

Article

High-Efficiency and High-Precision Ship Detection Algorithm Based on Improved YOLOv8n

Kun Lan ¹ , Xiaoliang Jiang ^{1,*}, Xiaokang Ding ¹, Huan Lin ¹ and Sixian Chan ²¹ College of Mechanical Engineering, Quzhou University, Quzhou 324000, China; 36116@qzc.edu.cn (K.L.); dingxk@qzc.edu.cn (X.D.); linhuan_design@zju.edu.cn (H.L.)² College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China; sxchan@zjut.edu.cn

* Correspondence: jxl_qzu@qzc.edu.cn

Abstract: With the development of the intelligent vision industry, ship detection and identification technology has gradually become a research hotspot in the field of marine insurance and port logistics. However, due to the interference of rain, haze, waves, light, and other bad weather, the robustness and effectiveness of existing detection algorithms remain a continuous challenge. For this reason, an improved YOLOv8n algorithm is proposed for the detection of ship targets under unforeseen environmental conditions. In the proposed method, the efficient multi-scale attention module (C2f_EMAM) is introduced to integrate the context information of different scales so that the convolutional neural network can generate better pixel-level attention to high-level feature maps. In addition, a fully-concatenate bi-directional feature pyramid network (Concatenate_FBiFPN) is adopted to replace the simple superposition/addition of feature map, which can better solve the problem of feature propagation and information flow in target detection. An improved spatial pyramid pooling fast structure (SPPF2+1) is also designed to emphasize low-level pooling features and reduce the pooling depth to accommodate the information characteristics of the ship. A comparison experiment was conducted between other mainstream methods and our proposed algorithm. Results showed that our proposed algorithm outperformed other models by achieving 99.4% of accuracy, 98.2% of precision, 98.5% of recall, 99.1% of mAP@.5, and 85.4% of mAP@.5:.95 on the SeaShips dataset.

Keywords: ship detection; YOLOv8; multi-scale attention; bi-directional; spatial pyramid pooling fast**MSC:** 68T45

Citation: Lan, K.; Jiang, X.; Ding, X.; Lin, H.; Chan, S. High-Efficiency and High-Precision Ship Detection Algorithm Based on Improved YOLOv8n. *Mathematics* **2024**, *12*, 1072. <https://doi.org/10.3390/math12071072>

Academic Editors: Xiangtao Zheng, Jinchang Ren and Ling Wang

Received: 17 March 2024

Revised: 31 March 2024

Accepted: 1 April 2024

Published: 2 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As an important part of human comprehensive transportation, water transportation has become the main mode of bulk cargo transportation because of its advantages of low cost, small pollution, large loading capacity, and long average transport distance. However, with the increasing number of ships, with the increasingly complex water transport environment, ship collision, illegal escape, overloading, illegal berthing, and other accidents occurring frequently, great threats might consequently be caused to social economy, ecological, environmental protection, and people's property safety. At present, in real ocean surveillance, due to the huge amount of video information, manual monitoring can no longer meet the needs of ship detection. In recent years, with the rapid development of GPU and artificial intelligence, researchers have applied image processing technology to the field of maritime supervision for ship detection and identification. Different from land traffic, the existing ship detection methods are prone to errors or neglect because of background interference such as waves and clutters, as well as the scale change of different ships. Therefore, effective monitoring and positioning of ships in complex situations is an important challenge that urgently needs to be addressed in the field of computer vision.

Currently, traditional ship detection algorithms are mainly classified into background modeling [1,2], temporal difference [3,4], optical flow [5,6], and template matching [7,8]. Most of these algorithms are bottom-up and differentiate targets from background regions by designing a large number of image-related features (such as grayscale, texture, edges, etc.), resulting in generally unsatisfactory performance in a complex environment. In the past decade, research of convolutional neural networks has been thriving throughout various fields, and a large number of object detection algorithms based on deep learning have emerged, such as Faster R-CNN [9], SSD [10], and YOLO [11]. Inspired by these technologies, Lin et al. [12] proposed an improved Faster R-CNN network, which further increases detection performance by using squeeze and excitation mechanisms. Later, considering the dense distribution and different size constraints, Zhou et al. [13] proposed an R-Libra CNN approach that could better integrate higher-level feature semantic information with lower-level features. Based on the principle of relay amplification, Li et al. [14] proposed a lightweight Faster R-CNN, which has better real-time performance. Similarly, Wen et al. [15] proposed a multi-scale SSD network, which effectively enhanced the feature representation capability and detection efficiency. Yang et al. [16] optimized the prediction frame suppression strategy of SSD to effectively solve the problem of occlusion or missed detection of small targets. Compared with Faster R-CNN and SSD, YOLO is the first single-stage target detection method, which divides the input image into grids according to the size of the feature map and sets different numbers of anchor boxes for each grid, so it has better real-time performance. Subsequently, researchers added a new feature extraction network, straight-through layer, feature pyramid, and other operations to the original YOLO and proposed Lira-YOLO [17], LMO-YOLO [18], LK-YOLO [19] and a series of improved versions for ship detection, which further improved target detection accuracy and reduced reasoning time.

Although the abovementioned algorithms have achieved fair results in ship detection, they still face challenges in practical application, mainly including (1) interference of rain, haze, wave, illumination, and other bad weather; (2) diversification of dimensions such as size and aspect ratio; (3) the complexity of orientation, position, posture, and other factors; (4) interference of background areas such as floating objects, shorelines, buildings. As shown in Figure 1, under these conditions, the detection accuracy of ship targets would be greatly affected.

In order to address these problems, we propose an improved YOLOv8n network to better detect ships from numerous misleading targets. In our algorithm, there are three advanced modules: efficient multi-scale attention module, fully connected weighted feature fusion mechanism, and spatial pyramid pooling fast structure. The combination of these modules has efficient processing performance, and it can achieve 99.4% accuracy, 98.2% precision, 98.5% recall, 99.1% mAP@.5, and 85.4% mAP@.5:.95. The contributions of this study are shown as follows:

1. The proposed efficient multi-scale attention module (C2f_EMAM) achieves the fusion of contextual information at different scales and significantly improves the attention of high-level feature maps;
2. The fully-concatenate bi-directional feature pyramid network (Concatenate_FBiFPN) is used to optimize the classification and regression structure to better solve the problem of feature propagation and information flow in target detection;
3. The spatial pyramid pooling fast structure (SPPF2+1) is redesigned to emphasize the low-level pooling features and learn the target features more comprehensively;
4. The ablation study of C2f_EMAM, Concatenate_FBiFPN, and SPPF2+1 is conducted, and the experimental results verified the feasibility and effectiveness of these modules.

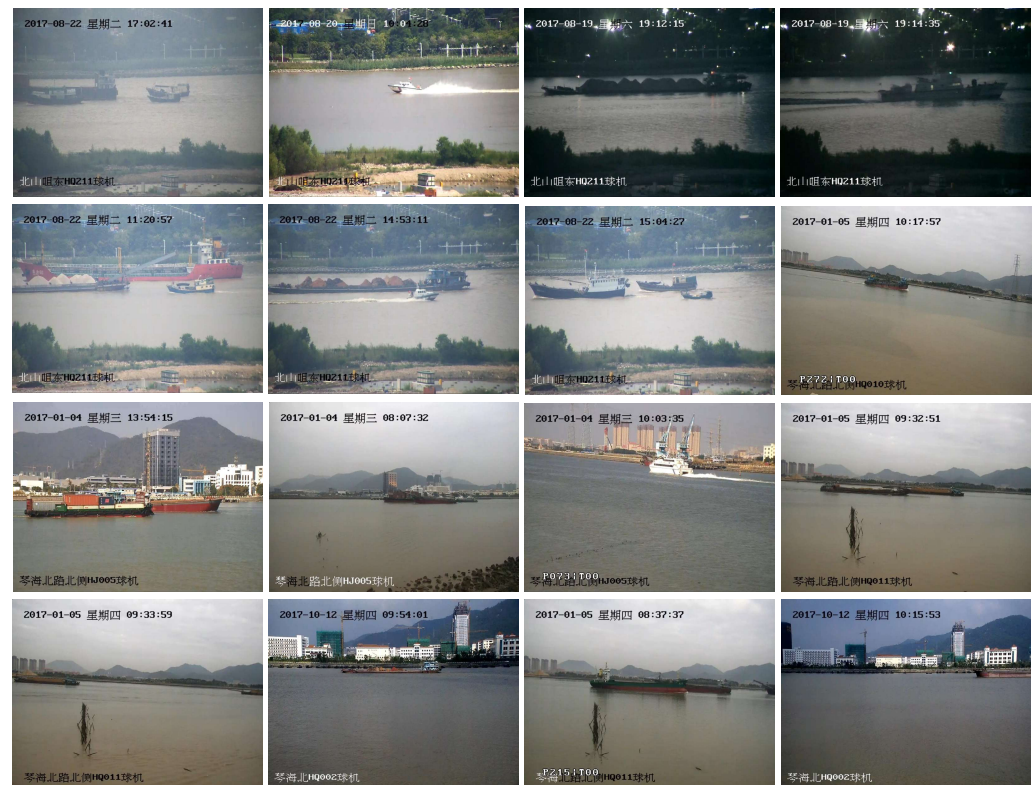


Figure 1. Challenges and limitations of ship detection. **The first row:** interference of rain, haze, wave, illumination and other bad weather; **The second row:** diversification of dimensions such as size and aspect ratio; **The third row:** the complexity of orientation, position, posture and other factors; **The last row:** interference of background areas such as floating objects, shorelines, buildings.

2. Materials and Methods

2.1. Overview of the Improved YOLOv8n

YOLOv1 (You Only Look Once) [11] is the initially developed version of the YOLO series, which was proposed by Redmon et al. in 2015. Compared with traditional target detection methods, it essentially transforms the target border positioning problem into a regression problem, which utilizes the whole graph as the network input to directly return the position and category of the bounding box in the output layer. More recently, YOLOv5 [20] and YOLOv7 [21] have become two of the most trending algorithms in the YOLO series. The main difference between them is that YOLOv5 is a lightweight target detection model compared to the more complex architecture of YOLOv7; it adopts the backbone network structure based on the feature pyramid network and the anchor-free detection mode. This method reduces the computational complexity and the number of parameters and greatly enhances the speed and accuracy. As a new training strategy, YOLOv7 adopts a deeper network structure and introduces some new technical means, such as the bottleneck attention module, which can significantly improve the accuracy and generalization ability of the target detector. In 2023, the Ultralytics team launched the latest version of YOLOv8 (Ultralytics, Frederick, MD, USA) [22]. By optimizing the C3 module in YOLOv5, removing the convolutional structure in the up-sampling stage, and using a new loss strategy, YOLOv8 has a faster detection speed and can detect and locate the target object more accurately.

Because of these advantages, we chose the lightweight version YOLOv8n (Ultralytics, Frederick, MD, USA) as the baseline, which mainly includes feature extraction modules (backbone and neck networks), a recognition branch, and a detection branch. The network architecture of the improved YOLOv8n is shown in Figure 2. Firstly, replacing the original C2f convolutional with C2f_EMAM lightweight structure can aggregate multi-scale spa-

tial structure information; thereby, the convolutional neural network can generate better pixel-level attention to advanced feature maps. Secondly, using an efficient bi-directional connection and weighted feature fusion strategy, the neck network is improved to Concatenate_FBiFPN so that the feature information can be propagated in both top-down and bottom-up directions. Finally, the SPPF structure has been improved by using two Max-pool 5 in parallel and performing secondary pooling on one of the pooling features. This has attracted more attention to low-level pooling features while also taking into account different receptive field features. The introduction details of each module are as follows.

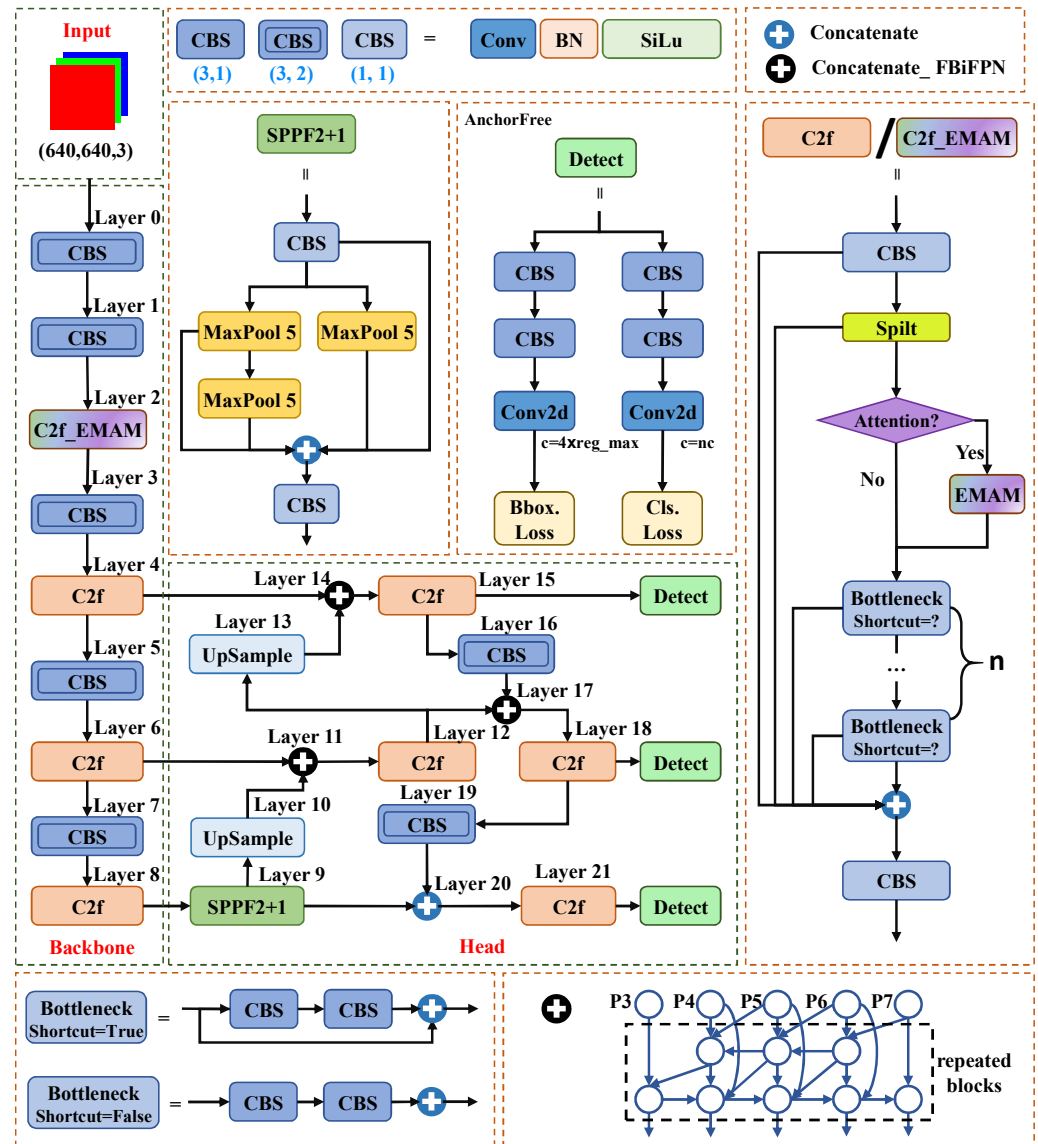


Figure 2. The network architecture of the improved YOLOv8n.

2.2. Architecture of C2f_EMAM

According to the literature [22], the CSP structure of YOLOv5 was replaced by the C2f structure in YOLOv8, which added more residual connections and thus enriched the gradient flow. However, due to different ship types and complex weather disturbances, the robustness and effectiveness of the YOLOv8 algorithm remain problematic. In deep learning and computer vision, the proposal of attention mechanisms is proven to be a very effective way to help models focus on important parts of the input. However, the traditional attention mechanism often works in a certain spatial scope and cannot obtain information from the global context. Inspired by Ouyang et al. [23], we add an efficient

multi-scale attention module to C2f and propose a new C2f_EMAM structure. Specifically, C2f_EMAM performs a convolution operation on input features, followed by the split function and EMAM, and then enters the last half into a bottleneck block and finally concatenates it through a convolution output; the overall structure of C2f_EMAM is shown in Figure 3. By fusing attention mechanisms, convolution nuclei of different sizes, and grouping convolution, C2f_EMAM aims to improve model performance while maintaining computational efficiency.

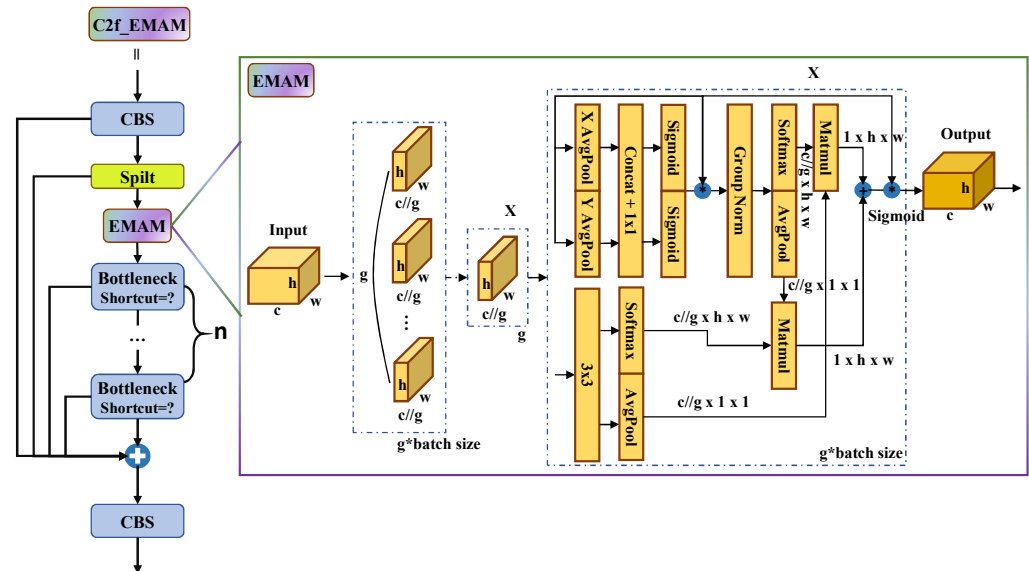


Figure 3. Network architecture of C2f_EMAM.

2.3. Architecture of Concatenate_FBiFPN

Concatenate_FBiFPN is a unique feature pyramid network that combines the bidirectional cross-scale connections and weighted feature fusion strategies of FPN, PANet, and BiFPN, as shown in Figure 4. The feature information can be propagated from top to bottom and connected efficiently according to certain rules, which can accomplish the aggregation of deep and shallow features easily and quickly. FPN uses a simple top-down approach to combine multi-scale features, while PANet adds a simple bottom-up secondary fusion path on the basis of FPN. On the basis of PANet, BiFPN removes nodes with only one input edge to simplify the network and adds an extra path from the original input to the output node on the same horizontal route. Since different input feature maps have different effects on the fused multi-scale feature maps, the Concatenate_FBiFPN structure is the solution to tackle this problem. By incorporating more features without increasing the number of parameters and computational cost, Concatenate_FBiFPN adds a skip connection between some input nodes and output nodes to achieve adequate multi-scale feature information fusion.

2.4. Architecture of SPPF2+1

The SPPF module is the spatial pyramid pooling module used in YOLOv5. Its mechanism is to pool feature maps of different scales without changing the size of the feature maps so as to improve the accuracy of receptor field extraction and target detection. The original SPPF is pooled three times layer by layer and concatenated at the end. Although shallow semantic features are considered in the final concatenate, the concatenated features are weakened. Considering that the shallow semantics contain more image features, we parallel two Maxpools 5 and only perform secondary pooling on one of the first pooling and concatenate the various features to obtain SPPF2+1. In this way, more attention is paid to the pooling characteristics of the lower layers, and different receptive field characteristics are also taken into account, as shown in Figure 5.

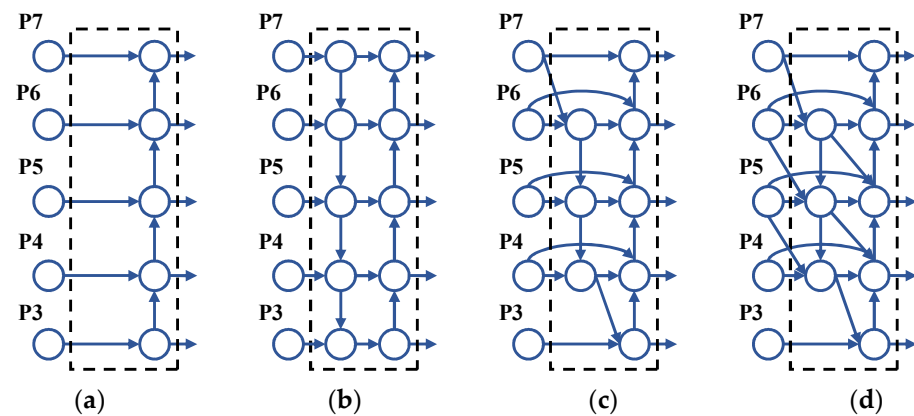


Figure 4. The network architecture of Concatenate_FBiFPN: (a) FPN; (b) PANet; (c) BiFPN; (d) Concatenate_FBiFPN.

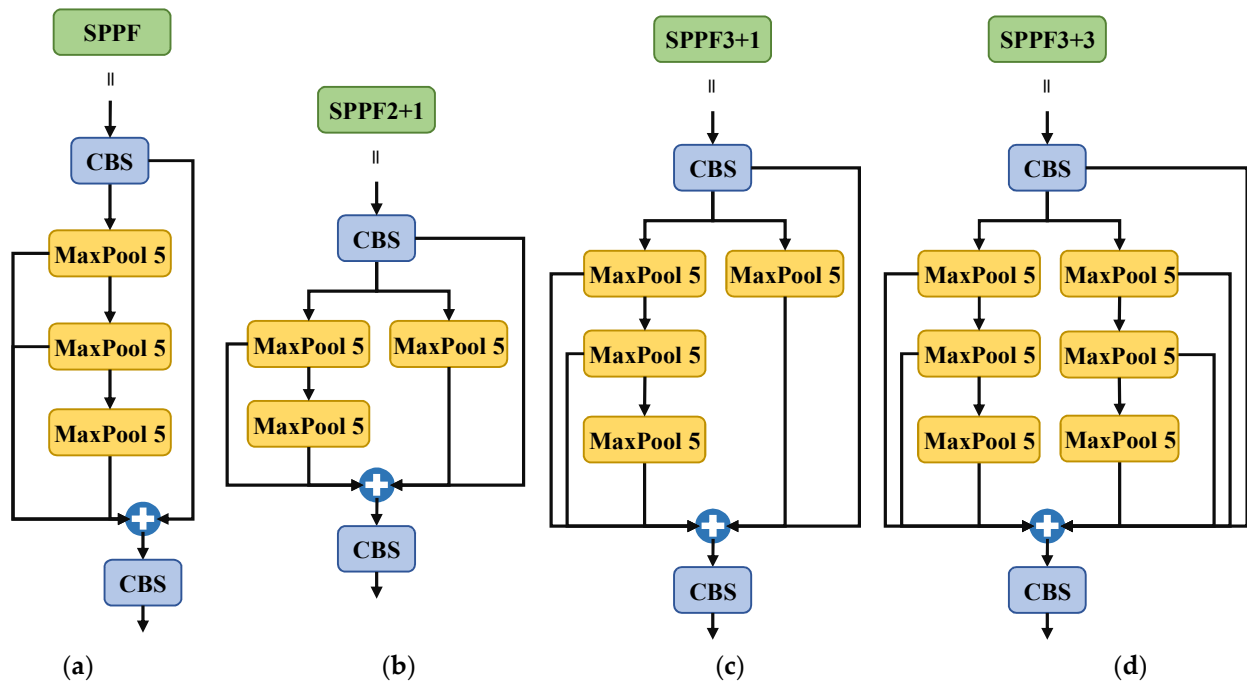


Figure 5. Network architecture of SPPF series: (a) SPPF; (b) SPPF2+1; (c) SPPF3+1; (d) SPPF3+3.

3. Results

The convolutional neural network is needed for training object detection models, so it is necessary to build an operating environment based on the Pytorch deep learning framework. In the Windows 10 environment, we use 24GB NVIDIA RTX6000 GPU memory and support CUDA to accelerate the model during the training process. In the training process, Adam Optimizer [24] and Tensorboard were used as a training visualization tool to visualize the training state of the model based on the Pytorch framework by reading the generated training log. The batch size was adjusted to 16, the learning rate was 0.001, and the number of iterations was 300. We used the publicly available SeaShips [25] dataset to verify the effect of the object detection methods. The images in the dataset were divided into three parts by the ratio of 3:1:1, which were respectively used as the training set, validation set, and test set. The data set contains a total of six classes: ore carrier, bulk cargo carrier, general cargo ship, container ship, fishing boat, and passenger ship, which can generally cover the main types of ships encountered at sea.

3.1. Evaluation Metrics

To evaluate the performance of these methods, accuracy, precision [26], recall [27], $mAP@.5$ [28], and $mAP@.5:.95$ [29] are used; equations are as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$AP = \int_0^1 P(R) dR \quad (4)$$

$$mAP@.5 = \frac{\sum_{i=1}^N AP_i}{N}, \quad IoU = 0.5 \quad (5)$$

$$mAP@.5 : .95 = \frac{\sum_{i=1}^N AP_i}{N}, \quad IoU \in [0.5, 0.95] \quad (6)$$

where TP and TN are the counts of positive and negative pixels that are correctly classified. FP and FN are the counts of misclassified positive and negative pixels. $mAP@.5$ means that when IoU is set to 0.5, the AP of all images in each category is calculated, and then all categories are averaged. $mAP@.5:.95$ represents the average mAP across different IoU thresholds ranging from 0.5 to 0.95 with steps of 0.05.

3.2. Results of the Improved YOLOv8n

As shown in Figure 6, the loss of the training set and verification set and the change of evaluation index in the training process of our method were illustrated. Observing Figure 6a–c, after 200 epochs, although the training set index showed a continuous downward trend, the loss of the validation set had become flat, which indicates that the model reached a convergence state within 300 epochs. In Figure 6d, it is shown that although precision and recall continue to show oscillations after 200 epochs, mAP has become flat.

Table 1 shows the detection indexes of various types of ships, and the results indicate that the detection of the proposed method is not lower than 0.988 at $mAP@.5$ for all types of detection. Figure 7 shows that the model can accurately locate and classify each type of ship.

3.3. Ablation Experiment

3.3.1. C2f_EMAM

Table 2 shows the comparison between the recognition results of C2f integrated with EMAM structure and the original C2f. The results show that the comprehensive index of C2f_EMAM is improved by 0.002 compared with the original YOLOv8n model $mAP@.5$ and $mAP@.5:.95$.

Figure 8 shows the change values of the recall of each ship type before and after the integration of EMAM, and it is found that the model integrated with EMAM improves the Recall of all types of ships.

In order to better show the role of the EMAM mechanism, we demonstrate the change of characteristic heatmap of passing the C2f_EMAM layer through Grad-CAM. As shown in the first row of Figure 9, after the EMAM structure is added, the extracted features are mostly focused around the ship, while the original structure still distributes a lot of attention on the sea surface and coast. As shown in the second line, the EMAM structure can improve the ability to search for small targets. Compared with the original structure, the EMAM has a higher attention level on the main object—the ships, on the images. However, even for some seemingly undistributed heat, the background features are suppressed away

from background areas, as shown in the last line of Figure 9. In summary, the C2f structure integrated into EMAM improves the feature extraction ability of the model.

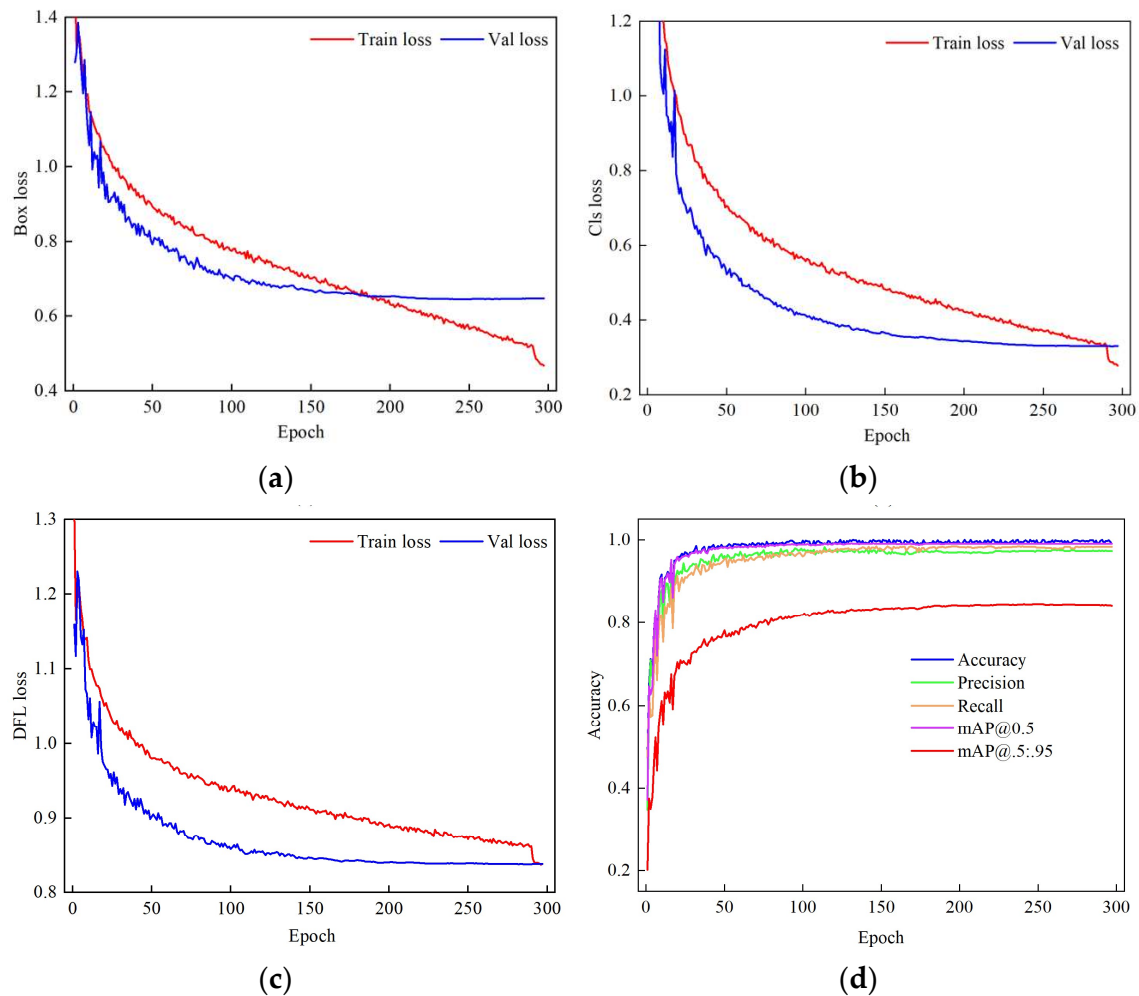


Figure 6. Each index in the training process: (a) Box loss; (b) Cls loss; (c) DFL loss; (d) Metrics.

Table 1. Performance results of the model on test set.

Ships	Accuracy	Precision	Recall	mAP@.5	mAP@.5:.95
Ore carrier	0.996	0.986	0.989	0.994	0.863
general cargo ship	0.995	0.987	0.994	0.993	0.875
bulk cargo carrier	0.992	0.968	0.979	0.989	0.873
Container ship	0.996	0.993	0.995	0.995	0.886
Fishing boat	0.992	0.970	0.965	0.989	0.801
Passenger ship	0.991	0.989	0.988	0.988	0.818
Average	0.994	0.982	0.985	0.991	0.854

3.3.2. Concatenate_FBiFPN

The evaluation metrics of three Concatenate structures are shown in Table 3, and the results show that the proposed Concatenate_FBiFPN structure effectively improves general performance of the model. Among them, mAP@.5 is increased by 0.002, mAP@.5:.95 is increased by 0.004.

Figure 10 shows the heatmap of a randomly selected testing sample image being processed by Layer 4 and Layer 13 after passing through different Concatenate structures. The results show that compared with the original Concatenate and Concatenate_BiFPN structure, Concatenate_FBiFPN has a higher attention level for the two tested objects judg-

ing from the higher density of the highlighted areas. This indicates that Concatenate_BiFPN can have heavier and more concentrated weights than the target.

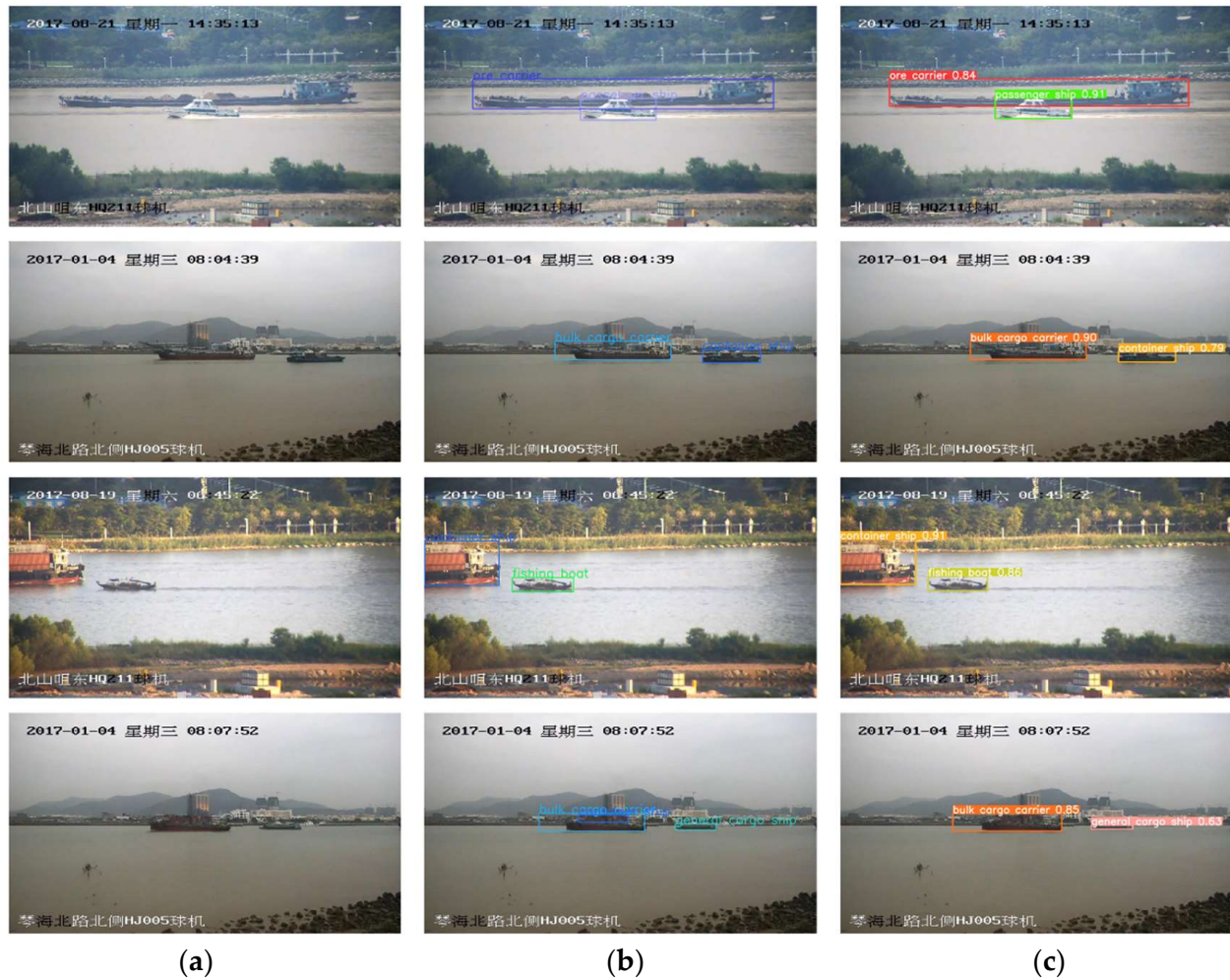


Figure 7. Detection results of the improved YOLOv8n: (a) Image; (b) Annotation; (c) Detection.

Table 2. Comparison of detection results of different C2f structures.

Structure	Accuracy	Precision	Recall	mAP@.5	mAP@.5:.95
Yolov8n-C2f	0.990	0.982	0.979	0.988	0.848
Yolov8n-C2f_EMAM	0.993	0.981	0.982	0.990	0.950

3.3.3. SPPF2+1

Table 4 shows the influence of four SPPF structures on the detection accuracy. The data shows that compared to the original SPPF structure, the SPPF2+1 structure significantly improves the detection accuracy of the model. However, after further deepening the model structure, the detection accuracy is relatively low, as shown in the test indexes of SPPF3+1 and SPPF3+3.

From the heat map shown in Figure 11, we can observe that the SPPF2+1 structure focuses highly on the detected object. However, SPPF3+1 focuses on more features in the background region, and SPPF3+3 makes the features of the tested region biased or even missing due to the excessive hybridity of features. The results show that the proposed SPPF2+1 structure can effectively adapt to ship target detection and distribute the weight of the detected area well.

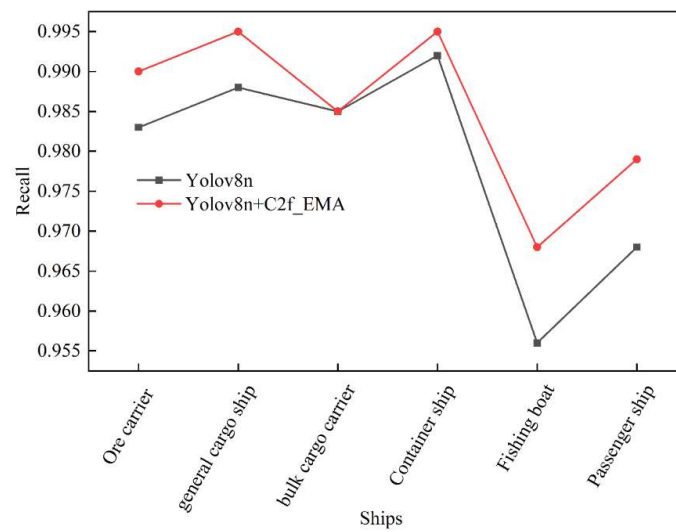


Figure 8. Influence of C2f_EMA structure on ship identification recall.

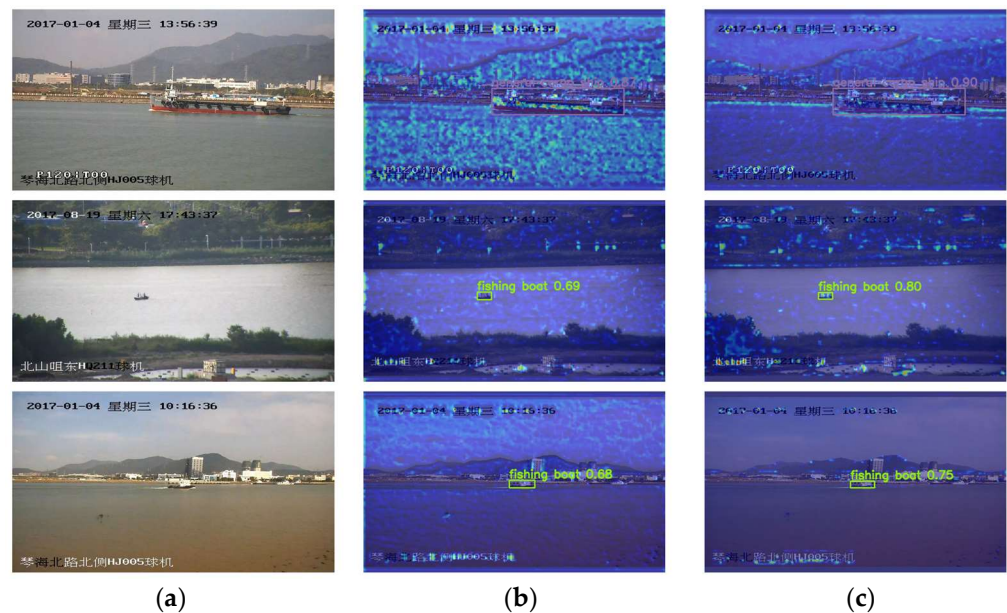


Figure 9. The C2f_EMA structure heatmap displays: (a) Image; (b) YOLOv8n-C2f; (c) YOLOv8n-C2f_EMA.

Table 3. Comparison of detection results of different Concatenate structures.

Structure	Accuracy	Precision	Recall	mAP@.5	mAP@.5:.95
YOLOv8n-Concatenate	0.989	0.982	0.979	0.988	0.848
YOLOv8n-Concatenate_BiFPN	0.991	0.981	0.980	0.989	0.851
YOLOv8n-Concatenate_FBiFPN	0.992	0.983	0.981	0.990	0.852

Table 4. Comparison of detection results of different SPPF structures.

Structure	Accuracy	Precision	Recall	mAP@.5	mAP@.5:.95
YOLOv8n-SPPF	0.989	0.982	0.979	0.988	0.848
YOLOv8n-SPPF2+1	0.992	0.980	0.982	0.990	0.951
YOLOv8n-SPPF3+1	0.987	0.979	0.978	0.985	0.837
YOLOv8n-SPPF3+3	0.988	0.978	0.979	0.985	0.838

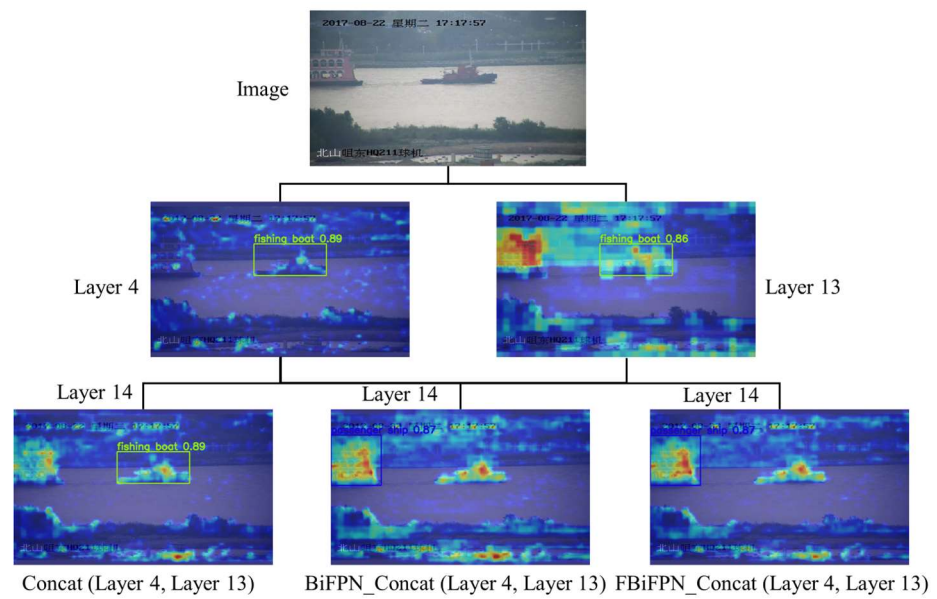


Figure 10. Heatmap display of different Concatenate structures.

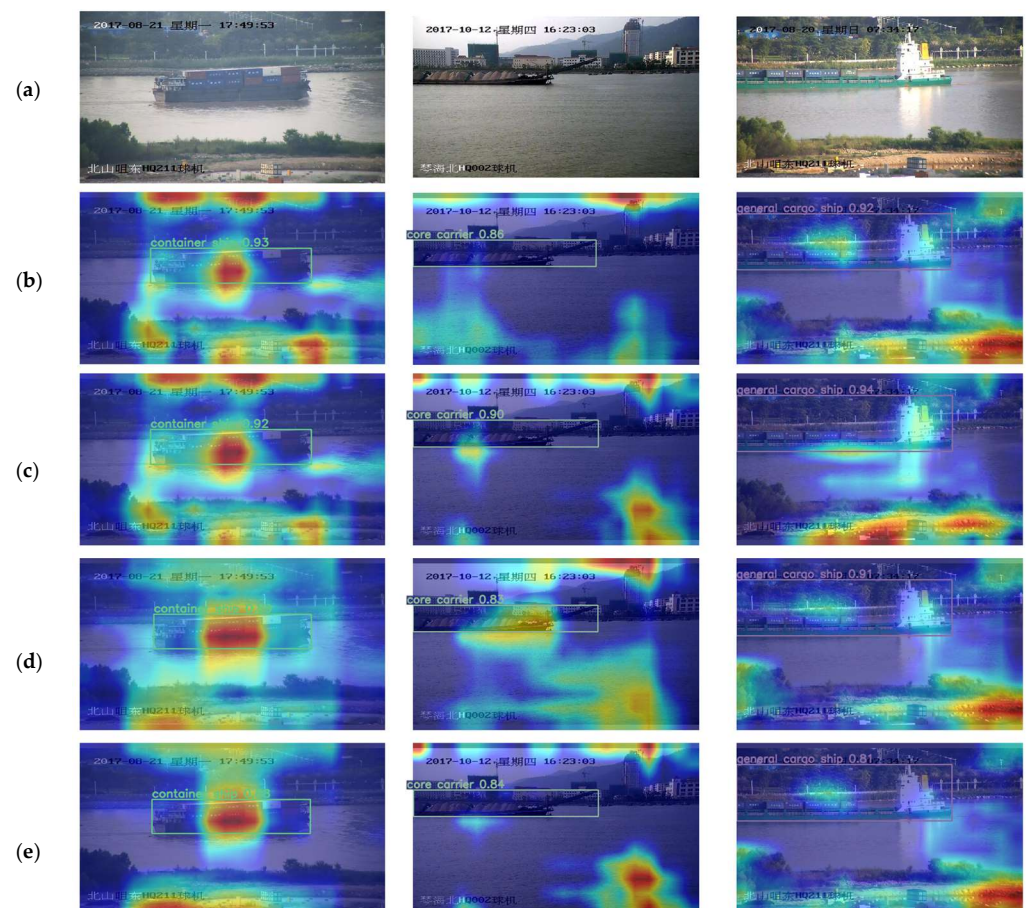


Figure 11. Heatmap display of different SPPF structures: (a) Image; (b) Yolov8n-SPPF; (c) Yolov8n-SPPF2+1; (d) Yolov8n-SPPF3+1; (e) Yolov8n-SPPF3+3.

4. Discussion

To test the performance of the model more comprehensively, our improved YOLOv8n was compared with some classic object detection methods based on the same dataset, and the results are shown in Table 5. The data shows that the mAP@.5 of our YOLOv8n gets

0.991 on the data set, which is much higher than other models. In terms of detection speed, its FPS score of 83.33 is also the highest. In addition, the number of parameters and the size of GFLOPs are kept to a minimum. The results show that our YOLOv8n on SeaShips dataset has better general performance.

Table 5. Comparison of detection results of different models.

Models	mAP@.5	mAP@.5:.95	Parameters (M)	GFLOPs (G)	FPS
Faster RCNN	0.850	0.698	39.56	101.80	20.40
EfficientDet-D1	0.950	0.761	6.67	12.43	19.62
Yolov5s	0.960	0.754	6.74	16.50	78.56
Yolov5m	0.975	0.760	20.10	50.70	64.40
Yolov5l	0.983	0.767	44.50	114.60	41.50
Yolov7	0.986	0.821	35.50	105.53	73.58
Yolov8n	0.988	0.848	2.87	8.21	83.33
Yolov8s	0.990	0.851	10.62	28.71	71.42
Yolov8m	0.991	0.854	24.66	79.1	47.62
Improved Yolov8n	0.991	0.854	2.87	8.30	83.33

5. Conclusions

This paper proposes a deep learning architecture based on YOLOv8n for the objective of ship detection under complex environmental conditions. Our conclusions are as follows: firstly, the proposed C2f_EMAM can integrate the context information of different scales and significantly improve the attention of high-level feature maps. Secondly, the fusion of Concatenate_FBiFPN into the model can better solve the problem of feature propagation and information flow in target detection. Finally, the SPPF2+1 is redesigned to emphasize the low-level pooling features and learn the target features more comprehensively. In the comparison experiments we conducted between our algorithm and other similar purpose models, our algorithm achieved 99.4% accuracy, 98.2% precision, 98.5% recall, 99.1% of mAP@.5 and 85.4% of mAP@.5:.95, which is superior to other traditional detection models. The results show that our proposed method possesses research significance in the field of automatic and intelligent ship detection.

However, this study does not fully discuss the representativeness and diversity of the datasets used, and the robustness of the model is still to be further improved. Currently, we are building a more comprehensive and diverse ship detection dataset and adopting more diverse evaluation indicators to improve the generalization ability and practicability of the model, which is also the focus of research in the future.

Author Contributions: Conceptualization, K.L. and X.J.; methodology, K.L. and X.J.; software, X.J.; validation, X.D., H.L. and S.C.; formal analysis, X.D.; investigation, H.L.; resources, S.C.; data curation, X.D.; writing—original draft preparation, K.L.; writing—review and editing, X.J.; visualization, X.D.; supervision, X.J.; project administration, X.J.; funding acquisition, S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 62102227, 62302263, Zhejiang Basic Public Welfare Research Project, grant number LTGC23E050001, LZY24E060001, ZCLTGS24E0601, Science and Technology Major Projects of Quzhou, grant number 2023K221.

Data Availability Statement: The data presented in this study are openly available in Seaships at <https://github.com/jiaming-wang/SeaShips>, accessed on 1 March 2024.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhou, A.; Xie, W.; Pei, J. Background modeling combined with multiple features in the Fourier domain for maritime infrared target detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 4202615. [\[CrossRef\]](#)
2. Zhou, A.; Xie, W.; Pei, J. Maritime infrared target detection using a dual-mode background model. *Remote Sens.* **2023**, *15*, 2354. [\[CrossRef\]](#)
3. Yoshida, T.; Ouchi, K. Improved Accuracy of Velocity Estimation for Cruising Ships by Temporal Differences Between Two Extreme Sublook Images of ALOS-2 Spotlight SAR Images with Long Integration Times. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 11622–11629. [\[CrossRef\]](#)
4. Yao, S.; Chang, X.; Cheng, Y.; Jin, S.; Zuo, D. Detection of moving ships in sequences of remote sensing images. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 334. [\[CrossRef\]](#)
5. Larson, K.M.; Shand, L.; Staid, A.; Gray, S.; Roesler, E.L.; Lyons, D. An optical flow approach to tracking ship track behavior using GOES-R satellite imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 6272–6282. [\[CrossRef\]](#)
6. Bahrami, Z.; Zhang, R.; Wang, T.; Liu, Z. An end-to-end framework for shipping container corrosion defect inspection. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 5020814. [\[CrossRef\]](#)
7. He, H.; Lin, Y.; Chen, F.; Tai, H.-M.; Yin, Z. Inshore ship detection in remote sensing images via weighted pose voting. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3091–3107. [\[CrossRef\]](#)
8. Chen, X.; Qiu, C.; Zhang, Z. A Multiscale Method for Infrared Ship Detection Based on Morphological Reconstruction and Two-Branch Compensation Strategy. *Sensors* **2023**, *23*, 7309. [\[CrossRef\]](#)
9. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [\[CrossRef\]](#)
10. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part I 14. 2016; pp. 21–37.
11. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
12. Lin, Z.; Ji, K.; Leng, X.; Kuang, G. Squeeze and excitation rank faster R-CNN for ship detection in SAR images. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 751–755. [\[CrossRef\]](#)
13. Guo, H.; Yang, X.; Wang, N.; Song, B.; Gao, X. A rotational libra R-CNN method for ship detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 5772–5781. [\[CrossRef\]](#)
14. Li, Y.; Zhang, S.; Wang, W.-Q. A lightweight faster R-CNN for ship detection in SAR images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 4006105. [\[CrossRef\]](#)
15. Wen, G.; Cao, P.; Wang, H.; Chen, H.; Liu, X.; Xu, J.; Zaiane, O. MS-SSD: Multi-scale single shot detector for ship detection in remote sensing images. *Appl. Intell.* **2023**, *53*, 1586–1604. [\[CrossRef\]](#)
16. Yang, Y.; Chen, P.; Ding, K.; Chen, Z.; Hu, K. Object detection of inland waterway ships based on improved SSD model. *Sh. Offshore Struct.* **2023**, *18*, 1192–1200. [\[CrossRef\]](#)
17. Long, Z.; Suyuan, W.; Zhongma, C.; Jiaqi, F.; Xiaoting, Y.; Wei, D. Lira-YOLO: A lightweight model for ship detection in radar images. *J. Syst. Eng. Electron.* **2020**, *31*, 950–956. [\[CrossRef\]](#)
18. Xu, Q.; Li, Y.; Shi, Z. LMO-YOLO: A ship detection model for low-resolution optical satellite imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4117–4131. [\[CrossRef\]](#)
19. Sun, S.; Xu, Z. Large kernel convolution YOLO for ship detection in surveillance video. *Math. Biosci. Eng.* **2023**, *20*, 15018–15043. [\[CrossRef\]](#)
20. Karthi, M.; Muthulakshmi, V.; Priscilla, R.; Praveen, P.; Vanisri, K. Evolution of yolo-v5 algorithm for object detection: Automated detection of library books and performace validation of dataset. In Proceedings of the 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSSES), Chennai, India, 24–25 September 2021; pp. 1–6.
21. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. 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, BC, Canada, 17–24 June 2023; pp. 7464–7475.
22. Jocher, G.; Chaurasia, A.; Qiu, J. YOLO by Ultralytics. 2023. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 1 March 2024).
23. Ouyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhan, J.; Huang, Z. Efficient multi-scale attention module with cross-spatial learning. In Proceedings of the ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
24. Ji, M.; Wu, Z. Automatic detection and severity analysis of grape black measles disease based on deep learning and fuzzy logic. *Comput. Electron. Agric.* **2022**, *193*, 106718. [\[CrossRef\]](#)
25. Shao, Z.; Wu, W.; Wang, Z.; Du, W.; Li, C. Seaships: A large-scale precisely annotated dataset for ship detection. *IEEE Trans. Multimed.* **2018**, *20*, 2593–2604. [\[CrossRef\]](#)
26. Zhang, L.; Ding, G.; Li, C.; Li, D. DCF-Yolov8: An Improved Algorithm for Aggregating Low-Level Features to Detect Agricultural Pests and Diseases. *Agronomy* **2023**, *13*, 2012. [\[CrossRef\]](#)

27. Yang, G.; Wang, J.; Nie, Z.; Yang, H.; Yu, S. A lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention. *Agronomy* **2023**, *13*, 1824. [[CrossRef](#)]
28. Li, S.; Wang, S.; Wang, P. A small object detection algorithm for traffic signs based on improved YOLOv7. *Sensors* **2023**, *23*, 7145. [[CrossRef](#)] [[PubMed](#)]
29. Chen, J.; Li, P.; Xu, T.; Xue, H.; Wang, X.; Li, Y.; Lin, H.; Liu, P.; Dong, B.; Sun, P. Detection of cervical lesions in colposcopic images based on the RetinaNet method. *Biomed. Signal Process. Control* **2022**, *75*, 103589. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.