Hindawi
Complexity
Volume 2020, Article ID 5476142, 14 pages
https://doi.org/10.1155/2020/5476142

Research Article

A Marine Object Detection Algorithm Based on SSD and Feature Enhancement

Kai Hu,[1,2] **Feiyu Lu**,[1,2] **Meixia Lu**,[1,2] **Zhiliang Deng**,[1,2] **and Yunping Liu**[1,2]

[1]*College of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, China*
[2]*Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), Nanjing University of Information Science & Technology, Nanjing 210044, China*

Correspondence should be addressed to Kai Hu; nuistpanda@163.com

Received 5 July 2020; Accepted 3 August 2020; Published 30 September 2020

Guest Editor: Zhijie Wang

Autonomous detection and fishing by underwater robots will be the main way to obtain aquatic products in the future; sea urchins are the main research object of aquatic product detection. When the classical Single-Shot MultiBox Detector (SSD) algorithm is applied to the detection of sea urchins, it also has disadvantages of being inaccurate to small targets and insensitive to the direction of the sea urchin. Based on the classic SSD algorithm, this paper proposes a feature-enhanced sea urchin detection algorithm. Firstly, according to the spiny-edge characteristics of a sea urchin, a multidirectional edge detection algorithm is proposed to enhance the feature, which is taken as the 4th channel of image and the original 3 channels of underwater image together as the input for the further deep learning. Then, in order to improve the shortcomings of SSD algorithm's poor ability to detect small targets, resnet 50 is used as the basic framework of the network, and the idea of feature cross-level fusion is adopted to improve the feature expression ability and strengthen semantic information. The open data set provided by the National Natural Science Foundation of China underwater Robot Competition will be used as the test set and training set. Under the same training and test conditions, the AP value of the algorithm in this paper reaches 81.0%, 7.6% higher than the classic SSD algorithm, and the confidence of small target analysis is also improved. Experimental results show that the algorithm in this paper can effectively improve the accuracy of sea urchin detection.

1. Introduction

Autonomous detection and fishing by underwater robots will be the main way to obtain aquatic products in the future, and sea urchins are the main research object of aquatic product detection. To complete the autonomous detection and salvage of underwater robots [1], a series of basic research and scientific problems such as underwater communication [2], underwater positioning [3], information perception [4], target detection and identification [5], and target grasping need to be solved [6], which is an important field of concern for many researchers. Sea urchins are one of the mainstream objects in the current research of aquatic product detection. Detecting and identifying sea urchins in the video is an important prerequisite for salvaging it, and has strong engineering realization value and scientific research significance.

At present, there is no detection algorithm specifically for sea urchins, and target detection methods are based on traditional machine learning and deep learning [7]. The traditional machine learning methods first select some candidate regions on a given underwater image [8], then use methods such as Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG) [9] to define features, and finally, use the Support Vector Machine (SVM) [10], Adaptive Boosting(AdaBoost) [11], and other technologies for classification.

However, the traditional target detection method has the following problems: when the underwater light changes rapidly, the algorithm is not effective [12]; when the slow motion and the background color are consistent, the feature pixel cannot be extracted; the time complexity is high; and the noise resistance performance is poor. Compared with

traditional machine learning algorithms, deep learning has the advantages of large data volume, strong scalability, good adaptability, and easy conversion. It can perform end-to-end target detection without specially defined features. It is a powerful method for automatically learning feature representations from data. Based on this, this article uses deep learning methods to carry out related research on sea urchins.

Current target detection methods based on deep learning can be mainly divided into two categories: one is a two-stage target detection algorithm based on candidate regions, such as the Region-based Convolutional Neural Network (R-CNN) [13], Fast Region-based Convolutional Neural Network (Fast R-CNN) [14], and Faster Region-based Convolutional Neural Network (Faster R-CNN) [15]. The other is a one-stage target detection algorithm based on regression, such as SSD (Single-Shot MultiBox Detector) [16] and YOLO (You Only Look Once) [17].

In 2013, Girshick et al. proposed the region-based convolutional neural network R-CNN, and it is a target detection algorithm based on deep learning. The convolutional neural network, which can be applied to image classification tasks, has been successfully applied to image detection tasks. R-CNN target detection achieved an accuracy rate of 53.3% on the Pascal VOC 2012 test set. Compared with the previous best target detection algorithm, the accuracy has been improved by 30%.

In 2015, Ross Girshick improved the previously proposed R-CNN algorithm and proposed Fast R-CNN. Fast R-CNN combines CNN feature extraction and subsequent SVM classification and used a new network to achieve classification and regression. The Fast R-CNN can obtain a one-stage training process through multitask learning, which greatly reduces read and write operations; and after Fast R-CNN sends the whole image into the CNN, the CNN characteristics of different candidate regions in the image are calculated through the mapping relationship so that only one calculation is needed to avoid the problem of repeated calculation. Because of these improvements, Fast R-CNN is 9 times faster than R-CNN in training.

Nonetheless, because Fast R-CNN still uses a selective search strategy, it does not reach the industrial level in speed. In 2016, Ren Shaoqing and Joseph Redmon et al. proposed the Faster R-CNN and YOLO algorithms, respectively. The former was improved on the basis of the Fast R-CNN. The Faster R-CNN built a Region Proposal Network (RPN), this network replaces the selective search method used in the R-CNN and Fast R-CNN, it can train the neural network model end-to-end, and then, accelerate the speed of target detection, so that the detection speed can basically meet the real-time requirements. The latter is a target detection method based on regression. For finding the candidate target box and determining the category of the target, YOLO is carried out on the output layer at the same time, which greatly accelerates the detection speed and reaches 45 fps/s, and it can meet the requirements of real-time target detection.

In the same year, Wei Liu et al. proposed SSD; it combined the anchor mechanism in the Faster R-CNN and the regression idea in YOLO, as the input image feature extraction using a small convolution filter, and the feature of the different scales with different aspect ratio classification prediction. Compared with the Faster R-CNN, the average detection accuracy is basically the same, but the training speed of the model is faster. Compared with YOLO, the detection speed is slightly improved, and the average detection accuracy is increased from 63.4% to 72.1%. Therefore, this article chooses the SSD algorithm to conduct sea urchin detection.

In the preliminary work of applying SSD to sea urchin detection, we conduct experiments on the public data set provided by the National Natural Science Foundation of China Underwater Robot Competition. After data analysis, we found that the existing classic SSD algorithm has a disadvantage of inaccurate detection of small sea urchin targets, and the overall detection performance of the sea urchin has room for further improvement (the AP value of classic SSD is 73.4%).

Considering that, in traditional machine learning, a sea urchin has several important features such as black, round, and spiny, we analyze that deep learning should have no pressure on the extraction of black and round features. Due to the different angles of the thorns and different convolution sizes, it may be that deep learning has the possibility of further improving the detection effect of thorns. Therefore, this paper intends to use feature enhancement methods, trying to enhance the analysis ability of deep learning on the feature learning of spiny, thereby improving the performance of detection and recognition of sea urchins; then, in order to improve the shortcomings of the SSD algorithm's poor ability to detect small targets [18], we use Resnet50 [19] as the basic architecture for network feature extraction to replace the original VGG16 [20] infrastructure of the SSD algorithm and use the feature cross-level fusion idea to improve the feature expression ability and strengthen the semantic information. This idea intends to use 3 fusions to complete the connection between the high-level network and the low-level network, which can expand the scope of the target detection field of view while enhancing the context information of small target prediction.

This article is organized as follows: the second part introduces the background information of the SSD algorithm; the third part describes the improved network algorithm in detail; the fourth part shows the experimental setup, provides the results, and discusses these results; and the fifth part summarizes the full text.

2. Background

The convolutional neural network is a type of feedforward neural network, and it includes convolution calculation and has a deep structure [21]. It has three core ideas: local network connection, convolution kernel parameter sharing, and pooling. The joint effect of local connection and parameter sharing is to reduce the number of parameters, make the operation simple and efficient, and be able to operate on very large data sets. Pooling is to aggregate the characteristics of different locations to obtain lower

dimensions, and it can prevent the problem of overfitting [22]. On this basis, a slight adjustment to the network framework can improve the generalization ability and robustness of the model [23].

The Single-Shot MultiBox Detector (SSD) algorithm is a one-stage algorithm; it is one of the most real-time and advanced target detection algorithms at present. The SSD algorithm has completed the task of classifying and locating the target using only a full-convolution network. The structural framework of SSD is shown in Figure 1. It uses VGG16 as the basic architecture and introduces the design concept of a prior frame. On the basis of VGG16, a new cascaded convolution layer is added to obtain multiscale feature maps to detect the target. All the prediction results are merged together, and the final detection result is obtained by Nonmaximum Suppression (NMS).

The design philosophy of the SSD model is as follows.

2.1. SSD Area Candidate Box. The SSD adopts the method of the multiscale feature map. Region candidate boxes of different sizes and aspect ratios will be set on feature maps of different scales. The regional candidate box definition is calculated as follows:

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m-1}(k-1), \quad k \in [1, m]. \tag{1}$$

In formula (1), m is the number of feature layers; $s_{\min}$ is the lowest feature map scale (default value is 0.2); $s_{\max}$ is the highest feature map scale (default value is 0.9); and the intermediate feature map scales are evenly distributed. The candidate boxes in the region have different aspect ratios of $a_r \in \{1, 2, 3, 1/2, 1/3\}$. The width and height of the region candidate box are $w_k^a = s_k\sqrt{a_r}$ and $h_k^a = s_k/\sqrt{a_r}$, and for the region candidate box with an aspect ratio of 1, an additional scale $s_k' = \sqrt{s_k s_{k+1}}$ is added. The center coordinates of each region candidate box are $((i+0.5)/w_{fk}, (j+0.5)/h_{fk})$. Of them, w_{fk} is the width of the k feature map, h_{fk} is the height of the k feature map, $j \in [0, h_{fk})$, and $i \in [0, w_{fk})$.

Finally, SSD uses conv4_3, fc7, conv8_2, conv9_2, conv10_2, and conv11_2 as prediction layers. With the deepening of the network, the size of the feature map decreases gradually and the size of the candidate frame increases continuously. Therefore, the shallow feature map is used to detect small targets, and the deep feature map is used to detect large targets.

2.2. SSD Loss Function. During the SSD training process, the target position and category are regressed. The target loss function includes two parts: Location Loss (Loc) and Confidence Loss (Conf). The expression is as follows:

$$L(x, c, l, g) = \frac{1}{N}\left(L_{\text{conf}}(x, c) + aL_{\text{loc}}(x, l, g)\right). \tag{2}$$

In formula (2), N is the number of matches between the regional candidate box and the real box. If $N = 0$, then $L = 0$; x is the matching result of the regional candidate box and the real box of different categories. If it matches $x = 1$, $x = 0$; c is the confidence of the prediction box; l is the position offset information of the prediction box; g is the offset information of the real box and the regional candidate box; and a is the position loss weight usually set to 1.

3. Feature Enhancement

3.1. Multidirectional Edge Feature Enhancement Algorithm. In the preliminary work of applying SSD to sea urchin detection, we found that the detection performance still has room for further improvement (the AP value of classic SSD is 73.4%).

After the analysis, we believe that the sea urchin has several important features such as black, round, and spiny, and we think that deep learning should have no pressure on the extraction of black and round features. Due to the different angles of the thorns and different convolution sizes, it may be that deep learning has the possibility of further improving the detection effect of thorns. Therefore, this paper proposes a sea urchin detection algorithm based on feature enhancement. According to the spiny-edge characteristics of a sea urchin, a multidirectional edge detection algorithm is proposed to enhance the feature, which is taken as the 4th channel of image and the original 3 channels of underwater image together as the input for the further deep learning.

Feature enhancement is to enhance the useful information in the image [24], and it can be a distortion process. The purpose is to improve the analysis effect of the image for the given image application situation [25]. We purposefully emphasize the overall or local characteristics of the image, make the original unclear image clear or emphasize some interesting features [26], expand the difference between the features of different objects in the image, and suppress uninteresting features. The image quality is improved [27], the amount of information is more abundant [28], and the image interpretation and recognition effects are strengthened [29] to meet the needs of some special analyses.

The sea urchin is black overall and has a round shape with thorns. If the edge features of the sea urchin can be effectively extracted, it will have a certain effect on its identification and positioning. The sea urchin spines have a small angle. The detection angle of each operator in the edge detection algorithm is 180 degrees in one direction, the detection angle in 2 directions is 90 degrees, and the detection angle in 4 directions is 45 degrees. It is difficult to accurately identify most sea urchin spines. According to this, this paper proposes 8 directions and 16 directions to improve the detection angle resolution to improve the accuracy of detecting sea urchin edge thorns and, then, use the edge detection channel as the 4th channel to enhance the edge characteristics of the sea urchin.

This paper presents a $5 * 5$ size, 16-direction Prewitt operator, compared with other classic Sobel operator 2-direction, Laplace operator unidirectional, Prewitt operator 2-direction, Prewitt operator 4-direction, and Prewitt operator 8-direction. The results are shown in Figures 2(b)–2(g) and 3(b)–3(g).

Figures 2(a) and 3(a) are two sea urchin maps randomly selected from the public data set provided by the National

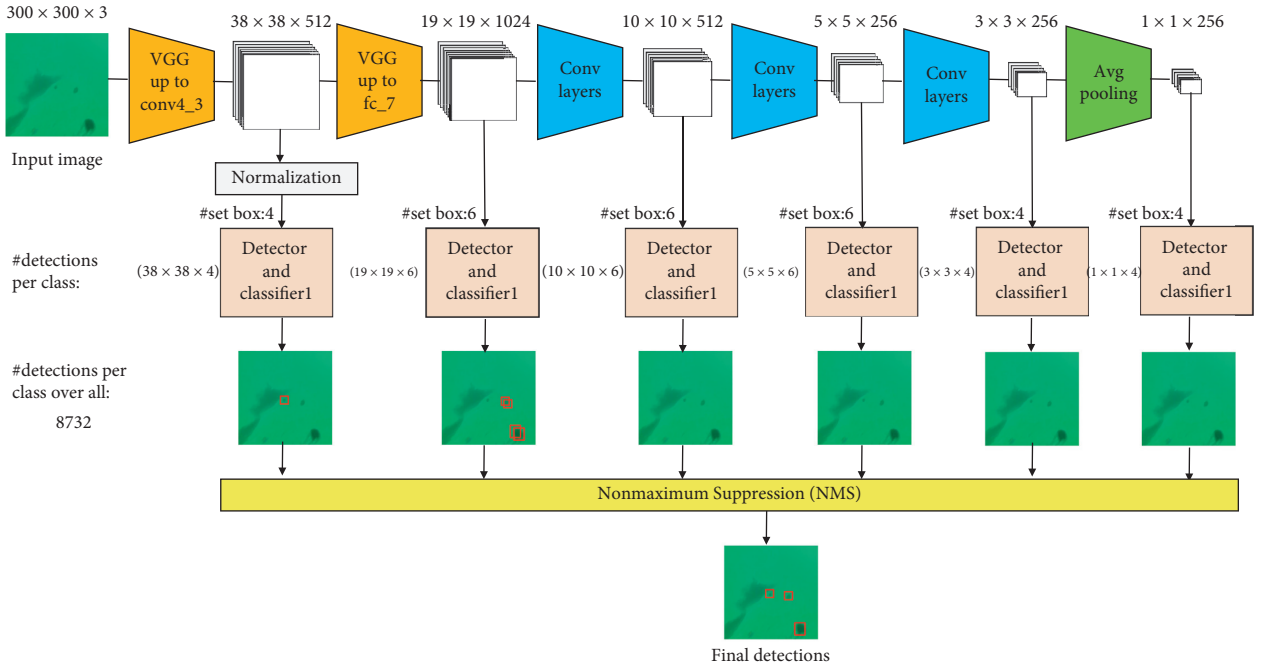

Figure 1: SSD structure frame.

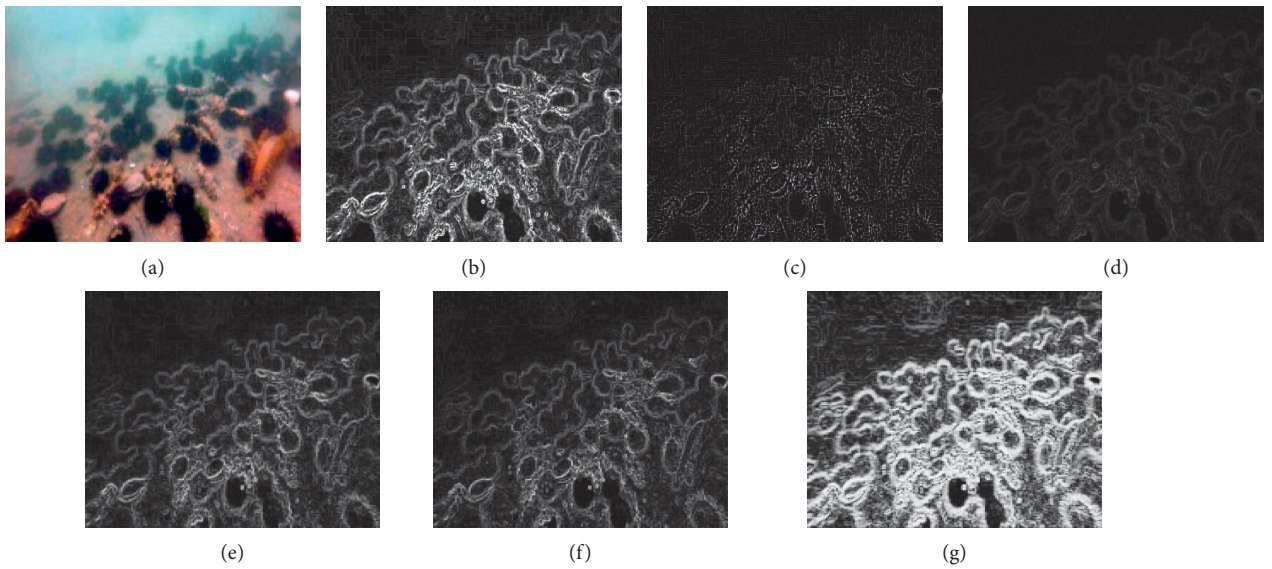

Figure 2: (a) The image of underwater sea urchins; (b) the image of Sobel operator 2-direction edge detection; (c) the image of Laplace operator single-direction edge detection; (d) the image of Prewitt operator 2-direction edge detection; (e) the image of Prewitt operator 4-direction edge detection; (f) the image of Prewitt operator 8-direction edge detection; and (g) the image of Prewitt operator 16-direction edge detection.

Natural Science Foundation of China Underwater Robot Competition. Figure 2(a) is a scene of a large number of small target sea urchins. Figure 3(a) is a scene of multi-category aquatic products interference, and it is more representative. Comparing the image edge extraction result maps in different scenes, it is possible to compare the effect differences of the three operators. Figures 2(b)2(g) and 3(b) 3(g) are the effect diagrams after performing edge detection for Sobel, Prewitt and Laplace operators in Figures 2(a) and 3(a). It can be found that the Laplace operator single-direction edge detection effect is not as good as the Sobel operator 2-direction, Prewitt operator 4-direction, and Prewitt operator 8-direction, but the edge spines of sea urchin can be better extracted. In contrast, the Prewitt

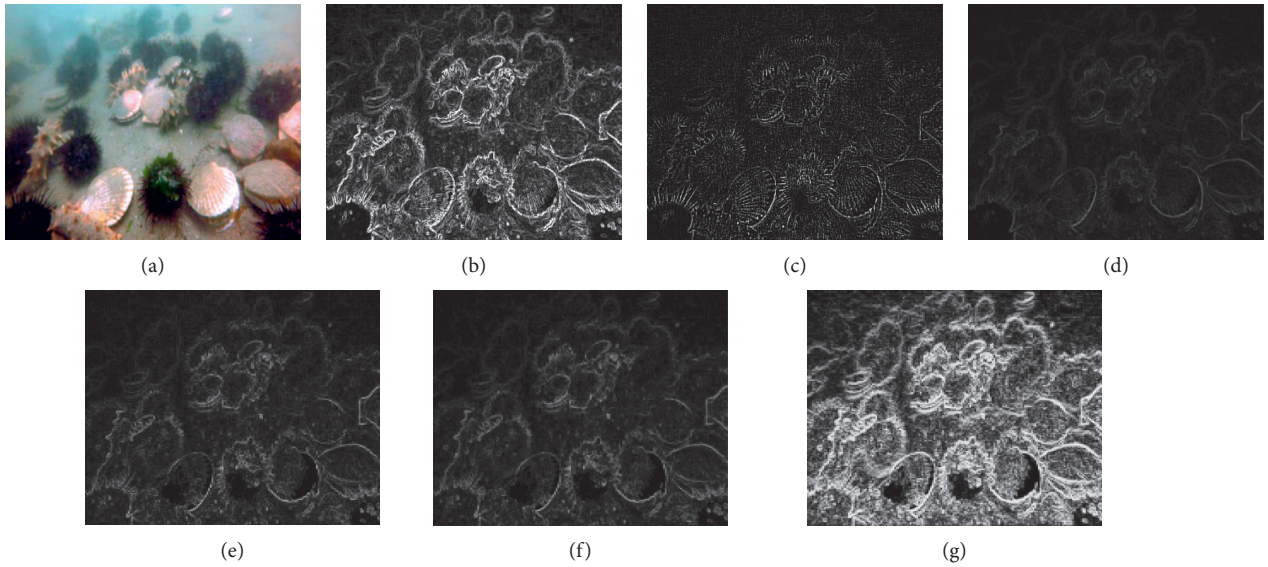

Figure 3: (a) The image of underwater sea urchins; (b) the image of Sobel operator 2-direction edge detection; (c) the image of Laplace operator single-direction edge detection; (d) the image of Prewitt operator 2-direction edge detection; (e) the image of Prewitt operator 4-direction edge detection; (f) the image of Prewitt operator 8-direction edge detection; and (g) the image of Prewitt operator 16-direction edge detection.

operator 2-direction of edge detection effect is poorer, it is hard to find the edge information of the sea urchin, but the noise is less. Among them, the Prewitt operator 16-direction edge detection has the best effect and the edge gray value is high, it can detect the edge information of the sea urchin better, but the noise is larger, while the SSD algorithm has a strong ability to suppress noise, so the noise is not considered for the time-being problem.

The visualization effect of Prewitt operator 16-direction edge detection is the best. The multidirectional edge detection process takes Prewitt operator 16-direction as an example, as shown in feature enhancement program. Taking the output image as the 4th channel, it can be clearly seen that the thorny area on the edge of the sea urchin is highlighted, and the background area is almost white, so that the background and urchin can be better segmented.

Feature enhancement program:

Step 1. Read the image

Step 2. Change the color map into grayscale

Step 3. Input Prewitt operator 16-direction stencil, such as $x1$, $x2$, $x3, \ldots, x16$; the gradient images are obtained by using the Prewitt gradient operator in 16 directions and converted into CV_8UC1

Step 4. OTSU binarization is performed on the converted gradient image to obtain the binarization image

Step 5. Carry out and operation on 16 binary images according to corresponding pixels

Step 6. Conduct multiple iterations of expansion and corrosion operations on the binaries

Step 7. Contour search and fill for small area blocks

Step 8. Background corrosion

Step 9. Output the mask diagram

3.2. Feature Cross-Level Fusion Mode. In the preliminary work of applying SSD to the detection of sea urchin, we found that the detection performance has room for further improvement (the AP value of classic SSD is 73.4%).

Through in-depth analysis, we think the main problem is that some sea urchins are small targets, and the traditional SSD algorithm has a relatively poor detection effect on small-sized objects. The feature map representation capability of shallow extraction is not strong enough. In this way, there will be misdetection and missed detection of small targets of the sea urchin. According to this, in order to improve the characteristics of the poor recognition ability of the SSD algorithm for small targets, the improved SSD algorithm draws on the idea of the residual network and uses Resnet 50 instead of VGG16 as the basic framework of the network. The network architecture is shown in Figure 4. Deepening the neural network by learning the residuals can avoid the problems of overfitting and the disappearance of the network gradient, learn more abstract texture features and semantic features, and strengthen the expression ability of features, so as to improve the ability of target classification and location. At the same time, a feature cross-level fusion method is proposed to improve the feature expression ability and strengthen the semantic information and further improve the problem of poor detection ability of the SSD algorithm for small targets.

In Figure 4, three fusion modules are used to complete the connection between the high-level network and the low-level network. The context information of small target prediction is enhanced, and the field of view of target detection is expanded. The fusion feature of fusion module 1, that is, the skip connection of res2_3 and res5_3, is fed into conv6, as shown in Figure 5. In order to fuse the feature maps of res2_3 and res5_3, the feature maps of res5_3 need to be upsampled. First, the res5_3 feature map is upsampled

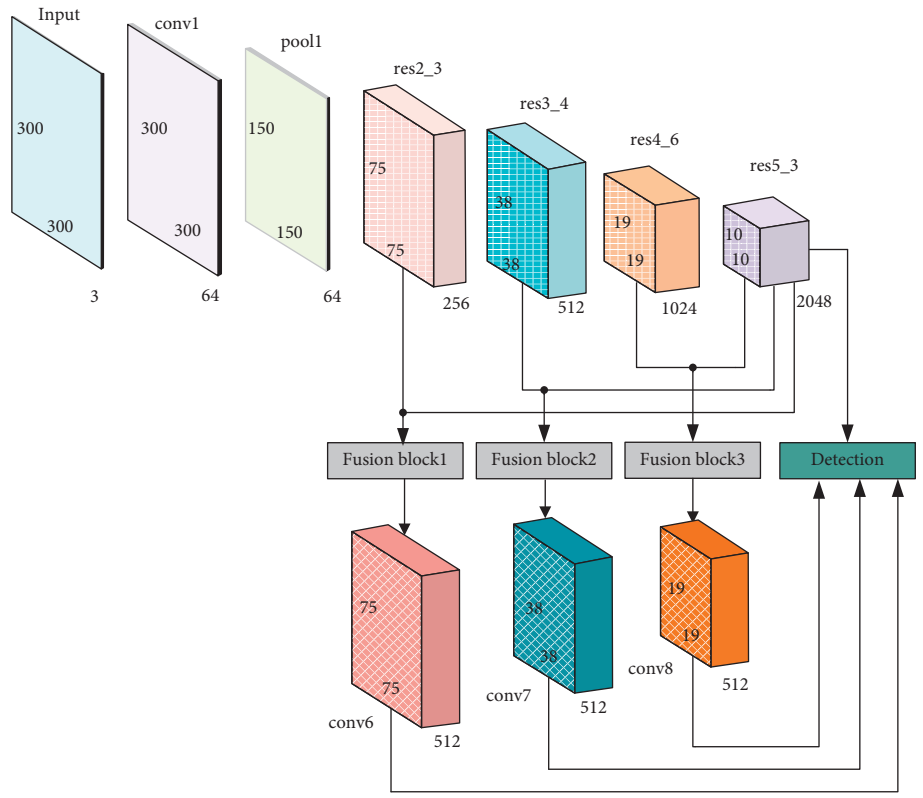

Figure 4: Network architecture diagram.

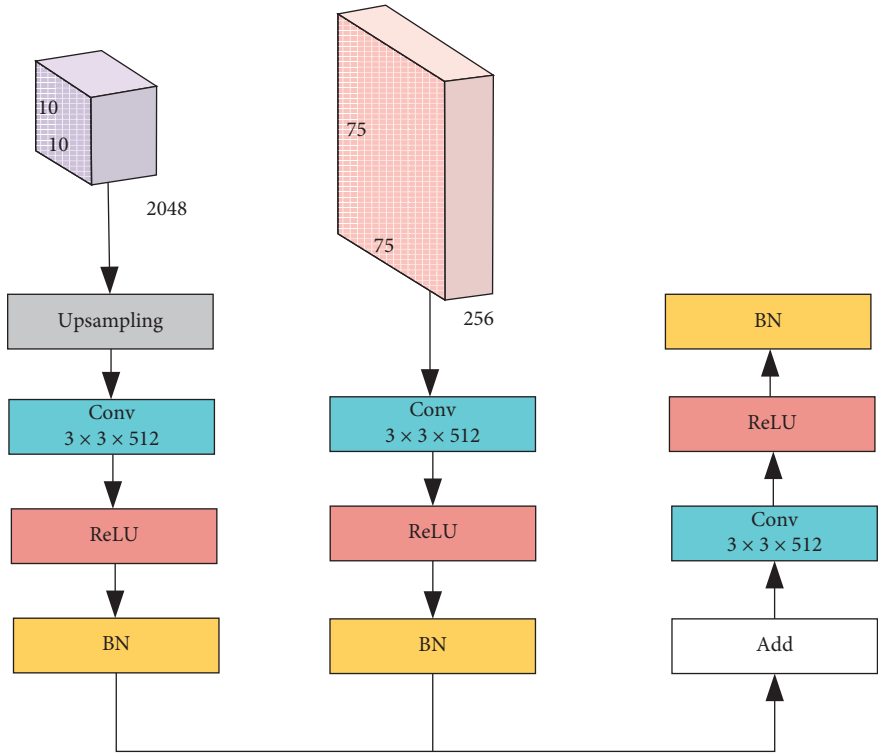

Figure 5: Features of cross-level fusion modules.

to the same size as res2_3 by interpolation and upsampling. The output of the upsampling is mapped to the modified activation function layer (Rectified Linear Unit (ReLU)) [22] through a convolution layer with a convolution kernel of 3×3. Then, go through the L2 regularization layer for normalization (Batch Normalization (BN)). Res2_3 is

directly mapped to the Relu activation function layer through a 3×3 convolution kernel map and, then, input to the L2 regularization layer. The output between the two partitions is summed and passed to the Relu layer after merging. Finally, $256 \cdot 3 \times 3$ convolution kernels are used to ensure that the detected features are distinguishable, and the fusion function is realized after a Relu layer. The fusion module 2 is the feature fusion of res3_4 and res5_3 jump-level connection to conv7. The fusion module 3 is the feature fusion of res4_6 and res5_3 jump-level connection to conv8. Finally, the four feature maps of conv6 (75×75), conv7 (38×38), conv8 (19×19), and res5_3 (10×10) after feature fusion are sent to the prediction module for prediction.

3.3. Overall Improvement Process. Figure 6 is the overall flow chart of underwater image detection, which simply describes the calculation and operation process of the improved SSD algorithm. After analysis, we believe that the sea urchin has several important features such as being black, round, and spiny, among which deep learning should have no pressure on the extraction of black and round features. Due to the different angles of the thorns and different convolution sizes, it may be that deep learning has the possibility of further improving the detection effect of thorns. Therefore, this paper proposes a sea urchin detection algorithm based on feature enhancement. According to the spiny-edge characteristics of sea urchin, a Prewitt operator 16-direction edge detection algorithm is proposed to enhance the feature, which is taken as the 4th channel of image and the original 3 channels of underwater image together as the input for the further deep learning.

At the same time, some sea urchins belong to small targets, and the traditional SSD algorithm has a relatively poor detection effect on small-sized objects. According to this, in order to improve the characteristics of the SSD algorithm's poor ability to recognize small targets, the improved SSD algorithm draws on the idea of residual network and uses Resnet 50 and replaces VGG16 as the basic framework of the network, and it can avoid the problems of overfitting and the disappearance of the network gradient. It adopts the feature cross-level fusion idea to improve the feature expression ability and strengthen the semantic information. Finally, the feature map is sent to the trained model for prediction, and the sea urchin detection result map is obtained.

4. Experimental Analysis

CPU: Inter i7-9700k, memory: 16G DDR4, GPU: Nvidia Geforce GTX2080Ti, operating system: 64-bit Ubuntu 16.04, and the experimental framework is Pytorch open source framework. Stochastic gradient descent is used as the learning rate, the initial learning rate is 0.001, and the learning rate is reduced by 10 times when the number of iterations is 100, 150, and 200 cycles; the momentum is set to 0.9, the weight attenuation coefficient is set to 0.0001, and the training batch size is 32, training 300 cycles.

This article uses the public data set provided by the National Natural Science Foundation of China Underwater Robot Competition [30] as the training set and test set. The public data set contains 3000 training pictures and 800 test pictures, a total of 3800 pictures, and the picture size is uniformly $300 * 300$. Among them, the sea urchins vary in size and type. The 3800 pictures are all underwater images, with different lighting conditions, complex backgrounds [31], and varying degrees of occlusion. The detection of sea urchins and other targets in this article is actually a two-category problem. The ultimate goal is to correctly detect all sea urchins and reduce the missed detection rate and the false detection rate.

In order to better evaluate the model, we set TP (True Positives) to represent the real class, which is the positive sample predicted correctly by the model, and FP (False Positives) to represent the true negative class, which is the positive sample predicted by the model to be negative. FN (False Negatives) represents a false negative class, which is the negative sample predicted by the model as positive, and TN (True Negatives) represents a true negative class, which is the negative sample predicted by the model as negative. The formulas of accuracy and recall rate are as follows:

$$\begin{aligned} \text{precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}. \end{aligned} \tag{3}$$

For each category of target detection, a check line-recall rate (P-R) curve can be obtained, and the accuracy under the curve is Average Precision (AP). For this task, there is only one type, so AP is mAP (mean Average Precision). The AP formula is as follows:

$$\text{AP} = \int_0^1 P(R)\mathrm{d}R. \tag{4}$$

4.1. Analysis on the Rationality of the Multidirectional Edge Feature Enhancement Algorithm. In the preliminary work of applying SSD to sea urchin detection, we found that the detection performance still has room for further improvement (the AP value of classic SSD is 73.4%).

After analysis, we believe that the sea urchin has several important features such as being black, round, and spiny, and we think that deep learning should have no pressure on the extraction of black and round features. Due to the different angles of the thorns and different convolution sizes, it may be that deep learning has the possibility of further improving the detection effect of thorns. Therefore, this paper proposes a sea urchin detection algorithm based on feature enhancement. According to the spiny-edge characteristics of sea urchin, a multidirectional edge detection algorithm is proposed to enhance the feature, which is taken as the 4th channel of image and the original 3 channels of underwater image together as the input for the further deep learning.

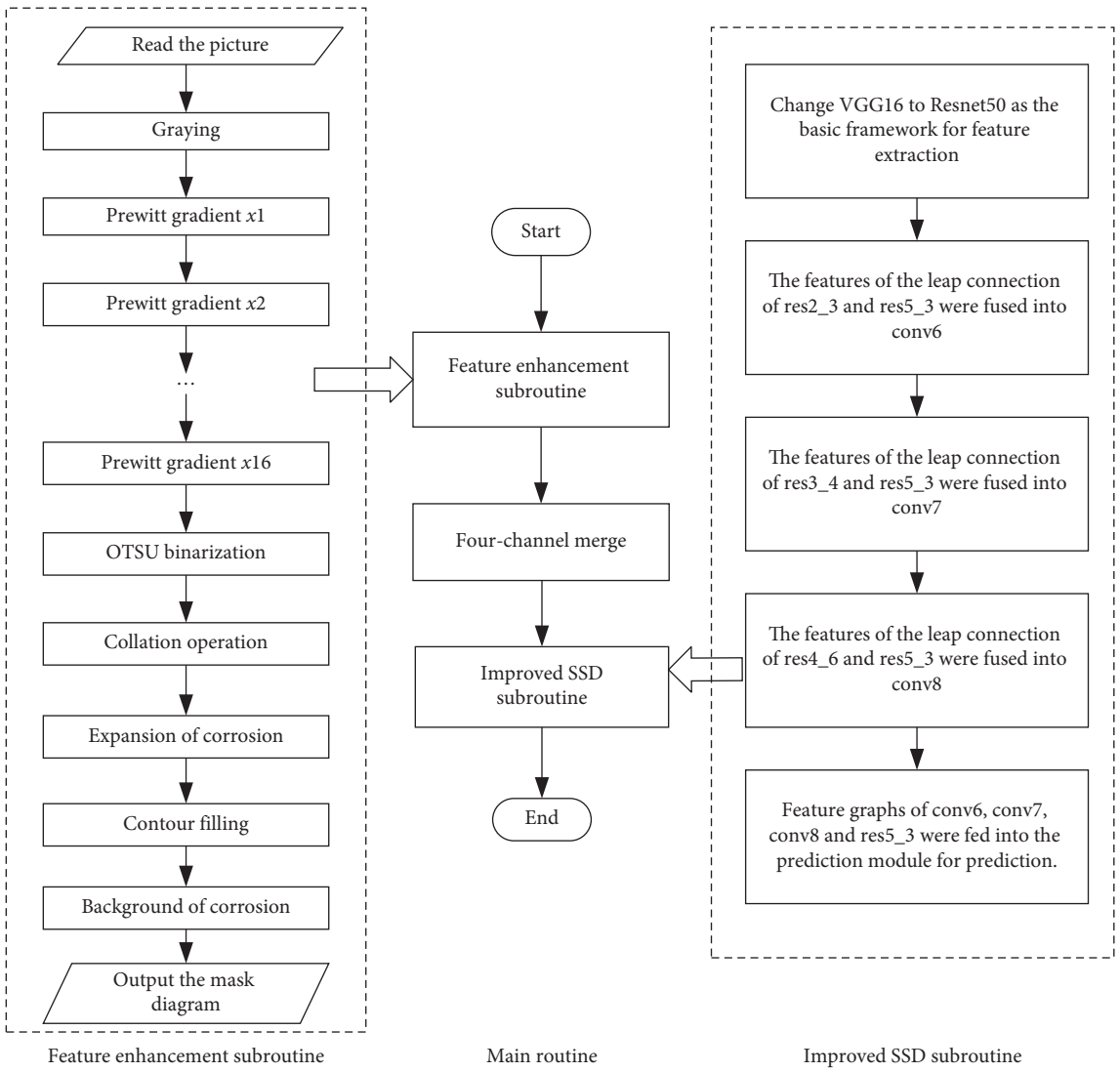

Figure 6: Underwater image detection flow chart.

Figure 7 shows the image after the Laplace operator single-direction, Prewitt operator 2-direction, Prewitt operator 4-direction, Prewitt operator 8-direction, Prewitt operator 16-direction, and Sobel operator 2-direction edge detection process. The improved algorithm model of this paper is compared with the traditional SSD algorithm loss function, the x-axis represents the epoch, and the y-axis represents the training loss. The loss function of the traditional SSD algorithm fluctuates greatly. When the training reaches 200 cycles, the performance of the two algorithms is basically stable. After the image has undergone edge detection in multiple directions, the training loss of this algorithm is significantly lower than that of the SSD algorithm.

Table 1 shows the performance indicators of sea urchin recognition with and without feature enhancement for different test models. During underwater image edge extraction through multiple directions, the performance indicators of Sobel operator 2 directions, Prewitt operator 2 directions, Prewitt operator 4 directions, Prewitt operator 8 directions, Prewitt operator 16 directions, and Laplace operator single direction are compared. The data shows that the performance of the algorithm detection target framework [32] of the Sobel operator 2-direction edge detection has been improved to 82.9 AP, which is 9.5 percentage points higher than that without the edge channel. The performance of the algorithmic target detection framework in this paper of the Prewitt operator 4-direction and 8-direction edge detection is close, and among them, Prewitt operator 8-direction is improved to 82.3% AP which is 8.9% points higher than that without the edge channel. The performance of the algorithmic target detection framework in this paper of the Prewitt operator 16-direction is improved to 83.1% AP, which is 9.7% points higher than that without the edge channel, while the edge channel of the Prewitt operator 2-direction is less effective. Using the Prewitt operator 16-direction can better extract the edge features of the sea urchin and ultimately improve the detection accuracy in the later stage.

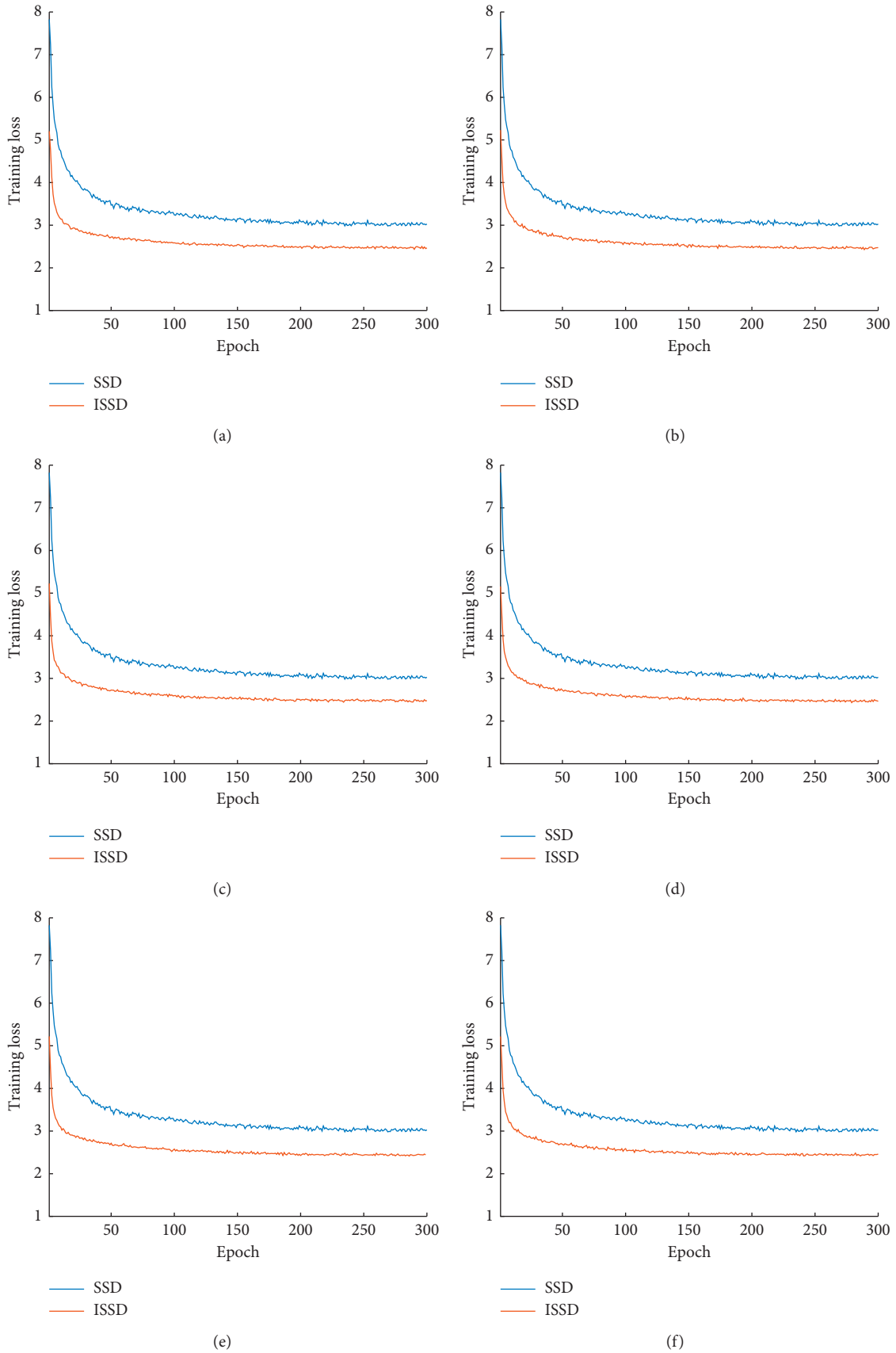

Figure 7: Comparison of loss functions of multidirection detections. (a) Loss function comparison of Laplace operator unidirectional. (b) Loss function comparison of Prewitt operator 2 directions. (c) Loss function comparison of Prewitt operator 4 directions. (d) Loss function comparison of Prewitt operator 8 directions. (e) Loss function comparison of Prewitt operator 16 directions. (f) Loss function comparison of Sobel operator 2 directions.

Table 1: Performance of sea urchin recognition under different test models with or without feature enhancement.

Model	Input size	(4th channel)	AP (%)
SSD	300 × 300	NO	73.4
ISSD	300 × 300	Sobel operator 2 directions	82.9
		Prewitt operator 2 directions	82.1
		Prewitt operator 4 directions	82.3
		Prewitt operator 8 directions	82.3
		Prewitt operator 16 directions	83.1
		Laplace operator unidirectional	82.4

4.2. Rationality Analysis of Feature Cross-Level Fusion. Analyzing the preliminary work of applying SSD to sea urchin detection, we think the main problem is that some sea urchins are small targets, and the traditional SSD algorithm has a relatively poor detection effect on small-sized objects. The feature map representation capability of shallow extraction is not strong enough. In this way, there will be misdetection and missed detection of small targets of the sea urchin. According to this, in order to improve the characteristics of the poor recognition ability of the SSD algorithm for small targets, the improved SSD algorithm draws on the idea of the residual network and uses Resnet50 instead of VGG16 as the basic framework of the network. Deepening the neural network by learning the residuals can avoid the problems of overfitting and the disappearance of the network gradient, learn more abstract texture features and semantic features, and strengthen the expression ability of features, so as to improve the ability of target classification and location. At the same time, a feature cross-level fusion method is proposed to improve the feature expression ability and strengthen the rational analysis of the semantic information feature cross-level fusion idea, mainly looking at the loss function and P-R curve during training. The convergence curve of the loss function during training is shown in Figure 8(a).

The training results of this algorithm are compared with the traditional SSD, RFBNet (Receptive Field Block Net) [33], FSSD (Feature Fusion Single-Shot Multibox Detector) [34], RefineDet (Single-Shot Refinement Neural Network for Object Detection) [35], and M2Det (Multilevel and Multiscale Detector) [36] algorithm, where the x-axis represents the period (epoch) and the y-axis represents the training loss. In the early stage of training, the RefineDet uses the Refine_multibox_loss, which is the second operation of the multibox_loss, so the convergence speed is slow; the other 5 curves use the multibox_loss as the loss function, and their convergence speeds are very fast. When the training reaches 200 cycles, the loss function of the original SSD algorithm value remains stable and no longer converges, maintaining a high loss value, and the positioning and classification losses are very large; RFBNet, FSSD, and M2Det algorithms are based on the SSD algorithm and all have a certain network architecture and feature fusion optimization to obtain better training effects; even if it reaches 200 cycles, the loss function continues to converge; the ISSD algorithm proposed in this paper adopts a method of feature cross-level fusion, which can ensure that objects of smaller scales will not have the problem of target disappearing after convolutions in deeper network layers, and it can provide improvement for the performance of small targets in the later period. It was very helpful and achieved better results during the training phase.

The AP value during training is shown in Table 2. The P-R curve is shown in Figure 8(b). The SSD algorithm has a low classification accuracy rate and recall rate index, and the highest recall rate is only 0.83. In contrast, although the highest recall rate of the RFBNet is similar to that of the SSD, its classification accuracy is higher. That is, the sea urchin in the RFBNet-based sea urchin identification system has higher confidence. The RefineDet uses the quadratic multibox_loss as the loss function, and it is similar to the Faster R-CNN detection method. In the first stage, it performs two classifications to filter a large number of samples, and then, in the second stage, it performs multiclassification to obtain the detection results, which greatly improves the accuracy rate and reduces the recall rate. The highest recall rate is only 0.80, and the lowest accuracy rate is as high as 0.72. The FSSD and M2Det algorithms have obvious advantages over the SSD and the RFBNet. The recall rate reaches 0.86, but it is difficult to achieve the desired performance. The ISSD algorithm proposed in this paper combines the advantages and disadvantages of the abovementioned three algorithms and has made a series of improvements. Finally, the performance of the algorithm has been greatly improved. Figure 8(c) shows the test results of the underwater target detection test set, and it can be seen from the figure that as the training period becomes larger, the detection accuracy of the six algorithms is constantly improving, the detection performance of the SSD algorithm fluctuates greatly, and its convergence is the fastest. In 200 training cycles, the performance of the six algorithms is basically stable. The detection accuracy of the algorithm in this paper is significantly better than that of the SSD algorithm, and the accuracy of the final test reaches 0.81. Table 2 is the AP performance index obtained by the ISSD algorithm model and the SSD, RFBNet, FSSD, RefineDet, and M2Det target detection model on the sea urchin test data set. The algorithm proposed in this paper has made a series of improvements, and finally, the performance of the algorithm has been further improved. The original SSD algorithm did not obtain the semantic information of the context due to the direct prediction of the multilayer feature map. When the confidence level is low, the recall rate is lower than that of the other three algorithms. When the confidence level is increased, the overall recall rate converges very quickly, and the confidence level of the sea urchin obtained by using the SSD algorithm is very low.

4.3. Analysis of Experimental Results. Analyzing the preliminary work of applying SSD to sea urchin detection, we believe that the sea urchin has several important features such as being black, round, and spiny, and we think that deep learning should have no pressure on the extraction of black and round features. Due to the different angles of the thorns and different convolution sizes, it may be that deep learning has the possibility of further improving the detection effect of thorns. Therefore, this paper proposes a sea

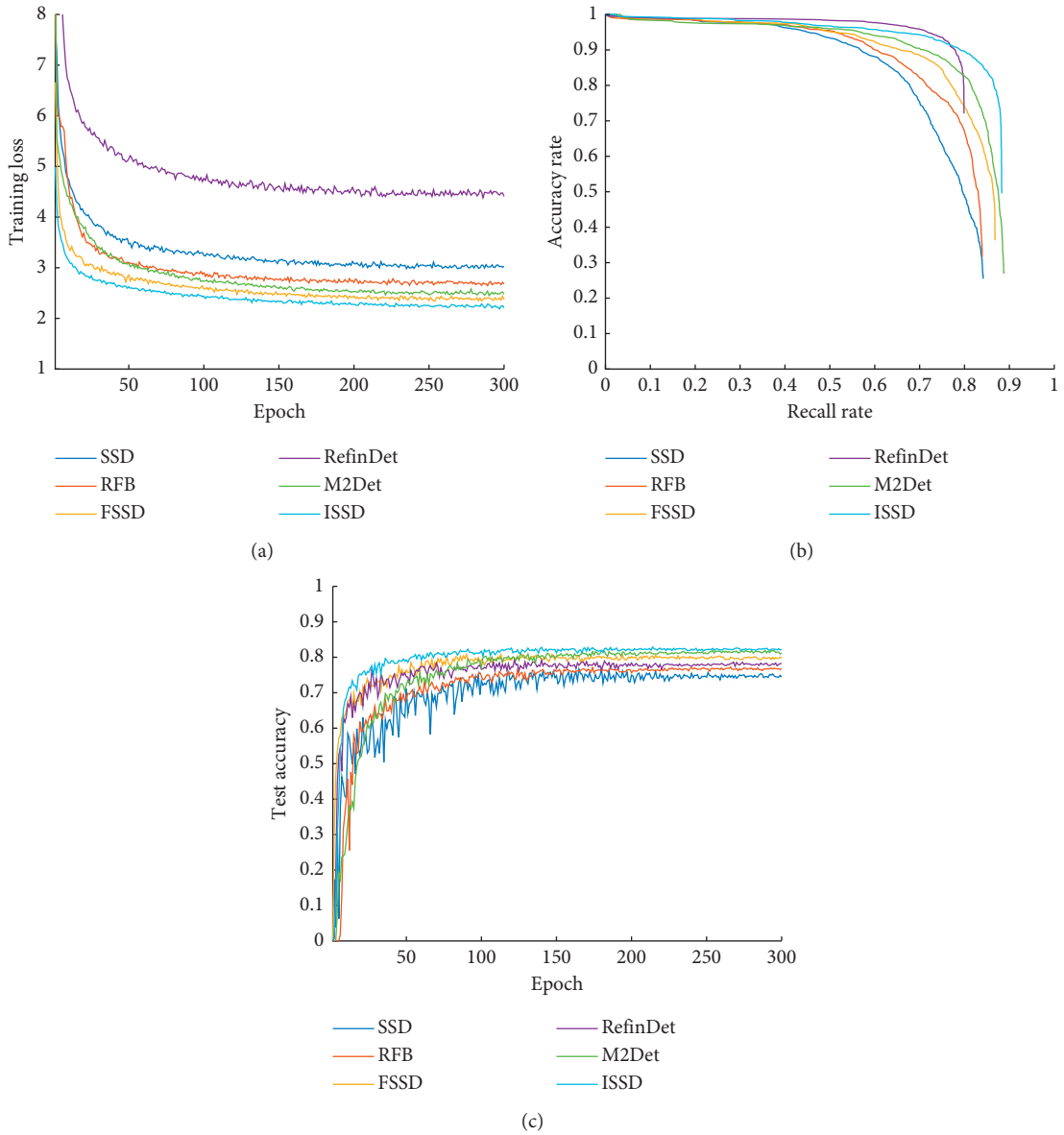

Figure 8: (a) Network loss function diagram; (b) P-R curve; and (c) accuracy curve of the test set.

Table 2: Detection and identification performance of different test models on the sea urchin test set.

Model	Input size	Model basic architecture	AP (%)
SSD	300 × 300	VGG16	73.4
RFB	300 × 300	VGG16	76.7
FSSD	300 × 300	VGG16	80.1
RefineDet	320 × 320	VGG16	78.6
M2Det	320 × 320	VGG16	80.4
ISSD	300 × 300	Resnet 50	81.0

urchin detection algorithm based on feature enhancement. According to the spiny-edge characteristics of sea urchin, a multidirectional edge detection algorithm is proposed to enhance the feature, which is taken as the 4th channel of image and the original 3 channels of underwater image together as the input for the further deep learning. At the same time, we think the main problem is that some sea urchins are small targets, and the traditional SSD algorithm has a relatively poor detection effect on small-sized objects. The feature map representation capability of shallow extraction is not strong enough. In this way, there will be misdetection and missed detection of small targets of the sea urchin. According to this, in order to improve the characteristics of the poor recognition ability of the SSD algorithm for small targets, the improved SSD algorithm draws on the idea of the residual network and uses Resnet50 instead of VGG16 as the basic framework of the network. Deepening the neural network by learning the residuals can avoid the problems of overfitting and the disappearance of the network gradient, learn more abstract texture features and semantic features, and strengthen the expression ability of features, so as to improve the ability of target classification

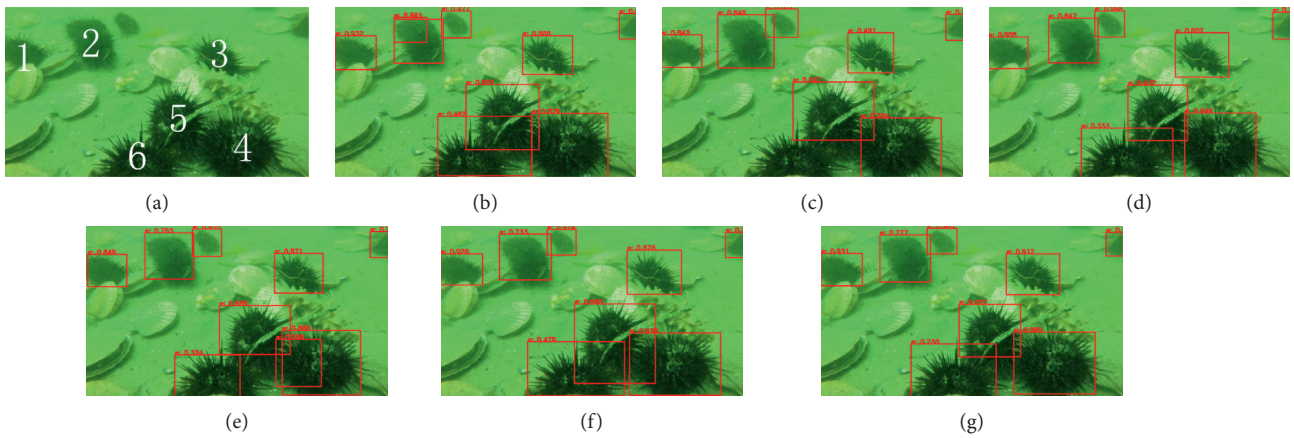

Figure 9: Multidirectional detection target detection result graph of ISSD, (a) the confidence test graph of sea urchin (Laplace operator unidirectional), (c) the confidence test graph of sea urchin (Prewitt operator 2 directions), (d) the confidence test graph of sea urchin (Prewitt operator 4 directions), (e) the confidence test graph of sea urchin (Prewitt operator 8 directions), (f) the confidence test graph of sea urchin (Prewitt operator 16 directions), and (g) the confidence test graph of sea urchin (Sobel operator 2 directions).

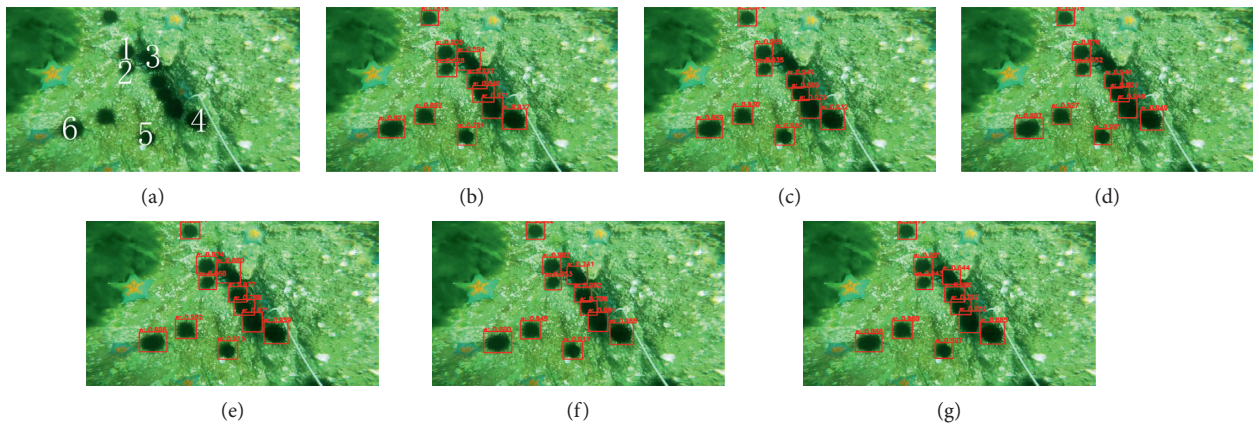

Figure 10: Multidirectional detection target detection result graph of ISSD, (a) the confidence test graph of sea urchin (Laplace operator unidirectional), (c) the confidence test graph of sea urchin (Prewitt operator 2 directions), (d) the confidence test graph of sea urchin (Prewitt operator 4 directions), (e) the confidence test graph of sea urchin (Prewitt operator 8 directions), (f) the confidence test graph of sea urchin (Prewitt operator 16 directions), and (g) the confidence test graph of sea urchin (Sobel operator 2 directions).

and location. Figures 9 and 10 are the effect diagram of the improved SSD algorithm model detection after performing the edge detection processing of the Sobel operator 2-direction, Prewitt operator 2-direction, Prewitt operator 4-direction, Prewitt operator 8-direction, Prewitt operator 16-direction, and Laplace operator single-direction. Obviously, it can be found from the figure that, after performing multidirectional edge detection on the image proposed in this paper, the algorithm of this paper has better performance for detecting small targets. In summary, the multidirectional detection algorithm proposed in this paper has better performance in sea urchin detection.

Table 3 shows the sea urchin confidence of the ISSD algorithm for multidirectional detection in Figure 9(a). Table 4 shows the sea urchin confidence of the ISSD algorithm for multidirectional detection in Figure 10(a). In the table, the values that are bold and underlined are the best and the values that are underlined are the 2nd best. It can be clearly found that the Prewitt operator 16-direction edge detection has the highest sea urchin confidence. The edge detection effect of the Prewitt operator in the 2 directions is poor, and there will be cases of missed detection.

Analyzing the preliminary work of applying SSD to sea urchin detection, we believe that the sea urchin has several important features such as being black, round, and spiny. According to the spiny-edge characteristics of sea urchin, a multidirectional edge detection algorithm is proposed to enhance the feature. The comparison of data in Tables 3 and 4 can more clearly illustrate the correctness of the sea urchin detection algorithm based on feature enhancement proposed in this paper, which is taken as the 4th channel of image and the original 3 channels of underwater image together as the input for the further deep learning. Using the Prewitt operator 16-direction can better extract the edge

Table 3: Sea urchin confidence in the ISSD algorithm for multidirectional detection in Figure 9(a).

Scene 1 sea urchin confidence	Sea urchin 1	Sea urchin 2	Sea urchin 3	Sea urchin 4	Sea urchin 5	Sea urchin 6
Laplace operator unidirectional	**0.932**	0.753	**0.886**	0.879	0.552	**0.462**
Prewitt operator 2 directions	0.843	0.648	0.491	0.294	0.247	None
Prewitt operator 4 directions	0.905	0.647	0.802	0.949	0.432	0.333
Prewitt operator 8 directions	0.849	*0.783*	0.871	0.888	**0.669**	0.324
Prewitt operator 16 directions	**0.928**	0.733	0.878	**0.939**	**0.666**	**0.479**
Sobel operator 2 directions	0.931	**0.777**	**0.912**	**0.889**	0.622	0.230

Table 4: Sea urchin confidence in the ISSD algorithm for multidirectional detection in Figure 10(a).

Scene 2 sea urchin confidence	Sea urchin 1	Sea urchin 2	Sea urchin 3	Sea urchin 4	Sea urchin 5	Sea urchin 6
Laplace operator unidirectional	0.903	0.828	0.594	0.917	0.791	**0.923**
Prewitt operator 2 directions	0.969	0.935	None	0.932	**0.930**	0.829
Prewitt operator 4 directions	0.976	0.952	None	0.949	0.901	0.887
Prewitt operator 8 directions	0.974	**0.968**	**0.680**	**0.959**	0.913	0.886
Prewitt operator 16 directions	**0.982**	**0.953**	**0.741**	**0.966**	0.917	**0.889**
Sobel operator 2 directions	**0.981**	0.943	0.644	0.865	**0.927**	0.838

features of the sea urchin and ultimately improve the detection accuracy in the later stage.

5. Conclusions

Autonomous detection and fishing by underwater robots will be the main way to obtain aquatic products in the future, and sea urchin is the main research object of aquatic product detection. The preliminary work of applying classic SSD to sea urchin detection, the existing shortcomings of inaccurate detection of small sea urchin targets, and the overall detection performance of sea urchin have room for further improvement. Therefore, this paper uses feature enhancement methods to enhance the analysis ability of deep learning on the feature learning of the thorny edge and improve the performance of detection and recognition of sea urchins. We used resnet 50 as the basic architecture for network feature extraction to replace the original VGG16 of the SSD algorithm.

ccording to the analysis of experimental data, the improvement of the classic SSD in this paper effectively improves the ability of the SSD in the sea urchin recognition task. However, the improved model still has shortcomings and will not meet the real-time requirements, mainly because the multidirectional edge detection has a large calculation amount and a long running time in image processing, so the algorithm is optimized and the calculation amount is compressed to meet the real-time requirements, and use of image enhancement and target detection for underwater robots, real-time detection, and recognition of sea urchins by underwater robots will be the focus and main direction of the next research.

Data Availability

The code used to support the findings of this study are available from the corresponding author upon request (nuistpanda@163.com, 001600@nuist.edu.cn). The data are from the open data set of the National Natural Science Foundation of China Underwater Robot Competition (http://www.cnurpc.org/a/xwjrz/2019/0808/129.html).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

All authors drafted the manuscript and read and approved the final manuscript.

Acknowledgments

The research in this article was supported by the National Natural Science Foundation of China (61773219 and 61701244) and the key special project of the National Key R&D Program (2018YFC1405703), and the authors would like to express their heartfelt thanks.

References

[1] A. Olmos and E. Trucco, "Detecting man-made objects in unconstrained subsea videos," in *Proceedings of the British Machine Vision Conference*, pp. 1–10, Cardiff University, Cardiff, UK, September 2002.

[2] M. Xia, W. A. Liu, Y. Xu, K. Wang, and X. Zhang, "Dilated residual attention network for load disaggregation," *Neural Computing and Applications*, vol. 31, no. 12, pp. 8931–8953, 2019.

[3] M. Xia, W. Liu, B. Shi, L. Weng, and L. Jia, "Cloud/snow recognition for multispectral satellite imagery based on a multidimensional deep residual network," *International Journal of Remote Sensing*, vol. 40, no. 1, pp. 156–170, 2019.

[4] J. Qian, M. Xia, and X. Yue, "Parallel knowledge acquisition algorithms for big data using MapReduce," *International Journal of Machine Learning & Cybernetics*, vol. 1, no. 2, pp. 1–15, 2015.

[5] L. Weng, X. Sun, M. Xia, J. Liu, and Y. Xu, "Portfolio trading system of digital currencies: a deep reinforcement learning with multidimensional attention gating mechanism," *Neurocomputing*, vol. 402, pp. 171–182, 2020.

[6] L. Weng, Y. Xu, M. Xia, Y. Zhang, J. Liu, and Y. Xu, "Water areas segmentation from remote sensing images using a

separable residual Segnet network," *ISPRS International Journal of Geo-Information*, vol. 9, no. 4, p. 256, 2020.

[7] Y. Bengio, "Deep learning of representations: looking forward," in *Proceedings of the International Conference on Statistical Language and Speech Processing*, Springer, Berlin, Germany, pp. 1–37, July 2013.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection,"vol. 1, pp. 886–893, in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, IEEE, San Diego, CA, USA, June 2005.

[10] T. Joachims, "Making large-scale SVM learning practical," Technical Report 1998, 28, MIT Press, Cambridge, MA, USA, 1998.

[11] D. D. Margineantu and T. G. Dietterich, "Pruning adaptive boosting," in *Proceedings of the Fourteenth International Conference on Machine Learning*, vol. 97, pp. 211–218, Nashville, TN, USA, July 1997.

[12] M. Xia, W. Song, X. Sun, J. Liu, T. Ye, and Y. Xu, "Weighted densely connected convolutional networks for reinforcement learning," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 4, Article ID 2052001, 2020.

[13] R. Girshick, J. Donahue, T. Darrell et al., "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Columbus, OH, USA, June 2014.

[14] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, Santiago, Chile, December 2015.

[15] S. Ren, K. He, R. Girshick et al., "Faster R-CNN: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 39, no. 6, pp. 1137–1149, 2015.

[16] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multibox detector," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Springer, Amsterdam, The Netherlands, pp. 21–37, October 2016.

[17] J. Redmon, S. Divvala, R. Girshick et al., "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, Las Vegas, NV, USA, June 2016.

[18] M. Bakratsas, P. Basaras, and D. Katsarosb, "Take me to SSD : a hybrid block-selection method on HDFS based on storage type," in *INNS Conference on Big Data*, pp. 111–119, Elsevier, Amsterdam, Netherlands, 2016.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.

[21] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang, "Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 10, pp. 2421–2433, 2015.

[22] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 315–323, Fort Lauderdale, FL, USA, April 2011.

[23] X. Li, H. Hu, L. Zhao et al., "Image recovery method combining histogram stretching for underwater imaging," *Scientific Reports*, vol. 8, no. 1, pp. 1–10, 2018.

[24] J. Liang, L. Ren, E. Qu, B. Hu, and Y. Wang, "Method for enhancing visibility of hazy images based on polarimetric imaging," *Photonics Research*, vol. 2, no. 1, pp. 38–44, 2014.

[25] C. Liu, J. Zhao, Y. Shen, Y. Zhou, X. Wang, and Y. Ouyang, "Texture filtering based physically plausible image dehazing," *The Visual Computer*, vol. 32, no. 6–8, pp. 911–920, 2016.

[26] J. Ahn, S. Yasukawa, T. Sonoda, T. Ura, and K. Ishii, "Enhancement of deep-sea floor images obtained by an underwater vehicle and its evaluation by crab recognition," *Journal of Marine Science and Technology*, vol. 22, no. 4, pp. 758–770, 2017.

[27] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 379–393, 2018.

[28] J. Y. Chiang and Y.-C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1756–1769, 2012.

[29] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, "Automatic red-channel underwater image restoration," *Journal of Visual Communication and Image Representation*, vol. 26, pp. 132–145, 2015.

[30] The National Natural Science Foundation of China Underwater Robot Competition, http://www.cnurpc.org/a/xwjrz/2019/0808/129.html.

[31] Y. Zhang, J. Zhang, P. J. Smith, M. Shafi, and P. Zhang, "Reduced complexity channel models for IMT-advanced evaluation," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, no. 1, pp. 1–13, 2009.

[32] P. Liu and Z. Li, "Task complexity: a review and conceptualization framework," *International Journal of Industrial Ergonomics*, vol. 42, no. 6, pp. 553–568, 2012.

[33] S. Liu, D. Huang, and Y. Wang, "Receptive field block net for accurate and fast object detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 385–400, Munich, Germany, September 2018.

[34] A. Belfodil, A. Belfodil, A. Bendimerad et al., "FSSD-a fast and efficient algorithm for subgroup set discovery," in *Proceedings of the 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Washington, DC, USA, October 2019.

[35] S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li, "Single-shot refinement neural network for object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June 2018.

[36] Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen, L. Cai et al., "M2Det: a single-shot object detector based on multi-level feature pyramid network," in *Proceedings of the AAAI Conference on Artificial Intelligence*, January 2019.