

# An Infrared Ship Detection Algorithm based on YOLOv8n

Haiyang Qin<sup>1,2</sup>, Gongquan Tan<sup>1,2</sup>, Hao Deng<sup>1,2</sup>, Dayang Cai<sup>1,2</sup>, Yao Wang<sup>1,2</sup>, Guobin Mao<sup>1,2</sup>,

Teng Hu<sup>1,2</sup>

<sup>1</sup> School of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibin 644000, China

<sup>2</sup> Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science & Engineering, Yibin 644000, China

---

**Abstract:** With the development of infrared technology, infrared imaging technology has also been widely used in marine ship detection. However, factors such as low contrast and high noise in infrared images result in poor detection performance. This article proposes an infrared ship detection algorithm based on YOLOv8n to address this issue. Firstly, by adding a small target detection layer, the detection accuracy of small target ships has been significantly improved. Secondly, a Focaler MPDIoU loss function was designed to address the issue of imbalanced sample categories and reduce excessive attention to easily classified samples. Finally, the introduction of a lightweight V7DownSampling downsampling module further improves detection accuracy and reduces the number of model parameters and model size. The experimental results show that the improved algorithm has improved the average accuracy on the publicly available InfiRay infrared ship dataset by 3.4 percentage points compared to the original YOLOv8n, while reducing the number of parameters by 11.8% and the model size by 8%, significantly improving detection accuracy and making it easy to deploy on resource limited platforms.

**Keywords:** Infrared Ship Detection; YOLOv8n; V7DownSampling; Loss Function.

---

## 1. Introduction

Ship detection technology plays a crucial role in modern ocean monitoring and maritime safety. Especially in the context of the gradual maturity and widespread application of infrared imaging technology, it is of great practical significance to efficiently and accurately detect ships at sea using infrared images. Infrared imaging technology has the unique advantage of being able to work at night and in harsh weather conditions, which makes it play an important role in maritime monitoring and search and rescue operations. However, infrared images also have problems such as low contrast, high noise, and small differences between targets and backgrounds, which pose great challenges to ship detection.

Traditional infrared ship detection in the past mainly relied on traditional image processing techniques. Technologies such as background modeling [3], image enhancement [4], edge detection [5], etc. have to some extent solved the detection problem in specific environments. However, due to their reliance on manually designed features and fixed rules, it is difficult to capture the complex features of targets in infrared images, resulting in high rates of false positives and missed detections. With the development of deep learning technology, infrared ship detection methods based on deep learning have become an effective technical means. Especially for object detection algorithms based on Convolutional Neural Networks (CNN), such as the YOLO [6-8] (You Only Look Once) series, they achieve end-to-end object detection through a single-stage detection framework, greatly improving detection speed while maintaining high accuracy.

Zhang et al. proposed an EGISD-YOLO and designed a deconvolution channel attention module (DCA) by improving the CSP module of YOLO, effectively reducing missed detections of ships. However, its model is large and not

suitable for deployment on platforms with limited computing resources; Gu Jiaojiao et al. proposed an infrared ship target detection algorithm based on improved Faster R-CNN. By modifying the backbone network and optimizing the number and size of Anchor boxes, the accuracy has been improved to some extent, but there is still a lot of room for improvement in terms of accuracy; Miao et al. proposed a robust method for detecting ships in infrared images using multi-scale feature extraction and convolutional neural networks (CNN). The effectiveness of this method largely depends on the quality of the infrared images, which can lead to a decrease in detection performance when the image quality is compromised; Zhang Shen et al. embedded the MobileNetv3 network into the YOLOv7 backbone network and introduced SE attention mechanism, Wise IoU loss function, and bipfn feature pyramid, effectively achieving high-speed and high-precision ship detection. However, its computational and parameter requirements are still relatively large.

In summary, although some progress has been made in the current field of infrared ship detection, there are still problems such as missed detection, false detection, and large model size. Therefore, this article proposes an infrared ship detection model based on YOLOv8n by adding a small target layer, improving the backbone downsampling and loss function, which effectively improves detection performance and is easy to deploy on devices with limited computing resources.

## 2. Introduction to YOLOv8n Algorithm

YOLOv8 is a new generation SOTA model released by Ultralytics. It refers to the ELAN design concept of YOLOv7, replaces the C3 structure of YOLOv5 with C2f structure, adopts Anchor Free detection head, and the loss calculation includes two parts: classification and regression, excluding objective branches. The classification branch uses binary

cross entropy (BCE) loss, while the regression branch uses distribution focus loss (DFL) and CIoU loss functions. The network structure of YOLOv8 includes an input part, a backbone network, a neck module, and an output part. The input part performs mosaic data enhancement, adaptive anchor point calculation, and adaptive grayscale filling on the input image. The neck module adopts FPN+PAN structure to enhance the feature fusion ability of the model, and combines up sampling and down sampling techniques with high-level and low-level feature maps to improve the detection performance of objects of different scales. It provides five scaled versions (N/S/M/L/X) to meet the needs of different computing capabilities and application scenarios. Taking into account all factors, this article improves YOLOv8n, and the YOLOv8 structure diagram is shown in Figure 1.

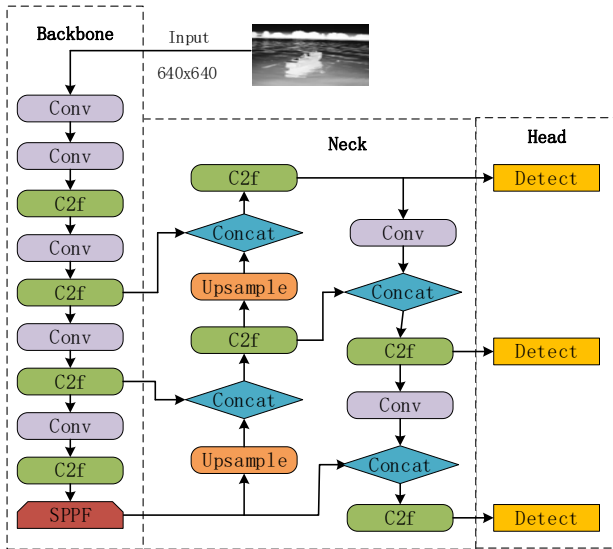


Figure 1. YOLOv8 Structure Diagram

### 3. Improved YOLOv8n Models

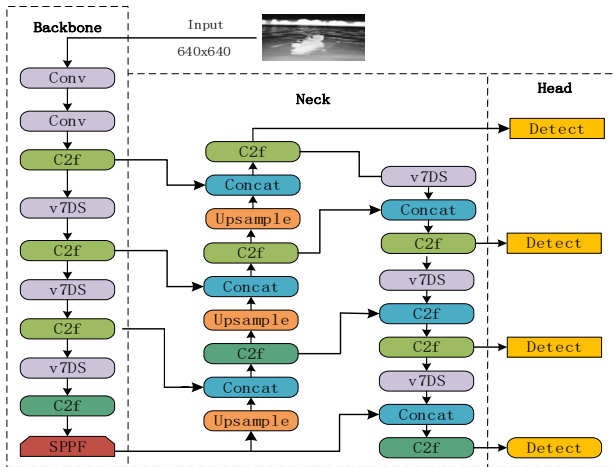


Figure 2. Improved Model Structure Diagram

The improved model structure is shown in Figure 2. Firstly, adding a small target detection layer at the head enhances the ability to extract small target features, thereby improving the detection accuracy of small ship targets. Secondly, the Focaler MPDIoU loss function is used to replace the default CIoU loss function, improving the performance and convergence speed of the model on small or severely occluded objects. Finally, the v8 downsampling is improved to v7DS, which improves the accuracy of ship detection and reduces the number of model parameters and computational

complexity of the small target detection layer to a certain extent.

#### 3.1. Adding a Small Object Detection Layer

In many ship images or videos, the target appears smaller in proportion due to its distance. Due to the limitation of receptive field, the  $80 \times 80$  maximum feature layer used in the original YOLOv8 often misses some targets with fewer pixels. To address this issue, we have added a  $160 \times 160$  extra large feature layer specifically designed for detecting smaller targets. This improvement addresses the shortcomings of the original algorithm in small object detection. Although adding this detection head increases the computational and memory overhead of the model, it significantly improves the accuracy of small object detection from ships and greatly reduces the missed detection rate of small object ships. The improved head is shown in Figure 3 below.

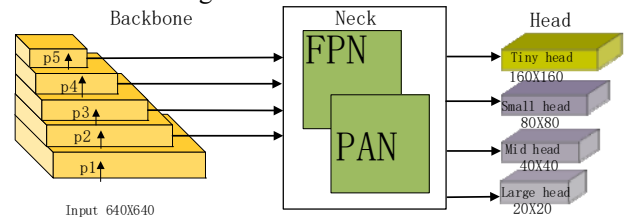


Figure 3. Improved Head Structure Diagram

#### 3.2. New Loss Function-Focal MPDIoU

In object detection, IOU is a very important metric used to evaluate the degree of overlap between predicted bounding boxes and real bounding boxes. IoU values range from 0 to 1, with higher values indicating greater overlap between predicted and real bounding boxes, resulting in higher accuracy in object detection. The calculation formula is as follows:

$$IoU = \frac{B_{pred} \cap B_{gt}}{B_{pred} \cup B_{gt}} \quad (1)$$

Among them,  $B_{pred}$  is the predicted bounding box and  $B_{gt}$  is the real bounding box.

Focaler IoU [13] is a loss function that focuses on samples of different difficulty levels. It is based on the IoU loss function and adjusts the loss value through linear interval mapping method, making the model pay more attention to difficult to detect samples (such as small targets or occluded targets) during the training process, while reducing the focus on easy to detect samples. Focaler-IoU, The calculation formula is as follows:

$$IoU^{focaler} = \begin{cases} 0, & IoU < \alpha \\ \frac{IoU - \alpha}{\beta - \alpha}, & \alpha \leq IoU \leq \beta \\ 1, & IoU > \beta \end{cases} \quad (2)$$

$$L_{Focaler-IoU} = 1 - IoU^{focaler} \quad (3)$$

Among them,  $\alpha$ ,  $\beta$  are hyperparameters,  $[\alpha, \beta] \in [0, 1]$ , in this paper  $\alpha=0$ ,  $\beta=0.95$

MPDIoU [14] is a bounding box regression loss function that optimizes the regression of bounding boxes by minimizing the distance between the top left and bottom right points of the predicted and real boxes. MPDIoU also considers the distance between the center

points of two bounding boxes, which helps to further optimize the position of the bounding boxes. The calculation formula is as follows:

$$MPDIoU = IoU - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \quad (4)$$

$$d_1 = \sqrt{(x_1^{Bprd} - x_1^{Bgt})^2 + (y_1^{Bprd} - y_1^{Bgt})^2} \quad (5)$$

$$d_2 = \sqrt{(x_2^{Bprd} - x_2^{Bgt})^2 + (y_2^{Bprd} - y_2^{Bgt})^2} \quad (6)$$

In the formula,  $w, h$  are the length and width of the input image,  $(x_1^{Bprd}, y_1^{Bprd}), (x_2^{Bprd}, y_2^{Bprd})$  are the coordinates of the predicted box and the real box at the top left corner, and  $(x_1^{Bgt}, y_1^{Bgt}), (x_2^{Bgt}, y_2^{Bgt})$  are the coordinates of the predicted box and the real box at the bottom right corner, representing the Euclidean distance between the real box and the predicted box at the top left and bottom right corners, respectively.

This article combines the advantages of Focaler IoU and MPDIoU to design the Focaler MPDIoU loss function calculation formula as follows

$$MPDIoU^{focaler} = \begin{cases} 0, & MPDIoU < \alpha \\ \frac{MPDIoU - \alpha}{\beta - \alpha}, & \alpha \leq IoU \leq \beta \\ 1, & MPDIoU > \beta \end{cases} \quad (7)$$

$$L_{Focaler-MPDIoU} = 1 - MPDIoU^{focaler} \quad (8)$$

By combining the characteristics of difficult and easy samples of Focaler IoU with the minimum point distance metric of MPDIoU, the detection accuracy is further improved, especially for small targets, and the convergence of the model is accelerated, thereby improving the positioning and boundary accuracy of sea surface ship detection.

### 3.3. Lightweight V7DownSampling Module

Infrared images usually have low resolution and contrast. In the original YOLOv8, downsampling is mainly achieved using convolution with a stride of 2, which may lead to the loss of fine-grained information, thereby affecting the detection standards for small or low contrast objects (such as ships) in infrared images. The convolution layer contains more parameters and computational complexity, and is prone to losing fine-grained information, resulting in poor detection of small ship targets. To solve this problem, the V7DownSampling module in YOLOv7 [8] is introduced. The steps are as follows: the input feature map  $X$  is first fed into a max pooling sequence, which includes a max pooling layer and a  $1 \times 1$  convolutional layer  $Conv(k=1)$ . The max pooling layer halves the width and height of the feature map, while the  $1 \times 1$  convolutional layer converts the number of input channels to half the number of output channels. Next, the input feature map  $x$  is also fed into a convolutional sequence, which includes a  $1 \times 1$  convolutional layer  $Conv(k=1)$  and a  $3 \times 3$  convolutional layer  $Conv(k=3, s=2)$ , further halving the width and height of the feature map. This is equivalent to a lightweight depthwise separable convolution, which helps reduce computational and parameter complexity, making the network more efficient. Finally, the output maxpool ( $x$ ) of the max pooling sequence and the output conv ( $x$ ) of the convolutional sequence are concatenated in the channel dimension to form the final output feature map. Compared to a single convolution with a stride of 2, the V7DownSampling module helps preserve more spatial information. The

V7DownSampling structure is shown in Figure 4.

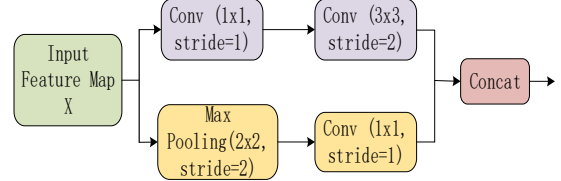


Figure 4. V7DownSampling Module Structure

## 4. Experimental Results and Analysis

### 4.1. Experimental Environment

This article uses the 64 bit Windows 11 operating system, CUDA version 10.2, programming language Python 3.8, deep learning framework PyTorch 1.12.1, and graphics card RTX6000. The experimental parameters are shown in Table 1 below.

Table 1. Experimental Parameter Settings

Parameters	Setup
Input image size	640×640
Initial learning rate	0.01
Epochs	200
Batch size	16
Optimizer	SGD
Momentum	0.937
Weight decay	0.0005

### 4.2. Dataset and Evaluation Indicators

This article uses a publicly available infrared ship dataset created by InfiRay, which includes 8402 images of infrared ships in 7 categories: liner, bulk carrier, warship, sailboat, canoe, container ship, and fishing boat. This article randomly divides the dataset into training, validation, and testing sets in a ratio of 7:2:1. Figures 5 and 6 show the distribution of the number of category labels and the size of label bounding boxes, respectively.

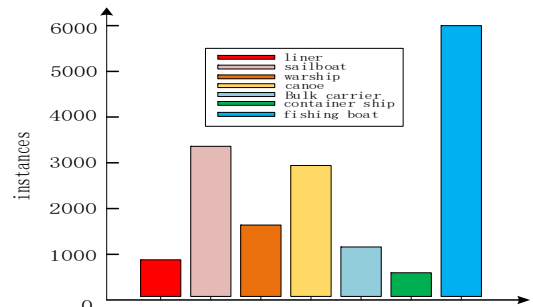


Figure 5. Quantity of Various Labels

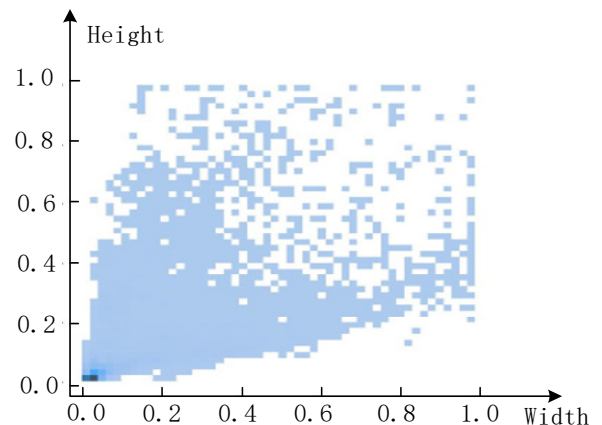


Figure 6. Distribution of label bounding box sizes

In order to accurately evaluate the performance of the proposed detection method, precision (P), recall (R), and average precision (mAP50) were used. The higher the P, R, and mAP, the more accurate the object detection. To comprehensively evaluate efficiency and performance, this study also considered the model's parameter count, computational complexity, and model size. Among them, mAP0.5 represents the average detection accuracy of all target categories when the IoU threshold is 0.5. The calculation formulas for P, R, AP (the area enclosed by the PR curve), and mAP are as follows.

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$AP = \int_0^1 P(R)dR \quad (11)$$

$$mAP = \frac{1}{k} \sum_{i=0}^k AP_i \quad (12)$$

$$AP = \int_0^1 P(R)dR \quad (13)$$

Among them: represents the number of correctly predicted positive classes, represents the number of incorrectly predicted negative classes, represents the number of incorrectly predicted positive classes.

Figures 7 (a) and (b) represent the PR curves before and after the improvement, respectively. From the figures, it can be seen that the improved algorithm has varying degrees of detection accuracy on seven types of ships, including liners mAP@0.5 Improved from 0.889 to 0.923, overall, the improved YOLO model has an average precision mean across all categories( mAP@0.5 )The improvement from 0.889 to 0.923 indicates that the improved model performs better in infrared ship detection and can more accurately detect and identify different types of ships

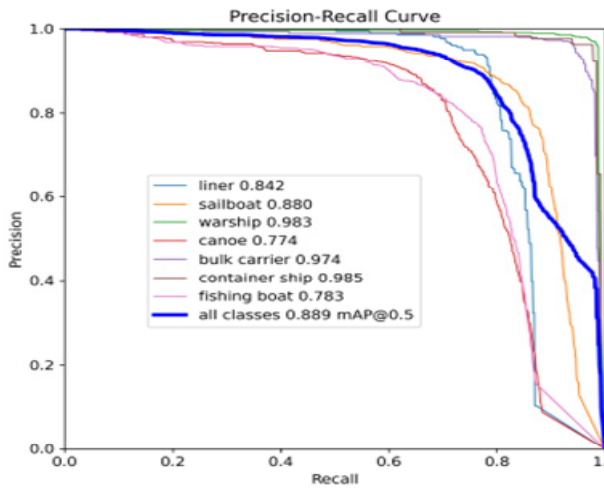


Figure 7. (a) PR curve before improvement

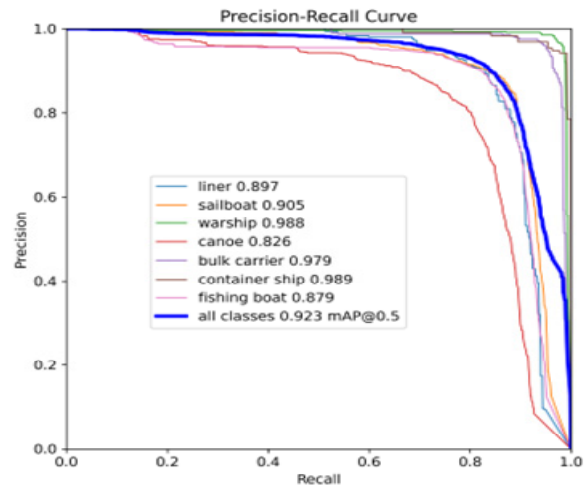


Figure 7. (b) PR curve after improvement

### 4.3. Module Ablation Experiment

In order to verify the effectiveness of the small target detection layer, Focaler MPDIoU loss function, and YOLOv5 downsampling module, the following ablation experiments were conducted using YOLO v8n as the baseline model. The experimental results are shown in Table 2 (where p2, F, and v represent the P2 small target detection layer, Focaler MPDIoU loss function, V7 DownSampling, and 'v' represents the use of this module in this set of experiments).

In experiments A, B, and C, the addition of p2, F, and v modules to YOLO v8n resulted in varying degrees of improvement in average accuracy. Among them, the addition of a small target layer led to mAP@0.5 Reached 91.5%, an increase of 2.6 percentage points, significantly improving the detection performance of small targets and effectively reducing missed detections. Introducing Focaler MPDIoU separately, mAP@0.5 By increasing by 0.5 percentage points, P and R also showed varying degrees of improvement, while indicators such as computational complexity remained unchanged. This indicates that the Focaler MPDIoU loss function is lossless in improving the detection performance of the model. Only add V7DownSampling, mAP@0.5 An increase of 0.1 percentage points, a 9% reduction in parameter count, and an 8% reduction in model size indicate that it can improve detection accuracy while reducing model complexity.

When the three improved modules were combined in the last set of experiments, the model achieved optimal performance, mAP@0.5 The highest value of 92.3% was achieved on YOLOv8n, which increased by 3.4 percentage points compared to YOLOv8n. At the same time, the number of parameters decreased by 11.8% and the model size decreased by 8%. Although it slightly increased the computational load, the significant performance improvement it brought was also worth it.

In summary, each improved module has been validated through ablation experiments for its positive role in improving detection accuracy and optimizing model complexity. Especially after integrating all the improvement points, the overall performance of the model has been significantly improved, proving the effectiveness and rationality of these improvement methods.

## 5. Comparative Experiments of Different Models

To further highlight the superiority of the algorithm proposed in this article, a comparative experiment was conducted with several mainstream deep learning based object detection algorithms. In this comparative experiment, the focus was on the accuracy (P%) and recall (R%) of the model mAP@0.5 %These three core indicators, as well as the

three efficiency indicators of parameter quantity, model size, and computational complexity. The experimental results are

shown in Table 3.

**Table 2.** Ablation Experiment

model	P2	F	V	P/%	R/%	mAP@0.5/%	Parameter quantity /106	Model size / MB	Calculated amount (GFLOPS)
YOLOv8n				90.3	84.3	88.9	3.007013	6.3	8.1
A	√			90.7	87.5	91.5	2.921964	6.3	12.2
B		√		90.8	84.5	89.4	3.007013	6.3	8.1
C			√	91.5	83.4	89	2.739045	5.8	7.8
D	√	√		90.7	87.9	92	2.921964	6.3	12.2
E	√		√	90.7	87.6	91.7	2.648124	5.8	11.9
Ours	√	√	√	90.4	88.7	92.3	2.648124	5.8	11.9

**Table 3.** Comparison experiments of different models

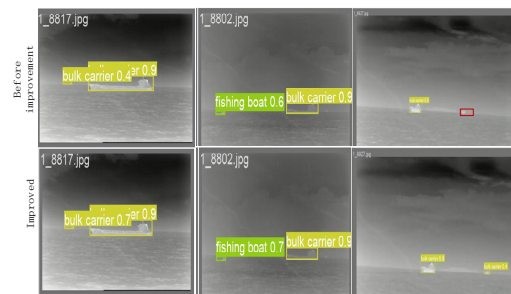
Models	P/%	R/%	mAP@0.5/%	Parameter /10 <sup>6</sup>	Model Size/MB	FLOPS/G
YOLOv3 [15]	88.7	89.7	90.9	61.529748	123.5	154.6
YOLOv3-tiny	87.4	83	86.1	8.680552	17.4	12.9
YOLOv5s [16]	91.6	88.4	91.6	7.029004	14.5	15.8
YOLOv7-tiny [8]	87.8	85.1	89.4	6.023832	12.3	13.1
rtetr-r18 [17]	91.4	89.8	92.5	19.880748	40.5	57
literature [12]	92.3	90.9	93.5	22.900000	-	36.3
YOLOv8n	90.3	84.3	88.9	3.007013	6.3	8.1
Ours	90.4	88.7	92.3	2.648124	5.8	11.9

Firstly, in terms of precision (P%) and recall (R%), although the ours algorithm did not reach the highest level in these two indicators, it still performed well, at 90.4% and 88.7%, respectively. This indicates that the ours algorithm has high accuracy and recall in detecting targets, which can meet the needs of most applications. Next, we will focus on the average precision mean(mAP@0.5%) This is an important indicator for comprehensive evaluation of algorithm performance. From the table, it can be seen that the ours algorithm mAP@0.5 % It is 92.3%, second only to 92.5% of rtdetr-r18 and 93.5% of reference [11], ranking third. However, it is worth noting that the ours algorithm has significant advantages in terms of parameter count and model size. Specifically, the parameter count of our algorithm is only 2648124, far lower than most other models such as YOLOv3's 61529748 and literature [11]'s 22900000. Meanwhile, the model size of our algorithm is only 5.8MB, which is smaller than YOLOv8n's 6.3MB. This means that the ours algorithm is more efficient in terms of computing resources and storage space, and is suitable for deployment in resource constrained environments. Finally, in terms of computational complexity (GFLOPS), the ours algorithm has a computational complexity of 11.9 GFLOPS, Compared with some larger models such as the 57GFLOPS of rtdetr-r18 and the 154.6GFLOPS of YOLOv3, our algorithm also performs well in computational efficiency, only slightly lower than YOLOv8n's 8.1 GFLOPS.

In summary, the ours algorithm maintains high accuracy and recall while having a small number of parameters and model size, as well as low computational complexity. This makes the ours algorithm have a high performance advantage in object detection tasks, especially in resource constrained environments, where the advantages of the ours algorithm are more pronounced. Therefore, it can be said that the ours

algorithm is an efficient and lightweight object detection algorithm.

In order to compare the detection performance of the improved model and the original model more intuitively, Figure 8 shows some of the detection results before and after the model improvement. In the original model, there were cases of false positives and false negatives (marked in red), and the confidence level was lower than that of the improved model. However, the improved model significantly improved this problem. Overall, the improved model showed higher confidence and lower false negatives in detecting ships, which makes it more effective and reliable in the field of ship detection.



**Figure 8.** Comparison of detection effects before and after improvement

## 6. Summarize

This article proposes an infrared ship detection algorithm based on improved YOLOv8n. By adding a small target detection layer, introducing the Focler MPDIoU loss function, and improving the downsampling module, the model has significantly improved its detection accuracy when dealing with small or occluded targets. The ablation

experiment further validated the effectiveness of various improvement measures, demonstrating the positive role of these methods in improving detection accuracy and optimizing model complexity. Compared with other mainstream object detection algorithms, the improved YOLOv8n maintains high accuracy and recall while having smaller parameter count, smaller model size, and lower computational complexity, making it more suitable for deployment in resource constrained environments. In summary, the improved model has better performance and reliability in the field of ship detection, and can provide technical support for practical applications such as ocean monitoring and maritime search and rescue. The main follow-up work will consider further lightweighting the model while ensuring detection accuracy.

## References

- [1] Gu Jing Research on Infrared Ship Target Detection Method Based on Deep Learning [D]. Jiangsu University of Science and Technology, 2023. DOI: 10.27171/d.cnki. ghdc. 2023.000280.
- [2] Jia Chunrong, Yang Fan, Gao Jianxin, et al. Research on Infrared Detection Technology for Marine Ship Targets [J/OL]. Laser Journal: 1-11 [2400-06-16].
- [3] Yang Heng, Wang Chao, Jiang Wentao, Liu Peizhen, Sun Xiaowei, Ji Ming Object detection algorithm based on random background modeling [J] Applied Optics, 2015, 36 (6): 880-887 DOI: 10.5768/JAO201536.0602001.
- [4] Liu S, Chen P, Woźniak M. Image enhancement-based detection with small infrared targets[J]. Remote Sensing, 2022, 14(13): 3232.
- [5] He Qian, Liu Boyun. Overview of Infrared Image Edge Detection Algorithms [J]. Infrared Technology, 2021, 43 (03): 199-207.
- [6] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
- [7] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arxiv preprint arxiv:2004.10934, 2020.
- [8] 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[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 7464-7475.
- [9] Zhan W, Zhang C, Guo S, et al. EGISD-YOLO: Edge Guidance Network for Infrared Ship Target Detection[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024.
- [10] Gu Jiaojiao, Li Bingzhen, Liu Ke, et al. Infrared Ship Target Detection Algorithm Based on Improved Faster R-CNN [J]. Infrared Technology, 2021, 43 (02): 170-178.
- [11] Miao R, Jiang H, Tian F. Robust ship detection in infrared images through multiscale feature extraction and lightweight CNN[J]. Sensors, 2022, 22(3): 1226.
- [12] Zhang Shen, Hu Lin, Sun Xiang'e, et al. Infrared ship detection based on attention mechanism and multi-scale fusion [J]. Advances in Laser and Optoelectronics, 2023, 60 (22): 256-262.
- [13] Zhang H, Zhang S. Focaler-IoU: More Focused Intersection over Union Loss[J]. arxiv preprint arxiv:2401.10525, 2024.
- [14] Siliang M, Yong X. Mpdio: a loss for efficient and accurate bounding box regression[J]. arxiv preprint arxiv:2307.07662, 2023.
- [15] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arxiv preprint arxiv:1804.02767, 2018.
- [16] Jocher G, Chaurasia A, Stoken A, et al. ultralytics/yolov5: v7.0-yolov5 sota realtime instance segmentation[J]. Zenodo, 2022.
- [17] Zhao Y, Lv W, Xu S, et al. Detsr beat yolos on real-time object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 16965-1697.