

Article Detection of Floating Garbage on Water Surface Based on PC-Net

Ning Li^{1,2}, He Huang¹, Xueyuan Wang¹, Baohua Yuan¹, Yi Liu¹ and Shoukun Xu^{1,*}

- School of Computer Science and AI Aliyun School of Big Data & School of Software, Changzhou University, Changzhou 213164, China
- ² School of Computer and Information Engineering, Hohai University, Nanjing 210098, China

Correspondence: xsk@cczu.edu.cn

Abstract: In the detection of surface floating garbage, the existence of complex backgrounds and the small target sizes make the surface floating garbage easy to mis-detect. Existing approaches cannot yet provide a solution to the aforementioned problems and they are typically limited to addressing specific issues. This paper proposes a PC-Net algorithm for floating garbage detection. First, a pyramid anchor generation approach is proposed, which makes the anchor to be generated centrally near the target and reduces the interference of background information in the anchor generation. Then, in the RoI Pooling feature map import stage, the classification map is used as the feature map. This approach generates feature maps with a higher resolution and more distinct features, thereby enhancing the feature information of small targets and enhancing the classification accuracy. Experimental results on floating garbage dataset indicate that the average detection accuracy of the proposed approach is 86.4%. Compared with existing detection approaches, such as Faster R-CNN, YOLOV3, YOLOX, and Dynamic R-CNN, the average accuracy of detection is increased by 4.1%, 3.6%, and 2.8%, respectively.

Keywords: object detection; anchor mechanism; classification discrimination diagram; floating garbage

1. Introduction

At present, there is much floating garbage in rivers, lakes and sea surfaces [1]. If this garbage can be recycled, it will improve the ecological environment and provide economic benefits [2,3].

However, the manual salvaging of floating garbage on water surfaces is inefficient and costly. Therefore, mechanized salvage will be the future trend, and research into floating garbage detection algorithms on the water surface will promote mechanized salvage over manual salvage. On the basis of its processing, the floating garbage detection algorithm for deep learning on water surfaces can be divided into one-phase and two-phase processes. Common first-stage algorithms include YOLOv2, YOLOv3, YOLOv4 [4], SSD, etc., and common second-stage algorithms include R-CNN, etc. Zhang et al. [5] proposed a network model using a combination of low-level and high-level features that has superior real-time performance for floating garbage detection on the water surface. Lin et al. [6] introduced Soft-NMS based on the YOLOX algorithm model to improve occlusion target detection. Wang et al. [7] proposed a lightweight Ylov4 target detection network based on an efficient Net-B0 fusion ECA mechanism. This method improves the speed of model detection by reducing the parameters of the network model. Verma V et al. [8] studied the use of symmetry during garbage image sampling. Because symmetry is applied to extract its features, image resizing is uniform. Ma et al. [9] proposed an enhanced single-lens multibox detector (SSD) with a lightweight and novel feature fusion module. Deng H et al. [10] introduced the idea of expansion convolution into the feature pyramid network to enhance the feature extraction ability of small objects. Secondly, the spatial channel attention mechanism is used to make features learn adaptively. Zheng et al. [11] proposed inland river ship recognition based on binocular stereo vision (BSV), and taking into account



Citation: Li, N.; Huang, H.; Wang, X.; Yuan, B.; Liu, Y.; Xu, S. Detection of Floating Garbage on Water Surface Based on PC-Net. *Sustainability* **2022**, *14*, 11729. https://doi.org/10.3390/ su141811729

Academic Editor: Hossein Bonakdari

Received: 28 July 2022 Accepted: 15 September 2022 Published: 19 September 2022

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



Copyright: © 2022 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/). the computational pressure caused by the huge network parameters of the classic YOLOv4 model, the MobileNetV1 network was used as the feature extraction module of YOLOv4 model. Dan Zeng et al. [12] used the unsupervised area proposal generation algorithm to selectively search and non-maximum suppression (NMS) to extract the location and size of garbage areas. Li et al. [13] in During training, the anchor boxes were re-clustered to replace the inappropriate anchor boxes. Li et al. [14] proposed a SAR ship feature enhancement method based on high-frequency sub-band channel fusion, which makes full use of the contour information, aiming at the speckle noise and ship contour ambiguity caused by SAR special imaging mechanism. Cheng et al. [15] proposed a saliency enhancement algorithm based on the difference of anisotropic pyramid (DoAP). Considering the limitation of IoU in small target detection, we design a detection framework based on bhattacharyalike distance (BLD). On the basis of the improved RefineDet [16] model, Zhang et al. proposed a real-time detection approach for floating objects on the water surface. This approach exhibits a good real-time performance, but the detection accuracy, particularly for relatively small floating objects on the water surface, needs improvement. Zhang et al. [17] introduced the structure of recursive feature pyramid (RFP) and deformable convolution network (DCN) into the learning framework in order to optimize the basic backbone of the network and construct a feature map with high-level semantics and lowlevel positioning information of the network. Based on YOLOv5S deep learning, Du et al. [18] adopted the BiFPN network structure to enhance the feature extraction ability of the original PANet network for unsafe objects in transmission line images. Zeng et al. [12] proposed a garbage detection method for airborne hyperspectral data based on multiscale CNN. Zhang et al. [19] introduced multibranch expansion convolution to enhance the characteristic information of small targets and replaced cascade RCNN with a multilayer deformable convolution network to improve the speed of the network model. Tian et al. [20] converted YOLOv4 into a four-scale detection method. In order to improve the detection speed, the new model was pruned.

Wen et al. [21] proposed a multiframe detection method of small targets on the sea surface based on a deep convolution neural network. Zhou et al. [22] proposed an improved YOLO-SASE detection algorithm, which combines a SASE module, SPP module and multilevel receptive field structure. GU et al. [23] proposed a small target detection method for ocean surveillance radar based on multifeature and principal component analysis. Gao et al. [24] proposed a high-precision detection algorithm based on feature mapping depth neural network of spindle network structure for dim targets with few pixel features in complex and diverse backgrounds. Jia et al. [25] added a center loss function on the basis of SSD (Single Shot MultiBox Detector) network to better deal with the problem that the intra-class difference is greater than the inter-class difference. This solves the problem of insufficient sensitivity to small objects. On the basis of the SSD network, Liu et al. [26] used residual network as the basic network of single lens multibox detector (SSD) target detection network model. Sha et al. [27] used ResNet50 instead of VGG16 in a fast-RCNN backbone network to increase the training depth of the network, and used soft-NMS instead of NMS, and modified the classifier layer of fast-RCNN. Sharma et al. [28], based on the fast-RCNN network model, introduced saliency detection to better detect and identify targets. On the basis of the Mask RCNN network model, Huang et al. [29] used ResNet as a feature extraction network, and effectively combined the feature pyramid network (FPN), ROIAlign and full convolution network (FCN) and other modules. Li et al. [30] used a high-resolution network (HRNet) as the backbone network of image feature extraction, and adopted focus loss as the classification loss and region of interest alignment instead of region of interest merging. Shi et al. [31] introduced a cascade strategy and adaptive threshold strategy, and proposed a domain-based adaptive fast RCNN method. Li et al. [32] put forward GPR-RCNN based on RCNN network model, which can robustly detect defective areas even in the presence of obvious noise. On the basis of Faster RCNN, Zhao et al. [33] improved the fast regional convolution neural network model by using the characteristic pyramid network (FPNs) to realize the insulator location in complex background images. Yu et al. [34], based on the Mask RCNN network model, took Resnet101 as the backbone network and adopted the feature pyramid network (FPN) structure for feature extraction. Han et al. [35] proposed a real-time small traffic sign detection method based on improved Faster RCNN. Xie et al. [36] integrated a deconvolution module on the basis of the Faster RCNN network model to provide additional context information, which is helpful to improve the detection accuracy of small-scale pedestrian examples. On the basis of the Faster RCNN network model, Sun et al. [37] combined various strategies to improve it, including feature concatenation, hard negative mining and multiscale training. On the basis of the YOLOv3 network model, Wang et al. [38] proposed AS-CBAM (Adaptive Selection Convolution Block Attention Module) and innovatively combined with HDC (Hybrid Extended Convolution) to maximize the receptive field and fine-tune the features, aiming to solve the problem that the original CBAM maximum pool operation can be easily used to introduce background noise. Nevertheless, some problems with complex contexts and too small a proportion of target size cannot be effectively resolved.

This paper focuses on two contributions and proposes a PC-Net-based algorithm for floating garbage detection on water surfaces. In the detection of floating garbage on the water surface, the presence of a complex background and a large aspect ratio gap between targets (such as branches and bottles) presents the first challenge.

A solution for the generation of pyramidal anchors is proposed based on the concept of the Faster R-CNN network. The fundamental concept is to generate an anchor centered on the target and to adjust the anchor's parameter settings based on the size and aspect ratio distribution of the target in order to better match the target. This strategy effectively reduces complex background interference, improves the overlap between the positive sample and the target, and enhances the performance of the anchor mechanism. Second, to address the issue that the target in the floating garbage on the water surface has an insufficient size share in the feature map and an unbalanced foreground and background share, which can easily result in the loss of feature information during the classification stage, thereby severely affecting the classification accuracy. The purpose of the proposed new classification discrimination map is to import a classification discrimination map with a higher resolution during the RoI Pooling feature map import stage in order to improve the accuracy of floating garbage detection on water surfaces.

2. PC-Net

In this paper, we employ the idea of a Faster R-CNN network, and Figure 1 depicts the model structure. First, ResNet-50 is used as the feature extraction network in this paper [39]. Second, a pyramidal anchor generation approach is employed during the anchor generation phase. Third, during the RoI pooling feature map import stage, in order to provide clearer features for the upcoming classification operation, the classification discrimination map is lead into the RoI pooling stage. The details of these three strategies are given in the following sections.

2.1. Feature Extraction Network

ResNet-50 is used as the feature extraction network in this paper. Figure 2 displays the network structure of ResNet-50, which includes residual learning and applies the concept of directly connected channels to the feature extraction network. Before ResNet-50, the feature extraction network transforms each layer nonlinearly before proceeding to the next layer of operation. The directly connected channel concept, on the other hand, permits a specific percentage of information from the previous levels to be transmitted to the subsequent layers; i.e., a jump connection is accomplished. Prior to ResNet-50, feature extraction networks suffer from issues such as the loss of feature information when extracting features and, in some circumstances, gradient disappearance and explosion, making it impossible to train deeper-layer feature networks. Additionally, ResNet-50 effectively overcomes the aforementioned issue by transferring the original data to the output and thus indirectly addressing the issue of feature information loss.



Figure 1. PC-Net model structure.



Figure 2. ResNet-50 network structure.

2.2. Pyramidal Anchor Generation Approach

In floating garbage identification on the water surface, the complicated water surface background results in inaccurate target detection. The fundamental problem is that ripples on the water surface, uneven illumination, and other various interfere with the extraction of positive and negative sample feature information, resulting in significant discrepancies between the retrieved feature information and the target. Based on this, this research proposes an approach for generating pyramidal anchors. This approach employs the semantic knowledge of image features to guide the construction of anchors. In other words, we jointly estimate the probable locations and shapes of the target centers, build anchors with associated locations, grades, and forms, and then forecast based on these anchors.

2.2.1. Center Area Selection

As indicated in Figure 3, the central area was selected. Using the label information, we first extracted the coordinate value of the center point, the length, and the width of the object to be detected. This information was then utilized for center area selection.



Figure 3. Schematic diagram of determining the center area of anchor frame.

The implementation process can be divided into the following three steps.

Step 1: The coordinate information of the labeled box is used to build a binarized label map for each image, where the portion of the labeled box containing the target is coded as 1 and the remainder is coded as 0; i.e., the foreground portion is labeled as 1 and the background portion is labeled as 0.

Step 2: Apply the coordinate location information of the designated box to various feature map scales yields the coordinate position information (x'_g, y'_g, w'_g, h'_g) , where x_g is the horizontal coordinate of the box's center and y_g is the vertical coordinate of the box's center. The box is then divided into three categories: ignore area, center area, and outer area.

The center area (CA) = $(x'_g, y'_g, \sigma_1 w'_g, \sigma_1 h'_g)$ defines the central region of the annotation box, i.e., the yellow portion of the region in the image above, which is the most central portion of the annotation box and the anchor box constructed with this portion as its center corresponds to the positive sample.

 $(x'_g, y'_g, \sigma_2 w'_g, \sigma_2 h'_g)$ is the ignore area (IA). IA is a broader region, represented in the picture above by the green area. If the anchor point center is created in this section, its IoU is rather low. Hence, this section is disregarded and used as a buffer.

The outer area (OA) is where CA and IA are removed in the whole feature map. If its center is in this part, the anchor is a negative sample.

Step 3: The above is mapped to other feature maps; i.e., this method realizes selects the central area of all feature maps.

2.2.2. Anchor Grade Classification

In the detection process of floating garbage on the surface of the water, there is an issue with the aspect ratio and size of the target, such as branches, plastic bags, and bottles, etc. Therefore, in this paper, the anchor grade approach is adopted to generate an anchor with corresponding sizes according to the different sizes of targets to be detected. As shown in Figure 4, it is a schematic diagram of anchor grade. The square is the set length and width, and the rectangular length and width are, respectively, enlarged by two times and reduced by one half. Refer to Formula (1) for the specific setting method.



Figure 4. Diagram of anchor grade.

In addition, as stated in Table 1, each level generates three sizes of anchor boxes. Each anchor box size is generated in three proportions: 1:1, 1:2, and 2:1. Formula (1) for grade classification is as follows:

$$l = \left\lceil 10 \times \left(e^{\frac{\min(w_g, h_g)}{\varepsilon}} - 1 \right) \right\rceil \tag{1}$$

where $\min(w_g, h_g)$ represents the smallest width w_g and height h_g values of the target g. ε is the hyperparameter, while *l* represents the anchor generation level. When *l* is equal to or greater than 3, the third rank is also chosen.

Table 1. Anchor grade.

Grade	Anchor Size
1	32 imes32, $64 imes64$, $128 imes128$
2	64 imes 64, $128 imes 128$, $256 imes 256$
3	128 $ imes$ 128, 256 $ imes$ 256, 512 $ imes$ 512

The anchors generated in this manner, regardless of whether they are positive or negative samples, are still mostly clustered around the target. So, the strategy efficiently eliminates background interference. In addition, anchors of a particular size and aspect ratio are formed based on the size of the target to be identified, so the approach effectively implements the anchor mechanism.

2.3. Classification Discrimination Diagram

In the detection of floating garbage on the surface of the water, there are many problems in the data set used in this paper; for example, the size of floating garbage is too small, and the information features retrieved during the feature extraction step are insufficiently distinct, hence reducing the accuracy of the categorization of floating garbage. This paper proposes a classification discrimination map to aid the classification of floating garbage on the water surface.

First of all, during the conventional processing of the classification operation in RoI Pooling, generating about 300 suggestion boxes after the feature map is processed by RPN. After that, these suggestion boxes are pooled and become the characteristic diagram of size 7×7 . Finally, they are identified and classified.

The structure of the RPN network is shown in Figure 5. Its main function is to generate an anchor based on the feature graph, and screen out candidate anchors that may contain targets from the generated anchor.



Figure 5. RPN network structure.

As a result, for a normal-sized target, as depicted in Figure 6a, when d is more than 102, the size of its labeled box is approximately the same as the feature network processing, and the subsequent operation will not result in the loss of feature data. However, if the designated box is too small for small targets, as depicted in Figure 6b, when m is less than 64, then after the feature network processing, its m1 value will be less than 4. Because 4×4 is smaller than 7×7 , an interpolation-based modification (bilinear interpolation in this case) is required for the feature map. Obviously, the semantic content of the image will be lost after such processing, and the original feature information will not be sufficiently clear, which will result in a less robust model after training, leading to misdetection and a reduction in detection accuracy. Therefore, when a target is to be detected but its length and width are less than a certain threshold, misdetection will occur. In addition, an evaluation of the dataset revealed that there are more small targets than large ones (the definition of small targets for the dataset in this paper is described in detail in Section 3.1).



Figure 6. Schematic diagram of RPN feature processing. (a) Schematic diagram of RPN feature processing without interpolation; (b) Schematic diagram of RPN feature processing requiring interpolation.

This paper proposes a classification discrimination diagram for the above problems (as shown in Figure 7). In the first step of the specific implementation procedure, Conv4 executes the upsampling operation, utilizing nearest-neighbor interpolation. In the second phase, Conv3 is processed by 1×1 convolution kernel to adjust the number of channels, so that the number of channels is the same as that of Conv4 convolution layer. In the third phase, the concat operation is executed to complete the feature fusion of Conv3 and Conv4 in order to generate the classification discrimination map. In the fourth phase, the suggestion box generated by RPN processing is expanded and mapped to the classification discrimination diagram in the same proportion. In the fifth phase, the classification discrimination diagram is imported into RoI Pooling stage for subsequent classification operation.



Figure 7. Classification discrimination diagram.

The modification is intended to increase the resolution of the discrimination map. The reason of this modification is based on the assumptions that the object detection network requires less information to establish if a proposal is a foreground object compared to identifying the type of object in the proposal region. Therefore, when the proposal is a foreground object, better results can be achieved by evaluating the class of objects with higher-resolution features.

3. Experimental Analysis

3.1. Experimental Environment and Algorithm Evaluation Metrics

ResNet-50 is used as the feature extraction network in this paper. ResNet-50 features deeper network layers than VGG16, ZFNet, and LeNet, and incorporates a residual module to reduce information loss and extract more comprehensive semantic feature information. As shown in Table 2, it is the environment configuration used in this experiment.

Category	Environment
Graphics Card	GeForce GTX 1080Ti
Memory	12 GB
Operating System	Ubuntu16.04
Deep Learning Framework	PyTorch 1.7.0
CUDA Version	CUDA 10.1
Scripting language	Python 3.7

Table 2. Illustrates the experimental environment.

The evaluation whether the approach of this paper is effective for detecting floating garbage on the water surface. In this experiment, recall and accuracy were used as evaluation metrics, and the formulas for calculating recall and accuracy are presented in Formulas (2) and (3).

$$recall = \frac{TP}{TP + FN}$$
(2)

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$
(3)

TP (True Positive) is a positive sample with a positive prediction in the above formula. *FP* (False Positive) is a positive sample with a negative forecast. *FN* (False Negative) is a positive sample with a negative prediction, whereas *TN* (True Negative) is a negative sample with a positive forecast.

According to the definition provided by the international organization SPIE, a small target is one whose image resolution is less than 32×32 . The size of the dataset utilized in this paper is 416×416 ; hence, a target whose size is less than 208 pixels is a small target. Based on this criterion measurement, there are 643 small targets in this paper's dataset.

3.2. Dataset

Since there is no public dataset for studies on the identification of floating litter on water surfaces. In total, 2400 sheets of data from a subset of the VOC2012 dataset, an online collection, and field photography are utilized for this investigation. Later, it was expanded to 9600 images by data augmentation-related techniques, containing pictures of different types of floating garbage in various water scenes. According to the various sorts of floating garbage, there are eight test categories: bottles, grass, branches, plastic bags, milk cartons, balls, plastic garbage, and leaves. Figure 8 depicts a portion of manual annotation of several types of floating garbage.



(g)leaf

(h)ball

Figure 8. Sample dataset image of different target categories.

3.3. Ablation Experiments

This experiment compares the performances of several anchor generation approaches based on Faster R-CNN for detecting floating garbage on water surfaces. As demonstrated in Table 3, the average accuracy increased by 3.4% with the production of pyramidal

Small Target Detection Number Model Backbone Recall/% Accuracy/% Quantity 1 Baseline ResNet-50 81.1 82.3 187 2 199 Pyramid anchor frame ResNet-50 84.9 85.7 Classification 3 ResNet-50 81.3 83.5 386 discrimination diagram 4 85.8 392 PC-Net ResNet-50 86.4

anchors, from 82.3% to 85.7%. The pyramidal anchor generation approach of the Faster R-CNN framework may effectively optimize the model, as is evident.

Table 3.	Ablation	experiments.	

A 1. 1 . . .

This experiment also compares the performance on the job of detecting floating garbage on the surface of the water after the classification discrimination map is included. According to Table 3, the accuracy increases by 1.2% when only the categorical discrimination map is introduced and increased by 0.7% when it is introduced after the pyramidal anchor has been built. After adopting the classification discrimination map within the Faster R-CNN framework, it is clear that the model may be effectively optimized.

Pyramidal anchor generation is used to reduce the disturbance of anchors caused by complex background information, such as water surface fluctuations and light reflections during the generation stage, which affects the quality of positive and negative samples, and then affects the subsequent classification and recognition. In this study, the classification of anchor classes is eventually accomplished using discrete terms, and the hyperparameter ε is employed to define the pyramidal anchor class differentiation. In this study, the experimental analysis of ε values is conducted, with the experimental results of the hyperparameter ε displayed in Table 4. This research focuses on values within the interval (500, 2000). As shown in Table 4, the best detection and accuracy are achieved when $\varepsilon = 1000$, followed by $\varepsilon = 1100$, and the worst result is obtained when $\varepsilon = 500$. It can be seen that choosing appropriate coefficients is very important for the anchor box generation stage.

Table 4. Accuracy with different ε values.

	500	700	900	1000	1100	1300	1500	2000
Accuracy/%	83.15	84.38	85.94	86.40	86.34	86.11	85.82	85.33

ACC is typically used to evaluate the classifier's classification performance. The ACC value represents the proportion of correctly identified samples among the target samples to be evaluated relative to the total number of target samples. Figure 9 depicts the ACC curves for the four approaches in Table 3, where the horizontal axis represents the number of iterations and the vertical axis represents the ACC value. It is evident from Figure 9 that the other three approaches in Table 3 are an improvement over the approach used as the baseline. Specifically, the approach presented in this research has the biggest development, leading to the conclusion that its classification results are superior and its mistake detection rate is lower. As depicted in Figure 9, the ACC values of the approach presented in this research and approach 2 (baseline + pyramidal anchor) are vastly superior to those of the other two ways. Comparing their differences reveals that both the approach presented in this research and approach 2 (baseline + pyramidal anchor) employ the generation approach for pyramidal anchors. Because it generates anchors surrounding the target to be identified, the pyramidal anchor creation can reduce background interference, increase the quality of positive samples, and hence reduce the mistake detection rate. Based on the preceding study, it is clear that the approach presented in this paper is superior in its ability to detect floating garbage on the water surface.



Figure 9. Accuracy comparison of different models.

Figure 10 displays the Faster R-CNN baseline, the pyramidal anchor, the classification discrimination plot, and the curve representing the variation of loss value with the number of iterations during model training. As seen in Figure 10, each model presented in this paper has converged. The proposed approach compared to the baseline approach, the number of parameters and model complexity are enhanced. Figure 10 demonstrates that the initial value of the approach presented in this research is slightly larger at the start of training, but the converged loss value is lower, the fluctuation in training is also reduced, and the overall convergence process does not change majorly.



Figure 10. Loss comparison of different models.

Figure 11 shows the detection findings of three distinct forms of floating debris on the surface of the water. There is complicated background interference in the first panel, including water reflection and light reflection. The second panel contains the interference of water surface fluctuations, and the target to be detected is small. In the third panel, there are many objects to be inspected and the size is small, as well as the interference of water surface fluctuation. Figure 11a depicts the baseline Faster R-CNN test results, while Figure 11b depicts the test results of the proposed approach. Figure 11 demonstrates that only one target was detected in the baseline of the first test image. Additionally, the approach described in this study generates positive and negative samples surrounding the target to be investigated, which effectively decreases environmental interference and finds three targets to be inspected in the second test image, for the milk carton, the approach presented in this research improves

by 0.56 compared to the baseline approach due to the absence of feature information due to the target being too small. Using classification discrimination maps, the approach described in this research augments the semantic information of small target characteristics. Similarly, in the third test image, the approach presented in this research effectively eliminates the complicated background interference of the water surface fluctuation class and enhances the semantic feature information of small targets.



---8-----

Figure 11. Test results of baseline method and this method.

3.4. Experimental Comparison

In Figure 12, the accuracy performance of each type of floating garbage in the proposed approach is displayed.

By comparing the results in Table 5, the proposed approach generates much less of a detection anchor than the Faster R-CNN approach, yet the detection rate of each category is improved to various degrees. Objects that are particularly susceptible to environmental interference, such as water grass and tree branches, have a higher detection rate of boost. Therefore, it can be confirmed from this aspect that the revised model presented in this study is resistant to external elements such as light and rain. In addition, the reduction in the detection frame of the approach presented in this research suggests that the improved model's time efficiency will not be much enhanced in comparison to the original model. In conclusion, the approach presented in this paper improves detection accuracy while lowering the anchor generation number significantly compared to the baseline (Faster R-CNN) approach.





	Transit Amelian	Detect A	nchor	Accuracy%	
Class	larget Anchor	Faster R-CNN	PC-Net	Faster R-CNN	PC-Net
bottle	3318	6291	4020	89.4%	90.0%
grass	375	2191	676	68.8%	83.9%
branch	880	4288	1463	75.7%	79.8%
plastic-bag	830	2125	1261	87.8%	88.2%
milk-box	497	1971	788	86.5%	88.3%
ball	105	703	144	84.9%	87.3%
plastic-garbage	434	2430	674	84.6%	87.5%
leaf	554	2531	901	81.4%	86.0%

According to Table 6, the approach suggested in this paper increases accuracy 4.1% more than Faster R-CNN. Another comparison with various target identification algorithms reveals that the approach presented in this study outperforms the SSD algorithm by 7.3%, the YOLOV3 algorithm by 6.6%, the YOLOX algorithm by 3.6%, and the Dynamic R-CNN [40] algorithm by 2.8% in terms of accuracy.

Table 6. Results of recall and accuracy under different models.

Method	Recall/%	Accuracy/%
SSD	72.3	79.1
YOLOv3	76.4	79.8
YOLOX	80.5	82.8
Faster R-CNN	81.1	82.3
Dynamic R-CNN [40]	84.8	83.6
PC-Net	85.8	86.4

As shown in Figure 13, it is the experimental result of different algorithms. As can be seen from the figure, PC-NET has a good effect on detecting small targets and targets reflected by light. However, the detection frame error for large targets is relatively obvious. As shown in the sixth row in Figure 13. Although the target can be detected, the detection frame is relatively small and does not completely cover the target. The reason for this phenomenon may be that PC-Net pays less attention to edge information. When the anchor is generated, the PC-Net network refers too much to the intermediate information of the target, so the weight ratio of boundary information is too low.



Original

(a) SSD

(b)YOLOX (c)Faster R-CNN

(e)PC-Net

(d)Dynamic R-CNN

Figure 13. Comparison of detection effects of different models.

4. Conclusions

(b)YOLOv3

The existing water surface floating garbage detection techniques cannot easily detect floating garbage in conditions involving complicated backdrops, small size shares, and other variables. On the basis of the existing framework model of target detection, additional experimental analysis and enhancement have been conducted to present the PC-Net network for identifying and classifying floating garbage on the water surface. The effect of the production of pyramidal anchors on the accuracy of identifying and classifying floating garbage on the surface of water has also been investigated. After the implementation of a classification discrimination map, the detection effect of small targets is improved. This paper presents an effective solution to the problems of a complicated background, an excessive number of small targets, and an excessive disparity in target size within the dataset. The algorithmic model suggested in this study is 4.1% more efficient than the Faster R-CNN algorithm, 7.3% more efficient than the SSD algorithm, 6.6% more efficient than the YOLOV3 algorithm, 3.6% more efficient than the YOLOX algorithm, and 2.8% more efficient than the Dynamic R-CNN algorithm. The future will be based on the PC-Net

network model, as some progress has been made in this research on the detection of small targets. Future research will conduct further investigations on the detection of more complicated scenarios and problems involving floating garbage, mutual block caused by floating garbage and floating garbage accumulation. Future approaches will further improve the precision of floating debris detection on the water surface.

Author Contributions: Methodology, N.L. and B.Y.; Formal analysis, N.L., Y.L. and S.X.; Investigation and Resources, N.L. and S.X.; Writing—original draft preparation, H.H.; Writing—review and editing and supervision: N.L. and X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Jiangsu province petrochemical process key equipment digital twin technical engineering research center, project number DTEC202103. The first batch of cooperation projects in Industry-University-Research, Jiangsu Province in 2022, project number BY2022218.

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. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Themistocleous, K.; Papoutsa, C.; Michaelides, S.; Hadjimitsis, D. Investigating detection of floating plastic litter from space using sentinel-2 imagery. *Remote Sens.* 2020, 12, 2648. [CrossRef]
- 2. Dickens, C.; McCartney, M.; Tickner, D.; Harrison, I.; Pacheco, P.; Ndhlovu, B. Evaluating the global state of ecosystems and natural resources: Within and beyond the SDGs. *Sustainability* **2020**, *12*, 7381. [CrossRef]
- Cucui, G.; Ionescu, C.A.; Goldbach, I.R.; Coman, M.D.; Marin, E.L.M. Quantifying the economic effects of biogas installations for organic waste from agro-industrial sector. *Sustainability* 2018, 10, 2582. [CrossRef]
- 4. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOV4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
- Zhang, L.; Zhang, Y.; Zhang, Z.; Shen, J.; Wang, H. Real-Time Water Surface Object Detection Based on Improved Faster R-CNN. Sensors 2019, 19, 3523. [CrossRef]
- Lin, J.; Yang, C.; Lu, Y.; Cai, Y.; Zhan, H.; Zhang, Z. An Improved Soft-YOLOX for Garbage Quantity Identification. *Mathematics* 2022, 10, 2650. [CrossRef]
- Wang, C.; Zhou, Y.; Li, J. Lightweight Yolov4 Target Detection Algorithm Fused with ECA Mechanism. *Processes* 2022, 10, 1285. [CrossRef]
- 8. Verma, V.; Gupta, D.; Gupta, S.; Uppal, M.; Anand, D.; Ortega-Mansilla, A.; Alharithi, F.S.; Almotiri, J.; Goyal, N. A Deep Learning-Based Intelligent Garbage Detection System Using an Unmanned Aerial Vehicle. *Symmetry* **2022**, *14*, 960. [CrossRef]
- 9. Ma, W.; Wang, X.; Yu, J. A lightweight feature fusion single shot multibox detector for garbage detection. *IEEE Access* 2020, *8*, 188577–188586. [CrossRef]
- 10. Deng, H.; Ergu, D.; Liu, F.; Ma, B.; Cai, Y. An Embeddable Algorithm for Automatic Garbage Detection Based on Complex Marine Environment. *Sensors* **2021**, *21*, 6391. [CrossRef]
- Zheng, Y.; Liu, P.; Qian, L.; Qin, S.; Liu, X.; Ma, Y.; Cheng, G. Recognition and Depth Estimation of Ships Based on Binocular Stereo Vision. J. Mar. Sci. Eng. 2022, 10, 1153. [CrossRef]
- 12. Zeng, D.; Zhang, S.; Chen, F.; Wang, Y. Multi-scale CNN based garbage detection of airborne hyperspectral data. *IEEE Access* **2019**, *7*, 104514–104527. [CrossRef]
- 13. Li, X.; Tian, M.; Kong, S.; Wu, L.; Yu, J. A modified YOLOv3 detection method for vision-based water surface garbage capture robot. *Int. J. Adv. Robot. Syst.* 2020, *17*, 1729881420932715. [CrossRef]
- 14. Li, S.; Fu, X.; Dong, J. Improved Ship Detection Algorithm Based on YOLOX for SAR Outline Enhancement Image. *Remote Sens.* **2022**, *14*, 4070. [CrossRef]
- 15. Cheng, J.; Xiang, D.; Tang, J.; Zheng, Y.; Guan, D.; Du, B. Inshore Ship Detection in Large-Scale SAR Images Based on Saliency Enhancement and Bhattacharyya-like Distance. *Remote Sens.* **2022**, *14*, 2832. [CrossRef]
- 16. Zhang, L.; Wei, Y.; Wang, H.; Shao, Y.; Shen, J. Real-Time Detection of River Surface Floating Object Based on Improved RefineDet. *IEEE Access* **2021**, *9*, 81147–81160. [CrossRef]
- 17. Zhang, Z.; Gui, F.; Qu, X.; Feng, D. Netting Damage Detection for Marine Aquaculture Facilities Based on Improved Mask R-CNN. J. Mar. Sci. Eng. 2022, 10, 996. [CrossRef]

- Du, F.; Jiao, S.; Chu, K. Research on Safety Detection of Transmission Line Disaster Prevention Based on Improved Lightweight Convolutional Neural Network. *Machines* 2022, 10, 588. [CrossRef]
- 19. Zhang, C.; Zhang, X.; Tu, D.; Wang, Y. Small object detection using deep convolutional networks: Applied to garbage detection system. *J. Electron. Imaging* **2021**, *30*, 043013. [CrossRef]
- Tian, M.; Li, X.; Kong, S.; Wu, L.; Yu, J. A modified YOLOv4 detection method for a vision-based underwater garbage cleaning robot. *Front. Inf. Technol. Electron. Eng.* 2022, 23, 1217–1228. [CrossRef]
- 21. Wen, L.; Ding, J.; Xu, Z. Multiframe Detection of Sea-Surface Small Target Using Deep Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* 2021, 60, 1–16. [CrossRef]
- Zhou, X.; Jiang, L.; Hu, C.; Lei, S.; Zhang, T.; Mou, X. YOLO-SASE: An Improved YOLO Algorithm for the Small Targets Detection in Complex Backgrounds. Sensors 2022, 22, 4600. [CrossRef]
- 23. Gu, T. Detection of small floating targets on the sea surface based on multi-features and principal component analysis. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 809–813. [CrossRef]
- Gao, Z.; Dai, J.; Xie, C. Dim and small target detection based on feature mapping neural networks. J. Vis. Commun. Image Represent. 2019, 62, 206–216. [CrossRef]
- Jia, D.; Zhou, J.; Zhang, C. Detection of cervical cells based on improved SSD network. *Multimed. Tools Appl.* 2022, *81*, 13371–13387. [CrossRef]
- Liu, Y.; Liu, R.; Wang, S.; Yan, D.; Peng, B.; Zhang, T. Video Face Detection Based on Improved SSD Model and Target Tracking Algorithm. J. Web Eng. 2022, 2, 545–568. [CrossRef]
- 27. Sha, G.; Wu, J.; Yu, B. The improved faster-RCNN for spinal fracture lesions detection. J. Intell. Fuzzy Syst. 2022. Preprint. [CrossRef]
- Sharma, V.; Mir, R.N. Saliency guided faster-RCNN (SGFr-RCNN) model for object detection and recognition. J. King Saud Univ.-Comput. Inf. Sci. 2022, 34, 1687–1699. [CrossRef]
- 29. Huang, H.; Feng, X.; Jiang, J.; Chen, P.; Zhou, S. Mask RCNN algorithm for nuclei detection on breast cancer histopathological images. *Int. J. Imaging Syst. Technol.* **2022**, *32*, 209–217. [CrossRef]
- 30. Li, Z.; Li, Y.; Yang, Y.; Guo, R.; Yang, J.; Yue, J.; Wang, Y. A high-precision detection method of hydroponic lettuce seedlings status based on improved Faster RCNN. *Comput. Electron. Agric.* **2021**, *182*, 106054. [CrossRef]
- 31. Shi, X.; Li, Z.; Yu, H. Adaptive threshold cascade faster RCNN for domain adaptive object detection. *Multimed. Tools Appl.* **2021**, *80*, 25291–25308. [CrossRef]
- 32. Li, H.; Li, N.; Wu, R.; Wang, H.; Gui, Z.; Song, D. Gpr-rcnn: An algorithm of subsurface defect detection for airport runway based on gpr. *IEEE Robot. Autom. Lett.* **2021**, *6*, 3001–3008. [CrossRef]
- 33. Zhao, W.; Xu, M.; Cheng, X.; Zhao, Z. An insulator in transmission lines recognition and fault detection model based on improved faster RCNN. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–8.
- Yu, Y.; Zhang, K.; Yang, L.; Zhang, D. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Comput. Electron. Agric.* 2019, 163, 104846. [CrossRef]
- 35. Han, C.; Gao, G.; Zhang, Y. Real-time small traffic sign detection with revised faster-RCNN. *Multimed. Tools Appl.* **2019**, *78*, 13263–13278. [CrossRef]
- 36. Xie, H.; Chen, Y.; Shin, H. Context-aware pedestrian detection especially for small-sized instances with Deconvolution Integrated Faster RCNN (DIF R-CNN). *Appl. Intell.* **2019**, *49*, 1200–1211. [CrossRef]
- Sun, X.; Wu, P.; Hoi, S.C. Face detection using deep learning: An improved faster RCNN approach. *Neurocomputing* 2018, 299, 42–50. [CrossRef]
- Wang, K.; Liu, M. YOLOv3-MT: A YOLOv3 using multi-target tracking for vehicle visual detection. *Appl. Intell.* 2022, 52, 2070–2091. [CrossRef]
- Li, X.; Ding, L.; Wang, L.; Cao, F. FPGA accelerates deep residual learning for image recognition. In Proceedings of the 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 15–17 December 2017; pp. 837–840.
- Zhang, H.; Chang, H.; Ma, B.; Wang, N.; Chen, X. Dynamic R-CNN: Towards high quality object detection via dynamic training. In Proceedings of the European conference on computer vision 2020, Virtual, 23–28 August 2020.