

Coastal Vessel Target Detection Model Based on Improved YOLOv7

Guiling Zhao^{1,2} and Ziyao Xu^{1,2}

Received: 03 September 2024 / Accepted: 07 November 2024
© Harbin Engineering University and Springer-Verlag GmbH Germany, part of Springer Nature 2025

Abstract

To address low detection accuracy in near-coastal vessel target detection under complex conditions, a novel near-coastal vessel detection model based on an improved YOLOv7 architecture is proposed in this paper. The attention mechanism Coordinate Attention is used to improve channel attention weight and enhance a network's ability to extract small target features. In the enhanced feature extraction network, the lightweight convolution algorithm Grouped Spatial Convolution is used to replace MPConv to reduce model calculation costs. EIou Loss is used to replace the regression frame loss function in YOLOv7 to reduce the probability of missed and false detection. The performance of the improved model was verified using an enhanced dataset obtained through rainy and foggy weather simulation. Experiments were conducted on the datasets before and after the enhancement. The improved model achieved a mean average precision (mAP) of 97.45% on the original dataset, and the number of parameters was reduced by 2%. On the enhanced dataset, the mAP of the improved model reached 88.08%. Compared with seven target detection models, such as Faster R-CNN, YOLOv3, YOLOv4, YOLOv5, YOLOv7, YOLOv8-n, and YOLOv8-s, the improved model can effectively reduce the missed and false detection rates and improve target detection accuracy. The improved model not only accurately detects vessels in complex weather environments but also outperforms other methods on original and enhanced SeaShip datasets. This finding shows that the improved model can achieve near-coastal vessel target detection in multiple environments, laying the foundation for vessel path planning and automatic obstacle avoidance.

Keywords Vessel target detection; YOLOv7; Attention mechanism; Lightweight convolution; Data enhancement

1 Introduction

Enhancing the ability to identify near-coastal targets is crucial for safeguarding maritime rights and interests and advancing the construction of a maritime power (He et al., 2024; Ramsay et al., 2023). The timely and accurate collection of near-coastal vessel data and vessel classification anal-

ysis plays a pivotal role in ensuring maritime traffic safety (Xu and Guedes Soares, 2023; Wang et al., 2023a). Owing to the complex marine environment, vessel target detection is susceptible to environmental disturbances, such as rain and fog, which considerably impair detection performance (Steccanella et al., 2020; Kong et al., 2022). Therefore, how to efficiently and accurately detect and classify near-coastal vessel targets in complex weather environments is an urgent problem to be solved in the field of maritime target detection.

Current vessel detection methods can be primarily categorized into two types: methods based on feature design and convolutional neural networks. The former can be further divided into statistical thresholding for image segmentation and methods based on visual saliency, shape and texture matching, and anomaly detection (Zhou et al., 2021; Kanjir et al., 2018). These traditional methods are robust in handling complex scenarios, such as sea wave noise and dark-colored vessels, but have low detection accuracy and efficiency and are susceptible to environmental interference when detecting near-coastal targets, failing to meet the stringent requirements of practical applications.

Convolutional neural network-based object detection models have shown exceptional performance in vessel target detection (Guo et al., 2023; Elvidge et al., 2015; Zhang

Article Highlights

- The low accuracy of near-coastal vessel target detection in complex environments is addressed by improving the YOLOv7 model.
- Three improved methods are proposed and compared for vessel target detection on clear days.
- Rainy and foggy weather conditions are simulated, the dataset is enhanced, and the detection effects of different improvement methods are compared.
- Compared with the other seven detection models, the improved model reduces the missed and false detection rates and has higher detection accuracy.

✉ Guiling Zhao
zhaoguiling@lntu.edu.cn

¹ School of Surveying, Liaoning Technical University, Fuxin 123000, China

² School of Geomatics, Liaoning Technical University, Fuxin 123000, China

et al., 2021). Two-stage detection algorithms, such as Fast R-CNN (Girshick, 2015) and Faster R-CNN (Ren et al., 2016), and one-stage detection algorithms, such as SSD (Liu et al., 2016) and YOLO series (Redmon et al., 2016; Redmon and Farhadi, 2017, 2018; Zhu et al., 2021; Li et al., 2021b) have been widely applied to practical vessel detection tasks. Qi et al. (2019) applied Faster R-CNN to a vessel detection task in the Yangtze River Basin and improved the model detection speed through a hierarchical narrowing method. Wang et al. (2020) proposed an improved YOLOv3 vessel detection method using the K-means++ clustering algorithm to reduce prediction box localization loss and missed detection rate. Huang et al. (2023) introduced the attention mechanism Convolutional Block Attention Module into the YOLOv4 model to enhance the model's ability to detect small ships by adjusting feature map weights. Zhang et al. (2022) embedded the attention mechanism Squeeze-and-Excitation into the YOLOv5 model, improving the detection accuracy of large surface vessels. Lastly, Wu et al. (2023) improved multiscale ship target detection accuracy and robustness in nearshore ship datasets by improving the anchor boxes in the YOLOv7 model. However, these methods are only suitable for vessel detection experiments conducted under good weather conditions and have not been experimentally verified for vessel detection in complex environments, such as rainy and foggy environments. Near-coastal target detection in complex weather environments differs from general target detection. First, the detection scene is open, with a wide variety of targets and large size differences (Del-Rey-Maestre et al., 2018; Tsuda et al., 2023). Second, environmental factors, such as rain and fog, blur image features, impeding target detection (Li et al., 2021a). Therefore, exploring a near-coastal vessel target detection model in complex environments is essential to maritime traffic safety.

This study proposes a near-coastal vessel detection model based on the improved YOLOv7. To address target size differences and blurred image features in complex weather, such as rainy and foggy weather, this paper introduces the attention mechanism Coordinate Attention (CA) and integrated width and height attention weights to enhance feature extraction capability for input images. To improve vessel detection accuracy, reduce the model size, and enhance the network's perception of global information, the lightweight convolution Grouped Spatial Convolution (GSConv) is used to replace MPConv. Then, the network's prediction box localization ability is enhanced by replacing CIOU Loss with EIoU Loss, which reduces missed and false detection rates.

2 Improved YOLOv7 model structure

The improved YOLOv7 model is divided into three parts:

the backbone feature extraction network (Backbone), the feature pyramid network for enhancing feature extraction, and YOLOhead, which serves as a classifier and regressor (Wang et al., 2023b). The network structure is shown in Figure 1. The Backbone extracts features from image information to obtain feature layers. A CBS module is composed of a convolution layer, a batch normalization layer, and an activation function layer. It changes the number of channels according to the convolution kernel used, extracts features, and performs downsampling (Liu et al., 2021). In a feature pyramid network, the CA and GSConv are introduced to optimize network structures, reduce network size, facilitate the extraction of feature information from an input feature map, and ultimately improve the detection capability of vessel targets without affecting detection time.

2.1 Attention mechanism

Attention mechanisms mimic human visual perception by enhancing the representation of important features and reducing the representation of unnecessary features on the basis of the human visual system's varying degrees of attention to different regions (Hou et al., 2021). Object-detection algorithms leveraging visual attention mechanisms can effectively distinguish targets from backgrounds. In near-coastal detection tasks, in which vessel targets are often obscured by noise and lighting conditions, CA adaptively learns interchannel relationships and assigns attention weights to different channels, thereby enhancing the extraction of discriminative features for vessel targets and suppressing background noise. This feature considerably improves the performance of small target vessel detection. The structure of CA is shown in Figure 2.

The CA has two parallel processes. First, an input feature map is subjected to average pooling along the width and height dimensions, yielding feature maps corresponding to width and height. The input feature layer shape is (C, H, W) . After average pooling along the width and height dimensions, the feature layers of $(C, 1, W)$ and $(C, H, 1)$ are obtained. These features are then mapped to the width and height dimensions.

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} X_c(j, w) \quad (1)$$

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} X_c(h, i) \quad (2)$$

where $z_c^w(w)$ and $z_c^h(h)$ represent the output results after pooling along the width and height dimensions, respectively, W and H denote the width and height of the input feature layer, and $X_c(h, i)$ and $X_c(j, w)$ represent the feature vectors at the i th row and j th column.

By performing a dimensional transformation on the output feature maps of the two parallel processes, the width

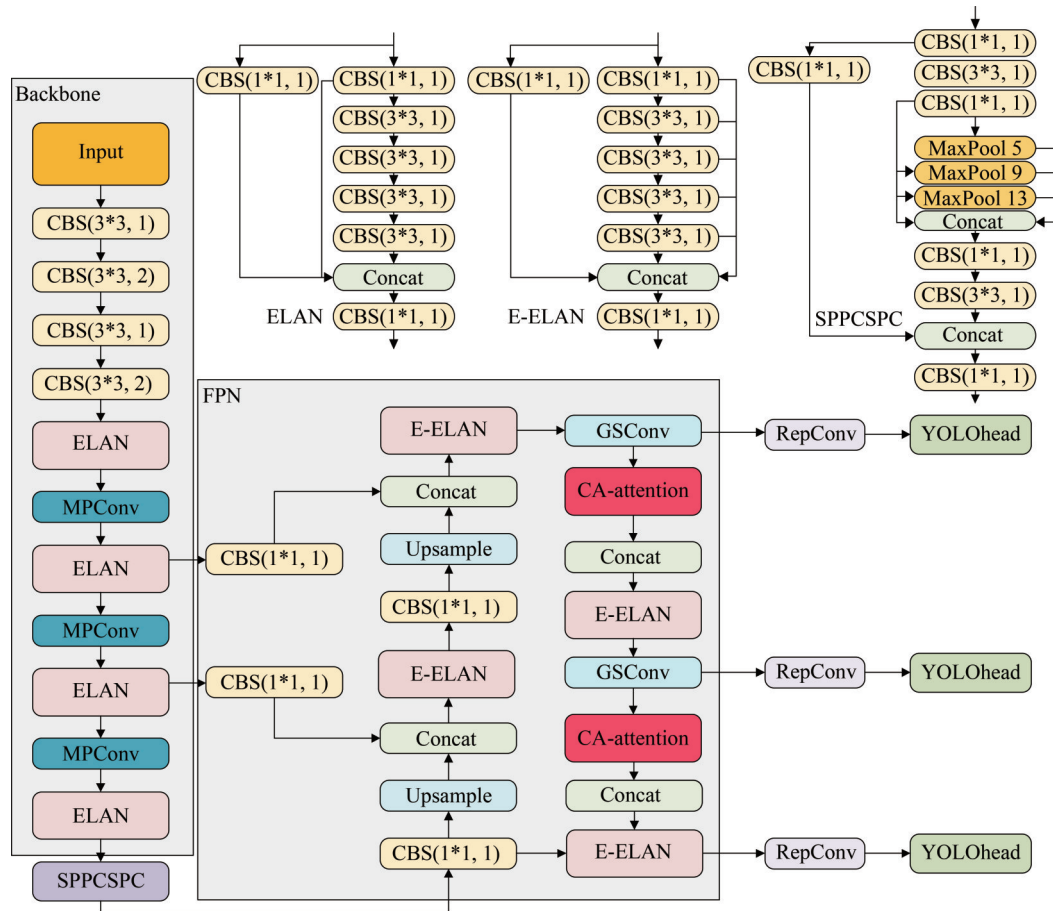


Figure 1 Structure of improved YOLOv7 model

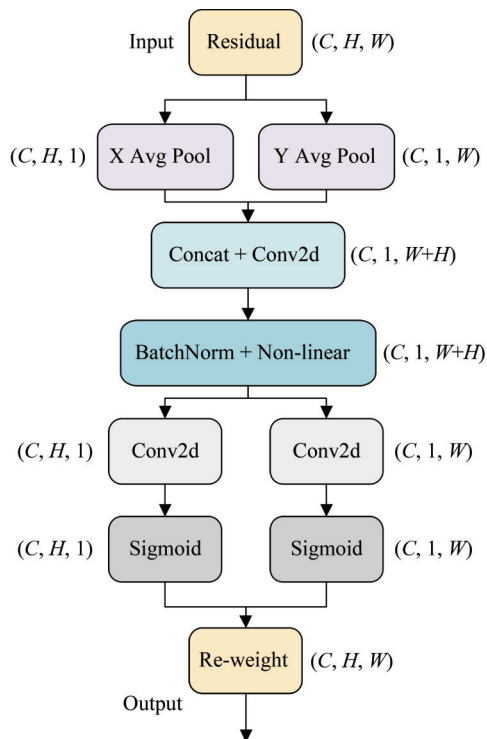


Figure 2 CA structure

and height dimensions are merged into a single dimension, achieving feature extraction and fusion in the width and height directions. The resulting feature map has a shape of $(C, 1, H+W)$. Finally, convolutional, normalization, and activation operations are applied to yield input features for subsequent networks.

$$f = \delta \left(F_1 \left([z^h, z^w] \right) \right) \quad (3)$$

where f represents the obtained feature map, δ is a non-linear activation function, and F_1 represents the convolution of 1×1 .

CA is again divided into two parallel submodules. An input feature undergoes 1×1 convolution to generate two feature maps: f^w and F_h , with sizes of $(C, 1, W)$ and $(C, H, 1)$, respectively. Subsequently, a sigmoid function is applied to the feature maps to generate attention weights for the width and height dimensions.

$$g^w = \sigma \left(F_w(f^w) \right) \quad (4)$$

$$g^h = \sigma \left(F_h(f^h) \right) \quad (5)$$

where σ represents the sigmoid function, and F_w and F_h are 1×1 convolutional layers that transform feature maps.

A multiplicative weighting operation produces an output feature map where each element is associated with an attention weight, indicating its importance in the width and height dimensions.

$$y_c(i, j) = X_c(i, j) \times g_c^w(j) \times g_c^h(i) \quad (6)$$

where $y_c(i, j)$ is the output obtained after applying attention weights to the input feature map $X_c(i, j)$, with $g_c^w(j)$ and $g_c^h(i)$ representing the attention maps computed for the width and height dimensions, respectively.

2.2 Lightweight convolution

Near-coastal vessel detection demands real-time and accurate results in complex environments. However, traditional convolutional operations incur high computational costs when processing high-resolution images, resulting in large model sizes and high complexity and hindering their deployment on resource-constrained devices (Li et al., 2024). To address this issue, the lightweight convolution GSConv is introduced in the paper. GSConv effectively reduces computational costs by dividing input feature maps into groups and performing convolution operations independently on each group. Additionally, GSConv's shuffle convolution mechanism enhances feature extraction by mixing information across channels and making it suitable for complex nearshore scenarios.

GSConv adopts a novel convolutional architecture that integrates spatial convolution (SC), depthwise separable convolution (DSC), and channel shuffle to boost model performance. By partially convolving feature maps, SC reduces computational overhead. DSC and channel shuffle enhance feature diversity through information separation and mixing. The structural difference between MPConv and GSConv is shown in Figure 3.

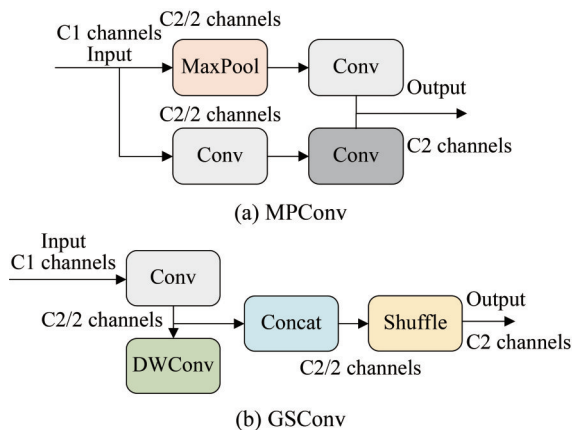


Figure 3 MPConv and GSConv structure

Compared with MPConv, which is used in the original YOLOv7, GSConv is more robust in handling image blur and feature loss caused by adverse weather conditions, such as rain and fog. Replacing MPConv with GSConv not only considerably reduces the model's computational cost and improves inference speed but also prevents an increase in the model's complexity by introducing additional network structures. In terms of model performance, GSConv effectively reduces model size and improves the accuracy of vessel detection under complex weather conditions, demonstrating its suitability for deployment on edge devices.

2.3 Loss function

The YOLOv7 network initially employs CIOU Loss as the bounding box regression loss function. CIOU considers the overlap area, center point distance, and aspect ratio of bounding boxes to measure discrepancies between predicted and ground truth boxes. However, CIOU Loss has limitations when dealing with scenes with changes in aspect ratio. Specifically, when the width and height of predicted and ground truth boxes scale proportionally, CIOU Loss cannot accurately adjust the width and height of a predicted box. To overcome this issue, EIOU Loss is adopted as a replacement for CIOU Loss. Introducing fine-grained aspect ratio constraints and Focal Loss enables EIOU Loss to handle variations in target box aspect ratios, thereby improving the model's detection accuracy in complex scenarios.

EIOU Loss consists of two components: IoU regression loss and Euclidean distance between the centers of the bounding boxes. IoU represents the ratio of the area of overlap between the predicted and ground truth boxes (A and B) to the area of their union.

$$\text{IoU} = \frac{A \cap B}{A \cup B} \quad (7)$$

CIOU Loss aligns well with the bounding box regression mechanism, considering the distance, overlap, and scale between predicted and ground truth boxes. By incorporating the center point distance and aspect ratio into the conventional IoU Loss, CIOU Loss enables accurate and stable bounding box regression.

$$L_{\text{CIOU}} = 1 - \text{IoU} + \frac{p^2(b, b^{\text{gt}})}{c^2} + av \quad (8)$$

where b and b^{gt} denote the center points of the predicted and ground truth boxes, respectively, p represents the Euclidean distance between the two center points, c represents the diagonal length of the smallest enclosing box that contains predicted and ground truth boxes, v is used in measuring the consistency of the relative scale of the two

boxes, and α represents the weight coefficient.

EIoU Loss extends CIoU Loss by decoupling the aspect ratio influence factor, enabling the separate calculation of the width and height of predicted and ground truth boxes. This modification effectively addresses false-positive and false-negative issues inherent in CIoU Loss during object detection.

$$L_{\text{EIoU}} = L_{\text{IoU}} + L_{\text{dis}} + L_{\text{asp}} \quad (9)$$

$$L_{\text{EIoU}} = 1 - \text{IoU} + \frac{p^2(b, b^{\text{gt}})}{c^2} + \frac{p^2(\omega, \omega^{\text{gt}})}{c_{\omega}^2} + \frac{p^2(h, h^{\text{gt}})}{c_h^2} \quad (10)$$

where L_{IoU} , L_{dis} , and L_{asp} represent the IoU, distance, and length losses, respectively; ω and ω^{gt} represent the width of predicted and ground truth boxes, respectively; h and h^{gt} represent the height of predicted and ground truth boxes, respectively; and c_{ω} , c_h represent the dimensions of their minimum bounding box.

3 Design and analysis of experiments

The experiments were conducted on a Windows 11 system equipped with an NVIDIA GTX 4080 GPU, an Intel i9-13900 CPU, 32 GB of RAM, CUDA 11.7, and PyTorch 1.13.1. To assess the model's convergence capability, all experiments were performed without pretrained models or frozen training. The batch size was set at 8, the number of epochs was set at 100, the input image size was 640×640, and the initial learning rate was 0.01.

3.1 Near-coastal target dataset

The SeaShip dataset was employed as the benchmark

for the experiments. This dataset comprises 7 000 images with a resolution of 1 920×1 080 and is divided into training, validation, and testing sets in 8:1:1 ratio. The SeaShip dataset encompasses six categories of vessels, including ore carriers, bulk cargo carriers, container ships, general cargo ships, fishing boats, and passenger ships. These vessels exhibit diversity in terms of type, size, and distance, providing a realistic simulation of vessel detection scenarios in near-coastal waters under clear weather conditions. Figure 4 presents sample images from the dataset.

To comprehensively evaluate the performance of object detection models in complex near-coastal vessel detection scenarios, this study leverages data augmentation techniques to simulate adverse weather conditions. Given the scarcity of samples depicting adverse weather in the SeaShip dataset, this study randomly selected 5 000 images and applied OpenCV to simulate rainy and foggy conditions by overlaying random raindrop textures and adding Gaussian noise. The augmented dataset, encompassing clear, rainy, and foggy weather conditions, more closely resembles real-world near-coastal vessel detection scenarios. The specific parameters used for data augmentation and illustrative examples of augmented images are presented in Table 1 and Figure 5.

3.2 Evaluation indicators

To evaluate the performance of the proposed model, a comprehensive set of evaluation metrics is adopted, including *F1*-score (*F1*), average precision (AP), mean average precision (mAP), precision (*P*), recall (*R*), and frames per second (FPS).

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (11)$$

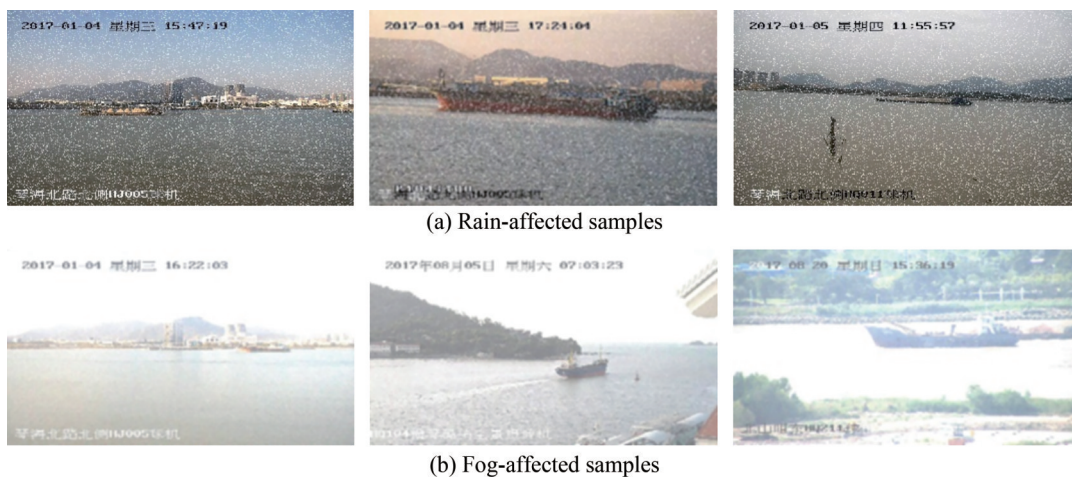
$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (12)$$



Figure 4 Samples of the SeaShip dataset

Table 1 Data enhancement parameter settings

Parameter	Affected variables	Value range	Value
Num drops	Number of raindrops	50–150	Image shape[0]/35
Drop length	Raindrop length	10–30 pixels	Image shape[1]/30
Colors	Raindrop color	List of RGB elements	(200, 200, 200), (150, 150, 150)
Loc	Mist brightness mean	0–255	128
Scale	Mist standard deviation	10–30	10

**Figure 5** Processed image dataset

$$AP = \int P(R) dR \quad (13)$$

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (14)$$

$$mAP = \frac{1}{n} \sum_{j=1}^n AP \quad (15)$$

where TP, FN, and FP represent true positives, false negatives, and false positives, respectively; n denotes the number of classes, and $\sum_{j=1}^n AP$ signifies the sum of average precision across all classes.

3.3 Ablation study

To prevent the influence of data augmentation on the model, experiments were conducted on the original SeaShip dataset with various methods. The experimental results are shown in Table 2 and Figure 6.

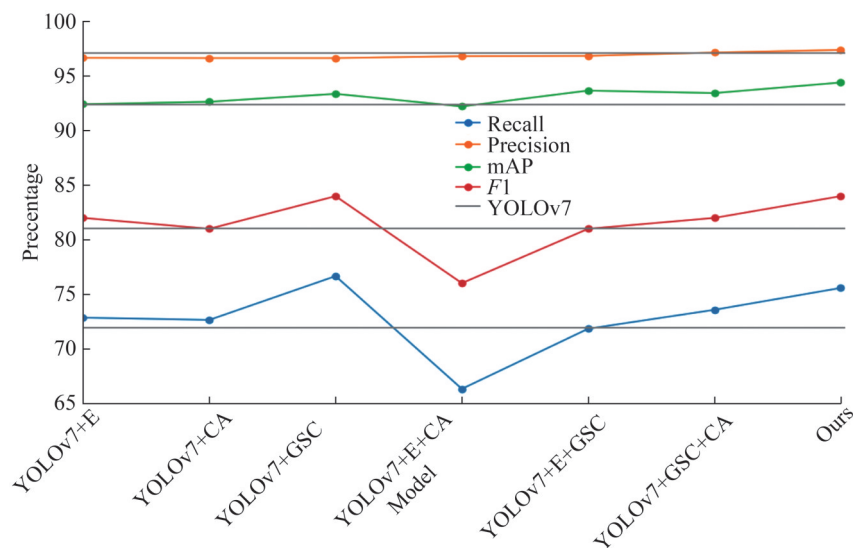
As shown in Table 2 and Figure 6, replacing the convolution layers with GSConv successfully reduced the model size, decreasing the number of parameters and computa-

tional cost. CA enhanced the model's feature extraction capability, especially for certain categories of vessel features. However, this effect leads to a lack of considerable improvement in overall performance because of the imbalance in weight among different classes. Furthermore, the improved EIou Loss and CA may be in conflict with optimization objectives, hindering joint optimization and decreasing the recall rate. CA focuses on feature interactions across channel and spatial dimensions, whereas GSConv reduces the number of parameters through grouped convolutions, thereby enhancing feature extraction. The combination of GSConv and CA enables effective feature extraction, improving the model's representational capacity.

Compared with the original model, the proposed model achieves an mAP of 94.45%, representing a 2.03% improvement relative to the baseline. Moreover, the model exhibits a 3.68% increase in recall and 0.26% increase in precision while reducing the number of parameters by 0.74 M and computational cost by 0.48 G. Detection speed is maintained at approximately 61.65 f/s. These results demonstrate the effectiveness of the proposed improvements, a considerable increase in the accuracy of near-coastal vessel detection, and a reduction in false positives and false negatives.

Table 2 Comparative ablation study on the SeaShip dataset

Model	Parameters (M)	Computational cost (G)	<i>R</i> (%)	<i>P</i> (%)	mAP (%)	<i>F1</i> (%)	FPS (f/s)
YOLOv7	37.22	52.59	71.88	97.19	92.42	82.64	63.79
YOLOv7+EIou	37.22	52.59	72.83	96.72	92.46	83.09	63.84
YOLOv7+CA	37.28	52.60	72.62	96.70	92.68	82.94	60.90
YOLOv7+GSConv	36.41	52.10	76.65	96.70	93.41	85.51	62.87
YOLOv7+EIou+CA	37.28	52.60	66.28	96.87	92.25	78.70	62.97
YOLOv7+EIou+GSConv	36.41	52.10	71.81	96.90	93.71	82.48	62.87
YOLOv7+GSConv+CA	36.48	52.11	73.55	97.21	93.48	83.74	60.71
Ours(YOLOv7+EIou+GSConv+CA)	36.48	52.11	75.56	97.45	94.45	85.12	61.65

**Figure 6** Line chart of evaluation indicators of different models

To visually demonstrate the effectiveness of the proposed improvements in the original dataset, six categories of vessel images were selected (Figure 7).

As demonstrated in Figure 7, the improved model can accurately recognize six categories of vessel images. The addition of CA and the replacement of convolution layers with GSConv effectively enhance the model's feature extraction capability, enabling the model to detect even vessels with missing feature information. The improved model demonstrates exceptional overall performance in detecting the six categories of vessel targets, demonstrating the effectiveness of the proposed algorithm improvements.

The effectiveness of the improved model was validated on the original SeaShip dataset. To further investigate the impact of data augmentation, ablation experiments were conducted on the augmented SeaShip dataset. The results are presented in Table 3.

The improved model achieves an *F1* score of 12.5%

higher than that of the original YOLOv7 model on the augmented dataset, with an mAP of 88.08%. Among the six classes of vessels, only the AP values of container and passenger ships are respectively 0.04% and 2.84% lower than the model with the improved CA and GSConv. The AP values of bulk cargo carriers, fishing boats, general cargo ships, and ore carriers respectively increase by 1.88%, 5.09%, 2.15%, and 3.12% and reach 86.80%, 92.54%, 92.06%, and 72.66%. These results show that the improved model exhibits excellent detection performance under complex weather conditions, indicating its high practical value.

3.4 Comparison experiment

To validate the superiority of the improved model, eight target detection models were evaluated on the augmented dataset. The results are presented in Table 4.



Figure 7 Comparison of six types of vessel detection effects

Table 3 Comparison of ablation experiments on SeaShip data set after data augment

Model	AP (%)						mAP (%)	F1 (%)
	Bulk cargo carriers	Container ships	Fishing boats	General cargo ships	Ore carriers	Passenger ships		
YOLOv7	80.71	97.37	87.46	90.65	72.99	72.65	83.64	64.16
YOLOv7+EIoU	82.02	98.14	88.94	91.74	79.72	73.81	85.73	54.11
YOLOv7+CA	82.37	95.15	87.39	90.65	77.00	55.98	81.42	65.54
YOLOv7+GSConv	80.28	98.13	85.41	91.99	76.07	58.86	81.79	68.64
YOLOv7+EIoU+CA	79.95	94.88	86.58	92.89	74.74	62.64	81.95	64.16
YOLOv7+EIoU+GSConv	83.09	97.39	88.20	92.31	78.30	75.53	85.80	71.84
YOLOv7+GSConv+CA	84.92	99.70	87.45	89.91	79.54	77.60	86.52	72.80
Ours(YOLOv7+EIoU+GSConv+CA)	86.80	99.66	92.54	92.06	82.66	74.76	88.08	76.66

Table 4 Comparison of different target detection models

Model	Parameters (M)	Computational cost (G)	R (%)	F1 (%)	mAP (%)	FPS (f/s)
Faster R-CNN	137.07	201.07	86.79	64.62	80.29	22.83
YOLOv3	61.55	77.67	61.91	68.60	77.14	51.94
YOLOv4	63.96	70.99	79.60	93.83	86.60	52.65
YOLOv5	46.65	57.32	53.11	68.30	84.87	87.17
YOLOv7	37.22	52.59	48.56	64.16	80.78	63.79
YOLOv8-n	3.02	4.10	45.42	58.89	83.04	112.68
YOLOv8-s	11.14	14.33	49.03	63.83	84.69	87.27
Ours	36.48	52.11	64.4	76.66	88.08	61.65

As shown in Table 4, the proposed model achieves an mAP of 88.08%, outperforming the other seven object detection models. The model exhibits significantly lower model complexity with 36.48 M parameters and 52.11 G computational cost, approximately 1/4 and 1/2 of Faster R-CNN and YOLOv4, respectively. Moreover, the model achieves a detection speed of 61.65 f/s, which is 270% and 117% faster than Faster R-CNN and YOLOv4, respectively. Thus, the model is suitable for real-time vessel detection in complex near-coastal environments.

To visually demonstrate the detection performance of different models on near-coastal vessels, a selection of detection results is presented from each model in Figure 8. Each group of images shows the detection results under clear,

rainy, and foggy weather conditions.

As shown in Figure 8, Faster R-CNN exhibits relatively low detection accuracy for vessels under clear conditions and produces false positives in rainy and foggy weather. YOLOv3 fails to detect vessels in clear and foggy weather, and YOLOv7 misses detections in rainy weather. By contrast, YOLOv4, YOLOv5, YOLOv8-n, YOLOv8-s, and the proposed model can detect targets under the three weather conditions. Notably, the proposed model demonstrates excellent detection performance in clear, rainy, and foggy weather, exhibiting considerable advantage over the other models in vessel detection under complex weather conditions.



Figure 8 Performance comparison of various detection models

4 Conclusions

To address low detection accuracy, high missed detection rates, and false alarms in near-coastal vessel detection under complex weather conditions, this paper proposes a novel vessel detection model based on the improved YOLOv7. By incorporating the CA into the reconstruction of channel attention weights, the network is able to extract vessel target features. The replacement of MPConv in the feature learning network with the lightweight convolution algorithm GSConv improves the vessel localization accuracy, generalization, and robustness of the model. Moreover, EIoU Loss is adopted to replace CIoU Loss in the original YOLOv7 to enhance the accuracy of prediction box localization and reduce missed detection and false alarm rates. Experimental results demonstrate that the proposed model can accurately detect vessels under complex weather conditions and outperforms other methods on the original SeaShip dataset and data-augmented SeaShip dataset. However, detection accuracy for small vessels should be further improved. Future research will concentrate on refining prediction box localization and optimizing model architecture to enhance detection performance for small vessels.

Competing interest The authors have no competing interests to declare that are relevant to the content of this article.

References

- Del-Rey-Maestre N, Mata-Moya D, Jarabo-Amores MP, Gomez-del-Hoyo PJ, Barcena-Humanes JL (2018) Artificial intelligence techniques for small boats detection in radar clutter. *Real data validation. Engineering Applications of Artificial Intelligence* 67: 296-308
- Elvidge CD, Zhizhin M, Baugh K, Hsu FC (2015) Automatic boat identification system for VIIRS low light imaging data. *Remote Sensing* 7(3): 3020-3036
- Girshick R (2015) Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 1440-1448. DOI: 10.1109/ICCV.2015.169
- Guo Y, Yu H, Ma L, Zeng L, Luo X (2023) THFE: A triple-hierarchy feature enhancement method for tiny boat detection. *Engineering Applications of Artificial Intelligence* 123: 106271. <https://doi.org/10.1016/j.engappai.2023.106271>
- He K, Pan Z, Zhao W, Wang J, Wan D (2024) Overview of research progress on numerical simulation methods for turbulent flows around underwater vehicles. *Journal of Marine Science and Application* 23(1): 1-22. DOI:10.1007/s11804-024-00403-8
- Hou Q, Zhou D, Feng J (2021) Coordinate attention for efficient mobile network design. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 13713-13722
- Huang Z, Jiang X, Wu F, Fu Y, Zhang Y, Fu T, Pei J (2023) An improved method for ship target detection based on YOLOv4. *Applied Sciences* 13(3): 1302
- Kanjir U, Greidanus H, Oštir K (2018) Vessel detection and classification from spaceborne optical images: A literature survey. *Remote Sensing of Environment* 207: 1-26
- Kong Z, Cui Y, Xiong W, Xiong Z, Xu P (2022) Ship target recognition based on context-enhanced trajectory. *ISPRS International Journal of Geo-Information* 11(12): 584
- Li B, Xie X, Wei X, Tang W (2021a) Ship detection and classification from optical remote sensing images: A survey. *Chinese Journal of Aeronautics* 34(3): 145-163
- Li H, Deng L, Yang C, Liu J, Gu Z (2021b) Enhanced YOLO v3 tiny network for real-time ship detection from visual image. *IEEE Access* 9: 16692-16706
- Li H, Li J, Wei H, Liu Z, Zhan Z, Ren Q (2024) Slim-neck by GSConv a lightweight-design for real-time detector architectures. *Journal of Real-Time Image Processing* 21(3): 62
- Liu RW, Yuan W, Chen X, Lu Y (2021) An enhanced CNN-enabled learning method for promoting ship detection in maritime surveillance system. *Ocean Engineering* 235: 109435
- Liu W, Angelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) SSD: Single shot multibox detector. In *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, Netherlands, 21-37*
- Qi L, Li B, Chen L, Wang W, Dong L, Jia X, Huang J, Ge C, Xue G, Wang D (2019) Ship target detection algorithm based on improved faster R-CNN. *Electronics* 8(9): 959
- Ramsay W, Fridell E, Michan M (2023) Maritime energy transition: Future fuels and future emissions. *Journal of Marine Science and Application* 22(4): 681-692. DOI:10.1007/s11804-023-00369-z
- Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 779-788*
- Redmon J, Farhadi A (2017) YOLO9000: better, faster, stronger. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA, 7263-7271*
- Redmon J, Farhadi A (2018) YOLOv3: An incremental improvement. *ArXiv Preprint, arXiv:1804.02767*
- Ren S, He K, Girshick R, Sun J (2016) Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39(6): 1137-1149
- Steccanella L, Bloisi DD, Castellini A, Farinelli A (2020) Waterline and obstacle detection in images from low-cost autonomous boats for environmental monitoring. *Robotics and Autonomous Systems* 124: 103346
- Tsuda ME, Miller NA, Saito R, Park J, Oozeki Y (2023) Automated VIIRS boat detection based on machine learning and its application to monitoring fisheries in the East China Sea. *Remote Sensing* 15(11): 2911
- Wang C, Pei J, Luo S, Huo W, Huang Y, Zhang Y, Yang J (2023a) SAR ship target recognition via multiscale feature attention and adaptive-weighted classifier. *IEEE Geoscience and Remote Sensing Letters* 20: 1-5
- Wang CY, Bochkovskiy A, Liao HYM (2023b) YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 7464-7475*
- Wang F, Yang X, Zhang Y, Yuan J (2020) Ship target detection algorithm based on improved YOLOv3. 3rd International Conference on Big Data Technologies, Qingdao, China, 162-166
- Wu W, Li X, Hu Z, Liu X (2023) Ship detection and recognition based on improved YOLOv7. *Comput. Mater. Contin.* 76(1):

- 489-498
- Xu H, Guedes Soares C (2023) Review of path-following control systems for maritime autonomous surface ships. *Journal of Marine Science and Application* 22(2): 153-171. DOI: 10.1007/s11804-023-00338-6
- Zhang D, Zhan J, Tan L, Gao Y, Župan R (2021) Comparison of two deep learning methods for ship target recognition with optical remotely sensed data. *Neural Computing and Applications* 33: 4639-4649
- Zhang X, Xu Z, Qu S, Qiu W, Di Z (2022) Recognition algorithm of marine ship based on improved YOLOv5 deep learning. *Journal of Dalian Ocean University* 37(5): 866-872. (in Chinese)
- Zhou J, Jiang P, Zou A, Chen X, Hu W (2021) Ship target detection algorithm based on improved YOLOv5. *Journal of Marine Science and Engineering* 9(8): 908
- Zhu X, Lyu S, Wang X, Zhao Q (2021) TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. 2021 IEEE International Conference on Computer Vision (ICCV), Montreal, Canada, 2778-2788