

Article

Ship Target Detection Algorithm Based on Improved YOLOv5

Junchi Zhou ¹, Ping Jiang ^{1,*}, Airu Zou ¹, Xinglin Chen ² and Wenwu Hu ^{1,*}

¹ College of Mechanical and Electrical Engineering, Hunan Agricultural University, Changsha 410128, China; jcz9858@stu.hunau.edu.cn (J.Z.); zar@stu.hunau.edu.cn (A.Z.)

² College of Information and Intelligent Science and Technology, Hunan Agricultural University, Changsha 410128, China; jayson@stu.hunau.edu.cn

* Correspondence: jiangping@hunau.net (P.J.); 1233087@hunau.edu.cn (W.H.)

Abstract: In order to realize the real-time detection of an unmanned fishing speedboat near a ship ahead, a perception platform based on a target visual detection system was established. By controlling the depth and width of the model to analyze and compare training, it was found that the 5S model had a fast detection speed but low accuracy, which was judged to be insufficient for detecting small targets. In this regard, this study improved the YOLOv5s algorithm, in which the initial frame of the target is re-clustered by K-means at the data input end, the receptive field area is expanded at the output end, and the loss function is optimized. The results show that the precision of the improved model's detection for ship images was 98.0%, and the recall rate was 96.2%. Mean average precision (*mAP*) reached 98.6%, an increase of 4.4% compared to before the improvements, which shows that the improved model can realize the detection and identification of multiple types of ships, laying the foundation for subsequent path planning and automatic obstacle avoidance of unmanned ships.

Keywords: machine vision; target detection; YOLOv5; loss function; unmanned ship



Citation: Zhou, J.; Jiang, P.; Zou, A.; Chen, X.; Hu, W. Ship Target Detection Algorithm Based on Improved YOLOv5. *J. Mar. Sci. Eng.* **2021**, *9*, 908. <https://doi.org/10.3390/jmse9080908>

Academic Editors: Haitong Xu, Lúcia Moreira and Carlos Guedes Soares

Received: 2 August 2021

Accepted: 20 August 2021

Published: 22 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As intelligent platforms that can be used for marine monitoring, unmanned surface ships need to complete complex and orderly autonomous operation tasks such as target recognition and obstacle avoidance when operating at high speeds on complex and uncertain surface environments. Accurate recognition and automatic obstacle avoidance place high requirements on the high-speed information processing capabilities of the vision system of an unmanned ship [1].

In recent years, deep learning has been widely used in the entire target detection field including in face recognition [2,3], in vehicle detection and recognition [4,5], in autonomous driving [6], and in the medical industry [7]. Compared to the SIFT algorithm [8] and what was proposed by David in the texture extraction algorithm [9–11], which are from among the representative traditional algorithms, as well as the HOG algorithm [12] proposed by the Navneet team, the deep learning target detection algorithm has made a great leap in performance and accuracy, and its model network's anti-scale change and anti-translation capabilities have been significantly improved.

Unmanned platforms are developing rapidly and becoming more mature. Equipment such as unmanned aerial vehicles and unmanned vehicles has gradually become more widely used. Research on unmanned offshore equipment has gradually begun to receive more attention, especially regarding surface unmanned boats, which has caused extensive research by scholars such as that on automatic collision avoidance [13] and path planning [14,15]. Environmental perception and target recognition technology are not only the core keys to realizing the autonomous decision-making functions and autonomous obstacle avoidance functions of surface unmanned boats, but they also improve the safety guarantee for the navigation of the unmanned boat. Therefore, the establishment of a visual inspection system for ships has become a hot issue for autonomous ships at sea. In terms

of ship detection, considering real-time requirements, current mainstream algorithms include TWO-STAGE and ONE-STAGE algorithms. In an algorithm based on area detection, Su J. [16] and Wang G. H. [17] used feature enhancement, pre-training model parameter tuning, and fine-tuning of the classification framework to achieve higher detection accuracy with the SSD algorithm for inland watercraft. This detection algorithm is slower because it needs to generate a region candidate frame first. In 2016, Redmon proposed YOLO (you only look once) [18]. This kind of regression-based algorithm is used to locate and identify achieved outstanding performance in the field of target detection. Yu Y. [19] and Jiang W. Z. [20] improved YOLOv2 and YOLOv3 by adjusting the network structure and changing the input scale to increase *mAP* to about 80%. However, this method still has room for improvement in the detection of small targets with complex maritime conditions.

Real-time detection of ship targets has high requirements for accuracy. As the latest representative algorithm of the YOLO series, YOLOv5 is characterized by faster speed, higher recognition accuracy, and smaller-sized files, and it can be carried on mobile devices with lower configurations [21], which gives it high research value. In this research, the model was applied to ship detection based on an unmanned ship platform. Aiming at the problem of poor detection of small targets, structure and detection accuracy were improved.

2. Experimental Platform

2.1. Hardware Platform

Figure 1 shows the perception platform based on the target visual detection system, which was an intelligent, water-fishing, unmanned speedboat that integrated water quality detection, automatic bait throwing, automatic obstacle avoidance, unmanned driving, image processing, and other technologies. The mechanical structure of the device was mainly composed of a 304 stainless steel hull and a 304 stainless steel drive shaft seal. The size was 800 mm × 280 mm × 320 mm, and it used a V-shaped bow structure design, which was beneficial for reducing resistance, reducing wake, lowering the center of gravity, enhancing stability, and accommodating more components.

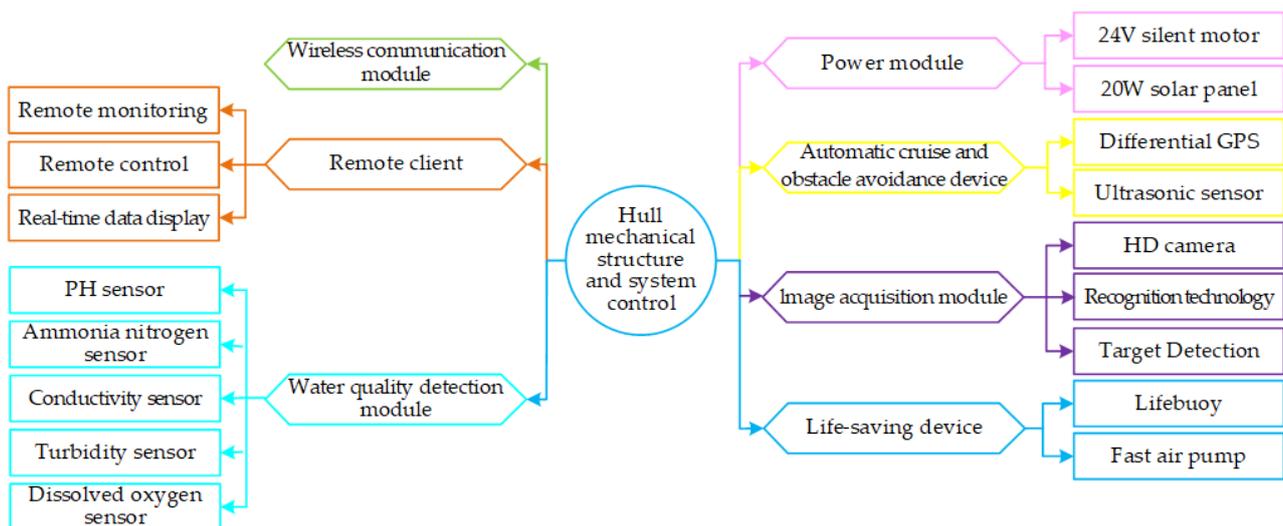


Figure 1. Cont.

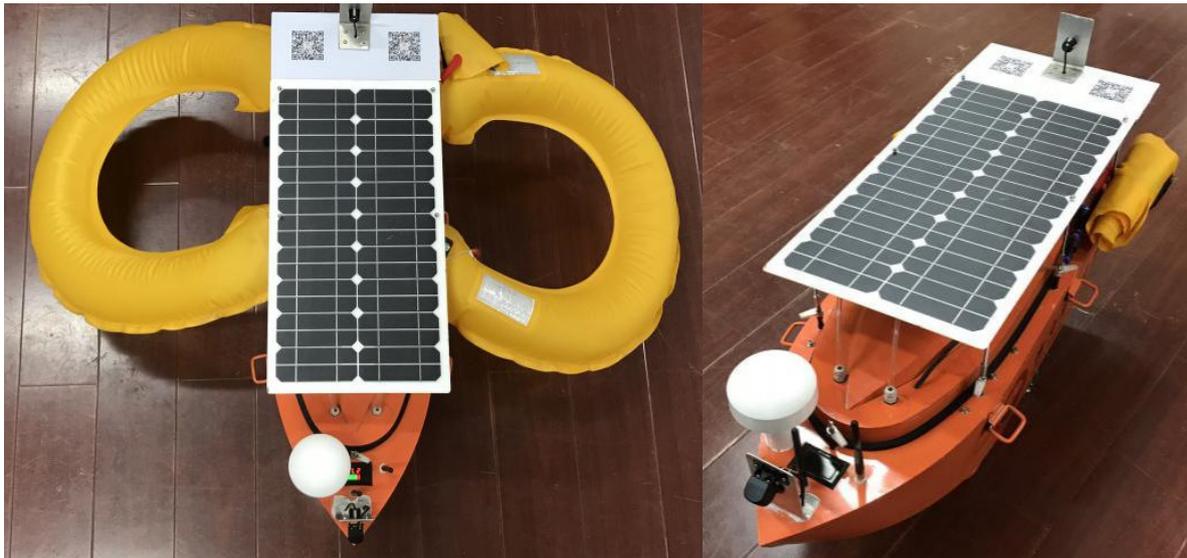


Figure 1. Main functions and structure of the unmanned ship.

2.2. Vision Platform System

This article mainly focuses on image processing target detection. The image recognition module was an embedded Jetson nano development board, as shown in Figure 2, which embedded the improved model algorithm that had been trained in advance and realized wireless communication, remote monitoring, and remote control through a 4G network module.



Figure 2. Image recognition module equipment.

The communication system was divided into an unmanned ship terminal, a cloud server terminal, and a client terminal, which realized the transmission and storage of information and could also realize the remote wireless control of the ship. We performed “end-to-end” calculations through the captured videos and pictures and returned the results to the terminal to issue instructions to the ship. The detection steps are shown in Figure 3.

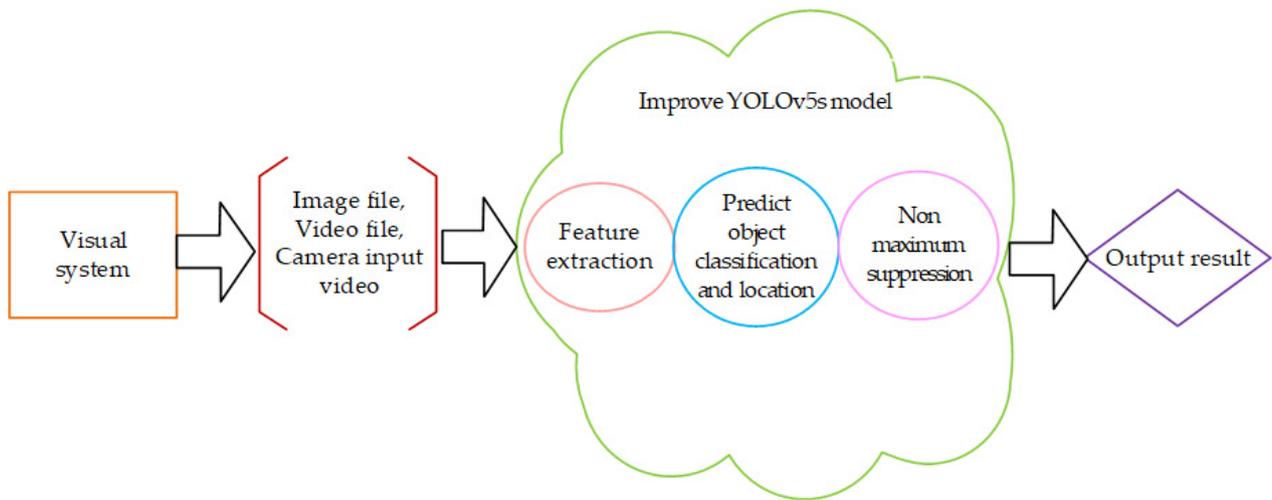


Figure 3. Vision system inspection process.

2.3. Vision Platform System

The graphics card used was an NVIDIA GeForce GTX 1660Ti; the CPU was INTEL Core I7-9750H@2.60 GHz six-core with 16GB of memory. The environment configuration was Windows 10, Python3.8, Pytorch1.8.1, and Cuda10.1, and the framework was TensorFlow. The parameter settings are shown in Table 1.

Table 1. Training parameter settings.

Parameter	Value
Momentum	0.937
Weight_decay	0.0005
Batch_size	45
Learning_rate	0.0001
Epochs	500
Thresh	0.4

3. Principles and Methods

The YOLOv5 model structure is similar to that of other YOLO algorithm series divided into four parts: input, backbone, neck, and prediction. Figure 4 shows the main structure of YOLOv5s.

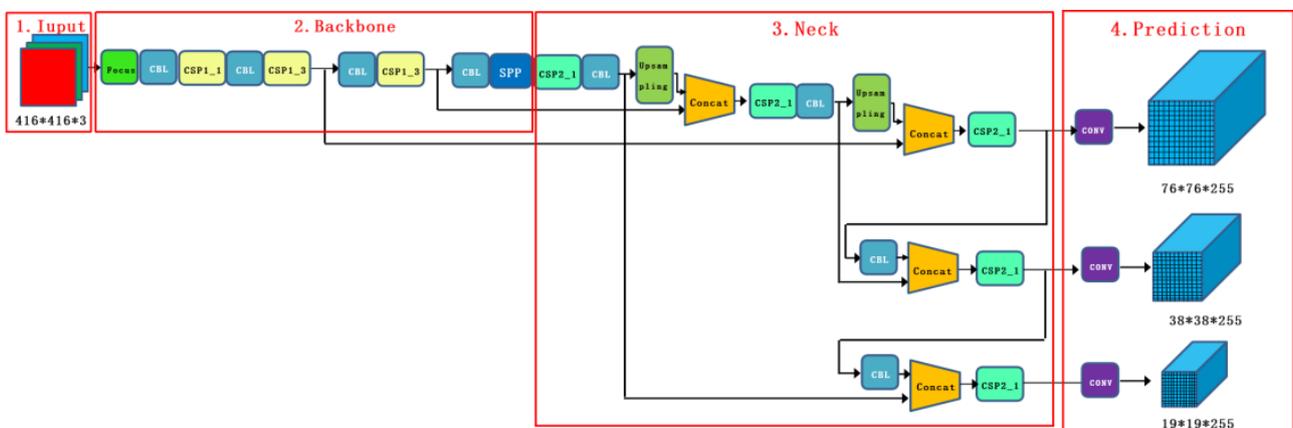


Figure 4. The main structure of the YOLOv5s model.

The input part can realize data enhancement, adaptive anchor frame calculation, and adaptive image scaling. The feature extraction part mainly adopts the focus structure that can complete slicing and convolution operations and the CSP structure that enhances the learning ability of the feature network. Because the Focus and CBL of different networks have different numbers of convolution kernels, and the number of residual modules of the CSP is different, the model can show different performances by controlling the width and depth of the network. The neck part uses FPN and PAN structures, using the information extracted from the backbone part to strengthen the network feature fusion ability. The output layer is divided into three convolutional layer channels, which are calculated through the loss function, and the result is subjected to maximum value suppression processing to give the prediction result.

3.1. Dataset Preparation and Preprocessing

The experiments in this article are divided into public datasets and self-made datasets. The public dataset is the SeaShips dataset, in which the images are from a monitoring system deployed on the coastline, and the pictures obtained from each frame of the image have been intercepted. The self-made dataset was collected from common ships in the river.

The mosaic enhancement method was used to randomly select four pictures for random scaling and then randomly distribution for splicing, which greatly enriched the detection dataset, especially because the random scaling added a lot of small targets, making the network more robust. The enhanced effect is shown in Figure 5.



(a)



(b)

Figure 5. Mosaic data enhancement effect: (a) original picture; (b) mosaic data enhancement.

When the image was zoomed on the input end, there were different black borders around it as well as information redundancy, which affected the training speed. We used Equation (1) to calculate the adaptive zoom:

$$\begin{aligned}
 \frac{416}{x} &= a \\
 \frac{416}{y} &= b \\
 x \times \min(a, b) &= c \\
 y \times \min(a, b) &= d \\
 c - d &= e \\
 np.mod(e, 2^5) &= f
 \end{aligned}
 \tag{1}$$

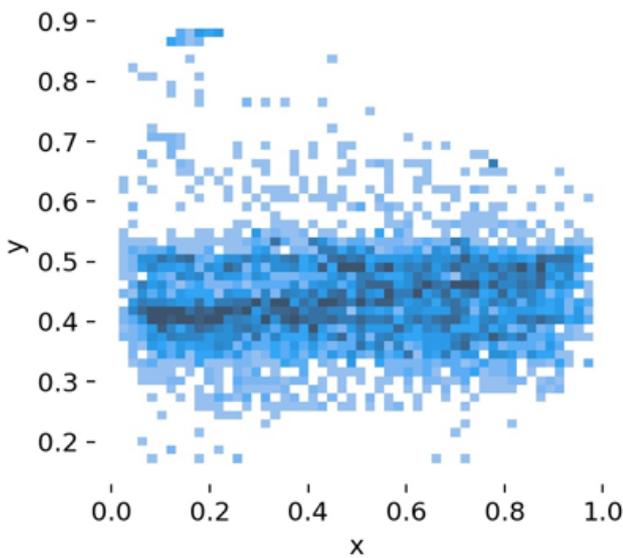
where x and y represent the length and width of the input, respectively; c and d represent the scaled size; e is the original height that needs to be filled; and f is the sum of the two sides that need to be filled.

3.2. YOLOv5s Algorithm Network Structure Improvement

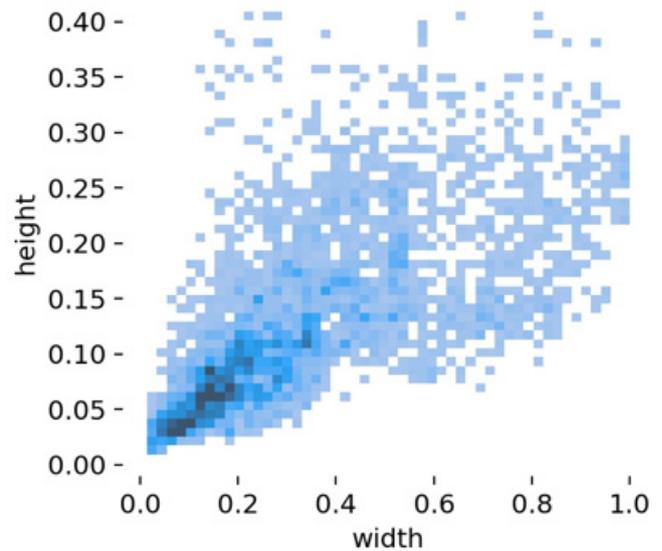
Figure 6a is an anchor frame distribution map to show the intuitive situation of data labels, and an overall analysis of the target position and target size on the label data obtained a target relative position map, as shown in Figure 6b, as well as a target relative size map, as shown in Figure 6c.



(a)



(b)



(c)

Figure 6. Statistical results of sample data: (a) anchor frame distribution map; (b) normalized target location map; (c) normalized target size map.

Figure 6b shows that the lower left corner of the data set picture was set as the coordinate origin to establish a rectangular coordinate system, and the relative coordinate values of the abscissa x and the ordinate y were used to evaluate the relative position of the target. The results show that the horizontal direction of the target runs through the entire coordinate axis, and the vertical direction is more concentrated but somewhat discrete.

Figure 6c shows that the width of the target mostly occupied 2~5% of the image width, and the target height mostly occupied 5~8% of the image height.

It can be seen from the above analysis that there was a large gap between the initial set of regional candidate frames and the distribution of the dataset because the target sample dataset had a rich variety of objects in different sizes, resulting in the insufficient detection of small targets and unbalanced targets. Therefore, the initial frame of the target was clustered first, and the loss function module and the receptive field area were improved.

3.2.1. K-Means Dimensional Clustering

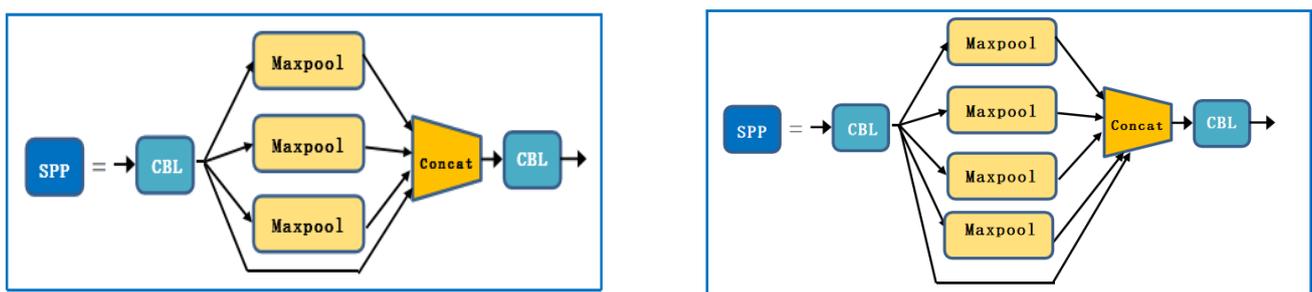
To improve the accuracy of ship identification, the direct use of the original a priori box cannot fully meet demands. Therefore, the K-means clustering algorithm was used to cluster the target frame of the labeled dataset. The purpose was to give the anchor frame and the detection frame a greater intersection ratio to select the best a priori frame. The calculation formula is as Equation (2):

$$d = 1 - IOU \tag{2}$$

where IOU represents the intersection ratio of the predicted frame and the true frame. The prior boxes obtained by re-clustering were (12,16), (17,39), (30,52), (54,60), (33,26), (126,183), (227,283), (373,326), and (407,486). The allocation was carried out according to the principle of using large a priori boxes for small scales and small a priori boxes for large scales.

3.2.2. Expanding the Receptive Field Area

In many vision tasks, the size of the receptive field is a key issue because each pixel in the output feature map must respond to a large enough area in the image to capture information about large objects. Therefore, we chose to add a maximum pooling layer in the space pyramid to improve multiple receptive fields fusion, thereby improving the detection accuracy of small targets. The improved structure is shown in Figure 7.



(a)

Figure 7. Cont.

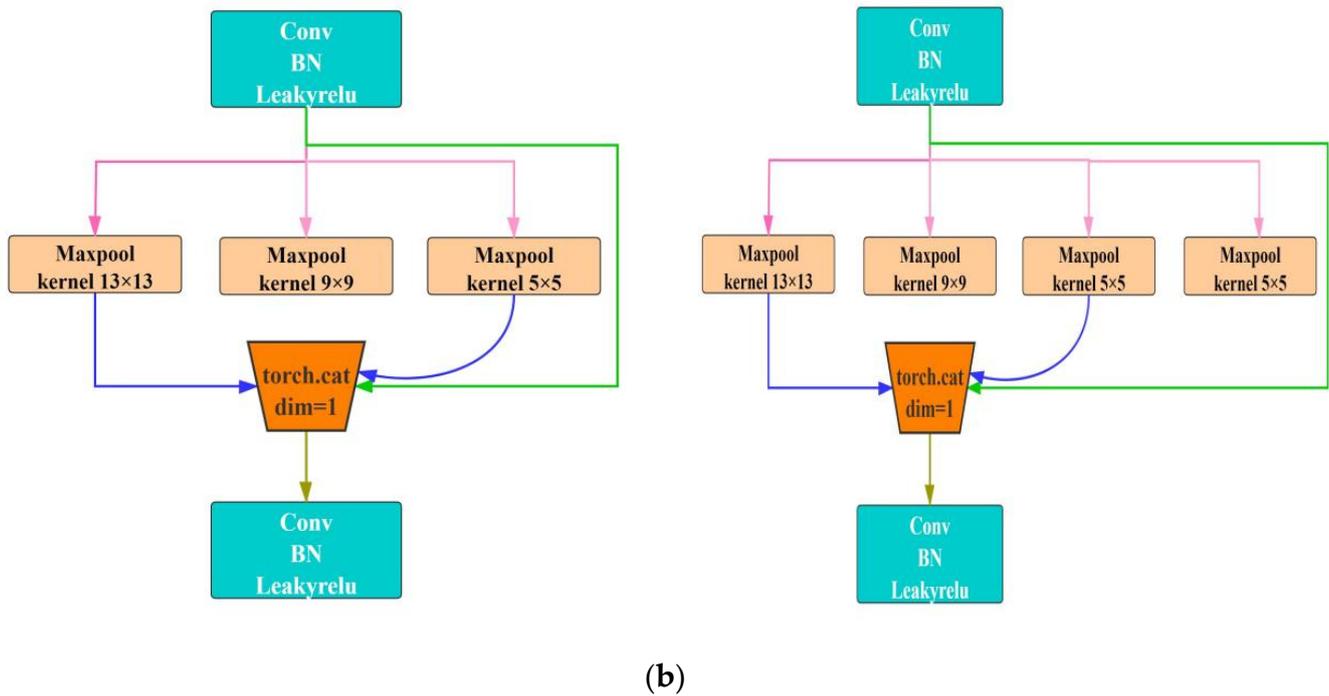


Figure 7. Comparison before and after pooling layer improvement: (a) macro structure; (b) micro structure.

Figure 7a is the macro structure, which visually shows that a maximum pooling layer has been added. Figure 7b shows the microstructure. In the figure, SPP is a spatial pyramid pooling module, and CBL is a combination module comprising a convolutional layer, a BN layer, and an activation function layer. From a microscopic point of view, we increased the receptive field of the model by adding a 3×3 maximum pooling filter.

3.2.3. Improved Loss Function

Equations (2)–(5) are the loss functions of the original YOLOv5 algorithm that was used for the bounding box, $GIOU_Loss$, which has certain limitations. When there is a phenomenon contained between the detection box and the real box, the overlapping part is unable to be optimized. For confidence and category loss, the original algorithm uses a two-category, cross-entropy loss function, which, to a certain extent, is not conducive to the classification of positive and negative samples.

$$Loss = GIOU_Loss + Loss_{conf} + Loss_{class} \tag{3}$$

$$GIOU_Loss = 1 - GIOU = 1 - (IOU - \frac{|Q|}{C}) \tag{4}$$

where C represents the smallest bounding rectangle between the detection frame and the prior frame and Q represents the difference between the smallest bounding rectangle and the union of the two boxes.

$$Loss_{conf} = \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - \hat{C}_i^j)] - \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - \hat{C}_i^j)] \tag{5}$$

$$Loss_{class} = \sum_{i=0}^{S^2} I_{ij}^{obj} \sum_{c \in classes} [\hat{P}_i^j \log(P_i^j) + (1 - \hat{P}_i^j) \log(1 - \hat{P}_i^j)] \tag{6}$$

where I_{ij}^{obj} and I_{ij}^{noobj} indicate whether there is a target in the j th detection frame in the i th grid, λ_{noobj} is the loss weight of the positioning error, C_j^i and P_j^i are training values, and \hat{C}_i^j and \hat{P}_i^j are predicted values.

According to the above problems, the improved loss function of Equations (6)–(8) was adopted. The bounding box of the improved algorithm used the *CIOU_Loss* loss function to increase the restriction mechanism for the aspect ratio so that the prediction box would be more in line with the real box. Confidence and category loss functions adopted an improved cross-entropy function, which made the separation of positive and negative samples flexible by changing their weights and reduced the impact on them.

$$CIOU_Loss = 1 - (IOU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v) \tag{7}$$

where $\rho()$ is the Euclidean distance between the center point of the detection frame and the prior frame, c is the diagonal length of the two smallest enclosing rectangles, and α is the weight coefficient.

The distance between the overlapping area and the center point is considered, but the aspect ratio is not considered, so the following parameters are added to the penalty term of DIOU:

$$v = \frac{4}{\pi^2} (\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w^p}{h^p})^2$$

$$\alpha = \frac{v}{(1-IOU)+v} \tag{8}$$

where v is a parameter for measuring the consistency of the aspect ratio.

$$Focal_Loss = \begin{cases} -\alpha(1-p)^{\gamma} \log P' & y = 0 \\ -(1-\alpha)p'^{\gamma} \log(1-p') & y = 1 \end{cases} \tag{9}$$

where α and γ represent coordination parameters.

4. Results and Discussion

The evaluation index system of this experiment included mean average precision, recall rate, and precision rate. The closer a *mAP* value was to 1, the better the overall performance of the model. There were six types of ships in the dataset used in this study, so the *mAP* calculation was the average of the six types of AP, the value of which was the area enclosed by the recall and precision curves, as in Equation (10):

$$recall = \frac{TP}{TP+FN}$$

$$precision = \frac{TP}{TP+FP} \tag{10}$$

$$mAP = \frac{\sum_{i=0}^{N-1} \int_0^1 P(R) dR}{N}$$

where *TP* represents the number of correctly identified ship images, *FP* represents the number of misrecognized ship images, and *FN* represents the number of missed ship images.

4.1. Model Training

By controlling the depth and width of the model, the four models could be trained in groups to determine which model was suitable for the detection of ships on the water. The four models (s, m, l, x) ranged from shallow to deep and from narrow to wide. The depth of the model was related to the number of residual components, and the width was related to the number of convolution kernels. The parameter settings are shown in Table 2.

Table 2. Model structure parameter settings.

Model	Depth_Multiple	Width_Multiple
YOLO v5s	0.33	0.50
YOLO v5m	0.67	0.75
YOLO v5l	1.0	1.0
YOLO v5x	1.33	1.25

The results of group training are shown in Table 3. Although the YOLOv5s model performed slightly worse, the *mAP* values of the other three models were all around 98%.

Table 3. Score of each model.

Model	Precision/%	Recall/%	<i>mAP</i> @0.5/%	<i>mAP</i> @[.5:.95]/%	Time/ms
YOLOv5s	93.3	92.5	94.2	72.5	30.6
YOLOv5m	97.3	97.4	98.0	77.7	93.7
YOLOv5l	97.8	98.2	98.8	78.9	128.5
YOLOv5x	98.2	97.1	98.7	78.0	204.6

Each parameter of the 5× model had a strong fluctuation in the 0–50 rounds; it was judged that the model had great instability for the detection of small targets. The specific situation is shown in Figure 8. The abscissa in the two figures is the epoch, and the ordinate is the value of the loss and *mAP*@0.5.

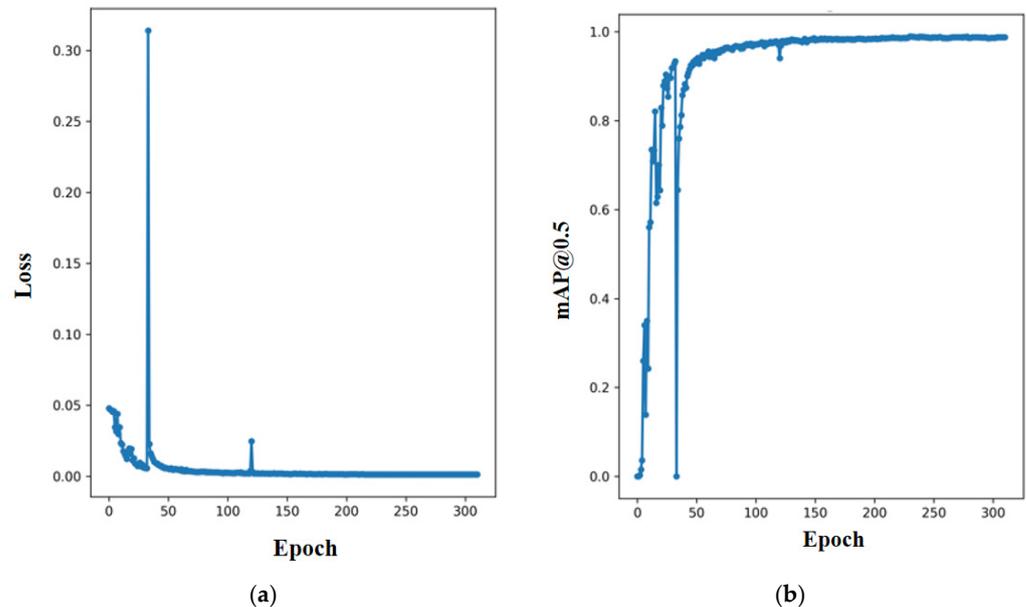


Figure 8. YOLO v5x training results: (a) loss function curve of YOLO v5x; (b) YOLO v5x *mAP*@0.5 curve.

Among them, the 5l and 5m model detection times were too long and did not have good real-time performance; the 5S model had a short detection time, so it had real-time requirements. A reason for its poor accuracy may be that the model is not effective for small target recognition, and the output frame is biased. This study has made improvements to this situation.

4.2. Improved Model Result Analysis and Comparison

Figure 9 shows the improved PR curve of the 5s model. It can be seen that the improved model achieved good recognition results for all types of ships, and the AP value for container ships reached 99.5%.

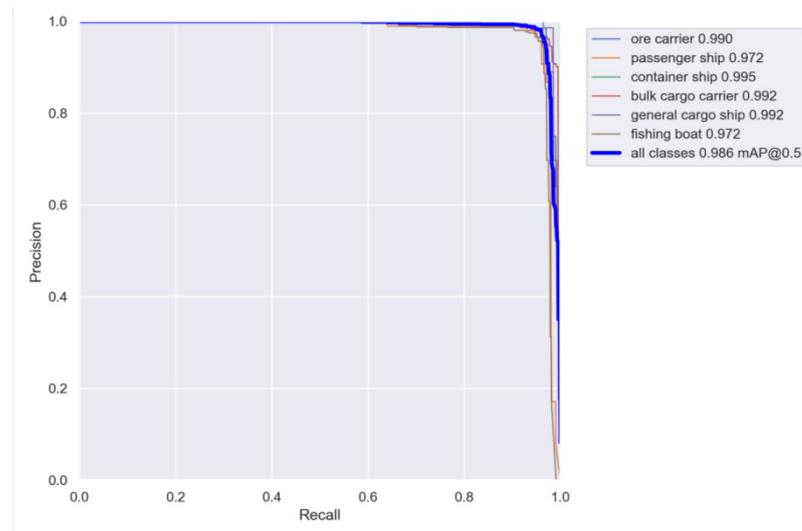


Figure 9. PR curve of YOLOv5s.

The confusion matrix displayed in Figure 10, each column of which representing the predicted category, the total number and value in each column indicating the number of data predicted to be the category and the number of real data predicted to be the category, each row representing the true attribution category of the data, and the total amounts of data in each row representing the number of data instances of that category, shows good stability in detecting various types of ships.

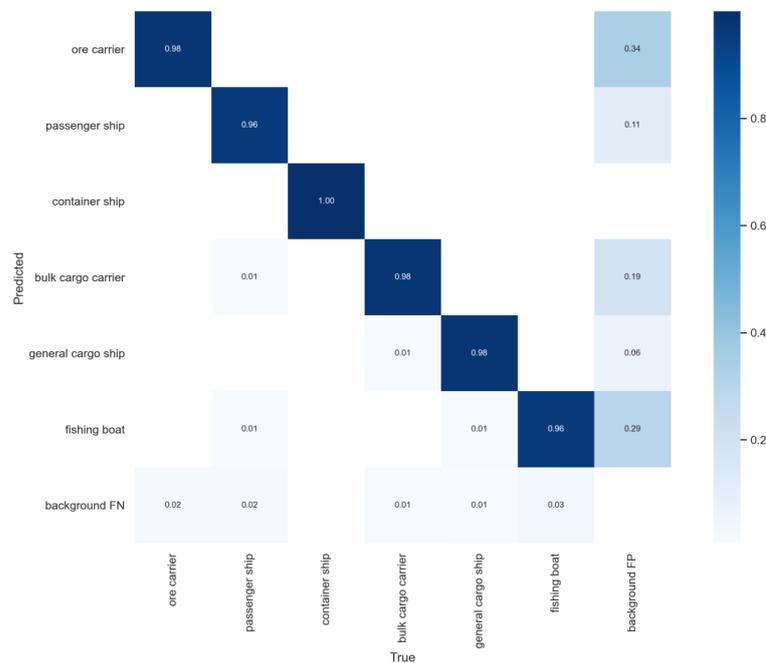


Figure 10. Confusion matrix of YOLOv5s.

Figure 11 shows comparisons of the pictures before and after model detection. Through the comparison pictures, it was found that the small target of the passenger ship that was not originally recognized was detected after the improvement in Figure 11a, which improved the ability for small target detection.

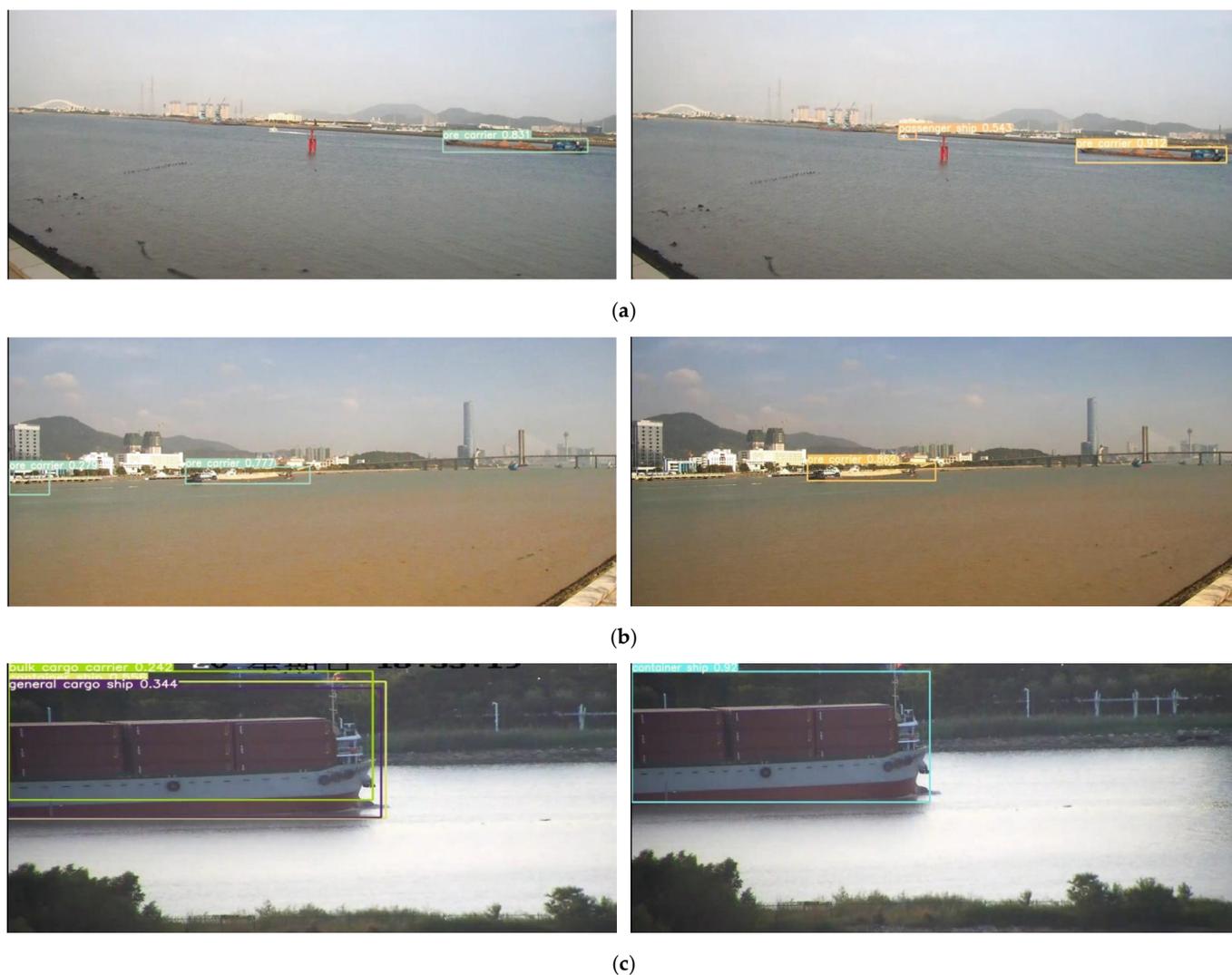


Figure 11. (a) Missed identification improvement; (b) misunderstanding improvement; (c) forecast improvement.

The original algorithm of Figure 11b misidentified the distant shore as an ore ship; the improved algorithm corrected the misidentification of the target and improved the confidence of the original algorithm for the ore ship.

The original algorithm of Figure 11c output multiple sets of prediction boxes, predicting that the target object was a cargo ship, a container ship, or a bulk carrier, but the confidence was low. The improved algorithm improved this situation and made a correct prediction.

Based on the above situation, the algorithm's ability to detect small targets and various types of ships was significantly improved, and the error rate was reduced. Although the detection time increased by 2.2 ms, the *mAP* increased by 4.4% compared with the original algorithm, which indicates that the improved network performance can meet the needs of real-time and accuracy and shows greater improvement compared to YOLOv2 and YOLOv3. The performance comparison is shown in Table 4.

Table 4. Comparison of improved model evaluation.

Model	Precision/%	Recall/%	<i>mAP</i> @0.5/%	<i>mAP</i> @[.5:.95]/%	Time/ms
Improved model	98.0	96.2	98.6	75.9	32.8
YOLOv5s	93.3	92.5	94.2	72.5	30.6
YOLOv3	80.6	79.8	77.2	51.6	50.9
YOLOv2	81.0	83.2	77.6	53.2	16.0

5. Conclusions

Autonomous navigation of unmanned ships at sea is inseparable from accurate detection of maritime targets. The images returned by a camera combined with accurate image analysis techniques can provide powerful preconditions for the perception systems of unmanned ships.

This study analyzed four models by adjusting the width and depth of YOLOv5. The results showed that the 5S model had a low accuracy rate, which may be due to insufficient detection capabilities for small targets, resulting in low accuracy. Therefore, to retain its high detection speed, advantages to improve it are required. By performing K-means dimensional clustering on the target frame of the dataset, the input end adopted mosaic enhancement and image scale transformation, added the largest pooling layer, and optimized the improvement method of the loss function, so that the *mAP* of the improved YOLOv5s reached 98.6%, which was an increase of 4.4% compared to the original; this improved the problem of low detection accuracy for small targets, indicating that the proposed improved method has a better recognition effect and can provide a strong guarantee for automatic driving of unmanned ships.

This research largely concerns the detection of several common ship types. Multi-frame recognition of dynamic targets is the key to dynamic obstacle avoidance at sea. The next step in this research will be to analyze the correlations between the data to identify a variety of other types of targets through transfer learning, improving the generalization ability of the model. This article will provide information support for future research:

- (1) A combination of ultrasonic and Doppler radars for water and underwater detection combined with an onboard camera and drone shooting to achieve obstacle avoidance and real-time path planning;
- (2) Weather detection equipment used to transmit weather conditions in a rescue area in real-time;
- (3) Real-time water depth and velocity detection through echo sounders and ADCP to achieve water rescue in difficult conditions.

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

Funding: This research was funded by the Youth Fund Project of the Hunan Science and Technology Department (No. 2020JJ5234), the Innovation and Entrepreneurship Training Project for College Students in Hunan Province (S202110537052), and the Excellent Youth Project of the Hunan Education Department (No. 20B292).

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. Liu, H. Research on Detection and Segmentation Algorithm of Unmanned Ship Surface Target Based on Deep Learning. Master's Thesis, Wuhan University of Technology of China, Wuhan, China, 2019.
2. Wang, Y.Q.; Feng, Y.W. An Improved Face Recognition Algorithm Based on Convolutional Neural Network. *Semicond. Optoelectron.* **2020**, *41*, 582–586.
3. Tian, X.; Su, H.; Liu, G.; Liu, T. Improved Classroom Face Recognition Algorithm Based on Insight Face and Its Application. *Laser Optoelectron. Prog.* **2020**, *57*, 221501. [[CrossRef](#)]
4. Wang, D.; Chen, X.; Zhao, F.; Sun, H. Vehicle Detection Algorithm Based on Convolutional Neural Network and RGB-D Images. *Laser Optoelectron. Prog.* **2019**, *56*, 181003. [[CrossRef](#)]
5. Yang, L.; Luo, J.; Song, X.; Li, M.; Wen, P.; Xiong, Z. Robust Vehicle Speed Measurement Based on Feature Information Fusion for Vehicle Multi-Characteristic Detection. *Entropy* **2021**, *23*, 910. [[CrossRef](#)] [[PubMed](#)]
6. Wang, Z.Y.; Ni, X.Y.; Shang, Z.D. Autonomous driving semantic segmentation with convolution neural networks. *Opt. Precis. Eng.* **2019**, *27*, 2429–2438. [[CrossRef](#)]
7. Dang, H.; Lu, X.; Sun, J.; Wu, C.; Tao, Y. Design of Vision-Based Drug-Grabbing Robot System. *Laser Optoelectron. Prog.* **2021**, *58*, 0210009. [[CrossRef](#)]
8. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV), Kerkyra, Greece, 20–27 September 1999; pp. 1150–1157.
9. Lin, C.; Xu, G.L.; Cao, Y.; Liang, C.H. Improved contour detection model with spatial summation properties based on nonclassical receptive field. *J. Electron. Imaging* **2016**, *25*, 043018. [[CrossRef](#)]
10. Suh, H.K.; Hofstee, J.W.; De Jsselmui, N.J.; Henten, E.V. Sugar beet and volunteer potato classification using Bag-of-Visual-Words model, Scale-Invariant Feature Transform, or Speeded Up Robust Feature descriptors and crop row information. *Biosyst. Eng.* **2018**, *166*, 210–226. [[CrossRef](#)]
11. Mukherjee, P.; Lall, B. Saliency and KAZE features assisted object segmentation. *Image Vis. Comput.* **2017**, *65*, 82–97. [[CrossRef](#)]
12. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 July 2005; pp. 886–893.
13. Deng, F.; Jin, L.; Hou, X.; Wang, L.; Li, B.; Yang, H. COLREGs: Compliant Dynamic Obstacle Avoidance of USVs Based on the Dynamic Navigation Ship Domain. *J. Mar. Sci. Eng.* **2021**, *9*, 837. [[CrossRef](#)]
14. Xu, H.; Paulo, O.; Guedes Soares, C. L1 adaptive backstepping control for path-following of underactuated marine surface ships. *Eur. J. Control.* **2020**, *58*, 357–372. [[CrossRef](#)]
15. Xu, H.; Hinostroza, M.A.; Guedes Soares, C. Modified Vector Field Path-Following Control System for an Underactuated Autonomous Surface Ship Model in the Presence of Static Obstacles. *J. Mar. Sci. Eng.* **2021**, *9*, 652. [[CrossRef](#)]
16. Su, J.; Yang, L.; Huang, H.; Jin, G. Improved SSD algorithm for small-sized SAR ship detection. *Syst. Eng. Electron.* **2020**, *42*, 1026–1034.
17. Wang, G.; Xie, S.; Chu, X.; Luo, T. Image Recognition Method of Ships in front of Unmanned Surface Vessel Based on Deep Learning. *Ship Eng.* **2018**, *40*, 19–22, 99.
18. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
19. Yu, Y.; Li, S.; Chen, L.; Liu, Y. Ship Target Detection Based on Improved YOLO v2. *Comput. Sci.* **2019**, *46*, 332–336.
20. Jiang, W.; Li, B.; Gu, J.; Liu, K. Ship Target Detection Algorithm based on Improved YOLO V3. *Electron. Opt. Control* **2021**, *28*, 52–56, 67.
21. Tan, S.; Bie, X.; Lu, G.; Tan, X.H. Real-time detection for mask-wearing of personnel based on YOLOv5 network model. *Laser J.* **2021**, *42*, 147–150.