

Article

An Algorithm for Ship Detection in Complex Observation Scenarios Based on Mooring Buoys

Wenbo Li ^{1,2}, Chunlin Ning ^{1,3,4,5,*}, Yue Fang ^{1,3,4,5}, Guozheng Yuan ^{1,2}, Peng Zhou ²  and Chao Li ^{1,3,4,5}

¹ First Institute of Oceanography, Ministry of Natural Resources, Qingdao 266061, China; z22160122@s.upc.edu.cn (W.L.); yfang@fio.org.cn (Y.F.); z21160050@s.upc.edu.cn (G.Y.); lichao@fio.org.cn (C.L.)

² College of Oceanography and Space Informatics, China University of Petroleum, Qingdao 266580, China; zhoupeng@upc.edu.cn

³ Key Laboratory of Marine Science and Numerical Modeling, Ministry of Natural Resources, Qingdao 266061, China

⁴ Shandong Key Laboratory of Marine Science and Numerical Modeling, Qingdao 266061, China

⁵ Laboratory for Regional Oceanography and Numerical Modeling, Qingdao Marine Science and Technology Center, Qingdao 266237, China

* Correspondence: cning@fio.org.cn; Tel.: +86-151-6669-6857

Abstract: Marine anchor buoys, as fixed-point profile observation platforms, are highly susceptible to the threat of ship collisions. Installing cameras on buoys can effectively monitor and collect evidence from ships. However, when using a camera to capture images, it is often affected by the continuous shaking of buoys and rainy and foggy weather, resulting in problems such as blurred images and rain and fog occlusion. To address these problems, this paper proposes an improved YOLOv8 algorithm. Firstly, the polarized self-attention (PSA) mechanism is introduced to preserve the high-resolution features of the original deep convolutional neural network and solve the problem of image spatial resolution degradation caused by shaking. Secondly, by introducing the multi-head self-attention (MHSA) mechanism in the neck network, the interference of rain and fog background is weakened, and the feature fusion ability of the network is improved. Finally, in the head network, this model combines additional small object detection heads to improve the accuracy of small object detection. Additionally, to enhance the algorithm's adaptability to camera detection scenarios, this paper simulates scenarios, including shaking blur, rain, and foggy conditions. In the end, numerous comparative experiments on a self-made dataset show that the algorithm proposed in this study achieved 94.2% mAP50 and 73.2% mAP50:95 in various complex environments, which is superior to other advanced object detection algorithms.

Keywords: ship detection; anchor buoy; complex environment; attention mechanism; YOLOv8



Citation: Li, W.; Ning, C.; Fang, Y.; Yuan, G.; Zhou, P.; Li, C. An Algorithm for Ship Detection in Complex Observation Scenarios Based on Mooring Buoys. *J. Mar. Sci. Eng.* **2024**, *12*, 1226. <https://doi.org/10.3390/jmse12071226>

Academic Editors: Mustafa Kemal Cambazoglu and Zafer Defne

Received: 29 May 2024

Revised: 2 July 2024

Accepted: 5 July 2024

Published: 20 July 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

A marine mooring buoy is a floating device located on the surface of the sea that has traditionally been used to mark shipping lanes or to indicate potentially hazardous objects, such as coral reefs or underwater shipwrecks [1]. Since the 1920s, it has been an important multifunctional platform carrying various types of observational equipment for observing information on the marine environment, meteorological conditions, and navigational safety [2–4]. If the ocean buoy is damaged, it not only leads to the loss of key data but also may collide with ships during drifting, causing secondary accidents [5,6]. Therefore, in order to reduce the damage of marine buoys due to ship collisions and other damages, the design of a buoy warning and forensic system is crucial.

Traditional buoy monitoring systems use CCTV technology monitoring or multi-sensor fusion technology to detect and warn ships. For example, Zheng et al. [7] used a CCTV video surveillance system to continuously collect image information for real-time

vessel detection. Zhao et al. [8] designed a control system that combines geomagnetic detection and infrared sensing, which can detect ship targets around buoys. Chen et al. [9] used underwater acoustic signals to locate the position of ships and provided different levels of warning based on the position of the ship to the buoy. Hwang et al. [6] used AIS to determine the distance between ships and buoys and then took photos of surrounding ships for evidence collection. Although the traditional detection methods mentioned above can detect the presence of ships, it is difficult to recognize the type of ship directly at the buoy end, and it is necessary to upload the pictures to the shore station system and subsequently identify the type of ship.

As artificial intelligence technology develops, it is becoming more commonplace to process photos or videos using deep learning techniques to detect ship targets. However, ship detection tasks based on the buoy end have unique features when compared to other types of object detection:

1. Based on its special application environment, the swaying and shaking of the buoy itself will cause the ship to appear shaky and blurred in the camera. This leads to the easy loss of key details of the ship during the extraction of edges and textures, thereby affecting the detection effectiveness;
2. The marine environment is complex and variable, and cameras are often disturbed by weather such as rain and fog, which can lead to ship targets being easily occluded, posing significant challenges to accurate positioning and recognition of ships;
3. Small ships that frequently damage buoys are more likely to exhibit smaller scales in the image, which may lead to missed detection of small targets. Therefore, ship detection algorithms need to have stronger small object detection capabilities to effectively identify and detect these small ships.

Currently, object detection methods based on deep learning are primarily categorized into two-stage detection models and one-stage detection models. The target detection process in the two-stage detection model involves two phases: creating candidate regions initially and then refining and classifying them to achieve the ultimate detection outcomes. This model usually has high accuracy but a slow detection speed. R-CNN [10], Fast-RCNN [11] and Faster-RCNN [12] are some of the top two-stage models. Gu et al. [13] improved the Faster R-CNN by combining the scene semantic narrowing and topic narrowing sub-networks to solve the problems of insufficient ship feature extraction and repeated detection. However, the two-stage algorithm is not suitable for application at the buoy end due to its inherently slower detection speed.

Conversely, one-stage detection models such as YOLO [14–19] and SSD [20] eliminate the step of generating candidate boxes, greatly simplifying the process and reducing computational complexity. In addition, these models also demonstrate good performance in handling small object detection in complex environments, making them an ideal choice for ship detection at the buoy end. Zhang et al. [21] improved the YOLOv4 algorithm by introducing Swin Transformer to extract deep features, effectively improving the accuracy of ship detection. However, the model has a complex structure and needs further simplification. Zheng et al. [22] proposed the MC-YOLOv5s algorithm, which uses MobileNetV3-Small as a lightweight feature extraction backbone network to improve the detection speed. Although the above improvement methods have shown some improvement in detection speed and accuracy, further optimization is still needed.

In recent years, the importance of the attention mechanism in the field of target detection has increased. This mechanism improves the generalization ability of detection models by focusing attention on regions and features of interest. For example, Shang et al. [23] proposed to improve the YOLOv5 algorithm by utilizing the convolutional block attention module (CBAM) and coordinate attention (CA), which improves the maritime target detection accuracy by assigning higher weights to the tensor regions where the target is more likely to appear. Although this method has been effective in improving the accuracy of ship detection, they have not proposed effective solutions when dealing with dense ship identification problems. In response, Wang et al. [24] optimized the YOLOv5

model by combining the CNeB2 module and the separated and enhanced attention module (SEAM) attention mechanism, successfully solving the problem of difficult positioning when ships are dense. Zhao et al. [25] significantly enhanced the model's performance in detecting occlusion and targets of different scales by integrating the ECA module into YOLOX's backbone network and improving non-maximum suppression. However, they did not provide targeted optimization to address recognition issues caused by rain, fog, and visual blurring. Si et al. [26] first used the K-means method to improve the clustering performance of ship data and then introduced the squeeze-and-extraction (SE) mechanism in the feature extraction network to enhance the feature extraction ability. This innovation effectively improves the recognition performance of the model for ships under sea fog conditions, but no effective improvement measures have been proposed for the problem of shaking blur. Wang et al. [27] proposed a feature fusion module incorporating a GT module, which effectively suppresses the noise introduced by shallow features, and introduced an SPD-Conv module in order to improve the detection accuracy of low-resolution images. This method has achieved significant results in improving the image resolution reduction caused by shaking, but it has limitations in processing visible light images required in this study, as it is mainly suitable for infrared image processing. Although the above target detection algorithms are able to achieve high-precision detection of ships in fixed surveillance environments, they fail to effectively deal with the problems of recognizing ships on shaky platforms and under complex environmental conditions at the buoy end at the same time. In view of this, this article proposes an improved YOLOv8 object detection algorithm aimed at effectively overcoming issues such as camera shake and rain and fog occlusion at the buoy end, thereby accurately identifying ship types at the buoy end. The following are this study's primary contributions:

1. Aiming at the problem of blurring ship images and a reduction in target spatial resolution due to camera shake, this study redesigned the bottleneck structure of the C2f module, combined with the polarized self-attention mechanism, to reduce information loss due to blurring through dual enhancement of spatial and channel dimensions, and to ensure a more comprehensive extraction and fusion of ship features;
2. For the feature of rain and fog obscuring ship targets in bad sea conditions, this study introduces the multi-head self-attention module in the neck network. This approach captures the relative position relationship between features by introducing relative position coding, which fully exploits the correlation between features, effectively weakens the interference of rain and fog, and significantly improves the accuracy of ship detection;
3. In response to the common problem of low accuracy in small vessel detection, this study combines an independently designed small target detection head and a larger size feature map to enhance the detection accuracy of small ships by extracting richer feature information from the shallow feature map.

2. Materials and Methods

2.1. Materials

2.1.1. Dataset Description

The dataset used in this study was collected from the Jiaozhou Bay area of Qingdao, China. The HIKVISION model DS-2TD5167-50H4/W/GLT camera was used as the equipment, as shown in Figure 1. The resolution of the device is 2688×1520 (visible light images), and it captures 25 frames per second.

The captured videos were then analyzed and processed, resulting in 3732 images. These images cover six different types of vessels: container ships (CS), general cargo ships (GCS), fishing boats (FB), passenger ships (PS), small wooden boats (SWB) and sailboats (SB). Figure 2 shows the various ship images captured.



Figure 1. (a) HIKVISION camera; (b) capture bracket.

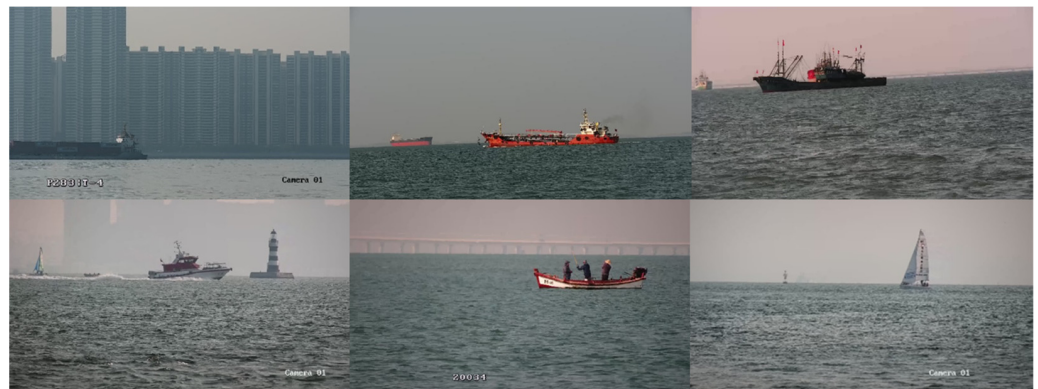


Figure 2. Example of image acquisition of various types of ships.

However, after collation, it was found that the amount of data for some types of ships was too small. In order to enrich the dataset of ship types, 162 and 313 images were selected from the SMD [28] and SeaShips [29] datasets for supplementation. Finally, the enriched dataset will be named BaiLongOnBoardShips, hereinafter referred to as the BLOBS dataset, which includes a total of 4207 images.

After acquiring the images, the Labelling annotation tool was used to annotate the images and generate the corresponding XML files, and the number of specific targets is shown in Table 1. Then, the corresponding annotation file for each image was converted to a txt file. Finally, this study randomly divides the dataset into a training set, validation set and test set according to the ratio of 8:1:1.

Table 1. Number of vessels of each type.

Name	Number (SMD)	Number (SeaShips)	Number (All)
container ship	98	179	1066
general cargo ship	183	0	1289
fishing boat	0	24	845
passenger ship	86	158	1639
small wooden boat	0	0	806
sailboat	17	0	655

2.1.2. Data Preprocessing

Since the camera is located on a buoy and has to experience a variety of rough sea conditions, this study preprocesses the dataset before training in order to better simulate the actual ship images and to improve the algorithm's generalization ability. Figure 3 shows a sample of the processed image, and the following are the specific operations of the preprocessing:

1. For the motion blur caused by camera shake, this paper uses a blur kernel of size 50×50 for simulation and normalizes the blur kernel to ensure that the total brightness of the processed image remains unchanged;

2. For the effect of rainy days, this paper first generates a noise layer on the input image to simulate raindrops and then rotates and stretches the noise layer to increase the real sensation and dynamic effect of the simulated raindrops. Finally, we use the transparency blending technique to blend the simulated raindrop layer with the original image, which finally enhances the visual impact of the rainy day;
3. In order to simulate the effect of a real foggy day, this paper adopts the method of synthesizing fog at the center point. The specific operation is to set the fog center at the center of the image, then calculate the Euclidean distance of each pixel point relative to the fog center, then use the attenuation coefficient to adjust the pixel brightness accordingly in order to simulate the natural attenuation of the light when it passes through the fog, and finally multiply the luminance value of each pixel by the attenuation coefficient, so as to achieve the simulation of the fog effect.

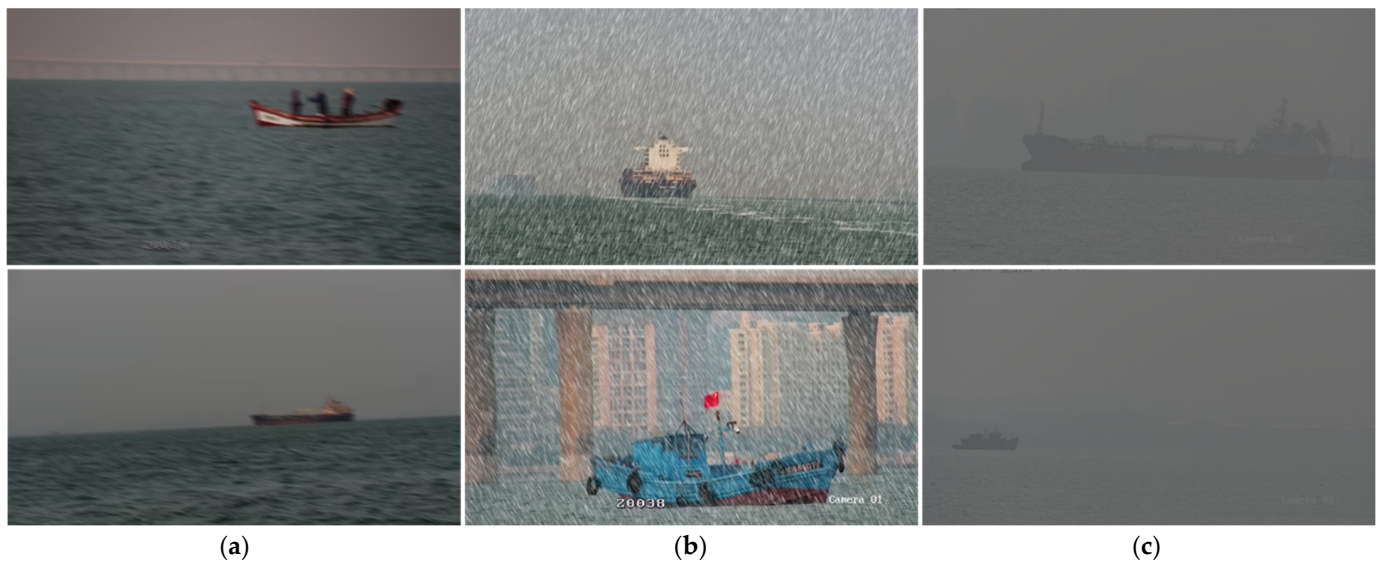


Figure 3. Preprocessing image simulation under different sea conditions. (a) Visual blur processing; (b) simulation of rain impacts; (c) center point synthetic fog treatment.

2.2. Methods

2.2.1. Overall Network Introduction

YOLOv8 [19], as an advanced target detection model, uses one-stage regression to quickly and efficiently locate and identify targets in images through convolutional operations, thus achieving an optimal balance between detection speed and accuracy. The original YOLOv8 network architecture is shown in Figure 4.

However, ship targets are often blurred and heavily occluded in images captured by camera shake and rainy and foggy days, and the existing YOLOv8 network structure cannot meet the higher detection demands. To address these problems, this paper takes YOLOv8n as the base model and optimizes the model in terms of feature extraction, feature fusion and small target detection head.

Firstly, in order to help the model accurately understand the spatial information of ships, this study embedded the polarized self-attention mechanism into the C2f module, thereby creating a new C2f_PSA module that can retain the high-resolution information in the original deep convolutional neural network through polarization filtering. Furthermore, in order to enhance the model's ability to capture remote structural information from images and further optimize the effect of feature fusion, this paper introduces the MHSA mechanism into the neck network of YOLOv8, thereby not only improving the receptive field of the network but also significantly enhancing its feature expression capability. Finally, in order to further enhance the detection capability of small ships in complex environments, this paper first concatenates shallow and deep feature maps to obtain richer feature information and then adds smaller scale detection heads for detection. Figure 5

shows the YOLOv8-PMH network architecture proposed in this study, with improvements represented in color.

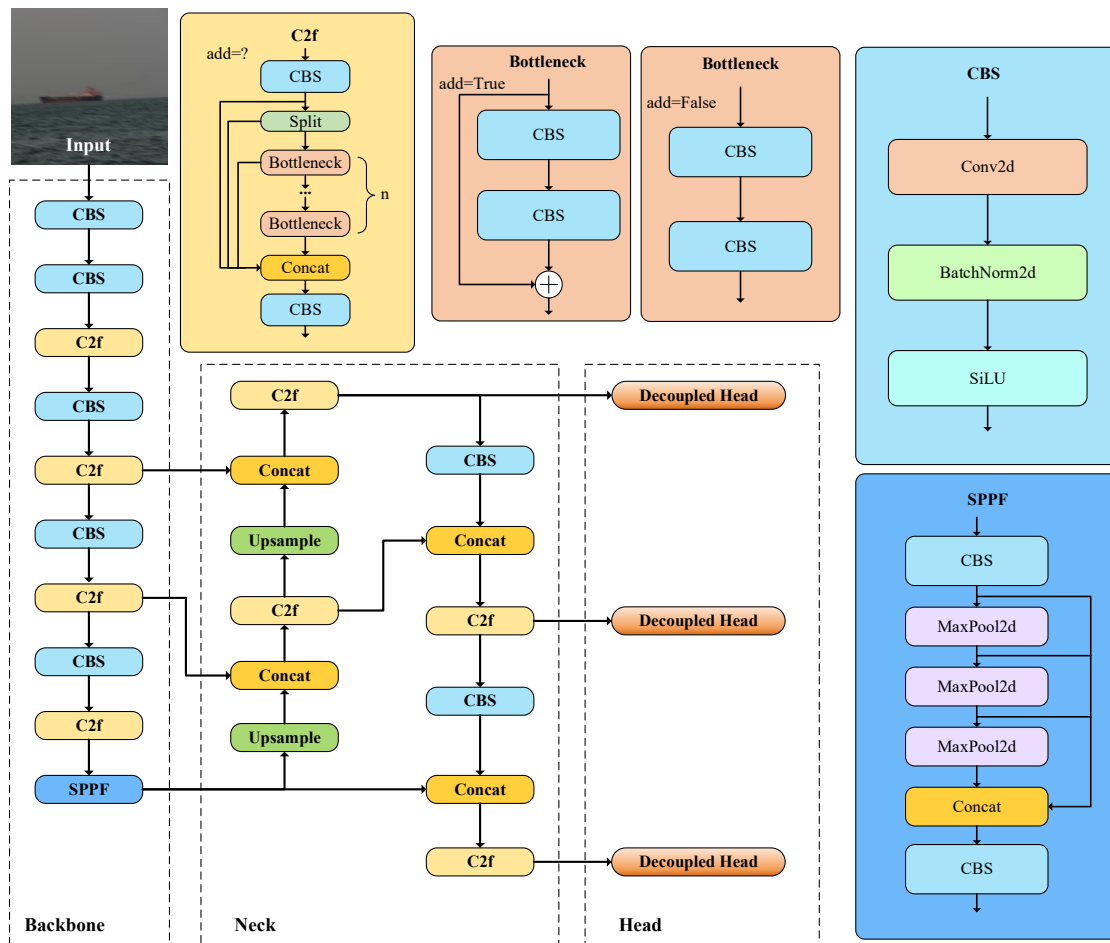


Figure 4. YOLOv8 network architecture diagram [19].

2.2.2. C2f_PSA Module

The key aspect of the attention mechanism is to enable the algorithm to focus on the most critical information. Channel attention mechanisms such as SENet [30], ECANet [31] and GCNet [32] are mainly achieved by assigning different weights to different channels, but they remain consistent in processing spatial location information. Therefore, the channel attention mechanism ignores the differences in spatial dimensions, which may lead to insufficient utilization of information from different regions in the image, thereby affecting the performance of the model in identifying ships in harsh sea conditions. Although the spatial attention mechanism can capture the location information features of important areas well, its ability to distinguish color differences between ships and the sea surface is relatively limited.

After the emergence of spatial and channel attention mechanisms, dual attention mechanisms that integrate these two dimensions have emerged to further improve the performance of models, such as DANet [33] and CBAM [34]. Although these two attention mechanisms combine channel information and spatial information, they are sensitive to input noise. Therefore, a polarized self-attention (PSA) mechanism [35] has been proposed for achieving high-quality pixel-by-pixel regression. Compared with existing channel spatial combination methods, PSA attention does not prefer specific layouts, thus demonstrating greater flexibility and adaptability. Combined with the principle of the PSA module, in this paper, channel self-attention and spatial self-attention are fused in parallel to obtain the PSA module shown in Figure 6.

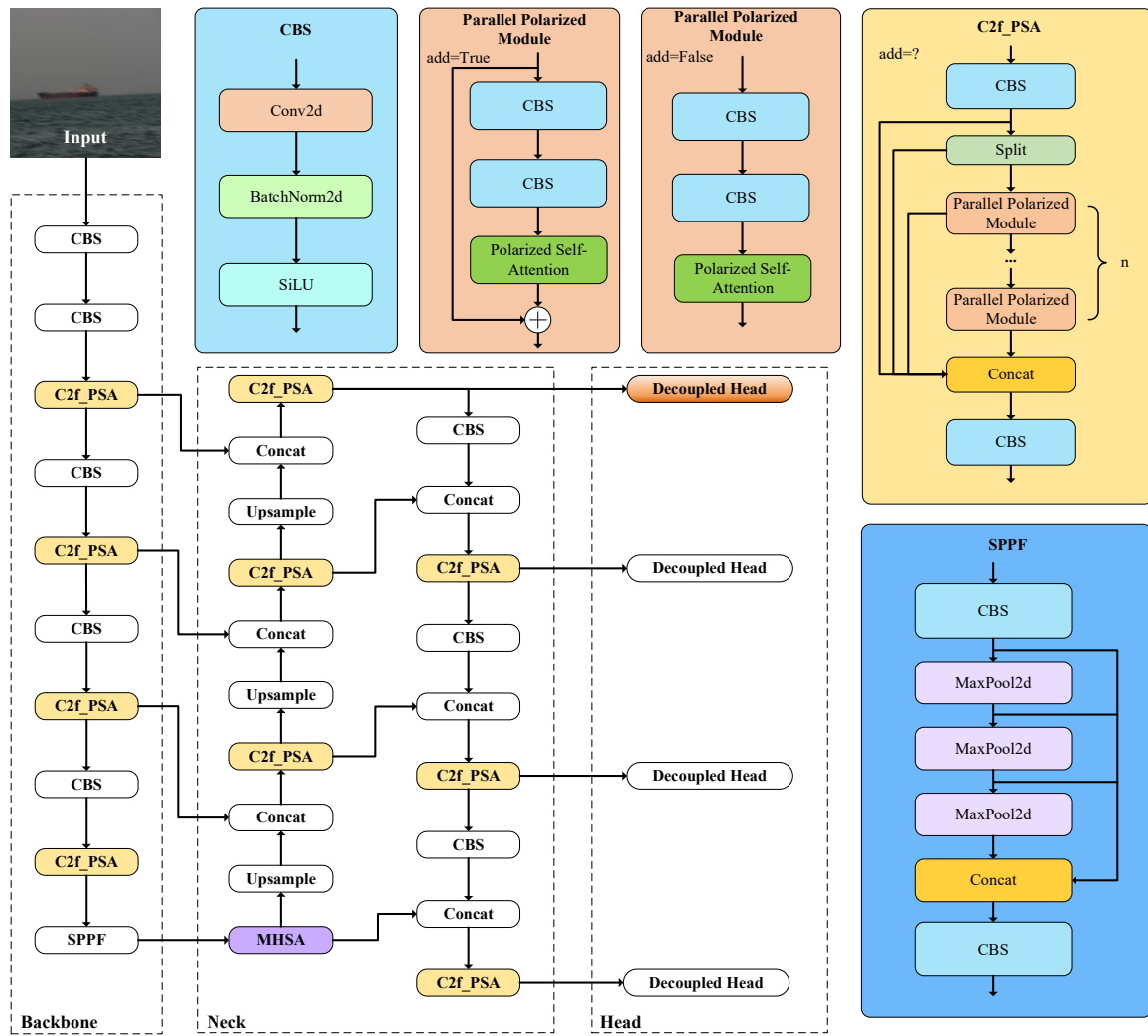


Figure 5. YOLOv8-PMH network architecture diagram.

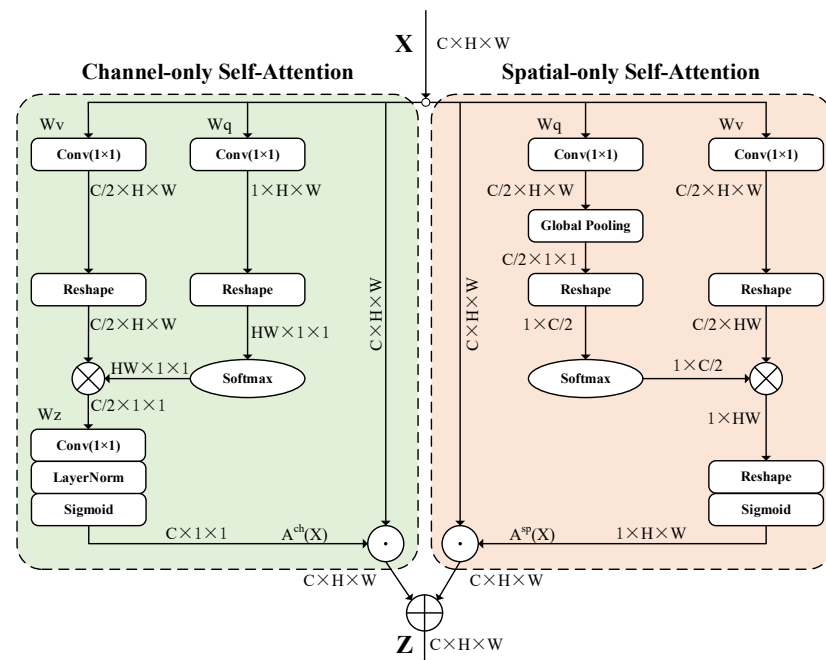


Figure 6. PSA module [35].

PSA attention, in order to preserve the potential loss of high-resolution information in the original deep convolutional neural network by downsampling, which maintains the dimension of $C/2$ in the channel and $[H, W]$ in the space, minimizes information loss due to blurring and ensures that the network can capture more efficiently, even in the case of bad sea conditions that cause the buoys to rock violently, resulting in blurred imaging ship information.

As illustrated in Figure 6, the PSA module first divides the input feature map $X \in \mathbb{R}^{C \times H \times W}$, and thus in the channel part, the input feature X is first transformed into two parts, q and v , by a 1×1 convolution operation. During this process, the channel of q is compressed, while the channel of v remains at a high $C/2$ level to retain high-resolution color information. In order to further enhance the information of q , the Softmax function is used to enhance the information of q . Next, q and v are matrices multiplied, and the result is used to raise the channel from $C/2$ to C dimensions by convolution and LayerNorm operations. And finally, the Sigmoid function is used so that all the parameters are kept between 0 and 1, to complete the accurate adjustment of the weights. By enhancing the channel information, the model is able to accurately separate the ship from the sea surface, relying on the color difference. The final output of the channel part is shown in Equation (2):

$$A^{ch}(X) = F_{SG} \left[W_{Z|\theta_1} \left((\sigma_1(W_v(X))) \times F_{SM}(\sigma_2(W_q(X))) \right) \right] \quad (1)$$

$$Z^{ch}(X) = A^{ch}(X) \odot^{ch} X \quad (2)$$

Here, W_q , W_v and W_z are the 1×1 convolutional layers, σ_1 and σ_2 are the tensor reshaping operations, θ_1 is the intermediate parameter for channel convolution, $F_{SM}(\cdot)$ is the Softmax operator, $\times$ is the matrix dot-product operation, $\odot^{ch}$ is the multiplication operator for channel branching and $F_{SG}(\cdot)$ is the Sigmoid function.

In the spatial part of the PSA module, unlike the channel part, for the q feature, global average pooling is used in this step to compress the spatial dimension and convert it to a size of 1×1 . Next, q and v are subjected to matrix dot-product operations followed by the Reshape and Sigmoid operations. By enhancing the spatial part, the high-resolution spatial information of the ship can be well preserved, even when encountering shaking blur. The calculation process of the final spatial part is shown in Equations (3) and (4):

$$A^{sp}(X) = F_{SG} \left[\sigma_3 \left(F_{SM}(\sigma_1(F_{GP}(W_q(X)))) \times \sigma_2(W_v(X)) \right) \right] \quad (3)$$

$$Z^{sp}(X) = A^{sp}(X) \odot^{sp} X \quad (4)$$

Here, σ_1 , σ_2 and σ_3 are the tensor reshaping operations, $F_{GP}(\cdot)$ is a global average pooling operation, and $\odot^{sp}$ is a multiplication operator for spatial branches.

After connecting in parallel mode, the final output of the PSA module is:

$$PSA_p(X) = A^{ch}(X) \odot^{ch} X + A^{sp}(X) \odot^{sp} X \quad (5)$$

This article uses the PSA attention mechanism to improve the C2f module, as shown in Figure 7. After embedding the polarized self-attention mechanism behind the original bottleneck structure, a new Parallel Polarized module is formed. Among them, the backbone network partly adopts the residual structure, while the neck network does not have the residual structure, and finally, the C2f_PSA module is constructed. Through this improvement, the model preserves the high-resolution information of ship targets as much as possible, thereby improving the precise positioning of ship targets.

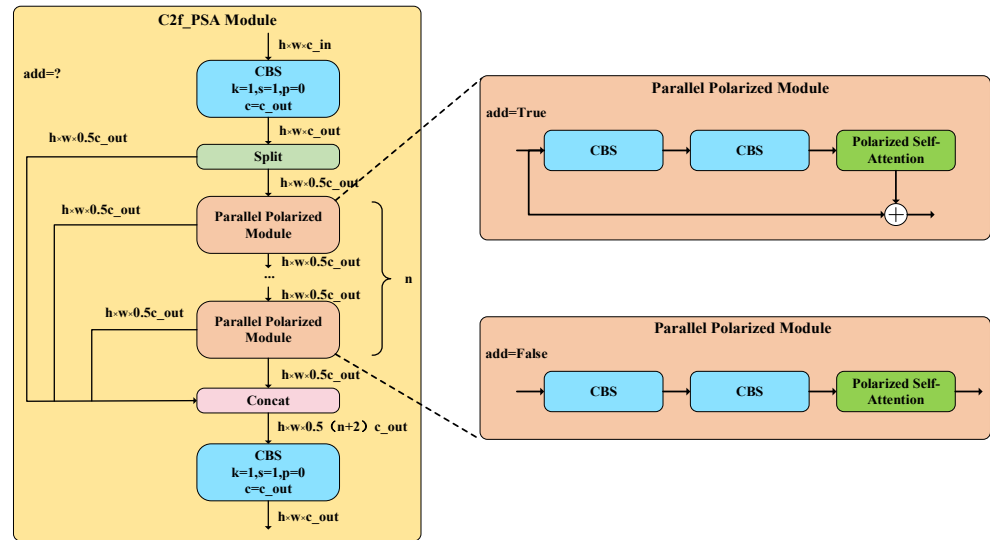


Figure 7. C2f_PSA module.

2.2.3. MHSA Module

Transformer was initially used for natural language processing tasks and then widely referenced in various aspects with its powerful representation capability and global vision [36], and its representative works include Vision Transformer (ViT) [37], data-efficient image transformers [38], Swin Transformer [39], etc. The multi-head self-attention mechanism (MHSA) [40] introduced in this study is the benchmark of attention mechanisms in Transformer.

The MHSA mechanism assigns higher weights to key target features by analyzing the correlation between features while assigning lower weights to irrelevant background features, thereby improving the feature fusion effect of the network and significantly reducing the interference of background factors such as rain and fog on ship detection. This makes the MHSA module highly applicable in ship recognition tasks in rainy and foggy environments. Therefore, this article introduces the MHSA mechanism into the YOLOv8 ship detection model, and the single-layer structure of the module is shown in Figure 8.

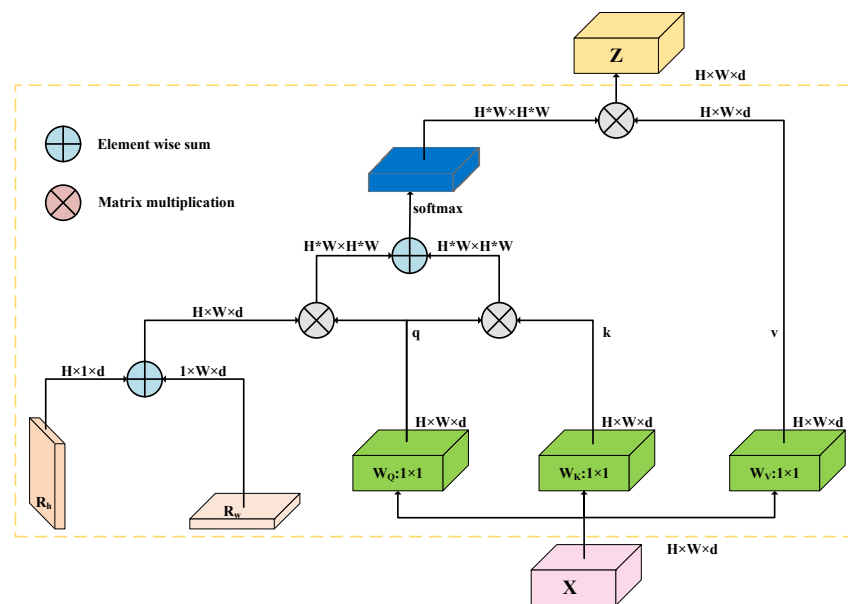


Figure 8. Single-layer structure of MHSA module [40].

The MHSA module uses relative position encoding R_h and R_w to capture the relative position relationship between features. The input X size is $H \times W \times d$, where H and W represent the height and width of the feature matrix, and d represents the dimension of each label.

MHSA performs pointwise convolution on input features to obtain the query code W_Q , key code W_K and value code W_V . The content information is obtained by multiplying the query code and key code matrices, and the position information is obtained by multiplying the relative position code with the query code matrices. Then, the position information and content information are matrix summed, the Softmax operation is performed, and then matrix multiplication is performed with the value encoding, and the final output result is shown in Equation (6):

$$Z(X) = F_{SM}(W_Q(X) \times (R_h + R_w) + W_Q(X) \times W_K(X)) \times W_V \quad (6)$$

The MHSA module takes into account not only the content information but also the relative distances between features at different locations, which enables it to effectively correlate cross-object information with location awareness. This unique design allows the MHSA module to fully utilize the correlation between features, thereby improving the network's focus on ship targets. By computing multiple heads in parallel, the model is able to better understand the environmental information around the ship and reduce the sensitivity to disturbances such as rain and fog, thus improving the model's ability and robustness to perceive complex scenes.

2.2.4. Small Ship Target Detection Head

Because in practical situations, it is usually fishing boats and small wooden boats that cause damage to buoys, the self-made BLOBS dataset contains a large number of small vessels. When using the original YOLOv8 algorithm for detection, its downsampling factor is large, which leads to the loss of microscopic information as the number of network layers increases and the receptive field expands. In this case, information about small-scale targets is spatially aggregated to a single point, which leads to a decrease in the accuracy of the model during the detection process.

In order to address the above issues, this study introduces an additional detection layer for smaller-size targets in the head section of YOLOv8. By upsampling the operation on the 80×80 feature map, the feature map size is changed from 80×80 , 40×40 and 20×20 to 160×160 , 80×80 , 40×40 and 20×20 , and the detection heads are changed from the original three to four. These modifications utilize the relatively comprehensive feature information in shallow feature maps, effectively ensuring a reduction in the network's receptive field, making the network pay more attention to the micro features, such as the shape and texture of small ships, thereby improving the recognition ability of small ships under adverse sea conditions.

3. Experiments

3.1. Evaluation Metrics

In order to comprehensively evaluate the performance of the YOLOv8-PMH algorithm in visible image ship detection, precision, recall, average precision (AP) and mAP are selected as evaluation metrics in this study. Where "m" in mAP denotes the average value of AP for all categories. In addition, mAP50 denotes the mAP value when the IoU threshold is 0.5, while mAP50:95 represents the mAP value when the IoU is incremented from 0.5 to 0.95 in steps of 0.05. These two metrics are the key indexes for evaluating the performance of the model, and the higher value of mAP implies that the model's detection effect is more excellent. The expression of each evaluation index is as follows:

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$AP = \int_0^1 P(R) dR \quad (9)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (10)$$

In the formula, TP represents the correct number of positive samples for prediction, which means that the YOLOv8-PMH algorithm correctly detects and locates ship targets. FP refers to an object that is actually a negative sample, but the network incorrectly predicts it as a positive sample. FN refers to an object that is actually a positive sample, but the network incorrectly predicts it as a negative sample.

3.2. Experimental Platform

All the experimental code sources in this study are in Python, using the Pytorch neural network architecture, with software installation versions of Python 3.8.10, torch 1.11.1 and cuda11.3. The NVIDIA GeForce RTX 3080 graphics card and 50 G RAM were used for the computational unit. Under the above experimental conditions, this study conducted network model training, and the specific parameters of the training are shown in Table 2.

Table 2. Configuration table for dataset training parameters.

Name	Number
resolution	640×640
optimizer	Adam
learning rate	0.01
momentum	0.937
batch size	16
epochs	200

3.3. Ablation Experiment

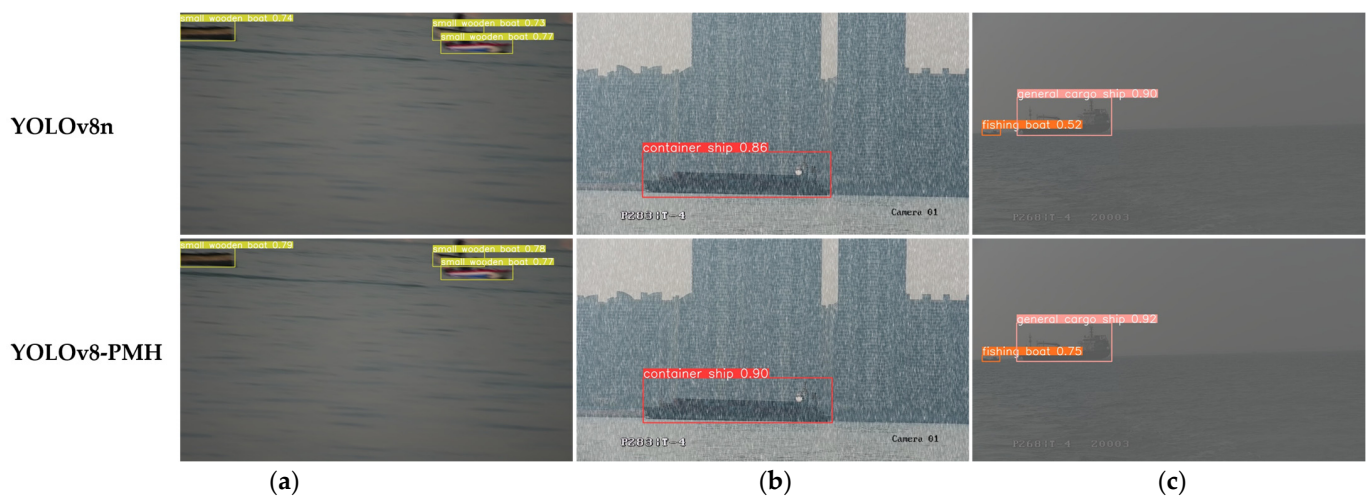
In order to verify the effectiveness of the improved method in identifying ship targets under complex sea conditions, seven ablation experiments were conducted on the BLOBS dataset using YOLOv8n as the baseline to analyze the impact and effectiveness of different modules on the YOLOv8-PMH algorithm. The comparison results of the ablation experiment are shown in Table 3.

It is obvious from Table 3 that in the ablation experiments, compared to the original YOLOv8 target detection algorithm, by adding the PSA mechanism to the original network architecture, the mAP50 is improved from 92.2% to 93.1%, and the mAP50:95 is improved from 70.4% to 71.6%, and these enhancements prove that the PSA mechanism has an augmentation of the spatial and channel dimensions. These improvements demonstrate the enhancement of the PSA mechanism in both spatial and channel dimensions, which ensures that the system is still able to effectively recognize ship information, even when the spatial resolution of the image is reduced due to shaking. Furthermore, after introducing the MHSA mechanism and small object detection head, the experimental results not only demonstrated stable performance in detecting large ships but also significantly improved the detection accuracy of small ships. The final comprehensive experimental data shows that the overall mAP50 has increased by 2%, and mAP50:95 has increased by 2.8%, which proves that the improvement strategy proposed in this paper has achieved certain improvements in detection performance indicators.

In addition, in order to show the effect of the improved model more intuitively, this paper also shows a visualization example of the original YOLOv8 target detection algorithm and the YOLOv8-PMH algorithm proposed in this paper in Figure 9.

Table 3. Ablation experiment.

Model	AP(%)						mAP50 (%)	mAP50:95 (%)
	CS	GCS	FB	PS	SWB	SB		
YOLOv8n	93.7	94.7	86.5	91.4	93.3	90.1	92.2	70.4
YOLOv8n + PSA	97.5	94.7	88.4	92.5	93.5	92	93.1	71.6
YOLOv8n + MHSA	97.4	94.6	88.1	92.2	93.4	91.7	92.9	71.3
YOLOv8n + 4-Head	97.2	94.5	88.9	92.3	93.5	91.8	93	71.3
YOLOv8n + PSA + MHSA	97.6	94.7	89.1	92.8	93.4	92.1	93.3	72.5
YOLOv8n + PSA + 4-Head	97.4	94.5	89.5	93.8	93.5	92.4	93.5	72.4
YOLOv8n + MHSA + 4-Head	97.5	94.6	90.5	93.4	93.5	92.2	93.6	72.8
YOLOv8n + PSA + MHSA + 4-Head	97.8	94.7	91.3	94.9	93.7	92.7	94.2	73.2

**Figure 9.** Comparison of detection results in different situations. (a) Fuzzy detection results; (b) rainy day detection results; (c) fog detection results.

By comparing the images, it can be found that the reduction in ship spatial resolution and partial occlusion caused by the shaking of the buoy significantly reduces the recognition accuracy of the YOLOv8 object detection algorithm. However, the YOLOv8-PMH algorithm improves this situation very well; even in images with low clarity or occlusion, it can still maintain high detection accuracy. In rainy and foggy weather, especially on foggy days with insufficient visibility, the problem of unimpressive detail features and difficult extraction of small target vessels is significant, and in practical situations, it is often small wooden boats as well as fishing boats and other small vessels that destroy buoys. However, the original YOLOv8 algorithm suffers from poor prediction of target size and position. In contrast, the YOLOv8-PMH algorithm proposed in this study possesses better feature extraction capabilities and has a better detection effect on all types of vessels. Therefore, the improved algorithm enables better evidence collection for such small vessels.

The PR curve is formed by the precision and recall of the model as the coordinate axes. Usually, due to the mutual constraint between precision and recall, the comprehensive performance of the model can be comprehensively understood by observing the enclosing area of the PR curve. In Figure 10, the PR curve of the original YOLOv8 model is shown on the left, and the PR curve of the YOLOv8-PMH model is shown on the right. By comparison, it is evident that the PR curve of the YOLOv8-PMH algorithm has a larger encircling area, especially the green curve about the fishing boat is obviously close to the upper-right corner, and this result intuitively demonstrates that the YOLOv8-PMH algorithm has significantly improved the detection results.

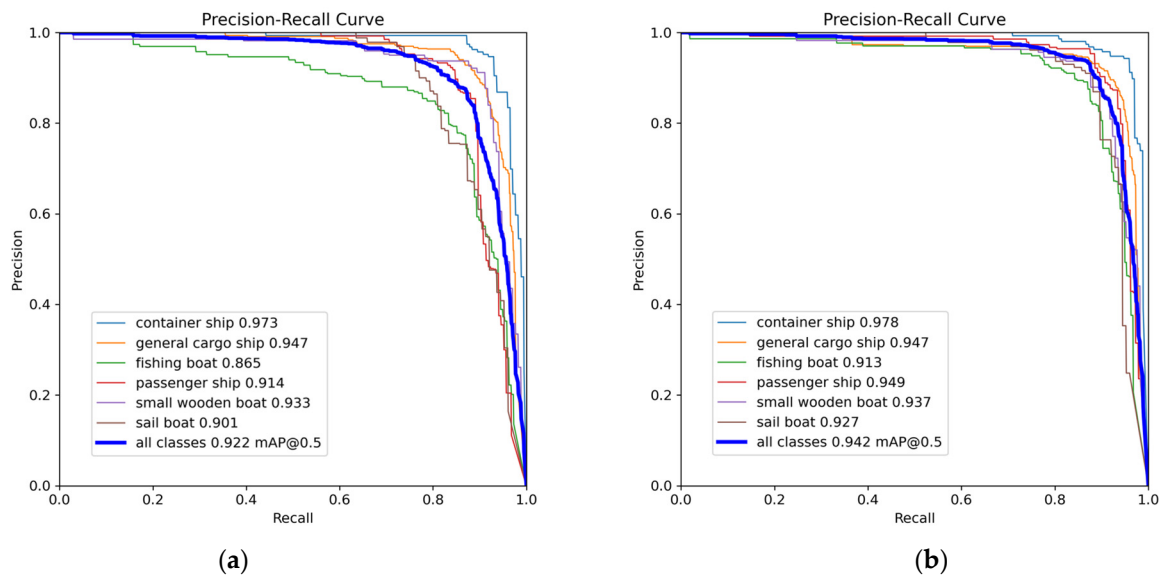


Figure 10. PR Curve comparison chart. (a) YOLOv8; (b) YOLOv8-PMH.

3.4. Performance Comparison of Multiple Models

In order to fully validate the effectiveness of the improved algorithm, this study conducted several sets of comparative experiments to compare and analyze its performance with a series of superior target detection algorithms, including Faster R-CNN (ResNet50), YOLOv3-tiny, YOLOv4-tiny, YOLOv5n, YOLOv6n and YOLOv8n. The experimental results are shown in Table 4.

Table 4. Comparison of ship detection results between YOLOv8-PMH and other networks.

Model	AP(%)						mAP50 (%)	mAP50:95 (%)
	CS	GCS	FB	PS	SWB	SB		
Faster R-CNN	97.4	94.5	87.4	90.9	86.5	89.4	92.2	-
SSD	96.1	93.9	82.8	90.3	84.2	85.5	88.8	-
YOLOv3-tiny	96.3	94.3	82.1	91.6	89.6	89.2	90.5	65.5
YOLOv4-tiny	97.3	94.6	88.2	91.4	91.8	79.5	90.5	64.4
YOLOv5n	97.2	93.6	86.3	92.6	91.6	85.3	91.1	68.2
YOLOv6n	97.5	94.6	85.6	92.8	90.8	88.2	91.6	70.5
YOLOv8n	97.3	94.7	86.5	91.4	93.3	90.1	92.2	70.4
YOLOv8-PMH	97.8	94.7	91.3	94.9	93.7	92.7	94.2	73.2

Based on the comparison results of the different models mentioned above, the YOLOv8-PMH algorithm proposed in this study achieves 94.2% for mAP50 and 73.2% for mAP50:95, which collectively shows the best detection performance on the homemade dataset. When detecting large ships, such as container ships and regular cargo ships, the detection performance of each model is quite good. However, for smaller vessels such as fishing boats, passenger ships, and sailboats that are more likely to cause damage to buoys, the YOLOv8-PMH algorithm exhibits superior performance.

In order to more intuitively highlight the superiority of the algorithm proposed in this article and its performance comparison with other detection algorithms, a visual example of the YOLOv8-PMH algorithm and other existing algorithms is shown in Figure 11.

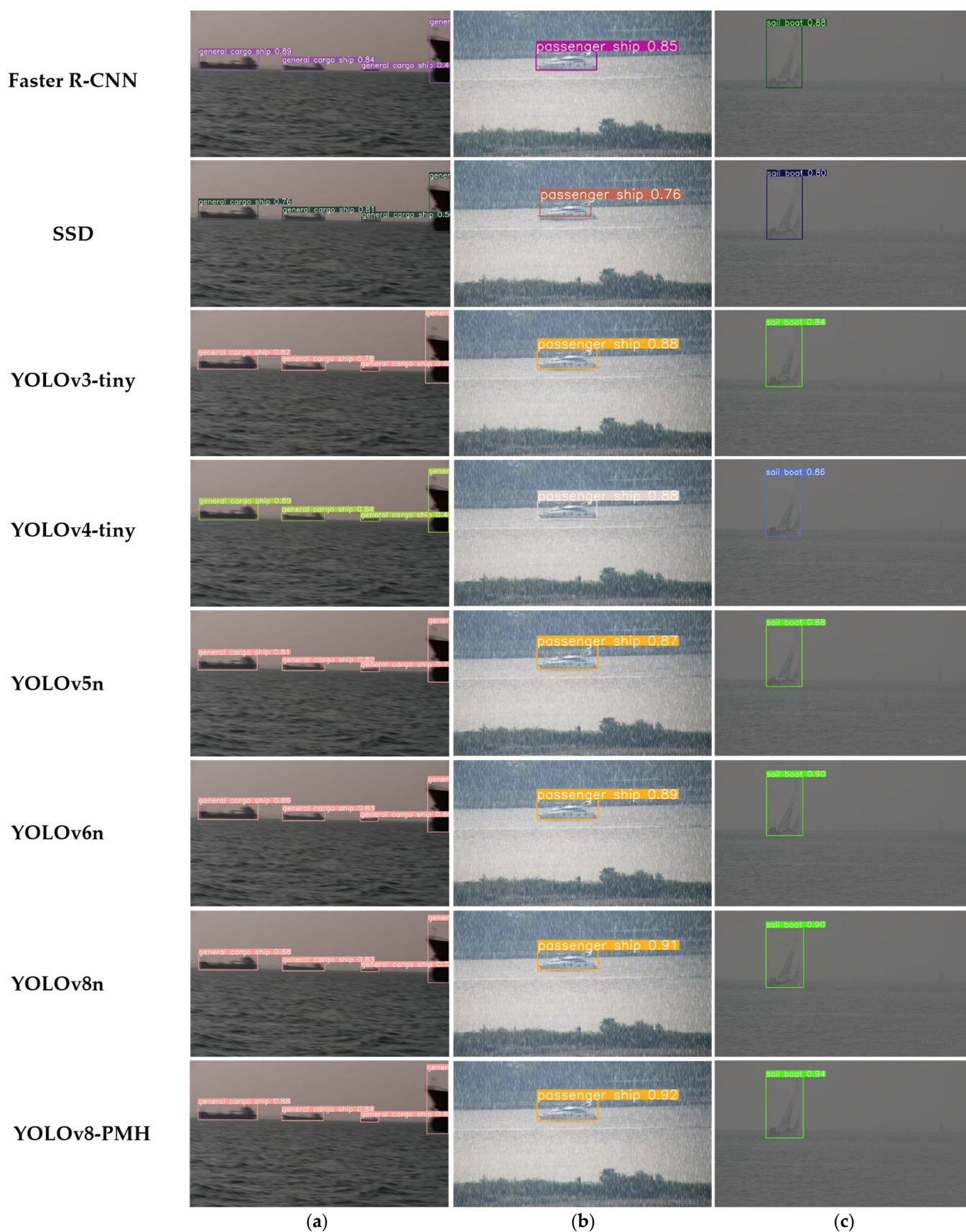


Figure 11. Comparison of experimental results of YOLOv8-PMH with other algorithms. (a) Fuzzy detection results; (b) rainy day detection results; (c) fog detection results.

By comparing the images, it can be found that YOLOv3-tiny and YOLOv5n have a slightly inferior detection performance compared to other object detection algorithms in images with lower resolution caused by blur. Faster R-CNN performs well in detecting large targets but performs poorly in detecting small targets. In the comprehensive analysis, the YOLOv8-PMH algorithm proposed in this paper is able to accurately localize and identify the targets in various test scenarios compared to other algorithms, demonstrating excellent detection performance, and can be well applied to the ship warning task at the end of the buoy.

In order to better demonstrate the performance of the algorithm on different datasets, the SeaShips dataset was processed and experimented with in this study using the same method. The experimental results are shown in Table 5.

Table 5. Comparison of detection accuracy in SeaShips dataset.

Model	AP(%)						mAP50 (%)	mAP50:95 (%)
	CS	GCS	FB	PS	BCC	OC		
YOLOv8n	98.4	95.4	93	87	91.7	94.6	93.3	68.6
YOLOv5 + BiFPN [24]	98	96.7	94.3	87.5	94.7	96.1	94.6	69.5
YOLOX + ECA [25]	98.2	95.5	93.7	86.9	95.7	96.3	94.4	69.3
YOLOv8-PMH	98.6	96.5	94.1	88.9	94.9	97.5	95.1	70.3

The experimental results show that the improved algorithm's public dataset still maintains better detection results.

4. Conclusions

This study aims to improve the ship detection capability of ocean buoys in complex weather conditions in order to provide technical support for subsequent warning and evidence collection at the buoy end. In order to achieve this goal, this paper proposes an improved YOLOv8 target detection algorithm by simulating the actual working situation for the buoy platform, which has been in a wet environment for a long time at sea and keeps swaying. Firstly, in the image preprocessing stage, a 50×50 blurring kernel is used to simulate the visual blur caused by buoy swaying, and Gaussian noise and center point synthetic fog are used to simulate a rainy and foggy environment. Subsequently, this study compared and analyzed the target detection accuracy of the original algorithm and the YOLOv8-PMH algorithm, and the research results showed that:

1. The YOLOv8-PMH algorithm improved with PSA attention maintains $C/2$ in the channel dimension and $[H, W]$ in the spatial dimension, a mechanism that effectively preserves the high-resolution information in the original deep convolutional neural network and significantly reduces the effect of blurring due to camera shake on information loss. Especially in the ship target recognition task, the C2f module with integrated PSA attention leads to a 0.9% improvement in the mAP50 of the algorithm and a 1.2% improvement in mAP50:95;
2. The MHSA attention mechanism based on the Transformer architecture effectively reduces rain and fog background interference and enhances the ability of feature fusion. After introducing this attention into the neck module, the algorithm achieved a 0.7% improvement in mAP50 and a 0.9% improvement in mAP50:95;
3. Based on the original algorithm, the newly designed small ship target detection head significantly improves the feature extraction capability for small ships without affecting the detection performance for large ship targets. With this improvement, the mAP50 of the algorithm realizes a 0.8% improvement, and the mAP50:95 obtains a 0.9% improvement;
4. After integrating the above improvements, compared with the original YOLOv8 algorithm, the mAP50 of the YOLOv8-PMH algorithm has increased by 2%, and the mAP50:95 has increased by 2.8%.

Through extensive experiments and comparisons with other benchmark algorithms, the YOLOv8-PMH algorithm proposed in this paper has significant advantages in terms of detection performance. However, there are still some aspects that can be improved. Firstly, the dataset is still insufficient, and more ship data will continue to be collected and labeled in the future. Secondly, methods such as target ranging and buzzer alarms will be fused in the future to provide hierarchical warnings and forensics for ship targets. In addition, this article has improved the original algorithm by adding additional small object detection heads, but this has added a greater computational load. Thus, although the algorithm has a high detection accuracy and can be well monitored near the shore, energy will be a major problem when located in the deep and distant sea, so future work will consider simplifying the model structure and improving the reasoning speed, with a view to realizing better real-time monitoring and warning and forensics.

Author Contributions: Conceptualization, W.L. and C.N.; Methodology, W.L., C.N. and Y.F.; Software, W.L.; Validation, W.L.; Formal analysis, C.N. and P.Z.; Investigation, W.L.; Resources, C.N., C.L. and G.Y.; Data curation, G.Y., C.N. and C.L.; Writing—original draft, W.L., C.N. and G.Y.; Writing—review and editing, C.N., C.L., Y.F. and P.Z.; Visualization, W.L.; Supervision, C.N., C.L. and Y.F.; Project administration, C.N.; Funding acquisition, C.N. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China [2022YFC3104301] and finally supported by the Laoshan Laboratory [LSKJ202201601].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Xu, R.; Wang, H.; Xi, Z.; Wang, W.; Xu, M. Recent Progress on Wave Energy Marine Buoys. *J. Mar. Sci. Eng.* **2022**, *10*, 566. [\[CrossRef\]](#)
2. Canepa, E.; Pensieri, S.; Bozzano, R.; Faimali, M.; Traverso, P.; Cavaleri, L. The ODAS Italia 1 buoy: More than forty years of activity in the Ligurian Sea. *Prog. Ocean.* **2015**, *135*, 48–63. [\[CrossRef\]](#)
3. Park, Y.W.; Kim, T.W.; Kwak, J.S.; Kim, I.K.; Park, J.E.; Ha, K.H. Design of Korean Standard Modular Buoy Body Using Polyethylene Polymer Material for Ship Safety. *J. Mater. Sci. Chem. Eng.* **2016**, *4*, 65–73. [\[CrossRef\]](#)
4. Li, X.; Bian, Y. Modeling and prediction for the Buoy motion characteristics. *Ocean Eng.* **2021**, *239*, 109880. [\[CrossRef\]](#)
5. Teng, C.; Cucullu, S.; Mcarthur, S.; Kohler, C.; Burnett, B. Buoy Vandalism Experienced by NOAA National Data Buoy Center. In Proceedings of the OCEANS 2009, Biloxi, MS, USA, 26–29 October 2009; pp. 1–8.
6. Hwang, H.G.; Kim, B.S.; Kim, H.W.; Gang, Y.S.; Kim, D.H. A development of active monitoring and approach alarm system for marine buoy protection and ship accident prevention based on trail cameras and AIS. *J. Korea Inst. Inf. Commun. Eng.* **2018**, *22*, 1021–1029.
7. Zheng, Y. Application of CCTV monitoring equipment in navigation mark management. *China Water Transp.* **2019**, *19*, 101–102.
8. Zhao, T.; Qi, J.; Ruan, D.; Shan, R. Design of Marine buoy early warning system based on geomagnetic and infrared dual mode detection. *J. Mar. Technol.* **2017**, *36*, 15–21.
9. Chen, S.; Xiang, H.; Gao, S. Intelligent anti-collision warning method of fairway buoy based on passive underwater acoustic positioning. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Hulun Buir, China, 28–30 August 2020.
10. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
11. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
12. 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.* **2017**, *39*, 1137–1149. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Gu, J.; Li, B.; Liu, K. Infrared ship target detection Algorithm based on Improved Faster R-CNN. *Infrared Technol.* **2021**, *43*, 170–178.

14. Redmon, J.; Divvala, S.; Girshick, R.; Farhad, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
15. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
16. Bochkovskiy, A.; Wang, C.; Liao, H. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
17. Ultralytics. YOLOv5: Object Detection. Available online: <https://github.com/ultralytics/yolov5> (accessed on 18 May 2020).
18. Wang, C.; Bochkovskiy, A.; Liao, H. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696.
19. Ultralytics/Ultralytics. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 20 November 2023).
20. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In *Computer Vision—ECCV 2016, Proceedings of the 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016*; Proceedings, Part I 14; Lecture Notes in Computer Science; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
21. Zhang, Q.; Li, Y.; Zhang, Z.; Yin, S.; Ma, L. Marine target detection for PPI images based on YOLO-SWFormer. *Alex. Eng. J.* **2023**, *82*, 396–403. [\[CrossRef\]](#)
22. Zheng, Y.; Zhang, Y.; Qian, L.; Zhang, X.; Diao, S.; Liu, X.; Huang, H. A lightweight ship target detection model based on improved YOLOv5s algorithm. *PLoS ONE* **2023**, *18*, e0283932. [\[CrossRef\]](#)
23. Shang, Y.; Yu, W.; Zeng, G.; Li, H.; Wu, Y. StereoYOLO: A Stereo Vision-Based Method for Maritime Object Recognition and Localization. *J. Mar. Sci. Eng.* **2024**, *12*, 197. [\[CrossRef\]](#)
24. Wang, J.; Pan, Q.; Lu, D.; Zhang, Y. An Efficient Ship-Detection Algorithm Based on the Improved YOLOv5. *Electronics* **2023**, *12*, 3600. [\[CrossRef\]](#)
25. Zhao, Q.; Wu, Y.; Yuan, Y. Ship Target Detection in Optical Remote Sensing Images Based on E2YOLOX-VFL. *Remote Sens.* **2024**, *16*, 340. [\[CrossRef\]](#)
26. Si, J.; Song, B.; Wu, J.; Lin, W.; Huang, W.; Chen, S. Maritime Ship Detection Method for Satellite Images Based on Multiscale Feature Fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 6642–6655. [\[CrossRef\]](#)
27. Wang, Y.; Wang, B.; Huo, L.; Fan, Y. GT-YOLO: Nearshore Infrared Ship Detection Based on Infrared Images. *J. Mar. Sci. Eng.* **2024**, *12*, 213. [\[CrossRef\]](#)
28. Prasad, D.K.; Rajan, D.; Rachmawati, L.; Rajabally, E.; Quek, C. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 1993–2016. [\[CrossRef\]](#)
29. 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. [\[CrossRef\]](#)
30. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
31. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11534–11542.
32. Cao, Y.; Xu, J.; Lin, S.; Wei, F.; Hu, H. Gcnet: Non-local networks meet squeeze-excitation networks and beyond. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27–28 October 2019; pp. 1971–1980.
33. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 3146–3154.
34. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
35. Liu, H.; Liu, F.; Fan, X.; Huang, D. Polarized self-attention: Towards high-quality pixel-wise regression. *arXiv* **2021**, arXiv:2107.00782.
36. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.
37. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Houtsby, N. An image is worth 16×16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
38. Touvron, H.; Cord, M.; Douze, M.; Massa, F.; Sablayrolles, A.; Jégou, H. Training data-efficient image transformers & distillation through attention. *arXiv* **2021**, arXiv:2012.12887.
39. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. *arXiv* **2021**, arXiv:2103.14030.
40. Srinivas, A.; Lin, T.Y.; Parmar, N.; Shlens, J.; Abbeel, P.; Vaswani, A. Bottleneck transformers for visual recognition. *arXiv* **2021**, arXiv:2101.11605.

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.