Journal of
Marine Science and Engineering

MDPI

Article

A Novel Intelligent Ship Detection Method Based on Attention Mechanism Feature Enhancement

Yingdong Ye, Rong Zhen *, Zheping Shao, Jiacai Pan and Yubing Lin

Navigation College, Jimei University, Xiamen 361021, China
* Correspondence: zrandsea@163.com

Abstract: The intelligent perception ability of the close-range navigation environment is the basis of autonomous decision-making and control of unmanned ships. In order to realize real-time perception of the close-range environment of unmanned ships, an enhanced attention mechanism YOLOv4 (EA-YOLOv4) algorithm is proposed. First of all, on the basis of YOLOv4, the convolutional block attention module (CBAM) is used to search for features in channel and space dimensions, respectively, to improve the model's feature perception of ship targets. Then, the improved-efficient intersection over union (EIoU) loss function is used to replace the complete intersection over union (CIoU) loss function of the YOLOv4 algorithm to improve the algorithm's perception of ships of different sizes. Finally, in the post-processing of algorithm prediction, soft non-maximum suppression (Soft-NMS) is used to replace the non-maximum suppression (NMS) of YOLOv4 to reduce the missed detection of overlapping ships without affecting the efficiency. The proposed method is verified on the large data set SeaShips, and the average accuracy rate of $mAP^{0.5–0.95}$ reaches 72.5%, which is 10.7% higher than the original network YOLOv4, and the FPS is 38 frames/s, which effectively improves the ship detection accuracy while ensuring real-time performance.

Keywords: water transportation; target detection; unmanned ship; deep learning; attention mechanism

Citation: Ye, Y.; Zhen, R.; Shao, Z.; Pan, J.; Lin, Y. A Novel Intelligent Ship Detection Method Based on Attention Mechanism Feature Enhancement. *J. Mar. Sci. Eng.* **2023**, *11*, 625. https://doi.org/10.3390/jmse11030625

Academic Editor: Claudio Ferrari

Received: 23 February 2023
Revised: 12 March 2023
Accepted: 14 March 2023
Published: 16 March 2023

1. Introduction

With the development of artificial intelligence and unmanned driving technology, unmanned ships have become an important research field in intelligent maritime transportation. The development of unmanned ships includes four stages: perception, understanding, decision-making, and control [1–4], among which perception is the basis for autonomous decision-making and control. Traditional environmental perception methods are affected by timeliness and cannot meet the requirements of real-time and accuracy of unmanned ships. Constructing a visible light detection system composed of a panoramic vision system can enable real-time and intuitive monitoring of dynamic water environments [5–7]. Furthermore, using computer vision and deep learning methods to detect and track ships in the system can effectively improve the efficiency of unmanned ship environmental perception. Therefore, the panoramic vision-based visible light ship detection method has received wide attention from scholars [3].

Currently, ship target detection under visible light can be divided into two categories: traditional feature-based methods and deep learning-based methods using convolutional neural networks. Traditional methods mostly rely on modeling based on target shape, background saliency contrast, and other factors. The algorithm depends on given parameters that satisfy specific regional conditions, which limits its accuracy and generalization ability in complex scenes [8]. In recent years, with the development of deep learning technology, many scholars have begun to use deep learning frameworks to recognize maritime targets. Deep learning detection methods can be classified into two categories: one-stage and two-stage methods [9–11]. Two-stage methods divide the detection process into two stages: first, generating a series of regions where targets may exist, and then searching

and identifying targets within these regions. For example, Liu et al. [12] modified the region extraction network in Fast-RCNN [13] to ResNet-101 [14] and combined the global and local features of proposed regions using the multi-region feature fusion module to improve detection performance of ships in different directions. Yu et al. [15] improved the feature weighting method in Faster-RCNN to provide more suitable feature fusion for the background differentiation of difficult ships in busy waters. In the far shore scene with low discrimination between complex background and ship target, ship target can be recognized. Two-stage methods have high accuracy in detecting ships, but these methods take longer time and cannot achieve real-time detection. Therefore, to improve detection speed, many scholars use one-stage methods to improve ship detection efficiency. These methods use regression analysis principles to obtain the position and category of the target in a single detection. For example, Liu et al. [16] redesigned ship anchor box size based on YOLOv3 [17], introduced soft non-maximum suppression, and reconstructed mixed loss functions to improve the network's learning and expression ability for ship features. Hong et al. [18] used a residual network instead of continuous convolution operation in YOLOv4 to solve the problem of network degradation and gradient disappearance, and established a nonlinear target tracking model based on UKF method, which improved the accuracy of ship detection.

Although the deep learning method has made some progress, it still cannot solve the following problems well in practical use, such as: (1) distant ship targets are smaller and are often obstructed, (2) the background areas typically contain floating objects, shorelines, and other interferences, leading to occurrences of false positives and false negatives, and (3) video images captured under fixed monitoring can be easily affected by adverse weather conditions, such as heavy fog, rain, or snow, further increasing the difficulty of ship detection. To effectively address the aforementioned problems and further improve the accuracy of ship detection using deep learning algorithms, this paper proposes a ship detection algorithm EA-YOLOv4 based on attention mechanism feature enhancement. The main contributions of the proposed algorithm are as follows:

(1) In order to improve the accuracy of ship type identification and the recall of ships in bad weather, a CBAM module is embedded in YOLOv4 structure. The spatial attention mechanism and channel attention mechanism in this module change the search weight of YOLOv4, making the network structure focus on the unique characteristics and effective channels of ships.
(2) In terms of loss function, EIoU is used to replace the CIoU of YOLOv4. CIoU performs well in general target detection, but ship targets have relatively fixed aspect ratio characteristics. The use of EIoU can better identify the ship position, speed up the algorithm fitting, and improve the ship positioning ability.
(3) Overlapping of ship targets at sea is common. In order to improve the detection ability of occluded targets, Soft-NMS is used to replace the NMS of YOLOv4 to post-process the algorithm output, improve the network's attention to overlapping targets, and further improve the ship detection performance while ensuring the solution efficiency.

The rest of this article is organized as follows. In Section 2, the algorithm EA-YOLOv4 in this paper is proposed, including the CBAM to change the network structure, the EIoU loss function replacement and the improved Soft-NMS. In Section 3, all the improved algorithms are ablated, and the best set of results are selected as the algorithm EA-YOLOv4 in this paper. EA-YOLOv4 is compared with other similar algorithms, and the advantages of the algorithm are analyzed. Section 4 presents the conclusions and future research directions.

2. EA-YOLOv4 Algorithm

In this section, we will provide a detailed introduction to the algorithm from four aspects: algorithm overview, multi-dimensional attention mechanism feature enhancement extraction, loss function improvement, and non-maximum suppression improvement.

2.1. Algorithm Overview

The algorithm proposed in this paper is illustrated in Figure 1, with YOLOv4 [19] selected as the backbone network. To enhance the network's ability to perceive the scale features of ships of different sizes, a spatial attention mechanism is adopted to increase the network's focus on ship scale features. In order to avoid the network ignoring different dimensional channel information, a combination of channel attention mechanisms is inserted to form the CBAM [20] structure, obtaining $f_3 - f_5$ as the feature representation. Next, in order to improve the detection position optimization problem of traditional CIoU [21] loss function, which only focuses on aspect ratio, an EIoU [22] loss function is employed to increase the network's attention to ship size and enhance its perception of ships of different sizes. Finally, to enhance the detection ability of overlapping targets, an improved Soft-NMS [23] is used to perform secondary screening on the output results, while ensuring detection speed and accuracy. The CBAM, Soft-NMS, and EIoU loss will be discussed in detail in subsequent sections.

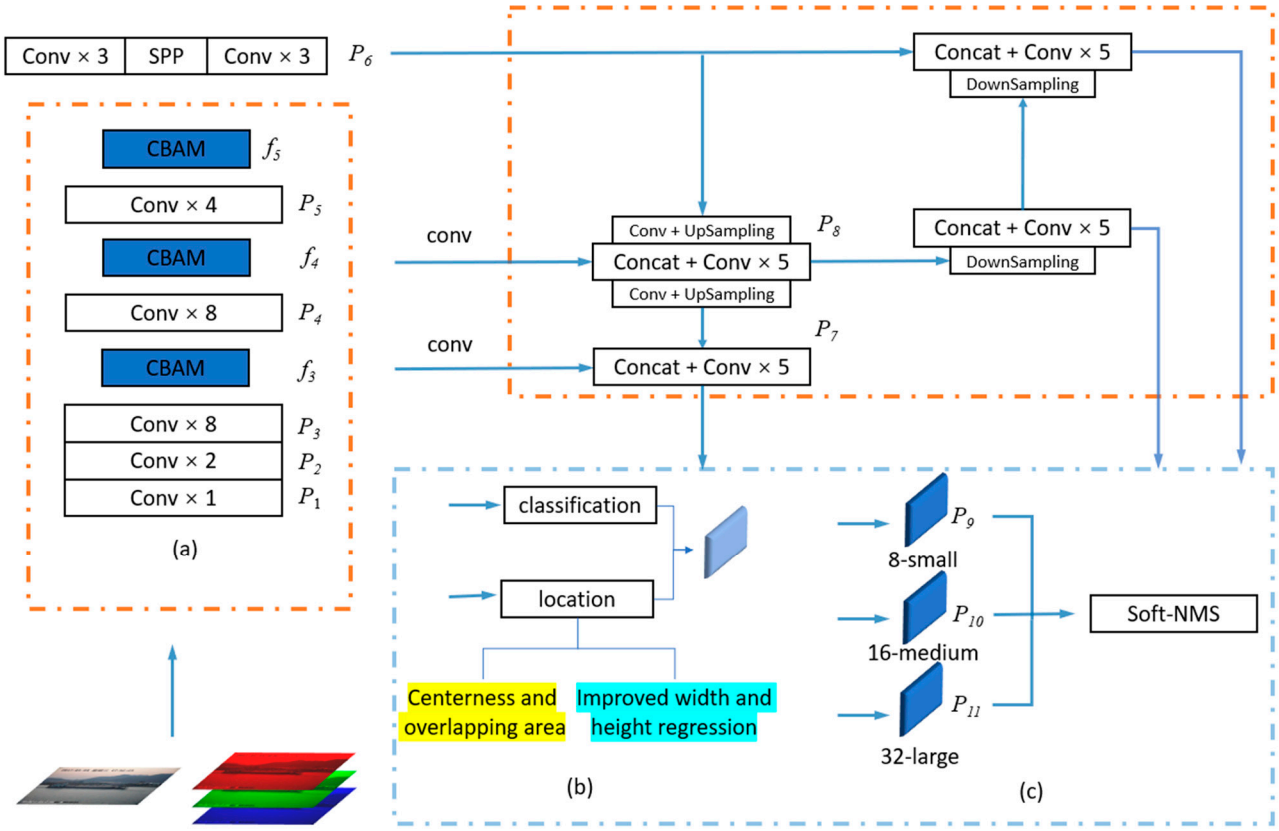


Figure 1. Algorithm structure diagram of EA-YOLOv4; (**a**) improvement of feature extraction network; (**b**) improvement of loss function; (**c**) improvement of loss function.

2.2. Multi-Dimensional Attention Mechanisms and Their Feature Enhancement

To enhance the feature representation capability of ships in complex backgrounds and improve their salience, this paper inserts attention mechanisms for feature selection in the network skeleton. Hu et al. [24] proposed using channel attention mechanisms to strengthen network feature extraction capability, but using only channel attention mechanisms ignores the spatial semantic information of ships. Since different ships have obvious differences in spatial representation, searching for ship spatial information can help the algorithm to quickly and accurately determine the ship type. M et al. [25] introduced a new learnable module, spatial transformer, which enables the neural network to actively transform the feature map according to the feature map itself, but this algorithm did not consider the semantic information of feature map channels, which resulted in too many computing

resources being occupied by irrelevant channels that negatively affected detection results. The above research shows that the use of both spatial and feature attention mechanisms can enhance the algorithm's recognition capability. To comprehensively consider ship spatial information and feature map channel information, this paper introduces CBAM as an embedded module of multi-dimensional attention mechanisms to improve the feature extraction network's extraction capability.

As shown in Figure 2, the schematic diagram of CBAM includes a channel attention mechanism in Figure 2a and a spatial attention mechanism in Figure 2b. The CBAM attention mechanism combines the advantages of channel and spatial attention mechanisms. A CBAM unit takes any tensor $X = [x_1, x_2, \ldots, x_c] \in R^{C \times H \times W}$ as input and outputs a tensor $Y = [y_1, y_2, \ldots, y_c] \in R^{C \times H \times W}$ of the same size. In order to make CBAM focus on both channel and spatial information, channel attention mechanism is first implemented on the input feature map as follows:

$$F_c = \sigma\{MLP[GAP(X)] + MLP[GMP(X)] \tag{1}$$

where X is the input feature map, GAP is the global average pooling, GMP is the global maximum pooling, MLP is the multi-layer perceptron, and σ is the activation function defined as:

$$\sigma = \frac{1}{1 + e^{-x}} \tag{2}$$

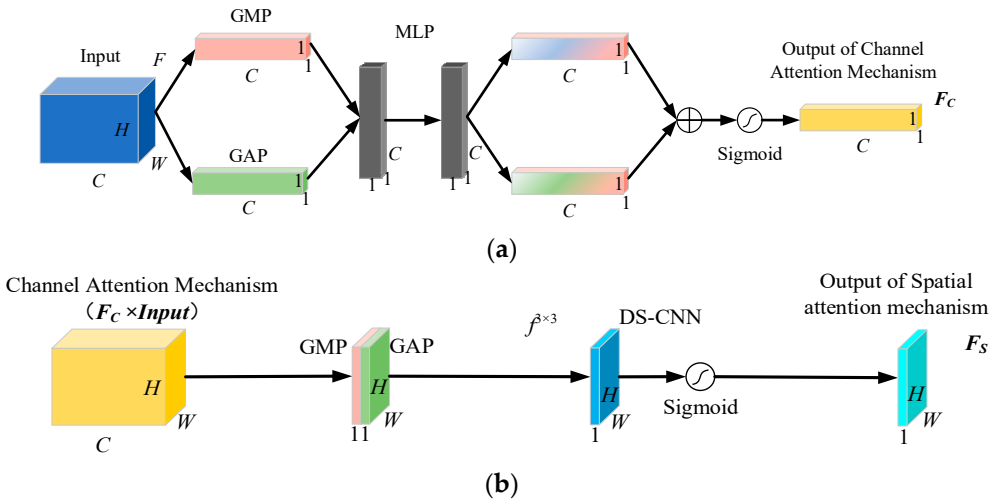


Figure 2. Multi-dimensional attention mechanism CBAM; (**a**) channel attention mechanism; (**b**) spatial attention mechanism.

After the channel attention, a vector $F_c = (\alpha_1, \alpha_2, \ldots, \alpha_c) \in R^{C \times 1 \times 1}$ of dimension C is obtained, where $\alpha_i, i = 1, 2, \ldots, c$ represents the allocation weights of different channels in the original feature map. When the detection platform encounters situations such as heavy fog or water surface reflection, attention tracking on channels helps the algorithm focus on effective information and reduce the impact of noise on recognition results.

The spatial attention mechanism is added after the channel attention, as shown in Figure 2b, and is implemented as follows:

$$F_s = \sigma\left\{f^{3\times3}\begin{bmatrix} GAP(Y_1) \\ GMP(Y_1) \end{bmatrix}\right\} \tag{3}$$

where Y_1 comes from channel attention mechanism output Y_c dot product with the original feature map X, which reflects the channel region of interest in the original feature map.

$f^{3\times3}$ is the convolutional layer with 3×3 kernel size, and F_s is the weight matrix obtained after the spatial attention mechanism, whose expression is:

$$F_s = \begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,w} \\ \beta_{2,1} & \beta_{2,2} & \cdots & \beta_{2,w} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{H,1} & \beta_{H,2} & \cdots & \beta_{H,W} \end{pmatrix} \tag{4}$$

where $\beta_{i,j}, i = 1, 2, \ldots, H; j = 1, 2, \ldots, W$ represents the allocation weights of spatial coordinates, which will be adaptively adjusted according to the search target. Finally, the matrix is multiplied by Y_1 to obtain the final output $Y = [y_1, y_2, \ldots, y_c] \in R^{C\times H\times W}$. The output can simultaneously resample the spatial structure information of ships on the basis of focusing on channel information.

2.3. Optimizing the Loss Function

The loss function affects the convergence speed of the model and the fitting performance of the evaluated algorithm. In the original YOLOv4 network, the bounding box regression loss is calculated using the CIoU function, as shown in Equation (5):

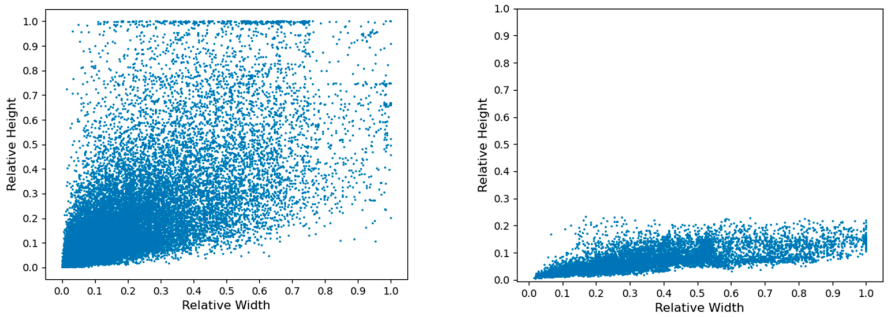
$$L_{CIoU} = 1 - CIoU = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha\frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2 \tag{5}$$

In the equation, b, b^{gt} denotes the predicted box and the reference box, where the reference box is the true target used to guide the model learning. IoU represents the intersection over union, which reflects the overlapping area between the two boxes. $\rho^2(b, b^{gt})$ represents the Euclidean distance between the two boxes.c is the diagonal distance between the minimum enclosing region containing the two boxes. α is a correction parameter. $\frac{w^{gt}}{h^{gt}}$ and $\frac{w}{h}$ represent the aspect ratios of the reference boxes and predicted boxes, respectively.

On the basis of IoU, CIoU adds center point distance detection and predicted box width and height detection, making the relative position between the predicted box and the reference box more accurate. However, compared with conventional detection tasks, ship detection in complex marine environments involves many small targets, and the shapes of marine targets are relatively fixed. Therefore, the aspect ratio used by CIoU is not suitable for the bounding box regression loss in ship target detection. In order to analyze the differences between ship targets and traditional targets and optimize the loss function, this paper compares the objects in the CoCo2017test (CoCo) dataset [26] and the SeaShips dataset [27]. The data is processed using linear normalization, and the results are shown in Figure 3. The results show that the aspect ratio distribution of the objects in the CoCo dataset is relatively balanced, while the aspect ratio distribution of the objects in the SeaShips dataset is close to a linear function, and the aspect ratios of the SeaShips dataset objects are relatively close. Therefore, the aspect ratio used by CIoU cannot accurately locate the ship's position. To improve this, this paper introduces EIoU, and its formula is given by Equation (6):

$$L_{EIoU} = 1 - EIoU = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \tag{6}$$

where b, b^{gt} denotes the predicted box and the reference box, $\rho(w, w^{gt})$ is the difference in width between the two boxes, $\rho(h, h^{gt})$ is the difference in height between the two boxes, and C_w, C_h is the width and height of the minimum outer box that covers the two boxes. When the EIoU value is larger, it means that the predicted box is closer to the position of the reference box, and the network loss is smaller. During network training, the network parameters are adjusted based on the overlap area, center point, and width and height values to make the network converge faster.



(a) (b)

Figure 3. Relative shape of targets under different data sets; (**a**) CoCo dataset target box relative value; (**b**) relative value of SeaShips dataset target box.

2.4. Improvement of Soft-Non-Maximum Suppression

In object detection algorithms, NMS is used to filter out overlapping detection boxes for the same object. The specific processing method is as follows:

$$s_i = \begin{cases} s_i & IoU(M, b_i) < N_t \\ 0 & IoU(M, b_i) > N_t \end{cases} \tag{7}$$

where M is the box with the highest score, IoU represents the degree of overlap between the highest-scoring box and other candidate boxes, and N_t is the threshold. NMS removes detection boxes within the threshold of the highest-scoring box, but the problem is that it will force high-scoring detection boxes adjacent to the target box to be removed, resulting in missed detection of occluded ship targets with high confidence. In order to balance the problem of repeated detection of targets and missed detection of occluded ship targets, Soft-NMS is used instead of traditional NMS for post-processing. The Soft-NMS processing method is shown in the equation below:

$$s_i = \begin{cases} s_i & IoU(M, b_i) < N_t \\ s_i[1 - IoU(M, b_i)] & IoU(M, b_i) > N_t \end{cases} \tag{8}$$

Soft-NMS simultaneously considers the score and overlap degree, and sets a penalty term for the occluded ship targets with higher scores to avoid missed detection while also ensuring that the same target is not detected repeatedly. As shown in Figure 4a, the two ships to be detected are a cargo ship and an occluded target ore ship, and the IoU is about 0.45. When using traditional NMS algorithm, the cargo ship and ore ship detection boxes are too close, causing the algorithm to consider the detected cargo ship and ore ship as the same target, thus directly ignoring the cargo ship. Figure 4b uses Soft-NMS to process the same target, and when the IoU of the cargo ship and ore ship is greater than given threshold, the score of the smaller cargo ship is reduced by such IoU. At this time, the algorithm considers the probability of the existence of a cargo ship here to be 0.53, which meet the set threshold and the occluded target is detected.

(a) (b)

Figure 4. Processing results of overlapping targets by NMS and Soft-NMS; (**a**) NMS processing mode; (**b**) Soft-NMS processing mode.

3. Experiment and Analysis

3.1. Experimental Platform and Dataset

The experimental environment in this article is based on the Windows platform, with an i7-10700F CPU, 32GB of memory, and an NVIDIA® GeForce® RTX 2070 Super GPU processor with 8GB of video memory. The experimental framework is built using the Python programming language. The deep learning development environment includes Python 3.7.8, PyCharm2019, Anaconda3.4.1, TensorFlow-GPU 2.3.0, CUDA 10.1.234, and cuDNN 7.6.5.

The SeaShips dataset was used to validate the ship detection algorithm. All images in the dataset come from approximately 1080 real video clips, including 7000 ship images of six categories. The six categories of images are ore carrier (OC), fishing boats (FB), container ships (CS), bulk carrier (BC), general cargo ship (GCS), and passenger ships (PS). The notable challenges include small targets, high overlap between coastline and ships, changes in brightness, and ship occlusion. The data set is randomly divided into training set, verification set and test set according to 2:1:1 ratio. That is, 3500 pictures in the training set, 1750 pictures in the verification set and 1750 pictures in the test set. The validation set is used to test the model training results. When the loss of the model on the verification set does not decrease for 20 consecutive rounds, the optimal solution is considered to be reached. The distribution of ship numbers in each set is shown in Figure 5. The proportion of the three datasets is consistent across different ship targets, ensuring that the algorithm has good robustness.

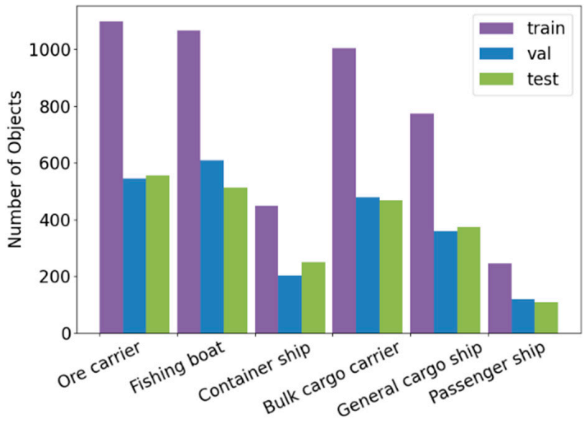


Figure 5. Distribution of six categories of ships in training set, verification set and test set.

3.2. Hyperparameters and Evaluation Metrics

The hyperparameters set for the experiment are shown in Table 1, where SGD refers to stochastic gradient descent. Before training, various data augmentation strategies were employed, such as horizontal rotation, mosaic, image cropping, and adjustments to brightness and contrast, to expand the training set. During training, all epochs were saved, and the epoch with the best performance on the validation set was selected as the test model for the current model. In the testing phase, the target confidence threshold and Soft-NMS threshold were both set to 0.5. The maximum training epoch is set to 300.

Table 1. Experiment parameter setting.

Hyperparameter	Value
Epoch	300
Batch size	4
Optimizer	SGD
First learning rate	1×10^{-4}
Last learning rate	1×10^{-5}
Learning rate decay rate	0.1
Weight decay	5×10^{-4}

To fully verify the ship detection capability of the model in complex scenes, the mean average precision (mAP) was used to evaluate the algorithm, which reflects the mean of AP under different IoU settings. At the same time, the frames per second (FPS) was used to measure the detection speed of the algorithm, which indicates the number of images processed per second. The AP can be expressed as:

$$mAP = \frac{1}{N}\frac{1}{10}\sum_{c=1}^{N}\sum_{IoU=0.5}^{0.95} AP_c^{IOU} \tag{9}$$

where N is the number of ship categories, AP_c^{IOU} is the AP for a given IoU threshold for a specific ship category, and mAP is the mean AP of the algorithm at 0.5–0.95 IoU thresholds, reflecting the recognition performance of the algorithm under different configurations. Specifically, the AP for a specific category can be expressed as:

$$AP = \int_0^1 p(r)dr \tag{10}$$

where $p(r)$ is the precision-recall curve.

3.3. Experiment

3.3.1. Ablation Study

In order to verify the best location for adding CBAM, this paper embedded CBAM in the backbone, FPN feature pyramid, and head, respectively, and the three embedding methods were represented by C1, C2, and C3. C1, C2, and C3 were used for recognition of six types of ships, and the results are shown in Table 2.

Table 2. Test results of three different embedding methods.

Algorithm	$AP^{0.5–0.95}$/%						$mAP^{0.5–0.95}$/%	FPS
	OC	FB	CS	BC	GCS	PS		
YOLOv4	58.7	56.9	67.6	60.7	64.7	62.0	61.8	45
C1	66.1	62.7	71.1	66.8	70.7	67.6	67.5	40
C2	64.3	62.1	72.4	67.3	69.4	66.3	66.9	42
C3	61.5	60.8	70.3	66.0	68.9	65.7	65.5	43

As can be seen from the results, the addition of CBAM structure in the backbone, FPN, or head has little effect on the overall recognition speed of the algorithm. The C1 algorithm, which adds CBAM to the backbone, improves the recognition accuracy due to its use of shallow features. Among them, the highest detection accuracy was achieved for ore ships, fishing boats, general cargo ships, and passenger ships. The C2 algorithm, which adds CBAM to the FPN, performed well in recognizing container ships and bulk cargo ships. The C3 algorithm, which adds CBAM to the head, did not achieve ideal results because the high-dimensional semantic information of the head feature map was disrupted by the addition of CBAM, which destroyed the original network representation of high-dimensional semantic information. Overall, the C1 algorithm performed the best with the highest average precision of 67.5, which was 5.7 percentage points higher than the original algorithm without CBAM, and the FPS was only 5 percentage points lower than the original algorithm.

In order to further investigate the differences in target representation resulting from adding the CBAM algorithm in different positions, the gradient-weighted class activation mapping (grad-CAM) method [28] was used to analyze three different attention embedding methods. The images were related to three common challenges in detection: foggy weather, small targets, and occluded ships, as shown in Figure 6. The heat map shows the algorithm's focus region on the image. By comparing the visualized feature maps of the improved

networks, it can be observed that: (1) C1 has stronger recognition ability for small targets. In the scene of recognizing small targets, C1 has a larger interested area for small targets, and a higher response to the main target area, indicating that C1 can recognize smaller targets well. (2) C1 enhances recognition ability in foggy weather. Under the foggy conditions, C1's response to the target region is more concentrated, while C2, C3 and the original network have a large amount of dispersed semantics, which means that the channel attention mechanism used in the backbone network avoids the image noise generated by foggy weather, enabling the algorithm to focus on ship features. (3) C1 has stronger recognition ability for overlapping targets. In the highly overlapping target recognition scene, C2, C3, and the original network have severe feature confusion, while C1 can accurately identify two different ship categories. This indicates that C1 has a higher focus on the target contour around it, and retains a large amount of ship's original spatial semantic information in the low-level feature map of the backbone, which enables the spatial attention mechanism to recognize and filter this information, improving the network's recognition accuracy for different categories of ships.

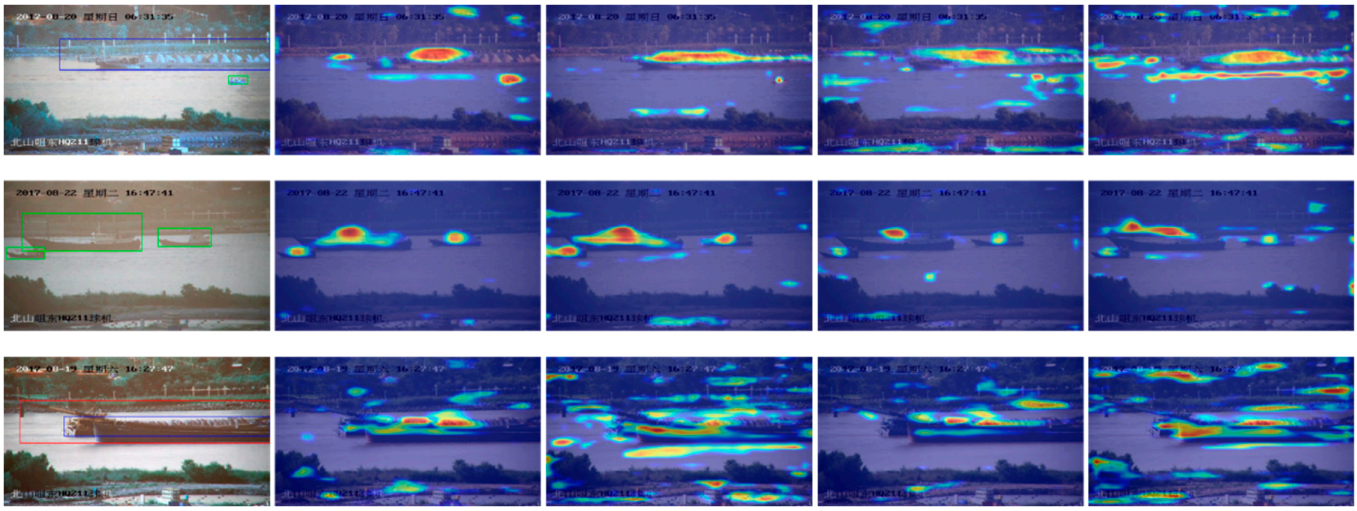


Figure 6. Grad-CAM renderings under three conditions; (**a**) original drawing; (**b**) C1; (**c**) C2; (**d**) C3; (**e**) not embedded.

Therefore, in different situations, the C1 algorithm with CBAM added to the backbone can improve the network's recognition ability to a certain extent. In the subsequent comparative experiments, this paper uses C1 as the basic framework, and tests the performance with the addition of Soft-NMS and the use of the improved EIoU loss function, as shown in Table 3. Comparing the two improved algorithms, the Soft-NMS algorithm (C1-1) improves the overall mAP by 2.5 percentage points, mainly due to the presence of many overlapping targets in the navigational ship images. The use of the EIoU improvement algorithm (C1-2) increases the overall mAP by 3.5 percentage points, indicating that correcting the aspect ratio of the target box separately can effectively determine the ship's position and improve the ship recognition accuracy. Therefore, this paper combines the two (C1-3) to improve the algorithm, and the final mAP reaches 72.5, which is an improvement of 10.7 percentage points compared to the original algorithm.

Table 3. Ablation experiment.

Algorithm	Soft-NMS	EIoU	$AP^{0.5–0.95}$/%	FPS
C1			67.5	40
C1-1	√		70.0	40
C1-2		√	70.7	38
C1-3 (this paper)	√	√	72.5	38

3.3.2. Comparative Experiments

To further demonstrate the superiority of the proposed algorithm, this paper conducted comparative experiments between the deep learning algorithms commonly used in object detection and EA-YOLOv4, including YOLOv3, SSD [29], Retina-Net [30] for one-stage methods, and Faster RCNN, Mask-RCNN [31] for two-stage methods. The experimental results are shown in Table 4. It shows that the lowest mAP came from YOLOv3, which is only 57.3. Faster-RCNN has an AP of 70.2% for container ships, but only 55.7% for fishing boats. The highest FPS came from SSD, which is 78, and the corresponding AP for passenger ships is 70.4%, which is only 1.2% lower than that of proposed algorithm, but its detection performance for other ships is unacceptable. Mask-RCNN has a relatively average detection performance for six types of ships. The overall performance of the proposed algorithm is better, and the detection performance for container ships, general cargo ships, and passenger ships is more prominent, with AP values of 75.7, 74.8, and 72.6, respectively. At the same time, the FPS of the proposed algorithm is 38, which can meet the real-time detection requirements for maritime ship targets.

Table 4. Comparison results between EA-YOLOv4 and other mainstream algorithms.

Algorithm	$AP^{0.5–0.95}$/%						$mAP^{0.5–0.95}$/%	FPS
	OC	FB	CS	BC	GCS	PS		
YOLOv3	52.7	57.1	59.6	51.8	60	62.3	57.3	37
SSD	63.5	60.7	68.5	61.4	61.8	70.4	64.4	78
RetinaNet	63.4	63.5	70.1	66.5	64.9	67.3	66.0	11
Faster-RCNN	64.1	55.7	70.2	57.8	63.7	61	62.1	5
Mask-RCNN	64.5	60.8	69.3	67.0	68.8	67.5	66.3	7
EA-YOLOv4 (this paper)	71.7	68.1	75.7	71.9	74.8	72.6	72.5	38

The detection results of different algorithms under various conditions are shown in Figure 7. It can be found that the scale of the ship in Figure 7a is small. YOLOv3, Faster-RCNN, and SSD cannot capture the features of small fishing boats well. EA-YOLOv4 improves the loss algorithm by using EIoU, which allows it to pay attention to the structural characteristics of small targets, thus improving the recognition ability for small targets. In Figure 7b, there are overlapping targets, namely a bulk carrier and a cargo ship. YOLOv3, Faster-RCNN, SSD, and Mask-RCNN all have prediction errors due to the unique shapes of bulk carriers and cargo ships. The CBAM mechanism embedded in EA-YOLOv4 has good feature representation capability for ship spatial features, and the introduced Soft-NMS mechanism prevents overlapping targets from being incorrectly removed. Therefore, EA-YOLOv4 has more accurate localization of overlapping ships. In Figure 7c, the scene has low brightness. YOLOv3, SSD, and Retina-Net in the one-stage methods cannot extract ship targets well. However, the CBAM mechanism of EA-YOLOv4 retains the high-value channels that contain ship information, enabling it to accurately search for ship targets in low brightness environments, achieving ship detection under low light conditions.

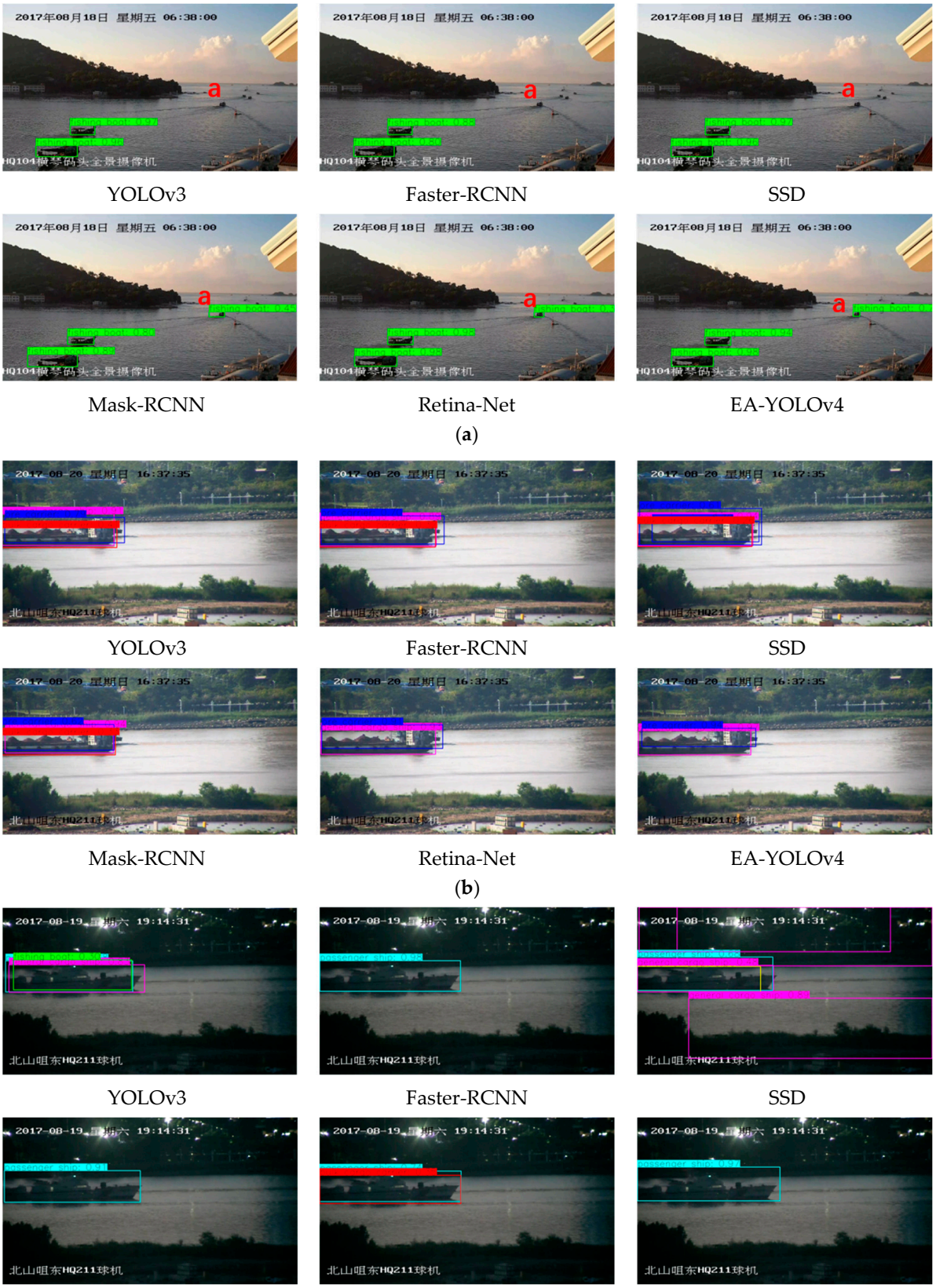


Figure 7. Ship detection results of different algorithms; (**a**) small size target scene; (**b**) multi-target overlapping scene (ore ship and general cargo ship); (**c**) underlit scene.

4. Conclusions

To improve the real-time perception capability of unmanned ships in close-range environments and further enhance the positioning accuracy and recognition accuracy of ship detection algorithms, this paper proposes an EA-YOLOv4 algorithm based on YOLOv4. By using the multidimensional attention mechanism CBAM to improve the algorithm network framework, the extraction of shape features and suppression of background interference are achieved. The improved EIoU loss function is used to enhance the algorithm's perception ability for ships of different scales, accelerate network convergence, and enhance the algorithm's detection ability for small targets. Soft-NMS is used for post-processing of detection results to find missed targets from overlapping ships. The experiment evaluates the influence of all improvements on EA-YOLOv4 in terms of the MSCOCO $AP^{0.5–0.95}$ index and frame rate, and compares it with other similar algorithms. The results show that EA-YOLOv4 performs best among all algorithms. Without significantly affecting the detection speed, EA-YOLOv4 has improved the accuracy of the identification of six types of ships in the real-time monitoring of the nearshore, and can be used as the technical basis for unmanned ship detection and environmental identification. At present, due to the limitation of data sets, this study is limited to ship detection in some severe weather. In the future, we will collect optical ship images under more conditions, increase the types of sea targets that can be detected, and continue to optimize the model, while ensuring the detection speed of the algorithm can realize the full target perception of the unmanned ship navigation environment.

Author Contributions: Conceptualization, Y.Y., R.Z. and J.P.; data curation, Y.Y. and R.Z.; investigation, R.Z., Y.Y. and Y.L.; writing—original draft, Y.Y., R.Z. and J.P.; methodology, R.Z., Y.Y. and Z.S.; writing—review and editing, Y.Y., R.Z. and J.P.; supervision, R.Z. and Z.S.; visualization, Y.Y. and Y.L. funding acquisition, R.Z. and Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: The research described in this paper is supported by the National Natural Science Foundation of China (No. 52001134) and Fuzhou-Xiamen-Quanzhou Independent Innovation Region Cooperated Special Foundation (No: 3502ZCQXT2021007).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Yan, X.P.; Wang, S.W.; Ma, F. Review and prospect for intelligent cargo ships. *Chin. J. Ship Res.* **2021**, *16*, 1–6.
2. Xiao, G.N.; Cui, W.Y. Evolutionary game between government and shipping companies based on shipping cycle and carbon quota. *Front. Mar. Sci.* **2023**, *10*, 1132174. [CrossRef]
3. Chen, X.Q.; Wu, S.B.; Shi, C.J.; Huang, Y.G.; Yang, Y.S.; Ke, R.M.; Zhao, J.S. Sensing Data Supported Traffic Flow Prediction via Denoising Schemes and ANN: A Comparison. *IEEE Sens. J.* **2020**, *20*, 14317–14328. [CrossRef]
4. Liu, Z.; Zhang, B.Y.; Zhang, M.Y.; Wang, H.L.; Fu, X.J. A quantitative method for the analysis of ship collision risk using AIS data. *Ocean Eng.* **2023**, *272*, 113906. [CrossRef]
5. Chen, X.Q.; Liu, S.H.; Liu, R.W.; Wu, H.F.; Han, B.; Zhao, J.S. Quantifying Arctic oil spilling event risk by integrating an analytic network process and a fuzzy comprehensive evaluation model. *Ocean Coast. Manag.* **2022**, *228*, 106326. [CrossRef]
6. Zhen, R.; Shi, Z.Q.; Liu, J.L.; Shao, Z.P. A novel arena-based regional collision risk assessment method of multi-ship encounter situation in complex waters. *Ocean Eng.* **2022**, *246*, 110531. [CrossRef]
7. Zhen, R.; Shi, Z.Q.; Shao, Z.P.; Liu, J.L. A novel regional collision risk assessment method considering aggregation density under multi-ship encounter situations. *J. Navig.* **2022**, *75*, 76–94. [CrossRef]
8. Liu, D.; Zhang, Y.; Zhao, Y.; Shi, Z.G.; Zhang, J.H.; Zhang, Y. Multi-Scale Inshore Ship Detection Based on Feature Re-Focusing Network. *Acta Opt. Sinica.* **2021**, *41*, 137–149.
9. Shao, Z.F.; Wang, L.G.; Wang, Z.Y.; Du, W.; Wu, W.J. Saliency-aware convolution neural network for ship detection in sur-veillance video. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 781–794. [CrossRef]

10. Zhen, R.; Ye, Y.; Chen, X.; Xu, L. A Novel Intelligent Detection Algorithm of Aids to Navigation Based on Improved YOLOv4. *J. Mar. Sci. Eng.* **2023**, *11*, 452. [CrossRef]
11. LA, T. YOLOv4-5D: An enhancement of YOLOv4 for autonomous driving. *Towards Data Sci.* **2021**. Available online: https://towardsdatascience.com/yolov4-5d-an-enhancement-of-yolov4-for-autonomous-driving-2827a566be4a (accessed on 3 February 2023).
12. Liu, Q.W.; Xiang, X.Q.; Yang, Z.; Hu, Y.; Hong, Y.M. Arbitrary Direction Ship Detection in Remote-Sensing Images Based on Multitask Learning and Multiregion Feature Fusion. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1553–1564. [CrossRef]
13. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 11–18 December 2015; pp. 1440–1448.
14. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 27–30 June 2016; pp. 770–778.
15. Yu, M.; Han, S.J.; Wang, T.F.; Wang, H.Y. An approach to accurate ship image recognition in a complex maritime transportation environment. *J. Marine Sci. Eng.* **2022**, *10*, 1903. [CrossRef]
16. Liu, R.W.; Yuan, W.; Chen, X.; Lu, Y. An enhanced CNN-enabled learning method for promoting ship detection in maritime surveillance system. *Ocean Eng.* **2021**, *235*, 109435. [CrossRef]
17. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804. 02767.
18. Hong, X.; Cui, B.; Chen, W.; Rao, Y.; Chen, Y. Research on Multi-Ship Target Detection and Tracking Method Based on Camera in Complex Scenes. *J. Mar. Sci. Eng.* **2022**, *10*, 978. [CrossRef]
19. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
20. Woo, S.H.; 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; Volume 11211, pp. 3–19.
21. Zheng, Z.H.; Wang, P.; Liu, W.; Li, J.Z.; Ye, R.G.; Ren, D.W. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12993–13000.
22. Zhang, Y.-F.; Ren, W.Q.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T.N. Focal and Efficient IOU Loss for Accurate Bounding Box Regression. *Neurocomputing* **2022**, *506*, 146–157. [CrossRef]
23. Bodia, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS-improving object detection with one line of Code. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 5562–5570.
24. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks (CVPR). In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
25. Jaderberg, M.; Simonyan, K.; Zisserman, A.; Kavukcuoglu, K. Spatial Transformer Networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; Volume 28.
26. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollar, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; Volume 8693, pp. 740–755.
27. Shao, Z.; Wu, W.; Wang, Z.; Du, W.; Li, C. SeaShips: A Large-Scale Precisely Annotated Dataset for Ship Detection. *IEEE Trans. Multimedia* **2018**, *20*, 2593–2604. [CrossRef]
28. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2020**, *128*, 336–359. [CrossRef]
29. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Lecture Notes in Computer Science: Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer: Cham, Switzerland, 2016; Volume 9905, pp. 21–37.
30. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.M.; Dollar, P. Focal Loss for Dense Object Detection. In Proceedings of the International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2999–3007.
31. He, K.M.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397. [CrossRef] [PubMed]