

YOLO-OSD: Optimized Ship Detection and Localization in Multiresolution SAR Satellite Images Using a Hybrid Data-Model Centric Approach

Muhammad Farhan Humayun ¹, Faryal Aurooj Nasir ¹, *Student Member, IEEE*, Farrukh Aziz Bhatti ¹, Madiha Tahir ¹, and Khurram Khurshid ¹

Abstract—With the advancements in space technology and the development of lightweight synthetic aperture radar (SAR) satellites by commercial companies, such as ICEYE, Capella Space and Umbra, SAR images have become available on a wide scale. Ship detection is a classic problem in the interpretation and analysis of satellite images and has its significance both in maritime as well as defense applications. In the case of SAR images, ship detection becomes even more challenging due to the presence of large-scale distortions as well as interclass similarity signature problem. Moreover, the state-of-the-art (SOTA) object detection models have weak generalization capability over SAR datasets. To overcome these challenges, we propose a You Only Look Once (YOLO)-based optimized ship detection model called YOLO-OSD. Our optimized ship detector is based on a hybrid data-model centric approach, which utilizes the statistical characteristics of the datasets under observation and has an efficient model architecture. We also carry out a detailed comparative analysis of our proposed model with other SOTA deep learning models on three well-known publicly available datasets. Our results show that the proposed YOLO-OSD outperforms YOLO5, YOLO7, and RetinaNet on all datasets under observation in terms of F1 score and mean average precision. YOLO-OSD also has approximately 16% fewer network parameters as compared with the original YOLO5. Moreover, our proposed model is at least 37.7% faster than YOLO7 and 41.02% faster than the YOLO8 model in terms of training time and thus suitable for real-time satellite-based SAR ship detection.

Index Terms—Bounding boxes (B-Boxes), deep learning (DL), multiresolution satellite images, ship detection, synthetic aperture radar (SAR), you only look once (YOLO).

I. INTRODUCTION

SPACE technology has shown remarkable progress in the past few decades from the manifestation of reusable launch vehicles to the development of compact and small-scale satellites. Satellite-based remote sensing has also received growing attention due to the development of lightweight and less power

Manuscript received 8 July 2023; revised 12 November 2023 and 2 January 2024; accepted 9 February 2024. Date of publication 14 February 2024; date of current version 29 February 2024. (*Corresponding author: Muhammad Farhan Humayun.*)

The authors are with the Department of Electrical Engineering and Computer Science, Institute of Space Technology, Islamabad 44000, Pakistan (e-mail: farhan.humayun2014@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2024.3365807

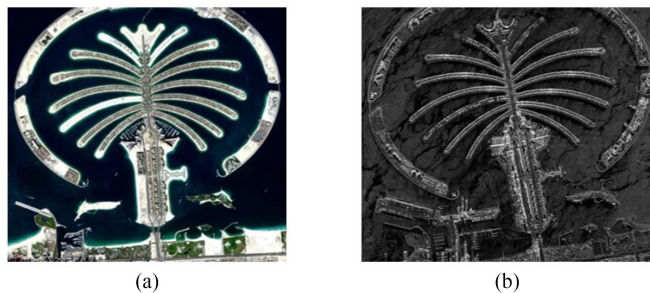


Fig. 1. Palm Island Dubai, UAE. (a) Optical Satellite image (Maxar Technologies). (b) SAR Satellite image (Capella Space).

hungry satellites capable of taking high-resolution images of Earth's surface. With the launch of new firms, such as ICEYE, Capella Space, and Umbra, the availability of high resolution synthetic aperture radar (SAR) satellite images has increased manifolds both in public and private sector domains. [1]. The problem of object detection in satellite images has important applications in civil as well as military domains. Small objects including ships, cars, air crafts, etc., can be detected using high resolution satellite imagery, whereas, satellite images with relatively medium resolution can be used to detect relatively bigger objects, such as airports, fields, and buildings [2], [3].

Among the remote sensing satellites, the SAR sensors also known as active sensors work on the principle of radars. They transmit electromagnetic waves and pick up the reflected waves from the target object in order to form the shape of the object [4]. These operate in the microwave region of the electromagnetic spectrum and therefore it is not possible to generate true color red, green, and blue images in the case of SAR sensors. However, SAR imaging is advantageous in the sense that it can penetrate clouds and even capture images during nighttime. Due to this capability, they are also called all weather, all time sensors [5]. Fig. 1 shows a visual comparison between electro-optical (courtesy Maxar Technologies) and SAR (courtesy Capella Space) satellite images of the same area.

Satellite-based ship detection has significant importance in the maritime domain awareness and can assist in various applications including monitoring of illegal fishing activities,

pirate threats, vessel traffic management along the coastlines and identification of noncooperative ships with no automatic identification system data [6], [7], [8]. SAR sensors are considered particularly useful for maritime applications due to their all-weather and day–night imaging capability. Apart from that they have wider swath coverage as compared with their optical counterparts. This enables them to monitor large areas of oceans and coastlines within shorter amounts of time [9], [10]. Despite these benefits, there still remain various challenges associated with SAR images, which makes satellite-based SAR ship detection a relatively arduous task. These are summed up below.

- 1) Involvement of deformed object boundaries due to various distortions and uneven scattering phenomena.
- 2) For applications related to the maritime domain, SAR images may involve complex backgrounds including sea clutter, islands, harbors, and ports, which negatively affect detection accuracy.
- 3) No specific color information is included in SAR images, so details in different bands cannot be leveraged as is the case in optical images.
- 4) The issues of interclass similarity signature i.e., two very different objects may have similar reflectance patterns and thus appear to be similar in a SAR image.
- 5) Most of the state-of-the-art (SOTA) object detection algorithms are tailored for optical images and thus possess weak generalization ability over SAR data.
- 6) Lack of publicly available, labeled SAR datasets as compared with the optical datasets.

In this article, we propose an you only look once optimized ship detection model named (YOLO-OSD) for multiresolution SAR satellite images using a hybrid data-model centric approach. Our main contributions are as follows.

- 1) We carry out detailed statistical analysis of three popular open source SAR ship datasets including SSDD, SAR-Ships, and iVision-MRSSD to perform anchor box customization for ease of model training and achieving better intersection over union (IoU) and mean average precision (mAP) scores.
- 2) We optimize the network architecture of YOLO5 model to significantly reduce the model network parameters and training time while simultaneously improving the ship detection accuracy in satellite-based SAR images.
- 3) We conduct extensive experiments, including detailed comparative and cross-dataset validation analysis, to assess the performance of our proposed YOLO-OSD approach. Our evaluations include quantitative and qualitative comparisons with SOTA object detectors, including YOLO5, YOLO7, YOLO8, and Retinanet.

The rest of this article is organized as follows. Section II includes a summary of past and recent works related to the domain. Section III discusses the methodology of the proposed YOLO-OSD approach including statistical analysis of the datasets involved. Section IV describes the experimental setups and details. Section V shows the quantitative and qualitative results of the experiments. Section VI provides a discussion

on the results and analysis. Finally, Section VII concludes this article, with current limitations and future research directions.

II. RELATED WORKS

Satellite-based SAR ship detection has been the focus of researchers since the past two decades. Traditional approaches for ship detection in SAR images are based on the constant false alarm rate (C-FAR) algorithm [11], [12], [13], [14]. It is an algorithm based on the statistical distribution of SAR image pixels and uses an adaptive thresholding strategy based on the false alarm rate. The thresholding mechanism employs the fact that in a SAR image, ships are normally characterized by the brightest pixels and other pixels can be treated either as background, sea clutter or other land-based features. However, the method heavily relies on predefined distributions to make detections and becomes irrelevant with changing backgrounds and imaging conditions, which is a frequently occurring case in satellite-based SAR images. Hence, techniques based on C-FAR algorithms generally have low detection performance especially in complex, in-shore scenes. Recently, Zhang et al. [15] proposed a novel ship detection method based on adaptive C-FAR for fully polarized SAR images with better detection performance and less false alarm rate as compared with previous methods.

A lot of work has also been carried out on autodetection of objects including ships using deep learning (DL) techniques. In the past decade, models based on convolutional neural networks (CNNs) have become popular in the field of computer vision, especially since the AlexNet [16] was proposed. Since then, majority of the work has been related toward the development of deeper and more complex neural network models in order to attain better accuracy [17], [18]. With the advancements in DL, various improved models have been proposed including region-based R-CNN [19], visual geometry group (VGG) [20], you only look once (YOLO) [21], fast R-CNN [22], single shot detector SSD [23], mask R-CNN [24], RetinaNet [25], fully convolutional one stage (FCOS) detector [26], YOLO-R [27], and gate recurrent CNN (GR-CNN) [28]. These models can be grouped into two broad categories, i.e., two-stage detectors e.g., R-CNN and one-stage detectors, such as SSD and YOLO. Table I enlists the popular models/algorithms used for object detection/image classification and their corresponding proposed years over the past decade.

Detailed works have also been carried out on SAR ship detection based on the DL models listed in Table I. For example, Fan et al. [29] studied ship detection in polarimetric SAR images using a modified faster RCNN model. Similarly, the authors in [30] and [31] performed SAR ship detection using modified architectures of YOLO2 and YOLO5, respectively. Wu et al. [32] carried out ship detection using a modified version of mask R-CNN. Wang et al. [33] discussed the possibility of combining SSD model with transfer learning for SAR ship detection. Xu et al. [34] studied large scale ship detection in SAR images using lite-YOLO5 model. Yang et al. [35] proposed a detection model based on coordinate attention and enhanced receptive fields and compared its performance with faster RCNN, SSD, FCOS, RetinaNet, and YOLO models. Similarly, Cui et al. [36]

TABLE I
POPULAR DL-BASED OBJECT DETECTORS IN THE PAST DECADE

Serial No.	Model / Algorithm Name	Year Proposed
1	AlexNet	2012
2	RCNN	2013
3	VGG	2014
4	Faster-RCNN	2015
5	SSD	2015
6	YOLO	2015
7	Mask R-CNN	2017
8	RetinaNet	2017
9	FCOS	2019
9	YOLOV5	2020
10	YOLOR	2021
11	G-RCNN	2021
12	YOLOV7	2022
14	YOLO8	2023
15	YOLO NAS	2023

proposed a model based on dense attention pyramid network and also compared its performance with faster RCNN and SSD. Comprehensive surveys have also been conducted on SAR ship detection using DL techniques [37], [38].

More recent research has been focused on the development of lightweight models with fewer layers and less network parameters to reduce the computational costs. For instance, Pang et al. [39] proposed a lightweight model called YOLO5-MNE by replacing the sigmoid linear unit (Silu) activation functions with rectified linear unit (ReLU) activation functions and also added a channel attention module to compensate for the loss in accuracy. The authors performed experiments on SSDD and AirSAR-Ship-1.0 datasets and compared the performance of their proposed approach with YOLO4 and YOLO5 models in terms of precision, network parameters, and GPU memory utilized. Yan et al. [40] proposed LssDet to reduce the floating point operations (FLOPs) and the number of network parameters by introducing a cross sidelobe attention module as well as a lightweight path aggregation feature pyramid network module. They performed experiments on SSDD and Ls-SSDD-v1.0 datasets and performed comparative analysis between different YOLO models in terms of average precision, network parameters, and FLOPs. Yang et al. [41] discussed a soft quantization approach to make the overall ship detection model small. The authors proposed a split bidirectional feature pyramid network to improve accuracy and a feature fusion module based on linear transformation to reduce the network size. They also performed extensive experiments on SSDD, SAR-Ships, and Air-SAR-Ship datasets and analyzed the results in terms of precision, recall, mAP, and network parameters. Similarly, Zhang and Zhang [42] proposed a ship detection model with only 20 convolutional layers for real-time SAR ship detection applications and performed experiments on SSDD dataset for comparative analysis. Yang et al. [43] proposed an algorithm/hardware co-design framework for on-board SAR ship detection with a focus on

simultaneously increasing detection accuracy by increasing output feature sizes while minimizing the need for computational resources through implementation of less expensive operations. They also performed experiments on the SSDD dataset using evaluation metrics of average precision and network parameters. Gao et al. [44] provided a comprehensive overview for on-board processing of satellite images as well as information fusion including current challenges and future prospects in the domain.

Other works have focused on developing models with better feature extraction and integration strategies to improve the performance of SAR ship detection in complex cases. For instance, Zhao et al. [45] proposed a novel visual transformer-based network for extraction of global features in the case of multisatellite SAR images. Ai et al. [46] proposed a mechanism to extract low level features based on modified C-FAR algorithm and fuse them with high level features extracted from CNNs. Cui et al. [47] proposed a spatial shuffle group enhance attention model to extract better semantic features and suppress unnecessary features. Zhou et al. [48] proposed MSSDNet comprising modules for multiscale feature extraction as well as adaptive feature fusion. Wang et al. [49] also proposed a feature transformer module for CNNs to extract global features from SAR ship images. Similarly, Gao et al. [50] proposed a dualistic cascade CNN comprising of basic geometric feature extraction network, and polarization feature enhancement network for comprehensive feature extraction and fusion.

Apart from that, anchor free ship detection models have also been proposed in an effort to reduce the dependency of model detection on the size of anchor boxes to reduce computational cost [51], [52]. Similarly, researchers have also explored the possibility of optical to SAR transfer learning to augment ship detection in SAR images. Bao et al. [53] proposed an optical-SAR pretraining approach to transfer characteristics of optical images to SAR images through common representation to improve model learning. Gao et al. [54] also proposed a novel method comprising of dense connection module and convolutional block attention module for enhanced feature extraction in the case of optical to SAR transfer learning for sparsely labeled datasets.

One of the major issues related to the applications of SAR imagery is the sparse availability of public datasets as compared with optical datasets. Majority of the traditional pre-trained object detection models are trained on optical data or simple camera images, which become less relevant in the case of SAR images. Due to these reasons, researchers have also published their own satellite-based SAR image datasets for ship detection. Significant satellite-based SAR ship datasets include SSDD [55], SAR-Ships dataset [56], AirSAR-Ship-1.0 [57], HRSID [58], Ls-SSDD-v1.0 [59], SRSSD-v1.0 [60], and the latest iVision-MRSSD [61], [62]. The SRSSD-v1.0 is based on rotated bounding boxes (B-Boxes) whereas, rest of the datasets comprise of regular upright B-Boxes. Table II provides a comparative analysis of the publicly available SAR ship datasets in terms of various parameters including the number of images, individual image size, number of sensors used and coverage of imaging frequencies. From Table II, it can be inferred that the latest iVision-MRSSD dataset is the most diverse in terms of satellite sensors and frequency bands involved.

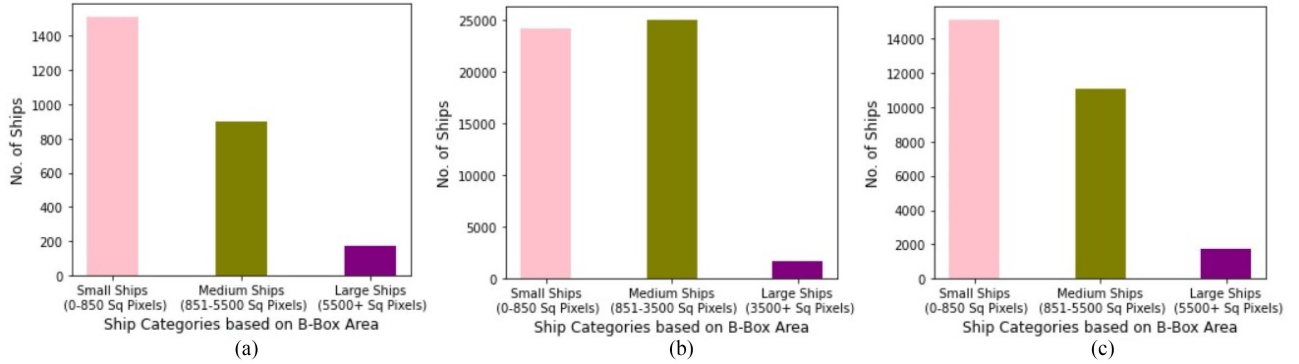


Fig. 2. Categorization of B-Boxes with respect to area (a) SSDD, (b) SAR-Ships, and (c) iVision-MRSSD.

TABLE II
COMPARISON OF PUBLICLY AVAILABLE SATELLITE BASED SAR SHIP DATASETS

Dataset	No. of Image Tiles	Tile Size	No. of Sensors Used	Frequency Bands Covered	Year of Release
SSDD	1160	Variable	03	C, X	2017
SAR-Ships	43,819	256 × 256	02	C	2019
Air-SAR-Ship	31	3000 × 3000	01	C	2019
HRSID	5604	800 × 800	03	C, X	2020
L-SSDD	9000	800 × 800	01	C	2020
SRSSD	666	1024 × 1024	01	C	2021
iVision-MRSSD	11,590	512 × 512	06	C, L, X	2023

III. MATERIALS AND METHODS

This section describes the main idea of the proposed YOLO-OSD approach. Section II-A comprises detailed statistical analysis of the three SAR ship detection datasets under observation for the generation of customized anchor boxes with respect to each dataset. Section II-B discusses the proposed architectural changes pertaining to YOLO-OSD and compares them with the original YOLO5 architecture.

A. Custom Anchor Box Generation Strategy

We have taken into account three different datasets named SSDD, SAR-Ships, and iVision-MRSSD to implement our proposed strategy. The details of these datasets are already described in Table II. The core idea of this strategy is to optimize the initial anchor box sizes used by the SOTA object detection models based on the inherent dataset distributions. Anchor boxes play a

vital role in object detection algorithms, such as YOLO and RetinaNet. They represent a set of predefined B-Boxes with different aspect ratios to facilitate object detection corresponding to different sizes and dimensions. A careful and systematic selection of anchor box sizes can allow the models to learn robust features and better adapt to the diversities within the datasets. It can also facilitate in the faster training process.

Initially, we extract the areas of the ship B-Boxes from labels of each dataset and categorize them into small, medium and large categories to visualize their distribution. Fig. 2 shows the area-wise categorization of B-Boxes in SSDD, SAR-Ships, and iVision-MRSSD, respectively. It can be seen that the proportion of small B-Boxes (covering less area) in all the datasets is large as compared with big B-Boxes.

Fig. 3 shows the point-wise distribution of bounding box areas in each dataset. It can be seen that majority of the points are closer to the x -axis again depicting a significant number of small ships.

To categorize the B-Boxes into a suitable number of clusters, we perform Elbow analysis. It is a method to determine the number of suitable clusters based on the dataset for a clustering problem. Elbow analysis basically runs K-Means clustering on the given dataset by setting different values of K i.e., the number of clusters. In our case, the value of K for the Elbow analysis ranged between 2 and 14. After that, it plots the sum of squared differences of the data samples with their allotted cluster centers also known as inertia. The point where the trend of the graph changes is called the elbow point and the number of clusters at that point are considered to be suitable for the dataset under observation. Fig. 4 shows the plot of inertia with respect to the number of possible clusters for each dataset. It can be seen that all the plots follow a similar trend with a typical elbow like shape being formed around 3 to 4 clusters. Looking at the trend in the elbow plots, we determine 4 to be the suitable number of clusters for all datasets.

After the Elbow analysis, we apply K-Means clustering with $K = 4$ to form suitable clusters. Fig. 5 shows a visualization of 4 possible B-Box clusters for each dataset.

We further analyze the ship B-Boxes on the basis of their width–height ratios. Fig. 6 depicts the width–height bounding box plots of SSDD, SAR-Ships, and iVision-MRSSD, respectively, with width on the x -axis and corresponding height plotted on the y -axis. It can be seen that most of the points are

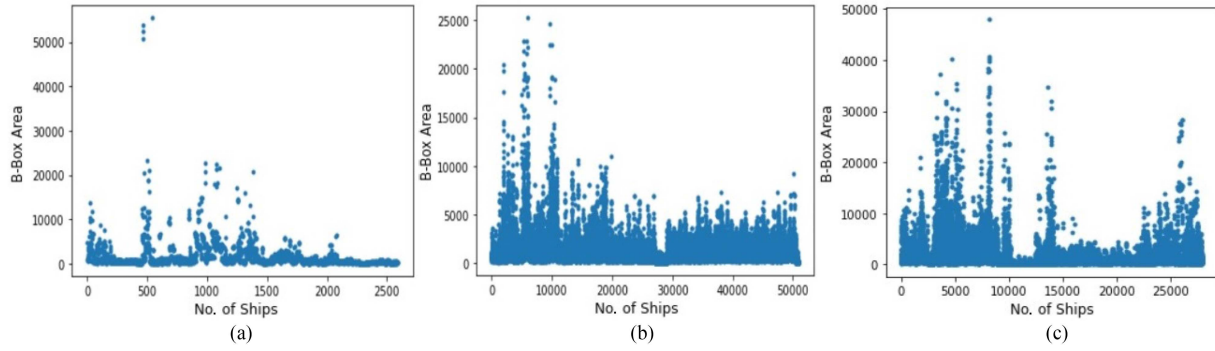


Fig. 3. Point-wise distribution B-Boxes with respect to area (a) SSDD, (b) SAR-Ships, and (c) iVision-MRSSD.

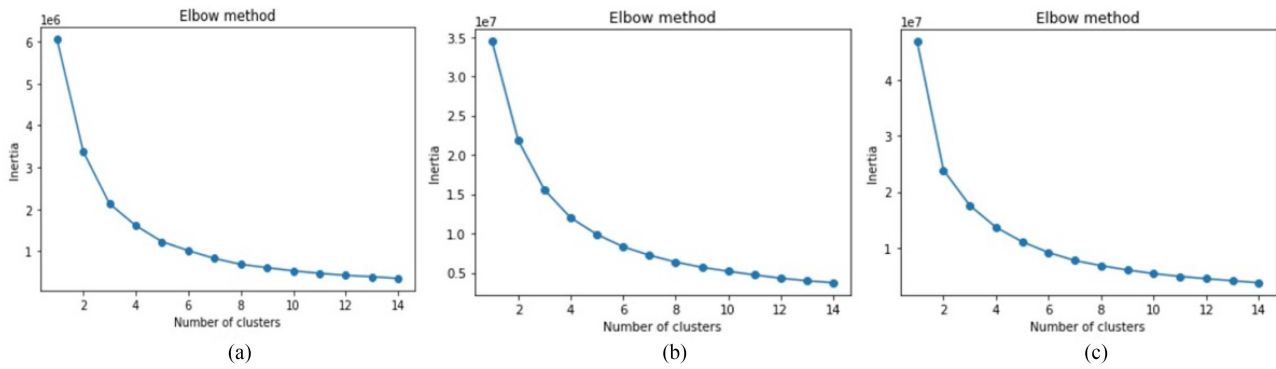


Fig. 4. Elbow plots of B-Boxes area values (a) SSDD, (b) SAR-Ships, and (c) iVision-MRSSD.

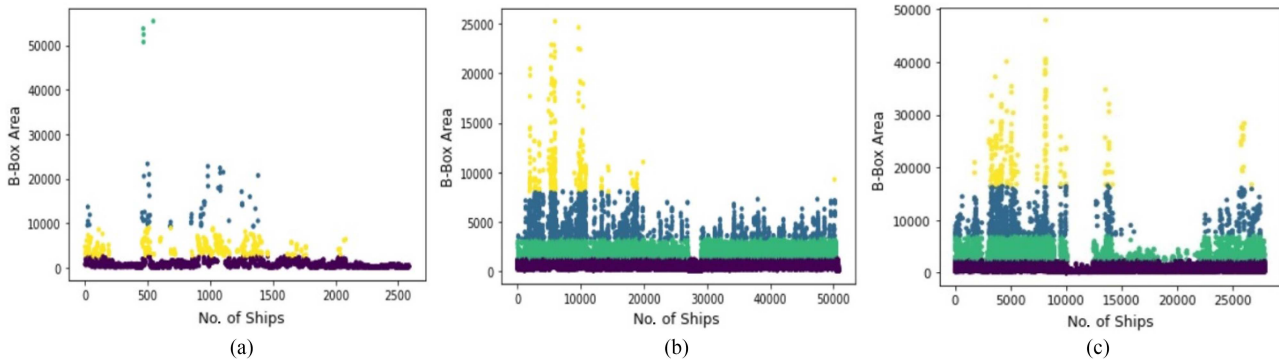


Fig. 5. Output of K-Mean clustering on B-Box area values (K=4) (a) SSDD, (b) SAR-Ships, and (c) iVision-MRSSD.

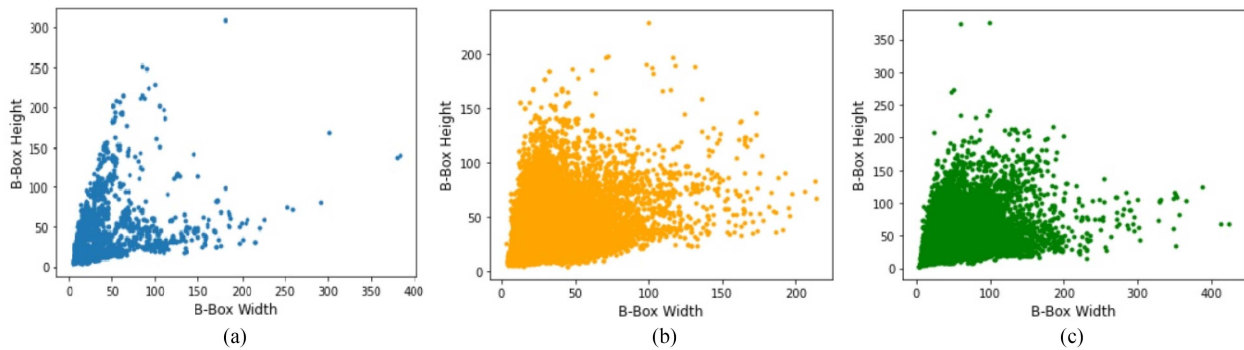


Fig. 6. Point-wise distribution of B-Box width–height (a) SSDD, (b) SAR-Ships, and (c) iVision-MRSSD.

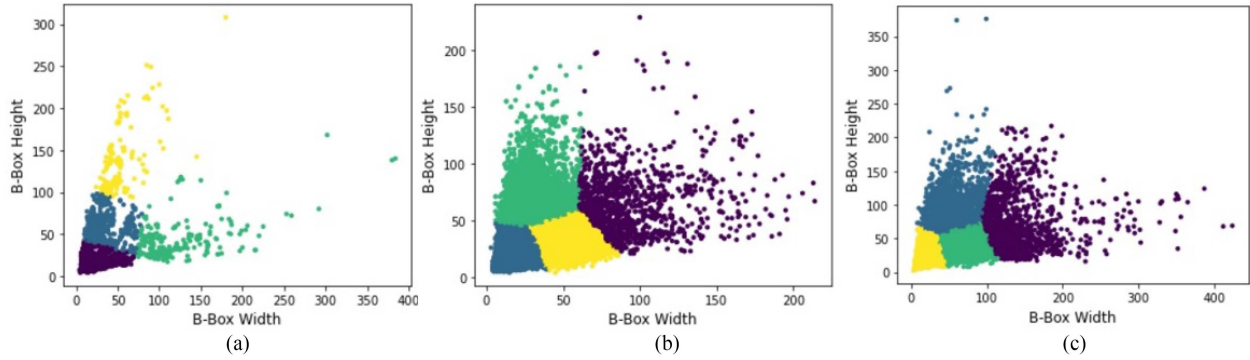


Fig. 7. Output of K-Means clustering ($K = 4$) on B-Box width–height values (a) SSDD, (b) SAR-Ships, and (c) iVision-MRSSD.

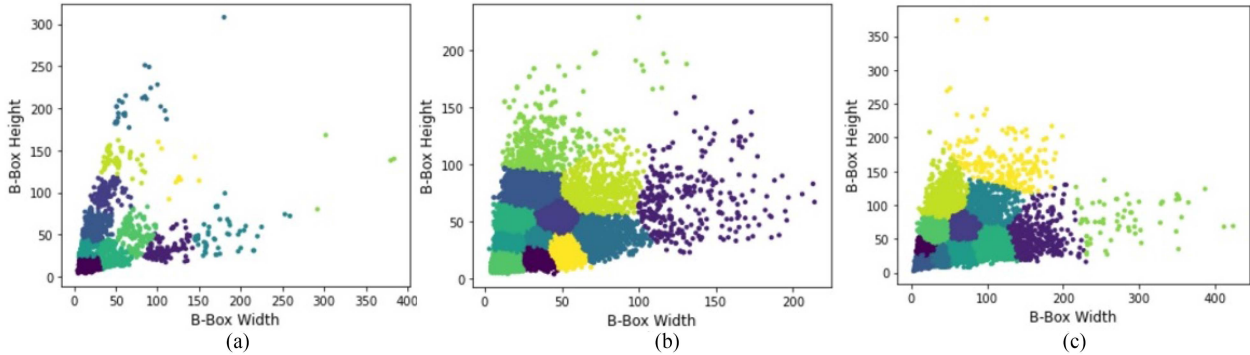


Fig. 8. Output of K-Means clustering ($K = 12$) on B-Box width–height values (a) SSDD, (b) SAR-Ships, and (c) iVision-MRSSD.

TABLE III
SET OF FINAL ANCHOR BOX SIZES BASED ON DATA DISTRIBUTION

Dataset	Model	No. of Anchors	Anchor Box Sizes
Generic	YOLO5 (Default)	9	(10,13), (16,30), (33,23), (30,61), (62,45), (59,119), (116,90), (156,198), (373,326)
SSDD	YOLO-OSD (Proposed)	12	(11,12), (24,13), (13,29), (16,44), (44,18), (37,35), (23,59), (107,31), (40,102), (79,53), (54,169), (199,67)
SAR-Ships	YOLO-OSD (Proposed)	12	(12,14), (22,16), (13,30), (37,17), (24,28), (20,44), (39,27), (37,41), (53,35), (27,73), (42,60), (74,68)
iVision-MRSSD	YOLO-OSD (Proposed)	12	(12,12), (15,12), (25,19), (18,34), (45,21), (34,35), (25,22), (64,33), (56,57), (42,90), (117,43), (107,107)

concentrated in the lower left corner of each plot indicating a higher concentration of relatively small B-Boxes corresponding to small ships. This is a crucial finding, which also influenced the idea behind our proposed architectural changes in the YOLO5 backbone.

We again apply K-Mean clustering with $K = 4$ on the width–height data of B-Boxes to determine the final clusters. The argument behind this is that clusters based on width–height ratios are a better representation of the data as compared with clusters based on area. This is because detection performance is sensitive to the widths and height ratios of B-Boxes and not just their area. Fig. 7 shows a visualization of probable B-Box clusters for each

dataset in terms of width–height ratios for the three datasets under observation.

An important thing to consider here is that, the original YOLO5 model utilizes a total of 9 initial anchor box sizes, 3 for each scale pertaining to small, medium, and large scale detections. Based on our findings, we propose 12 anchor boxes i.e., 4 anchor boxes for each scale of detection. Fig. 8 shows the outputs of K-Means clustering with $K = 12$.

Table III summarizes the final values of anchor box dimensions achieved after minor adjustments with respect to all the datasets. The default values for the original YOLO5 model are also included for comparison purposes.

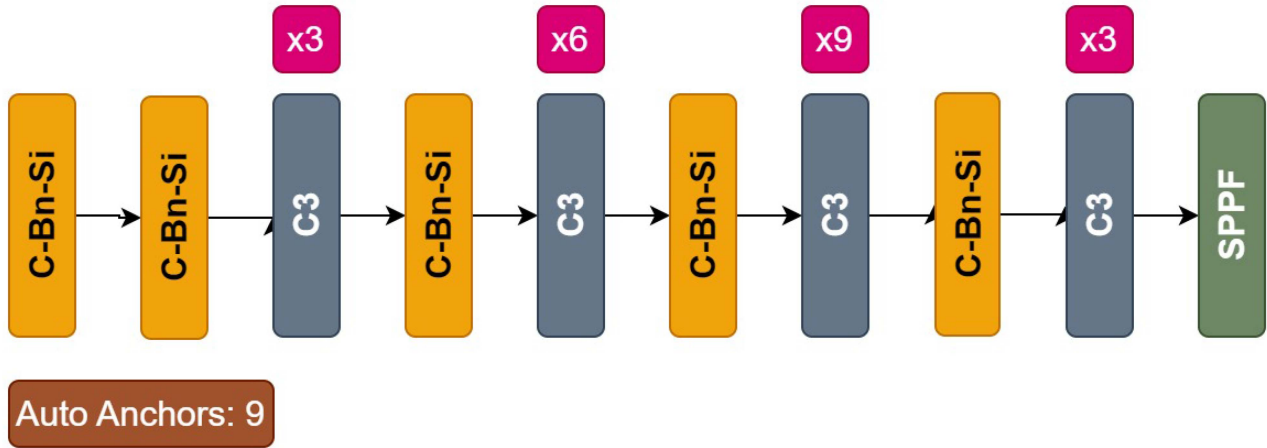


Fig. 9. Network backbone of original YOLO5 model.

B. Network Architecture of the Proposed YOLO-OSD

This section focuses on the architectural aspects of the newly proposed model. We propose an optimized SAR ship detector, suitable for real-time ship detection in multiresolution SAR satellite images by modifying the YOLO5 backbone to achieve faster execution/training time as well as improve the model performance in terms of evaluation metrics including precision, recall, F-1 score, and mAP. The architectures of YOLO-based object detectors are comprised of three primary parts including the network backbone, neck, and head. The backbone comprising of various convolutional layers is responsible for feature extraction and learning. In YOLO5, the backbone is based on cross stage partial network [63]. The neck of YOLO5 is based on the path aggregation network [64], which involves a feature pyramid network to enhance learning of the low-level features. The YOLO5 head responsible for final detection results is based on the principle of multiscale detection and uses three different sizes of feature maps to detect small, medium, and large objects. This is very important in our case as the sizes of ships vary throughout the datasets from very small boats to large ships, as described in the previous section. To develop an optimized SAR ship detection model, we first reduce the number of repetitive C3 modules in the YOLO5 backbone. C3 modules consist of triple convolution operations and are computationally very expensive. Reducing the number of C3 modules has two immediate effects: 1) it significantly reduces the overall number of network parameters thus making the model lightweight and 2) it results in increased size of feature maps because reducing the C3 modules means reducing the number of convolution operations. The size of intermediate feature maps and the number of convolution layers have an inverse relationship with one another. This is due to the fact that the convolution layers typically involve strided and down-sampling operations to reduce the computational costs but this also results in the reduction of spatial size of the feature maps. In case of multiple layers of convolutions the spatial size of the intermediate feature maps gets reduced too much, therefore, features pertaining to small objects do not propagate well through the network, effectively blocking the learning of

small objects. This in turn leads to increased false negatives in the case of small objects [43], [47], [65]. This issue becomes even more important in the case of ship detection, as detailed anchor analysis of ship datasets in the previous section indicates the presence of a large number of small ships. Fig. 9 shows the typical structure of the YOLO5 model consisting of C-Bn-Si, triple convolution (C3), and spatial pyramid pooling fast blocks. C-Bn-Si stands for Convolution-2D, batch normalization, and SiLU activation function, respectively. It can be seen in Fig. 9 that the C3 module is repeated a total of 21 times. Also, the default number of anchors in YOLO5 is 9.

Apart from that, we also change the normal C3 blocks with the triple cross-convolution (C3x) blocks. The C3 modules involve triple convolutions with a concatenation function at the end. On the other hand, the C3x module comprises of C3x. The basic idea behind cross convolutions also termed as asymmetric or spatial shuffle convolutions is that the normal 2-D convolutions are decomposed into two 1-D convolutions. For instance, a regular 2-D convolution filter of size $n \times n$ is modularized into two convolution filters of size $1 \times n$ and $n \times 1$, respectively. This results in a significant decrease in the overall network parameters, hence making the model lightweight [66]. In our case, the C3x modules involve 1×3 and 3×1 convolutions instead of single 3×3 operations. These are also considered better feature extractors as compared with the regular convolutions when it comes to dealing with target objects oriented at varying angles, which is indeed the case with ship detection [67]. Fig. 10 shows the comparison of C3 and C3x computation blocks, respectively.

The modified backbone of YOLO-OSD is shown in Fig. 11. It can be seen that we replaced the simple triple convolution blocks with C3x blocks to enhance the feature extraction process. Moreover, we reduced the number of triple convolution blocks from 21 to 15. This effectively reduced the number of network parameters by approximately 16% making the model lightweight as compared with the original YOLO5 model without compromising the detection performance. Table IV summarizes the comparisons of various network attributes of the original YOLO5 and the proposed YOLO-OSD model. It can

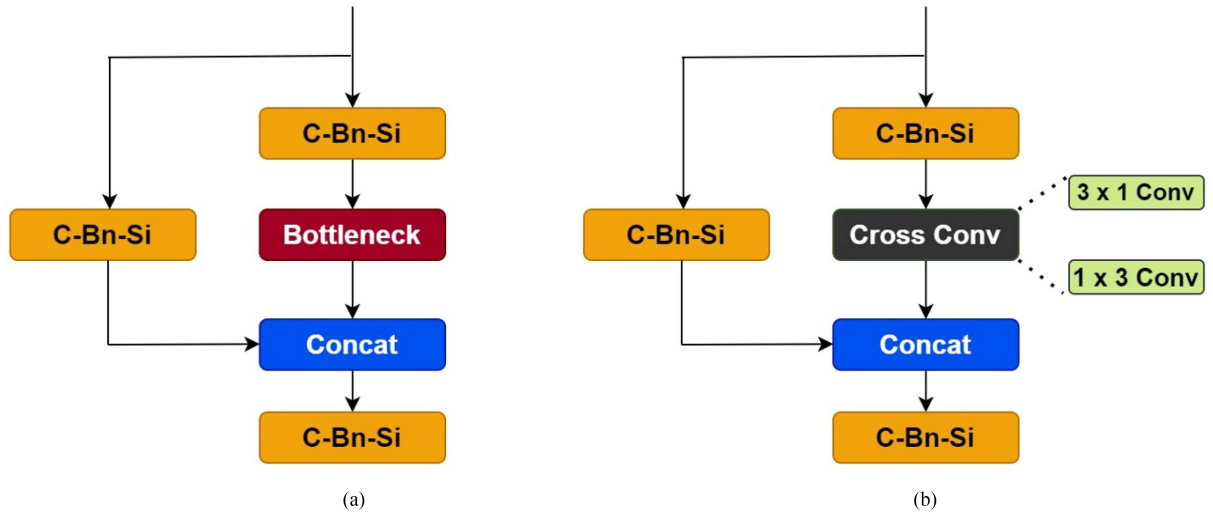


Fig. 10. Structure of (a) C3 block and (b) C3-x block.

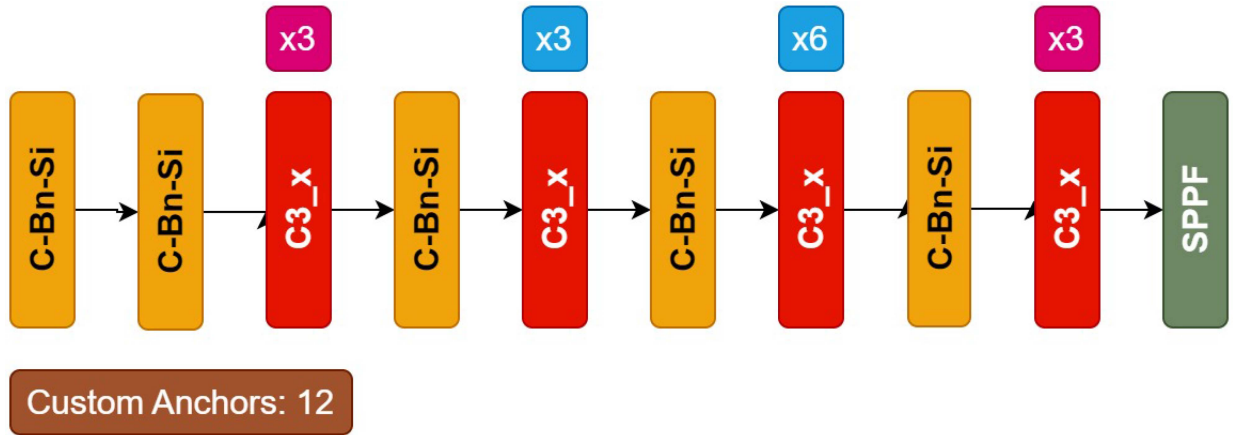


Fig. 11. Network backbone of modified YOLO-OSD model.

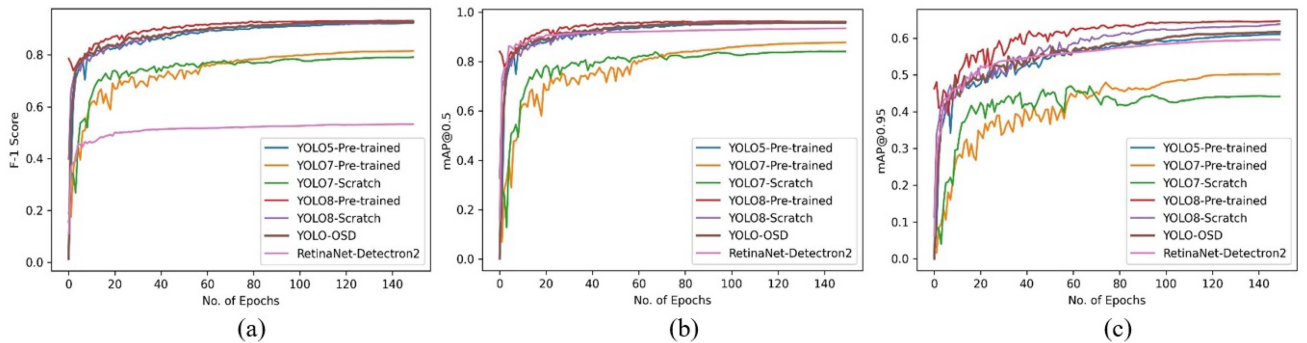


Fig. 12. (a)–(c) F-1, mAP@0.5 and mAP@0.95 graphs of all models trained on iVision-MRSSD, respectively.

be seen that the number of layers in the YOLO-OSD is reduced to 326 as compared with 368 of the original YOLO5 model. Consequently, the network parameters are also reduced significantly. Similarly, our model also requires fewer floating point operations per second (GFLOPs).

IV. EXPERIMENTAL DESIGN AND SETUP

This section describes various experimental aspects of our work. Specifically, Section III-A delineates the environmental setup and Section III-B discusses various evaluation metrics

TABLE IV
COMPARISON OF NETWORK ATTRIBUTES OF YOLO5 AND YOLO-OSD

Model	No. of Layers	No. of Parameters	GFLOPS
YOLO5 (Original)	368	46 138 294	108.2
YOLO-OSD (Proposed)	326	38 722 504	83.0

used to assess the performance of different DL models on the datasets under observation.

A. Environmental Setup

All the experiments including statistical analysis, model training, validation, and testing of the datasets were performed on a local machine having Windows 10 Pro operating system, Intel core-i7 CPU, 32 GB RAM, 2 TB HDD, and NVIDIA RTX GeForce 2080 GPU with 8 GB graphics memory. All the programming was done in Python language using the Anaconda development environment. Moreover, a train: validation: test set ratio of 70:20:10 was used throughout the experiments. For SAR-Ships dataset, the training, validation, and test sets were generated by randomly picking tiles in the proposed ratios. In the case of iVision-MRSSD and SSDD datasets, they already contained training, validation and test sets, so no further processing was required. The batch size for SSDD and iVision-MRSSD was set to 16, whereas for SAR-Ships, it was set to 32 image tiles per batch. All the models were trained for 150 epochs.

B. Performance Evaluation Metrics

Following evaluation metrics were used to assess the performance of SAR ship detection on various datasets involved.

- 1) *Precision*: Determines how many positively detected ships are actually correct

$$\text{Precision} = \frac{\text{True Detections}}{\text{True Detections} + \text{False Detections}}. \quad (1)$$

- 2) *Recall*: It is also termed as sensitivity of the model, determines that how many of the actual ships have been detected correctly

$$\text{Recall} = \frac{\text{True Detections}}{\text{True Detections} + \text{False Omissions}}. \quad (2)$$

- 3) *F-1 Score*: It merges the two metrics i.e., precision and recall into a single metric. To achieve a high F-1 score, both precision and recall need to be high

$$\text{F-1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (3)$$

- 4) *mAP*: It is the mean of average precisions of all classes, where the average precision is simply the area under the precision-r-ecall curve. In our case, we have only a single class and hence $N = 1$

$$\text{Mean Avg Precision} = \frac{1}{N} \int_{k=1}^n \text{AP}_k. \quad (4)$$

V. RESULTS AND ANALYSIS

This section discusses the results of various models trained on the three datasets under observation and analyzes their performance, both quantitatively and qualitatively based on the evaluation metrics. We trained eight models, i.e., two variants each for the YOLO5, YOLO7, and YOLO8 object detectors, one for RetinaNet and one for our proposed YOLO-OSD on all the three datasets. A total of 24 models were trained. For the YOLO models, one variant was trained from scratch without using any pretrained weights, while the second variant was trained using transfer learning with pretrained weights of the MS COCO dataset. The MS COCO dataset [68] comprises of optical images and contains 80 classes. It also includes a class labeled “boat.” We did this to analyze the effect of transfer learning on the task of ship detection in SAR images.

A. Performance Evaluation on iVision-MRSSD Dataset

Table V summarizes the results of YOLO5, 7, 8, RetinaNet, and YOLO-OSD object detectors on the iVision-MRSSD dataset in terms of various evaluation metrics. All the models were trained for 150 epochs and a batch size of 16. It can be inferred from Table V that the proposed YOLO-OSD outperforms all the models in terms of training time, whereas, it also performs better than YOLO5, YOLO7, and RetinaNet in terms of F1 score and mAP. YOLO8 (pretrained) model has the overall best detection performance and YOLO7 model shows the lowest performance in terms of mAP scores. It can also be seen that the pretrained models of YOLO7 and YOLO8 have slightly better performance as compared with the models trained from scratch. Apart from that the training time for RetinaNet model is the highest among all models.

Similarly, Fig. 12(a)–(c) shows F1-Confidence curves, mAP@0.5, and mAP@0.95 graphs of all the models with respect to the number of training epochs on the iVision-MRSSD dataset. The plots pertaining to YOLO8 in red color show highest F1 and mAP values. Our proposed YOLO-OSD (brown color) is very close second to YOLO8 on these values.

Fig. 13 shows the detection performance of various DL object detectors on a small subset of iVision-MRSSD test set. B-Boxes in blue color represent ground truths, whereas yellow, red, cyan, and green colors represent detections from YOLO7, 5, 8, RetinaNet, and YOLO-OSD, respectively.

B. Performance Evaluation on SAR-Ships Dataset

We also trained all the YOLO variants and the RetinaNet on the open source dataset named SAR-Ships for comparison purposes. Table VI sums up the results of object detectors on SAR-ships dataset. It can be seen from Table VI that the YOLO8 (pretrained) model also outperforms other models on the SAR-Ship dataset with our proposed model YOLO-OSD also outperforming YOLO5, YOLO7, and RetinaNet. On the other hand, the training time of YOLO-OSD is significantly less as compared with all the other models.

Fig. 14(a)–(c) shows F1, mAP@0.5, and mAP@0.95 plots of all models with respect to the number of training epochs on the

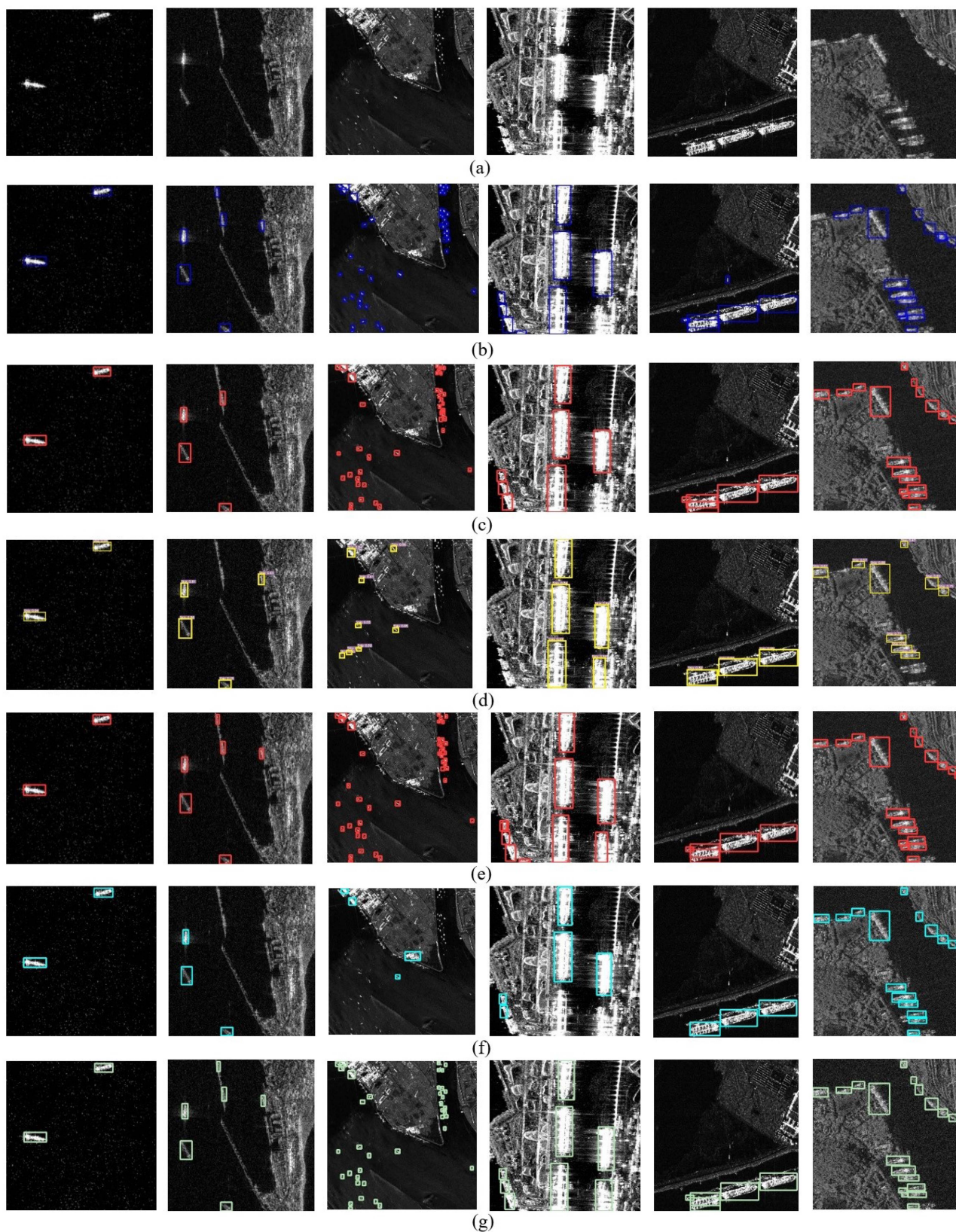


Fig. 13. Detection results on iVision-MRSSD. (a) Raw images. (b) Ground truths. (c) Yolo5 detections. (d) Yolo7 detections. (e) Yolo8 detections. (f) RetinaNet detections. (g) Yolo-OSD detections.

TABLE V
PERFORMANCE COMPARISON OF YOLO VARIANTS ON IVISION-MRSSD DATASET

Model	Precision	Recall	F-1 Score	mAP@0.5	mAP@0.95	Training Time (Hrs)
YOLO5 (Pre-trained)	0.927	0.917	0.922	0.956	0.61	6.249
YOLO5 (Scratch)	0.927	0.917	0.922	0.956	0.61	6.279
YOLO7 (Pre-trained)	0.852	0.781	0.815	0.877	0.502	9.767
YOLO7 (Scratch)	0.821	0.763	0.791	0.841	0.441	9.588
YOLO8 (Pre-trained)	0.939	0.925	0.932	0.962	0.646	10.021
YOLO8 (Scratch)	0.929	0.921	0.925	0.960	0.638	9.988
RetinaNet - Detectron2	0.957	0.368	0.532	0.934	0.595	28.322
YOLO-OSD (Proposed)	0.933	0.921	0.925	0.959	0.616	5.583

TABLE VI
PERFORMANCE COMPARISON OF YOLO VARIANTS ON SAR-SHIPS DATASET

Model	Precision	Recall	F-1 Score	mAP@0.5	mAP@0.95	Training Time (Hrs)
YOLO5 (Pre-trained)	0.925	0.918	0.921	0.956	0.591	7.846
YOLO5 (Scratch)	0.925	0.918	0.921	0.956	0.591	7.96
YOLO7 (Pre-trained)	0.91	0.874	0.892	0.935	0.53	14.516
YOLO7 (Scratch)	0.896	0.873	0.884	0.928	0.526	14.368
YOLO8 (Pre-trained)	0.945	0.938	0.941	0.973	0.672	12.062
YOLO8 (Scratch)	0.923	0.93	0.926	0.965	0.635	11.749
RetinaNet - Detectron2	0.96	0.602	0.740	0.952	0.546	34.366
YOLO-OSD	0.923	0.926	0.924	0.960	0.598	7.092

SAR-Ships dataset. Again graphs of YOLO8 in red and purple colors depict the highest values with our proposed YOLO-OSD in brown color coming up very close second.

Fig. 15 shows the detection performance of various DL object detectors on a small subset of the SAR-Ships test set. B-Boxes in blue color represent ground truths, whereas yellow, red, cyan, and green colors represent detections from YOLO7, 5, 8, RetinaNet, and YOLO-OSD models, respectively.

C. Performance Evaluation on SSDD Dataset

On a similar terms, we also trained the seven variants of YOLO and one RetinaNet on the open source dataset named SSDD for comparison purposes. Table VII summarizes the results on the SSDD dataset. All the models were trained for 150 epochs.

Table VII also confirms the superiority of YOLO8 (pretrained) model on the SSDD dataset as it gives best performance in terms of all evaluation metrics when compared with other models. Our

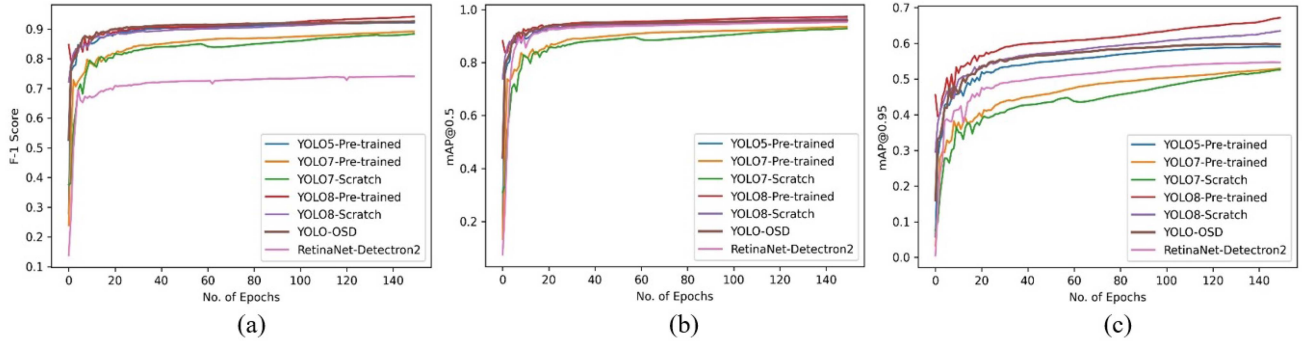


Fig. 14. (a)–(c) F-1, mAP@0.5, and mAP@0.95 graphs of all models trained on SAR-Ships, respectively.

TABLE VII
PERFORMANCE COMPARISON OF YOLO VARIANTS ON SSDD DATASET

Model	Precision	Recall	F-1 Score	mAP@0.5	mAP@0.95	Training Time (Hrs)
YOLO5 (Pre-Trained)	0.969	0.918	0.943	0.97	0.628	0.873
YOLO5 (Scratch)	0.927	0.917	0.922	0.973	0.638	0.882
YOLO7 (Pre-Trained)	0.861	0.844	0.852	0.894	0.516	1.301
YOLO7 (Scratch)	0.883	0.824	0.852	0.885	0.498	1.312
YOLO8 (Pre-Trained)	0.963	0.945	0.954	0.98	0.741	3.487
YOLO8 (Scratch)	0.962	0.927	0.944	0.981	0.712	3.529
RetinaNet - Detectron2	0.983	0.423	0.591	0.961	0.627	7.128
YOLO-OSD	0.963	0.932	0.947	0.977	0.638	0.781

YOLO-OSD also outperforms all the models in terms of training time whereas it has better detection performance than YOLO5, YOLO7, and RetinaNet in terms of F-1 score and mAP.

Similarly, Fig. 16(a)–(c) shows F1, mAP@0.5, and mAP@0.95 plots of YOLO5, 7, 8, RetinaNet, and YOLO-OSD models with respect to the number of training epochs on the SSDD dataset. Similar trends can be seen in the performances, as YOLO8 and the proposed YOLO-OSD outperform others on SSDD as well.

Fig. 17 shows the detection performance of various DL object detectors on a small subset of SSDD test set. B-Boxes in blue color represent ground truths, whereas yellow, red, cyan, and green colors represent detections from YOLO7, 5, 8, RetinaNet, and YOLO-OSD models, respectively.

D. Ablation Study of the Proposed YOLO-OSD

This section aims to determine the performance improvements brought by each modification in the proposed

YOLO-OSD. As mentioned in the previous section, a total of three improvements/modifications have been carried out in the newly proposed approach. These include:

- 1) optimization of the initial anchor box sizes based on the statistical analysis of the datasets,
- 2) reduction of the repetitive triple convolution (C3) modules in the YOLO5 backbone, and
- 3) exchanging the C3 blocks with the C3x (C3x) modules.

We have carried out ablation experiments with respect to all the three datasets under observation. Table VIII summarizes the results of the ablation study. It can be seen from Table VIII that the first rows pertaining to each dataset correspond to the results of the unmodified YOLO5. The next three rows show the effect of adding each individual modification as described above. It can be inferred from the results of the ablation study that reduction of C3 modules reduces the training time of the model significantly with similar detection performance in case of the MRSSD dataset, whereas the detection performance reduced slightly for SAR-Ships and SSDD datasets. Following that the

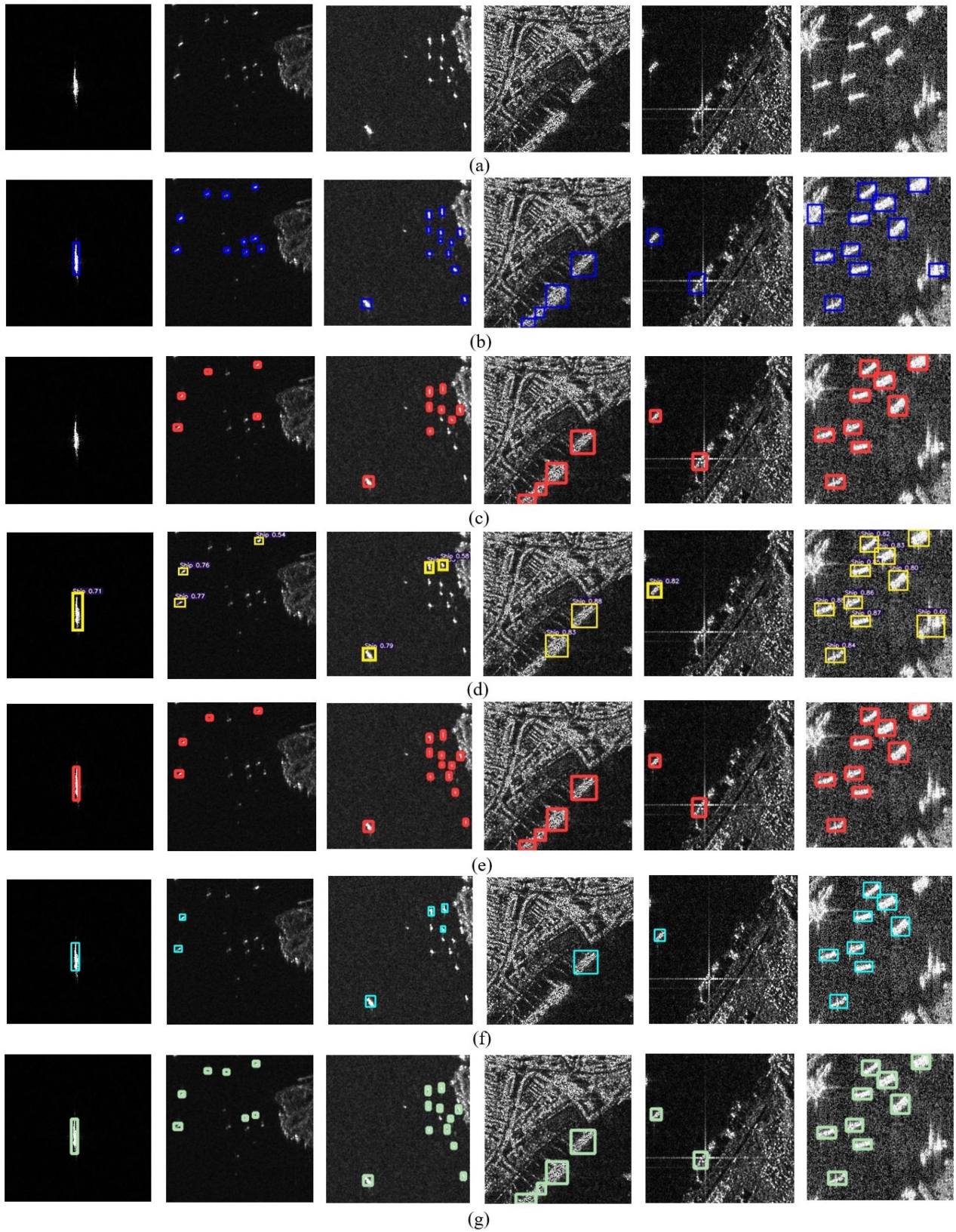


Fig. 15. Detection Results on SAR-Ships: (a) Raw images. (b) Ground truths. (c) Yolo5 detections. (d) Yolo7 detections. (e) Yolo8 detections. (f) RetinaNet detections. (g) Yolo-OSD detections.

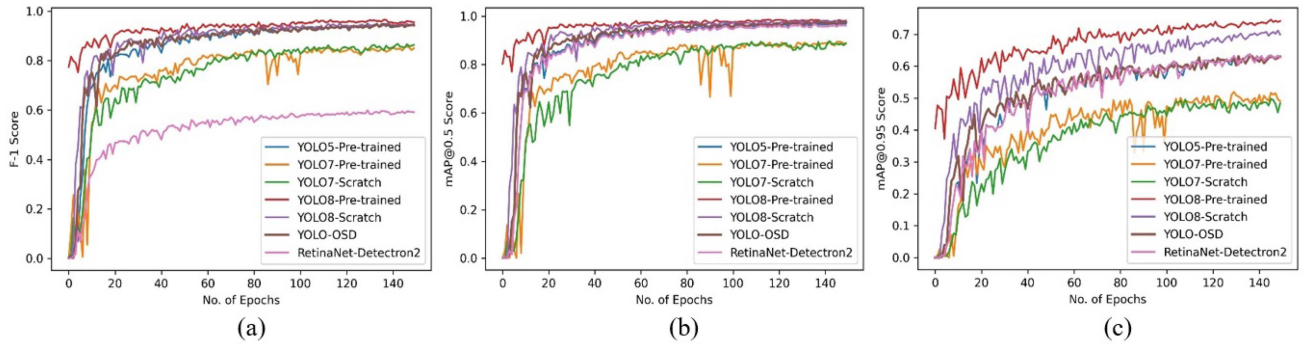


Fig. 16. (a)–(c) F-1, mAP@0.5 and mAP@0.95 graphs of all models trained on SSDD, respectively.

TABLE VIII
SUMMARY OF ABLATION EXPERIMENTS RELATED TO THE PROPOSED YOLO-OSD MODEL

Dataset	Optimized Anchor Boxes	Reduced C3 Modules	C3x Modules	Precision	Recall	F1-Score	mAP@0.5	mAP@0.95	Trg Time (Hrs)
MRSSD	X	X	X	0.927	0.917	0.922	0.956	0.61	6.249
	X	✓	X	0.935	0.915	0.925	0.957	0.605	5.417
	X	✓	✓	0.936	0.901	0.918	0.954	0.598	5.310
	✓	✓	✓	0.934	0.922	0.928	0.959	0.616	5.483
SAR-Ships	X	X	X	0.925	0.918	0.921	0.956	0.591	7.846
	X	✓	X	0.923	0.913	0.918	0.954	0.585	6.808
	X	✓	✓	0.916	0.92	0.918	0.954	0.588	6.754
	✓	✓	✓	0.927	0.927	0.927	0.960	0.598	6.738
SSDD	X	X	X	0.969	0.918	0.943	0.97	0.628	0.873
	X	✓	X	0.959	0.896	0.926	0.962	0.629	0.778
	X	✓	✓	0.965	0.920	0.942	0.973	0.630	0.773
	✓	✓	✓	0.963	0.932	0.947	0.977	0.638	0.781

Bold values represent the best results pertaining to performance criteria mentioned in column headings for a particular dataset.

exchanging of C3 modules with C3x modules further reduces the training time while minimally improving the detection performance in the case of SAR-Ships and SSDD datasets. Finally, the inclusion of optimized anchor boxes along with the other modifications leads to best detection performance with respect to all datasets with significant improvements in training times as well, signifying the efficacy of our proposed approach.

E. Cross Validation Analysis

This section focuses on the verification and finding the effective combination of DL models and the datasets on unseen data distributions. We performed cross dataset validation of all the YOLO models on the three corresponding datasets under observation. Effectively, each YOLO model trained on one dataset was validated on the validation sets of the other two datasets under observation. Table IX summarizes the results of cross validation analysis.

It can be seen from Table IX that the YOLO5 model trained on the iVision-MRSSD dataset performs comparatively better when evaluated on the validation sets of SAR-Ships and SSDD datasets. Similar trends can be seen in the case of YOLO7, YOLO8, and YOLO-OSD models. On the other hand, models trained on SAR-Ships and SSDD datasets have relatively low F-1 scores and mAP values when evaluated on the validation set of iVision-MRSSD. From these results, it can be concluded that the iVision-MRSSD dataset is more diverse and complex as compared with other datasets. This is also evident from the fact that it comprises of data from six different satellite sensors and covers a large range of spatial resolutions and imaging frequencies. Therefore, models trained on the iVision-MRSSD dataset learn features that generalize well on the unseen data with varying distributions. Similarly, it can be seen that the YOLO-OSD model, especially in combination with the iVision-MRSSD dataset has a better F-1 score as compared with other models in the cross validation results.

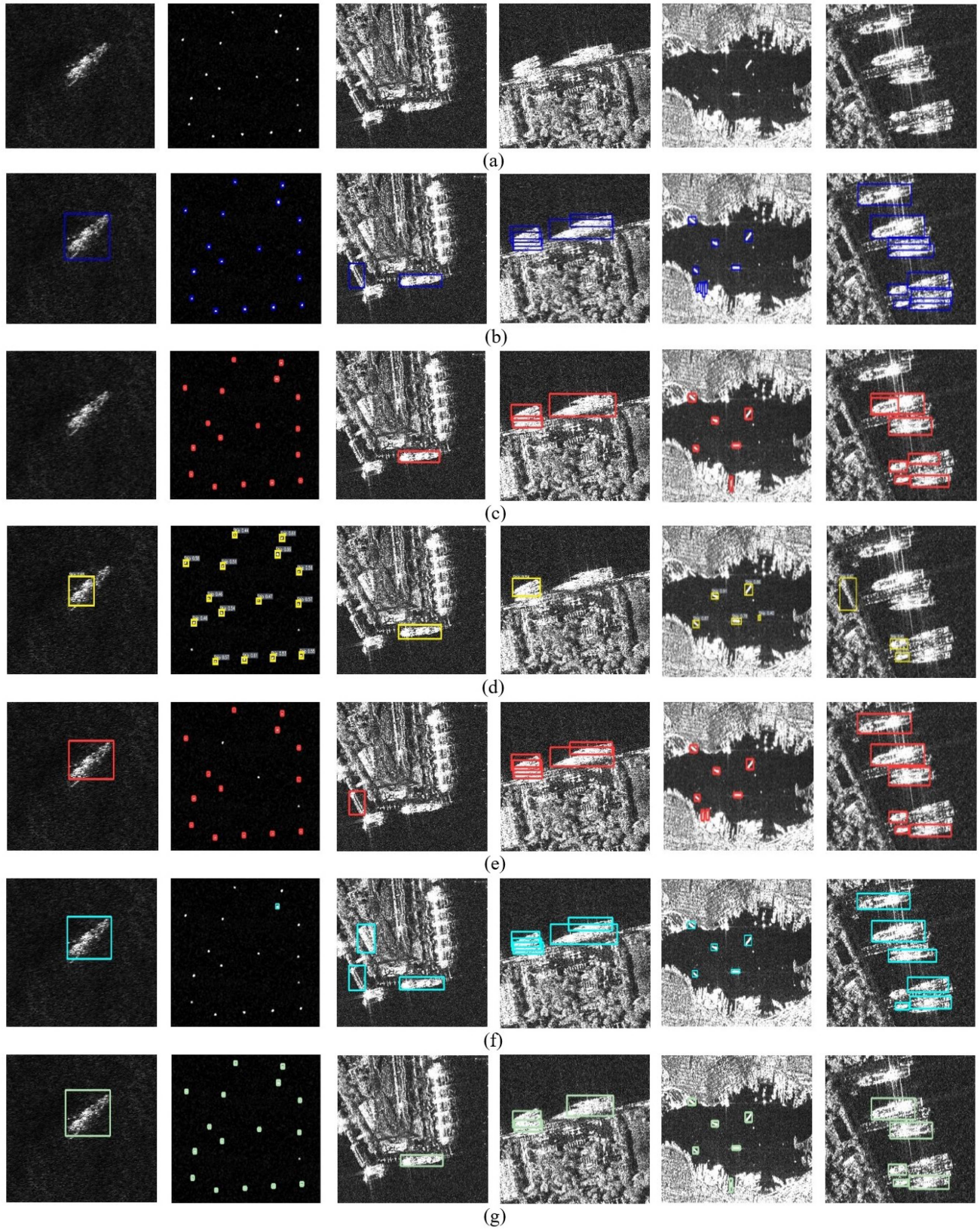


Fig. 17. Detection Results on SSDD. (a) Raw images. (b) Ground truths. (c) Yolo5 detections. (d) Yolo7 detections. (e) Yolo8 detections. (f) RetinaNet detections. (g) Yolo-OSD detections.

TABLE IX
CROSS VALIDATION RESULTS OF YOLO5, 7, 8 AND YOLO-OSD MODELS

	Model Trained On	Validation Set	Precision	Recall	F1-Score	mAP@0.5	mAP@0.95	Inference Time (mSec)
YOLO5	SAR-Ships	SSDD	0.827	0.63	0.715	0.697	0.337	8.1
		MRSDD	0.709	0.493	0.582	0.584	0.251	9.4
	SSDD	SAR-Ships	0.715	0.506	0.593	0.547	0.209	2.9
		MRSDD	0.725	0.304	0.428	0.374	0.167	9.4
	MRSSD	SAR-Ships	0.828	0.697	0.757	0.756	0.312	2.9
		SSDD	0.853	0.794	0.822	0.869	0.442	8.2
YOLO7	SAR-Ships	SSDD	0.858	0.69	0.765	0.803	0.399	7.1
		MRSDD	0.704	0.55	0.618	0.628	0.287	6.4
	SSDD	SAR-Ships	0.635	0.512	0.567	0.515	0.156	2.9
		MRSDD	0.63	0.412	0.498	0.438	0.165	6.4
	MRSSD	SAR-Ships	0.811	0.749	0.779	0.8	0.328	2.9
		SSDD	0.871	0.726	0.792	0.824	0.383	6.9
YOLO8	SAR-Ships	SSDD	0.859	0.579	0.692	0.677	0.328	10.8
		MRSDD	0.715	0.406	0.518	0.465	0.212	12.3
	SSDD	SAR-Ships	0.681	0.456	0.546	0.533	0.207	3.8
		MRSDD	0.614	0.223	0.327	0.283	0.134	12.4
	MRSSD	SAR-Ships	0.809	0.705	0.753	0.76	0.32	3.8
		SSDD	0.821	0.834	0.827	0.873	0.468	10.3
YOLO-OSD	SAR-Ships	SSDD	0.803	0.67	0.730	0.734	0.36	8.1
		MRSDD	0.731	0.436	0.546	0.509	0.232	8.7
	SSDD	SAR-Ships	0.722	0.508	0.596	0.541	0.2	2.6
		MRSDD	0.72	0.299	0.423	0.355	0.153	8.7
	MRSSD	SAR-Ships	0.831	0.704	0.762	0.759	0.313	2.6
		SSDD	0.856	0.813	0.834	0.862	0.433	7.8

Bold values represent the cross validation results achieved when, models are trained on MRSSD dataset and evaluated on SAR Ships and SSDD datasets respectively.

VI. DISCUSSION

From the different experiments described in the previous section, it can be deduced that the YOLO-OSD has better detection results on all the three datasets as compared with YOLO5, YOLO7, and RetinaNet, whereas it has significantly improved training times as compared to all other models. Both data-centric and model-centric approaches contribute to the improvement of results. The data-centric approach involves the customization of initial anchor boxes based on the statistical analysis of the data distributions, whereas, the model-centric approach focuses on the optimization of YOLO5 backbone by replacing the C3 blocks with C3x block and reducing the overall number of repetitive blocks. This change in architecture reduced the number of layers as well as the overall network parameters while facilitating the learning of small targets by increasing the size of intermediate feature maps.

In terms of detection accuracy, YOLO8 has better performance as it is an anchor free algorithm but has significantly large training times. Overall, the RetinaNet model has the lowest performance in terms of almost all evaluation metrics. It also has the highest training times on all datasets. This is due to the fact that RetinaNet has a deeper and complex network architecture based on ResNet and utilizes two subnetworks for classification and regression as compared with the highly optimized YOLO architectures. Due to its complex and large backbone, RetinaNet requires more computational resources and training time to achieve results comparable to the YOLO models. For the sake of fair comparison and bench-marking, we had set the number of training epochs for each model to be 150 in our experiments and RetinaNet needs significantly large number of training epochs to produce better results. Also the inference speed of RetinaNet is less as compared with the YOLO models. Therefore, it can

be suggested that it is not suitable for real-time ship detection in SAR images. Furthermore, it can be seen from Figs. 13, 15, and 17 that RetinaNet is not very robust against small targets and hence has a significantly high number of false negatives for small ships. This also explains the remarkably low recall values of RetinaNet as compared with the other models. The reason for high false negatives in case of RetinaNet is because it uses a focal loss function, which is designed to address the class imbalance problem by down-weighting the easy examples. However, this also potentially, reduces the sensitivity of the model to the hard examples, such as small ships in our case. thus leading to low recall values. Apart from that, it is evident from Tables V–VII that the pretrained versions of all models (on MS COCO dataset) especially YOLO7, and YOLO8 perform slightly better as compared with the models, which are trained from scratch. This hints that certain features from the MS COCO dataset are still relevant for the task of ship detection in SAR satellite images and the initial weights trained on MS COCO dataset somewhat slightly facilitate in better learning of the ships in SAR datasets.

Moreover, the cross-validation analysis of all the models on different datasets show that the newly proposed iVision-MRSSD dataset can provide SOTA DL models with good generalization capability due to its rich diversity in terms of satellite sensors, spatial resolutions, imaging frequencies, and different scene types. Also our proposed YOLO-OSD has a slightly better capability of learning generalized features when trained on a complex and diverse dataset, such as iVision-MRSSD.

The reason for choosing the three latest models of the YOLO series instead of other object detection models to carry out comparative analysis is that it has already been established in previous studies that the YOLO models outperform others on the task of ship detection in SAR satellite images [30]. The authors in [32] and [41] compared the performance of YOLO3 and YOLO4 with Faster-RCNN, SSD512 and MobileNetV3 on different SAR ship datasets and showed that the YOLO4 outperforms other models in terms of the evaluation metrics. We also trained the RetinaNet model using the Detectron2 library and as expected, it under-performs on almost all evaluation metrics when compared with the YOLO models. Moreover, the decision for further optimization of YOLO5 model among other YOLO models was taken because during the initial experiments, it was evaluated that the YOLO5 due to its simpler and relatively lightweight architecture had the least training times as compared with YOLO7 and YOLO8, which is also evident from the results summarized in Tables V–VII. Therefore, it was decided to further optimize the YOLO5 model. The decision was taken to support the case of real-time SAR ship detection without compromising the detection accuracy. Furthermore, since YOLO8 is an anchor free model, the idea of anchor box adjustment could not be incorporated into it.

VII. CONCLUSION

Ship detection in SAR satellite images poses various challenges including distorted ship boundaries, interclass similarity issues and weak generalization capabilities of SOTA object detection algorithms. To rectify some of these challenges, we

proposed an optimized ship detection model named YOLO-OSD based on a hybrid data-model centric approach. The newly proposed model has approximately 16% less network parameters as compared with YOLO5 and is approximately 40% faster as compared with YOLO8 with comparable detection performance. Moreover, the proposed YOLO-OSD outperforms YOLO5, YOLO7, and RetinaNet in terms of F-1 score and mAP on all three datasets included in the study. We further analyzed the effect of transfer learning on the problem of SAR ship detection and conclude that the pretrained models on MS COCO dataset have slightly better detection performance when compared with models that are trained from scratch by assigning random weights. We also performed cross-validation of models on all datasets under observation to determine that the iVision-MRSSD dataset has rich diversity and in combination with our proposed YOLO-OSD provides better generalization capabilities over unseen data with different distributions. Our current work is limited in the sense that it analyzes the results of only one anchor-free model YOLO8. In future, we would like to investigate the possibility of anchor-free optimized SAR ship detection. Furthermore, we would also like to address the case of ship detection using oriented B-Boxes, which can significantly increase the IoU and mAP scores.

REFERENCES

- [1] Q. Zhao et al., "An overview of the applications of earth observation satellite data: Impacts and future trends," *Remote Sens.*, vol. 14, no. 8, 2022, Art. no. 1863. [Online]. Available: <https://www.mdpi.com/2072-4292/14/8/1863>
- [2] K. Uehara, H. Sakanashi, H. Nosato, M. Murakawa, H. Miyamoto, and R. Nakamura, "Object detection of satellite images using multi-channel higher-order local autocorrelation," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2017, pp. 1339–1344.
- [3] J. Kang, S. Tariq, H. Oh, and S. S. Woo, "A survey of deep learning-based object detection methods and datasets for overhead imagery," *IEEE Access*, vol. 10, pp. 20118–20134, 2022.
- [4] R. Bamler, "Principles of synthetic aperture radar," *Surv. Geophys.*, vol. 21, no. 2, pp. 147–157, Mar. 2000, doi: [10.1023/A:1006790026612](https://doi.org/10.1023/A:1006790026612).
- [5] X. Hu, P. Zhang, Y. Ban, and M. Rahmehoonfar, "Gan-based SAR and optical image translation for wildfire impact assessment using multi-source remote sensing data," *Remote Sens. Environ.*, vol. 289, 2023, Art. no. 113522. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425723000731>
- [6] S. Arivazhagan, W. S. Lilly Jebarani, R. Newlin Shebiah, S. V. Ligi, P. Hareesh Kumar, and K. Anil Kumar, "Significance based ship detection from SAR imagery," in *Proc. 1st Int. Conf. Innov. Inf. Commun. Technol.*, 2019, pp. 1–5.
- [7] B. LI, X. XIE, X. WEI, and W. TANG, "Ship detection and classification from optical remote sensing images: A survey," *Chin. J. Aeronaut.*, vol. 34, no. 3, pp. 145–163, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1000936120304544>
- [8] R. Zhang, S. Li, G. Ji, X. Zhao, J. Li, and M. Pan, "Survey on deep learning-based marine object detection," *J. Adv. Transp.*, vol. 2021, Nov. 2021, Art. no. 5808206, doi: [10.1155/2021/5808206](https://doi.org/10.1155/2021/5808206).
- [9] T. Zhang et al., "Balance learning for ship detection from synthetic aperture radar remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 182, pp. 190–207, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271621002781>
- [10] K. Kseňak, P. K. Bartoš, and P. Blišťan, "Assessment of the usability of SAR and optical satellite data for monitoring spatio-temporal changes in surface water: Bodrog river case study," *Water*, vol. 14, no. 3, 2022, Art. no. 299. [Online]. Available: <https://www.mdpi.com/2073-4441/14/3/299>
- [11] G. Gao, G. Kuang, Q. Zhang, and D. Li, "Fast detecting and locating groups of targets in high-resolution SAR images," *Pattern Recognit.*, vol. 40, no. 4, pp. 1378–1384, 2007. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031320306000409>

- [12] X. Leng, K. Ji, K. Yang, and H. Zou, "A bilateral CFAR algorithm for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1536–1540, Jul. 2015.
- [13] J. Ai, X. Yang, J. Song, Z. Dong, L. Jia, and F. Zhou, "An adaptively truncated clutter-statistics-based two-parameter CFAR detector in SAR imagery," *IEEE J. Ocean. Eng.*, vol. 43, no. 1, pp. 267–279, Jan. 2018.
- [14] C. Wang, F. Bi, W. Zhang, and L. Chen, "An intensity-space domain CFAR method for ship detection in HR SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 529–533, Apr. 2017.
- [15] C. Zhang et al., "A novel full-polarization SAR image ship detector based on scattering mechanisms and wave polarization anisotropy," *ISPRS J. Photogrammetry Remote Sens.*, vol. 190, pp. 129–143, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271622001630>
- [16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [17] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," in *Proc. 3rd IAPR Asian Conf. Pattern Recognit.*, 2015, pp. 730–734.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [19] 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.*, 2014, pp. 580–587.
- [20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2015. [Online]. Available: <https://dblp.org/rec/journals/corr/SimonyanZ14a.html>
- [21] 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.*, 2016, pp. 779–788.
- [22] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.
- [23] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.
- [24] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
- [25] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.
- [26] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9626–9635.
- [27] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-Yolov4: Scaling cross stage partial network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13024–13033.
- [28] J. Wang and X. Hu, "Convolutional neural networks with gated recurrent connections," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3421–3435, Jul. 2022.
- [29] W. Fan, F. Zhou, X. Bai, M. Tao, and T. Tian, "Ship detection using deep convolutional neural networks for PolSAR images," *Remote Sens.*, vol. 11, no. 23, 2019, Art. no. 2862. [Online]. Available: <https://www.mdpi.com/2072-4292/11/23/2862>
- [30] Y.-L. Chang, A. Anagaw, L. Chang, Y. C. Wang, C.-Y. Hsiao, and W.-H. Lee, "Ship detection based on Yolov2 for SAR imagery," *Remote Sens.*, vol. 11, no. 7, Art. no. 786, 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/7/786>
- [31] G. Tang, Y. Zhuge, C. Claramunt, and S. Men, "N-Yolo: A SAR ship detection using noise-classifying and complete-target extraction," *Remote Sens.*, vol. 13, no. 5, 2021, Art. no. 871. [Online]. Available: <https://www.mdpi.com/2072-4292/13/5/871>
- [32] Z. Wu, B. Hou, B. Ren, Z. Ren, S. Wang, and L. Jiao, "A deep detection network based on interaction of instance segmentation and object detection for SAR images," *Remote Sens.*, vol. 13, no. 13, 2021, Art. no. 2582. [Online]. Available: <https://www.mdpi.com/2072-4292/13/13/2582>
- [33] Y. Wang, C. Wang, and H. Zhang, "Combining single shot multibox detector with transfer learning for ship detection using sentinel-1 images," in *Proc. SAR Big Data Era: Models, Methods Appl.*, 2017, pp. 1–4.
- [34] X. Xu, X. Zhang, and T. Zhang, "Lite-Yolov5: A lightweight deep learning detector for on-board ship detection in large-scene Sentinel-1 SAR images," *Remote Sens.*, vol. 14, no. 4, 2022, Art. no. 1018. [Online]. Available: <https://www.mdpi.com/2072-4292/14/4/1018>
- [35] X. Yang, X. Zhang, N. Wang, and X. Gao, "A robust one-stage detector for multiscale ship detection with complex background in massive SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5217712.
- [36] Z. Cui, Q. Li, Z. Cao, and N. Liu, "Dense attention pyramid networks for multi-scale ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8983–8997, Nov. 2019.
- [37] J. Li, J. Chen, P. Cheng, Z. Yu, L. Yu, and C. Chi, "A survey on deep-learning-based real-time SAR ship detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3218–3247, 2023.
- [38] J. Li, C. Xu, H. Su, L. Gao, and T. Wang, "Deep learning for SAR ship detection: Past, present and future," *Remote Sens.*, vol. 14, no. 11, 2022, Art. no. 2712. [Online]. Available: <https://www.mdpi.com/2072-4292/14/11/2712>
- [39] 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, 2022, Art. no. 7088. [Online]. Available: <https://www.mdpi.com/1424-8220/22/18/7088>
- [40] G. Yan, Z. Chen, Y. Wang, Y. Cai, and S. Shuai, "LSSDET: A lightweight deep learning detector for SAR ship detection in high-resolution sar images," *Remote Sens.*, vol. 14, no. 20, 2022, Art. no. 5148. [Online]. Available: <https://www.mdpi.com/2072-4292/14/20/5148>
- [41] X. Yang, J. Zhang, C. Chen, and D. Yang, "An efficient and lightweight CNN model with soft quantification for ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5230713.
- [42] T. Zhang and X. Zhang, "ShipdeNet-20: An only 20 convolution layers and < 1-MB lightweight SAR ship detector," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 7, pp. 1234–1238, Jul. 2021.
- [43] G. Yang et al., "Algorithm/hardware codesign for real-time on-satellite CNN-based ship detection in SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5226018.
- [44] G. Gao, L. Yao, W. Li, L. Zhang, and M. Zhang, "Onboard information fusion for multisatellite collaborative observation: Summary, challenges, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 11, no. 2, pp. 40–59, Jun. 2023.
- [45] S. Zhao, Y. Luo, T. Zhang, W. Guo, and Z. Zhang, "A domain specific knowledge extraction transformer method for multisource satellite-borne SAR images ship detection," *ISPRS J. Photogrammetry Remote Sens.*, vol. 198, pp. 16–29, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271623000515>
- [46] J. Ai, R. Tian, Q. Luo, J. Jin, and B. Tang, "Multi-scale rotation-invariant haar-like feature integrated CNN-based ship detection algorithm of multiple-target environment in SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10070–10087, Dec. 2019.
- [47] 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.
- [48] K. Zhou, M. Zhang, H. Wang, and J. Tan, "Ship detection in SAR images based on multi-scale feature extraction and adaptive feature fusion," *Remote Sens.*, vol. 14, no. 3, 2022, Art. no. 755. [Online]. Available: <https://www.mdpi.com/2072-4292/14/3/755>
- [49] S. Wang et al., "Yolo-SD: Small ship detection in SAR images by multi-scale convolution and feature transformer module," *Remote Sens.*, vol. 14, no. 20, 2022, Art. no. 5268. [Online]. Available: <https://www.mdpi.com/2072-4292/14/20/5268>
- [50] G. Gao, Q. Bai, C. Zhang, L. Zhang, and L. Yao, "Dualistic cascade convolutional neural network dedicated to fully PolSAR image ship detection," *ISPRS J. Photogrammetry Remote Sens.*, vol. 202, pp. 663–681, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271623001879>
- [51] 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.
- [52] Y. Feng et al., "A lightweight position-enhanced anchor-free algorithm for SAR ship detection," *Remote Sens.*, vol. 14, no. 8, 2022, Art. no. 1908. [Online]. Available: <https://www.mdpi.com/2072-4292/14/8/1908>
- [53] W. Bao, M. Huang, Y. Zhang, Y. Xu, X. Liu, and X. Xiang, "Boosting ship detection in SAR images with complementary pretraining techniques," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8941–8954, 2021.
- [54] G. Gao, Y. Dai, X. Zhang, D. Duan, and F. Guo, "ADCG: A cross-modality domain transfer learning method for synthetic aperture radar in ship automatic target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5109114.

- [55] T. Zhang et al., “SAR ship detection dataset (SSDD): Official release and comprehensive data analysis,” *Remote Sens.*, vol. 13, no. 18, 2021, Art. no. 3690. [Online]. Available: <https://www.mdpi.com/2072-4292/13/18/3690>
- [56] 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, 2019, Art. no. 765.
- [57] S. Xian, W. Zhirui1, S. Yuanrui, D. Wenhui1, Z. Yue, and F. Kun, “Air-sarship-1.0: High-resolution SAR ship detection dataset,” *J. Radars*, vol. 8, no. 6, pp. 852–862, 2019.
- [58] 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.
- [59] T. Zhang et al., “LS-SSDD-V1.0: A deep learning dataset dedicated to small ship detection from large-scale sentinel-1 SAR images,” *Remote Sens.*, vol. 12, no. 18, 2020, Art. no. 2997. [Online]. Available: <https://www.mdpi.com/2072-4292/12/18/2997>
- [60] S. Lei, D. Lu, X. Qiu, and C. Ding, “SRSSDD-V1.0: A high-resolution SAR rotation ship detection dataset,” *Remote Sens.*, vol. 13, no. 24, 2021, Art. no. 5104. [Online]. Available: <https://www.mdpi.com/2072-4292/13/24/5104>
- [61] M. F. Humayun, “iVision-MRSSD: Multi-resolution satellite based SAR ship detection dataset,” 2023, doi: [10.7910/DVN/ONUWST](https://doi.org/10.7910/DVN/ONUWST).
- [62] M. F. Humayun, F. A. Bhatti, and K. Khurshid, “iVision MRSSD: A comprehensive multi-resolution SAR ship detection dataset for state of the art satellite based maritime surveillance applications,” *Data Brief*, vol. 50, 2023, Art. no. 109505. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352340923006054>
- [63] 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*, 2020, pp. 1571–1580.
- [64] K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, “PANet: Few-shot image semantic segmentation with prototype alignment,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9196–9205.
- [65] R. Hesse, S. Schaub-Meyer, and S. Roth, “Content-adaptive downsampling in convolutional neural networks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2023, pp. 4544–4553.
- [66] S.-Y. Lo, H.-M. Hang, S.-W. Chan, and J.-J. Lin, “Efficient dense modules of asymmetric convolution for real-time semantic segmentation,” in *Proc. ACM Multimedia Asia*, 2020, pp. 1–6, doi: [10.1145/3338533.3366558](https://doi.org/10.1145/3338533.3366558).
- [67] W. Xing, Z. Cui, and J. Qi, “HRCTNet: A hybrid network with high-resolution representation for object detection in UAV image,” *Complex Intell. Syst.*, vol. 9, pp. 6437–6457, May 2023, doi: [10.1007/s40747-023-01076-6](https://doi.org/10.1007/s40747-023-01076-6).
- [68] T.-Y. Lin et al., “Microsoft coco: Common objects in context,” in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.



Muhammad Farhan Humayun received the bachelor’s degree in information and communication systems engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2014, and the double master’s degree in Project Management from Bahria University, Islamabad, and in computer science with specialization in computer vision/deep learning from the Department of Electrical Engineering and Computer Science, Institute of Space Technology, Islamabad, in 2018.

He is currently working as a Machine Learning and Automation Engineer with the National Space Agency, Islamabad. His research interests include computer vision, deep learning, optimization problems, satellite remote sensing, applications of SAR imagery and autonomous and cognitive systems.

Mr. Humayun’s honors and awards include the National Innovation Award of the year 2014 at the National Invention for Innovation Summit, Pakistan, High Achiever’s Award from Rector NUST, and Team Performance Award for Best Project.



Faryal Aurooj Nasir (Student Member, IEEE) received the bachelor’s degree in avionics engineering from the College of Aeronautical Engineering, National University of Sciences and Technology, Risalpur, Pakistan, in 2010, master’s degree in artificial intelligence from the Capital University of Science and Technology, Islamabad, Pakistan, in 2022. He is currently working toward the Ph.D. degree in computer vision with the Institute of Space Technology, Islamabad.

She is a Researcher and has done multiple projects with the public sector in domains of radar signal processing and avionics system design and product development. Her research interests include pattern recognition, satellite remote sensing, and optimization of deep learning models.



Farrukh Aziz Bhatti received the B.E. and M.S. degrees in avionics engineering and electrical engineering from the National University of Sciences and Technology, Islamabad, Pakistan, in 2004 and 2008, respectively, and the Ph.D. degree in electrical and electronic engineering from The University of Auckland, Auckland, New Zealand, in 2014.

From 2014–23, he was an Assistant Professor with the Department of Electrical Engineering, Institute of Space Technology, Islamabad. He also worked on an EU (H2020) project as a Postdoctoral Research Fellow with CONNECT Centre for Future Networks, Trinity College Dublin, Dublin, Ireland (2017–18). Currently he is the Co-founder and CEO of NewVative, a funded startup that develops AI and IoT-based cold chain intelligence solutions for serving healthcare and food industries. His research interests include the application of machine learning to remote sensing, biosignal processing, and computer vision as well as IoT platform development for commercial use cases.



Madiha Tahir received the Ph.D. degree with specialization in natural language processing and algorithm optimization from the Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Swabi, Pakistan, in 2021.

She is currently working as Assistant Professor with the Electrical Engineering and Computer Science Department, Institute of Space Technology, Islamabad, Pakistan. Her research interests include computer vision, evolutionary computing, algorithm design, data analysis, deep learning, and NLP.



Khurram Khurshid received the Ph.D. degree in computer vision from Paris Descartes University, Paris, France, in 2009.

is currently serving as a Full Professor and Head of Department (Electrical Engineering and Computer Science) with the Institute of Space Technology (IST). He is also the Founder and Director of the Artificial Intelligence and Computer Vision (iVision) research lab at IST. He is also one of the Founding members of Pakistan Pattern Recognition Society, which is an official chapter of IAPR in Pakistan.

In addition, he is also the Co-PI of the Satellite Research and Technology Lab under the National Center of GIS and Space Applications. His research interests include machine vision, image/video processing and analysis, pattern recognition, and satellite remote sensing.

He is an Associate Editor of *Journal of Space Technology*, which is an HEC recognized annual peer-reviewed journal. He has also contributed as Reviewer and Committee Member for various top ranked international journals and conferences and recipient of various national and international research grants. He was the recipient of the “Best University Teacher” award in Pakistan, Jan. 2016 by the Higher Education Commission.