

## Article

# Sea Cucumber Detection Algorithm Based on Deep Learning

Lan Zhang <sup>1,2,†</sup>, Bowen Xing <sup>1,\*,†</sup> , Wugui Wang <sup>3,†</sup> and Jingxiang Xu <sup>1,†</sup> 

1 College of Engineering Science and Technology, Shanghai Ocean University, Shanghai 201306, China; zhang258708921@163.com (L.Z.); jxxu@shou.edu.cn (J.X.)

2 Shanghai Investigation Design & Research Institute, Shanghai 200335, China

3 China Ship Development and Design Center, Wuhan 430064, China; woxinyouyou99@163.com

\* Correspondence: bwxing@shou.edu.cn

† These authors contributed equally to this work.

**Abstract:** The traditional single-shot multiBox detector (SSD) for the recognition process in sea cucumbers has problems, such as an insufficient expression of features, heavy computation, and difficulty in application to embedded platforms. To solve these problems, we proposed an improved algorithm for sea cucumber detection based on the traditional SSD algorithm. MobileNetv1 is selected as the backbone of the SSD algorithm. We increase the feature receptive field by receptive field block (RFB) to increase feature details and location information of small targets. Combined with the attention mechanism, features at different depths are strengthened and irrelevant features are suppressed. The experimental results show that the improved algorithm has better performance than the traditional SSD algorithm. The average precision of the improved algorithm is increased by 5.1%. The improved algorithm is also more robust. Compared with YOLOv4 and the Faster R-CNN algorithm, the performance of this algorithm on the P-R curve is better, indicating that the performance of this algorithm is better. Thus, the improved algorithm can stably detect sea cucumbers in real time and provide reliable feedback information.

**Keywords:** sea cucumber fishing; image recognition; deep learning; single-shot multibox detector



**Citation:** Zhang, L.; Xing, B.; Wang, W.; Xu, J. Sea Cucumber Detection Algorithm Based on Deep Learning. *Sensors* **2022**, *22*, 5717. <https://doi.org/10.3390/s22155717>

Academic Editor: Loris Nanni

Received: 21 June 2022

Accepted: 28 July 2022

Published: 30 July 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/>).

## 1. Introduction

Recently, sea cucumber farming has been rapidly developed as an aquatic type of farming [1]. With the development of sea cucumber production, the breeding problems of sea cucumbers are becoming increasingly serious. Traditional sea cucumber fishing, which has low efficiency and high risk, is mainly dependent on manual work [2]. To promote the development of sea cucumber breeding automation, it is necessary to research the automatic identification of sea cucumbers in the natural underwater environment based on machine vision [3,4]. A sea cucumber target recognition by BP neural networks was proposed by Wang et al. They used RGB and depth images as prior knowledge to improve recognition accuracy [5]. A depth residual network with different configurations for sea cucumber target recognition was proposed by Guo et al. [6]. A real-time cultured sea cucumber detector attached to an autonomous underwater vehicle (AUV), with YOLOv4-tiny and transfer learning are proposed by Thao et al. [7]. After the study of sea cucumber target recognition, the above scholars and other researchers have proposed a series of methods with practical applications. However, the majority of research has not considered the application to embedded platforms and real-time issues.

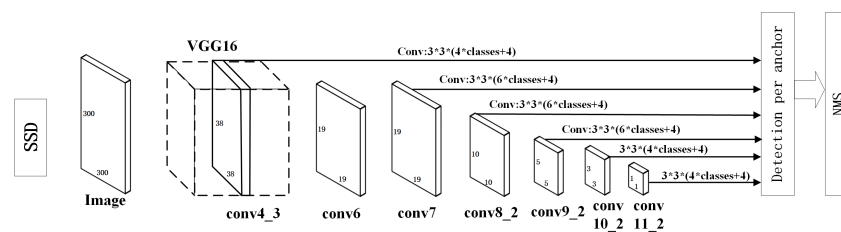
To solve the above problems, an improved SSD target detection algorithm is proposed. First, we use the MobileNetv1 to detect and locate sea cucumbers. Secondly, the shallow feature receptive fields are improved by RFB, and have more details and location information of small targets. This algorithm combines the attention mechanism to strengthen features at different depths, suppress irrelevant features, and perform feature fusion to further improve the accuracy of sea cucumber detection.

## 2. SSD Object Detection Algorithm

The core of the traditional SSD algorithm is to predict results by different convolution layers. The traditional SSD algorithm introduces the idea of regression and proposes the concept of the prior box [8]. The algorithm predicts targets on feature maps of different receptive fields, and completes target location and classification at one time.

### 2.1. Traditional SSD Model Structure

A traditional SSD network uses VGG16 as the backbone network. The two fully connected layers of VGG16, which lay the foundation for subsequent multi-scale feature extraction, are converted into convolution layers. The traditional SSD network has six feature maps at different dimensions. Conv4\_3, Conv7, Conv8\_2, Conv9\_2, Conv10\_2, and Conv11\_2 are connected to the final classification layer for regression prediction [9,10]. The network structure of the traditional SSD algorithm is shown in Figure 1.



**Figure 1.** Network-structure diagram of traditional SSD algorithm.

First, the traditional SSD algorithm generates six prior boxes on six feature maps. Secondly, the prior boxes are assembled from different feature maps. Finally, the final set of prior boxes is selected by non-maximum suppression (NMS) [11,12].

### 2.2. The Loss Function of Traditional SSD

Model prediction performance is tested by the loss function [13,14]. The traditional SSD algorithm loss function is divided into two parts, the classification loss function and the position loss function,

$$L(x, c, l, g) = \frac{1}{N} \left( L_{conf}(x, c) + \alpha L_{loc}(x, l, g) \right), \quad (1)$$

where  $L_{conf}$  is the classification loss function,  $L_{loc}$  is the position loss function,  $N$  is the number of samples,  $\alpha$  is the weighting coefficient,  $x$  is the matching information for the current prediction box category,  $c$  is the labeled category,  $l$  represents the coordinates of the search prediction boxes, and  $g$  represents the coordinates of marked boundary frames.

The position loss function is shown as follows. Where  $x_{ij}^k = \{1, 0\}$  represents whether the  $i$  search prediction box matches the  $J$  real box on category  $k$ ,  $l_i^m$  is the prediction box;  $\hat{g}_j^m$  is the ground-truth box,  $N$  is the number of matched samples,  $Pos$  is the positive sample,  $Box$  is a set of prediction box attribute parameters, and  $smooth_{L1}$  is the error function of L1:

$$L_{loc}(x, l, g) = \sum_{i \in Pos} \sum_{m \in Box} x_{ij}^k smooth_{L1} \left( l_i^m - \hat{g}_j^m \right). \quad (2)$$

The classification loss function is shown as follows. Where  $\hat{c}_p^i$  is the probability which is the target of the  $i$  prediction box is  $p$ ,  $\hat{c}_0^i$  is the probability that the target is not detected in the  $i$  prediction frame,  $x_{ij}^p$  represents whether the  $i$  search prediction box matches the  $j$  real box on category  $P$ ,  $Neg$  is the negative sample, and  $Pos$  is the positive sample:

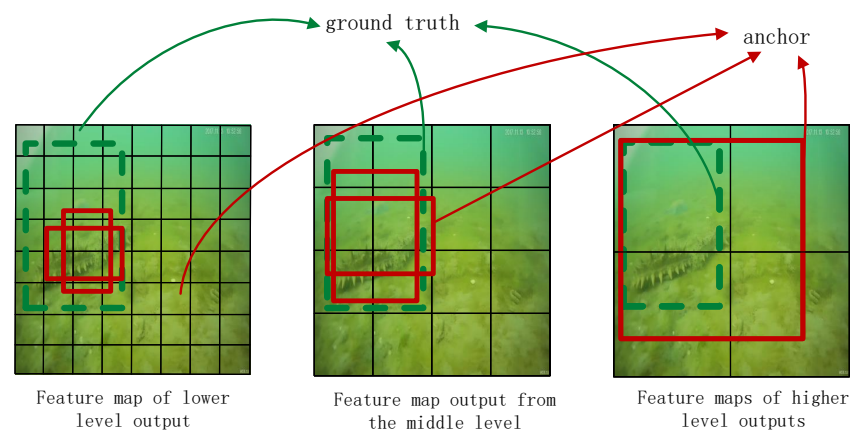
$$L_{conf}(x, c) = - \sum_{i \in Pos} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i^0) \quad (3)$$

$$\hat{c}_i^p = \exp(c_i^p) / \sum_p \exp(c_i^p). \quad (4)$$

### 2.3. Traditional SSD Performance Analysis

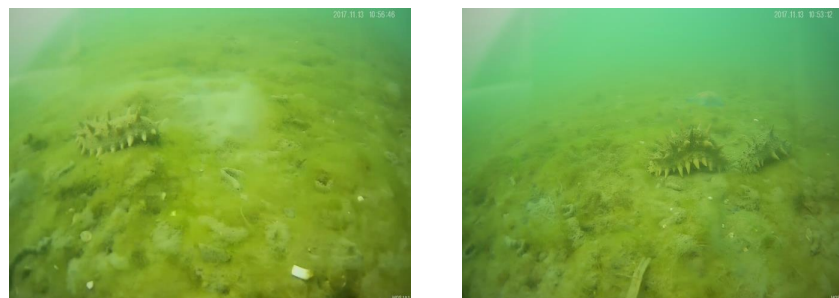
The traditional SSD algorithm uses multi-scale feature maps for detection, and generates six prior boxes which have different aspect ratios [15].

As shown in Figure 2, feature maps of the lower level are larger, but the receptive field of each unit is relatively small. Feature maps of the lower level are suitable for detecting small targets. Feature maps of the higher level are smaller, but the receptive field of each unit is relatively large. Feature maps of the larger level are suitable for detecting large targets. Therefore, the detection of small targets relies on lower-level feature maps. The number of lower-level-feature convolution layers is small, resulting in insufficient feature extraction and poor semantic distinction.



**Figure 2.** Output results from different levels of SSD.

As shown in Figure 3, the sea cucumber is a small target. When using the traditional SSD algorithm to identify sea cucumbers, the detection ability is insufficient, the robustness is poor, and it is impossible to accurately locate the underwater sea cucumber. In addition, the traditional SSD algorithm has a large computing capacity, and cannot be applied to embedded platforms. To overcome these problems, the traditional SSD algorithm is optimized.



**Figure 3.** Underwater robots taking photographs of sea cucumbers.

## 3. Sea Cucumber Detection Algorithm Based on Improved MobileNetv1 SSD

### 3.1. MobileNetv1 Structure

MobileNetv1, which uses depthwise separable convolution, is a lightweight CNN structure [16]. The depthwise separable convolution is mainly divided into two parts, depthwise convolution and pointwise convolution [17,18].

As shown in Figures 4–6, where  $D_F$  represents the height and width of the input matrix,  $D_K$  represents the size of the convolution kernel,  $M$  is the number of input feature matrix channels, and  $N$  is the number of output feature matrix channels:

$$\frac{D_K \times D_K \times D_F \times D_F \times M + M \times N \times D_F \times D_F}{D_K \times D_K \times M \times N \times D_F \times D_F} = \frac{1}{N} + \frac{1}{D_K^2}. \tag{5}$$

The depthwise separable convolution in MobileNetV1 networks uses a convolution kernel of  $3 \times 3$ . The computation of standard convolution parameters is about nine times that of depthwise separable convolution parameters. The introduction of the MobileNetV1 network reduces the number of calculation parameters and realizes a lightweight structure.

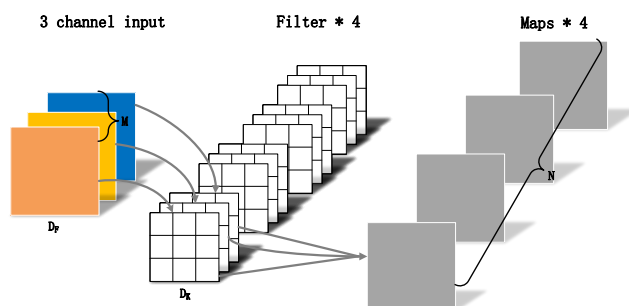


Figure 4. Standard convolution.

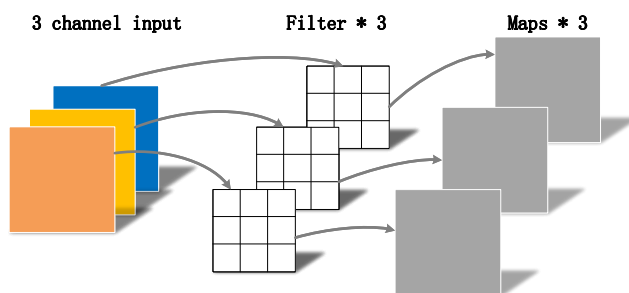


Figure 5. Depthwise convolution.

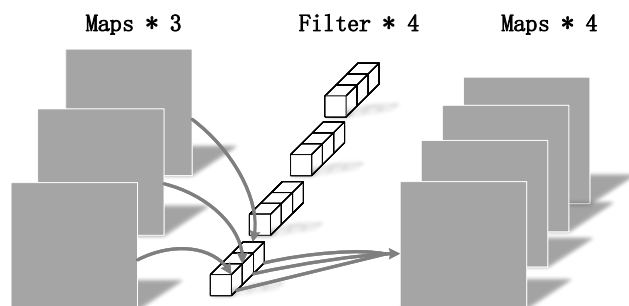


Figure 6. Pointwise convolution.

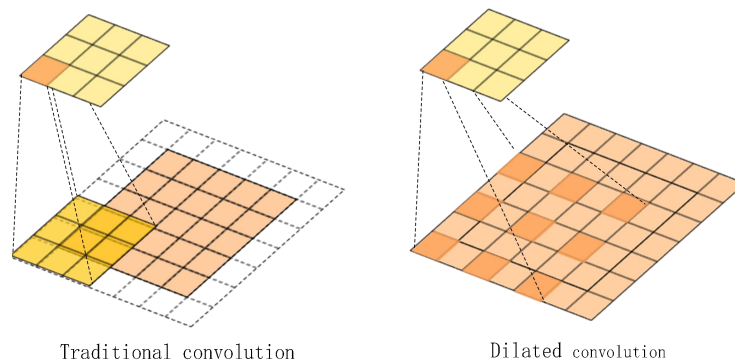
### 3.2. Introduction of SSD Network with Dilated Convolutional Structure

The RFB module, which is similar to the inception network, is a multi-branch convolution block [19]. The RFB module consists of a multi-branch convolution layer and a dilated convolution layer. The dilated convolution adds holes to increase the receptive field. The dilated convolution has a hyper-parameter  $K_d$  [20,21]:

$$K_d = l \times (K - 1) + 1, \tag{6}$$

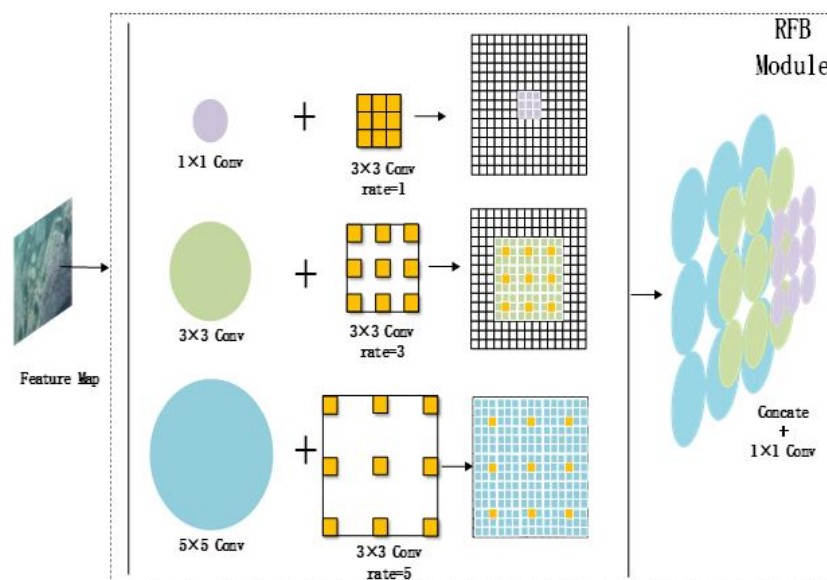
where  $K_d$  is the size of the dilated convolution kernel,  $L$  is a dilation factor, and  $K$  is the size of the original convolution kernel.

As shown in Figure 7, when the dilation factor is 2, the dilated convolution has larger receptive fields than the original convolution.



**Figure 7.** Comparison of traditional and expansive convolution.

The RFB module processing flow is shown in Figure 8. The RFB module consists of three branch modules [22]. The bottom layer in each branch is processed by convolutional nuclei of  $1 \times 1$ ,  $3 \times 3$ , and  $5 \times 5$ . Finally, three different receptive fields are obtained [23,24].



**Figure 8.** RFB module processing flowchart.

In this paper, the RFB module is introduced to process the Conv4\_3 layer and the FC7 layer. As a result, shallow semantic information is richer. Deep feature maps need feature processing, so introducing an attention mechanism to strengthen the feature maps at different depths.

### 3.3. Introduction of Attention Mechanisms

In this paper, spatial attention and channel attention mechanisms are introduced to strengthen the features at different depths [25]. Finally, the multi-layer feature maps are fused to achieve better detection results.

First, the spatial attention mechanism is introduced into the features at different depths. Secondly, when a multi-channel feature map is an input, spatial attention will learn and train relationships of different spatial domains in the feature map. After giving higher weights to more representative local features, a two-dimensional spatial weight map  $W$  is generated. Finally, the two-dimensional weight map  $W$  is multiplied by the corresponding position space to obtain a representative feature map. The training mechanism of spatial

attention is as follows. First, through global maximum pooling and global average pooling, feature representative values are obtained at each spatial position. Then, these feature representative values are fused by convolution operation to obtain the spatial attention map. Finally, spatial attention weights of 0–1 are generated by the sigmoid activation function [26–28]. The calculation formula is as follows:

$$W_s(F) = \text{sigmoid}\{f^{3 \times 3}\{[\text{MaxPool}(F); \text{AvgPool}(F)]\}\}, \quad (7)$$

where  $W_s(F)$  is the feature map after spatial attention mechanism processing,  $F$  is the input multi-channel feature map,  $f^{3 \times 3}$  is the convolution operation of  $3 \times 3$ ,  $\text{AvgPool}$  is the global average pooling, and  $\text{MaxPool}$  is the global maximum pooling.

First, the channel attention method filters out irrelevant channel features in a multi-channel feature map. Secondly, through the relationship between each channel in the feature map to learn the weight array. Finally, the weight array is multiplied by the corresponding channel [29,30]. The following formula is used to calculate the channel attention mechanism:

$$W'_s(F) = \text{sigmoid}\{\text{MLP}[\text{AvgPool}(F)] + \text{MLP}[\text{MaxPool}(F)]\}, \quad (8)$$

where  $W'_s(F)$  is the result characteristic diagram,  $\text{MLP}$  is the multilayer perceptron,  $\text{AvgPool}$  is the global average pooling, and  $\text{MaxPool}$  is the global maximum pooling.

### 3.4. MobileNetv1 SSD Network with Attention Mechanisms

The improved SSD algorithm can detect sea cucumbers by the MobileNetv1 SSD network. The size of the shallow feature receptive fields is increased by RFB [31]. First, the attention mechanism creates a model of the relationship between relevant feature channels and feature spaces. Secondly, the obtained weights between each feature channel and feature space, are multiplied by the original feature information. Finally, the obtained channel features map and spatial feature map not only contain the most representative features but also suppress irrelevant features. In short, the improved SSD algorithm not only improves the recognition accuracy of small target objects but also reduces the missed detection rate and false detection rate.

In this paper, the features of the Conv4\_3 and FC7 layers are selected to utilize RFB. Finally, P1( $19 \times 19$ ), P2( $10 \times 10$ ), P3( $5 \times 5$ ), P4( $3 \times 3$ ), P5( $2 \times 2$ ), and P6( $1 \times 1$ ) feature maps are obtained [32]. The improved SSD network structure is shown in Figure 9.

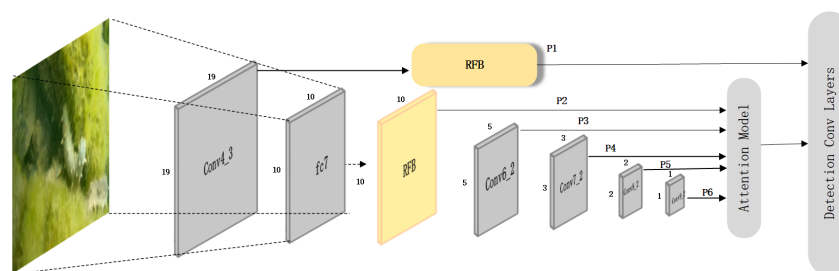


Figure 9. Improved SSD network structure.

## 4. Experimental Results and Analysis

To verify the feasibility of the proposed algorithm, the following computing environment was used: Intel(R)Core(TM)i7-9750H CPU, NVIDIA GeForce RTX 2060 graphical processing unit, Ubuntu 20.04 operating system, and Keras 2.1.5 deep-learning framework.

### 4.1. Experimental Data

The experimental data were extracted from a video of shallow sea cucumber farming. The video is recorded by a remotely operated vehicle. By framing the video, 1710 original sea cucumber images were collected by using data augmentation to increase the original

dataset. The number of images in the dataset is 4005. The filtered images were manually annotated by Labellmg software. In the dataset, 70% comprise the training set, and 30% comprise the test set. Labellmg software uses rectangular boxes to mark sea cucumber targets. The label information is stored in XML format. The information includes image name, category name, image size, and position information.

#### 4.2. Evaluation Index Setting

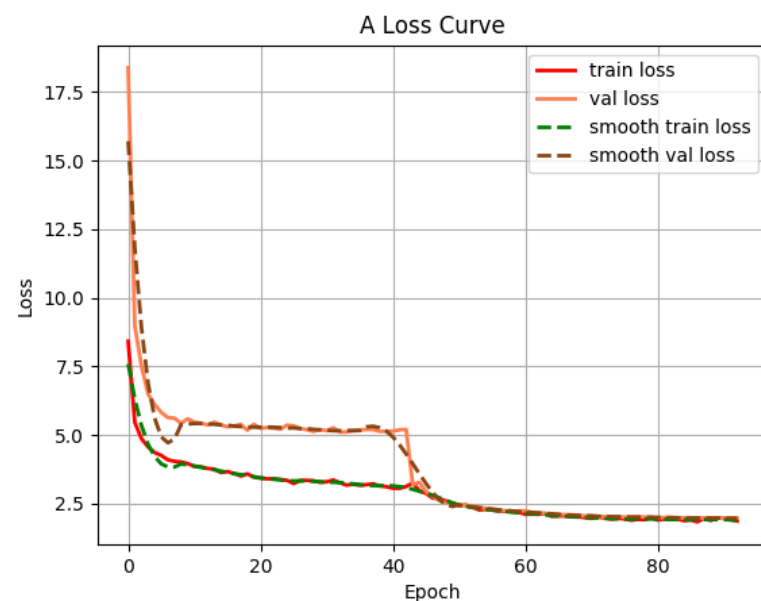
The intersection over union, which is the overlap rate between the candidate bound and the ground truth bound, is a concept in object detection. Owing to the complexity of the marine environment, this paper uses mean average precision (mAP) with an IOU threshold of 0.5 and detection frame rate as evaluation indicators. The mean average precision, which is calculated by the P-R curve, is an evaluation metric in object detection models. The P-R curve consists of a precision curve and a recall curve. The P-R curve reflects global performance [33]. The following equations are used to calculate precision, recall, and mAP:

$$\text{Precision} = TP / (TP + FP) \quad (9)$$

$$\text{Recall} = TP / (TP + FN) \quad (10)$$

$$mAP = \sum_{i=1}^k AP_i / k, \quad (11)$$

where  $FP$  is a false positive example,  $FN$  is a false negative example,  $TP$  is the real example,  $TN$  is a true negative example,  $AP_i$  is the average precision of a category; and  $k$  is the number of categories. Figure 10 is a loss curve.



**Figure 10.** The loss curve.

#### 4.3. Improved SSD Model Validation

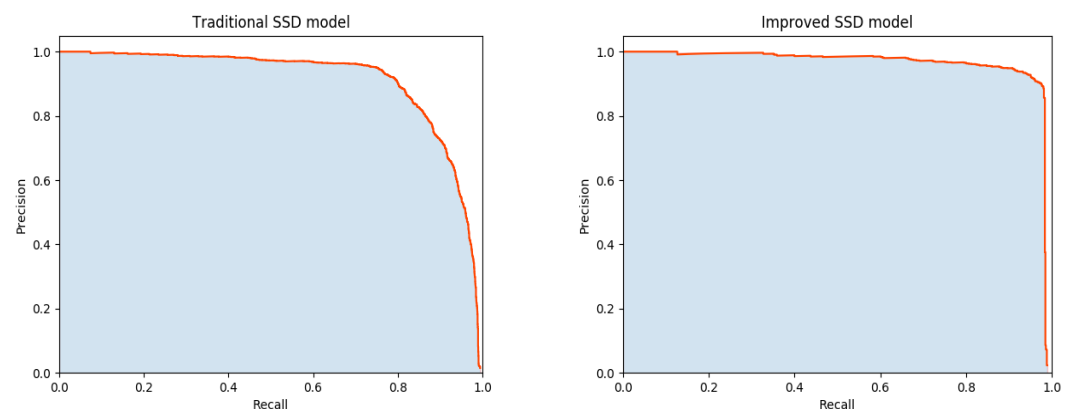
To verify the effectiveness of the improved SSD model, this paper conducts a comparative experiment on the recognition effect of the traditional SSD model. The improved SSD model uses the same training set and sets the same parameters as the traditional SSD model. The training is divided into the freezing stage and the thawing stage. In the freezing stage, the learning rate is  $5 \times 10^{-4}$ , the model backbone is frozen, and the network is fine-tuned. In the thawing stage, the learning rate was  $1 \times 10^{-4}$ , the model backbone is thawed, and the feature extraction network is adjusted.

As shown in Table 1, compared with the traditional algorithm, the average accuracy of the algorithm proposed in this paper is improved by 5.1%, the detection frame rate is improved by 18.514 frame/s. The detection accuracy of sea cucumber is improved effectively.

**Table 1.** Improved and traditional SSD algorithm comparison test results.

Detection method	$mAP_{50}$	Frames per Second	No. of Iterations
Traditional SSD algorithms	0.914	5.916	100
Improved algorithm	0.965	24.430	100

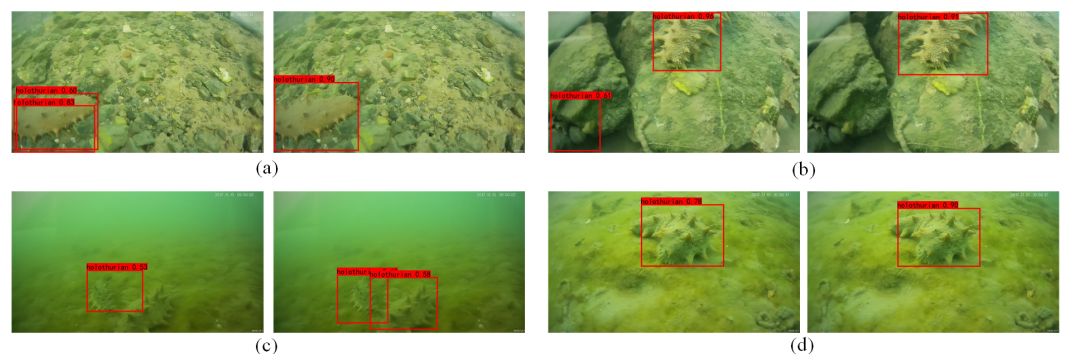
Figure 11 is the comparison chart of the P-R curve before and after the model improvement. The model performance is reflected by the P-R curve.



**Figure 11.** Comparison of P-R curves before and after model improvement.

As shown in Figure 11, the improved SSD model has the larger area between the P-R curve and coordinate axis. The equilibrium point of the improved SSD model is closer to coordinate (1,1), indicating that the system performance of the improved SSD model is better.

Figure 12 shows the comparison images of sea cucumbers detected by the SSD model before and after improvement. As shown in Figure 12a, the result of the traditional SSD algorithm has a repeat box. As shown in Figure 12b, the result of the traditional SSD algorithm has a false detection. As shown in Figure 12c, the result of the traditional SSD algorithm reveals a missed detection. The comparison results in Figure 12d show that the confidence of the improved SSD algorithm has increased by 12%. According to the results in Figure 12, the proposed algorithm reduces the missed detection rate. Compared with the traditional algorithm, the confidence of the proposed algorithm is higher. In addition, the portion of the target missed by the traditional SSD algorithm can be detected by the proposed algorithm.



**Figure 12.** Comparison of sea cucumber identification before and after model improvement.



#### 4.4. Comparison of Different Models

In order to further prove the effectiveness of the algorithm in this paper, the FasterR-CNN algorithm and the YOLOv4 algorithm are selected for comparison test. The Faster RCNN algorithm and the YOLOv4 are typical one-stage and two-stage target detection algorithms. The following graph is the P-R curve comparison graph of four algorithms. Under the same sea cucumber dataset, the performance of the proposed algorithm on the P-R curve is better than YOLOv4 and Faster R-CNN, indicating that the performance of the proposed algorithm is better (see Figure 13). In terms of detection speed, in the process of algorithm testing, the proposed algorithm is slower than YOLOv4 algorithm and faster than FasterR-CNN algorithm. Compared with the typical YOLOv4 and FasterR-CNN in the first and second stages, the proposed algorithm has better target detection ability and higher applicability for sea cucumber target recognition.

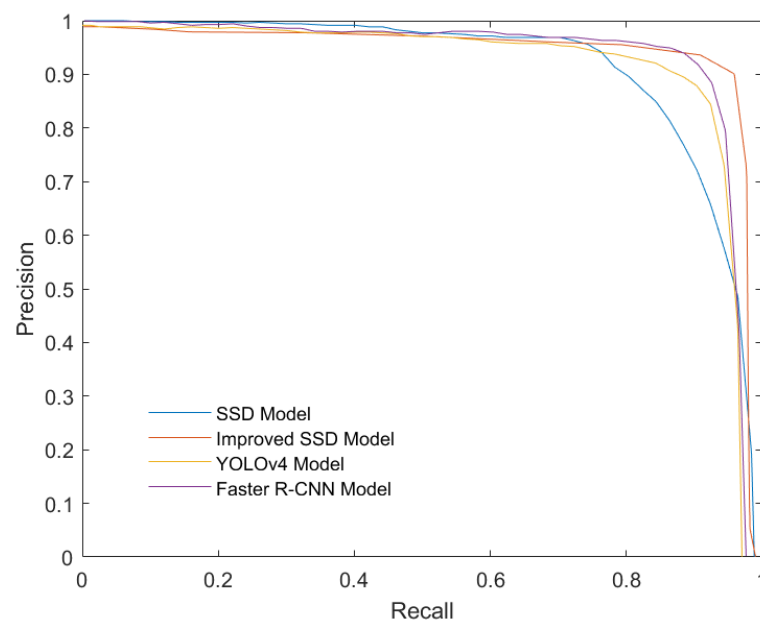


Figure 13. Four models comparison of P-R curves.

## 5. Conclusions

Aimed at the low accuracy and a large amount of computation required by traditional SSD algorithms in detecting sea cucumbers, an improved algorithm for sea cucumber detection is proposed. First, a MobileNetv1 SSD network is used to detect and locate sea cucumbers. Through a receptive field block, the shallow feature receptive field is improved to increase the detail and location information. The improved algorithm is combined with an attention mechanism to strengthen features of different depths. The experimental results show that, compared with the traditional SSD algorithm, the proposed algorithm has good robustness and recognition rate. Compared with the YOLOv4 and Faster R-CNN algorithms, the performance of this algorithm on the P-R curve is better, indicating that the performance of this algorithm is better. Underwater sea cucumbers have the characteristic of changing body color to match their environment, and future research will aim at solving the problems caused by this characteristic, focusing on finding an innovative pre-treatment method to achieve efficient identification of sea cucumbers.

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

**Funding:** This research was funded by Shanghai Science and Technology Committee (STCSM) Local Universities Capacity-building Project (No. 22010502200), Scientific Research Project of China Three Gorges Corporation (No. 202003111).

**Acknowledgments:** The authors would like to express their gratitude for the support of Fishery Engineering and Equipment Innovation Team of Shanghai High-level Local University.

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

### Abbreviations

The following abbreviations are used in this manuscript:

SSD Single Shot MultiBox Detector  
RFB Receptive Field Block

### References

1. Ru, X.; Zhang, L.; Li, X.; Liu, S.; Yang, H. Development strategies for the sea cucumber industry in China. *J. Ocean. Limnol.* **2019**, *37*, 300–312. [\[CrossRef\]](#)
2. Daniel Azari, B.G.; Daniel, R.S.; Walsalam, G.I. Sea cucumber aquaculture business potential in Middle East and South-East Asia-Pathways for ecological, social and economic sustainability. *Surv. Fish. Sci.* **2021**, *7*, 113–121. [\[CrossRef\]](#)
3. Li, J.; Xu, C.; Jiang, L.; Xiao, Y.; Deng, L.; Han, Z. Detection and Analysis of Behavior Trajectory for Sea Cucumbers Based on Deep Learning. *IEEE Access* **2020**, *8*, 18832–18840. [\[CrossRef\]](#)
4. Zhang, H.; Yu, F.; Sun, J.; Shen, X.; Li, K. Deep learning for sea cucumber detection using stochastic gradient descent algorithm. *Eur. J. Remote Sens.* **2020**, *53*, 53–62. [\[CrossRef\]](#)
5. Wang, Q.; Li, S.; Qin, H.; Hao, A. Super-Resolution of Multi-Observed RGB-D Images Based on Nonlocal Regression and Total Variation. *IEEE Trans. Image Process.* **2016**, *25*, 1425–1440. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Guo, X.; Zhao, X.; Liu, Y. Underwater sea cucumber identification via deep residual networks. *Inf. Process. Agric.* **2019**, *6*, 307–315. [\[CrossRef\]](#)
7. Ngo Gia, T.; Li, Y.; Jin, D.; Guo, J.; Li, J.; Tang, Q. Real-Time Sea Cucumber Detection Based on YOLOv4-Tiny and Transfer Learning Using Data Augmentation. In *Advances in Swarm Intelligence*; Tan, Y., Shi, Y., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2021; Volume 12690, pp. 119–128.
8. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
9. Ning, C.; Zhou, H.; Song, Y.; Tang, J. Inception Single Shot MultiBox Detector for object detection. In Proceedings of the 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Hong Kong, China, 10–14 July 2017; pp. 549–554.
10. Huanjie, C.; Qiqi, W.; Guowei, Y.; Jialin, H.A.N.; Chengjuan, Y.I.N.; Jun, C.; Yizhong, W. SSD Object Detection Algorithm with Multi-Scale Convolution Feature Fusion. *J. Front. Comput. Sci. Technol.* **2019**, *13*, 1049.
11. Shen, Z.; Liu, Z.; Li, J.; Jiang, Y.-G.; Chen, Y.; Xue, X. DSOD: Learning Deeply Supervised Object Detectors From Scratch. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1919–1927.
12. Kumar, A.; Zhang, Z.J.; Lyu, H. Object detection in real time based on improved single shot multi-box detector algorithm. *J. Wirel. Commun. Netw.* **2020**, *2020*, 1–18. [\[CrossRef\]](#)
13. Arora, A.; Grover, A.; Chugh, R.; Reka, S.S. Real Time Multi Object Detection for Blind Using Single Shot Multibox Detector. *Wirel. Pers. Commun.* **2019**, *107*, 651–661. [\[CrossRef\]](#)
14. Leng, J.; Liu, Y. An enhanced SSD with feature fusion and visual reasoning for object detection. *Neural Comput. Appl.* **2019**, *31*, 6549–6558. [\[CrossRef\]](#)
15. Nguyen, A.Q.; Nguyen, H.T.; Tran, V.C.; Pham, H.X.; Pestana, J. A Visual Real-time Fire Detection using Single Shot MultiBox Detector for UAV-based Fire Surveillance. In Proceedings of the 2020 IEEE Eighth International Conference on Communications and Electronics (ICCE), Phu Quoc Island, Vietnam, 13–15 January 2021; pp. 338–343.
16. Chen, H.Y.; Su, C.Y. An Enhanced Hybrid MobileNet. In Proceedings of the 2018 9th International Conference on Awareness Science and Technology (iCAST), Fukuoka, Japan, 19–21 September 2018; pp. 308–312.
17. Sinha, D.; El-Sharkawy, M. Thin MobileNet: An Enhanced MobileNet Architecture. In Proceedings of the 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 10–12 October 2019; pp. 0280–0285.
18. Hsiao, S.F.; Tsai, B.C. Efficient Computation of Depthwise Separable Convolution in MobileNet Deep Neural Network Models. In Proceedings of the 2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), Penghu, Taiwan, 15–17 September 2021; pp. 1–2.
19. Liu, S.; Huang, D.; Wang, Y. Receptive Field Block Net for Accurate and Fast Object Detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 385–400.
20. Yuan, Z.; Liu, Z.; Zhu, C.; Qi, J.; Zhao, D. Object Detection in Remote Sensing Images via Multi-Feature Pyramid Network with Receptive Field Block. *Remote Sens.* **2021**, *13*, 862. [\[CrossRef\]](#)

21. Feng, H.; Guo, J.; Xu, H.; Ge, S.S. SharpGAN: Dynamic Scene Deblurring Method for Smart Ship Based on Receptive Field Block and Generative Adversarial Networks. *Sensors* **2021**, *21*, 3641. [[CrossRef](#)] [[PubMed](#)]
22. XiaoFan, L.; HaiBo, P.; Yi, W.; JiangChuan, L.; HongXiang, X. Introduce GloU into RFB net to optimize object detection bounding box. In Proceedings of the Proceedings of the 5th International Conference on Communication and Information Processing, Chongqing, China, 15–17 November 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 108–113.
23. Jin, F.; Liu, X.; Liu, Z.; Rui, J.; Guan, K. A Target Recognition Algorithm in Remote Sensing Images. In Proceedings of the 6th International Symposium of Space Optical Instruments and Applications, Delft, The Netherlands, 24–25 September 2019; pp. 133–142.
24. Liu, S.; Huang, D.; Wang, Y. Adaptive NMS: Refining Pedestrian Detection in a Crowd. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 6459–6468.
25. Zhang, J.; Zhao, Z.; Su, F. Efficient-Receptive Field Block with Group Spatial Attention Mechanism for Object Detection. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 3248–3255.
26. Fukui, H.; Hirakawa, T.; Yamashita, T.; Fujiyoshi, H. Attention Branch Network: Learning of Attention Mechanism for Visual Explanation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 10705–10714.
27. Li, Y.; Zeng, J.; Shan, S.; Chen, X. Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism. *IEEE Trans. Image Process.* **2019**, *28*, 2439–2450. [[CrossRef](#)] [[PubMed](#)]
28. Choi, E.; Bahadori, M.T.; Sun, J.; Kulas, J.; Schuetz, A.; Stewart, W. RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism. In Proceedings of the Annual Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; Curran Associates, Inc.: Red Hook, NY, USA, 2016; Volume 29.
29. Tao, C.; Gao, S.; Shang, M.; Wu, W.; Zhao, D.; Yan, R. Get The Point of My Utterance! Learning Towards Effective Responses with Multi-Head Attention Mechanism. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; p. 4424.
30. Chen, Y.; Liu, L.; Phoneyilay, V.; Gu, K.; Xia, R.; Xie, J.; Zhang, Q.; Yang, K. Image super-resolution reconstruction based on feature map attention mechanism. *Appl. Intell.* **2021**, *51*, 4367–4380. [[CrossRef](#)]
31. Jiang, H.; Shi, T.; Bai, Z.; Huang, L. AHCNet: An Application of Attention Mechanism and Hybrid Connection for Liver Tumor Segmentation in CT Volumes. *IEEE Access* **2019**, *7*, 24898–24909. [[CrossRef](#)]
32. Shang, T.; Dai, Q.; Zhu, S.; Yang, T.; Guo, Y. Perceptual Extreme Super-Resolution Network With Receptive Field Block. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Virtual, 14–19 June 2020; pp. 440–441.
33. Li, X.; Yang, Z.; Wu, H. Face Detection Based on Receptive Field Enhanced Multi-Task Cascaded Convolutional Neural Networks. *IEEE Access* **2020**, *8*, 174922–174930. [[CrossRef](#)]