

Article

# A Novel Intelligent Detection Algorithm of Aids to Navigation Based on Improved YOLOv4

Rong Zhen <sup>1</sup> , Yingdong Ye <sup>1</sup>, Xinqiang Chen <sup>2</sup>  and Liangkun Xu <sup>1,\*</sup> 

<sup>1</sup> Navigation College, Jimei University, Xiamen 361021, China

<sup>2</sup> Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China

\* Correspondence: xuliangkun518@163.com

**Abstract:** Aiming at the problem of high-precision detection of AtoN (Aids to Navigation, AtoN) in the complex inland river environment, in the absence of sufficient AtoN image types to train classifiers, this paper proposes an automatic AtoN detection algorithm Aids-to-Navigation-YOLOv4 (AN-YOLOv4) based on improved YOLOv4 (You Only Look Once, Yolo). Firstly, aiming at the problem of an insufficient number of existing AtoN datasets, the Deep Convolutional Generative Adversarial Networks (DCGAN) is used to expand and enhance the AtoN image dataset. Then, aiming at the problem of small target recognition accuracy, the image pyramid is used to multi-scale zoom the dataset. Finally, the K-means clustering algorithm is used to correct the candidate box of AN-YOLOv4. The test on the test dataset shows that the improvement effect of AN-YOLOv4 is obvious. The accuracy rate of small targets is 92%, and the average accuracy (mAP) of eight different types of AtoN is 92%, which is 14% and 13% higher than the original YOLOv4, respectively. This research has important theoretical significance and reference value for the intelligent perception of the navigation environment under the intelligent shipping system.

**Keywords:** intelligent ship; navigation environment perception; detection of Aids to Navigation; machine vision; YOLOv4



**Citation:** Zhen, R.; Ye, Y.; Chen, X.; Xu, L. A Novel Intelligent Detection Algorithm of Aids to Navigation Based on Improved YOLOv4. *J. Mar. Sci. Eng.* **2023**, *11*, 452. <https://doi.org/10.3390/jmse11020452>

Academic Editors: Sébastien Lafond and Sepinoud Azimi

Received: 2 December 2022

Revised: 20 January 2023

Accepted: 14 February 2023

Published: 18 February 2023



**Copyright:** © 2023 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

The navigation environment of inland waterways is complex, and the light and light quality of AtoN are easily affected by poor visibility factors such as rainfall and heavy fog, which leads to potential safety hazards for ship navigation [1,2]. With the vigorous development of the marine economy and shipping industry, the intelligent supervision service of ship shore collaboration has put forward higher requirements for AtoN. At present, the target detection in the inland waterway transport field is mainly carried out for ships, and there is not much research on AtoN in the channel [3,4]. Although the position of inland AtoN is relatively fixed, their volume is small, and their motion characteristics are easily affected by wind, wave, current, and other factors, so the requirements for detection and classification are higher. As the infrastructure for marking the boundary of navigable waters, the accurate detection and identification of AtoN are of great significance for assisting ship navigation and safety [5]. At the same time, this research can also provide support for intelligent ships to carry out the intelligent perception of the navigation environment.

Traditional marine navigation environment perception takes the ship as the detection target, and the detection media can be divided into three categories: radar-related technology, infrared-related technology, and visible light imaging technology [6]. Traditional detection methods are slow in detection speed and low in accuracy and generally only apply to fixed sea areas with poor robustness [7,8]. With the development of deep learning technology, many scholars have begun to try to use a target detection framework for marine target recognition. The deep learning target recognition framework is mainly divided into a single-stage and a multi-stage method [9–13]. The single-stage method is represented

by You Only Look Once (YOLO) [14] and Single Shot MultiBox Detector (SSD) [15]. A bounding box is not formed by these algorithms in advance, and the bounding box and classification are considered together when the network output. Chang et al. [16] combined layers 23, 24, and 25 of YOLOv4 into one layer to avoid repeated operations, thus reducing network running time. The AP (Average Precision, AP) obtained on the SAR Ship Detection Dataset (SSDD) is basically the same as YOLOv4, and the test time of a single picture is more than twice shorter. Li et al. [17] introduced Convolutional Block Attention Module (CBAM) on a tiny YOLOv3 skeleton network to better adapt to complex background images such as onshore buildings and complex light waves on the water surface. The detection accuracy is basically the same as YOLOv3, and the detection speed is faster than YOLOv3. Tang et al. [18] used the HSV algorithm to replace the traditional RGB algorithm in color classification and obtained higher accuracy and a smaller missed detection rate on the public ship dataset High-Resolution Ship Collections 2016 (HRSC2016). In addition, some scholars have conducted research based on the multi-stage method. The multi-stage method, represented by Faster Regions with Convolutional Neural Networks (Faster-RCNN) [19], first obtains the candidate regions and then identifies and classifies the candidate regions. You et al. [20] introduced a semantic segmentation sub-network into Fast RCNN to identify the target area and achieved a small improvement in AP on GaoFeng-2(GF-2) image. Gao et al. [21] proposed an effective training strategy based on Fast-RCNN, which trains a large number of images containing only land areas as negative samples. The effectiveness of the strategy is verified on satellite images. Zhang et al. [22] first used a support vector machine (SVM) to divide the large detection area into small regions of interest (ROI) that may contain ships and then connects these regions to Fast-RCNN for positioning and classification. The experimental results show that this method can improve the accuracy and recall of optical satellite image recognition. Although the multi-stage method has high recognition accuracy, its efficiency is low, and the detection speed is slow. Therefore, the multi-stage method is often not suitable for tasks requiring real-time detection. In addition, some scholars have built their own models for specific needs to identify marine targets. Li et al. [23] build a deep learning-based rapid detection model for maritime targets and detect common targets at sea. It finally achieves an accuracy rate of 87.5% for ship detection, 82.5% for AtoN detection, and 80.0% for island detection. Although this model detects multiple types of targets, its AtoN detection accuracy is about ten percentage points lower than this paper.

Although the research on marine target recognition based on a deep learning framework has made some progress, most of the research is carried out on ship images, and there is a lack of research on AtoN recognition. Compared with the ship target, the size of the AtoN image is often small, and it is difficult to identify distant AtoN targets accurately. The ship candidate box is often a long horizontal bar, while the AtoN candidate box is generally a vertical bar, which is not consistent with the ship candidate box [24]. Moreover, at present, there is a public dataset built for ship detection tasks, but there is no public dataset for the detection of AtoN, which also limits the progress of research on AtoN detection to a certain extent. To solve the above problems and to ensure the accuracy and real-time detection of AtoN in an intelligent maritime environment, an automatic AtoN detection algorithm based on improved YOLOv4 (Aids to Navigatio-YOLOv4, AN-YOLOv4) is proposed in this paper. The main contributions of the proposed algorithm are as follows:

- (1) Aiming at the lack of an AtoN dataset of the AN-YOLOv4 algorithm, a joint dataset expansion method is proposed. This method first introduces DCGAN to augment the AtoN dataset and then uses the image pyramid network to increase the dataset by more than three times on the basis of the augmented pictures. The image pyramid network can enrich the AtoN scale information while supplementing the dataset. It helps to enhance the ability of the network to identify small targets.

- (2) Aiming at the problem that the candidate box of the ordinary Yolov4 algorithm cannot be effectively applied to the detection of the AtoN target, the K-means clustering al-

gorithm is used to modify the AtoN candidate box of the AN-YOLOv4. This method solves the difference between the AtoN candidate box and the common object candidate box.

(3) This method can accurately identify eight different types of AtoN, assist the ship-shore cooperative system in judging the navigation environment, and ensure the safety of ships in the inland river environment.

The rest of this article is organized as follows. In Section 2, the main algorithm YOLOv4 is stated. In Section 3, the algorithm AN-YOLOv4 is proposed, including the dataset joint expansion method using DCGAN and image pyramid and the use of k-means to change the candidate box. In Section 4, the effectiveness of the dataset joint expansion method and the k-means algorithm is verified, and the AN-YOLOv4 algorithm is compared with other algorithms. Section 5 presents the conclusions and future research directions.

## 2. YOLOv4 Algorithm

### 2.1. Algorithm Model Structure

Figure 1 shows the network structure of YOLOv4. The backbone network of YOLOv4 [25] is CSPDarknet53 [26]. The CSP structure is shown in Figure 1c. The purpose of this structure is to enrich the combination of gradients and reduce the amount of computation. The activation function of YOLOv4 is modified from the LeakeyReLU function to the Mish function. As shown in Figure 1b, Adding the Mish [27] function to the backbone network can make the gradient smoother and penetrate deep into the neural network while maintaining accuracy. The Mish function has the characteristics of smoothness, non-monotonicity, lower bounded and upper unbounded, and good performance. The expression of the Mish function is shown in Equation (1), where  $x$  and  $h$  represent the midpoint of the Mish function, respectively. Abscissa and ordinate.

$$Mish = xtanh[\ln(1 + e^x)] \tag{1}$$

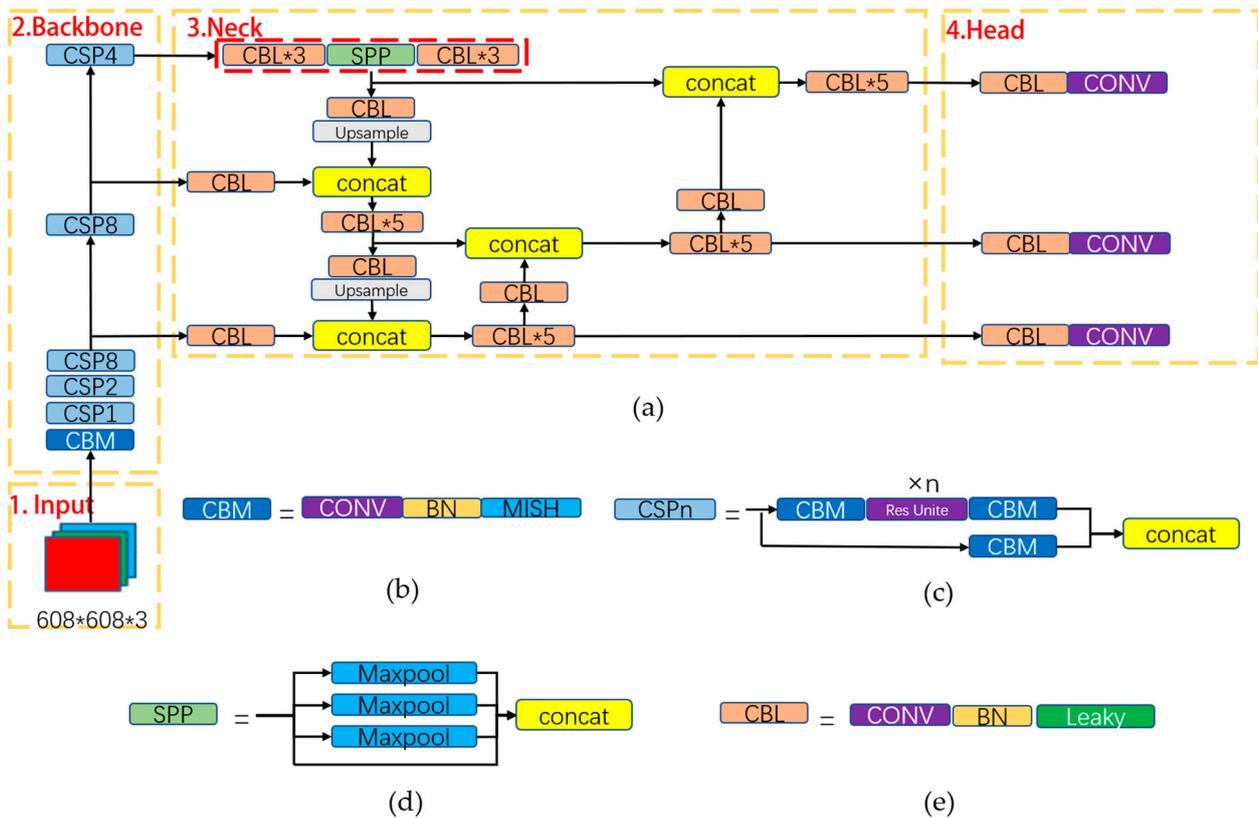


Figure 1. YOLOv4 network structure; (a) YOLOv4 network structure; (b) CBM module structure; (c) CSPn module structure; (d) SPP module structure; (e) CBL module structure.

### 2.2. Spatial Pyramid Pooling Structure

As shown in Figure 1d, the Spatial Pyramid Pooling (SPP) [28] structure is located after the last feature layer of YOLOv4. First, three regular convolutions are performed, and then four different scales of maximum pooling are used to perform the same again. The pooling kernel sizes of max pooling are  $13 \times 13$ ,  $9 \times 9$ , and  $5 \times 5$ , respectively. Finally, the three pooling layers are stacked with the original image. Using the SPP structure can highlight the salient features of the data, greatly improve the speed of generating candidate boxes, and save computational costs.

### 2.3. Path Aggregation Network Structure

YOLOv4 uses the Path Aggregation Network (PANet) [29] algorithm to replace the Feature Pyramid Networks (FPN) [30] algorithm of YOLOv3. The PANet algorithm was proposed by Liu et al. in 2018. PANet can accurately preserve spatial information, which helps to locate pixels correctly. Figure 2a,b are schematic diagrams of FPN and PANet algorithms. FPN has only one back-propagation, while PANet performs one forward-propagation after one back-propagation. The common advantage of FPN and PANet is that they can fuse shallow feature maps with rich, detailed features and deep feature maps with rich semantic features, which improves the problem of poor detection of small objects. However, compared with large target recognition, the detailed features of shallow feature maps are more helpful for localization. The distance from FPN shallow features to the top layer is too long, which affects the effect of deep feature maps locating large targets. Therefore, PANet enhances the path from the shallow layer to the deep layer, shortening the distance between the shallow layer and the deep layer. PANet enhances the localization effect of large objects while fusing shallow and deep features to improve the detection effect of small objects.

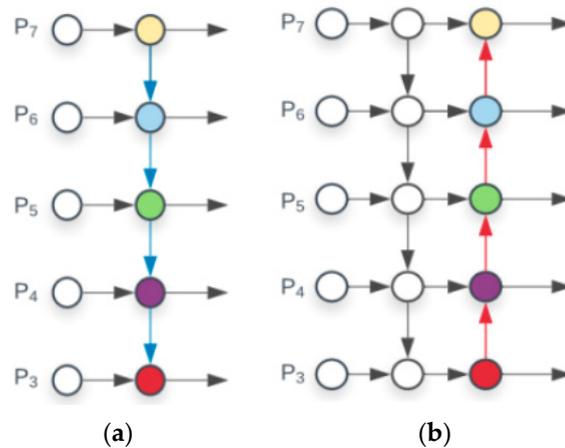


Figure 2. (a) FPN structure; (b) PANet structure.

The AtoN detection has its own particularity in the automatic detection of the intelligent ship platform. The AtoN is of great significance in guiding the ship and ensuring the safe navigation of the ship. Therefore, AtoN detection must be real-time, accurate, and effective. The YOLOv4 algorithm balances the detection speed and accuracy and integrates a large number of excellent technologies, which greatly improves the accuracy. For this reason, this paper realizes the rapid detection of AtoN based on the YOLOv4 algorithm.

### 2.4. CIoU of YOLOv4

The YOLOv3 loss function takes the center coordinates and width of the detection box as independent variables, which is not conducive to the mutual fitting between the dimensions of the detection box. Aiming at the shortcomings of the MSE loss function of YOLOv3, the YOLOv4 loss function comprehensively considers the information such as the length, width, and size of the detection frame, and uses the Intersection Over Union

(IoU) [31] loss instead of the MSE loss. On the basis of IoU, GloU [32] loss, DioU [33] loss, and CloU loss are extended. CloU loss takes into account the overlap area, center distance, and length-width ratio of the three geometric factors, which are in good agreement with the morphological characteristics of the ship. Therefore, this article selects the CloU loss function. The CloU function is:

$$L_{CloU} = 1 - IOU(A, B) + \frac{\rho^2(A_{ctr}, B_{ctr})}{c^2} + \alpha v \tag{2}$$

$$\alpha = \frac{v}{(1 - IOU) + v} \tag{3}$$

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w_{gt}}{h_{gt}} - \arctan \frac{w}{h} \right)^2 \tag{4}$$

where:  $IOU(A, B)$  is the IoU of the prediction frame and the real frame;  $\rho^2(A_{ctr}, B_{ctr})$  is the Euclidean distance between the center point of the prediction frame and the real frame;  $c$  is the diagonal distance of the minimum closure region containing both the prediction frame and the real frame; and  $w_{gt}$  and  $h_{gt}$  are the width and height of the real frame, respectively; and  $w$  and  $h$  are the width and height of the prediction frame, respectively;  $\alpha$  is a positive number;  $v$  is used to measure the consistency of aspect ratio.

### 3. AN-YOLOv4 Algorithm

#### 3.1. Data Enhancement Based on DCGAN

During the research process, we searched the internet and found that the current number of training samples about AtoN images is relatively small. Therefore, this paper uses DCGAN to enhance and expand AtoN dataset. The DCGAN originates from the Generation Adversarial Network (GAN). The GAN is a machine learning architecture proposed by Ian Goodflow [34] of the University of Montreal in 2014. The GAN algorithm takes Game Theory as the basic idea, and the main components are Generator(G) and Discriminator (D). The basic working principle of GAN is shown in Figure 3. The goal of the Generator is to generate a picture close enough to the original picture to give the Discriminator for judgment. The main task of the Discriminator is to check whether the newly generated picture is close enough to the original picture [35]. If the newly generated picture can pass the judgment of the discriminator, the newly generated picture can be used as a supplement to the dataset [36].

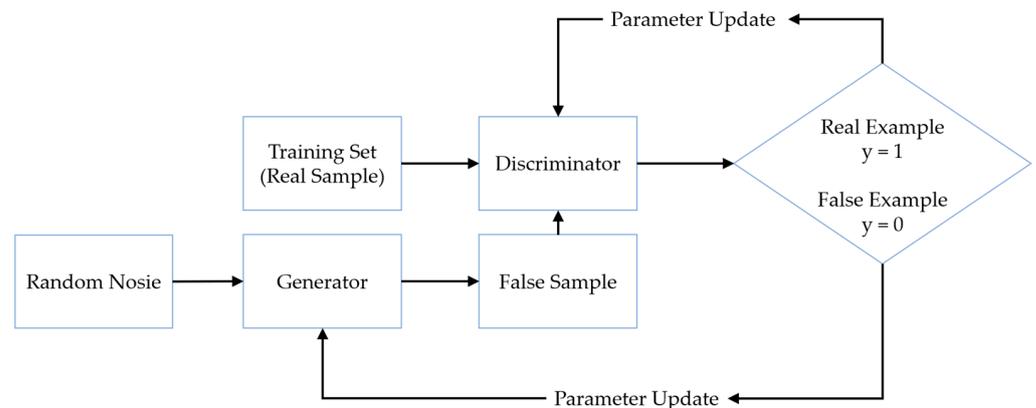
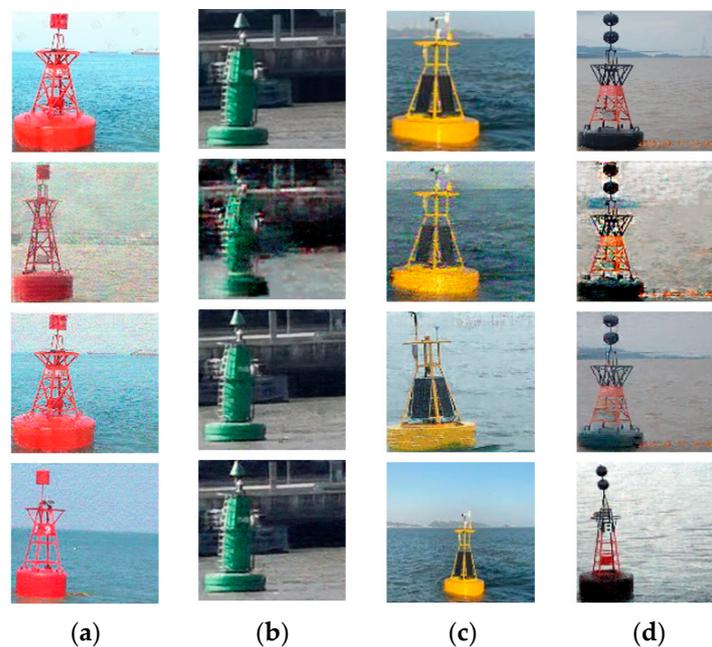


Figure 3. Generative adversarial network schematic.

The DCGAN takes GAN as the basic prototype. The DCGAN was proposed by Radford et al. [37] in 2015. By introducing CNN into Gan, it obtains a more stable training process and higher-quality image samples. In order to satisfy the input of AN-YOLOv4. This paper designs a DCGAN network model. Figure 4 shows the results of DCGAN. Table 1 shows the DCGAN structure designed in this paper. The first nine network layers

starting with G in the table are the generation network, and its function is the same as that of the GAN generator, which can generate AtoN pictures. Then the network starting with D in the next nine layers is the discriminant network, and its function is consistent with the discriminator of GAN, which can discriminate the generated pictures. When the network works, first, the generative network will generate a 100\*1 feature map of the AtoN image. Then, the image is resized to the same size as the original image through a seven-layer deconvolutional network. Finally, the generated image and the original image are put into the discriminant network at the same time, and the discriminant network is used to judge whether the input image is a qualified image. Similar to GAN’s generator and discriminator, the generative and discriminative networks will continuously adjust parameters. Under the mutual adjustment, the probability of the image generated by the generation network passing through the discriminant network remains within the preset target, making the newly generated image more effective.



**Figure 4.** The results of DCGAN; (a) left lateral mark; (b) right lateral mark; (c) bearing mark; (d) isolated danger mark. The first row, the second row, the third row, and the fourth row are the original picture, 50 epochs of training, 100 epochs of training, and 200 epochs of training, respectively. Select the image with appropriate noise and illumination brightness as the supplementary set of the dataset.

**Table 1.** The structure of the DCGAN model designed in this paper.

	Layers Name	Filter Size/Stride	Operations	Input Layer	Output Size (W × H × C)
Generator	G-random input		N/A		1 × 1 × 100
	Reshape		BN	G-random input	4 × 4 × 1024
	T-Conv1	4 × 4/2	BN, ReLU	Reshape	8 × 8 × 512
	T-Conv2	4 × 4/2	BN, ReLU	T-Conv1	16 × 16 × 512
	T-Conv3	4 × 4/2	BN, ReLU	T-Conv2	32 × 32 × 256
	T-Conv4	4 × 4/2	BN, ReLU	T-Conv3	64 × 64 × 256
	T-Conv5	4 × 4/2	BN, ReLU	T-Conv4	128 × 128 × 128
	T-Conv6	4 × 4/2	BN, ReLU	T-Conv5	256 × 256 × 128
T-Conv7	4 × 4/2	BN, tanh	T-Conv6	512 × 512 × 3	

Table 1. Cont.

	Layers Name	Filter Size/Stride	Operations	Input Layer	Output Size (W × H × C)
Discriminator	D-input	3 × 3/2	N/A		512 × 512 × 3
	Conv1	3 × 3/2	LeakyReLU	D-input	256 × 256 × 128
	Conv2	3 × 3/2	BN, LeakyReLU	Conv1	128 × 128 × 128
	Conv3	3 × 3/2	BN, LeakyReLU	Conv2	64 × 64 × 256
	Conv4	3 × 3/2	BN, LeakyReLU	Conv3	32 × 32 × 256
	Conv5	3 × 3/2	BN, LeakyReLU	Conv4	16 × 16 × 512
	Conv6	3 × 3/2	BN, LeakyReLU	Conv5	8 × 8 × 512
	Conv7	3 × 3/2	BN, LeakyReLU	Conv6	4 × 4 × 1024
	flatten		N/A	Conv7	1 × 1 × 16384
	Sigmoid		N/A	flatten	1 × 2 × 1

### 3.2. Data Expansion Based on Image Feature Pyramid

In actual navigation tasks, the scale of the AtoN targets collected by the ship-borne camera changes drastically, and the distant AtoN targets often occupy only a few pixels. Early detection of AtoN helps intelligent ships to make timely decisions, which requires the network model to have a strong ability to identify small targets. Due to the limited image size of the training set in this paper, the network model cannot meet the recognition of AtoN of various scales when it is actually used [38]. In order to solve the problem of AtoN recognition of different scales, the image expanded by DCGAN is reduced and enlarged by the image pyramid method [39]. As shown in Figure 5, the image size is changed by re-sampling so that the images of different scales contain the same content. Finally, the image pyramid is applied to training and testing. The effect of the image pyramid in detection is further analyzed.

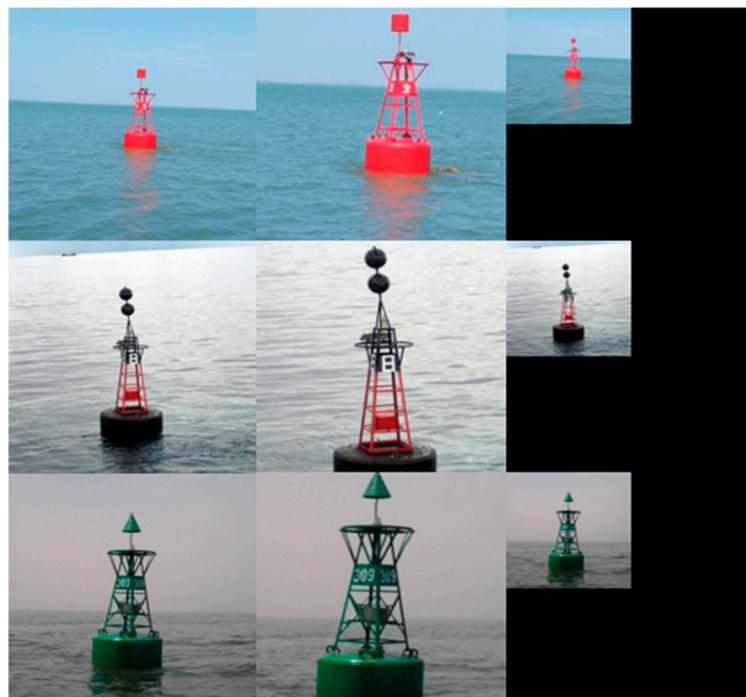


Figure 5. Image Pyramid Example.

### 3.3. Algorithm Improvement Based on k-means Clustering

#### 3.3.1. Introduction to the k-means Algorithm

The detection target in this paper is AtoN, which is not consistent with the target size of the public dataset. Therefore, it is necessary to establish a separate candidate box for the AtoN dataset [40]. The k-means algorithm is a classical clustering algorithm based on distance, which uses distance as the similarity evaluation index. That is, the closer the distance between two objects, the higher the similarity. The algorithm considers that clusters are composed of close objects, so the final goal is to obtain compact and independent clusters. When the algorithm starts running, first input an initial value K to the algorithm, which indicates how many classes into which data need to be divided. Then the algorithm will randomly select k initial points as the centroid. In the first iteration, the Euclidean distance between each point and the centroid is calculated, and the point is classified into the nearest class. Then, recalculate the centroids of the K clusters and repeat the above process until the center of the cluster does not change.

#### 3.3.2. Determination of AN-YOLOv4 Candidate Box Based on k-means

When applying the above k-means algorithm to the selection of candidate boxes, due to the different sizes of real candidate boxes, the larger real box is prone to greater error than the smaller real box during iteration 14. Therefore, it is necessary to change the distance judgment function of AN-YOLOv4 to:

$$d(box, centroid) = 1 - IoU(box, centroid) \tag{5}$$

where *box* represents the real box, *centroid* represents the cluster center, and *IoU* is the intersection union ratio. AN-YOLOv4 predicts feature maps of three scales, with three candidate boxes on each scale feature map, for a total of nine candidate boxes. Therefore, nine cluster centers need to be set, considering that in scenes at different scales, the size of each target box is different. Therefore, the width and height of the bounding box need to be normalized to the width and height of the image. After multiple iterations on the self-built dataset, the obtained clustering results are shown in Figure 6. After clustering, the optimal coordinates and ratios of nine candidate boxes are shown in Table 2. It can be seen from the table that there are huge differences between the candidate box of AtoN and other target recognition tasks. Selecting this candidate box will help the network model to fit faster and improve the accuracy of the model.

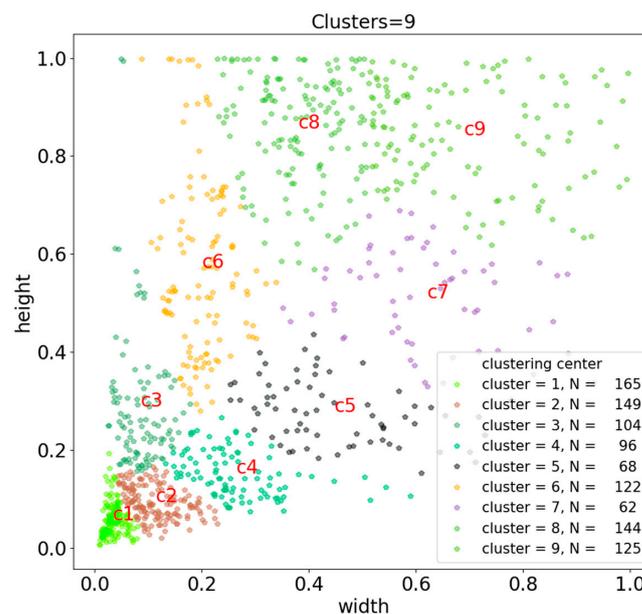


Figure 6. Selection of candidate boxes.

**Table 2.** The coordinates and scale of the candidate box of AtoN.

Number	Candidate Box Coordinates	Ration
1	17, 28	0.61:1
2	58, 47	1.23:1
3	44, 147	0.30:1
4	135, 77	1.75:1
5	229, 141	1.62:1
6	102, 292	0.35:1
7	318, 260	1.22:1
8	194, 438	0.44:1
9	352, 431	0.82:1

## 4. Evaluation

### 4.1. Data Preparation

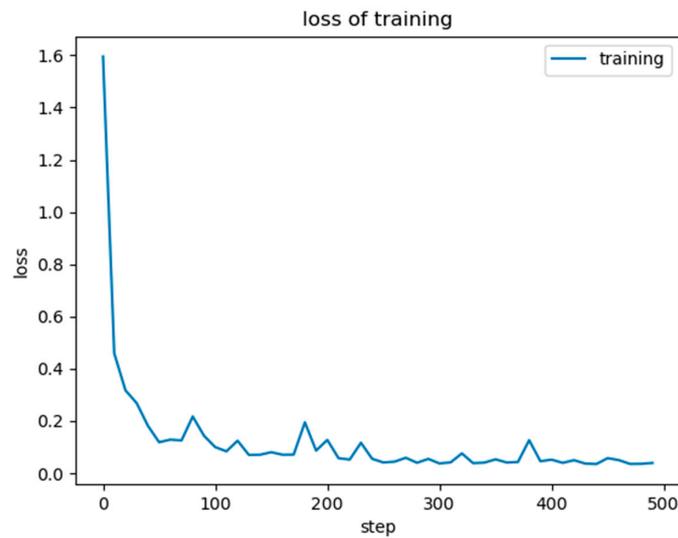
The AtoN dataset in this paper comes from the Internet, a total of 425. The training set and test set are divided, as shown in Table 3. Among them, there are 60 left lateral marks, 59 right lateral marks, 62 north side marks, 61 south side marks, 57 West Side marks, 55 East Side marks, and 71 isolated danger marks. In order to verify the effectiveness of the algorithm. Take 20 pieces of each category at random as the test set. The remaining pictures are used for training, a total of 285. After data expansion by DCGAN and multi-scale scaling of the image pyramid, the training set is expanded from 285 to 1035. The Labelling Annotation Tool is used to label the categories according to the PASCAL-VOC dataset format, and the training set of this paper is obtained.

**Table 3.** Division of training set and test set.

Types of AtoN	Original Training Set	Improving Training Set	Testing Set
left lateral marks	40	148	20
right lateral marks	39	148	20
north side marks	42	150	20
south side marks	41	149	20
west side marks	37	147	20
east side marks	35	138	20
isolated danger marks	51	155	20
total	285	1035	140

### 4.2. Experimental Environment and Parameter Configuration

The experimental simulation environment of this paper: Windows platform, CPU i7-10700f, memory 32GB, GPU processor NVIDIA<sup>®</sup> GeForce<sup>®</sup> RTX 2070 super, software environment: Python 3.7.8, pycharm2019, Anaconda 3.4.1, TensorFlow 2.3, CUDA 10.1.234, cudnn7.6.5. In this paper, 1035 pictures are divided into 207 batches. Each batch has five pictures. The SGD optimizer is used for optimization. The initial learning rate is 0.001, the attenuation coefficient is 0.0005, the random gradient drop is 0.9, and the confidence threshold is 0.3. After testing, when Epoch is set to 500 times, the algorithm loss will not decrease. The training process is shown in Figure 7. Save the model every 1 Epoch, evaluate the model performance through the test set, and select the model with the best performance in the test set for comparison.



**Figure 7.** Training loss.

#### 4.3. Analysis of Experimental Results

In this paper, the network performance is judged by four indicators: small target accuracy, category AP value, mAP value, and FPS (Frame Per Second). STA is small target accuracy. The frame rate is used to compare with other mainstream algorithms. The small target division method refers to the MS COCO dataset [41]. When the resolution is less than 32\*32, it is a small target. Considering that the recall rate of small targets is more important in practical applications, the small target accuracy is set as the percentage of all small targets found. The experimental IoU is selected as 0.5. The confidence threshold is selected as 0.5. That is, when the detection network believes that a predicted target has a probability of more than 0.5 to be the real target, it considers this target to be the predicted real target. The functions of AP and mAP are shown as follows:

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

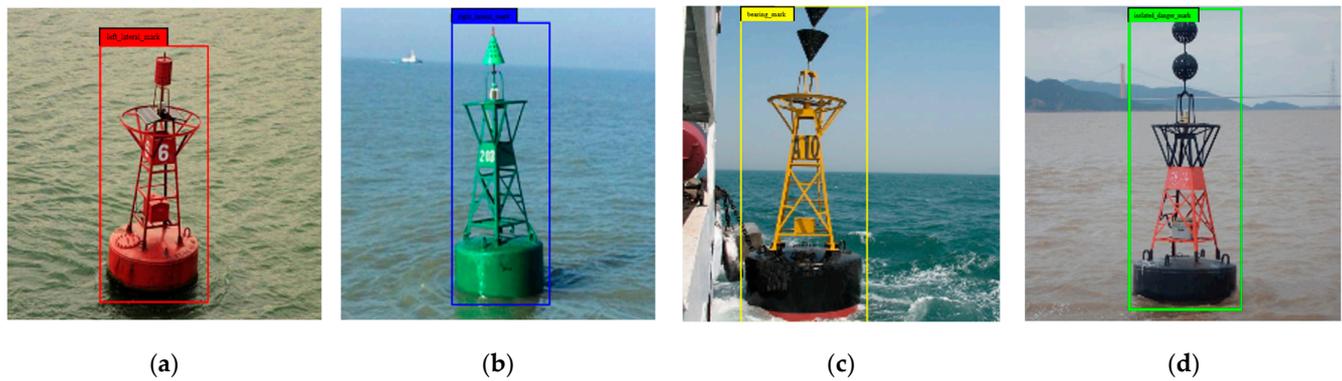
$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$AP = \int_0^1 p(r) dr \tag{8}$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \tag{9}$$

where  $TP$  represents the number of positive samples correctly identified as positive samples;  $FP$  represents the number of negative samples incorrectly identified as positive samples;  $FN$  represents the number of positive samples incorrectly identified as negative samples. The PR curve can be drawn by taking different precision and recall values, and the area under the PR curve is defined as a single category  $AP$ . The mean of all detection categories is  $mAP$ .

Part of the experimental results of AN-YOLOv4’s identification of AtoN types is shown in Figure 8. As discussed in Section 3, our ship detection framework is proposed by taking into consideration several modules, e.g., the DCGAN of the dataset(DG), the Image Pyramid Network(IPN), and the K-means function to modify the candidate box (K-m). The ablation experiments will thus be implemented to investigate which plays a more important role in improving detection performance. The numerical experiments are illustrated detailedly in Table 4.



**Figure 8.** Partial test image; (a) left lateral mark; (b) right lateral mark; (c) bearing mark (d) isolated danger mark.

**Table 4.** Model performance evaluation with different improvement methods.

Algorithm	DG	IPN	K-m	STA	AP <sub>left</sub>	AP <sub>right</sub>	AP <sub>east</sub>	AP <sub>west</sub>	AP <sub>south</sub>	AP <sub>north</sub>	AP <sub>danger</sub>	mAP
YOLOv4_0				0.78	0.88	0.90	0.77	0.78	0.75	0.83	0.60	0.79
YOLOv4_1	✓			0.82	0.95	0.93	0.80	0.82	0.76	0.88	0.71	0.84
YOLOv4_2	✓	✓		0.89	0.95	0.95	0.88	0.83	0.90	0.87	0.80	0.88
YOLOv4_3			✓	0.82	0.95	0.92	0.82	0.81	0.82	0.82	0.72	0.84
YOLOv4_4	✓		✓	0.85	0.95	0.95	0.90	0.91	0.88	0.92	0.77	0.90
AN-YOLOv4	✓	✓	✓	0.92	1	0.95	0.92	0.93	0.93	0.95	0.78	0.92

In Table 4, YOLOv4\_0 does not use any improvement method; YOLOv4\_1 uses the DCGAN method alone; YOLOv4\_2 uses the data joint improvement method; YOLOv4\_3 uses k-means alone; YOLOv4\_4 uses the combination of DCGAN and k-means; AN-YOLOv4 uses all improvements method. After testing, it was found that the mAP of YOLOv4\_1 and YOLOv4\_3 both reached 0.84, and the small target accuracy of AtoN reached 0.82, which were 0.05 and 0.04 percentage points higher than that of YOLOv4\_0 without any improvement, respectively. This proves that using DCGAN to improve the dataset and using k-means to improve the candidate box can both improve the detection accuracy of the algorithm. YOLOv4\_2 uses the data joint improvement method, and the mAP reaches 0.89, which is 0.05 better than DCGAN alone. However, the small target accuracy rate of YOLOv4\_2 has achieved a high score of 0.89, which is 7% higher than that of YOLOv4\_1, which shows that the image pyramid network is of great help in improving the accuracy of small AtoN targets. Based on all experiments, the three improvements proposed in this study can effectively improve the accuracy of network detection. The final network detection result has a small AtoN target accuracy of 0.92 and mAP of 0.92, which are 0.14 and 0.13 more than the original unimproved YOLOv4, respectively.

Although many scholars improved and tested the algorithm by migrating the feature extraction network, it is undeniable that some feature extraction structures performed better in the original algorithm [42]. Therefore, mainstream algorithms using different feature extraction networks are selected to test and compare with AN-YOLOv4 on the self-built Aids to Navigation dataset. These algorithms include YOLOv4, YOLOv4 tiny, Faster R-CNN, and Mask RCNN. Part of the test results is shown in Figure 9. The four scenes shown in the figure are complex background, fog, low brightness, and small target detection. In the complex background detection scene, YOLOv4-tiny detects the wrong target. In low brightness and small target detection scenarios, YOLOv4-tiny, YOLOv4, and Fast-RCNN have different degrees of missed detection, and Mask-RCNN has repeatedly detected the navigation target. In general, the proposed algorithm NM-YOLOv4 can spend less time on accurate detection in different scenarios.



**Figure 9.** (a) YOLOv4-tiny; (b) YOLOv4; (c) Mask-RCNN; (d) Faster-RCNN; (e) NM-YOLOv4(ours). In four different scenarios (complex background, foggy, low brightness, and small target), test results of algorithms YOLOv4-tiny, YOLOv4, Faster-RCNN Mask-RCNN, and NM-YOLOv4(ours).

The mAP value and PR curve of this algorithm and other algorithms for the detection of AtoN are shown in Table 5 and Figure 10, respectively. Table 5 lists the test of different algorithms on AIDS datasets. Through comparative experiments, it is found that the AN-YOLOv4 algorithm surpasses other algorithms in various accuracy indicators. Among them, YOLOv4-tiny runs faster than the algorithm in this paper, but its accuracy and mAP are too low to be suitable for intelligent ship tasks with high accuracy requirements. Mask-RCNN, as a multi-stage algorithm, has high accuracy and mAP, but its model size is too large, and FPS is low, so it is not suitable for real-time detection tasks. The improved AtoN detection algorithm AN-YOLOv4 based on YOLOv4 proposed in this paper has a

higher detection rate and accuracy of small targets and runs faster. Figure 10a shows the PR curve of different algorithms for a small target of AtoN. It can be seen from the PR curve that other algorithms have a high false positive rate for small targets. This is because other algorithms are easy to identify noise as small AtoN. Figure 10b–i shows the PR curves of different algorithms for different AtoN targets. It can be seen that the algorithm in this paper is superior to other algorithms in terms of accuracy and recall rate. Overall, the AN-YOLOv4 algorithm is superior to the current mainstream target detection algorithms.

Table 5. Comparison of algorithms superiority.

Algorithm	STA	APleft	APright	APeast	APwest	APsouth	APnorth	APdanger	mAP	FPS
YOLOv4	0.78	0.88	0.9	0.77	0.78	0.75	0.83	0.6	0.79	30.8
YOLOv4-tiny	0.54	0.7	0.67	0.6	0.61	0.6	0.59	0.6	0.62	62
Faster-RCNN	0.82	0.86	0.85	0.8	0.72	0.76	0.75	0.71	0.78	12.0
Mask-RCNN	0.89	0.93	0.95	0.87	0.82	0.88	0.87	0.8	0.87	9.5
AN-YOLOv4	0.92	1	0.95	0.92	0.93	0.93	0.95	0.78	0.92	30.8

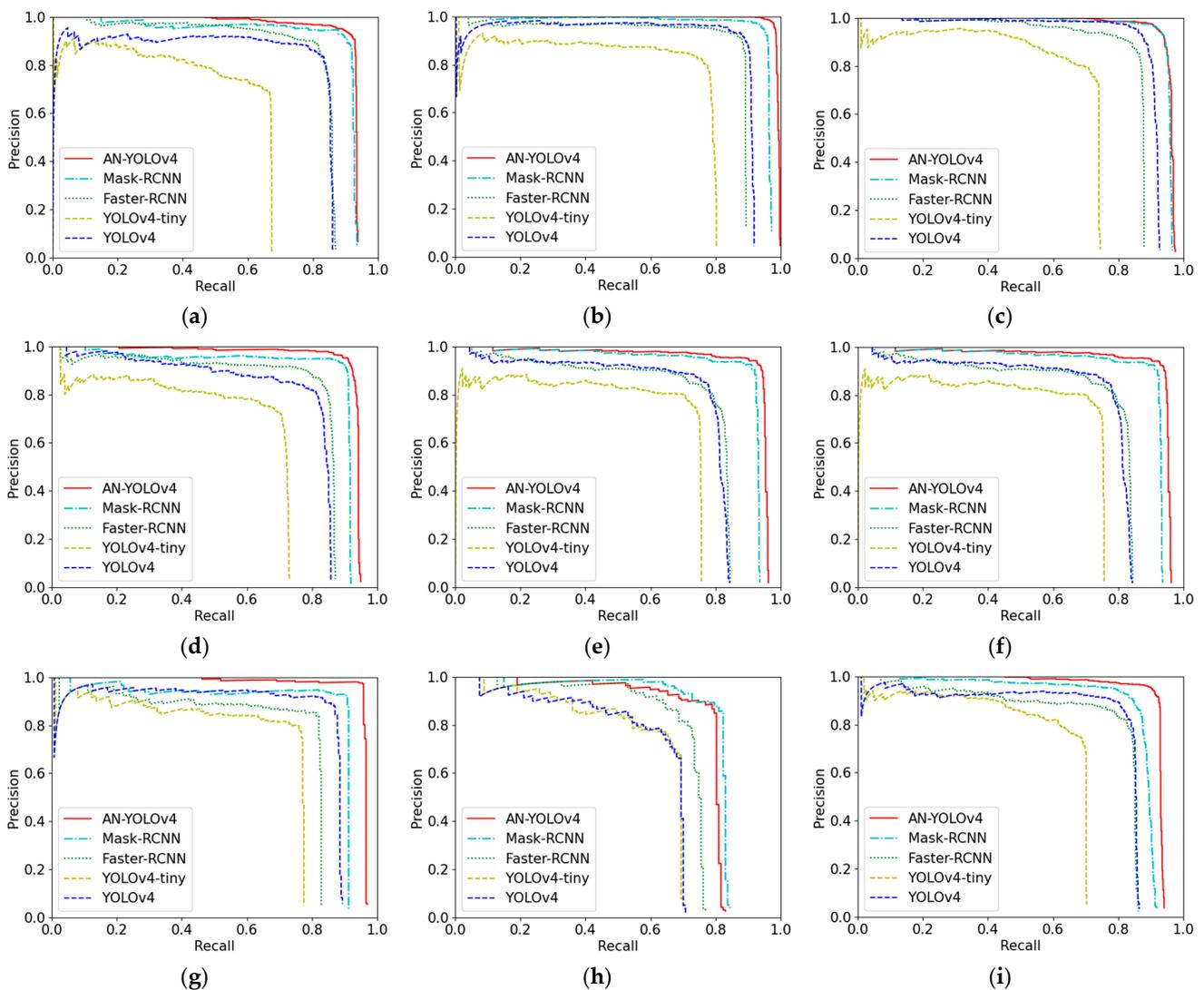


Figure 10. PR curve under five different algorithms. (a) small target accuracy; (b) left lateral marks; (c) right lateral marks; (d) east side marks; (e) west side marks; (f) south side marks; (g) north side marks; (h) isolated danger marks; (i) total.

## 5. Conclusions

This paper proposes an AtoN target detection algorithm AN-YOLOv4 based on improved YOLOv4. The DCGAN is used to enrich the AtoN dataset, and the image pyramid network is used to divide the AtoN data at multiple scales so that the number of the original dataset is expanded by more than three times. Considering that the candidate box used in the existing public dataset is not suitable for AtoN, K-means is used to improve the size of the candidate box. The experimental results show that the expansion of the AtoN dataset and the improvement of candidate box size makes the AN-YOLOv4 improve the accuracy of AtoN recognition and still maintain a high detection speed. The AN-YOLOv4 method optimizes and improves the algorithm on the framework of YOLOv4. Under the accuracy that meets the requirements of actual AtoN target recognition applications, the method in this paper has a simpler structure, consumes less computing resources, and has a faster detection speed. This method can be used in a smart ship for environment perception and monitoring systems to ensure the accurate position of the AtoN. At the same time, this method can identify the type of AtoN target and assist the ship-shore cooperative system in judging the channel environment.

In the future, smart ships can be equipped with AtoN monitoring terminals for real-time detection of the channel environment. Therefore, the AtoN target detection method based on AN-YOLOv4 has strong applicability and scalability. In addition, the algorithm only detects targets for AtoN. On the premise of ensuring accuracy and detection speed, how to detect ships, islands, and other sea targets and integrate information with other sensing algorithms to achieve an intelligent perception of the navigation environment is the follow-up research direction.

**Author Contributions:** Conceptualization, R.Z. and Y.Y.; Data curation, Y.Y. and X.C.; Investigation, R.Z., Y.Y. and X.C.; Writing—original draft, R.Z. and Y.Y.; Methodology, R.Z., Y.Y. and L.X.; Writing—review and editing, R.Z., Y.Y. and L.X.; Supervision, X.C. and L.X.; Visualization, Y.Y. and L.X. Funding acquisition, R.Z. and L.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research described in this paper is supported by the National Natural Science Foundation of China (No. 52001134) and The Educational Research Project of Young and Middle-Aged Teachers in Fujian Province (No. JAT200285).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We are grateful to the anonymous reviewers whose comments and suggestions have contributed to improving the quality of research described in this paper.

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

## References

1. Li, S.B.; Huang, Y.J.; Duan, J.; Huang, D.W. Research on key technologies of visualization of navigation safety information. *Hydrographic Surveyi.* **2020**, *40*, 73–77.
2. Zhu, B.; Wen, S.; Sun, H.F.; Wu, X.K. Application of nautical safety class notation on VLOC. *Ship Eng.* **2020**, *42*, 110–112+315.
3. Chen, X.Q.; Li, Z.B.; Yang, Y.S.; Qi, L.; Ke, R. High-Resolution vehicle trajectory extraction and denoising from aerial videos. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 3190–3202. [[CrossRef](#)]
4. Ai, W.Z.; Ding, T.M. Research on fairway layout in bridge waters. *J. Transp. Syst. Eng. Inf. Technol.* **2014**, *14*, 131–137.
5. Chen, X.Q.; Wang, S.Z.; Shi, C.J.; Wu, H.F.; Zhao, J.S.; Fu, J.J. Robust ship tracking via Multiview learning and sparse representation. *J. Navig.* **2019**, *72*, 176–192. [[CrossRef](#)]
6. Chen, X.Q.; Ling, J.; Wang, S.Z.; Yang, Y.S.; Luo, L.J.; Yan, Y. Ship detection from coastal surveillance videos via an ensemble Canny-Gaussian-morphology framework. *J. Navig.* **2021**, *74*, 1252–1266. [[CrossRef](#)]
7. Qiao, D.L.; Liu, G.Z.; Lv, T.Z.; Li, W.; Zhang, J. Marine vision-based situational awareness using discriminative deep learning: A survey. *J. Mar. Sci. Eng.* **2021**, *9*, 397. [[CrossRef](#)]

8. Cheng, G.; Han, J.W. A survey on object detection in optical remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 11–28. [[CrossRef](#)]
9. Zheng, W.; Tang, W.L.; Jiang, L.; Fu, C.W. SE-SSD: Self-Ensembling single-stage object detector from point cloud. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 14489–14498.
10. Zheng, L.Y.; Tang, M.; Chen, Y.Y.; Zhu, G.B.; Wang, J.Q.; Lu, H.Q. Improving multiple object tracking with single object tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–24 June 2021; pp. 2453–2462.
11. Feng, C.J.; Zhong, Y.J.; Gao, Y.; Scott, M.R.; Huang, W.L. TOOD: Task-aligned one-stage object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 3490–3499.
12. Xuan, S.Y.; Zhang, S.L. Intra-Inter camera similarity for unsupervised person re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 11921–11930.
13. Cengil, E.; Cinar, A. Poisonous Mushroom Detection using YOLOV5. *Turk. J. Sci.* **2021**, *16*, 119–127.
14. Redmon, J.; Diwala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
15. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Lecture Notes in Computer Science: Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer: Cham, Switzerland, 2016; Volume 9905, pp. 21–37.
16. Chang, Y.-L.; Anagaw, A.; Chang, L.; Wang, Y.C.; Hsiao, C.-Y.; Lee, W.-H. Ship detection based on YOLOv2 for SAR imagery. *Remote Sens.* **2019**, *11*, 786. [[CrossRef](#)]
17. Li, H.; Deng, L.B.; Yang, C.; Liu, J.B.; Gu, Z.Q. Enhanced YOLO v3 tiny network for real-time ship detection from visual image. *IEEE Access* **2021**, *9*, 16692–16706. [[CrossRef](#)]
18. Tang, G.; Liu, S.B.; Fujino, I.; Claramunt, C.; Wang, Y.D.; Men, S.Y. H-YOLO: A single-shot ship detection approach based on region of interest preselected network. *Remote Sens.* **2020**, *12*, 4192. [[CrossRef](#)]
19. Ren, S.Q.; He, K.M.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 7–12 December 2015; pp. 91–99.
20. You, Y.N.; Li, Z.Z.; Ran, B.H.; Cao, J.Y.; Lv, S.D.; Liu, F. Broad area target search system for ship detection via deep convolutional neural network. *Remote Sens.* **2019**, *11*, 1965. [[CrossRef](#)]
21. Gao, L.R.; He, Y.Q.; Sun, X.; Jia, X.P.; Zhang, B. Incorporating negative sample training for ship detection based on deep learning. *Sensors* **2019**, *19*, 684. [[CrossRef](#)] [[PubMed](#)]
22. Zhang, S.M.; Wu, R.Z.; Xu, K.Y.; Wang, J.M.; Sun, W.W. R-CNN-Based ship detection from high resolution remote sensing imagery. *Remote Sens.* **2019**, *11*, 631. [[CrossRef](#)]
23. Li, L.Y.; Wu, D.F.; Wu, Z.M.; Yang, R.F. Fast maritime target detection method based on deep learning. *Ship Eng.* **2020**, *42*, 94–99.
24. Liu, R.W.; Yuan, W.Q.; Chen, X.Q.; Lu, Y.X. An enhanced CNN-enabled learning method for promoting ship detection in maritime surveillance system. *Ocean Eng.* **2021**, *235*, 109435. [[CrossRef](#)]
25. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
26. Wang, C.-Y.; Liao, H.-Y.M.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W.; Yeh, I.-H. CSPNet: A new backbone that can enhance learning capability of cnn. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 1571–1580.
27. Mishra, D. Mish: A self regularized nonmonotonic neural activation function. *arXiv* **2019**, arXiv:1908.08681.
28. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In *Lecture Notes in Computer Science: Computer Vision—Eccv 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer: Cham, Switzerland, 2014; Volume 8691, pp. 346–361.
29. Liu, S.; Qi, L.; Qin, H.F.; Shi, J.P.; Jia, J.Y. Path aggregation network for instance segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
30. Lin, T.-Y.; Dollar, P.; Girshick, R.; He, K.M.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
31. Yu, J.H.; Jiang, Y.N.; Wang, Z.Y.; Cao, Z.M.; Huang, T. UnitBox: An advanced object detection network. In Proceedings of the 24th ACM international conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016; pp. 516–520.
32. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
33. Zheng, Z.H.; Wang, P.; Liu, W.; Li, J.Z.; Ye, R.G.; Ren, D.W. Distance-IoU Loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), New York, NY, USA, 7–12 February 2020; pp. 12993–13000.
34. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. Acn.* **2020**, *63*, 139–144. [[CrossRef](#)]

35. Yu, X.M.; Hong, S.; Yu, J.X.; Lu, Y.B.; Peng, Y. Research on a ship target data augmentation method of visible remote sensing image. *Chin. J. Sci. Instrum.* **2020**, *41*, 261–269.
36. Liu, W.; Yang, M.F.; Nie, J.T.; Zhang, Y.; Yang, H.L.; Xiong, Z.H. Low-Light maritime image enhancement based on local generative adversarial network. *Comput. Eng.* **2021**, *47*, 16–23.
37. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
38. Guo, W. Automatic Ship Detection in Optical Remote Sensing Images Based on Deep Learning. Master's Thesis, Wuhan University, Wuhan, China, 2019.
39. Zheng, J. The Object Detection Method for Pedestrian Video Based on YOLOv3. Master's Thesis, Xidian University, Xi'an, China, 2019.
40. Nie, X.; Liu, W.; Wu, W. Ship detection based on enhanced YOLOv3 under complex environments. *J. Comput. Appl.* **2020**, *40*, 2561–2570.
41. Lin, T.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollar, P.; Zitnick, C. Microsoft COCO: Common objects in context. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, 5–12 September 2014; Volume 8693, pp. 740–755.
42. Khasawneh, N.; Fraiwan, M.; Fraiwan, L. Detection of K-complexes in EEG signals using deep transfer learning and YOLOv3. *Clust. Comput.* **2022**, 1–11. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.