Journal of *Marine Science and Engineering*

MDPI

Article

An Approach to Accurate Ship Image Recognition in a Complex Maritime Transportation Environment

Meng Yu, Shaojie Han, Tengfei Wang and Haiyan Wang *

School of Transportation and Logistics Engineering, Wuhan University of Technology, Wuhan 430063, China
* Correspondence: hywang777@whut.edu.cn

Abstract: In order to monitor traffic in congested waters, permanent video stations are now commonly used on interior riverbank bases. It is frequently challenging to identify ships properly and effectively in such images because of the intricate backdrop scenery and overlap between ships brought on by the fixed camera location. This work proposes Ship R-CNN(SR-CNN), a Faster R-CNN-based ship target identification algorithm with improved feature fusion and non-maximum suppression (NMS). The SR-CNN approach can produce more accurate target prediction frames for prediction frames with distance intersection over union (DIOU) larger than a specific threshold in the same class weighted by confidence scores, which can enhance the model's detection ability in ship-dense conditions. The SR-CNN approach in NMS replaces the intersection over union (IOU) filtering criterion, which solely takes into account the overlap of prediction frames, while DIOU, also takes into account the centroid distance. The screening procedure in NMS, which is based on a greedy method, is then improved by the SR-CNN technique by including a confidence decay function. In order to generate more precise target prediction frames and enhance the model's detection performance in ship-dense scenarios, the proposed SR-CNN technique weights prediction frames in the same class with DIOU greater than a predetermined threshold by the confidence score. Additionally, the SR-CNN methodology uses two feature weighting methods based on the channel domain attention mechanism and regularized weights to provide a more appropriate feature fusion for the issue of a difficult ship from background differentiation in busy waters. By gathering images of ship monitoring, a ship dataset is created to conduct comparative testing. The experimental results demonstrate that, when compared to the three traditional two-stage target detection algorithms Faster R-CNN, Cascade R-CNN, and Libra R-CNN, this paper's algorithm Ship R-CNN can effectively identify ship targets in the complex background of far-shore scenes where the distinction between the complex background and the ship targets is low. The suggested approach can enhance detection and decrease misses for small ship targets where it is challenging to distinguish between ship targets and complex background objects in a far-shore setting.

Keywords: maritime management; ship monitoring; video image recognition; convolutional neural network; feature fusion

Citation: Yu, M.; Han, S.; Wang, T.; Wang, H. An Approach to Accurate Ship Image Recognition in a Complex Maritime Transportation Environment. *J. Mar. Sci. Eng.* **2022**, *10*, 1903. https://doi.org/10.3390/jmse10121903

Academic Editors: Claudio Ferrari, Nam Kyu Park and Kevin X Li

Received: 12 October 2022
Accepted: 25 November 2022
Published: 5 December 2022

1. Introduction

Due to the increase in inland waterway traffic, the probability of the occurrence of ship traffic accidents is rapidly increasing [1]. It is necessary for the maritime departments to strengthen the supervision and law enforcement of inland waterway transportation to ensure the safety of ship navigation through precise recognition and target location of passing ships in inland waterways. Inland waterway ship detection methods have been proposed by researchers employing a variety of technologies throughout the last few decades. At present, there are many kinds of ship detection methods, including automatic radar plotting aid (ARPA) [2,3], laser measurement [4], automatic identification system (AIS) [5–8], infrared imaging [9,10], video monitoring [11,12], light detection and ranging (LiDAR) [13], as well as some new ship monitoring technologies [14,15]. Each kind of

detection method has its advantages and disadvantages. Maritime radar can operate in all weather conditions, has a broad detection range, and can take active measurements, but it is readily hampered by impediments like buildings and has poor ability to block interference. Although the laser measurement method has a wide detection range and good range accuracy, it is expensive and requires a specific installation location. AIS can obtain the position and speed of a ship accurately and reliably, but it requires that the ship be properly installed and turned on with AIS equipment. Some small vessels are not installed. In addition, AIS signals have packet loss problems, and a lot of work must be done in data collation. Although infrared imaging is a method that can be used day or night and has great anti-interference capabilities, its low resolution has an impact on detection accuracy and makes it challenging to pinpoint the ship. Computer vision is mostly used in video surveillance to detect ships. It offers the benefits of flexible deployment, economy, and automatic target identification in challenging settings. Currently, effective monitoring of inland vessels employing video surveillance systems has emerged as a development trend due to the quick development of computer vision technology. It performs ship detection via visible light images, which might usefully fill the gap left by the lack of other monitoring techniques.

At present, the research on ship detection based on visible light image is mainly based on the following platforms: (1) Ship detection based on space-based visible remote sensing image [16], which mainly relies on satellite and aircraft platforms to obtain images and then detect ships. It has the advantage of a large detection range, but it is easy to cause small targets to be missed and cannot be monitored in real time. (2) Ship detection based on the monitoring image of a mobile water platform [17,18]. It mainly refers to the real-time monitoring of ships in the surrounding environment by relying on the water mobile platform to achieve autonomous navigation, which requires high real-time monitoring and model deployment to embedded devices. (3) Ship detection based on static shore platform monitoring image [19–22], which is usually used in ports and coasts. The background of the monitoring image remains basically unchanged, and it is used to monitor the ship's state in a specific area. The model can be deployed on desktop computers or cloud computers.

The following challenges with ship identification frequently arise from a perspective in permanent camera stations on interior riverbank bases to monitor traffic in congested waterways: (1) Background of the seaside region—buildings along the riverbank and on the shoreline are readily a threat to ship targets. (2) Significant variation in ship scale—inland rivers are home to enormous commercial ships, little fishing boats, and bamboo rafts. (3) Ships crossing paths—during times of high traffic, inland rivers are prone to dense ship arrangements that cause reciprocal occlusion of ships. Additionally, changes in the environment, such as those caused by light, rain, fog, etc., might impact how well ships can be identified and located. Therefore, accurately identifying ship targets in inland waterways is a difficult task.

This paper mainly focuses on the ship detection of inland river static shore platform based on monitoring images. Considering that the background is usually complex and the ships in the monitoring area are prone to mutual occlusion, this paper proposes a ship target detection algorithm SR-CNN with improved feature fusion and non-maximal suppression based on Faster R-CNN.

The main contributions of this paper are as follows:

(1) This paper proposes a Soft-DIOU-NMS with weighted box fusion to improve the accuracy and recall of ship detection.
(2) This paper proposes a hybrid weighted feature fusion method, which can reduce the miss detection and error detection of ships in complex background environment and improve the localization accuracy of ships.

The rest of this paper is organized as follows. Section 2 reviews previous literature on ship detection. Section 3 proposes a ship target detection algorithm based on improved feature fusion and non-maximal suppression. Section 4 introduces the details of the experiment preparation, the experimental results, and a comparative analysis. Section 5

analyzes the shortcomings of the current study and the next research directions. Section 6 proposes the conclusion of this paper.

2. Related Works

The current research on ship target detection based on visible light image is divided into detection methods based on traditional image processing technology and detection methods based on deep learning.

The traditional approach treats ship target detection as a classification problem. The main steps are region of interest segmentation, feature extraction, and classifier classification. Traditional ship detection methods include two main categories outlined below. (1) Ship detection based on sky-sea line (SSL): Zhang [23] et al. established a marine region model by detecting sky-sea line based on discrete cosine transform and then they used the background subtraction method and foreground segmentation method to detect the ship's target. Chen [24] et al. extracted the sky-sea line by using the maximum between-class variance (OTSU) algorithm and Hough transform and the ship's position was detected by the peak gray value. Shan et al. [25] proposed a maritime target detection algorithm that mainly includes three stages of SSL estimation, SSL detection, and target saliency detection. (2) Detection based on ship structure characteristics: Liu et al. [26] detected the light points in the video image by Laplacian of Gaussian (LoG) and filtered the invalid points using grayscale thresholding to achieve ship detection. Xin [27] proposed the use of OTSU for image segmentation to determine the target area, match scale invariant feature transform (SIFT) feature points to form a matching vector, and to analyze the matching vector in the frequency domain to determine the number and location of ships. However, the recognition accuracy of this method is low. Shafer et al. [28] combined anomaly detection and dictionary learning methods to develop a sparsely driven anomaly detector to detect and track ship targets in video surveillance. However, traditional ship detection methods tend to rely too much on the extraction of manual features and have poor generalization ability. Therefore, the recognition rate and detection accuracy of these methods cannot meet the practical requirements in complex environmental backgrounds such as extreme weather environments (e.g., rain, haze, strong and weak light).

The dependence on whether to rely on anchor, target detection algorithms based on deep learning can be divided into two categories: Anchor-Base and Anchor-Free. The target detection algorithm of Anchor-Base can be divided into a single-stage algorithm and a two-stage algorithm according to whether the candidate box is generated or not. The two-stage algorithm first generates a series of candidate boxes and then the filter high-quality candidate box is used for target classification and coordinate regression. It is characterized by higher detection accuracy, but real-time performance needs to be improved. Classical two-stage algorithms have R-CNN [29], Faster R-CNN [30], and Mask R-CNN [31]. The single-stage algorithm takes the equidistant sampling points on the detection graph as the center of the anchor box, and directly uses the anchor box and the real box to achieve the classification and coordinate regression of the target. Such methods are characterized by faster detection but slightly reduced accuracy, among which the representative algorithms have YOLO [32], SSD [33], etc. The Anchor-Free target detection algorithm includes a corner point based detection method and a key point based detection method. The corner point based algorithm identifies the target bounding box as a pair of key points and achieves target classification and localization by detecting corner point features, among which the representative algorithm is CornerNet [34]. The key point based algorithm predicts the probability that each location in the feature map belongs to the centroid and determines the target edges using the feature heat map, the classic of which is CenterNet [35]. Deep learning has now achieved notable success in target detection, and more and more researchers are introducing deep learning into ship detection [36]. Compared with traditional ship detection methods, deep learning-based detection methods greatly improve the effectiveness of ship detection. Shao et al. [37] developed a saliency-aware convolutional neural network-based model for extracting

discriminative ship features in coastal marine images. Kim [38] et al. proposed a new probabilistic ship detection and classification method to determine the class of ships by considering the confidence of the deep learning detector as probabilities and combining the probabilities of consecutive images over time through Bayesian fusion. In order to improve the detection capability of small target ships, Chen et al. [17] proposed a hybrid deep learning approach combining generative adversarial networks (GAN) and convolutional neural networks (CNN). Sun et al. [20] proposed a ship detector NSD-SSD based on a multiscale feature fusion (MFF), prediction module (PM) and a priori box reconstruction (RPB) visual images. For ship identification under different lighting and weather conditions, Liu et al. [21] achieved more reliable and robust ship target detection under adverse weather conditions by redesigning the anchor box size, predicting the localization uncertainty of the enclosing box, introducing soft non-maximal suppression, and reconfiguring the hybrid loss function. Based on Yolov3, Chang et al. [22] combined visible and infrared images, then selected a suitable input image size and detection scale, fewer convolution filters, and spatial pyramid pooling (SPP) to achieve effective detection of ships during day and night. Chen et al. [39] proposed an end-to-end fully convolutional Anchor-Free network by using key points to generate a ship envelope box and introducing feature fusion modules and feature enhancement modules. This method has better detection robustness for rain and fog occlusion, scale change, and neighboring ship interference. In addition, some researchers have combined traditional ship detection methods with detection methods based on deep learning. For example, Li et al. [18] combined traditional image processing methods with deep learning methods to propose a multi-level hybrid network-based ship detection and identification method, but this method can only detect ships near the sky-sea line. Qi et al. [40] combined image downscaling, scene semantic reduction and topic reduction with Faster R-CNN to speed up the model detection. However, the above studies are more about improving the inference speed of the model by sacrificing accuracy.

For stationary static monitoring systems, such as ports, where ships move slowly, there is no need to strictly limit the size and power consumption of the computing devices. The models can be deployed on desktop computers. Few studies have been conducted on this aspect. Therefore, based on the review and analysis of ship detection related work, this paper studies ship detection based on complex coastal scenarios to improve the detection and localization accuracy of ships while reducing ship misdetection and missed detection target.

3. Methodologies

3.1. Network Architecture

Faster region-based convolutional neural network (Faster R-CNN) is a two-stage detector. It consists of region proposal network (RPN) and fast region-based convolutional neural network (Fast R-CNN), which share the feature convolution network. RPN generates regions that may contain objects; Fast R-CNN divides these regions into objects or backgrounds and refines the boundaries of the regions. In this paper the benchmark model is Faster R-CNN with feature pyramid networks [41] (FPN). First, by analyzing the problems of the NMS screening mechanism and screening criteria, we propose a weighted box fusion Soft-DIOU-NMS to improve it. Second, the features are directly added in the feature fusion of the benchmark model, without considering the influence of different fusion methods on the model performance. We explored this issue and proposed a hybrid weighted feature fusion method by comparing different fusion methods. Finally, the improved two methods are applied to the benchmark model, and the optimized detection framework is called ship region-based convolutional neural network (SR-CNN), whose structure is shown in Figure 1.

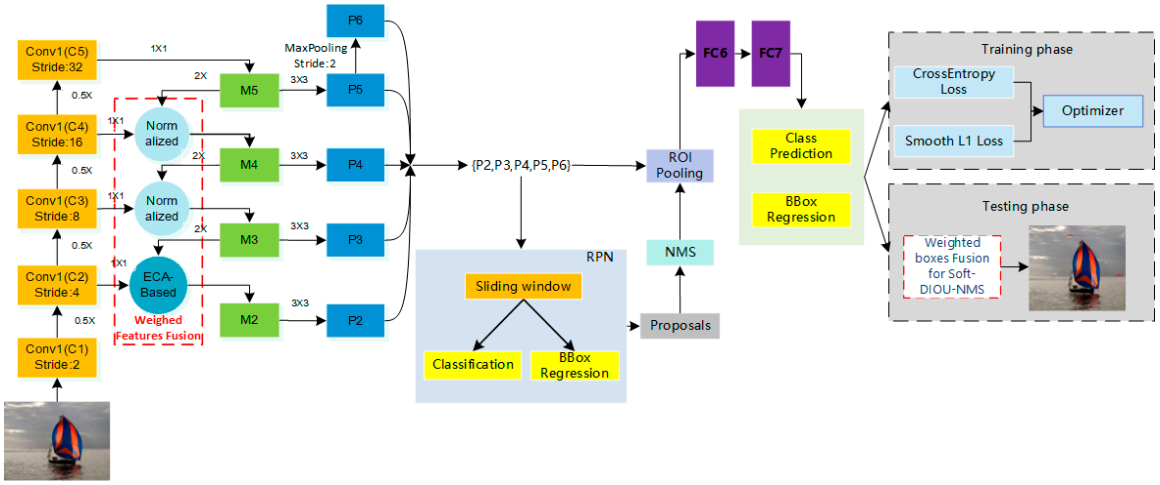


Figure 1. Architecture of SR-CNN for ship detection.

3.2. Soft-DIOU-NMS with Weighted Box Fusion

The model generates a large number of candidate boxes during ship detection. These boxes may overlap, so we need to filter them. Non-maximum suppression (NMS) [42] is the classical method for removing redundant candidate boxes. Its confidence reset formula is shown in Equation (1). However, it has the following shortcomings: (1) Too strict screening mechanisms may lead to missed detection target; (2) the screening criteria only consider the overlap area of the two boxes, ignoring the distance between them; (3) when determining the final result, only the prediction box with the maximum confidence of classification is considered, but the localization accuracy of the prediction box with the maximum confidence is not necessarily optimal.

$$s_i = \begin{cases} 0, IOU(b_m, b_i) \geq N_t \\ s_i, IOU(b_m, b_i) < N_t \end{cases} \tag{1}$$

where b_i is the i-th candidate box, s_i is the classification confidence corresponding to b_i, b_m is the candidate box with the highest confidence, and N_t is the non-maximum suppression threshold.

In order to solve the above problems, we propose a Soft-DIOU-NMS based on weighted frame fusion. On the one hand, inspired by Soft-NMS [43], a penalty function f(x) is introduced, and the DIOU [44] of the candidate box bm with the largest confidence and the prediction box bi is used as the input of f(x). Then, f(x) is multiplied by the confidence score s_i as the final confidence score, and the calculation process is shown in Equation (2).

$$s_i = \begin{cases} s_i f(DIOU(b_m, b_i)), DIOU(b_m, b_i) \geq 0 \\ s_i \qquad\qquad\qquad\qquad , DIOU(b_m, b_i) < 0 \end{cases} \tag{2}$$

DIOU adds a measure of the distance between two boxes based on IOU. It calculates the overlap area and the distance between the center points of the two boxes, which can more comprehensively describe the position relationship between the boxes. The schematic diagram is shown in Figure 2 and the calculation formula is shown in Equation (3).

$$DIOU = \frac{A \cap B}{A \cup B} - \left(\frac{d^2}{c^2}\right)^{\beta} \tag{3}$$

where A, B represent two ship candidate boxes, d represents the distance between the center points of two candidate boxes, c represents the diagonal distance of the smallest box C containing both ship candidates, and β is the center point distance penalty magnitude. After pre-experiment we found that the model performs best when the value of β is 0.6.

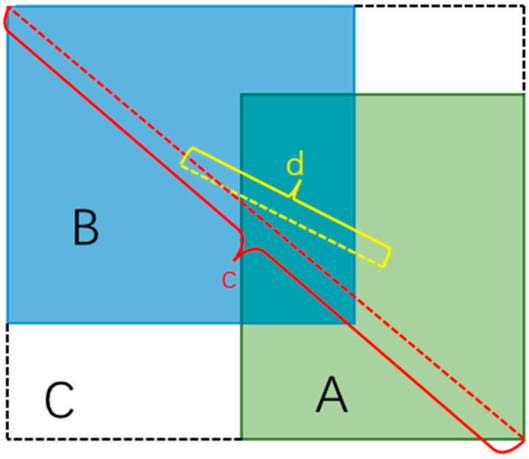


Figure 2. Schematic diagram of DIOU (A and B represent two ship candidate boxes, C is the smallest box containing A and B, c represents the diagonal distance of C, d represents the distance between the center points of two candidate boxes).

The penalty function f(x) is shown in Equation (4).

$$f(DIOU(b_m, b_i)) = exp(-\frac{DIOU(b_m, b_i)^2}{\sigma}) \quad (4)$$

where the parameter σ is taken as 0.5 according to Section 6.2 of Ref. [43].

On the other hand, when determining the final candidate box, we think that the classification confidence of the candidate box is not strongly correlated with the localization accuracy. Sometimes the low-confidence candidate box may have higher localization accuracy than the high-confidence candidate box. For example, in Figure 3, although the white box can detect well the main part of the ship, the edge part of the ship is not predicted well. However, the red weighted fusion box can detect well all components of the ship. Therefore, we take the weighted fusion box as the final candidate box. The process of weighted fusion box is as follows: First, the DIOU of each candidate box and the candidate box with the largest confidence are calculated. Then candidate boxes with DIOU greater than threshold α are selected. Finally, the selected candidate box and the candidate box with the largest confidence are weighted according to the confidence to determine the final candidate box.

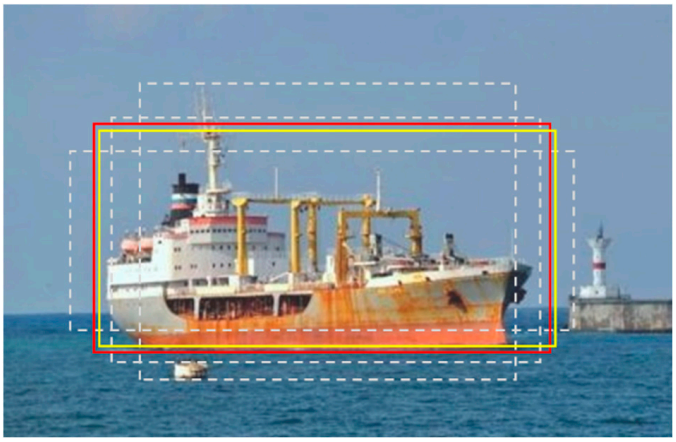


Figure 3. Schematic diagram of the weighted fusion boxes (Yellow box is the real target box, white box for prediction box, and red box is weighted fusion box).

The formula for the weighted fusion candidate box is shown in Equation (5).

$$M' = \frac{\sum_{i=1}^{n} s_i \cdot b_i + s_m \cdot b_m}{s_m + \sum_{i=1}^{n} s_i} \tag{5}$$

where M' is the weighted fusion candidate box, b_i is the i-th candidate box, s_i is the classification confidence corresponding to b_i, b_m is the maximum confidence candidate box, s_m is the classification confidence corresponding to b_m, and n is the number of candidate boxes finally selected.

The proposed Soft-DIOU-NMS with weighted fusion box algorithm pseudocode is shown in Algorithm 1.

Algorithm 1 Pseudocode of Soft-DIOU-NMS with weighted fusion box.
Soft-DIOU-NMS with weighted fusion box

Input: $B = \{b_1, \cdots, b_N\}, S = \{s_1, \cdots, s_N\}$, α
B is the list of one class initial detection boxes
S contains corresponding detection scores
α is bounding box location information fusion threshold
Output: D, S
$D \leftarrow \{\ \}$
while B $\neq$ empty. **do**
  $m \leftarrow argmax\ S$
  $M \leftarrow b_m$
  if f$(DIOU(b_m, b_i)) \geq \alpha$. **then**
    $M' = \frac{\sum_{i=1}^{n} s_i \cdot b_i + s_m \cdot b_m}{s_m + \sum_{i=1}^{n} s_i}$
  end if
  $D \leftarrow D \cup M'$; $B \leftarrow B - M$
  for b_i in B **do**
    if f$(DIOU(b_m, b_i)) \geq 0$. **then**
      $s_i \leftarrow s_i f(DIOU(b_m, b_i))$
    end if
  end for
end while

3.3. Hybrid Weighted Feature Fusion

In the feature extraction network, the shallow feature has low semantics due to less convolution, but it contains more image texture information such as location and details. The deep feature contains more global semantic information, but its resolution is low and its perception of details is poor [45]; FPN enhances the semantic information of low-level features by fusing high-level features with low-level features from top to bottom. This improves the ability to detect small targets. However, when feature fusion is carried out by FPN, the features from different feature layers are added directly according to the same weight. Different weights in the fusing features can have different effects on the performance of the model. Therefore, we use an attention mechanism and add additional weights to the feature layers to make the network learn the importance of different input features autonomously.

The first feature fusion method is ECA-based weighted feature fusion. It mainly uses the channel domain-based attention module efficient channel attention (ECA) [46] to learn the weight values of features in each channel. The main idea of ECA is to obtain the average value of each channel feature through the global average pooling, and then learn the weight of each channel through a weighted one-dimensional convolution. The size of the convolution kernel k represents the cross-channel information interaction rate of

the module, which can be adaptively determined by the total number of channels. The calculation formula is shown in Equation (6).

$$k = \varphi(C) = \left| \frac{\log_2 C + b}{\gamma} \right|_{odd} \tag{6}$$

where k is the size of the convolution kernel, C is the total number of channels, and odd is the nearest odd number. In this paper, we set γ and b to 2 and 1 according to Section 3.2.3 of Ref. [46].

The structure of the ECA module is shown in Figure 4, and its operation process is shown in Equation (7).

$$ECA(x) = \sigma(Conv_{1d}^{1x1}(GAP(x))) \times x \tag{7}$$

where σ is the Sigmoid activation function, $Conv_{1d}^{1X1}$ denotes the one-dimensional convolution operation with k kernel, GAP denotes the global average pooling operation, and x is the input tensor.

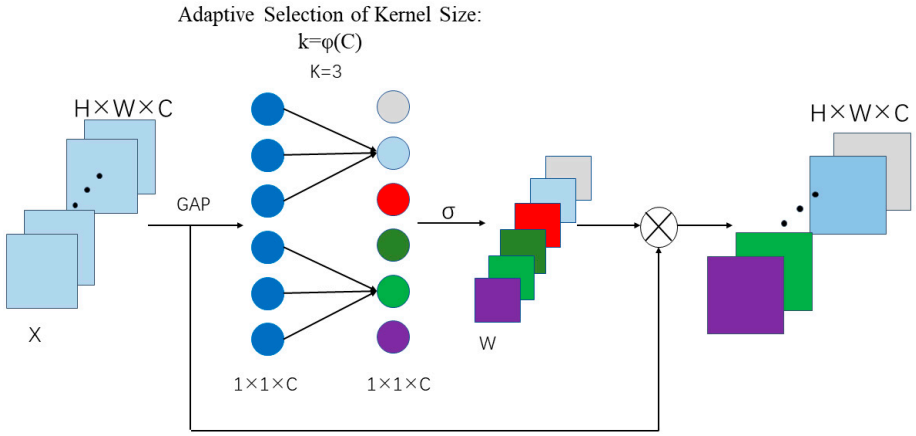


Figure 4. ECA module structure.

After learning the weight value w_i of each channel in the feature layer p_i through the ECA module, the original feature p_i is not multiplied directly by the learned weight w_i. Instead, the weights of different feature layers are weighted and fused to obtain the fused feature layer p'_i. The fusion structure of the feature layer is shown in Figure 5. The calculation process is as follows:

$$w_i = \sigma\left(Conv_{1d}^{1x1}(GAP(p_i))\right) \tag{8}$$

$$w_{i+1} = \sigma\left(Conv_{1d}^{1x1}(GAP(p_{i+1}))\right) \tag{9}$$

$$p'_i = \frac{p_i \cdot w_i + p_{i+1} \cdot w_{i+1}}{w_i + w_{i+1}} \tag{10}$$

where, p_i represents the shallow feature, p_{i+1} represents the deep feature after sampling, w_i represents the weight value of each channel learned by the attention mechanism ECA of the feature layer p_i, w_{i+1} is the same.

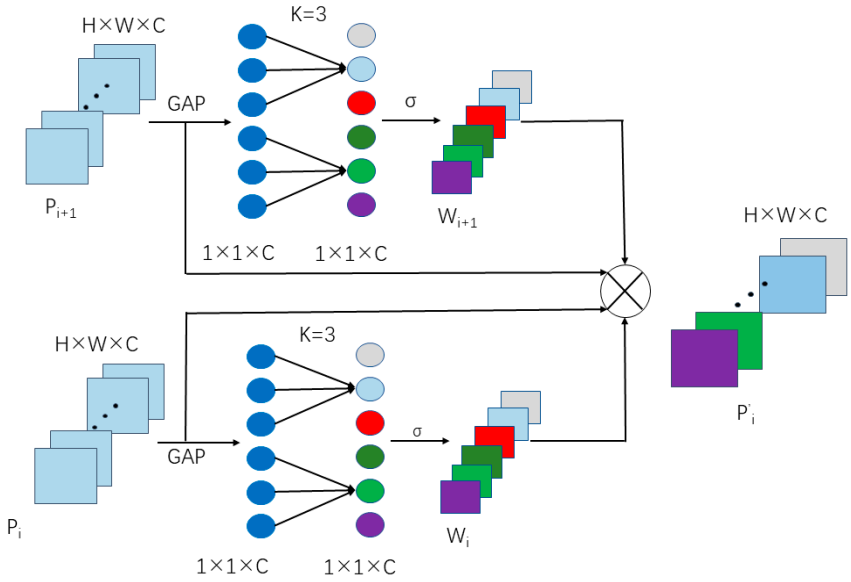


Figure 5. ECA-based weighted feature fusion module.

The second feature fusion method is normalized weight feature fusion. By adding an autonomously learnable weight for different input feature layers, it makes the network learn the importance of each feature layer. To ensure the stability of training and minimal computational cost [47], we use weight normalization to constrain the range of each weight, and the calculation process of each weight is shown in Equation (11):

$$w_{\mathrm{i}}' = \frac{f(\mathrm{w_i})}{\varepsilon + \sum_j^1 f(\mathrm{w_i})} \tag{11}$$

where w'_i denotes the normalized weight value of w_i', $f(w_i)$ denotes that $w_i \geq 0$ is guaranteed by applying the rule activation function to the weights, and ε represents a constant of 0.0001 here to avoid denominator 0.

The normalized weights are multiplied with different feature layers and then summed up. The calculation process is as in Equation (12).

$$p_i' = \sum_j w_{\mathrm{i}}' p_{\mathrm{i}} \tag{12}$$

where p_{i} represents the input feature map and w'_{i} represents the normalized weight value of feature layer p_{i}.

The ECA-based weighted feature fusion module can achieve fine feature fusion, especially for high-resolution feature fusion., but its computational cost is higher. The computational cost of the normalized feature fusion method is lower, but it also cannot realize the feature fusion at the channel level. Therefore, based on FPN, we apply two feature fusion methods to propose four hybrid weighted feature fusion schemes as shown in Figure 6.

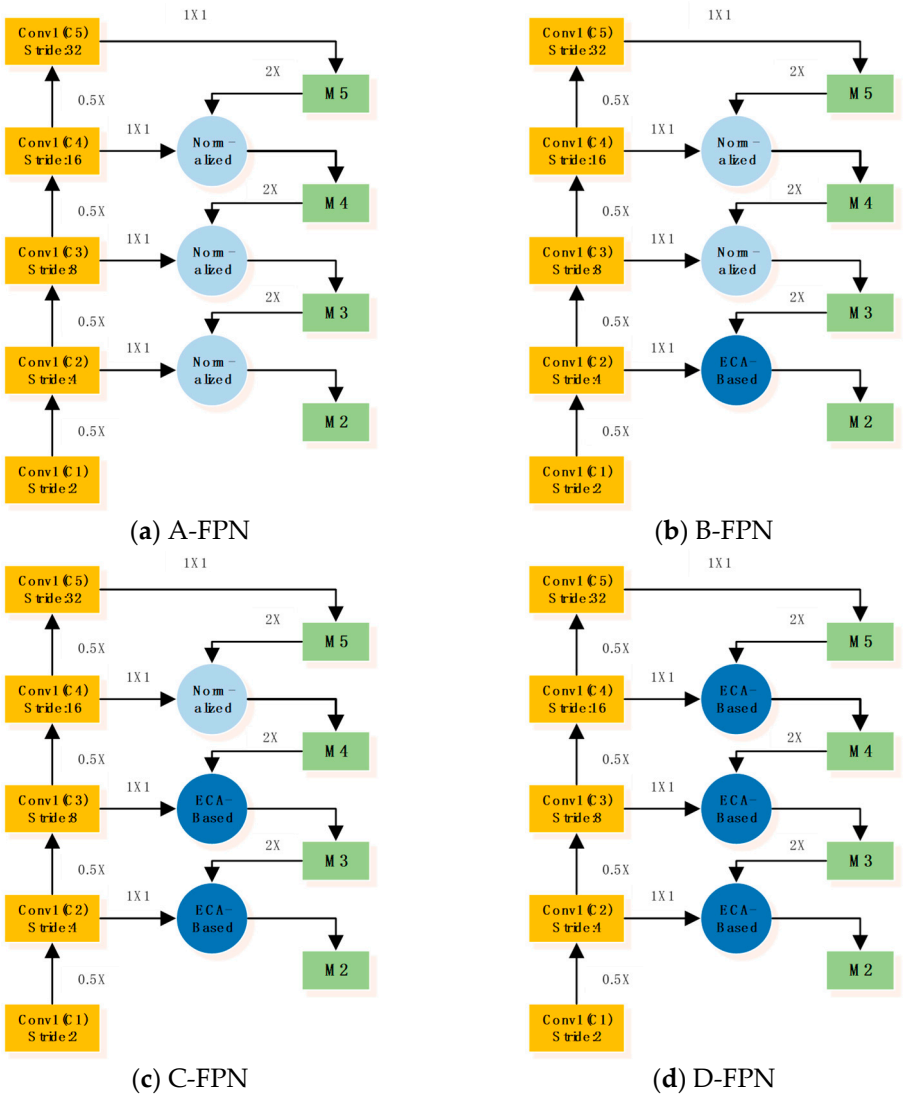


Figure 6. Hybrid weighted feature fusion of FPN.

4. Experiments and Results

4.1. Dataset Description

To validate the effectiveness of the proposed algorithm, the research team collected and labeled 10,000 pictures of ships. It includes six categories: liner, container ship, bulk carrier, island reef, other ships, and sailboat. In order to visualize the size and number distribution of ships in the dataset, we normalized the width and height of all ship targets to draw a visual distribution map, as shown in Figure 7. In this paper, the criteria for classifying target size are small targets with pixel sizes below 32^2, medium targets between 32^2 and 96^2, and large targets above 96^2. The dataset is in COCO standard format, and the ratio of training set to test set is 8:2 for the experiments.

4.2. Evaluation Metrics

All experimental evaluations were calculated using the COCO API and the evaluation metrics include AP, AP_{50}, AP_{75}, AP_S, AP_M, AP_L, $AR_{all\text{-}100}$. Among them, AP refers to the average accuracy, that is, the average precision rate of each category when predicting multiple categories; AP_{50} and AP_{75} mean the average accuracy at IOU thresholds of 0.5 and 0.75 respectively; AP_S, AP_M, AP_L refer to the average accuracy of the target size under different pixel area; $AR_{all\text{-}100}$ refers to the average recall rate of all categories when each

one picture does not exceed 100 test results, that is, the average recall rate of each category when predicting multiple categories.

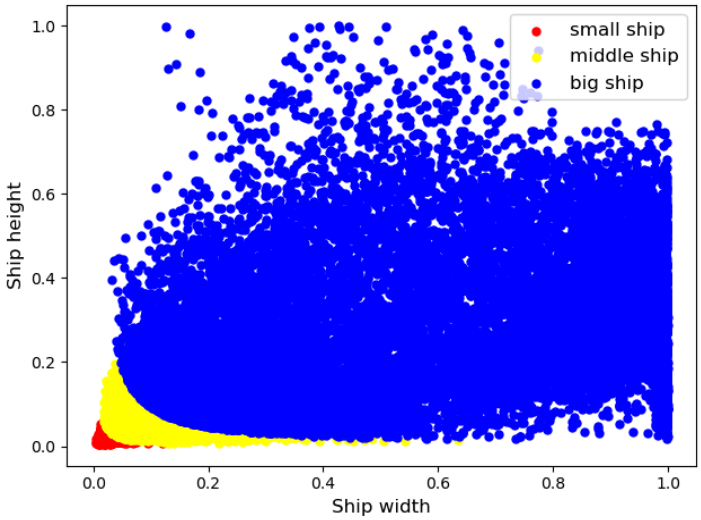


Figure 7. Visualizing the distribution of ship size and number in the dataset.

4.3. Experimental Environment and Parameter Settings

For the experimental platform, the CPU is Intel Xeon Silver 4210R @ 2.40 GHz with 32 GB of RAM and the GPU is GeForce RTX3090 with 24 GB of video RAM. In terms of the software environment, the operating system is a 64-bit Win10 and the deep learning framework is Pytorch1.10.0. In this paper, all experiments were conducted based on the open source object detection framework mmdetection [48]. All models were trained for 24 epochs, the initial learning rate was 0.02, the sixteenth epoch learning rate penaltized to 1/10 of the original, and the rest of the parameters were mmdetection default parameters.

4.4. Experimental Results and Analysis

4.4.1. Soft-DIOU-NMS with Weighted Box Fusion Experiment

The super parameter α was introduced into the Soft-DIOU-NMS with weighted box fusion post-processing algorithm. In order to study the influence of different α values on the Soft-DIOU-NMS with weighted box fusion: first, the original NMS was replaced by soft-NMS and the confidence attenuation function was selected as Gaussian function for the experiment; then the filter criteria from IOU to DIOU was replaced; finally, the influence of different α on the model's detection performance was researched.

As shown in Table 1, first, after replacing the post-processing algorithm NMS with Soft-NMS in Faster-RCNN, the mAP increased by 1.1%. Second, when replacing the original screening criterion IOU with DIOU, the model detection accuracy and recall rate are slightly improved compared with Soft-NMS, which are 0.1% and 0.7%, respectively. This shows that using DIOU as a screening criterion is more reasonable. Finally, when the weighted fusion boxes algorithm is introduced, the detection accuracy of the model is improved under different α values. The maximum value of mAP was 53.1% when α was 0.85, which was 1.0% higher than that of Soft-NMS. This shows that it is not reasonable to only select the prediction box with the highest classification confidence as the final prediction box in the post-processing algorithm. The classification confidence of the prediction box is not strongly correlated with the localization accuracy. It is necessary to consider the prediction box under different confidences to obtain the final target prediction box.

Table 1. The influence of different α on Soft-DIOU-NMS with weighted box fusion.

	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	$AR_{all\text{-}100}$
Baseline	51.0	**78.4**	55.3	6.1	23.0	58.8	60.8
Soft-NMS (Gaussian)	52.1	78.1	57.9	**6.4**	23.6	60.2	67.0
Soft-NMS (DIOU)	52.2	78.0	**58.1**	**6.4**	23.6	60.2	67.7
Candidate box fusion (α = 0.70)	52.5	78.0	57.6	6.1	23.9	60.6	66.4
Candidate box fusion (α = 0.75)	52.8	78.0	58.0	6.2	23.9	60.9	66.9
Candidate box fusion (α = 0.80)	**53.1**	78.0	57.8	6.3	**24.0**	61.3	67.6
Candidate box fusion (α = 0.85)	**53.1**	77.9	57.8	6.3	23.3	**61.4**	67.9
Candidate box fusion (α = 0.90)	52.9	78.0	57.8	6.3	**24.0**	61.1	**68.2**

As shown in Figure 8, when the fusion threshold α gradually increases, the detection accuracy first increases and then decreases, and is higher than that without fusion. However, the detection time decreases with the increase of threshold α. This is because the larger the threshold α, the less weighted are the prediction boxes. The detection accuracy is the same for threshold values α equal to 0.8 and 0.85, but the detection time of the model is shorter for α equal to 0.85. Therefore, the threshold value α of the subsequent experiment is 0.85.

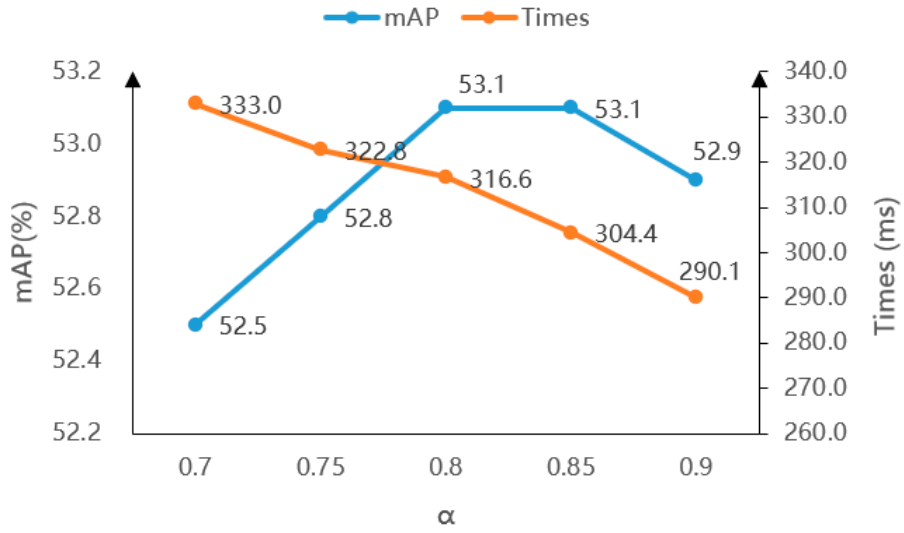


Figure 8. The influence of different α on model's detection accuracy and time.

4.4.2. Hybrid Weighted Feature Fusion Experiment

The ECA-based weighted feature fusion module learns the weights of different feature layers by using the channel-based domain attention module ECA. Then the learned weights are used to weight the features of different channels to achieve feature fusion. Adding additional weights to the feature layers can also autonomously adjust the importance of different feature layers to achieve feature fusion. In order to explore the influence of the two feature fusion methods on the detection ability of the model, on the basis of Faster-RCNN, the four feature fusion schemes proposed in Section 3.3 were verified, and also compared with the classical PAFPN [49] structure.

It can be seen from Table 2 that the four feature fusion schemes proposed, effectively improve the detection ability of the model. Among them, B-FPN has the greatest improvement. Compared with the original FPN, mAP increases by 2.5% and $AR_{all\text{-}100}$ increases by 1.5%. Compared with PAFPN, mAP increases by 0.7%, and $AR_{all\text{-}100}$ increases by 0.5%. This shows that the hybrid weighted feature fusion method proposed can effectively improve the localization accuracy of the model for the target and reduce the target miss detection.

Table 2. Comparison of different FPN structure detection results.

	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	$AR_{all\text{-}100}$
Baseline	51.0	78.4	55.3	6.1	23.0	58.8	60.8
PAFPN	52.8	**79.4**	57.1	6.7	24.7	60.5	61.8
A-FPN	52.0	78.4	55.4	5.2	22.4	60.2	61.2
B-FPN	**53.5**	79.0	**57.3**	**7.1**	**25.2**	**61.4**	**62.3**
C-FPN	52.2	78.3	56.2	5.8	23.6	60.3	61.2
D-FPN	52.5	78.7	56.8	6.2	24.4	60.6	61.7

In order to further analyze the role of each module of B-FPN, we tested the model detection effect when only normalized weight feature fusion is added to the upper two layers and only ECA-based weighted feature fusion is added to the lower layer.

As shown in Table 3, adding normalized weight feature fusion to only the top two layers, and ECA-based weighted feature fusion to only the bottom layer can both improve the localization accuracy and recall of the model. Moreover, only adding ECA-based weighted feature fusion at the lowest level can improve the detection ability of the model better, which further verifies that ECA-based feature fusion can achieve efficient feature fusion for high-resolution feature maps.

Table 3. B-FPN ablation experiments.

	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	$AR_{all\text{-}100}$
Baseline	51.0	78.4	55.3	6.1	23.0	58.8	60.8
Only_ Normalized	52.4	78.6	56.1	7.2	23.2	60.4	61.7
Only_ECA-based	52.7	**79.0**	57.1	**7.3**	23.5	60.6	62.0
B-FPN	**53.5**	**79.0**	**57.3**	7.1	**25.2**	**61.4**	**62.3**

4.4.3. Combination Experiment of Two Methods

In order to further verify the effect of the two methods, the combination of the B-FPN(Method1) and Soft-DIOU-NMS with weighted box fusion (Method2) was used in the benchmark algorithm Faster R-CNN for the experiments.

It can be seen from Table 4 that after the final combination of the two methods, the mAP of the model increases by 3.9% and $AR_{all\text{-}100}$ also increases to a certain extent. Therefore, the improved method proposed in this paper can effectively improve the detection ability of the model and achieve more accurate localization of the ship target.

Table 4. Combination experiment of two improved methods.

Method1	Method2	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	$AR_{all\text{-}100}$
		51.0	78.4	55.3	6.1	23.0	58.8	60.8
√		53.5	**79.0**	57.3	7.1	25.2	61.4	62.3
√	√	**54.9**	78.8	**59.5**	**7.4**	**26.1**	**63.0**	**68.3**

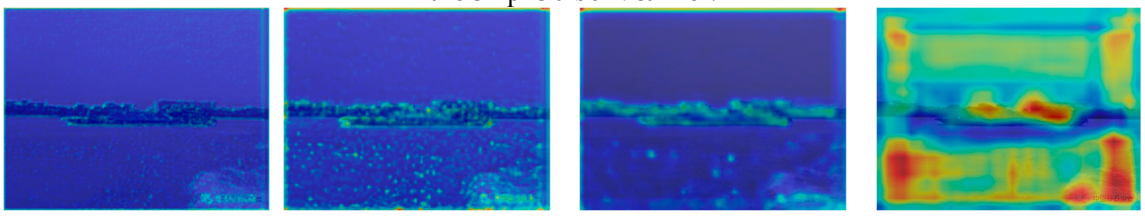
4.4.4. Visual Comparative Analysis

The model can be explained by the generated heat map. We use the process that each position on the feature map will be repeatedly extracted and discarded with the continuous stacking of convolution layers. The feature map finally retains useful feature information. The learned feature information is linearly combined to form an activation feature map. Then, the activation feature map is restored to the same size as the original map by using the up-sampling and the original image is added with the activation feature map. In this way, the proposed hybrid feature fusion method is visually interpreted and analyzed.

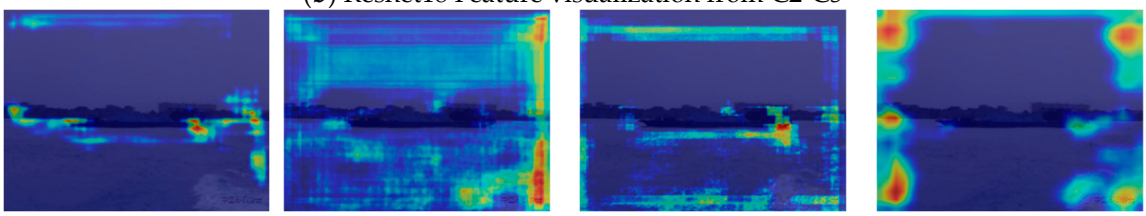
As shown in Figure 9b, the shallow features of the feature extraction network contain more texture position information, while the deep features contain more semantic informa-

tion. Compared with P2–P5 in Figure 9c–e, it can be found that the different feature fusion schemes have a great influence on the focus area of the model. Both PAFPN and FPN can learn effectively the target area. However, B-FPN is more accurate than PAFPN and FPN. To sum up, through visualization results of feature layer, B-FPN can more effectively use multi-scale structure to learn ship characteristics and more accurately detect ship targets.

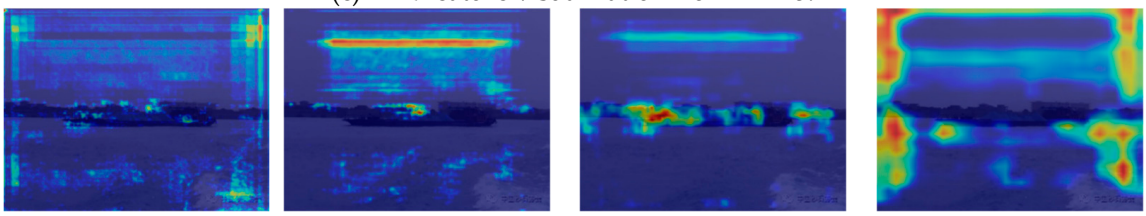
(**a**) Input image. The background of the picture is the building on the other side of the river and the ship is a bulk carrier.



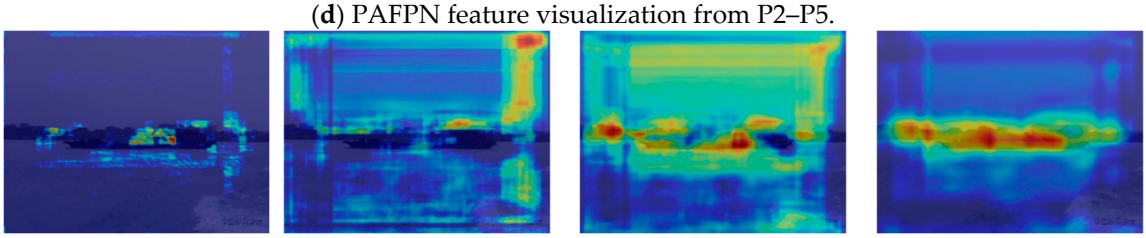
(**b**) Resnet18 Feature visualization from C2-C5



(**c**) FPN feature visualization from P2–P5.



(**d**) PAFPN feature visualization from P2–P5.



(**e**) B-FPN feature visualization from P2–P5.

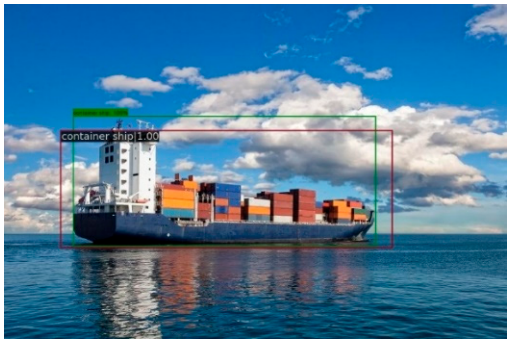
Figure 9. Visualization results of feature layer.

4.4.5. Comparative Analysis of the Detection Results of the Two Methods under Different Backgrounds

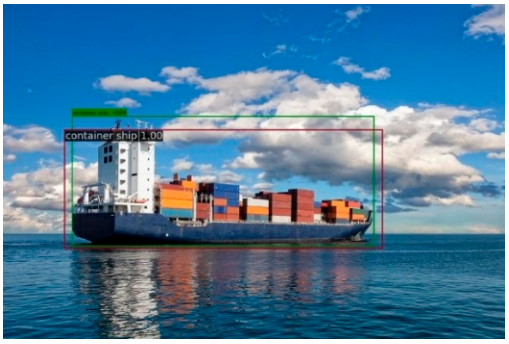
In order to reflect the impact of the two improved methods more intuitively on the performance of the target detection model, single-target images in simple scenes and

multi-target images in complex scenes were selected for testing. The images were first detected using the benchmark algorithm, and then the images were detected using the improved algorithm. In the following figure the green box indicates the real target box, the yellow box indicates the missed ship target, and the other color boxes indicate the predicted boxes of the model. The fonts above the anchor box represent the target category and the classification confidence, respectively.

The results of the detection of the baseline algorithm (a) and the improved algorithm (b) for simple and complex scenes are shown in Figure 10. It can be seen that the prediction box localization is more accurate after the Soft-DIOU-NMS with weighted box fusion processing; At the same time, for dense targets, the original NMS readily causes target miss detection, while using the improved algorithm can avoid target miss detection.

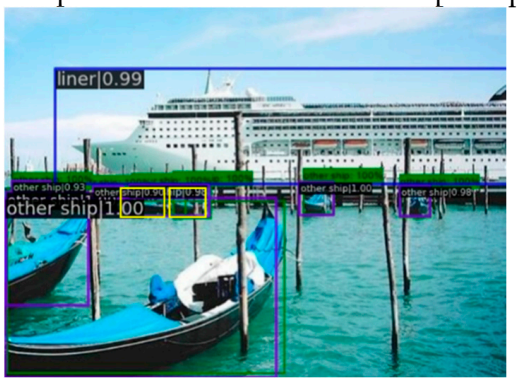


(**a1**) NMS

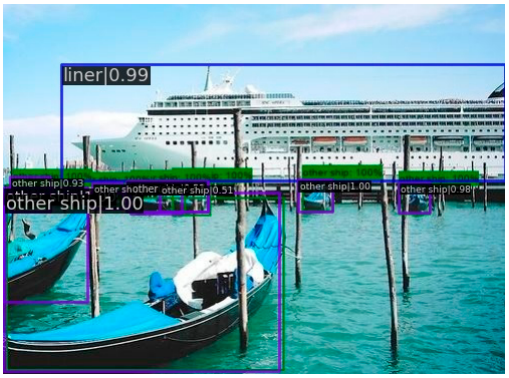


(**b1**) Soft-DIOU-NMS with weighted box fusion

The pictures show a container ship in open water.



(**a2**) NMS



(**b2**) Soft-DIOU-NMS with weighted box fusion

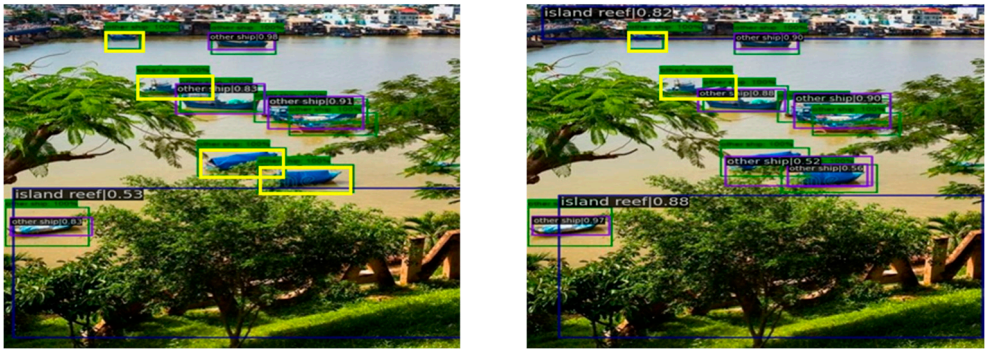
The pictures show a large number of ships in a dock with a large liner in the background.

Figure 10. Comparison of detection results of NMS and weighted box fusion for Soft-DIOU-NMS (Green box is real target box and yellow box is missed ship target).

As shown in Figure 11, in simple scenes, the hybrid weighted feature fusion (B-FPN) can effectively identify the ship target, while FPN cannot distinguish the ship body from the background and wrongly classifies the bulk carrier's head as another ship; in complex scenes, the FPN has missed the detection target, but B-FPN can better learn the ship features to reduce the occurrence of the missed detection target.

(**a1**) FPN (**b1**) B-FPN

The pictures show a large bulk carrier in a port with the container yard in the background.



(**a2**) FPN (**b2**) B-FPN

The pictures show the small ships sailing in the inland river with the buildings and trees on both sides of the river in the background.

Figure 11. Comparison of FPN and B-FPN detection results (Green box is real target box and yellow box is missed ship target).

4.4.6. Compare with Other Methods

In order to further verify the performance of the proposed algorithm for ship target detection, SR-CNN was compared with three representative two-stage target detection algorithms which are Faster R-CNN, Cascade R-CNN, and Libra R-CNN. These methods use the same data partitioning and optimization parameters, using pre-trained weights on ImageNet.

As shown in Table 5, compared with the other detection algorithms, SR-CNN is able to achieve better detection results under the same conditions. In fact, the parameters (Params) of Faster R-CNN are 28.147502 M and the Params of SR-CNN are 28.147514 M. Therefore, compared to the Faster R-CNN with backbone as Resnet18 and neck as FPN, the SR-CNN improves mAP by 3.9% with only a few parameters and floating-point operations per second (FLOPs) added. Compared to Faster R-CNN with backbone as Resnet50, neck as FPN and backbone as Resnet18, neck as PAFPN, when the Params and FLOPs of SR-CNN are far less than these two structures, the mAP of SR-CNN increases by 0.8% and 2.1%, respectively. Compared with Cascade R-CNN, although the mAP of the model is the same, the Params of SR-CNN are only half of Cascade R-CNN and the FLOPs are smaller. Compared with Libra R-CNN, SR-CNN also has smaller Params and FLOPs, while the mAP of the model is improved by 0.9%.

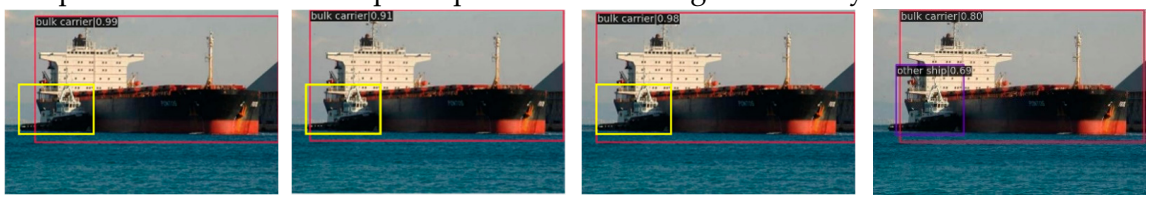
In order to compare the detection results of different algorithms more intuitively, two typical scenarios of nearshore and remote shore were selected from the ship dataset and Faster R-CNN benchmark algorithm, Cascade R-CNN algorithm, Libra R-CNN algorithm and SR-CNN algorithm were used for detection. The comparison results are shown in Figures 12 and 13. In the figures, the yellow box represents the missed ship target and the orange box represents the incorrect ship detection.

Table 5. Performance comparison of different detection algorithms.

Algorithm	Backbone	FLOPs (G)	Params (M)	mAP (%)
Faster R-CNN+FPN	Resnet18	148.00	28.15	51.0
Faster R-CNN+FPN	Resnet50	203.44	41.15	54.1
Faster R-CNN+PAFPN	Resnet18	174.02	31.69	52.8
Cascade R-CNN+FPN	Resnet18	150.02	55.94	**54.9**
Libra R-CNN+FPN	Resnet18	149.11	28.41	54.0
SR-CNN	Resnet18	**148.03**	28.15	**54.9**

(**a1**) Faster R-CNN (**b1**) Cascade R-CNN (**c1**) Libra R-CNN (**d1**) SR-CNN

The pictures show a bulk ship in a port with the background of a yard.



(**a2**) Faster R-CNN (**b2**) Cascade R-CNN (**c2**) Libra R-CNN (**d2**) SR-CNN

The pictures show that a large bulk ship goes into port under towing by a tow ship.

(**a3**) Faster R-CNN (**b3**) Cascade R-CNN (**c3**) Libra R-CNN (**d3**) SR-CNN

The pictures show a tow ship moored in front of a large liner in the passenger terminal.

Figure 12. Comparison of ship detection algorithms in nearshore complex background (Orange box is incorrect ship detection and yellow box is missed ship target).

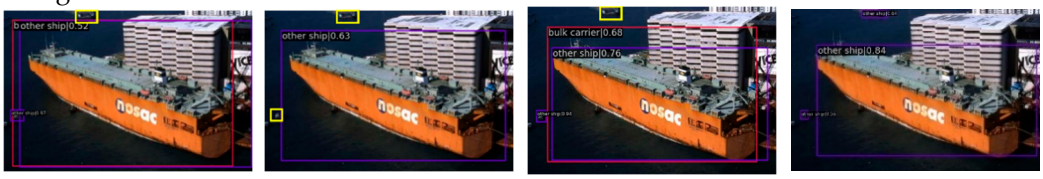
As can be seen from Figure 12, unreasonable feature fusion in Faster R-CNN algorithm leads to low localization accuracy in the nearshore complex background. The small ship has missed and false detections in Figure 12a. Cascade R-CNN sets different overlap thresholds for each R-CNN through the joint multi-level network structure to balance the distribution of positive and negative samples of the network, thus obtaining a higher precision prediction box. However, it still has missed detection of small target ships and ship targets in complex backgrounds, as shown in Figure 12b. Libra R-CNN has significantly improved the detection capability for small target ships, but there are still missed and false detections of ship targets in complex backgrounds, as shown in Figure 12c. SR-CNN can not only identify ship targets in the complex background of offshore scenes, but also reduce false detection, as shown in Figure 12d.

As can be seen from Figure 13, due to the complexity of the background in the remote shore scene and the high ambiguity of the ship itself, the discrimination between the ship and the background is low, so it is more likely to cause missed detection of ship targets. As shown in Figure 13a,b, Faster R-CNN algorithm and Cascade R-CNN algorithm are not sensitive to small-scale target ship targets in remote shore areas. Libra R-CNN has better

detection of small targets, but it still misses some small targets in complex backgrounds, as shown in Figure 13c. SR-CNN improves the top-down information flow through feature fusion to provide low-level details needed for coordinate regression and improve the positioning ability of ship features as shown in Figure 13d.

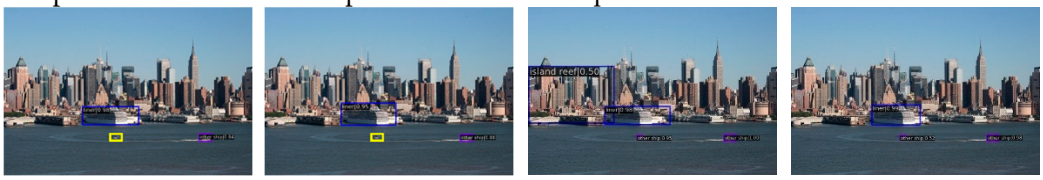
(**a1**) Faster R-CNN (**b1**) Cascade R-CNN (**c1**) Libra R-CNN (**d1**) SR-CNN

The pictures show two ships in an inland river with buildings across the river in the background.



(**a2**) Faster R-CNN (**b2**) Cascade R-CNN (**c2**) Libra R-CNN (**d2**) SR-CNN

The pictures show a ro-ro ship and some small ships at the dock.



(**a3**) Faster R-CNN (**b3**) Cascade R-CNN (**c3**) Libra R-CNN (**d3**) SR-CNN

The pictures show ships at passenger terminals and small ships travelling in inland rivers.

Figure 13. Comparison of ship detection algorithms in remote shore complex background (Blue box is liner and yellow box is missed ship target).

5. Discussion

With the development of economy and society, the demand for inland waterway cargo transportation is increasing, and ship transportation is becoming more and more frequent. However, due to the complex navigation environment, the large number of ships, and the frequent occurrence of overtaking and crossing, various types of traffic accidents occur frequently. This shows that the maritime supervision ability of inland ships is obviously insufficient. Therefore, it is particularly important to improve the existing water traffic safety supervision system. A video surveillance system can visually display the water traffic and ship movements. It has become an important monitoring method to detect ships. However, the processing of video surveillance at this stage usually relies mainly on manual viewing, which is inefficient. This is also not conducive to compatibility with other ship monitoring methods.

Based on the above reasons, this paper applies the target detection technology based on deep learning for the identification of ships in inland rivers. We optimized the existing methods according to the characteristics of the inland river environment. To a certain extent, the problems of false detection, missed detection, and inaccurate positioning in ship detection under the complex background of inland waterways were improved. For example, in the first example of Figures 10 and 11, the ship target boxes obtained by our method are closer to the real ship contour. Therefore, our method is more accurate for ship positioning. In the second and third examples of Figure 12, our method can detect missing ship targets in complex backgrounds. It indicates that our method improves the performance of the model in complex environments. However, this technology still has

some limitations: first, object detection techniques based on deep learning often require a large amount of data for training to obtain strong robustness. However, the lack of diversity of shooting angles of ships in the data collected in this paper affects the overall recognition performance of the model. Second, this paper only focuses on ship recognition under illumination conditions, and does not consider ship recognition under night conditions. At night, the ship navigation environment is more complex, and more effective means are needed to detect it. Finally, this study only considers the ship recognition method based on video images. The advantages of different methods are not combined to achieve more efficient and accurate ship identification. In view of these shortcomings, the next research work will be carried out from the following aspects: (1) Collecting multi-angle ship images to expand the data set to enhance its universality; (2) combining with multi-spectral image recognition technology to improve the stability of ship identification in different environments—for example, ship recognition based on infrared images can be considered; (3) combining different detection methods, comprehensively analyzing the advantages and disadvantages of different methods, and combining different methods to achieve more comprehensive ship monitoring, such as video, AIS, and radar.

6. Conclusions

In order to address the issues of false detection, missed detection, and inaccurate localization in ship identification for maritime video surveillance in the challenging environment of inland rivers, an improved feature fusion and non-maximum suppression ship target detection algorithm, SR-CNN, is proposed in this paper. By upgrading the post-processing technique through Soft-DIOU-NMS with weighted box fusion, it enhances the model's capacity to detect ships in surveillance photographs in dense settings. The relevance of various features is determined using ECA-based weighted feature fusion and normalized weighted feature fusion, which makes feature fusion more logical. This can enhance noise interference in complicated situations and lessen information loss during the fusion process. Finally, a ship dataset was produced by collecting ship monitoring images and comparing the proposed algorithm with the classical algorithm. The results show that the detection accuracy of SR-CNN is improved by 3.9% and the recall rate is improved by 1.5% compared with Faster R-CNN.

Author Contributions: Conceptualization, M.Y. and S.H.; methodology, M.Y. and S.H.; validation, M.Y., S.H., H.W. and T.W.; formal analysis, H.W.; resources, T.W.; data curation, S.H.; writing—original draft preparation, S.H. and T.W.; writing—review and editing, M.Y. and H.W.; visualization, S.H.; supervision, M.Y.; project administration, M.Y.; funding acquisition, M.Y. and H.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (2020YFB1712400) and the National Natural Science Foundation of China (No.52272423, No.71672137).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Zhang, M.; Zhang, D.; Fu, S.; Kujala, P.; Hirdaris, S. A predictive analytics method for maritime traffic flow complexity estimation in inland waterways. *Reliab. Eng. Syst. Saf.* **2022**, *220*, 108317. [CrossRef]
2. Qin, X.; Yan, M.; Zhu, D. Research on information fusion structure of radar and AIS. In Proceedings of the 2018 Chinese Control And Decision Conference (CCDC), Shenyang, China, 9–11 June 2018; pp. 3316–3322.
3. Lazarowska, A. Verification of ship's trajectory planning algorithms using real navigational data. *TransNav Int. J. Mar. Navig. Saf. Sea Transp.* **2019**, *13*, 559–564. [CrossRef]

4. Nosov, V.N.; Kaledin, S.B.; Ivanov, S.G.; Timonin, V.I. Remote Tracking to Monitor Ship Tracks at or near the Water Surface. *Opt. Spectrosc.* **2019**, *127*, 669–674. [CrossRef]
5. Zhang, M.; Conti, F.; Le Sourne, H.; Vassalos, D.; Kujala, P.; Lindroth, D.; Hirdaris, S. A method for the direct assessment of ship collision damage and flooding risk in real conditions. *Ocean Eng.* **2021**, *237*, 109605. [CrossRef]
6. Zhang, M.; Kujala, P.; Hirdaris, S. A machine learning method for the evaluation of ship grounding risk in real operational conditions. *Reliab. Eng. Syst. Saf.* **2022**, *226*, 108697. [CrossRef]
7. Zhang, M.; Montewka, J.; Manderbacka, T.; Kujala, P.; Hirdaris, S. A Big Data Analytics Method for the Evaluation of Ship—Ship Collision Risk reflecting Hydrometeorological Conditions. *Reliab. Eng. Syst. Saf.* **2021**, *213*, 107674. [CrossRef]
8. Wolsing, K.; Roepert, L.; Bauer, J.; Wehrle, K. Anomaly Detection in Maritime AIS Tracks: A Review of Recent Approaches. *J. Mar. Sci. Eng.* **2022**, *10*, 112. [CrossRef]
9. Li, L.; Liu, G.; Li, Z.; Ding, Z.; Qin, T. Infrared ship detection based on time fluctuation feature and space structure feature in sun-glint scene. *Infrared Phys. Technol.* **2021**, *115*, 103693. [CrossRef]
10. Farahnakian, F.; Heikkonen, J. Deep Learning Based Multi-Modal Fusion Architectures for Maritime Vessel Detection. *Remote Sens.* **2020**, *12*, 2509. [CrossRef]
11. Wang, Y. Development of AtoN Real-time Video Surveillance System Based on the AIS Collision Warning. In Proceedings of the 2019 5th International Conference on Transportation Information and Safety (ICTIS), Liverpool, UK, 14–17 July 2019; pp. 393–398.
12. Nalamati, M.; Sharma, N.; Saqib, M.; Blumenstein, M. Automated Monitoring in Maritime Video Surveillance System. In Proceedings of the 2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ), Wellington, New Zealand, 25–27 November 2020; pp. 1–6.
13. Song, H.; Lee, K.; Kim, D.H. Obstacle Avoidance System with LiDAR Sensor Based Fuzzy Control for an Autonomous Unmanned Ship. In Proceedings of the 2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS), Toyama, Japan, 5–8 December 2018; pp. 718–722.
14. Tassetti, A.N.; Galdelli, A.; Pulcinella, J.; Mancini, A.; Bolognini, L. Addressing Gaps in Small-Scale Fisheries: A Low-Cost Tracking System. *Sensors* **2022**, *22*, 839. [CrossRef]
15. Galdelli, A.; Mancini, A.; Tassetti, A.N.; Ferrà Vega, C.; Armelloni, E.; Scarcella, G.; Fabi, G.; Zingaretti, P. A cloud computing architecture to map trawling activities using positioning data. In Proceedings of the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Anaheim, CA, USA, 18–21 August 2019; American Society of Mechanical Engineers: New York, NY, USA, 2019; Volume 59292, p. V009T12A035.
16. Kanjir, U.; Greidanus, H.; Oštir, K. Vessel detection and classification from spaceborne optical images: A literature survey. *Remote Sens. Environ.* **2018**, *207*, 1–26. [CrossRef] [PubMed]
17. Chen, Z.; Chen, D.; Zhang, Y.; Cheng, X.; Zhang, M.; Wu, C. Deep learning for autonomous ship-oriented small ship detection. *Saf. Sci.* **2020**, *130*, 104812. [CrossRef]
18. Li, Z.; Zhang, Q.; Long, T.; Zhao, B. Ship target detection and recognition method on sea surface based on multi-level hybrid network. *J. BEIJING Inst. Technol.* **2021**, *30*, 1–10.
19. Chen, X.; Qi, L.; Yang, Y.; Postolache, O.; Yu, Z.; Xu, X. Port Ship Detection in Complex Environments. In Proceedings of the 2019 International Conference on Sensing and Instrumentation in IoT Era (ISSI), Lisbon, Portugal, 29–30 August 2019; pp. 1–6.
20. Sun, J.; Xu, Z.; Liang, S. NSD-SSD: A Novel Real-Time Ship Detector Based on Convolutional Neural Network in Surveillance Video. *Comput. Intell. Neurosci.* **2021**, *2021*, 7018035. [CrossRef]
21. 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]
22. Chang, L.; Chen, Y.-T.; Wang, J.-H.; Chang, Y.-L. Modified Yolov3 for Ship Detection with Visible and Infrared Images. *Electronics* **2022**, *11*, 739. [CrossRef]
23. Zhang, Y.; Li, Q.-Z.; Zang, F.-N. Ship detection for visual maritime surveillance from non-stationary platforms. *Ocean Eng.* **2017**, *141*, 53–63. [CrossRef]
24. Chen, Z.; Li, B.; Tian, L.F.; Chao, D. Automatic detection and tracking of ship based on mean shift in corrected video sequences. In Proceedings of the 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, China, 2–4 June 2017; pp. 449–453.
25. Shan, X.; Zhao, D.; Pan, M.; Wang, D.; Zhao, L. Sea–Sky Line and Its Nearby Ships Detection Based on the Motion Attitude of Visible Light Sensors. *Sensors* **2019**, *19*, 4004. [CrossRef] [PubMed]
26. Liu, L.; Liu, G.; Chu, X.M.; Jiang, Z.L.; Zhang, M.Y.; Ye, J. Ship Detection and Tracking in Nighttime Video Images Based on the Method of LSDT. *J. Phys. Conf. Ser.* **2019**, *1187*, 42074. [CrossRef]
27. You, X.; Yu, N. An Automatic Matching Algorithm Based on SIFT Descriptors for Remote Sensing Ship Image. In Proceedings of the 2011 Sixth International Conference on Image and Graphics, Hefei, China, 12–15 August 2011; pp. 377–381.
28. Shafer, S.; Harguess, J.; Forero, P.A. Sparsity-driven anomaly detection for ship detection and tracking in maritime video. In Proceedings of the Automatic Target Recognition XXV, SPIE, Baltimore, MD, USA, 22 May 2015; Volume 9476, pp. 75–82.
29. 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.

30. Zhang, L.; Lin, L.; Liang, X.; He, K. Is faster R-CNN doing well for pedestrian detection? In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 443–457.
31. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
32. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
33. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
34. Law, H.; Deng, J. Cornernet: Detecting objects as paired keypoints. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 734–750.
35. Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. Centernet: Keypoint triplets for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 6569–6578.
36. Wang, L.; Fan, S.; Liu, Y.; Li, Y.; Fei, C.; Liu, J.; Liu, B.; Dong, Y.; Liu, Z.; Zhao, X. A Review of Methods for Ship Detection with Electro-Optical Images in Marine Environments. *J. Mar. Sci. Eng.* **2021**, *9*, 1408. [CrossRef]
37. Shao, Z.; Wang, L.; Wang, Z.; Du, W.; Wu, W. Saliency-aware convolution neural network for ship detection in surveillance video. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 781–794. [CrossRef]
38. Kim, K.; Hong, S.; Choi, B.; Kim, E. Probabilistic ship detection and classification using deep learning. *Appl. Sci.* **2018**, *8*, 936. [CrossRef]
39. Chen, J.; Xie, F.; Lu, Y.; Jiang, Z. Finding arbitrary-oriented ships from remote sensing images using corner detection. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 1712–1716. [CrossRef]
40. Qi, L.; Li, B.; Chen, L.; Wang, W.; Dong, L.; Jia, X.; Huang, J.; Ge, C.; Xue, G.; Wang, D. Ship target detection algorithm based on improved faster R-CNN. *Electronics* **2019**, *8*, 959. [CrossRef]
41. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
42. Hosang, J.; Benenson, R.; Schiele, B. Learning non-maximum suppression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4507–4515.
43. Bodla, 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, Venice, Italy, 22–29 October 2017; pp. 5561–5569.
44. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. 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.
45. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 818–833.
46. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. *arXiv* **2020**, arXiv:1910.03151.
47. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.
48. Chen, K.; Wang, J.; Pang, J.; Cao, Y.; Xiong, Y.; Li, X.; Sun, S.; Feng, W.; Liu, Z.; Xu, J.; et al. MMDetection: Open MMLab Detection Toolbox and Benchmark. *arXiv* **2019**, arXiv:1906.07155. [CrossRef]
49. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.