



OPEN

Multi-ship detection and classification with feature enhancement and lightweight fusion

Ying Han¹, Hao Wang², Nick Renjin³, Jie Song¹, Gongxiang Cui¹, Yugang Wang¹✉ & Fengyu Zhou⁴

Ship target tracking and detection are essential procedures in the shipping industry that guarantee ship traffic and marine safety. However, issues including complicated background interference, multi-scale object recognition, and inadequate training for small sample recognition are common with classical detection techniques. To address these challenges, this study proposes an enhanced ship multi-target detection model utilizing a modified YOLOv8 algorithm. In addition to integrating the ESSE module and GSConvns technology into the YOLOv8 backbone, the YOLOv8n model acts as the baseline and integrates Wise-IoU technology. This improvement greatly increased multi-scale feature extraction's efficacy while maintaining the fewest possible parameters. With this change, lightweight fusion is accomplished, and the model's capacity to extract semantic characteristics from ship photos is enhanced, particularly when it comes to identifying targets against intricate backdrops. According to testing results on the Dockship, Seaships, and Infrared Offshore Ship datasets, the enhanced algorithm's average detection accuracy is 82.1%, 99.1%, and 91.7%, respectively. This is a considerable improvement over the baseline model. Furthermore, the model's ability to improve the features, lightweighting, and detecting capabilities of different ship kinds has been validated by IoU computation and ablation experimental analysis. These results highlight how the suggested approach could improve ship target detection's automation, dependability, and quality.

Keywords Artificial intelligence, YOLOv8, Ship detection and classification, Lightweighting, Multi-scale feature

With the development of the global marine economy, both the importance of ships as the leading carriers and the demand for ship safety and monitoring have attracted increasing attention¹. Although it has made remarkable achievements in ship inspection technology, there are still many challenges in practical application². The complex background environment of the coastal area, where ships are often mixed with disturbing objects such as cargo, cranes, and buildings, makes it difficult to identify targets³. Additionally, in the far sea area, although there are fewer interferences, the size of the ship target is relatively smaller than the background environment⁴. It is affected by the environment, such as light, and it is also challenging to identify.

Traditional ship detection and classification methods predominantly rely on manual visual observation or basic image processing techniques. However, these approaches suffer from notable drawbacks, including pronounced subjectivity, limited efficiency, and suboptimal accuracy. Li et al.⁵ highlighted edge contours to improve image detection by Canny operator optimization. An enhanced adaptive Canny method was presented by Liu et al.⁶ for ship infrared image edge detection. Cui et al.⁷ discovered that threshold segmentation techniques were effective and simple, while they are not resilient and need to be adjusted for complex and dynamic situations. Klimkowska et al.⁸ studied five color Spaces and each channel in the new image, using image segmentation to detect the target object, and experimentally demonstrated the effectiveness of this method. However, it relies more on color information. The above detection methods are sensitive to noise and may be significantly affected by light and the environment. Zhou et al.⁹ tried to detect better multi-scale ship targets in SAR images, which improved

¹Naval Architecture and Port Engineering College, Shandong Jiaotong University, Weihai 264200, People's Republic of China. ²Weichai Power Co., Ltd., Weifang 261000, People's Republic of China. ³Inspur Software Group Ltd., Jinan 250013, People's Republic of China. ⁴School of Control Science and Engineering, Shandong University, Jinan 250061, People's Republic of China. ✉email: sdu201413070@gmail.com

the feature extraction capability of the model in remote sensing datasets. However, the occlusions and complex background RGB images need to be improved. The template matching method was effectively applied to find fractures on ship hulls and bilge keels by Khumaidi et al.¹⁰. However it requires a large number of templates due to its sensitivity to size, rotation, and deformation. From above, traditional methods are generally sensitive to environmental changes and have limited accuracy.

With the further development of deep learning, the Two-Stage algorithm is continuously optimized to adapt to different scenarios. Zhang et al.¹¹ proposed a target detection framework, Faster-RCNN, which provides a very effective method for ship detection in high-resolution remote sensing images. Guo et al.¹² proposed a rotating Libra R-CNN method. As a remote sensing data set, the DOTA data set was significantly different from the data set required by our experiment in terms of data characteristics and processing methods. A ship recognition and segmentation approach based on an enhanced Mask R-CNN model was proposed by Nie et al.¹³, who focused on ship detection in satellite remote sensing images. Two object-oriented ship detectors based on Faster R-CNN were introduced by Loran et al.¹⁴ in response to the restricted visibility of maritime radar. Wolridge et al.¹⁵ proposed the development of Computer Video Analysis (CVA) technology, which uses a new algorithm to track the water surface near floating structures, successfully improving the accuracy and robustness of wave interaction measurements, and proving in experiments that it can effectively evaluate the response under different wave conditions. Two-stage detection algorithms, despite their accuracy advantages, face limitations including slower detection speed due to candidate region generation, high computational resource demands hindering deployment on resource-constrained devices, poor performance with small targets like ships, and insufficient robustness in complex backgrounds such as sea surfaces, resulting in false or missed detections.

With the advancement of technology, more advanced methods are explored to overcome the limitations in ship detection and classification. An improved ImYOLOv3 was presented by Chen et al.¹⁶ to address the difficulties associated with ship detection in computer vision. Ye et al.¹⁷ proposed the EA-YOLOv4 algorithm, which incorporated a convolution block attention module (CBAM) into YOLOv4 to enhance feature perception. Additionally, it adopts an improved effective Joint Crossing (EIoU) loss function to optimize ship detection of different sizes. Yang et al.¹⁸ developed a more lightweight network architecture, EL-YOLO, which achieved significant advances in detection accuracy and lightweight in different maritime scenarios. Wang et al.¹⁹ proposed a new ALF-YOLO architecture integrating AFPN, LSK and the fourth detection head, focusing on the key features of the ship and eliminating the interference of the complex environment. In addition, Cheng et al.²⁰ proposed a lightweight model, SGST-YOLOv8, for pedestrian and vehicle detection. Chen et al.²¹ proposed CD-YOLO, a multi-scale ship detection model in complex scenes based on YOLOv7, aiming at the problems of missed detection and misidentification in synthetic aperture radar (SAR) images. SED-YOLO is an enhanced remote sensing image small target identification network developed by Li et al.²², founded on YOLOv5s. They implement an efficient multi-scale attention (EMA) mechanism at the conclusion of the network to facilitate multi-scale feature learning.

Despite advancements in YOLO-based enhancement techniques for ship target recognition in recent years, EL-YOLO exhibits constrained feature extraction skills and inadequate robustness against small targets and intricate backdrops. ALF-YOLO exhibits significant computational demands and restricted generalization capabilities. SGST-YOLOv8 is tailored for pedestrians and vehicles, yet demonstrates inadequate adaptability for maritime applications and lacks a multi-scale enhancement mechanism. CD-YOLO is specifically engineered for SAR images, but shows subpar performance with optical images and possesses high model complexity. SED-YOLO's feature fusion suffers from low efficiency and presents challenges in achieving a balance between lightweight design and accuracy. These methods remain inadequate for practical application needs including feature extraction, small object recognition, background interference suppression, model optimization, and real-time performance balance. According to relevant research, the average detection accuracy of ship targets in complex sea conditions is only 78.3%, far below the ideal level. In addition, when the proportion of ship targets in the image is less than 0.5%, the missed detection rate is as high as 42.3%.

Therefore, an improved YOLOv8 ship detection model, GEW-YOLO (GSConvns, Efficient Ship Semantic Enhancement (ESSE) module, Wise-IoU), is proposed in this paper, which integrates high efficiency and lightweight. Firstly, based on YOLOv8, Wise-IoU loss function, lightweight GSConvns module and ESSE module are used to improve the model, which realizes high-precision detection in complex environments. To verify the validity of the model, three different scenario datasets—Dockship, SeaShips, and Infrared Offshore Ship datasets, are expanded and experimented with. The results show the significant advantages of the model in complex environments and in achieving efficient lightweight. Finally, the proposed algorithm is further analyzed by comparing different improved algorithms through ablation study. The experimental results show that the proposed algorithm has high detection accuracy for Marine ship image detection, and the calculation parameters and time complexity of the model are greatly reduced.

The article is organized as follows. Dataset and the proposed method are covered in section “[Dataset and the proposed method](#)”. Experiment results and discussion are presented in section “[Experiment results and discussion](#)”. Lastly, section “[Conclusions](#)” presents the conclusions.

Dataset and the proposed method

Dataset and preprocessing

To evaluate the effectiveness of the proposed method, this section explains the three ship datasets used in the experiment. These datasets are Dockship²³, SeaShips²⁴, and Infrared Offshore Ship Datasets²⁵. These datasets provide valuable benchmarks for comparing and evaluating algorithms and are widely used in ship identification. The sections that follow give a thorough overview of each dataset. Three datasets is divided into training, testing, and validation sets in a 7:2:1 ratio.

Dockship Dataset divides ships into seven types based on their appearance, structure, function, and size, including Venetian kayaks, kayaks, buoys, sailboats, ferries, cruise ships, and origami boats. To prove practical significance, origami boats are replaced with cargo ships. The updated dataset comprises 1296 images with a resolution of 1280×850 , though not exactly uniform.

To expand the scope of the experiment, two additional datasets, SeaShips and Infrared Offshore Ship datasets, are also conducted. The SeaShips dataset comprises 7000 images with a resolution of 1920×1080 . All images are captured by surveillance cameras deployed in the coastal video surveillance system. Numerous picture variants, such as various sizes, ship sections, lighting conditions, viewpoints, backdrops, and occlusions, are included in the collection. Six categories, ore ships, general cargo ships, bulk cargo ships, container ships, fishing boats, and passenger ships. The corresponding are 4900, 1400, and 700 images, respectively. The image size was adjusted to 640×640 to match the input size required by the method. The experimental results were analyzed and evaluated on the test set.

Different from the above two datasets, the infrared marine vessel dataset mainly comes from infrared sensors, which can detect objects well while concealing themselves. However, the fewer texture features and lower contrast of images make it more difficult to detect small targets. It contains 8402 image data in different scenarios. These images are divided into seven types—cruise ships, bulk carriers, warships, sailboats, canoes, container ships, and fishing boats. The image resolutions are 384×288 , 640×512 , and 1280×1024 , respectively. There is a well-balanced number of ships of various categories in the training, testing, and validation sets.

The proposed method

The uneven distribution of ships and the prevalence of small targets at sea present significant challenges for detection tasks. Furthermore, complex and variable weather conditions and geographical features further complicate the detection process. Consequently, targeted improvements have been implemented to enhance both the accuracy and efficiency of detection. A GSConvns-based slim neck, an ESSE module, and a Wise-*IoU* loss function are integrated into YOLOv8, which is termed GEW-YOLO.

Lightweight multi-ship detection network architecture

The overall architecture of the proposed GEW-YOLO is illustrated in Fig. 1. Which contains the input layer, the feature extraction backbone, the feature fusion neck, and the regression classification head. GEW-YOLO

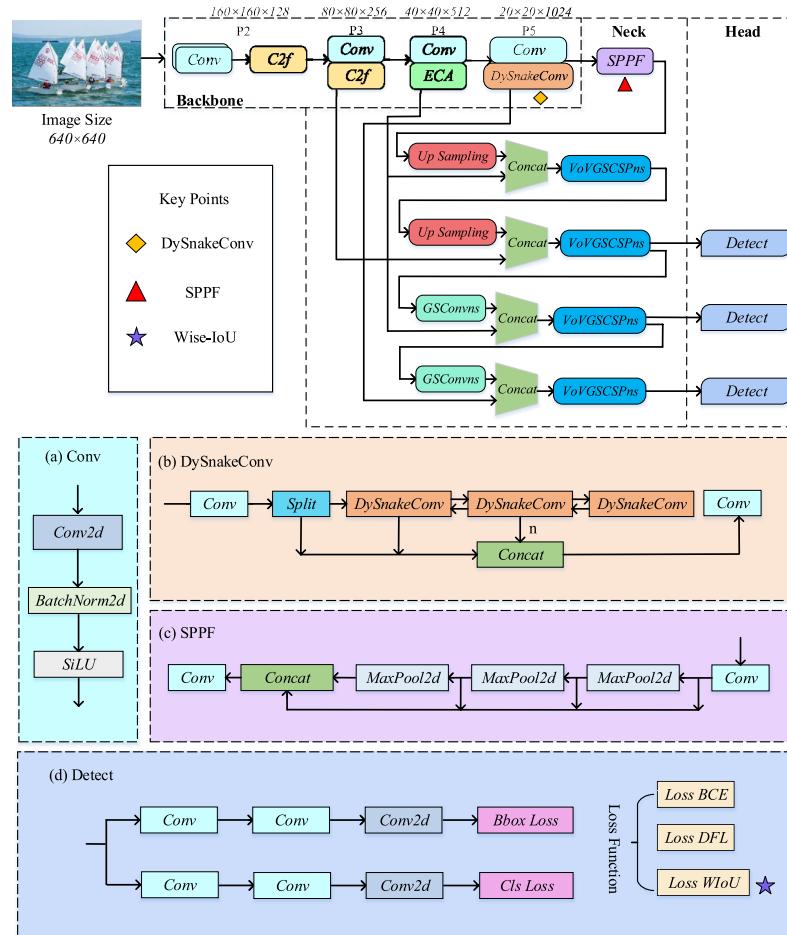


Fig. 1. Detailed architecture of the proposed GEW-YOLO.

distinguishes itself with a highly lightweight neck design, which makes it ideal for real-time applications like ship inspection. At the input layer, data augmentation techniques are employed to enhance the diversity of the input image data. These techniques include cropping, brightness correction, random rotation, and the addition of Gaussian noise, which improve the model's capability to detect ships under complex sea conditions. Cropping ratio and brightness adjustment are often used procedures in the context of data augmentation. For infrared ship datasets, random cropping and brightness tweaks work especially well since they accurately replicate changes in shooting distance and illumination. To improve the model's resilience to infrared photos, the brightness is set to plus or minus 30%, and the cropping ratio is set to 0.9, maintaining 90% of the original image area. Furthermore, ship target detection data is often supplemented with Gaussian noise, add or subtract 15° rotation, and flipping either horizontally or vertically. These techniques enhance the diversity of ship samples and enhance model generalization, making them particularly useful for the Dockship and Seaships visible light datasets.

The Resize technique is commonly used for processing image sizes. This method extracts and reorganizes pixel information according to the target dimensions using methods like pixel clustering and interpolation, going beyond basic cropping or padding. Consequently, the resized image is scaled to the desired size while maintaining the majority of the original information. To satisfy model input requirements, the Dockship dataset, for example, which has 1926 photos of size 1280×850 , has been shrunk to 640×640 . A "proportional scaling + padding" approach is used to avoid distortion. In particular, the image is scaled down by a factor of 0.5 to make it 640×425 in size, using the long side as the scaling reference. Padding is placed above and below to reach the desired height because it is less than 640. In order to guarantee conformity with the model's input criteria, the Infrared Offshore Ship and Seaships datasets also go through size normalization using the same method, finally downsizing all photos to 640×640 . After preprocessing, the image data is sent to the backbone network for precise detection and additional ship feature extraction.

From the Fig. 1, the backbone network is composed of Conv modules, C2f (shortcut) modules, DySnakeConv²⁶ modules, and SPPF (Spatial Pyramid Pooling Fast) modules. As shown in Fig. 1a, c, these modules are specifically designed to extract detailed ship information features from images. They also form the foundational framework of the entire network, ensuring that the model can effectively capture key ship features. Lightweight convolution methods like GSConvns (Group Shuffle Convolution with Neighborhood Sampling) and VoVGSCSPNs (Vision Optimized Group Shuffle Cross Stage Partial Network with Neighborhood Sampling) modules are selected for the neck layer.

GSConvns is an improved convolutional module designed to enhance feature extraction efficiency and improve the model's adaptability to complex backgrounds. By incorporating group convolution, this module significantly reduces computational complexity, while the channel shuffle mechanism facilitates cross-group information interaction, thereby mitigating feature isolation caused by grouping. Furthermore, GSConvns employs a neighborhood sampling strategy to effectively aggregate local contextual information, thereby strengthening the model's ability to perceive fine-grained features. This design is particularly suited for small object detection tasks in complex environments, such as ship detection in maritime scenes as addressed in this study.

VoVGSCSPNs is a feature fusion module built upon GSConvns. It splits the input feature map into two branches: one is processed by GSConvns for feature enhancement, while the other preserves the original information. These branches are subsequently fused through concatenation followed by convolution operations. This architecture not only alleviates the vanishing gradient problem in deep networks but also enhances the diversity and robustness of feature representations via a multi-scale contextual sampling mechanism. VoVGSCSPNs significantly improves the model's generalization capability in complex scenes while retaining critical information of small objects, demonstrating strong performance in both visible and infrared image fusion scenarios. By integrating GSConvns and VoVGSCSPNs into the YOLOv8 framework, the resulting model achieves a synergistic effect that optimizes both feature extraction and fusion. This combination effectively reduces the model's parameter count and computational cost, striking an optimal balance between high accuracy and lightweight design. Because of its effective design, the model can handle vast amounts of ship picture data more quickly, satisfying the requirements of real-time detection.

Furthermore, the incorporation of an ESSE attention module successfully avoids intricate dimensionality reduction and augmentation procedures. This results in lightweight features and improved efficiency. The ESSE attention module improves maritime focus in the model and lessens background noise in ship detection tasks. This allows the model to focus more on identifying and classifying ship targets.

Lastly, the loss function is refined in the detecting head by substituting the Wise-IoU loss function for the CIoU loss function. Ship target recognition and the processing of low-quality samples have significantly improved as a result of this modification. By successfully improving ship detection and classification accuracy, the Wise-IoU loss function offers a more dependable and durable assurance for ship detection jobs.

Lightweight neck network module– GSConvns

GSConvns is designed to optimize ship detection and classification tasks. Meanwhile, it strikes a balance between the accuracy and speed of the model. Firstly, it cleverly combines the advantages of ordinary convolution and depthwise separable convolution. Although ordinary convolution has strong feature extraction ability, its computational cost is high; And depthwise separable convolution significantly reduces the number of parameters and computational complexity by decomposing the standard convolution into depth convolution and point convolution. GSConvns utilizes this feature to reduce model size while maintaining feature extraction capability. Secondly, the computational efficiency is further optimized by combining grouped convolution and spatial convolution (SC). Grouping convolution divides the input channel into several groups and performs convolution operations separately for each group, effectively reducing the number of parameters and

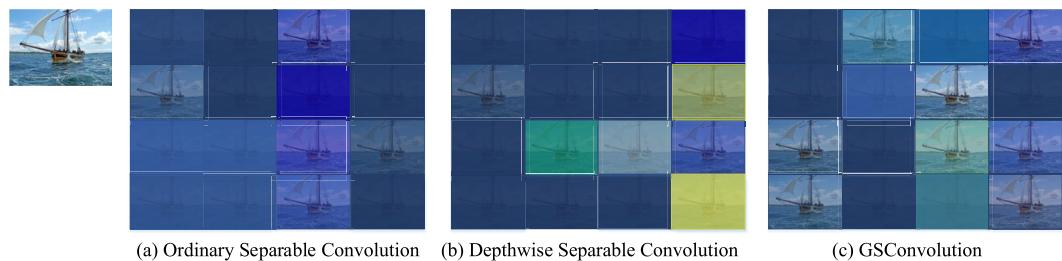


Fig. 2. SC, DSC, and GSConvolution.

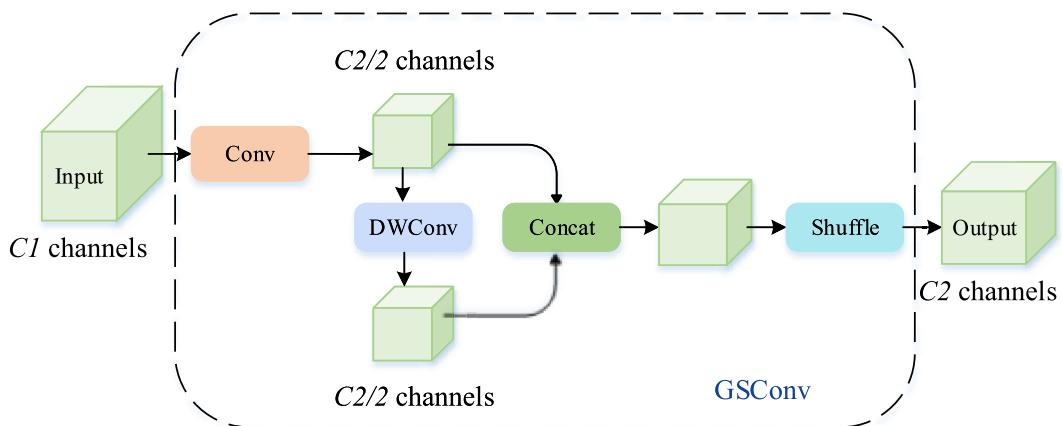


Fig. 3. GSConvns module.

computational complexity. And spatial convolution focuses on spatial feature extraction, combining the two to make the model more efficient in the process of ship feature extraction.

The core difference between them and traditional depthwise separable convolutions is that although traditional depthwise separable convolutions effectively reduce the number of parameters and computational complexity by decomposing standard convolutions into depthwise convolutions and pointwise convolutions, their information exchange between channels is limited, and they lack sufficient expression of multi-scale features; GSConvns enhances cross channel information flow and local feature extraction capabilities while maintaining lightweight by introducing grouping convolution, channel shuffling, and neighborhood sampling mechanisms. VoVGSCSPNs combines multi-path structure and Ghost convolution to achieve efficient fusion and information preservation of multi-scale features through spatial pyramid pooling. The synergistic effect of the two significantly reduces model parameters and computational overhead, while effectively improving the richness of feature expression.

ESSE module

To capture detailed features for ship recognition and classification, GSConvns should be applied to the neck layer. This prevents semantic information loss that may occur in the top layers. The visualization results of Separable Convolution(SC), Depthwise Separable Convolution(DSC), and GSConvns are shown in Fig. 2. The figure indicates that the feature maps of GSConvns and SC are more comparable than those of DSC and SC. This indicates that the accuracy of the model is very close to that of ordinary convolution²⁷. The Slim-Neck employs the GSConvns method to mitigate the adverse effects of DSC defects while fully utilizing its advantages. This approach enhances target localization and classification capabilities in ship detection and classification.

As shown in Fig. 3, GSConvns implements DSC using DWConv after processing a downscaled version of a SC. Following these two convolutions, the output is concatenated. Lastly, the information produced by regular convolution is integrated into each segment of the information provided by DSC using a shuffle operation. The proposed structure accelerates processing speed and improves model efficiency. Additionally, it enhances the model's ability to capture intricate ship details while maintaining detection accuracy.

The introduction of the attention mechanism can markedly enhance the model's focus on ship features in detection and classification tasks. Enhance the precision of categorization and detection throughout the procedure. The ECA module (Efficient Channel Attention)²⁸ is a streamlined channel attention method designed to improve the model's capacity to capture essential semantic information by adaptively assessing the significance of various channels and weighting the channel dimensions of feature maps. It solely pertains to interaction inside the channel dimension and does not handle spatial information, hence limiting its efficacy on targets with dense arrangements and significant scale variations. The ESSE module augments capabilities from both semantic and

spatial dimensions, effectively integrating the two forms of information via the Feature Fusion Module (FFO). The ESSE module possesses the semantic improvement capabilities of the ECA module while also markedly enhancing the model's awareness of spatial positioning, target boundaries, and detailed information.

In the ESSE module, 1×1 semantic convolution and 3×3 spatial convolution operate concurrently in parallel branches. The 1×1 convolution primarily facilitates information exchange and semantic augmentation within the channel dimension. It possesses minimal computational cost and effectively integrates semantic information across many channels, hence improving the model's comprehension of high-level semantics, including ship classifications and textures. The 3×3 convolution emphasizes the extraction of local spatial characteristics, improving the model's ability to discern target borders, forms, and spatial configurations by gathering contextual information from adjacent areas. The ESSE module enhances the model's semantic comprehension and spatial feature extraction through labor division and collaboration, without substantially elevating computational complexity. Figure 4 illustrates the network architecture and operational principle of the ESSE module. The feature fusion process of the ESSE module is executed by the FFO submodule, which primarily integrates the output features of 1×1 semantic convolution and 3×3 spatial convolution using residual connections and a weighted fusion technique. During the fusion process, FFO incorporates a residual connection mechanism to integrate the original input features with the fused features, thereby preventing information loss and improving gradient flow. Ultimately, FFO generates a collection of refined feature maps that encompass both substantial semantic information and accurate spatial details, yielding superior input for the detection head and enhancing the model's ability in detecting ship targets.

The ESSE module is easy to integrate into various CNN structures, avoiding the loss of feature information. Firstly, through average pooling, the spatial dimension is compressed to obtain the global semantic feature information. Secondly, a $1 - D$ convolution (as shown in Fig. 4, $K=3$) operation with the convolution kernel size of k is carried out, and the σ activation function obtains the weight w of each channel. Finally, normalized semantic weight of feature channel is multiplied with the corresponding elements of the original input feature map to get the final output feature map X'' . This process can strengthen the feature semantic information of the ship, suppress the background noise, and improve the ability of the model to recognize the ship.

Wise-IoU loss function

In maritime vessel detection, ship targets can display considerable scale variations resulting from disparities in distance and viewpoint. A multitude of small boats are closely positioned, rendering them particularly vulnerable to background noise interference. Wise-IoU dynamically modifies the loss weights to enhance the model's focus on small, challenging-to-detect targets and those that are tough to discover during training. The computational complexity of Wise-IoU is analogous to that of conventional IoU loss, without imposing any extra processing demands. This makes it more appropriate for the ship identification system discussed in this study.

Wise-IoU (WIoU)²⁹ is a loss function of Bounding Box Regression (BBR) in ship detection in this paper. The gradient is dynamically adjusted based on the anchor's mass. This helps the model focus more on ordinary-quality anchors and reduces over-fitting to low-quality examples, ultimately improving performance. The WIoU loss function uses an "outlier" instead of the traditional IoU index to enhance the anchor frame quality in ship inspection.

The WIoU v1 used in this article improves the generalization ability of the model in processing low-quality ship images through its unique focusing mechanism. The diagram of WIoU is shown in Fig. 5.

In Fig. 5, W_g and H_g represent the minimum size of the closed box. The formula is as follows³⁰.

$$L_{\text{WIoU}v1} = R_{\text{WIoU}} L_{\text{IoU}} \quad (1)$$

$$R_{\text{WIoU}} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 - H_g^2)^*}\right) \quad (2)$$

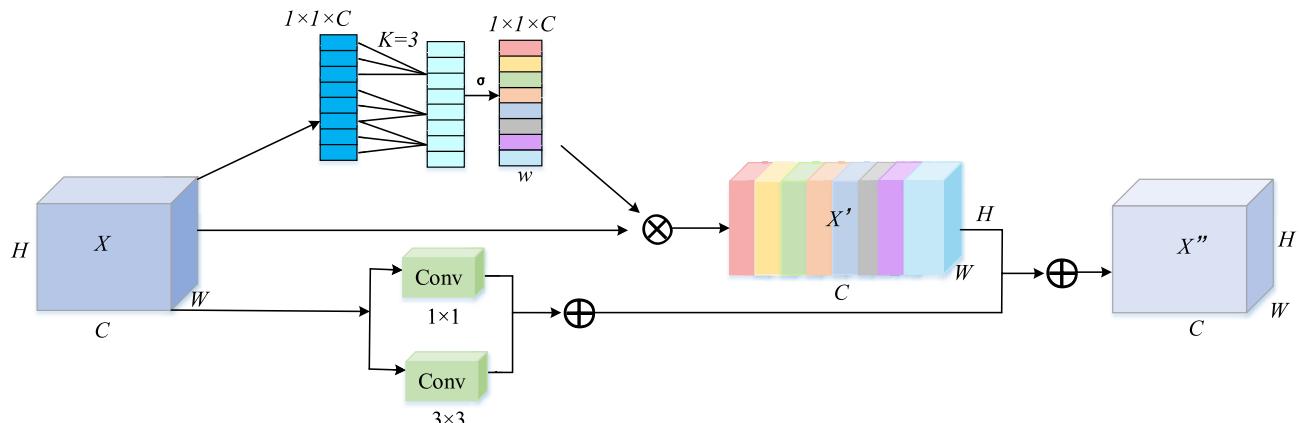


Fig. 4. Network structure of ESSE module.

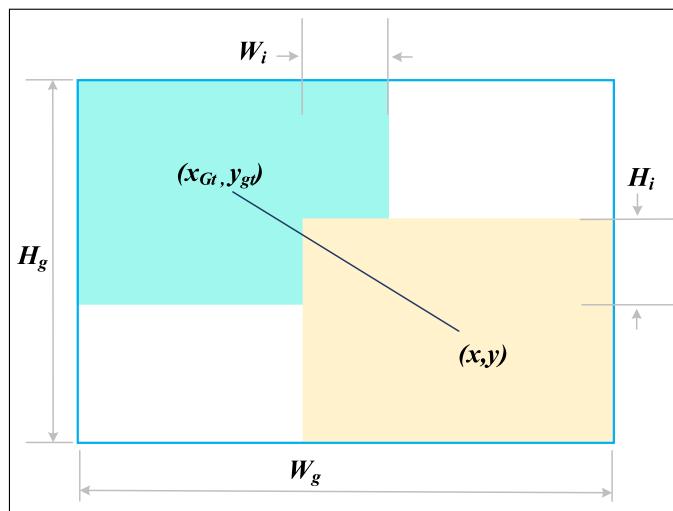


Fig. 5. Schematic diagram of WIoU.

where, L_{IoU} represents R_{WIoU} , which significantly reduces the quality of the anchor box, R_{WIoU} denotes a significantly enlarged anchor box L_{IoU} . The superscript * indicates that to prevent hindering convergence speed, W_g and H_g are separated from the calculation graph.

The formula for WIoU is:

$$\text{WIoU} = IoU^\gamma \times (1 + \alpha e^{-\beta \cdot IoU}) \quad (3)$$

Among them, γ , α , and β are hyperparameters obtained through experimental tuning. In this article, γ is set to 2.0, α is set to 0.5, and β is set to 1.0.

The introduction of the Wise-IoU loss function brings a significant improvement to our proposed GEW-YOLO model. This model incorporates a dynamic non-monotony focusing mechanism. It adjusts weights based on target-prediction alignment, thereby prioritizing samples that are difficult to detect and classify. Furthermore, the Wise-IoU loss function indirectly improves the classification ability of the model by optimizing the accuracy of the boundary box and reducing classification errors.

Experiment results and discussion

In this section, the performance of the proposed GEW-YOLO model is evaluated through several experiments.

Experimental environment and parameter settings

All dataset samples are annotated by the open-source software LabelImg. The image name, target category, and size and coordinates of the bounding rectangle are among the annotation labels that are generated by using rectangular boxes to mark the targets. After being saved in XML format, all annotation data is translated from PASCAL VOC format to YOLO format for model training.

The experimental environment is based on Python 3.9.19 and implemented using the PyTorch 2.3.1 framework on the Ubuntu 22.04.4 operating system. Use remote server NVIDIA GeForce RTX 3060 GPU (with 12 GB of memory) and CUDA12.2 and CUDNN 8.9.2 accelerated GPU for training and testing. In the image preprocessing process, the height and width of the image are set to 640×640 , and the batch size is 8. During the training process, model stability was observed around 200 epochs. In order to save computing resources, we adjusted the number of training periods for all models to 400. Use the stochastic gradient descent(SGD) optimizer with a learning rate of 0.01, momentum of 0.937, and weight decay set to 0.0005. Apply embedded data augmentation during the training phase. Data augmentation is turned off during the last 10 periods of the training process. The remaining parameters are set to default values in TensorFlow.

Evaluation indicators

Precision, *Recall*, *mAP50*, and *mAP0.5-0.95* are used as evaluation criteria for the performance of the detection model. The specific definitions are as follows.

- (1) P , represents *Precision*, which is the proportion of correctly predicted samples based on the expected results. The formula is as follows³¹.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

where the results of positive cases are indicated by *TP* (True Positive) and *FP* (False Positive). The higher the accuracy value, the more accurate the detection results obtained by the algorithm. This also means fewer cases are incorrectly marked as positive during detection or classification, leading to more precise detection results.

(2) *R*, represents the recall rate, which is the proportion of correctly predicted positive class samples to all actual positive class samples. The formula is as follows³².

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

where *FN* is False Negative, the recall rate plays a crucial role in ensuring that the model can fully cover key elements, especially in application scenarios where loss detection costs are high. To properly evaluate model performance, recall rate is typically used in conjunction with other measures like accuracy.

(3) *mAP* represents Mean Average Precision. *mAP* is an indicator used to evaluate object detection performance of the model that takes into account both the *R* and *P* of the model. The detection accuracy of the model increases with the size of the *mAP* value. The calculation formulas for *AP* and *mAP* are as follows³³.

$$AP = \sum_{k=0}^{k=n-1} [Recalls(k) - Recalls(k+1)] \times Precisions(k) \quad (6)$$

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k \quad (7)$$

where *AP* is the Average Precision, *n* is the total amount of data, and *k* is the index of each sample point. *mAP@0.5* indicates the average accuracy when the Intersection over Union (IoU) threshold is 0.5. To be more precise, acquire *mAP@0.5* by averaging the *AP* of all categories after calculating the *AP* of all types for each category. With a step size of 0.05, *mAP@0.5-0.95* is the average *mAP* at various IoU thresholds (varying from 0.5 to 0.95). *mAP@0.5* and *mAP@0.5-0.95* is an essential indicator for measuring the performance of object detection models.

(4) *F1_Curve*. The *F1* score refers to the harmonic mean of accuracy and recall, aiming to consider both accuracy and recall simultaneously. With a maximum value of 1 and a minimum value of 0, the better the model, the higher its value. The calculation formula is as follows³⁴.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

The *F1* Confidence Curve is a visual graph that shows the continuous changes in *F1* score as Confidence gradually increases. The *F1* Confidence curves for the Dockship, Seaships, and Infrared Offshore Ship datasets are represented by (a), (b), and (c) accordingly in Fig. 6.

(5) *P_Curve*. *P_Curve* is a chart used to evaluate the performance of the model, which shows the relationship between the accuracy and confidence of the model. As Fig. 7 illustrates, detection accuracy rises in tandem with the confidence level.

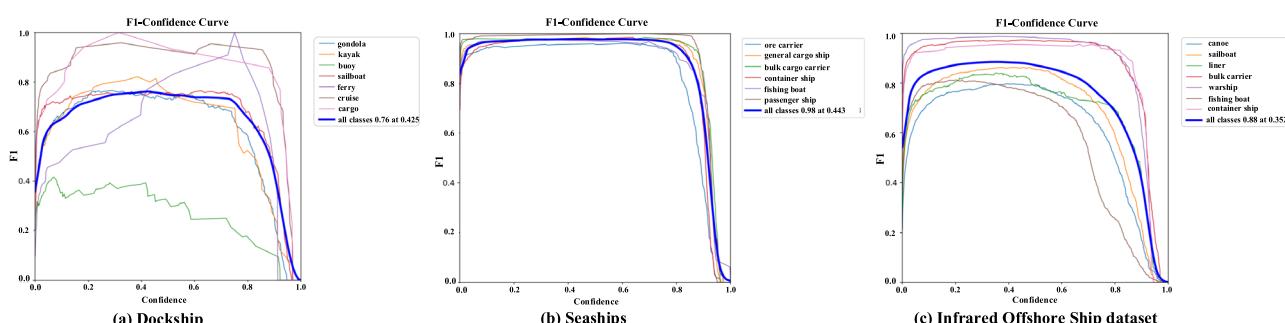


Fig. 6. *F1*-Confidence Curve. The curve in the figure shows the changes in *F1* scores for each type of vessel under different confidence thresholds.

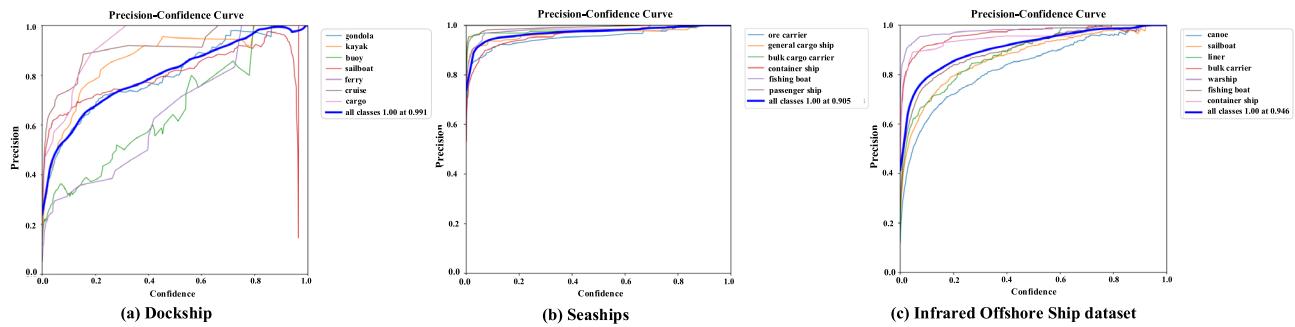


Fig. 7. P_Curve. The figure shows the accuracy of model predictions for different ships at different confidence thresholds.

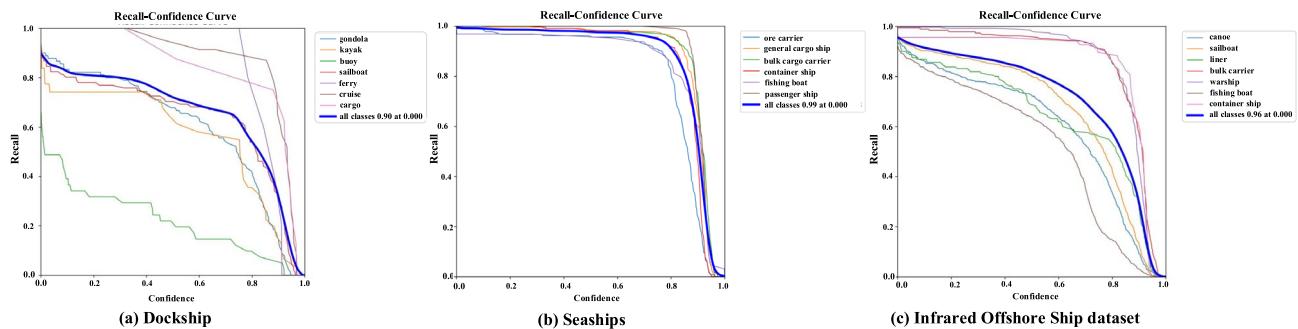


Fig. 8. R_Curve. The figure shows the recall rates predicted by the model for different ships at different confidence thresholds.

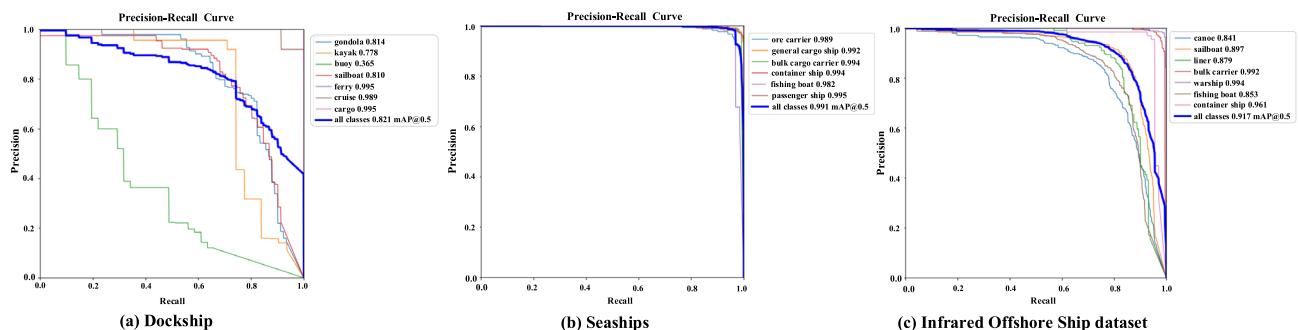


Fig. 9. PR_Curve. The curve in the figure represents the performance of different types of ships in PR_Curve, reflecting the comprehensive detection ability of the model for various types of ships.

- (6) R-Curve. The R-Curve demonstrates the relationship between the recall and confidence of the model. In R_curve, when the confidence threshold is 0, the recall rate does not necessarily have to be 1. As shown in Fig. 8, the height of the curve directly reflects the recall rate of the model at different confidence thresholds. A higher curve indicates that the model has a higher recall rate.
- (7) PR_Curve. PR_Curve describes the link between accuracy and recall at different classification thresholds. The model can sustain high accuracy and high recall while the PR_Curve is in a high position. The curve will move toward low precision values or low recall rates when the model shows tendencies toward high-precision values or high recall rates. The PR_Curves of the three datasets are shown in Fig. 9.
- (8) IoU. IoU measures the degree of overlap between predicted bounding boxes and real bounding boxes. Its value ranges from 0 to 1, with the closer it is to 1 indicating a higher degree of agreement between the predicted results and the actual situation. In addition, IoU can also be used as part of the loss function to help the model better learn bounding box predictions during the training process.

Model	GSConvns	ESSE	Wise-IoU	Precision(%)	Recall(%)	mAP@0.5(%)
YOLOv8	n	n	n	80	70.9	78.3
YOLOv8+GSConvns	y	n	n	86	60	76.6
YOLOv8+ESSE	n	y	n	79.4	67.4	77.8
YOLOv8+Wise-IoU	n	n	y	80.6	62	71.1
y/n GSConvns	n	y	y	78.5	69.6	79.2
y/n ESSE	y	n	y	75.4	70.6	74.1
y/n Wise-IoU	y	y	n	79.4	67.4	75.6
GEW-YOLO(ours)	y	y	y	79.4	76.9	82.1

Table 1. Different ablation experimental results of GEW-YOLO model on Dockship dataset. y/n represents none. Significant values are in [bold].

Model	GSConvns	ESSE	Wise-IoU	Precision(%)	Recall(%)	mAP@0.5(%)
YOLOv8	n	n	n	94.5	95.5	97.1
YOLOv8+GSConvns	y	n	n	95.6	93.8	97.6
YOLOv8+ESSE	n	y	n	97.4	96.4	98.3
YOLOv8+Wise-IoU	n	n	y	95.9	94	98.4
y/n GSConvns	n	y	y	97.4	97	98.9
y/n ESSE	y	n	y	95.1	95	98.1
y/n Wise-IoU	y	y	n	96.7	96.1	98.7
GEW-YOLO(ours)	y	y	y	97.6	98.1	99.1

Table 2. Different ablation experimental results of GEW-YOLO model on Seaships dataset. y/n represents none. Significant values are in [bold].

Model	GSConvns	ESSE	Wise-IoU	Precision(%)	Recall(%)	mAP@0.5(%)
YOLOv8	n	n	n	87.2	83.5	89.4
YOLOv8+GSConvns	y	n	n	90.7	85.7	90.5
YOLOv8+ESSE	n	y	n	91.8	84.2	90.2
YOLOv8+Wise-IoU	n	n	y	89.6	84.6	89.4
y/n GSConvns	n	y	y	90.4	86.6	91.1
y/n ESSE	y	n	y	89.3	86.4	91.2
y/n Wise-IoU	y	y	n	90.6	83.0	91.0
GEW-YOLO(ours)	y	y	y	90.7	86.6	91.7

Table 3. Different ablation experimental results of GEW-YOLO model on Infrared Offshore Ship dataset. y/n represents none. Significant values are in [bold].

Ablation study

Ablation studies are presented to evaluate the contributions of each module in the proposed GEW-YOLO model. The influence of different techniques on the detection performance of each method is studied. In this paper, the GEW-YOLO model and the YOLOv8-based comparison model are implemented to ensure that the hyperparameters are exactly the same in the training process. In addition, the impact of various improvement modules on the Dockship dataset, Seaships dataset, and Infrared Offshore Ship datasets are evaluated respectively.

YOLOv8+GSConvns means that the neck structure of YOLOv8 is replaced by the GSConvns module. YOLOv8+ESSE indicates that an attention mechanism is added. YOLOv8+Wise-IoU refers that the GEW-YOLO model replaces the original YOLOv8 neck structure. y/n ESSE means that the GEW-YOLO model is without an ESSE attention module added. y/n Wise-IoU indicates that the GEW-YOLO model retains the original loss function of YOLOv8.

Tables 1, 2 and 3 show the ablation experimental results of GEW-YOLO models on the Dockship, Seaships, and Infrared Offshore Ship datasets. The comparing methods refer to certain modules while omitting others.

The experimental results on the Dockship dataset have improved with the addition of the GSConvns module, ESSE, and Wise-IoU to the YOLOv8 network. In particular, the GSConvns module reduced *Recall* by 10%, while it improved *Precision* by around 6%. Wise-IoU increased the *Precision* of the model by 0.6%, demonstrating its efficacy across different IoU thresholds. With a minor 0.7% decrease in *Recall*, the GSConvns module increased ship recognition *Precision* by 1.1% and *mAP* by 0.5% for the Seaships dataset. Each metric increased by about 1% as a result of the ESSE module. The introduction of the ESSE module ensures that the detailed features of the



Fig. 10. Detection performance of GEW-YOLO on Dockship dataset.

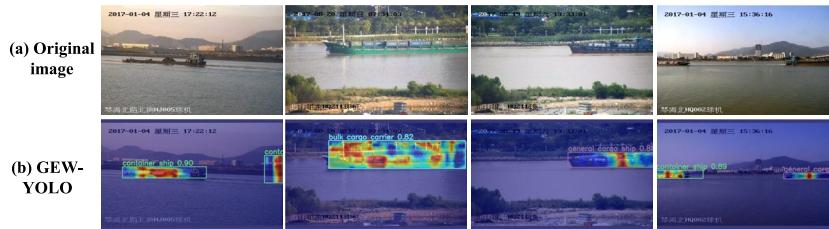


Fig. 11. Detection performance of GEW-YOLO on the Seaships dataset.

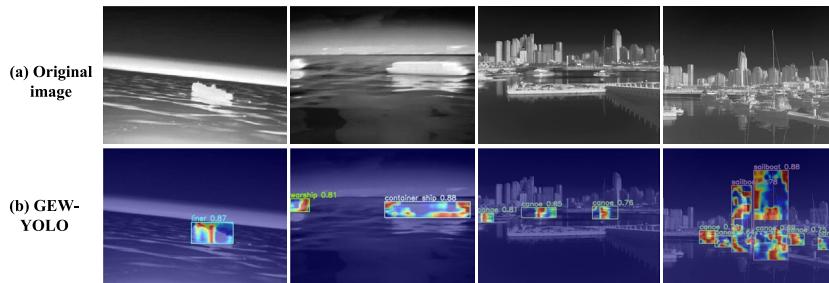


Fig. 12. Detection performance of GEW-YOLO on Infrared Offshore Ship dataset.

ship are taken into account, thereby improving the detection efficiency of the model. According to Table 3, *Recall* and *Precision* increased by 2.5% and 3.5%, respectively, and *mAP@0.5* increased by 2.2% when the integrated GSConvs module was compared to the baseline YOLOv8 model. It is worth noting that when the baseline model introduces the ESSE module, the accuracy and *mAP@0.5* are improved by 4.6% and 0.8%, respectively. When introduced separately, the Wise-IoU did not affect *mAP@0.5*, however, other metrics also showed modest gains. The improvement in model accuracy but the decrease in recall may be due to two reasons. Firstly, GSConvs and other modules enhance the ability of model to distinguish positive samples, while also increasing its sensitivity to negative samples, resulting in some positive samples being misjudged as negative samples; Secondly, these modules increase the complexity of the model, affecting its generalization ability in small targets or complex scenarios. Future research will delve deeper into this trade-off relationship.

The suggested GEW-YOLO model performs better overall in terms of *Precision*, *Recall*, *mAP@0.5*. Notably, in the Dockship dataset, *mAP@0.5* rose from 78.3% to 82.1%. The four associated indicators for the Seaships and another dataset (Table 3) demonstrated increases of 3.1%, 2.6%, 2%, 6.3% and 3.5%, 3.1%, 2.3%, 4.2%, respectively. It demonstrates the capacity of the model to improve average detection precision and its efficacy in ship identification tasks, even in intricate marine environments.

Experimental discussion

The heatmaps of GEW-YOLO for object detection on three different datasets—Dockship, Seaships, and Infrared Offshore Ship dataset are shown in Figs. 10, 11 and 12. The heat map results also indicate the ship type and confidence, and the detection degree of the model can be clearly seen.

Comparison with improved YOLO based detection methods

To highlight the advantages of the proposed GEW-YOLO model, other detection techniques based on YOLOv8 enhancement and ship field detection techniques are compared in this paper.

Model	Precision(%)	Recall(%)	mAP@0.5(%)	mAP@0.5:0.95(%)	Params(M)
YOLOv5	78.2	73.1	78.4	50.2	1.9
YOLOv6	78.7	74.7	79.5	51.4	4.3
YOLOv7	80.2	74.2	79.2	52.1	3.2
YOLOv9	80.6	69.5	76.4	49.2	3.3
YOLOv10	80.6	75.6	80.8	52.0	2.7
YOLOv11	79.3	76.7	81.2	53.5	3.3
Faster-RCNN ¹¹	80.1	73.4	79.8	51.6	100
YOLO-MS ³⁵	81.6	73.0	81.0	51.9	2.5
YOLO-NL ³⁶	74.7	71.7	72.5	46.7	4.5
YOLO-PL ³⁷	73.2	74.1	72.6	45.4	3.5
Gold-YOLO ³⁸	75.8	70.2	74.5	46.8	4.1
FFCA-YOLO ³⁹	73.6	71.2	73.9	47.4	4.6
SOD-YOLO ⁴⁰	75.7	74.4	74.5	46.9	4.5
GEW-YOLO(ours)	79.4	76.9	82.1	61.1	1.2

Table 4. Experimental results of YOLO based detection method on Dockship dataset. Significant values are in [bold].

Model	Precision(%)	Recall(%)	mAP@0.5(%)	mAP@0.5:0.95(%)	Params(M)
YOLOv5	93.6	93.5	95.2	78	1.9
YOLOv6	92.5	91.8	96.6	69.6	4.3
YOLOv7	95.4	93.7	93.7	79.8	3.2
YOLOv9	95.7	94	96.4	76.5	3.3
YOLOv10	92.4	94.9	94.7	80.8	2.7
YOLOv11	96.1	98.7	95.3	76.5	3.3
Faster-RCNN ¹¹	95.5	96.8	94.8	77.6	100
YOLO-MS ³⁵	98.2	93.1	94.1	78.4	2.5
YOLO-NL ³⁶	90.3	89.4	91.6	64.2	4.5
YOLO-PL ³⁷	91.5	90.2	92.3	64.6	3.5
Gold-YOLO ³⁸	93.4	96.8	94.6	67.9	4.1
FFCA-YOLO ³⁹	94.7	91.4	95.8	67.4	4.6
SOD-YOLO ⁴⁰	95.3	95.8	96.3	71.6	4.5
YOLO-HPSD ⁴¹	-	-	98.8	-	-
GEW-YOLO(ours)	97.6	98.1	99.1	86.3	1.2

Table 5. Experimental results of YOLO based detection method on Seaships dataset. Significant values are in [bold].

We experimented using the publicly accessible Dockship, Seaships, and Infrared Offshore Ship datasets to validate our suggested model. Numerous YOLO-based models, including YOLOv5 to YOLOv11, have been created. In addition, the detection accuracy of the two-stage algorithm Faster-RCNN was also considered. These are noteworthy contributions. Chen et al.³⁵ presented an effective target detector, YOLO-MS, exhibiting robust performance on the COCO dataset. Zhou³⁶ advanced YOLO-NL by augmenting technologies like CSPNet and PANet, leading to enhanced accuracy in multi-target detection. YOLO-PL³⁷ is a lightweight helmet detection technique that uses an improved PAN (E-PAN) structure with a specially tailored YOLO-P algorithm to increase accuracy. Wang et al.³⁸ presented the Gather-and-Distribute mechanism, tailored to tackle feature fusion issues in information fusion, and proposed the Gold-YOLO method. Zhang et al.³⁹ developed three novel modules to improve feature representation and background suppression for the detection of small targets in remote sensing. This work resulted in the creation of a lightweight method, FFCA-YOLO. Based on YOLOv8, SOD-YOLO by Li et al.⁴⁰ creates a new neck architecture, BSSI-FPN, to handle spatial information sparsity and incorporates RFCBAM modules to enhance feature extraction. By adding the C2F_iAFF module, MLCA mechanism, and BiFPN structure optimization feature fusion and information integration mAP@0.5, Zhu et al.⁴¹ introduced a high-precision ship target identification technique called YOLO-HPSD, which increases detection accuracy to 98.86%. The accuracy of the approach suggested in this article is higher than that of the YOLO-HPSD algorithm.

The experimental results of the detection on three datasets are shown in Tables 4, 5 and 6. It can be observed that GEW-YOLO still performs well in ship detection in complex water scenarios. These comparisons and experimental results collectively validate the significant advantages of the GEW-YOLO model in the field of ship detection.

Model	Precision(%)	Recall(%)	mAP@0.5(%)	mAP@0.5:0.95(%)	Params(M)
YOLOv5	85.7	90.4	90.3	61.2	1.9
YOLOv6	90.7	84.1	90.5	63.4	4.3
YOLOv7	87.3	84.4	90.2	61.3	3.2
YOLOv9	90.1	85.2	89.9	62.1	3.3
YOLOv10	90.5	86.3	91.0	64.6	2.7
YOLOv11	87.9	87.8	91.6	62.5	3.3
Faster-RCNN ¹¹	86.4	83.5	89.2	60.6	100
YOLO-MS ³⁵	91.3	82.9	90.4	63.9	2.5
YOLO-NL ³⁶	87.6	85.4	90.3	62.8	4.5
YOLO-PL ³⁷	89.3	84.4	89.7	62.1	3.5
Gold-YOLO ³⁸	86.2	82.5	88.4	60.2	4.1
FFCA-YOLO ³⁹	90.8	84.0	90.0	63.9	4.6
SOD-YOLO ⁴⁰	90.6	85.5	90.1	62.2	4.5
GEW-YOLO(ours)	90.7	86.6	91.7	65.4	1.2

Table 6. Experimental results of YOLO based detection method on Infrared Offshore Ship dataset. Significant values are in [bold].

Model	Precision(%)	Recall(%)	mAP@0.5(%)	mAP@0.5:0.95(%)	Params(M)
<i>S</i> ³ <i>Det</i> ⁴²	97.9	97.3	99	84	2.3
FE-YOLO ⁴³	-	-	97.6	72.6	2.1
YOLO-Ships ⁴⁴	89.2	90.2	93.5	-	-
YOLO-MBS ⁴⁵	97.2	96.9	98.8	-	-
GEW-YOLO(ours)	97.6	98.1	99.1	86.3	1.2

Table 7. Experimental results of YOLO based detection method on Infrared Offshore Ship dataset. Significant values are in [bold].

This study compared the performance of various enhanced YOLO models in object recognition tasks. YOLO-LITE achieved the best accuracy in accuracy and YOLOv11 performs well in *Recall*. However, *mAP@0.5* is far lower than the model proposed in this paper. Other models are slightly inferior on ship detection data sets. Overall, GEW-YOLO achieves the best balance between detection accuracy and speed, outperforming other YOLO-based methods.

Comparison with detection methods in specific ship fields

Due to the limited number of references available for the Dockship dataset and Infrared Offshore Ship datasets, we chose to compare some methods from specific ship domains on the Seaships dataset. Li et al.⁴² trained the *S*³ *Det* model and improved the detection performance of small ships through technological improvements. Experimental results have shown better recall and accuracy than the original model. Cai et al.⁴³ proposed the FE-YOLO model. By introducing channel attention and Ghostconv to improve the network structure, the detection accuracy has been improved. Zhang et al.⁴⁴ proposed a lightweight and effective method for YOLO-Ships. By improving the Ghost module, designing a ship feature enhancement module, and C3REGhost module, the feature extraction and detection capabilities have been enhanced. Lu et al.⁴⁵ proposed the YOLO-MBS algorithm. By using MobileNetv3 as the backbone network, introducing the E-IOU loss function, adding the SimAM attention mechanism, and employing the BiFPN feature pyramid network, the model parameters were significantly reduced. Additionally, both computational speed and detection accuracy were improved.

As shown in Table 7, the proposed method performs well in terms of *Precision*, *Recall*, *mAP@0.5*, and *mAP@0.5:0.95*. Compared with the *S*³ *Det* model, although the accuracy decreased by 0.3%, the *Recall*, *mAP@0.5* and *mAP@0.5:0.95* increased by 0.8%, 0.1% and 2.3%, respectively. For key index *mAP@0.5*, the improved modules of YOLO-Ships proposed by Zhang et al. have effectively reduced the model's weight. However, according to the experimental results, the detection performance still requires significant improvement. In summary, our proposed model strikes an effective balance between lightweight design and high-precision performance in ship target detection tasks.

Comparison of test results

Figure 13 illustrates the detection of the identical image from the Dockship dataset with the YOLOv8 and GEW-YOLO models to validate the influence of different models on detection results. GEW-YOLO clearly surpasses YOLOv8 in detection performance, achieving a 3.8% higher mAP (82.1% vs 78.3%) and effectively recognizing images with overlapping labels or objects, with a 6% improvement in recall rate for overlapping targets. Figure 14 illustrates that images were randomly selected from the Seaships test dataset for further analysis, identified

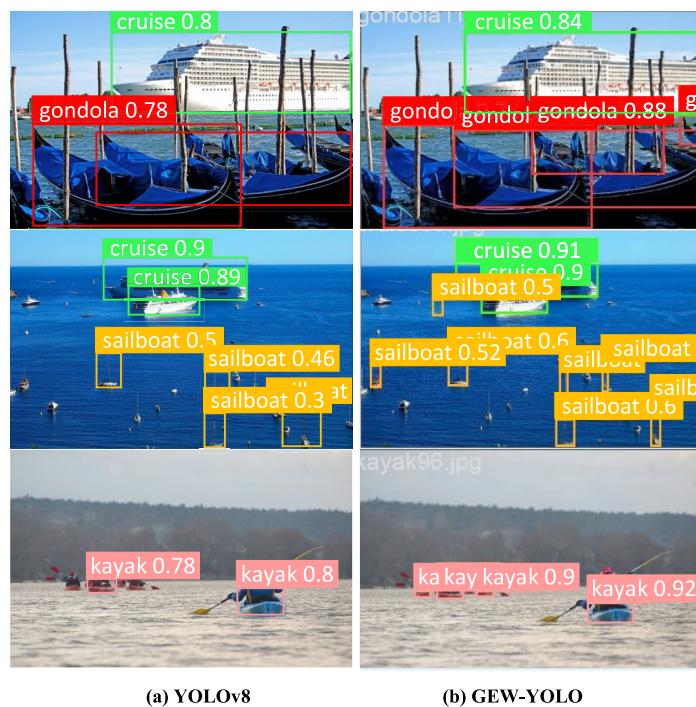


Fig. 13. Comparison results of dockship dataset experiments.



Fig. 14. Comparison results of seaships dataset experiments.

utilizing the YOLOv8 and GEW-YOLO networks. The three chosen photos, which depict sparse things, dense objects, and small, indistinct target objects, effectively illustrate the detecting capabilities. Additionally, images from the Infrared Offshore Ship dataset were randomly selected for evaluation using the YOLOv8 and GEW-YOLO networks; the results are presented in Fig. 15. GEW-YOLO demonstrates superior performance with an average detection confidence score of 0.87, compared to YOLOv8's 0.79, representing a 10.1% improvement in low-contrast conditions.

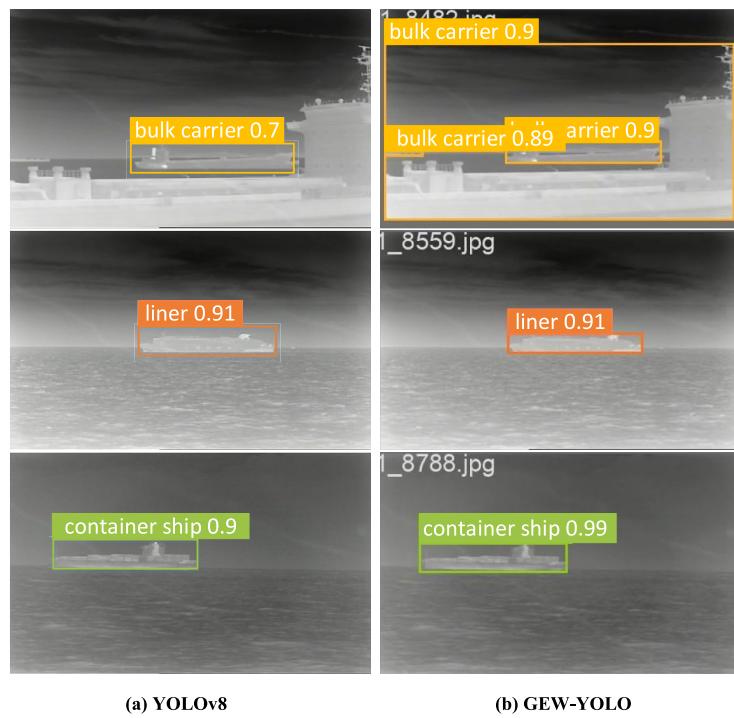


Fig. 15. Comparison results of infrared offshore ship dataset experiments.

The test results clearly indicate that GEW-YOLO outperforms YOLOv8 in object detection. GEW-YOLO has enhanced its detection precision and ability to identify densely packed small objects by the integration of the GSConvns module, ESSE module, and a revised loss function.

Conclusions

In order to meet the demand for a lightweight, real-time, and accurate multi-target detection system for ships, this study proposes a state-of-the-art method for detecting and classifying ship images using the improved YOLOv8 algorithm. Recognizing the limitations of existing algorithms in feature learning and their increased complexity, YOLOv8 is selected as the baseline model and enhanced with ship feature texture information using the ESSE module. The GSConvns module is also integrated into the neck network, which significantly enhances the multi-scale feature fusion while maintaining the lightweight architecture of the model. Finally, the Wise-IoU loss function is applied in the detection head, which exceeds the traditional CIoU and SIoU methods and helps the model achieve higher accuracy. Experimental results show that the improved algorithm has average detection accuracies of 82.1%, 99.1%, and 91.7% on the Dockship, SeaShips, and Infrared Offshore Ship datasets, respectively.

Compared to YOLOv8 and YOLOv11, the improved YOLOv8 outperforms these models in terms of mAP metrics, especially mAP 0.5:0.95, and the number of model parameters is smaller. In addition, the optimized architecture of the improved YOLOv8 ensures faster inference compared to YOLOv11 while reducing computational complexity. The ablation experimental analysis further validates the model's ability to focus on the ship's target area under complex backgrounds, small targets, and different orientations, demonstrating its robustness in accurately detecting and localizing ships.

This study proposes a more effective method for ship target detection and classification; however, several limitations remain. The improved module exhibits reduced accuracy when processing images captured under adverse weather conditions or with low visibility. Additionally, the proposed model's computational complexity increases substantially compared to baseline methods, potentially limiting its deployment in real-time systems with constrained hardware resources. The algorithm also demonstrates diminished performance when detecting small-scale ship targets or vessels with high aspect ratios, indicating challenges in feature extraction for such objects. Furthermore, the scope of this study was confined to a limited set of ship types, as evidenced by the experimental validation. Future research should therefore incorporate a more comprehensive dataset encompassing a broader spectrum of ship categories and diverse environmental conditions to enhance the model's generalization capabilities.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 26 February 2025; Accepted: 24 September 2025

Published online: 30 October 2025

References

- Ali, I. The world's maritime industry in the 21st century: challenges, expectations, and directions. *South East Asian Mar. Sci. J.* **2**, 64–75 (2025).
- Benson, G. A., Mitchell, P. D. & Henson, B. Localization of auvs for ship hull inspection: a review. *IEEE Access* **2025**, 256 (2025).
- Kermeneen, P. How and to what extent can Historical Ship Structural Components be Observed in a Shallow Dynamic Environment. Ph.D. thesis, Flinders University (2023).
- Xie, T., Ji, Q., Yu, P. & Zhang, J. Method for extracting ship shaft rate features by fusing acoustic and magnetic field. *Sci. Rep.* **15**, 27536 (2025).
- Li, Y. & Zhang, D. Toward efficient edge detection: a novel optimization method based on integral image technology and canny edge detection. *Processes* **13**, 293. <https://doi.org/10.3390/pr13020293> (2025).
- Liu, L., Liang, F., Zheng, J., He, D. & Huang, J. Ship infrared image edge detection based on an improved adaptive canny algorithm. *Int. J. Distrib. Sens. Netw.* **14**, 1550147718764639. <https://doi.org/10.1177/1550147718764639journals.sagepub.com/home/dsn> (2018).
- Cui, J., Jia, H., Wang, H. & Xu, F. A fast threshold neural network for ship detection in large-scene sar images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **15**, 6016–6032. <https://doi.org/10.1109/JSTARS.2022.3192455> (2022).
- Klimkowska, A. & Lee, I. A preliminary study of ship detection from uav images based on color space conversion and image segmentation. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **42**, 189–193. <https://doi.org/10.5194/isprs-archives-XLII-2-W-6-189-2017> (2017).
- Zhou, K., Zhang, M., Wang, H. & Tan, J. Ship detection in sar images based on multi-scale feature extraction and adaptive feature fusion. *Remote Sens.* **14**, 755. <https://doi.org/10.3390/rs14030755> (2022).
- Khumaidi, A., Adhitya, R. & Julianito, E. Automatic hull fracture detection system using template matching. In *2020 International Conference on Applied Science and Technology (iCAST)* 589–592 (IEEE, 2020). <https://doi.org/10.1109/iCAST51016.2020.9557676>.
- Zhang, S., Wu, R., Xu, K., Wang, J. & Sun, W. R-cnn-based ship detection from high resolution remote sensing imagery. *Remote Sens.* **11**, 631. <https://doi.org/10.3390/rs11060631> (2019).
- Guo, H., Yang, X., Wang, N., Song, B. & Gao, X. A rotational libra r-cnn method for ship detection. *IEEE Trans. Geosci. Remote Sens.* **58**, 5772–5781. <https://doi.org/10.1109/TGRS.2020.2969979> (2020).
- Nie, X., Duan, M., Ding, H., Hu, B. & Wong, E. K. Attention mask r-cnn for ship detection and segmentation from remote sensing images. *IEEE Access* **8**, 9325–9334. <https://doi.org/10.1109/ACCESS.2020.2964540> (2020).
- Loran, T., da Silva, A. B. C., Joshi, S. K., Baumgartner, S. V. & Krieger, G. Ship detection based on faster r-cnn using range-compressed airborne radar data. *IEEE Geosci. Remote Sens. Lett.* **20**, 1–5. <https://doi.org/10.1109/LGRS.2022.3229141> (2022).
- Wolrige, S. H., Howe, D. & Majidiyan, H. Intelligent computerized video analysis for automated data extraction in wave structure interaction; a wave basin case study. *J. Mar. Sci. Eng.* **13**, 617 (2025).
- Chen, L., Shi, W. & Deng, D. Improved yolov3 based on attention mechanism for fast and accurate ship detection in optical remote sensing images. *Remote Sens.* **13**, 660. <https://doi.org/10.3390/rs13040660> (2021).
- Ye, Y., Zhen, R., Shao, Z., Pan, J. & Lin, Y. A novel intelligent ship detection method based on attention mechanism feature enhancement. *J. Mar. Sci. Eng.* **11**, 625. <https://doi.org/10.3390/jmse11030625> (2023).
- Yang, D. et al. A streamlined approach for intelligent ship object detection using el-yolo algorithm. *Sci. Rep.* **14**, 15254. <https://doi.org/10.1038/s41598-024-64225-y> (2024).
- Wang, S., Li, Y. & Qiao, S. Alf-yolo: enhanced yolov8 based on multiscale attention feature fusion for ship detection. *Ocean Eng.* **308**, 118233. <https://doi.org/10.1016/j.oceaneng.2024.118233> (2024).
- Cheng, G., Chao, P., Yang, J. & Ding, H. Sgstd-yolov8: an improved lightweight yolov8 for real-time target detection for campus surveillance. *Appl. Sci.* **14**, 5341. <https://doi.org/10.3390/app14125341> (2024).
- Chen, Z., Liu, C., Filaretov, V. F. & Yukhimets, D. A. Multi-scale ship detection algorithm based on yolov7 for complex scene sar images. *Remote Sens.* **15**, 2071. <https://doi.org/10.3390/rs15082071> (2023).
- Wei, X., Li, Z. & Wang, Y. Sed-yolo based multi-scale attention for small object detection in remote sensing. *Sci. Rep.* **15**, 3125. <https://doi.org/10.1038/s41598-025-87199-x> (2025).
- Sathana, V., Daniel, S. P., Srikanth, C., Senthivel, K. & Santhoshraj, D. Revolutionizing waste management through the integration of iot and deep learning technology. In *Spectrum and Power Allocation in Cognitive Radio Systems* 214–223 (IGI Global, 2024). <https://doi.org/10.4018/979-8-3693-2893-4.ch014>.
- Shao, Z., Wu, W., Wang, Z., Du, W. & Li, C. Seaships: a large-scale precisely annotated dataset for ship detection. *IEEE Trans. Multimedia* **20**, 2593–2604. <https://doi.org/10.1109/TMM.2018.2865686> (2018).
- Zhang, T. et al. Infrared ship target segmentation based on adversarial domain adaptation. *Knowl.-Based Syst.* **265**, 110344. <https://doi.org/10.1016/j.knosys.2023.110344> (2023).
- Tang, Q. & Ren, X. A student classroom behavior detection system based on improved yolov10. In *2024 China Automation Congress (CAC)* 1402–1406 (IEEE, 2024). <https://doi.org/10.1109/CAC63892.2024.10865023>.
- Li, H. et al. Slim-neck by gconv: a better design paradigm of detector architectures for autonomous vehicles. *arXiv preprint arXiv:2206.02424*. <https://doi.org/10.48550/arXiv.2206.02424> (2022).
- Wang, Q. et al. Eca-net: efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11534–11542 (2020). <https://doi.org/10.1109/CVPR42600.2020.01155>.
- Tong, Z., Chen, Y., Xu, Z. & Yu, R. Wise-iou: bounding box regression loss with dynamic focusing mechanism. *arXiv preprint arXiv:2301.10051*<https://doi.org/10.48550/arXiv.2301.10051> (2023).
- Zhang, Z. et al. Wed-yolo: a detection model for safflower under complex unstructured environment. *Agric. Basel* **15**, 14523. <https://doi.org/10.3390/agriculture15020205> (2025).
- Miao, J. & Zhu, W. Precision-recall curve (prc) classification trees. *Evol. Intel.* **15**, 1545–1569. <https://doi.org/10.1007/s12065-021-0565-2> (2022).
- Chabbouh, M., Bechikh, S., Mezura-Montes, E. & Ben Said, L. Evolutionary optimization of the area under precision-recall curve for classifying imbalanced multi-class data. *J. Heuristics* **31**, 9. <https://doi.org/10.1007/s10732-024-09544-z> (2025).
- Shen, L., Tao, H., Ni, Y., Wang, Y. & Stojanovic, V. Improved yolov3 model with feature map cropping for multi-scale road object detection. *Meas. Sci. Technol.* **34**, 045406. <https://doi.org/10.1088/1361-6501/acb075> (2023).
- Takahashi, K., Yamamoto, K., Kuchiba, A. & Koyama, T. Confidence interval for micro-averaged f1 and macro-averaged f1 scores. *Appl. Intell.* **52**, 4961–4972. <https://doi.org/10.1007/s10489-021-02635-5> (2022).
- Chen, Y. et al. Yolo-ms: rethinking multi-scale representation learning for real-time object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, 562. <https://doi.org/10.1109/TPAMI.2025.3538473> (2025).
- Zhou, Y. A yolo-nl object detector for real-time detection. *Expert Syst. Appl.* **238**, 122256. <https://doi.org/10.1016/j.eswa.2023.122256> (2024).
- Li, H., Wu, D., Zhang, W. & Xiao, C. Yolo-pl: Helmet wearing detection algorithm based on improved yolov4. *Digital Signal Process.* **144**, 104283. <https://doi.org/10.1016/j.dsp.2023.104283> (2024).
- Wang, C. et al. Gold-yolo: efficient object detector via gather-and-distribute mechanism. *Adv. Neural Inf. Process. Syst.* **36**, 1478. <https://doi.org/10.48550/arXiv.2309.11331> (2024).

39. Zhang, Y. et al. Ffca-yolo for small object detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **62**, 1–15. <https://doi.org/10.1109/TGRS.2024.3363057> (2024).
40. Li, Y. et al. Sod-yolo: small-object-detection algorithm based on improved yolov8 for uav images. *Remote Sens.* **16**, 3057. <https://doi.org/10.3390/rs16163057> (2024).
41. Zhu, M., Han, D., Han, B. & Huang, X. Yolo-hpsd: a high-precision ship target detection model based on yolov10. *PLoS ONE* **20**, e0321863 (2025).
42. Li, L. et al. Spotlight on small-scale ship detection: Empowering yolo with advanced techniques and a novel dataset. In *Proceedings of the Asian Conference on Computer Vision* 784–799 (2024). <https://doi.org/10.1007/978-981-960-04-1>.
43. Cai, S., Meng, H. & Wu, J. Fe-yolo: Yolo ship detection algorithm based on feature fusion and feature enhancement. *J. Real-Time Image Proc.* **21**, 61. <https://doi.org/10.1007/s11554-024-01445-5> (2024).
44. Zhang, Y., Chen, W., Li, S., Liu, H. & Hu, Q. Yolo-ships: lightweight ship object detection based on feature enhancement. *J. Vis. Commun. Image Represent.* **101**, 104170. <https://doi.org/10.1016/j.jvcir.2024.104170> (2024).
45. Lu, Q. et al. Yolo-mbs: a lightweight ship target detection method. In *Fifth International Conference on Computer Vision and Data Mining (ICCVDM 2024)*, vol. 13272 677–683 (SPIE, 2024). <https://doi.org/10.1117/12.3048239>.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the Shandong Provincial Natural Science Foundation(ZR2022QF149), the Science Foundation of Shandong Jiaotong University(Grant No.Z202125) and the National Natural Science Foundation of China (Grant No. 61803227).

Author contributions

Ying Han: Investigation, Writing - Original draft. Yugang Wang and Ying Han: Conceptualization. Hao Wang and Nick Renjin: Datasets Resources. Jie Song and Gongxiang Cui: Investigation. Yugang Wang and Fengyu Zhou: Conceptualization, Writing Review & Editing.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.W.

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.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025