xu et al., 2022), and yolov5-X techniques all have detection accuracies that are higher for offshore scenes than for inshore situations. This is reasonable

TABLE 3 The identification outcomes of various state-of-the-art CNN based approaches on offshore and inshore ship scene for SSDD Dataset.

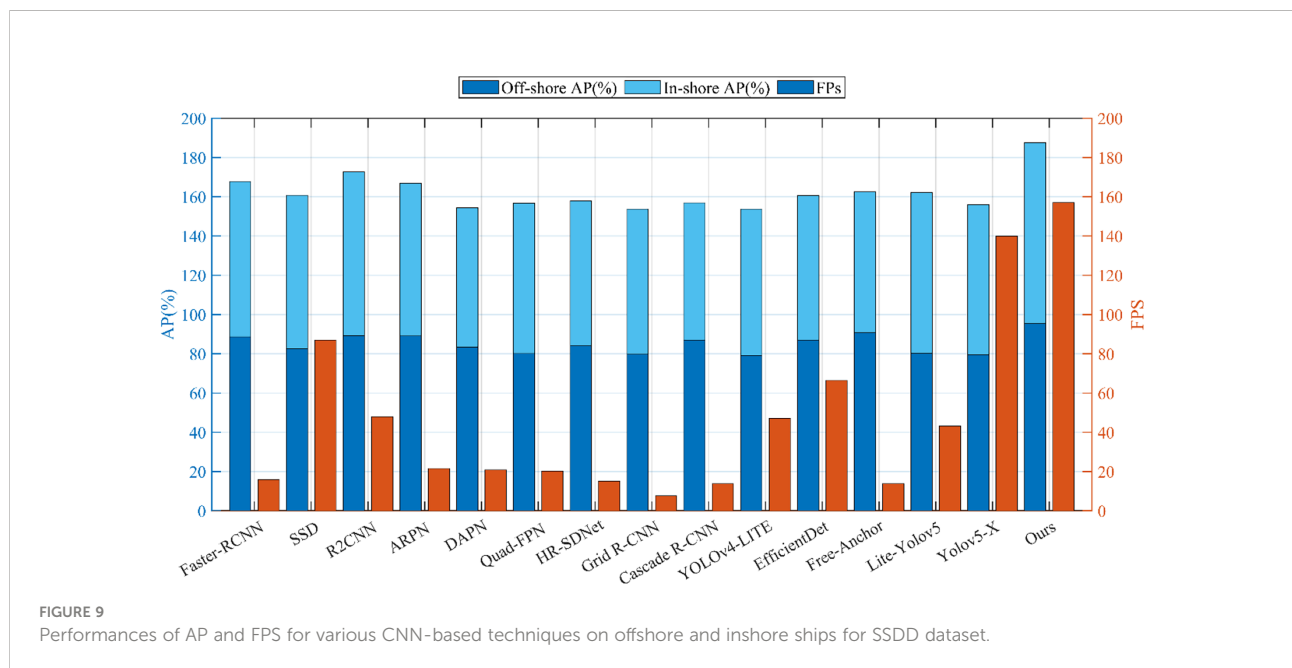
| Methods | Off-Shore | | | | In-Shore | | | | FPS |
|---|-----------|-------|--------------|--------|----------|-------|--------------|--------|------------|
| | P (%) | R (%) | AP (%) | F1 | P (%) | R (%) | AP (%) | F1 | |
| Faster-RCNN (Lin et al., 2017) | 88.60 | 91.81 | 88.60 | 0.8770 | 77.21 | 79.37 | 74.20 | 0.7377 | 16 |
| SSD (Liu et al., 2016) | 85.59 | 85.65 | 82.50 | 0.8170 | 74.67 | 71.45 | 78.25 | 0.7735 | 87 |
| R2CNN (Jiang et al., 2017) | 91.14 | 90.07 | <u>89.03</u> | 0.8853 | 78.99 | 70.88 | <u>83.53</u> | 0.8289 | 48 |
| ARPN (Zhao et al., 2020) | 91.64 | 90.38 | 88.10 | 0.8797 | 79.63 | 77.70 | 77.7 | 0.7690 | 21.55 |
| DAPN (Cui et al., 2019) | 91.89 | 89.89 | 83.37 | 0.8271 | 75.76 | 78.77 | 70.98 | 0.6950 | 20.81 |
| Quad-FPN (Zhang et al., 2021b) | 85.97 | 89.56 | 80.10 | 0.7989 | 52.90 | 77.78 | 76.59 | 0.7551 | 20.25 |
| HR-SDNet (Wei et al., 2020) | 86.47 | 88.28 | 84.03 | 0.8350 | 72.57 | 77.97 | 73.85 | 0.7225 | 15.16 |
| Grid R-CNN (Lu et al., 2019) | 89.10 | 91.18 | 79.73 | 0.7853 | 81.08 | 81.56 | 74.01 | 0.7332 | 7.55 |
| Cascade R-CNN (Cai and Vasconcelos, 2018) | 90.10 | 84.75 | 87.01 | 0.8623 | 69.78 | 79.03 | 69.89 | 0.6814 | 10.55 |
| YOLOv4-LITE (Liu et al., 2022) | 89.59 | 87.03 | 79.17 | 0.7835 | 54.80 | 85.70 | 74.62 | 0.7323 | 42.5 |
| EfficientDet (Tan et al., 2020) | 92.19 | 89.15 | 86.11 | 0.8517 | 79.01 | 77.64 | 73.63 | 0.7292 | 21.44 |
| Free-Anchor (Zhang et al., 2019) | 90.31 | 87.35 | 87.67 | 0.8677 | 75.70 | 76.07 | 71.80 | 0.7055 | 13.76 |
| Lite-Yolov5 (Xu et al., 2022) | 91.32 | 91.08 | 80.16 | 0.7901 | 75.78 | 80.62 | 81.92 | 0.8045 | 43.25 |
| Yolov5-X | 77.25 | 82.60 | 79.56 | 0.7862 | 70.89 | 81.61 | 76.49 | 0.7578 | 140 |
| Our model | 94.57 | 94.06 | 95.36 | 0.9413 | 89.44 | 90.77 | 92.27 | 0.9140 | 157 |

TABLE 4 The identification outcomes of various state-of-the-art CNN based approaches on offshore and inshore ship scene for AirSARship Dataset.

| Methods | Off-Shore | | | | In-Shore | | | | FPS |
|---|-----------|-------|--------|--------|----------|-------|--------|--------|-------|
| | P (%) | R (%) | AP (%) | F1 | P (%) | R (%) | AP (%) | F1 | |
| Faster-RCNN (Lin et al., 2017) | 83.72 | 89.89 | 84.12 | 0.8370 | 69.58 | 86.01 | 79.21 | 0.7877 | 16 |
| SSD (Liu et al., 2016) | 92.09 | 86.66 | 86.54 | 0.8550 | 64.76 | 88.08 | 81.41 | 0.8035 | 87 |
| R2CNN (Jiang et al., 2017) | 84.08 | 94.07 | 85.11 | 0.8653 | 72.15 | 90.92 | 79.22 | 0.6989 | 48 |
| ARPN (Zhao et al., 2020) | 82.40 | 80.83 | 82.22 | 0.8161 | 71.74 | 69.13 | 68.10 | 0.6754 | 21.55 |
| DAPN (Cui et al., 2019) | 90.18 | 87.98 | 85.83 | 0.8454 | 69.50 | 66.57 | 72.98 | 0.7141 | 20.81 |
| Quad-FPN (Zhang et al., 2021b) | 89.70 | 90.56 | 88.27 | 0.8789 | 62.10 | 78.15 | 80.77 | 0.7945 | 20.25 |
| HR-SDNet (Wei et al., 2020) | 83.40 | 89.82 | 79.27 | 0.7818 | 70.50 | 78.71 | 76.05 | 0.7506 | 15.16 |
| Grid R-CNN (Lu et al., 2019) | 89.89 | 89.91 | 81.03 | 0.8010 | 76.81 | 71.60 | 69.55 | 0.6854 | 7.55 |
| Cascade R-CNN (Cai and Vasconcelos, 2018) | 88.75 | 74.57 | 79.26 | 0.7854 | 70.87 | 81.03 | 70.91 | 0.6998 | 10.55 |
| YOLOv4-LITE (Liu et al., 2022) | 87.01 | 89.55 | 79.07 | 0.7889 | 68.60 | 84.09 | 82.02 | 0.8189 | 42.5 |
| EfficientDet (Tan et al., 2020) | 89.10 | 90.75 | 81.20 | 0.8081 | 87.29 | 76.91 | 63.35 | 0.6260 | 21.44 |
| Free-Anchor (Zhang et al., 2019) | 90.38 | 77.35 | 84.60 | 0.8354 | 70.65 | 78.60 | 74.12 | 0.7379 | 13.76 |
| Lite-Yolov5 (Xu et al., 2022) | 91.56 | 90.48 | 81.17 | 0.8054 | 78.90 | 74.82 | 84.94 | 0.8388 | 66.25 |
| Yolov5-X | 79.12 | 82.70 | 79.44 | 0.7822 | 76.35 | 79.69 | 73.46 | 0.7225 | 140 |
| Our model | 92.83 | 90.55 | 94.57 | 0.9313 | 90.14 | 89.08 | 91.11 | 0.9014 | 157 |

considering that the former has a more complicated background than the latter. Perhaps as a result of their poor small ship identification capabilities, the other alternative approaches have lower precision values than the suggested model. In this current research work, the recall values of the proposed model are occasionally less than those of other offered methods. As a result, a suitable score threshold can be further thought about in the future to balance missed detections and false alarms. Additionally, the current model appears to be faster than other

approaches based on the FPS data in Tables 3, 4, and Figures 9, 10, potentially as a result of the separable depth-wise and point-wise convolutions utilized in the backbone network. In conclusion, the offshore scene has greater accuracy, AP, and F1 scores for both datasets than the inshore scenario. This might be due to the inshore scene’s densely packed ships and increased backdrop interference from the land. Additionally, it demonstrates that it is more difficult to spot ships in the inshore scene than it is in the offshore environment.

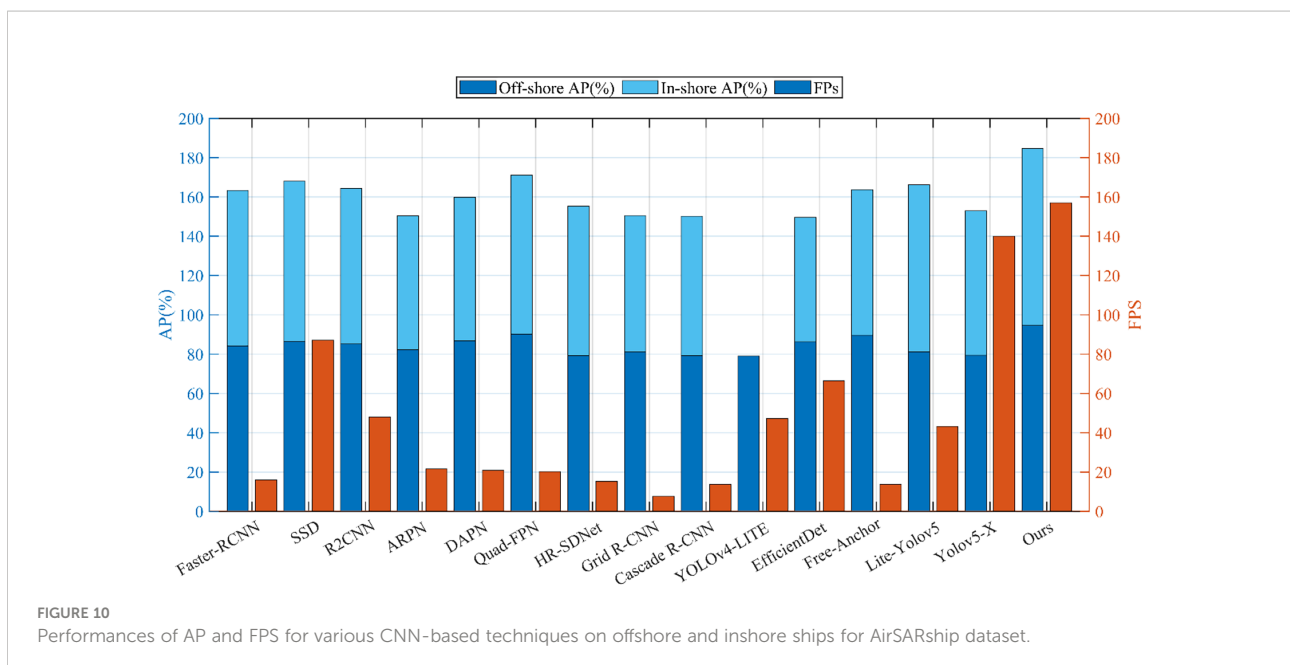


The proposed model’s performance in terms of detection results compared to existing convolutional neural network-based approaches tested on offshore ships and inshore ships using SSDD dataset and AirSARship dataset is shown in Figures 11, 12. The suggested model is capable of detecting different SAR ships with multiscale sizes under varied backgrounds. This demonstrates its great scale and scene adaptation together with excellent detection performance. The currently proposed model can increase the detection confidence scores when compared to the second-best CNN-based ship detector R2CNN. For instance, the suggested model raises the confidence score in Figure 11 first detection sample from 0.96 to 1.0. This can demonstrate the better trustworthiness of the newly proposed model. It is evident that the inshore scenario contains a sizable number of tightly packed ship targets. The other suggested solutions miss certain closely grouped inshore ships. The proposed model, however, is capable of accurately localizing and detecting these multiscale ships with high probabilities.

The other suggested CNN-based techniques in this research can precisely identify the port’s heavily docked ships. However, it can be observed that the suggested method is more accurate and can detect these ships better when comparing the detection outcomes of several proposed CNN-based systems. The detection outcomes of the Faster-RCNN (Lin et al., 2017), SSD (Liu et al., 2016), R2CNN (Jiang et al., 2017), ARPN (Zhao et al., 2020), DAPN (Cui et al., 2019), Quad-FPN (Zhang et al., 2021b), HR-SDNet (Wei et al., 2020), Grid R-CNN (Lu et al., 2019), Cascade R-CNN (Cai and Vasconcelos, 2018), YOLOv4-LITE (Liu et al., 2022), EfficientDet (Tan et al., 2020), Free-Anchor (Zhang et al., 2019), Lite-Yolov5 (Xu et al., 2022), and yolov5-X algorithms contain several false alarms for the offshore and inshore scene (Figures 11, 12). Additionally, there are a few missed ships in the detection findings, which could be a

result of how closely docked the ship targets are, making it more challenging for the framework to discriminate between them. Similarly, it is observed by comparing the suggested model’s detection results that they are more precise than those produced by existing CNN-based techniques. Figures 11, 12 displays the outcomes of various CNN-based object detection techniques in an offshore scenario created for SAR images. It is evident that the offshore landscape contains a substantial number of dense multi-scale ship targets (the first two column of Figures 11, 12).

A pink color circle denotes false alarms in the identification outcomes of other proposed CNN-based models in this study that are not the current model. This might be due to a small number of false alarms that closely resemble ships, creating it great challenging for the network to successfully recognize. Because there are wakes of ships and surroundings, such ship A in Figure 12, the Faster-RCNN (Lin et al., 2017), SSD (Liu et al., 2016), R2CNN (Jiang et al., 2017), ARPN (Zhao et al., 2020), DAPN (Cui et al., 2019), Quad-FPN (Zhang et al., 2021b), HR-SDNet (Wei et al., 2020), Grid R-CNN (Lu et al., 2019), Cascade R-CNN (Cai and Vasconcelos, 2018), YOLOv4-LITE (Liu et al., 2022), EfficientDet (Tan et al., 2020), Free-Anchor (Zhang et al., 2019), Lite-Yolov5 (Xu et al., 2022), and yolov5-X algorithms can distinguish between wakes of ships and their surroundings. Additionally, several ships are overlooked by the Faster-RCNN (Lin et al., 2017), SSD (Liu et al., 2016), R2CNN (Jiang et al., 2017), ARPN (Zhao et al., 2020), DAPN (Cui et al., 2019), Quad-FPN (Zhang et al., 2021b), HR-SDNet (Wei et al., 2020), Grid R-CNN (Lu et al., 2019), Cascade R-CNN (Cai and Vasconcelos, 2018), YOLOv4-LITE (Liu et al., 2022), EfficientDet (Tan et al., 2020), Free-Anchor (Zhang et al., 2019), Lite-Yolov5 (Xu et al., 2022), and yolov5-X algorithms, as seen by the red circles in Figures 11, 12. The inability to extract distinguishing characteristics of ships and interference may be to blame. However, the suggested



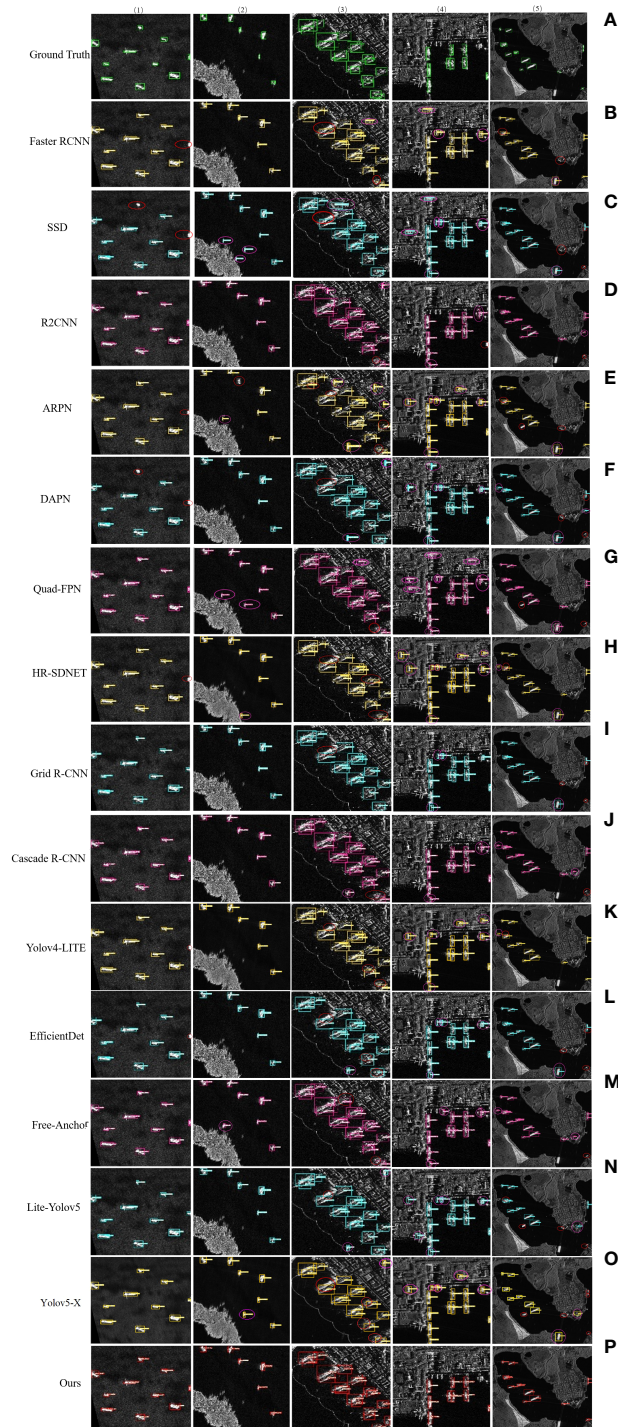


FIGURE 11

The visual detection outcomes of CNN-based approaches for offshore and inshore ships based on SSDD Dataset. Results from (A) is ground truth, results from (B) is Faster-R-CNN method, results from (C) SSD method, results from (D) is R2CNN method, results from (E) is ARPN, results from (F) is DAPN, results from (G) is Quad-FPN, results from (H) is HR-SDNet, results from (I) is Grid R-CNN, results from (J) is Cascade R-CNN, results from (K) is YOLOv4-LITE, results from (L) is EfficientDet, results from (M) is Free-Anchor, results from (N) is Lite-Yolov5, results from (O) is yolov5-X, and results from (P) is our proposed method. Note the pink circle show the false detection of ship and red circle is show the missing ship.

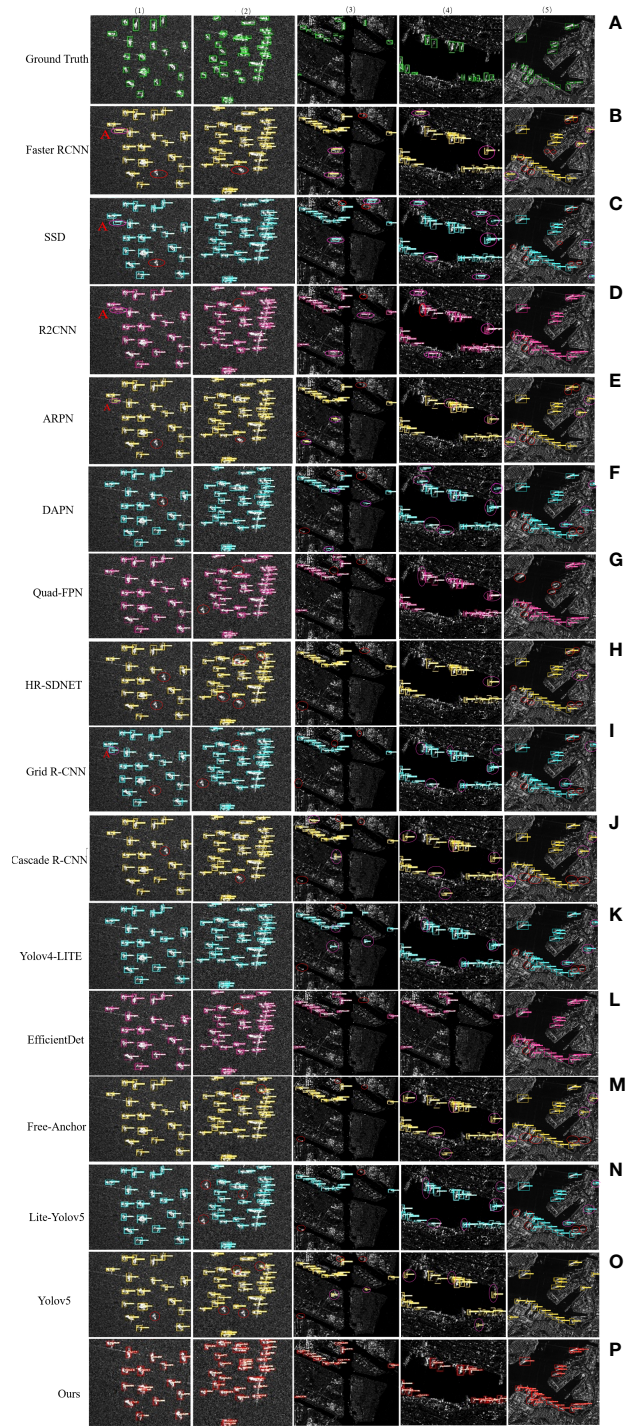


FIGURE 12

The visual detection outcomes of CNN-based approaches for offshore and inshore ships based on AirSARship Dataset. Results from (A) is ground truth, results from (B) is Faster-R-CNN method, results from (C) SSD method, results from (D) is R2CNN method, results from (E) is ARPN, results from (F) is DAPN, results from (G) is Quad-FPN, results from (H) is HR-SDNet, results from (I) is Grid R-CNN, results from (J) is Cascade R-CNN, results from (K) is YOLOv4-LITE, results from (L) is EfficientDet, results from (M) is Free-Anchor, results from (N) is Lite-Yolov5, results from (O) is yolov5-X, and results from (P) is our proposed method. Note the pink circle show the false detection of ship and red circle is show the missing ship.

model could identify these ships without any false alarms, which is also one of the driving forces for this paper and highlights the current proposed model's powerful and robust feature representation skills.

The suggested model is compared with other state-of-the-art CNN based methods SAR ship identification techniques. As can be observed, two-stage or multi-stage approaches typically execute detection tasks more effectively than single-stage methods. However, compared to these two stage or multistage detection methods, one stage detection methods clearly have a faster inference efficiency. This might be as a result of the two-stage detection network's sophisticated network architecture and increased computational load. Some one-stage detection techniques that are more effective at detecting ships have recently been proposed, including YOLOv4-LITE (Liu et al., 2022), EfficientDet (Tan et al., 2020), Free Anchor (Zhang et al., 2019), Lite-Yolov5 (Xu et al., 2022), and two-stage detectors like ARPN (Zhao et al., 2020), DAPN (Cui et al., 2019), Quad-FPN (Zhang et al., 2021b), HR-SDNet (Wei et al., 2020), Grid R-CNN (Lu et al., 2019), and Cascade R-CNN (Cai and Vasconcelos, 2018) are the performance comparisons of the current model with other cutting-edge detectors.

The detection accuracies of inshore scenes for the proposed model and the other CNN-based techniques are also all lower than those of offshore scenes. De-formable convolution can lessen the interference of complicated backgrounds, particularly for inshore sceneries, hence the recently presented technique seems to be robust to background interferences. The other state-of-the-art techniques are less precise and have lower recall values than the model now under study because of their poor small ship recognition capabilities. As a result, it will be possible to consider an acceptable score threshold in the future to balance missed

detections and false alarms. Additionally, accuracy needs to be further improved, for example, when striking military targets with precision. It might be suggested in the future to choose between speed and accuracy. The suggested model has a higher detection effectiveness. This might be the case because other methods overlook smaller targets since they do not consider the underlying data in the prediction layer. There are some false alarms in the identification outcomes of previous techniques for the complicated inshore scenarios. Particularly, several land features in the inshore scene are wrongly identified as targets by Faster-RCNN, SSD, R2CNN, ARPN, DAPN, Quad-FPN, HR-SDNet, Grid R-CNN, Cascade R-CNN, YOLOv4-LITE, EfficientDet, Free-Anchor, Lite-Yolov5. By doing several experiments using the SSDD and AirSARship datasets, we illustrate the effectiveness of our suggested model. The SSDD dataset ablation studies of FPN +PAN and attention mechanism modules have shown that each of them can enhance ship detection performance, and the combination of both can increase detection outcomes.

3.5 Detection performance on SAR large scene images

This section compares various CNN-based algorithms, such as Yolov5-X, Faster-RCNN (Lin et al., 2017), SSD (Liu et al., 2016), and R2CNN (Jiang et al., 2017), as well as current methods for object recognition in SAR images using large-scale scene SAR images. To validate the good migration ability of the suggested model, the actual ship identification in two more SAR large scene images has been carried out. Figure 13 displays the areas covered by the two SAR large scene images

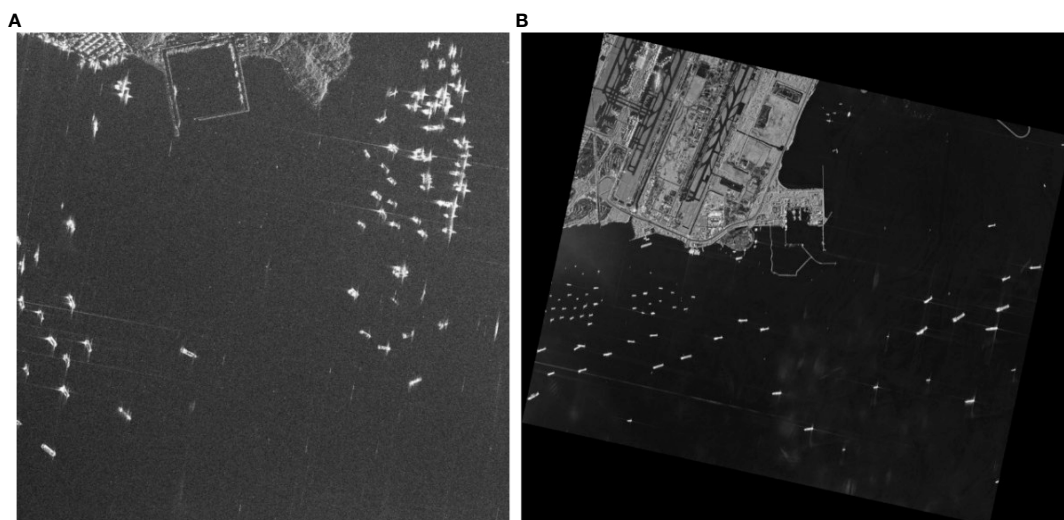


FIGURE 13

In this current research work, the two SAR large scenes images acquired from Chinese GF-3 satellite are utilized for ship detection. (A) AirSARship resolution is 1/3m, and (B) GF 3 satellite resolution is 1m.

acquired by the Chinese satellite GF3. These two SAR images were selected because they both lie along globally important routes named Malacca strait (Table 5). The VV polarization SAR images from Table 5 are considered because ships often exhibit higher backscattering values in VV polarization (Torres et al., 2012). Due to the restricted GPU memory, they are divided into 800 x 800 (Zhang et al., 2020a) and followed by (Xian et al., 2019) 1000 x 1000 of small sub images before being used for training and testing. SAR ships are ultimately added to the suggested model in order to actually detect them. The outcomes of the sub- image’s detection are then added to the original SAR large scene image (Figure 14). The detection accuracy and speed of various approaches are assessed, respectively, using the AP and FPS (Table 6).

The two SAR large scene images are used to depict the results of SAR ship detection using the current model and other CNN-based methods. The current model is able to successfully detect the majority of ships, demonstrating its strong migration application capabilities in ocean surveillance (Figures 15A, B). Features and a clear environment come first. In particular, just a few ships in the big panorama were missed by all of these convolutional neural network -based approaches, which are indicated by pink circles in Figures 15A, B. The identification

outcomes on the second SAR large scene image are displayed, and the identification results of various CNN-based approaches on the SAR large scene images are demonstrated on the left side of Figures 15A, B. Figures 15A, B’s right side enlarges and displays two particular regions designated with brown and blue rectangles. The comparison of the detection outcomes of various model performances on SAR large scene images is presented in Figures 15A, B. Offshore ships make up the majority of this SAR large-scene image. Inshore scenes have substantial clutter, which could cause false alarms. The results of another CNN-based approach suggested in this study show that there are few false alarms and missed targets in the offshore scenes, and the false alarms are repressed in the inshore scenes as well. However, the proposed approach does not have false alarms or miss target detection, which is one of the motivating factors for this paper. Figures 15A, B illustrates the visualization of the SAR ship identification performance of the suggested model on two SAR large scene images acquired from the Chinese GF3 satellite and employed in the current study. The suggested model can successfully detect the majority of ships, which demonstrates its good migration application capability in ocean surveillance, it can be deduced from Figures 15A, B. In conclusion, the identification outcomes on the two SAR large scene images at

TABLE 5 The details descriptions of two SAR large scene images.

| Images | Mode | Satellite | Resolution (m) | Image Size | Format |
|------------|--------|-----------|----------------|---------------|--------|
| AirSARship | SL/UFS | GF-3 | 1/3 | 3000 × 3000 | Tiff |
| GF3 | SL | GF-3 | 1 | 15350 × 13592 | Tiff |

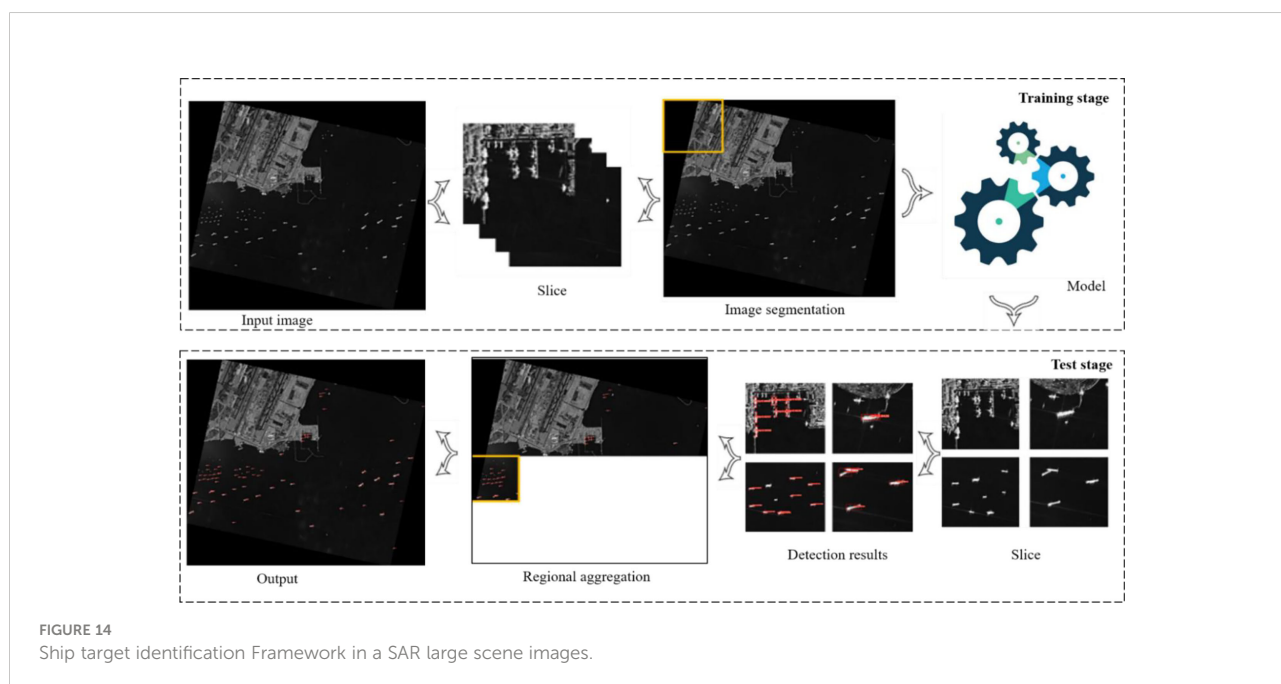


TABLE 6 Detection outcomes of various CNN-based approaches on two SAR large scene ship images.

| Method | Image-1 | | Image-2 | |
|--------------------------------|---------|----------|---------|----------|
| | AP | Time (s) | AP | Time (s) |
| Faster-RCNN (Lin et al., 2017) | 0.838 | 37s | 0.845 | 17s |
| SSD (Liu et al., 2016) | 0.790 | 14s | 0.785 | 25s |
| R2CNN (Jiang et al., 2017) | 0.879 | 65s | 0.830 | 39s |
| Yolov5-X | 0.765 | 140s | 0.749 | 140s |
| Ours | 0.956 | 157s | 0.948 | 156s |

various resolutions show that the present system identifies multi size ships with competitive outcomes and has a strong generalization capacity in comparison to the various CNN-based approaches created for identification object in SAR images. It demonstrates that the approach now under study can adapt to SAR images from various sources more effectively.

4 Ablation study

The ablation studies presented in this section, used to demonstrate the suggested FPN+PAN and attention mechanism module's effectiveness through removal and installation to better understand the behavior of the framework.

4.1 Ablation study on the FPN+PAN module

Table 7 illustrates the ablation research of YOLOv5 removal and installation of the FPN+PAN module. Table 7, “✗” denotes YOLOv5 without the FPN+PAN module, while “✓” denotes YOLOv5 with the FPN+PAN module (i.e., our suggested model). Experiments were carried out in offshore and inshore scenes respectively, to evaluate the identification achievement of the suggested approach for offshore and inshore scenes. In Table 7, the identification effectiveness of the model is approximately similar, with little modification in identification indicator, as a result of the relatively simple background of offshore ship and little interference, due to the identification achievement has been outstanding in the simple background, and only limited enhancement can be obtained. In contrast to offshore ships, inshore ships have a more complicated backdrop clutter. Moreover, wharfs as well as other structures on the shore significantly undermine detections, and SAR ship detection effectiveness generally declines. However, the detection performance is significantly enhanced by improving the feature representation capacity and minimizing the aliasing effect of fusion features. In Table 7, by installation of FPN +PAN modules, the

result of p, r and AP rate of the model are enhanced by about 2.9%, 4.6% and, 5.5% respectively for offshore ship detection, while for inshore ship detection the p, r and AP rate of the model are enhanced by about 6.08%, 15% and, 8.07% respectively.

4.2 Ablation study on the attention mechanism module

Table 8 presented the ablation research of YOLOv5 removal and installation of the attention mechanism module. Table 8, “✗” denotes YOLOv5 without the attention mechanism module, “✓” denotes YOLOv5 with the attention mechanism module (i.e., our suggested model). The installation of attention mechanism module to our model, the detection performance is significantly enhanced by improving the feature representation capacity and minimizing the aliasing effect of fusion features. In Table 8, through installation of attention mechanism modules, as a result the p, r and AP rate of the model are enhanced by about 7.22%, 1.36% and, 5.31% respectively for offshore ship detection, while for inshore ship detection the p, r and AP rate of the model are enhanced by about 1.03%, 10.97% and, 10.67% respectively.

5 Conclusions and future work

In this modern technological era, the advanced machine learning and artificial intelligence-based models have revolutionized diverse research domains with full spectrum. Due to its automatic feature extraction and strong identification skills, it can be used in a variety of study fields. An improved version of the unique one stage YOLOv5 for SAR ship identification has been proposed in this study work, drawing inspiration from the capabilities of these models in other research domains. The generic YOLOv5 model has been improved to address the major issues with the SAR ship detection process. These issues include complexity (complex background interferences, various size ship feature differences, and indistinct tiny ship characteristics), high-cost effectiveness, poor identification and

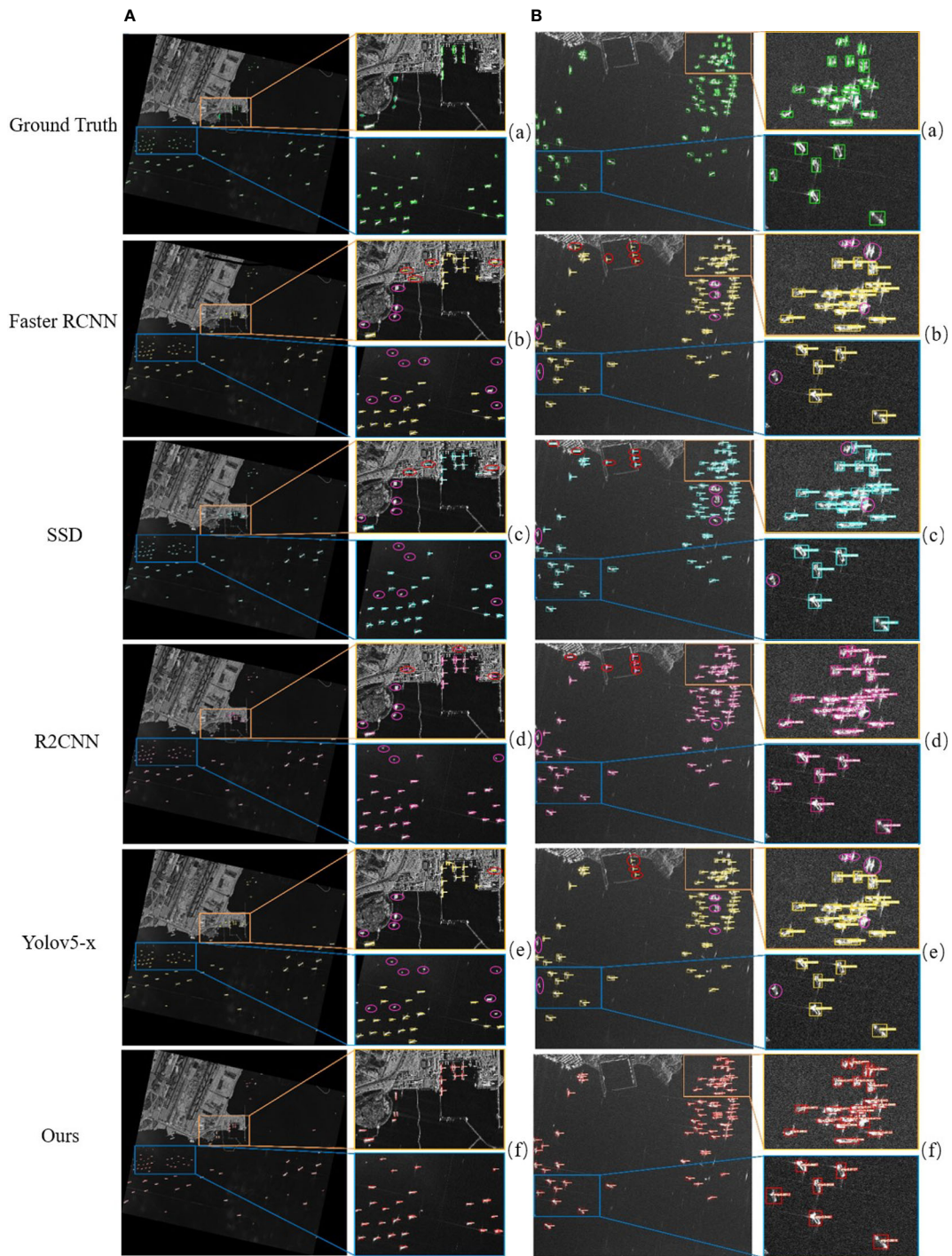


FIGURE 15

The identification outcomes of various CNN based approaches, **(A)** first large-scene SAR image based on SSDD Dataset, **(B)** second SAR large scene image based on AirSARship Dataset. Rectangles with green colors correspond to the ground truth ships while yellow, sky color, pink and red colors are referred to predictions ship, respectively. The red circle shows the false detection of ship and pink circle is showing the missing ship. The right side of an image displays two enlarged special areas that are marked by blue and brown rectangles respectively.

TABLE 7 The ablation study of our proposed model removal and installation the FPN+PAN module on the SSDD dataset.

| FPN+PAN | OFF-shore | | | In-shore | | |
|---------|-----------|-------|--------------|----------|-------|--------------|
| | p (%) | r (%) | AP(%) | p (%) | r (%) | AP(%) |
| ✗ | 91.67 | 89.43 | 89.78 | 83.36 | 75.56 | 84.20 |
| ✓ | 94.57 | 94.06 | 95.36 | 89.44 | 90.77 | 92.27 |

TABLE 8 The ablation study of our proposed model removal and installation the attention mechanism module on the SSDD dataset.

| Attention Mechanism | OFF-shore | | | In-shore | | |
|---------------------|-----------|-------|--------------|----------|-------|--------------|
| | p (%) | r (%) | AP(%) | p (%) | r (%) | AP(%) |
| ✗ | 87.35 | 92.70 | 90.05 | 88.41 | 79.80 | 81.60 |
| ✓ | 94.57 | 94.06 | 95.36 | 89.44 | 90.77 | 92.27 |

recognition rates, and implementation complexities. The changes to the generic YOLOv5 model in the neck region and backbone section employing C3 and PAN structure have been designed to address these major issues. The SSDD and AirSARship open SAR ship datasets, as well as two SAR large scene images acquired from the GF-3 Chinese satellite, are utilized to obtain the experimental results. After producing testing findings, it has been determined that the enhancement to the generic YOLOv5 model not only enhanced identification capabilities but also demonstrated that this model is not data-hungry (to provide optimum results even for a small amount of dataset). The applicability of this model is assessed using a variety of validation metrics, including accuracy, different training and test sets, and TF values, as well as comparisons with other cutting-edge classification models (ARPN, DAPN, Quad-FPN, HR-SDNet, Grid R-CNN, Cascade R-CNN, Multi-Stage YOLOv4-LITE, EfficientDet, Free-Anchor, Lite-Yolov5). Based on the performance values, it has been determined that the examined model exceeded the benchmark models targeted in this research work by producing high identification rates. Additionally, these high identification rates show how useful the suggested approach is for maritime surveillance. Recommended and forthcoming future work includes the following:

- To enhance the effectiveness of our model detection in the future, we will consider the challenges in SAR data, such as the azimuth ambiguity, sidelobes, and the sea condition.
- In the future, we will investigate optimizing the detection speed of our model.
- We might suggest merging contemporary deep CNN abstract features with conventional concrete ones to further improve detection accuracy.
- In order to further boost the identification speed and accuracy, we will focus on merging the backscattering

characteristics of ships in SAR images with convolutional network architecture and offering a robust constraint, such as a mask.

- Future research on instance segmentation and ship detection will be taken into consideration.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

Conceptualization, MY, LS. Methodology, MY and SH. software, MY, SH and LS. Validation, WJ, LS and XM. Formal analysis, AC and DW. Investigation, XM. Resources, WJ. Data curation, MY, KD and DW. Writing—original draft preparation, MY. Writing—review and editing, MY, XM and LS. Visualization, LS, KD. Supervision, WJ and LS. Project administration, LS. Funding acquisition, WJ and LS. All authors contributed to the article and approved the submitted version.

Funding

Funding National Key Research and Development Program of China (2017YFC1405600).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Baselice, F., and Ferraioli, G. (2013). Unsupervised coastal line extraction from SAR images. *IEEE Geosci. Remote Sens. Lett.* 10, 1350–1354. doi: 10.1109/LGRS.2013.2241013
- Cai, Z., and Vasconcelos, N. (2018). "Cascade r-cnn: Delving into high quality object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 6154–6162. doi: 10.48550/arXiv.1712.00726
- Caputo, S., Castellano, G., Greco, F., Mencar, C., Petti, N., and Vessio, G. "Human detection in drone images using YOLO for search-and-Rescue operations," in *International conference of the Italian association for artificial intelligence* (Springer), 326–337. doi: 10.1007/978-3-031-08421-8_22
- Chang, Y.-L., Anagaw, A., Chang, L., Wang, Y. C., Hsiao, C.-Y., and Lee, W.-H. (2019). Ship detection based on YOLOv2 for SAR imagery. *Remote Sens.* 11, 786. doi: 10.3390/rs11070786
- Chen, G., Li, G., Liu, Y., Zhang, X.-P., and Zhang, L. (2019). SAR image despeckling based on combination of fractional-order total variation and nonlocal low rank regularization. *IEEE Trans. Geosci. Remote Sens.* 58, 2056–2070. doi: 10.1109/TGRS.2019.2952662
- Chen, H., and Shi, Z. (2020). A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sens.* 12, 1662. doi: 10.3390/rs12101662
- Cui, Z., Li, Q., Cao, Z., and Liu, N. (2019). Dense attention pyramid networks for multi-scale ship detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* 57, 8983–8997. doi: 10.1109/TGRS.2019.2923988
- Du, L., Dai, H., Wang, Y., Xie, W., and Wang, Z. (2019). Target discrimination based on weakly supervised learning for high-resolution SAR images in complex scenes. *IEEE Trans. Geosci. Remote Sens.* 58, 461–472. doi: 10.1109/TGRS.2019.2937175
- Elfwing, S., Uchibe, E., and Doya, K. (2018). Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks* 107, 3–11. doi: 10.48550/arXiv.1702.03118
- Gao, Y., Gao, F., Dong, J., and Wang, S. (2019). Change detection from synthetic aperture radar images based on channel weighting-based deep cascade network. *IEEE J. selected topics Appl. Earth observations Remote Sens.* 12, 4517–4529. doi: 10.1109/JSTARS.2019.2953128
- Girshick, R. (2015). "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*. 1440–1448. doi: 10.1109/ICCV.2015.169
- Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 580–587. doi: 10.48550/arXiv.1311.2524
- Han, J., Li, G., and Zhang, X.-P. (2019). Refocusing of moving targets based on low-bit quantized SAR data via parametric quantized iterative hard thresholding. *IEEE Trans. Aerospace Electronic Syst.* 56, 2198–2211. doi: 10.1109/TAES.2019.2944707
- Jiang, Y., Zhu, X., Wang, X., Yang, S., Li, W., Wang, H., et al. (2017). R2CNN: Rotational region CNN for orientation robust scene text detection. doi: 10.48550/arXiv.1706.09579. arXiv preprint arXiv:1706.09579.
- Li, X., Chen, P., and Fan, K. (2020). "Overview of deep convolutional neural network approaches for satellite remote sensing ship monitoring technology," in *IOP conference series: Materials science and engineering* (IOP Publishing), 012071. doi: 10.1088/1757-899X/730/1/012071
- Li, T., Liu, Z., Xie, R., and Ran, L. (2017b). An improved superpixel-level CFAR detection method for ship targets in high-resolution SAR images. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 11, 184–194. doi: 10.1109/JSTARS.2017.2764506
- Lin, H., Chen, H., Jin, K., Zeng, L., and Yang, J. (2019). Ship detection with superpixel-level Fisher vector in high-resolution SAR images. *IEEE Geosci. Remote Sens. Lett.* 17, 247–251. doi: 10.1109/LGRS.2019.2920668
- Lin, T.-Y., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2117–2125. doi: 10.48550/arXiv.1612.03144
- Li, J., Qu, C., and Shao, J. (2017a). "Ship detection in SAR images based on an improved faster r-CNN," in *2017 SAR in big data era: Models, methods and applications (BIGSAR DATA)* (IEEE), 1–6. doi: 10.1109/BIGSAR DATA.2017.8124934
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). "Ssd: Single shot multibox detector," in *European Conference on computer vision* (Springer), 21–37. doi: 10.48550/arXiv.1512.02325
- Liu, S., Kong, W., Chen, X., Xu, M., Yasir, M., Zhao, L., et al. (2022). Multi-scale ship detection algorithm based on a lightweight neural network for spaceborne SAR images. *Remote Sens.* 14, 1149. doi: 10.3390/rs14051149
- Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. (2018). "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8759–8768. doi: 10.48550/arXiv.1803.01534
- Li, J., Xu, C., Su, H., Gao, L., and Wang, T. (2022). Deep learning for SAR ship detection: Past, present and future. *Remote Sens.* 14, 2712. doi: 10.3390/rs14112712
- Lu, X., Li, B., Yue, Y., Li, Q., and Yan, J. (2019). "Grid r-cnn," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7363–7372. doi: 10.48550/arXiv.1811.12030
- Ma, M., Chen, J., Liu, W., and Yang, W. (2018). Ship classification and detection based on CNN using GF-3 SAR images. *Remote Sens.* 10, 2043. doi: 10.3390/rs10122043
- Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., and Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *J. big Data* 2, 1–21. doi: 10.1186/s40537-014-0007-7
- Nepal, U., and Eslamiat, H. (2022). Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs. *Sensors* 22, 464. doi: 10.3390/s22020464
- Niedermeier, A., Romaneessen, E., and Lehner, S. (2000). Detection of coastlines in SAR images using wavelet methods. *IEEE Trans. Geosci. Remote Sens.* 38, 2270–2281. doi: 10.1109/36.868884
- Ouchi, K. (2013). Recent trend and advance of synthetic aperture radar with selected topics. *Remote Sens.* 5, 716–807. doi: 10.3390/rs5020716
- Patel, K., Bhatt, C., and Mazzeo, P. L. (2022). Deep learning-based automatic detection of ships: an experimental study using satellite images. *J. Imaging* 8, 182. doi: 10.3390/jimaging8070182
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788. doi: 10.48550/arXiv.1506.02640
- Redmon, J., and Farhadi, A. (2017). "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7263–7271. doi: 10.48550/arXiv.1612.08242
- Redmon, J., and Farhadi, A. (2018). Yolov3: An incremental improvement. doi: 10.48550/arXiv.1804.02767. arXiv preprint arXiv:1804.02767.
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* 28. doi: 10.48550/arXiv.1506.01497
- Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S. (2019). "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 658–666. doi: 10.48550/arXiv.1902.09630
- Salembier, P., Liesegang, S., and López-Martínez, C. (2018). Ship detection in SAR images based on maxtree representation and graph signal processing. *IEEE Trans. Geosci. Remote Sens.* 57, 2709–2724. doi: 10.1109/TGRS.2018.2876603
- Tan, M., Pang, R., and Le, Q. V. (2020). "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10781–10790. doi: 10.48550/arXiv.1911.09070
- Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., et al. (2012). GMES sentinel-1 mission. *Remote Sens. Environ.* 120, 9–24. doi: 10.1016/j.rse.2011.05.028
- Wang, X., Chen, C., Pan, Z., and Pan, Z. (2019). Fast and automatic ship detection for SAR imagery based on multiscale contrast measure. *IEEE Geosci. Remote Sens. Lett.* 16, 1834–1838. doi: 10.1109/LGRS.2019.2913873

- Wang, C.-Y., Liao, H.-Y. M., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., and Yeh, I.-H. (2020a). "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 390–391. doi: 10.48550/arXiv.1911.11929
- Wang, X., Li, G., and Zhang, X.-P. (2020c). "Contrast of contextual Fisher vectors for ship detection in SAR images," in *2020 IEEE international radar conference (RADAR) (IEEE)*, 198–202. doi: 10.1109/RADAR42522.2020.9114850
- Wang, X., Li, G., Zhang, X.-P., and He, Y. (2020d). Ship detection in SAR images via local contrast of Fisher vectors. *IEEE Trans. Geosci. Remote Sens.* 58, 6467–6479. doi: 10.1109/TGRS.2020.2976880
- Wang, W., Ren, J., Su, C., and Huang, M. (2021). Ship detection in multispectral remote sensing images via saliency analysis. *Appl. Ocean Res.* 106, 102448. doi: 10.1016/j.apor.2020.102448
- Wang, S., Wang, M., Yang, S., and Jiao, L. (2016). New hierarchical saliency filtering for fast ship detection in high-resolution SAR images. *IEEE Trans. Geosci. Remote Sens.* 55, 351–362. doi: 10.1109/TGRS.2016.2606481
- Wang, D., Wan, J., Liu, S., Chen, Y., Yasir, M., Xu, M., et al. (2022). BO-DRNet: An improved deep learning model for oil spill detection by polarimetric features from SAR images. *Remote Sens.* 14, 264. doi: 10.3390/rs14020264
- Wang, R., Xu, F., Pei, J., Zhang, Q., Huang, Y., Zhang, Y., et al. (2020b). "Context semantic perception based on superpixel segmentation for inshore ship detection in SAR image," in *2020 IEEE radar conference (RadarConf20) (IEEE)*, 1–6. doi: 10.1109/RadarConf2043947.2020.9266627
- Wei, S., Su, H., Ming, J., Wang, C., Yan, M., Kumar, D., et al. (2020). Precise and robust ship detection for high-resolution SAR imagery based on HR-SDNet. *Remote Sens.* 12, 167. doi: 10.3390/rs12010167
- Willburger, K., Schwenk, K., and Brauchle, J. (2020). Amaro—an on-board ship detection and real-time information system. *Sensors* 20, 1324. doi: 10.3390/s20051324
- Xian, S. Z., W., Yuanrui, S., Wenhui, D., Yue, Z., and Kun, F. (2019). Air-ship-1.0: High resolution sar ship detection dataset. *J. Radars* 8, 852–862. doi: 10.12000/JR19097
- Xiao, Q., Cheng, Y., Xiao, M., Zhang, J., Shi, H., Niu, L., et al. (2020). Improved region convolutional neural network for ship detection in multiresolution synthetic aperture radar images. *Concurrency Computation: Pract. Exp.* 32, e5820. doi: 10.1002/cpe.5820
- Xiong, G., Wang, F., Zhu, L., Li, J., and Yu, W. (2019). SAR target detection in complex scene based on 2-d singularity power spectrum analysis. *IEEE Trans. Geosci. Remote Sens.* 57, 9993–10003. doi: 10.1109/TGRS.2019.2930797
- Xiong, B., Sun, Z., Wang, J., Leng, X., and Ji, K. A. (2022). Lightweight Model for Ship Detection and Recognition in Complex-Scene SAR Images. *Remote Sens.* 14, 6053. doi: 10.3390/rs14236053
- Xu, P., Li, Q., Zhang, B., Wu, F., Zhao, K., Du, X., et al. (2021). On-board real-time ship detection in HISEA-1 SAR images based on CFAR and lightweight deep learning. *Remote Sens.* 13, 1995. doi: 10.3390/rs13101995
- Xu, X., Zhang, X., and Zhang, T. (2022). Lite-yolov5: A lightweight deep learning detector for on-board ship detection in large-scene sentinel-1 sar images. *Remote Sens.* 14, 1018. doi: 10.3390/rs14041018
- Yasir, M., Jianhua, W., Mingming, X., Hui, S., Zhe, Z., Shanwei, L., et al. (2022). Ship detection based on deep learning using SAR imagery: a systematic literature review. *Soft Computing*, 1–22. doi: 10.1007/s00500-022-07522-w
- Yekeen, S. T., Balogun, A. L., and Yusof, K. B. W. (2020). A novel deep learning instance segmentation model for automated marine oil spill detection. *ISPRS J. Photogrammetry Remote Sens.* 167, 190–200. doi: 10.1016/j.isprsjprs.2020.07.011
- Yuan, S., Yu, Z., Li, C., and Wang, S. (2020). A novel SAR sidelobe suppression method based on CNN. *IEEE Geosci. Remote Sens. Lett.* 18, 132–136. doi: 10.1109/LGRS.2020.2968336
- Yu, J., Jiang, Y., Wang, Z., Cao, Z., and Huang, T. (2016). "Unitbox: An advanced object detection network," in *Proceedings of the 24th ACM international conference on multimedia*, 516–520. doi: 10.1145/2964284.2967274
- Zhang, X., Liu, G., Zhang, C., Atkinson, P. M., Tan, X., Jian, X., et al. (2020b). Two-phase object-based deep learning for multi-temporal SAR image change detection. *Remote Sens.* 12, 548. doi: 10.3390/rs12030548
- Zhang, K., Luo, Y., and Liu, Z. (2021a). "Overview of research on marine target recognition," in *2nd international conference on computer vision, image, and deep learning (SPIE)*, 273–282. doi: 10.1117/12.2604530
- Zhang, X., Wan, F., Liu, C., Ji, R., and Ye, Q. (2019). Freeanchor: Learning to match anchors for visual object detection. *Adv. Neural Inf. Process. Syst.* 32. doi: 10.48550/arXiv.1909.02466
- Zhang, X., Xie, T., Ren, L., and Yang, L. (2020c). "Ship detection based on superpixel-level hybrid non-local MRF for SAR imagery," in *2020 5th Asia-pacific conference on intelligent robot systems (ACIRS) (IEEE)*, 1–6. doi: 10.1109/ACIRS49895.2020.9162609
- Zhang, T., Zhang, X., and Ke, X. (2021b). Quad-FPN: A novel quad feature pyramid network for SAR ship detection. *Remote Sens.* 13, 2771. doi: 10.3390/rs13142771
- Zhang, T., Zhang, X., Ke, X., Zhan, X., Shi, J., Wei, S., et al. (2020a). LS-SSDD-v1.0: A deep learning dataset dedicated to small ship detection from large-scale sentinel-1 SAR images. *Remote Sens.* 12, 2997. doi: 10.3390/rs12182997
- Zhao, Y., Zhao, L., Xiong, B., and Kuang, G. (2020). Attention receptive pyramid network for ship detection in SAR images. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 13, 2738–2756. doi: 10.1109/JSTARS.2020.2997081