scientific reports

Check for updates

OPEN

Ship detection using ensemble deep learning techniques from synthetic aperture radar imagery

Himanshu Gupta[1], Om Prakash Verma[2], Tarun Kumar Sharma[3], Hirdesh Varshney[4], Saurabh Agarwal[5]✉ & Wooguil Pak[5]✉

Synthetic Aperture Radar (SAR) integrated with deep learning has been widely used in several military and civilian applications, such as border patrolling, to monitor and regulate the movement of people and goods across land, air, and maritime borders. Amongst these, maritime borders confront different threats and challenges. Therefore, SAR-based ship detection becomes essential for naval surveillance in marine traffic management, oil spill detection, illegal fishing, and maritime piracy. However, the model becomes insensitive to small ships due to the wide-scale variance and uneven distribution of ship sizes in SAR images. This increases the difficulties associated with ship recognition, which triggers several false alarms. To effectively address these difficulties, the present work proposes an ensemble model (eYOLO) based on YOLOv4 and YOLOv5. The model utilizes a weighted box fusion technique to fuse the outputs of YOLOv4 and YOLOv5. Also, a generalized intersection over union loss has been adopted in eYOLO which ensures the increased generalization capability of the model with reduced scale sensitivity. The model has been developed end-to-end, and its performance has been validated against other reported results using an open-source SAR-ship dataset. The obtained results authorize the effectiveness of eYOLO in multi-scale ship detection with an F_1 *score* and *mAP* of 91.49% and 92.00%, respectively. This highlights the efficacy of eYOLO in multi-scale ship detection using SAR imagery.

Keywords Ensemble learning, Ship detection, Synthetic aperture radar (SAR), Weighted box fusion, YOLO

The recent technological advents in remote sensing, particularly with Synthetic Aperture Radar (SAR) imaging, have opened the doors of earth observations for many probable applications in both military and civilian domains[1]. Amongst these, ship detection has attracted significant interest because of its utmost importance in numerous surveillance and disaster management applications, such as maritime surveillance, military reconnaissance, fishery management, and seaborne traffic services[2]. Therefore, SAR-based ship detection has been considered one of the most lucrative applications of ocean surveillance. The shipping industry also plays a vital role in the global economy as most of the trade has been carried out by sea. Similarly, with more than 200 seaports in India, this industry has also contributed significantly toward sustainable development and economic well-being. However, these extensive coastlines also present some key challenges, such as effective border patrolling, traffic management, illegal migrations, etc.[3]. Therefore, efficient and robust ship detection has been the most demanding application field for SAR. Further, it can also help the authorities for overall navigation safety by providing useful information regarding pirates and traffickers.

Usually, with the availability of SAR-enabled satellites such as Gaofen-3, COSMO-SkyMed, ALOS, Sentinel-1, Kompsat-5, etc., a large number of ship-SAR datasets have been introduced[4]. However, only Gaofen-3 and Sentinel-1 have been popularly employed for maritime applications. Further, though both operate in C-band, their images vary in resolution, imaging mode, polarization, incidence angle, and background. The Sentinel-1 has four acquisition modes, whereas Gaofen-3 has 12 imaging modes that have been further classified into six groups. A detailed comparison between the SAR payload characteristics of Gaofen-3 and Sentinel-1 has been

[1]Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, Karnataka, India. [2]Department of Instrumentation and Control Engineering, Dr B R Ambedkar National Institute of Technology Jalandhar, Jalandhar 144027, Punjab, India. [3]Department of Computer Science and Engineering, Shobhit University Gangoh, Saharanpur 247341, Uttar Pradesh, India. [4]Department of Computer Science and Engineering, Babu Banarasi Das University, Lucknow 226010, Uttar Pradesh, India. [5]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea. ✉email: saurabh@yu.ac.kr; wooguilpak@yu.ac.kr

presented in Table 1. It reveals that Gaofen-3 offers high-resolution imaging, whereas Sentinel-1 provides broad area coverage. Therefore, Gaofen-3 has been found suitable for detailed local monitoring, whereas Sentinel-1 is for continuous environmental monitoring. Moreover, the open data policy of Sentinel-1 significantly enhances its accessibility and utility in various applications worldwide.

Traditionally, ship detection using SAR imagery has been accomplished by employing the Constant False Alarm Rate (CFAR) and generalized likelihood ratio test[5,6]. These frameworks detect ships by identifying the handcrafted features after segmenting the background (sea and land). However, these models greatly struggle to achieve acceptable accuracy, particularly in the inshore area of SAR images, because of the artificially designed features for identification[7]. On the contrary, the advancements in neural networks, especially Convolutional Neural Networks (CNNs), established their dominance in object detection[8]. Further, with the availability of large data, recently significant progress has been made in the deep learning (DL) based object detection frameworks for remote sensing images[9–11]. These modern DL models employ deep CNNs to automatically extract the vital discriminative features and become the primary choice for ship detection.

Earlier, the DL models have been employed to match the features of the cropped patch with the target object[12]. However, these methods find difficulties in handling the large geometric variation of ships. Therefore, to overcome this problem, the literature reveals many prominent works based on two-stage detectors, such as Region-based CNN (RCNN) and its family[13]. These models mostly employ feature fusion techniques before the region proposal network (RPN) to develop an end-to-end ship detection framework[14,15]. Though these frameworks achieved good accuracy, they require higher computation time. Therefore, in later literature, single-stage detectors, primarily You Only Look Once (YOLO), have been employed for object detection tasks[16]. Motivated by this, many versions and modifications have been developed and utilized for ship detection in SAR imagery. For example, the ship has been detected from low-resolution wide-band SAR images using YOLOv2[17] and YOLOv3[18]. The Hybrid YOLO model has been developed for ship detection from open-source SAR images[19]. Further, the noise level classifier has been appended to distinguish images with noise levels effectively[20]. The Duplicate Bilateral YOLO (DB-YOLO) has been evolved to detect multi-scale ships[21]. More recently, the SSS-YOLO model has been developed to detect small ships in SAR images[22]. These mentioned approaches utilized pre-defined anchors to localize the objects, whereas other algorithms based upon anchor-free mechanisms also exist and have been employed for ship detection[23].

The above-mentioned DL approaches consider only one detector at a time to recognize the ship in the image. These models generate their final prediction using the Non-Maximum Suppression (NMS) method. The NMS retains only that predicted box that has the maximum score and eliminates all other redundant boxes based on the predefined threshold[24]. This approach works efficiently in most cases but often results in high missed detection with objects near each other. Consequently, soft-NMS has been proposed to improve the final predictions[25]. Nevertheless, these filtering approaches lag in producing averaged localization of predictions by combining multiple models.

Though the traditional NMS approach struggles with fused models, the literature suggests that ensemble models can uplift performance[26]. On the contrary, the Weighted Box Fusion (WBF) technique fuses the predicted boxes of all the ensemble models based on their confidence score and, thus, constructs average boxes[27]. Therefore, the WBF approach has shown tremendous capability to improve the performance of object detection models by fusing multiple detectors for natural ground images. Additionally, the performance of any detection model has been highly influenced by the methodology employed to estimate and minimize the loss, usually Intersection over Union (IoU) loss. This approach works satisfactorily in most cases but fails to optimize the loss, especially when the predicted and ground truth bounding boxes do not overlap. Therefore, the literature reveals other loss functions such as Generalized IoU ($GIoU$)[28], Distance IoU ($DIoU$)[29], and Complete IoU ($CIoU$)[30] to handle this issue for object detection using natural images effectively. $DIoU$ focuses on minimizing the distance between the

S. no.	Parameter	Gaofen-3	Sentinel-1
1.	Orbit	Sun-synchronous orbit	Sun-synchronous orbit
2.	Orbit altitude	755 km	693 km
3.	Frequency band	C-band	C-band
4.	Center frequency	5.4 GHz	5.4 GHz
5.	Peak power	1.5 kW	4.7 kW
6.	Incidence angle	10°–60°	20°–45°
7.	Antenna size	15 m × 1.5 m	12.3 m × 0.84 m
8.	Signal bandwidth	0–240 MHz	0–100 MHz
9.	Polarization	Single/Dual/Full	Single/Dual
10.	Antenna system	Waveguide slot	Waveguide crack
11.	Angle of elevation	± 20°	± 11°
12.	Imaging modes	12	4
13.	Spatial resolution	1–500 m	5–20 m
14.	Swath width	10–650 km	20–400 km
15.	Design life of S/C	8 years	7.25 years

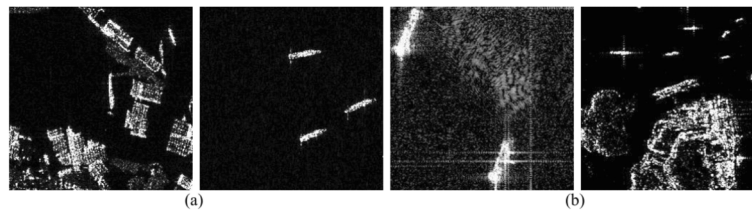**Table 1.** Comparison between SAR payload characteristics of Gaofen-3 and Sentinel-1.

Fig. 1. Sample images in SAR-ship dataset (**a**) Sentinel-1, (**b**) Gaofen-3.
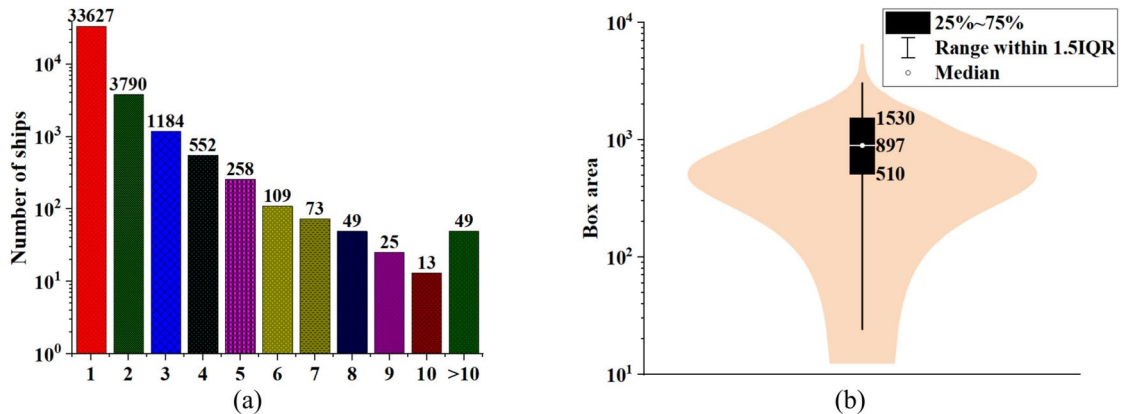

Fig. 2. Dataset (**a**) number of ships per image, (**b**) available box area.

centers of the predicted and ground-truth bounding boxes, whereas *CIoU* adds aspect ratio consideration. Both these are more effective when the bounding boxes already overlap. However, due to low object contrast and noisy backgrounds in SAR imagery, *GIoU* can be much more effective than *DIoU* and *CIoU*. Because it focuses more on the target objects, which helps maintain robustness against false object edges. Moreover, the effectiveness of WBF and *GIoU* loss functions in ship detection from SAR images has neither been exhaustively investigated nor well documented.

Based on these observations, it has been perceived that the generalization ability of DL models should be increased to develop an efficient ship detection model from SAR. Also, the limitation of *IoU* necessitates the exploration of other alternatives to optimize the losses more efficiently. Therefore, the present work focuses on the development of a highly robust generic ship detection framework with better loss optimization capabilities in a reliable manner. Summarizing, the main contribution of the present work has been listed as:

- The *GIoU* loss has been incorporated into the loss function, which reduces the scale sensitivity of the network and ensures a better multiscale feature learning capability of the model.
- Based on the modified loss function, YOLOv4 (G-v4) and YOLOv5 (G-v5) models have been developed to detect ships from SAR imagery.
- Considering the complexities associated with SAR images containing ships, an ensemble YOLO (eYOLO) framework has been proposed to enhance the generalization capability of the model. The eYOLO has been developed as an end-to-end network that fuses the G-v4 and G-v5 models via WBF.

The rest of this work has been organized as follows: section "Dataset description" describes the dataset being utilized. Section "Preliminary knowledge" briefly provides the preliminary information regarding YOLOv4 and YOLOv5 along with the loss function and employed fusion mechanism. The proposed methodology for the development of eYOLO has been presented in section "Proposed methodology". The results obtained by eYOLO and comparative analysis have been provided in section "Experiments". Finally, section "Conclusion and future scope" concludes the present work with directions for future work.

Dataset description

This study utilizes the open-source SAR-ship dataset provided by the Chinese Academy of Sciences, which contains ships of various shapes and sizes in relatively low-resolution SAR images[31]. This dataset considers 102 Gaofen-3 and 108 Sentinel-1 images to generate a total of 39,729 images with 50,885 ship appearances, and the size of all these ships has been kept at 256 pixels for both range and azimuth. The sample images from this dataset have been presented in Fig. 1. Further, the distribution of the dataset is depicted in Fig. 2, where the number of ships in each image and the bounding box area are demonstrated in Fig. 2a and b, respectively. This indicates that the employed dataset mostly has a single ship in each image chip with a median box size of 897. This median box area has been estimated to be only 1.37% of the total image size, whereas around 85.85% of

ships have less than 3.05% representation of image size. Further, the relative box size has been calculated by $\sqrt{(boxsize/imagesize)}$, which reflected that approximately 92.91% of ships have a relative size of less than 0.2. These observations reveal that the dataset has ships of various shapes and has been highly populated with small-sized ships. These small-sized ships added the difficulties associated with already complex ship detection from SAR because now the deep CNNs may miss the distinguishable attributes of ships from the background on the feature map, particularly after several rounds of sub-sampling. Additionally, similar scattering patterns have been witnessed in the presence of tides or ships near shore, which significantly uplifts the difficulties of this task.

The acquired dataset has been randomly subdivided into training, validation, and test datasets by adopting an empirically verified policy of 85:10:5. Therefore, the dataset has 33,769 images with 44,239 ship instances for training, 3993 images with 4589 ship appearances for validation, and 1967 images with 2057 ship occurrences for testing.

Preliminary knowledge

This section develops a brief concept of the two most popular versions of YOLO (v4 and v5) in sub-sections "YOLOv4" and "YOLOv5". The loss function utilized by both versions has been discussed in sub-section "Weighted box fusion", and the basic idea of WBF in sub-section "Weighted box fusion".

YOLOv4

Before the inception of YOLO, two-stage detectors have been commonly accepted. However, they failed to balance the speed-accuracy trade-off. Therefore, YOLO transforms object detection into regression to effectively mitigate this issue, which remains the core idea of all the versions of YOLO. The YOLOv4 incorporates a mosaic data augmentation technique to enhance the prediction capability of the model. Further, it optimizes and updates the backbone, incorporates advanced activation functions, and improves the training methodology to balance the speed-accuracy trade-off[32,33]. It employs CSPDarknet53 as the backbone network for feature extraction, PANet (Path Aggregation Network) to fuse the extracted attributes, and YOLOv3 as the head for object detection.

The basic structure of the ship detection model based on YOLOv4 has been presented in Fig. 3. It combines Convolution and Batch Normalization with the Leaky-ReLU activation function to develop a CBL block, whereas the MISH activation function for the CBM block. Further, Spatial Pyramid Pooling (SPP) transforms convolutional features into pooled features of the same length.

YOLOv5

YOLOv4 enhances the detection capability of object detection models, reduces the computational complexity, and can be trained with a single graphical processing unit (GPU). However, it still has a large model size because of the large number of training parameters. Therefore, to reduce the model size for lightweight characteristics, with high detection accuracy and less inference time, YOLOv5 evolved. It incorporates a focus module and redefines the CSP module in the existing backbone of YOLOv4. The focus module slices the image to retain more information with a reduced computational burden. It also has two CSP modules (CSP1_X and CSP2_X) modified from the CSPNet. The CSP module first divides the earlier layers' features into two parts, processes them, and then combines them hierarchically. The CSP1_X module has been utilized in the backbone, whereas CSP2_X in the neck to strengthen the integration capability of the network and reduce the computational complexities with improved accuracy. Besides these modules, the SPP module in the ninth layer of the backbone
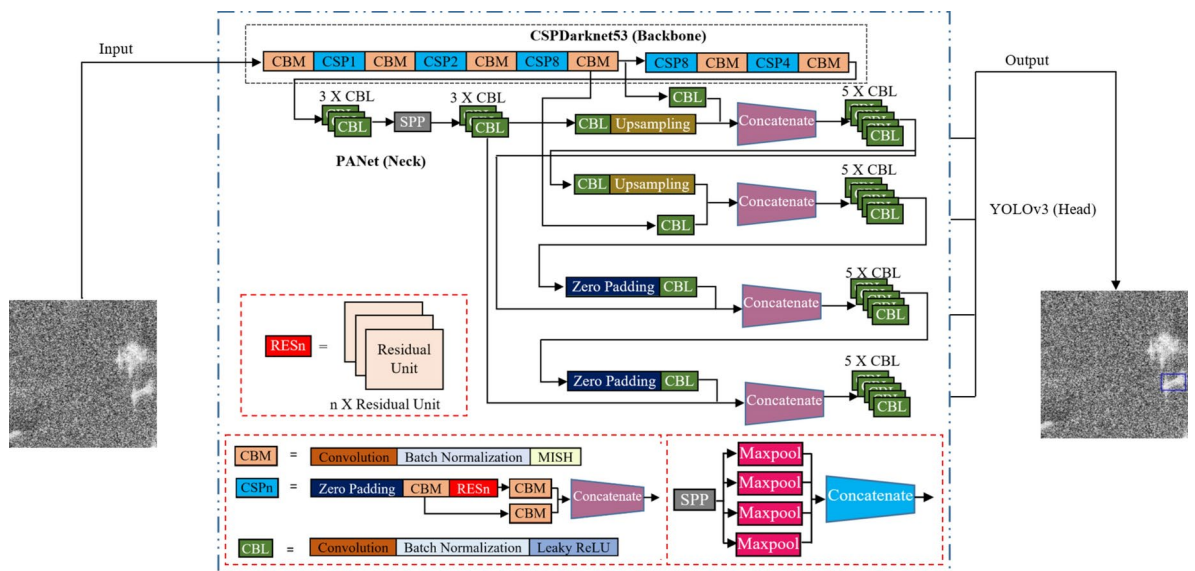

Fig. 3. YOLOv4-based ship detection model.

has been employed to generate the feature vector of the length of the fully connected layer. The basic structure of the YOLOv5 and various modules has been presented in Fig. 4.

Loss function

The loss function has been the most crucial part of any learning algorithm as it allows the model to learn the differences between the predicted and actual values. In both the studied versions of YOLO, the loss function has been typically defined by the summation of bounding box regression loss ($loss_1$), confidence loss ($loss_2$), and classification loss ($loss_3$). The $loss_1$ has been expressed by Eq. (1).

$$loss_1 = -\sum_{m=0}^{S^2}\sum_{n=0}^{B} 1_{m,n}^{obj} M \tag{1}$$

where, S^2 and B represent the number of grids and the number of bounding boxes in each grid, respectively. M denotes the mean square error (MSE) between the estimated and actual values of bounding boxes. The confidence loss has been mathematically represented by Eq. (2).

$$loss_2 = -\sum_{m=0}^{S^2}\sum_{n=0}^{B} T_{m,n}^{obj} \left[K_m^n \log\left(C_m^n\right) + (1 - K_m^n)\log(1 - C_m^n)\right]$$
$$-\lambda_{noobj} \sum_{m=0}^{S^2}\sum_{n=0}^{B} (1 - T_{m,n}^{obj}) \left[K_m^n \log\left(C_m^n\right) + (1 - K_m^n)\log(1 - C_m^n)\right] \tag{2}$$
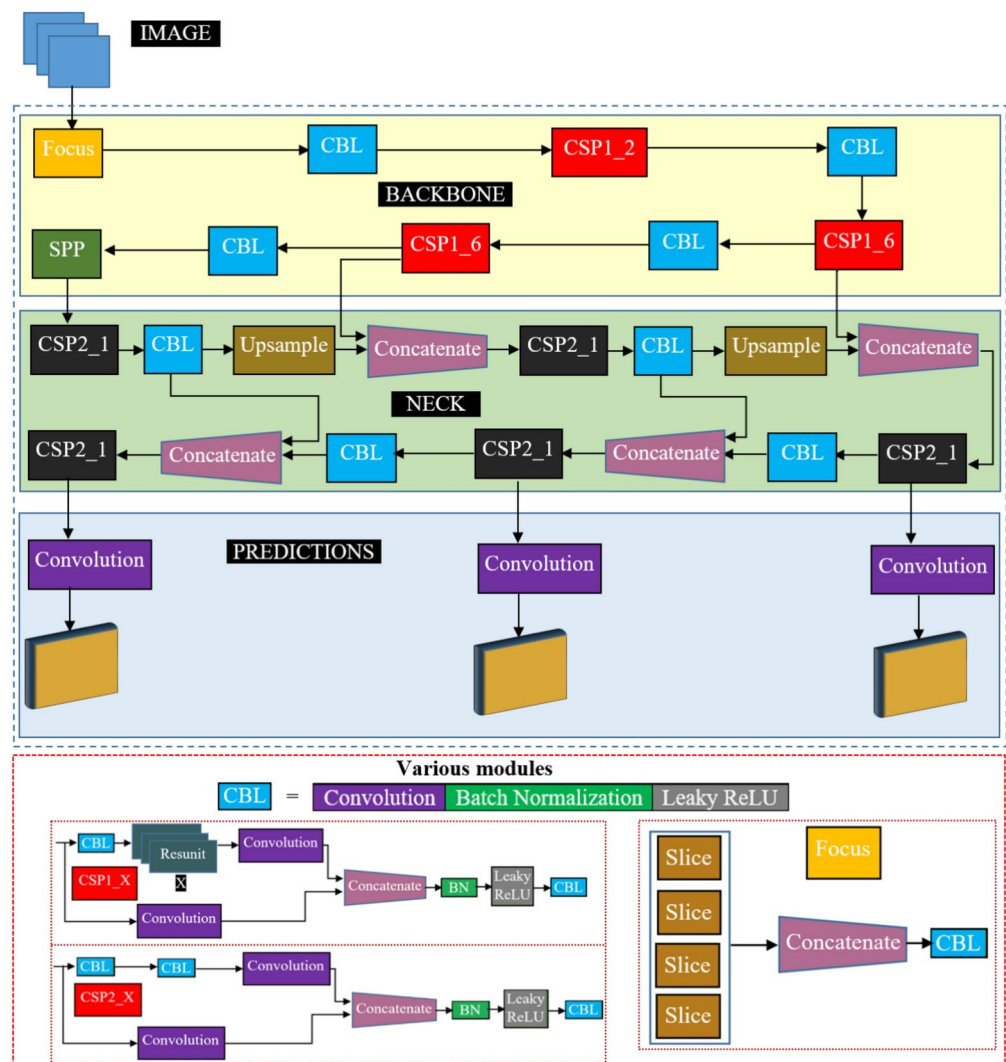

Fig. 4. The basic structure of YOLOv5 with various modules.

where, λ_{noobj} symbolizes the weight parameter. $T_{m,n}^{obj}$ represents the function of objects and $T_{m,n}^{obj} = 1$ if and only if the nth bounding box of the mth grid detects the current object. K_m^n and C_m^n have been used to represent the confidence score of predicted and target bounding boxes, respectively. The classification loss has been generally defined by Eq. (3).

$$loss_3 = -\sum_{m=0}^{S^2} \sum_{n=0}^{B} T_{m,n}^{obj} \sum_{c=1}^{C} [E_m^n(c) \log(F_m^n(c)) - (1 - E_m^n(c)) \log(1 - F_m^n(c))] \tag{3}$$

where, $E_m^n(c)$ and $F_m^n(c)$ denote the prediction and actual probability that the object belongs to the true category. Therefore, the final loss function has been defined as in Eq. (4).

$$Loss = loss_1 + loss_2 + loss_3 \tag{4}$$

Weighted box fusion

As discussed earlier, the NMS and soft-NMS provide acceptable results but have often encountered difficulties, especially for the ensemble approach. Therefore, techniques like non-maximum weighted (NMW) and WBF have been proposed[27]. In these techniques, the predicted bounding boxes from one or more than one model have been used to estimate the average bounding box so that a more reliable and robust detection algorithm can be developed. The NMW considers *IoU* values to build a weight matrix of predicted boxes. This method improves the drawbacks associated with NMS; however, as these weight matrices employ only *IoU*, they fail to attract general acceptability. On the other hand, WBF ensemble models by fusing multiple bounding boxes for a single object using their confidence score. Further, on every addition of the bounding box in the fused list, it recalculates the coordinates and confidence score of the fused box using Eqs. (5–7) and, therefore, performs very efficiently.

$$C = \frac{\sum_{n=1}^{N} C_n}{N} \tag{5}$$

$$X_{12} = \frac{\sum_{n=1}^{N} C_n \times X_{12,n}}{\sum_{n=1}^{N} C_n} \tag{6}$$

$$Y_{12} = \frac{\sum_{n=1}^{N} C_n \times Y_{12}}{\sum_{n=1}^{N} C_n} \tag{7}$$

where, *C* represents the confidence score. *X* and *Y* symbolize the x and y coordinates of the bounding box, respectively.

Proposed methodology

Generally, all object detection models use Intersection over Union (IoU) to estimate their accuracy. This measure the similarity between the predicted bounding boxes and the ground truth and is mathematically expressed by Eq. (8).

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{8}$$

where, *A* and *B* represent the area of the predicted and ground truth bounding boxes, respectively. The value of *IoU* lies in the range of [0, 1]. This method performs satisfactorily in most of the cases. However, it finds reliable assessment difficult, especially when the bounding boxes do not overlap or have different overlapping directions. Therefore, in this work, despite *IoU* loss (Eq. 1), Generalized *IoU* (*GIoU*) has been adopted to evaluate the bounding box regression loss[28]. The basic methodology to estimate the *GIoU* has been presented in Fig. 5, and its mathematical formulation by Eq. (9).

$$GIoU = \frac{|A \cap B|}{|A \cup B|} - \frac{|C \backslash (A \cup B)|}{|C|} \tag{9}$$

where, *A* and *B* represent two arbitrary-shaped bounding boxes. *C* depicts the smallest enclosing box, and *GIoU* lies between -1 and 1.

Therefore, if A^C represents the area of box *C*, then the box regression loss can be framed by Eq. (10), which replaces the $loss_1$ in Eq. 4.
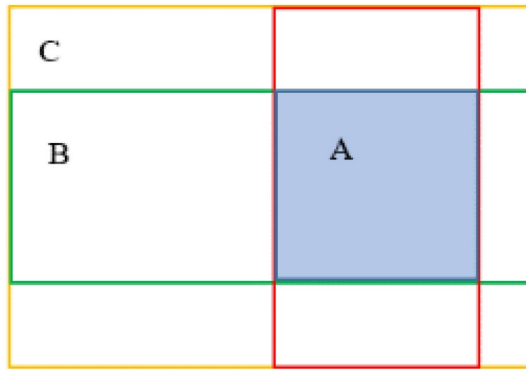
Fig. 5. Generalized Intersection over Union (*GIoU*).

$$loss_4 = \sum_{m=0}^{S^2} \sum_{n=0}^{B} 1_{m,n}^{obj} \left[1 - IoU + \frac{A^C - (A \cup B)}{A^C} \right] \tag{10}$$

The preprocessed dataset has 33,769 and 3993 images for training and validating the ship detection models. Further, data augmentation techniques such as rotation, flip, and crop have also been employed. These techniques help the model to achieve better generalization capability. Both YOLOv4 and YOLOv5 have been trained for 100 epochs with a batch size of 16 and an image size of 256×256. Once trained, the weights of these models have been saved. Then, the WBF technique has been employed to ensemble these models. Further, the performance of these models has been evaluated and compared with each other. This whole methodology has been represented in Fig. 6.

Experiments
Simulation settings
In the present work, the proposed model has been implemented by employing the PyTorch framework. For this purpose, a DL environment has been built on CUDA and cuDNN-enabled Ubuntu 18.04 LTS system having an Intel Xeon Gold processor with 48-GB QUADRO RTX-8000 GPU. Further, this work considers YOLOv5s as the baseline model. However, this model has been trained on ground-image-based datasets having entirely different viewpoints. Therefore, it cannot be directly employed to detect the object from SAR images. Hence, the baseline model has also been trained by fine-tuning the provided final weights (v5_T). Also, YOLOv5 has been trained from scratch (v5_S). Further, both YOLOv4 and YOLOv5 have been trained by employing *GIoU* loss to develop an ensemble object detection model. These ensemble models have been represented by G-v4 (*GIoU*+YOLOv4), G-v5 (*GIoU*+YOLOv5), and eYOLO (G-v4+G-v5), respectively. The image size has been fixed to 256×256 during the entire simulation work. All these models have been developed by employing the ADAM optimizer with a learning rate of 0.001, weight decay of 0.0005, and momentum of 0.937[34,35]. These models have been trained for 100 epochs, and the batch size has been set to 16.

Evaluation metrics
The performance of the developed ensemble model has been compared with the v5_T and v5_S based on mean Average Precision (*mAP*), considering it the most crucial factor in analyzing the generalization ability of any object detection model. Further, various other performance metrics such as precision (*P*), recall (*R*), and F_1 *score* have been utilized to compare the model performance with other reported works. These indicators have been computed by Eqs. (11–14). Also, as the F_1 *score* gives a higher weightage to *P*, another measure of the F_2 *score* has been introduced in Eq. (15), which gives a higher weightage to *R*.

$$mAP = \int_{0}^{1} P(R) \, dR \tag{11}$$

$$P = \frac{T_P}{T_P + F_P} \tag{12}$$

$$R = \frac{T_P}{T_P + F_N} \tag{13}$$

$$F_1 \; score = \frac{2 \times P \times R}{P + R} \tag{14}$$

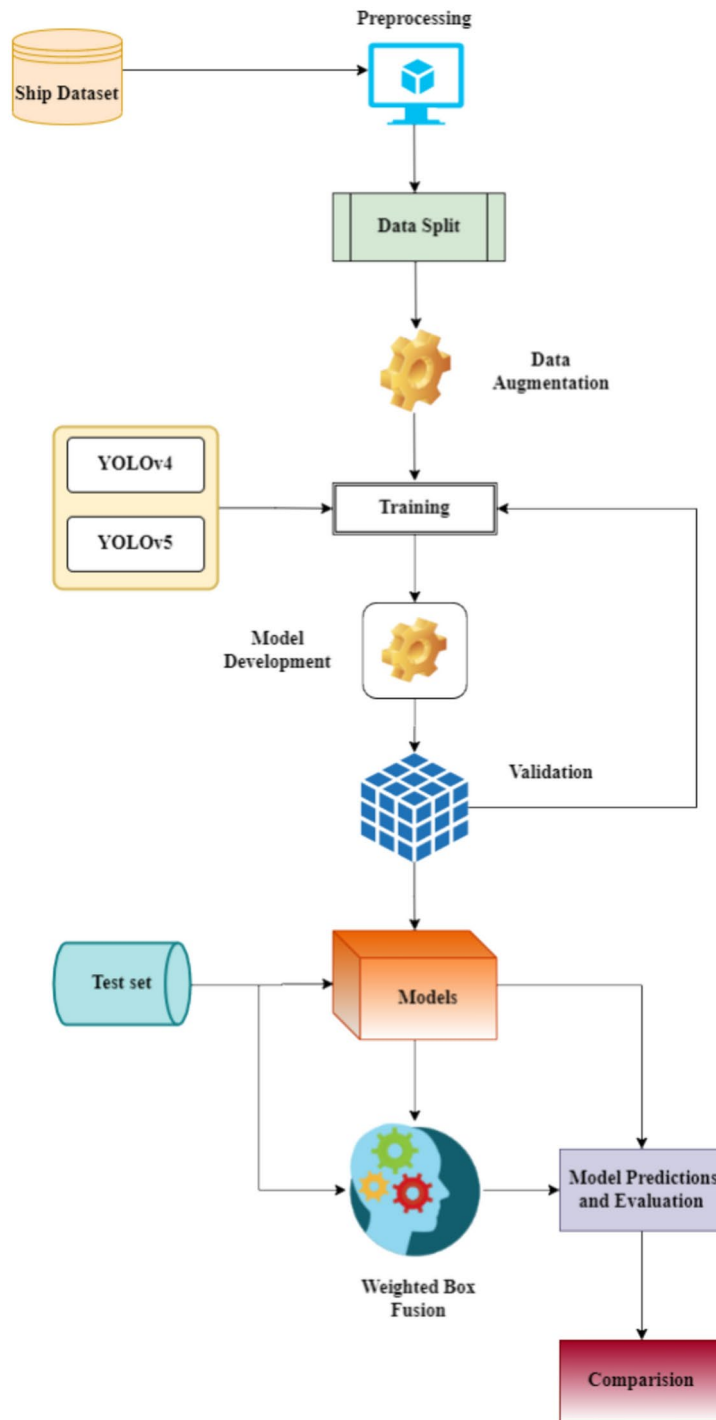$$F_2 \; score = \frac{5 \times P \times R}{4 \times P + R} \tag{15}$$

Fig. 6. Methodology for the development of eYOLO.

Result analysis

As mentioned earlier, the SAR images have significantly different characteristics. Therefore, the open-source DL object detection models cannot be directly applied. First, the YOLOv5 model with open-source pre-trained weights has been tested to review this critically. From this analysis, it has been revealed that although it achieved mAP of 0.689 on the COCO dataset, only 0.048 mAP has been attained on the employed dataset. Therefore, two models of YOLOv5 (v5_T and v5_S) have been trained with the aforementioned hyperparameters by employing transfer learning and from scratch, respectively. These models help to analyze the impact of these techniques on the model performance for an entirely different representative dataset. The performance of these models has been compared based on mAP, and it has been witnessed that v5_T achieved the mAP of 0.909. Therefore, it performs slightly better (0.44%) than v5_S on the unseen dataset. This also verifies that the transfer learning

Model	mAP@0.5
v5_P	0.048
v5_T	0.909
v5_S	0.905
G-v4	0.780
G-v5	0.912
eYOLO	0.920

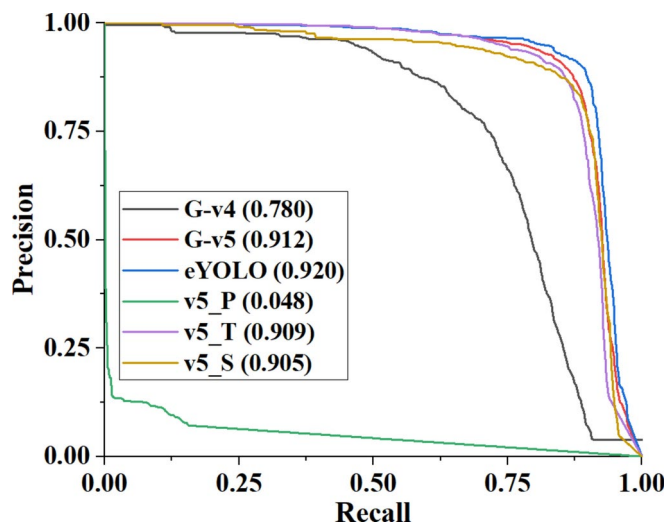**Table 2**. Performance comparison.

Fig. 7. *P-R* curve for ship detection models.

approach should be preferred over training from scratch because this approach with the pre-trained backbone has superior capabilities.

Further, two other models (G-v4 and G-v5) were trained to incorporate *GIoU* loss. These models have been trained from scratch because of the unavailability of their trustworthy weights with this loss. On comparing their performance, it has been observed that G-v5 dominates G-v4 by a comprehensive margin of 16.92% for *mAP*. Consequently, G-v5 has a more overwhelming ship detection capability than G-v4. Also, the *mAP* of G-v5 has been computed as 0.77% and 0.33% higher than v5_S and v5_T, respectively. Therefore, it has been perceived that the *GIoU* enhances the model's prediction ability and should be preferred over *IoU*. Lastly, the developed G-v4 and G-v5 models have been fused to create an eYOLO by employing the WBF technique. It has been witnessed that eYOLO attained the *mAP* of 0.920 and dominated G-v4, G-v5, v5_T, and v5_S with an encouraging margin of 17.95%, 0.88%, 1.21%, and 1.66%, respectively. Therefore, eYOLO has better generalization capability and can detect ships with higher *P* and *R*. The computed *mAP* values of all the models have been tabulated in Table 2, and their impact has been depicted by the *P-R* curve, as shown in Fig. 7.

The ships detected by these models have been visually compared against the ground-truths for test images, and a sample of these has been demonstrated in Fig. 8. The bounding boxes generated by all the models for the detected ships in each image have been compared against the ground truths, and it has been computed that eYOLO predicts better and crisper bounding boxes than all others. This also verifies that eYOLO has better detection capability because of the fusion. Occasionally, eYOLO also detects some false ships (Image 4); however, considering the complexities associated with SAR, these may be neglected because the number of these false detections has been computed very little.

The developed models have also been compared for qualitative analysis based on the false detection rate (FDR) and missed detection rate (MDR). The values of these rates for all the developed models have been mentioned in Table 3. Based on FDR, the eYOLO outperforms v5_T and v5_S by a significant margin of 35.16% and 54.70%, respectively. Similarly, it outclasses G-v4 and G-v5 for the same parameter by 160.49% and 25.2%, respectively. For MDR, the developed eYOLO also dominates others by a minimum margin of 28.42%. Therefore, the developed eYOLO has been examined as more robust than others.

Comparison and validation

The results obtained by eYOLO have been compared against other reported work on the same dataset to validate the proposed model. For this purpose, a comparative table has been framed and presented in Table 4. The developed eYOLO conquered the other developed models for *P* by the minimum and maximum margins
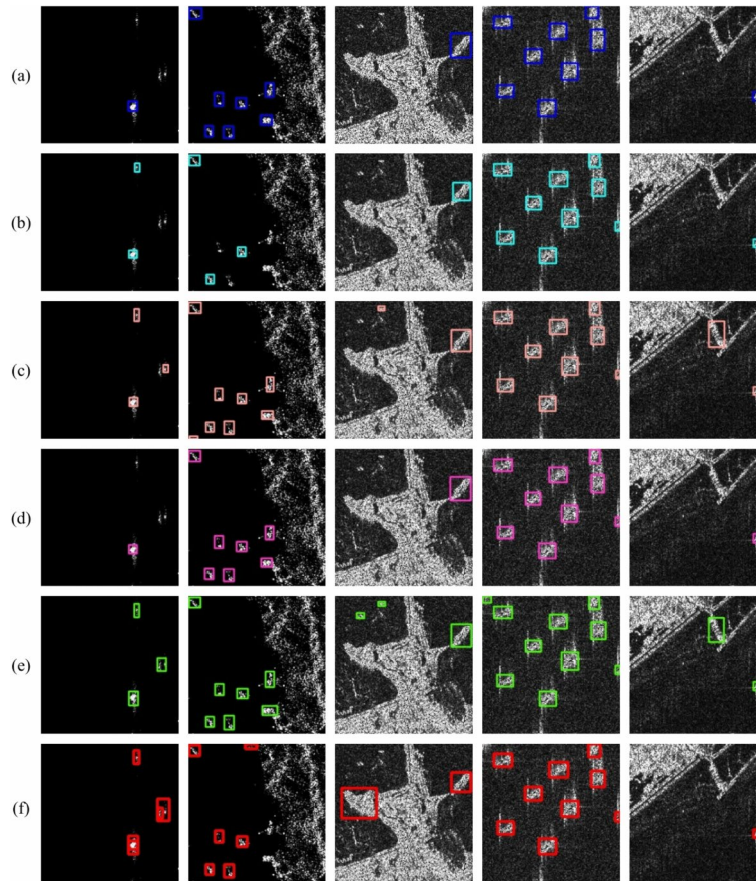
Fig. 8. Ships detected (**a**) ground truth, (**b**) G-v4, (**c**) G-v5, (**d**) eYOLO, (**e**) v5_T, (**f**) v5_S.

Model	FDR (%)	MDR (%)
v5_T	6.92	10.15
v5_S	8.32	11.62
G-v4	34.60	19.56
G-v5	6.73	9.40
eYOLO	5.24	7.51

Table 3. Qualitative analysis of developed ship detection models.

S. no.	Methodology	P (%)	R (%)	F_1score (%)	F_2score (%)	mAP (%)
1.	Faster RCNN[13]	91.00	91.00	91.00	x	91.00
2.	Libra RCNN[39]	87.80	91.40	89.60	x	91.50
3.	CR2A-Net[37]	91.70	92.20	91.90	x	90.10
4.	DAPN[40]	91.00	91.40	91.20	x	91.90
5.	SSD[29]	x	x	x	x	89.43
6.	YOLOv3[22]	x	x	x	x	91.27
7.	RetinaNet[36]	92.14	86.63	89.01	x	91.59
8.	NNAM[41]	77.6	95.4	85.5	x	89.90
9.	DCMSNN[15]	78.10	94.30	85.40	x	91.60
10.	eYOLO (ours)	92.90%	90.13%	91.49%	90.68%	92.00

Table 4. Comparative analysis with state-of-the-art models. 'x' represents the unavailability of the data.

of 0.82% and 19.72%, respectively. It also achieves a 4.28% higher value of *R* as compared with the reported literature[36] and has been found comparable with others. Based on the F_1 *score*, it outclasses most of the reported models by at least 0.32%. However, the achieved F_1 *score* has been found to be slightly lower than the value attained by CR2A-Net[37]. This may be because the hyperparameters of G-v4 can be further improved, and this model finds difficulty detecting small-size ships in complex scenarios, which has also been witnessed in the higher MDR (Table 3). On the basis of *mAP* also, the developed eYOLO significantly outperformed all the other compared models, which suggests that algorithmic enhancements, such as loss functions, optimization techniques, etc., might improve other aspects of model performance (robustness, generalization ability, etc.), with only slight improvement in *mAP* [38]. Therefore, it has been concluded that the developed eYOLO has better adaptability and generalization capability. Further, the higher values of both the F_1 *score* and F_2 *score* concealed that the developed eYOLO may be very effective for the marine industry.

Conclusion and future scope

In this work, a novel ensemble object detection framework has been proposed for efficient ship detection from SAR imagery, and the effectiveness of the model has been analyzed on the open-source SAR-Ship dataset. Firstly, the *GIoU* loss function has been incorporated to modify the loss, which has been further utilized to train G-v4 and G-v5. Then, the WBF technique has been employed to develop the proposed ensemble object detection model (eYOLO) by fusing G-v4 and G-v5. The developed eYOLO model has been compared against other developed models on the basis of *mAP*, FDR, and MDR, which establishes the superiority of eYOLO over others for this specific task. Further, the efficacy of eYOLO has been validated by comparing it with other reported works. The simulated results endorsed that the eYOLO has a better adaptability and generalization capability. Though the proposed eYOLO performed satisfactorily and has shown tremendous ability to detect ships of several scales, it still suffered from false alarms, particularly at the edges, which will be addressed in the future.

Data availability

No new dataset has been generated during the current study.

References

1. Wang, X., Li, G., Plaza, A. & He, Y. Ship detection in SAR images via enhanced nonnegative sparse locality-representation of fisher vectors. *IEEE Trans. Geosci. Remote Sens.* **59**, 9424–9438 (2021).
2. Li, L., Zhou, Z., Wang, B., Miao, L. & Zong, H. A novel CNN-based method for accurate ship detection in HR optical remote sensing images via rotated bounding box. *IEEE Trans. Geosci. Remote Sens.* **59**, 686–699 (2021).
3. Dwarakish, G. S. & Salim, A. M. Review on the role of ports in the development of a nation. *Aquat. Procedia* **4**, 295–301 (2015).
4. Zhang, T. et al. SAR ship detection dataset (SSDD): official release and comprehensive data analysis. *Remote Sens.* **13**, 3690 (2021).
5. Li, Y., Zhang, S. & Wang, W. Q. A lightweight faster R-CNN for ship detection in SAR images. *IEEE Geosci. Remote Sens. Lett.* **19**, 1–5 (2022).
6. Xie, T., Liu, M., Zhang, M., Qi, S. & Yang, J. Ship detection based on a superpixel-level CFAR detector for SAR imagery. *Int. J. Remote Sens.* **43**, 3412–3428 (2022).
7. Cui, Z., Wang, X., Liu, N., Cao, Z. & Yang, J. Ship detection in large-scale SAR images via spatial shuffle-group enhance attention. *IEEE Trans. Geosci. Remote Sens.* **59**, 379–391 (2021).
8. Kumar, S. et al. A Novel YOLOv3 algorithm-based deep learning approach for waste segregation: towards smart waste management. *Electronics* **10**, 14 (2020).
9. Gupta, H. & Verma, O. P. Monitoring and surveillance of urban road traffic using low altitude drone images: a deep learning approach. *Multimedia Tools Appl.* **2021**, 1–21. https://doi.org/10.1007/s11042-021-11146-x (2021).
10. Gupta, H., Jindal, P. & Verma, O. P. *Automatic Vehicle Detection from Satellite Images Using Deep Learning Algorithm* 551–562 (Springer, Singapore, 2021). https://doi.org/10.1007/978-981-16-1696-9_52.
11. Zhu, X. X. et al. Deep learning meets SAR: concepts, models, pitfalls, and perspectives. *IEEE Geosci. Remote Sens. Mag.* **9**, 143–172 (2021).
12. Cozzolino, D., Di Martino, G., Poggi, G. & Verdoliva, L. A fully convolutional neural network for low-complexity single-stage ship detection in Sentinel-1 SAR images. In *International Geoscience and Remote Sensing Symposium (IGARSS) 2017-July* 886–889 (2017).
13. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149 (2017).
14. Li, J., Qu, C. & Shao, J. Ship detection in SAR images based on an improved faster R-CNN. In *Proceedings of 2017 SAR in Big Data Era: Models, Methods and Applications, BIGSARDATA 2017, 2017-January* 1–6 (2017).
15. Jiao, J. et al. A densely connected end-to-end neural network for multiscale and multiscene SAR ship detection. *IEEE Access* **6**, 20881–20892 (2018).
16. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition vols 2016-Decem* 779–788 (IEEE Computer Society, 2016).
17. Chang, Y. L. et al. Ship detection based on YOLOv2 for SAR imagery. *Remote Sens.* **11**, 786 (2019).
18. Jiang, S. et al. Ship detection with sar based on Yolo. In *International Geoscience and Remote Sensing Symposium (IGARSS)* 1647–1650 (2020). https://doi.org/10.1109/IGARSS39084.2020.9324538.
19. Devadharshini, S., Kalaipriya, R., Rajmohan, R., Pavithra, M. & Ananthkumar, T. Performance investigation of hybrid YOLO-VGG16 based ship detection framework using SAR images. In *2020 International Conference on System, Computation, Automation and Networking, ICSCAN 2020.* https://doi.org/10.1109/ICSCAN49426.2020.9262440 (2020).
20. Tang, G., Zhuge, Y., Claramunt, C. & Men, S. N-YOLO: A SAR ship detection using noise-classifying and complete-target extraction. *Remote Sens.* **13**, 871 (2021).
21. Zhu, H. et al. DB-YOLO: A Duplicate Bilateral YOLO Network for Multi-Scale Ship Detection in SAR Images. *Sensors.* https://doi.org/10.3390/s21238146 (2021).
22. Wang, J., Lin, Y., Guo, J. & Zhuang, L. SSS-YOLO: towards more accurate detection for small ships in SAR image. *Remote Sens. Lett.* https://doi.org/10.1080/2150704X.2020.1837988 (2020).

23. Guo, H., Yang, X., Wang, N. & Gao, X. A CenterNet++ model for ship detection in SAR images. *Pattern Recogn.* **112**, 107787 (2021).
24. Neubeck, A. & Van Gool, L. Efficient non-maximum suppression. *Proc. Int. Conf. Pattern Recogn.* **3**, 850–855 (2006).
25. Bodla, N., Singh, B., Chellappa, R. & Davis, L. S. Soft-NMS—improving object detection with one line of code. In *Proceedings of the IEEE International Conference on Computer Vision* 5562–5570 (2017).
26. Mehbodniya, A. et al. Fetal health classification from cardiotocographic data using machine learning. *Expert Syst.* **2021**, 1–13. https://doi.org/10.1111/exsy.12899 (2021).
27. Solovyev, R., Wang, W. & Gabruseva, T. Weighted boxes fusion: ensembling boxes from different object detection models. *Image Vis. Comput.* **107**, 104117 (2021).
28. Rezatofighi, H. et al. Generalized intersection over union: a metric and a loss for bounding box regression. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 658–666 (2019).
29. Zheng, Z. et al. Distance-*IoU Loss: Faster and Better Learning for Bounding Box Regression* (Wiley, 2019).
30. Zheng, Z. et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Trans. Cybern.* https://doi.org/10.1109/TCYB.2021.3095305 (2021).
31. Wang, Y., Wang, C., Zhang, H., Dong, Y. & Wei, S. *A SAR Dataset of Ship Detection for Deep Learning under Complex Backgrounds* (Springer, 2019). https://doi.org/10.3390/rs11070765.
32. Bochkovskiy, A., Wang, C.-Y. & Liao, H.-Y. M. *YOLOv4: Optimal Speed and Accuracy of Object Detection* (Wiley, 2020).
33. Wu, D., Lv, S., Jiang, M. & Song, H. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Comput. Electron. Agric.* **178**, 105742 (2020).
34. Yu, Y., Yang, X., Li, J. & Gao, X. A cascade rotated anchor-aided detector for ship detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–14. https://doi.org/10.1109/TGRS.2020.3040273 (2022).
35. Zhao, Q., Wu, Y. & Yuan, Y. Ship target detection in optical remote sensing images based on E2YOLOX-VFL. *Remote Sens.* **16**, 340. https://doi.org/10.3390/rs16020340 (2024).
36. Gao, F., Shi, W., Wang, J., Yang, E. & Zhou, H. Enhanced feature extraction for ship detection from multi-resolution and multi-scene synthetic aperture radar (SAR) images. *Remote Sens.* **11**, 2694 (2019).
37. Yu, Y., Yang, X., Li, J. & Gao, X. A cascade rotated anchor-aided detector for ship detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–12 (2022).
38. Tian, Y., Su, D., Lauria, S. & Liu, X. Recent advances on loss functions in deep learning for computer vision. *Neurocomputing* **497**, 129–158 (2022).
39. Pang, J. et al. *Libra R-CNN: Towards Balanced Learning for Object Detection* (Springer, 2024)
40. Cui, Z., Li, Q., Cao, Z. & Liu, N. Dense attention pyramid networks for multi-scale ship detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* **57**, 8983–8997 (2019).
41. Chen, C., He, C., Hu, C., Pei, H. & Jiao, L. A deep neural network based on an attention mechanism for SAR ship detection in multiscale and complex scenarios. *IEEE Access* **7**, 104848–104863 (2019).

Author contributions

Conceptualization, H.G., O.P.V., T.K.S., H.V., S.A., and W.P.; writing—original draft preparation, H.G., O.P.V., T.K.S., H.V., S.A., and W.P.; writing—review and editing, H.G., O.P.V., T.K.S., H.V., S.A., and W.P.; Funding acquisition, W.P. All authors reviewed the manuscript.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.A. or W.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.