applied sciences

MDPI

Article

Underwater Object Detection Method Based on Improved Faster RCNN

Hao Wang * and Nanfeng Xiao

School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China
* Correspondence: kylewhao@aliyun.com; Tel.: +86-13533953032

Abstract: In order to better utilize and protect marine organisms, reliable underwater object detection methods need to be developed. Due to various influencing factors from complex and changeable underwater environments, the underwater object detection is full of challenges. Therefore, this paper improves a two-stage algorithm of Faster RCNN (Regions with Convolutional Neural Network Feature) to detect holothurian, echinus, scallop, starfish and waterweeds. The improved algorithm has better performance in underwater object detection. Firstly, we improved the backbone network of the Faster RCNN, replacing the VGG16 (Visual Geometry Group Network 16) structure in the original feature extraction module with the Res2Net101 network to enhance the expressive ability of the receptive field of each network layer. Secondly, the OHEM (Online Hard Example Mining) algorithm is introduced to solve the imbalance problem of positive and negative samples of the bounding box. Thirdly, GIOU (Generalized Intersection Over Union) and Soft-NMS (Soft Non-Maximum Suppression) are used to optimize the regression mechanism of the bounding box. Finally, the improved Faster RCNN model is trained using a multi-scale training strategy to enhance the robustness of the model. Through ablation experiments based on the improved Faster RCNN model, each improved part is disassembled and then the experiments are carried out one by one, which can be known from the experimental results that, based on the improved Faster RCNN model, mAP@0.5 reaches 71.7%, which is 3.3% higher than the original Faster RCNN model, and the average accuracy reaches 43%, and the *F*1-score reaches 55.3%, a 2.5% improvement over the original Faster RCNN model, which shows that the proposed method in this paper is effective in underwater object detection.

Keywords: underwater object detection; improved faster RCNN; Res2Net101; OHEM; GIOU; Soft-NMS

Citation: Wang, H.; Xiao, N. Underwater Object Detection Method Based on Improved Faster RCNN. *Appl. Sci.* **2023**, *13*, 2746. https://doi.org/10.3390/app13042746

Academic Editor: Sungho Kim

Received: 24 January 2023
Revised: 11 February 2023
Accepted: 15 February 2023
Published: 20 February 2023

1. Introduction

Marine biological resources on the earth are abundant and diverse [1], and it is especially important to effectively detect the growth of marine organisms and manage them scientifically in order to better understand, utilize and protect these limited marine resources. However, little is known about the seafloor, and darkness, high pressure, and extreme temperatures of the seafloor can cause many adverse effects on exploration efforts. For example, light decays quickly underwater and there is insufficient light, resulting in blurred and unclear images of marine organisms, and various noise and image distortion problems. In particular, there are various limitations of traditional manual means to explore underwater environments: they are time consuming, high cost, and have limited detection range and defects in acquired images. Therefore, how to conduct stable, reliable, and fast underwater object detection and monitoring in complex and harsh underwater environments is an urgent problem to be solved.

In view of complex underwater environments and the prevalence of image noises, this paper chose the Faster RCNN as the basis. Firstly, the initial image dataset is preprocessed and extended with data enhancement. Secondly, the feature extraction module of the Faster

RCNN is replaced by Res2Net101 [2]. The OHEM algorithm [3] is introduced to solve the imbalance of positive and negative samples of the bounding box. Thirdly, the regression mechanism of the bounding box is optimized. Finally, the training strategy is optimized so that the improved Faster RCNN model can detect the underwater objects accurately and effectively.

The main novel points of this paper are summarized as follows:

1. A two-stage underwater object detection method is proposed and improved, which is based on the Faster RCNN, and its initial image dataset is subjected to Mosaic [4] data enhancement and related preprocessing. The improved Faster RCNN model has stable performance and good accuracy, and is more conducive to underwater object detection.
2. VGG16 is replaced by Res2Net101 in the feature extraction module of the improved Faster RCNN, because the structure of Res2Net101 can enhance the expression ability of the receptive field in the network layer of the improved Faster RCNN. The OHEM algorithm is introduced to effectively solve the imbalance problem of the positive and negative samples of the candidate prior frame.
3. The IOU structure in the improved Faster RCNN is replaced by GIOU [5], and the non-overlapping area between the candidate prior frame and the real object is also taken into account, so that the weakness of the original IOU can be weakened and the improved Faster RCNN can better optimize the candidate prior frame in the training process.
4. The original NMS algorithm is replaced by the Soft-NMS algorithm [6], which only needs simple modification and no additional parameters, is easy to realize and can be easily applied to different object detection algorithms. The training strategy of the improved model is optimized by a multi-scale training approach [7], which can improve the robustness of the improved Faster RCNN detection algorithm to different object sizes.

2. Related Work

2.1. Development of Underwater Object Detection Technology

With the rapid development of deep learning methods, it is widely used in various fields, e.g., [8–14]. The successful cases aptly illustrate that deep learning methods can be fully applied to detect marine organisms and their growth environments. The deep learning-based object detection methods are generally divided into two main technical routes: one-stage object detection algorithms and two-stage object detection algorithms. Regarding the one-stage object detection methods, R. Joseph et al. invented the one-stage object detection algorithm YOLO in 2015 [15], which excels in the speed of detecting images. W. Liu proposed the SSD algorithm [16], which is based on the principle of detecting objects at different scales by setting different detection branches in the network. Subsequently, through the continuous efforts, YOLOV2 [13], YOLOV3 [17], YOLOV4 [18] and other algorithms of the YOLO series appeared one after another. Regarding the two-stage object detection algorithm, in 2014 R. Girshick firstly proposed the RCNN algorithm [19], which has good performance in detection accuracy but runs very slowly due to too many complex calculations. In the same year, Kaiming He invented the SPPNet [20] algorithm with spatial pyramidal pooling structure, which is 20 times faster than RCNN with little change in the accuracy. R. Girshick introduced the Fast RCNN [21] algorithm with substantial improvement in the speed of training and testing in 2015. Shaoqing Ren invented the Faster RCNN [22] algorithm, and Khasawneh N. also used the faster R-CNN and deep transfer learning to realize the detection of K-complexes in EEG waveform images [23], which are more accurate, faster, and very close to real-time performance.

On the other hand, Kashif Iqbal proposed an unsupervised color correction method for underwater image enhancement [24], which optimized the problems of low contrast and color distortion in the underwater images. David Zhang [25] proposes a robust and unsupervised deep learning algorithm to automatically detect fish, thereby reducing the burden of manual annotation. Fei Yuan [26] uses the LSTM (Long Short-Term Memory)

neural network for water quality classification, which improves the detection accuracy and real-time performance of the water quality monitoring system. Jung-Hua [27] proposed a method for detecting abnormal behavior of underwater fish by combining deep learning object detection, fish tracking and DTW (Dynamic Time Warping).

2.2. Network Structure of the Faster RCNN

The Faster RCNN usually works in three main parts, which are feature information extraction from input image, bounding box generation, classifier classification and object position correction by regressor. The overall structure of the faster RCNN is shown in Figure 1. The input image can be implemented by a convolutional neural network for feature information extraction, and the obtained convolutional feature information is used as the input of the RPN (Region Proposal Network), which in turn produces the region proposals. The main function of the regression layer is to predict the region proposal parameters corresponding to the anchor points of the bounding box, while the main function of the classification layer is to determine whether the object in the bounding box is an object or a background. The proposed regions generated by RPN can be mapped to the convolutional feature map according to their positions to form an ROI (Region of Interest) [28]. Then, the pooling operation of the ROI is performed to partition the mapped ROI regions into blocks of the same size, and then the maximum pooling operation is performed to adjust the size of the bounding boxes in each region. In the next step, the information of the bounding box in each region is transmitted to the next layer of the network, i.e., the fully connected layer, and after the fully connected layer the label classification score and the position of the corrected bounding box can be output by the softmax function [29].

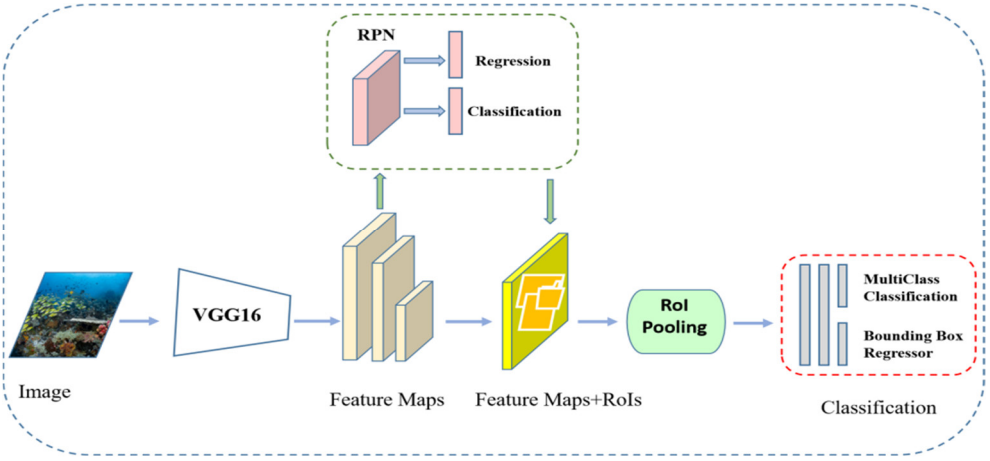


Figure 1. Network structure diagram of the Faster RCNN.

2.3. Loss Function of the Faster RCNN

The loss function of the Faster RCNN has two main components: (1) classification loss and bounding box regression loss of the regional network; (2) classification loss and loss of bounding box position correction at detection. The loss function of the Faster RCNN can be defined by Equation (1) [30].

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

where i is an index of anchor numbers in each small batch of data, and p_i is a prediction probability of the anchors to objects. p_i* is a label value of a category, which can be 0 or 1, 0 means a false object, and vice versa is a true object. λ is a weight coefficient, N_{cls} is a classification loss parameter, N_{reg} is regression loss parameter, L_{cls} is classification loss in object detection, and $L_{cls}(p_i, p_i$*) is the logarithmic loss between the detection object and

non-object, which is calculated as shown in Equation (2) [30]. R is a function named as $smooth_{L1}(x)$, and the expression is shown in Equation (3) [31]:

$$L_{cls}(p_i, p_i^*) = -\log[p_i^* p_i + (1 - p_i^*)(1 - p_i)] \quad (2)$$

$$smooth_{L1}(x) = \begin{cases} 0.5x^2 & if|x| < 1 \\ |x| - 0.5 & otherwise \end{cases} \quad (3)$$

where $L_{reg}(t_i, t_i^*)$ is the regression loss inside the object detection, which is represented by $L_{reg}(t_i, t_i^*) = R(t_i - t_i^*)$, t_i refers to the predicted coordinates, t_i^* is the coordinates of the detection object, and $x = t_i - t_i^*$.

3. Proposed Method

3.1. Improvement of the Faster RCNN

In this section, the network structure of the two-stage object detection algorithm Faster RCNN is improved. The backbone network module is selected as the backbone network by comparing VGG16Net [32], ResNet50 [33], ResNet101 [34], Res2Net101 [35], and Res2-Net101 feature extraction modules. The GIOU function is used to replace the original IOU. GIOU can overcome the shortcomings of the original IOU well, and when the predicted bounding box and the actual truth box do not overlap at all the GIOU still works and the model can continue to be optimized. In order to optimize the class imbalance problem of the data, the OHEM algorithm is introduced. After the feature map goes through the RPN module, a large number of preselected boxes are generated and the Soft-NMS algorithm is used in the screening of bounding boxes. The network structure of the improved Faster RCNN is shown in Figure 2.

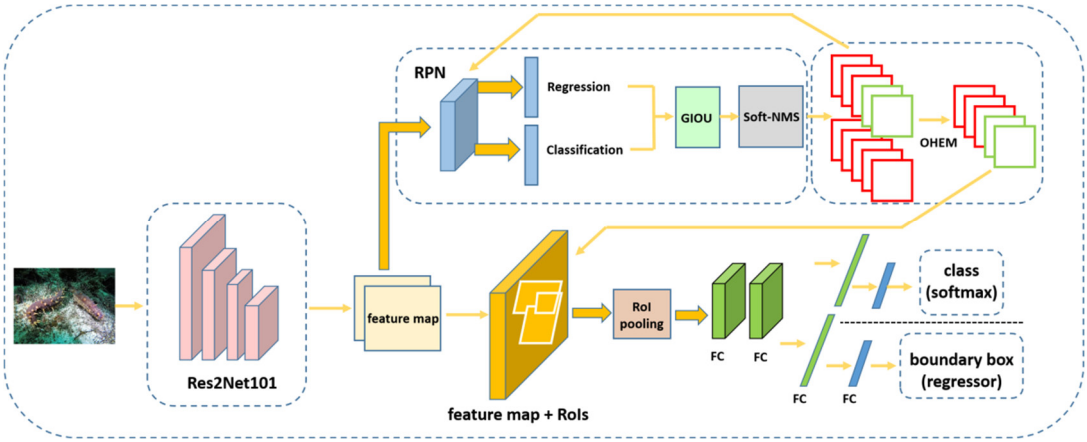


Figure 2. The network structure of the improved Faster RCNN.

3.2. Initial Screening of Bounding Boxes

In order to further improve the performance of the underwater object detection, a *k*-means++ algorithm [36] with *k*-means algorithm [37] is applied to implement the clustering of the bounding boxes sizes of the images in the experimental dataset. The *k*-means algorithm is a simple clustering algorithm, the algorithm is easy to implement, and the clustering results are good. However, the *k*-means algorithm calculates the location of the *k* original clustering centers, which has an important impact on the final clustering results as well as the execution time. Therefore, it is necessary to select the appropriate *k* initial clustering centers. The *k*-means++ algorithm is a further improvement of this problem. The *k* clustering centroids derived from the filtering by the *k*-means++ algorithm are considered as the initial values for the *k*-means algorithm to run. Finally, nine values are generated by the *k*-means algorithm as the initial values of the training model bounding boxes, which are 49 × 33, 52 × 75, 60 × 43, 66 × 60, 90 × 49, 103 × 89, 142 × 110, 192 × 243 and 459 × 401.

3.3. Improved Backbone Network

VGG16Net is the main module of the backbone network in the original Faster RCNN, which is used for feature information extraction from the input image. The ResNet [38] network improves the feature extraction network VGG19Net by adding a residual module with a short-circuit jump mechanism, while the network structure is changed by performing down-sampling using a convolution with a step size of two, while the original fully connected layer is changed to use global average pooling layer. The most common ResNet network structures are ResNet50 and ResNet101, and ResNet101 is a further stacking of the fourth layer of convolutional blocks on top of the structure of ResNet50, increasing the number of layers from the original six to twenty-three. Res2Net [39] creates a multi-layered residual-like network connection inside the original residual module, which enables a finer region to describe multi Res2Net, builds a multi-layered residual class network connection inside the original residual module, which can describe the multi-scale feature information in a finer area, and greatly improve the feature expression capability of the network layers.

In the experiments in Section 4 of this paper, the Faster RCNN (VGG16), the Faster RCNN (ResNet50), the Faster RCNN (ResNet101), and the Faster RCNN (Res2Net101) network structures are all tested through experiments, and according to the results of the comparison experiments the Res2Net101 network structure has the best detection effect; therefore, the Res2Net101 is chosen as the backbone network structure of the improved Faster RCNN.

3.4. Positive and Negative Sample Imbalance Improvement

In the process of training the model by the Faster RCNN, a large number of the bounding boxes will be generated when the image feature information passes through the RPN network structure. In the Faster RCNN, the value of the IOU is used to determine the positive and negative attributes of the bounding box samples; generally the bounding box with the IOU value greater than 0.7 is a positive sample, while the bounding box with the IOU value less than 0.3 is a negative sample. In these large numbers of bounding boxes there is often a serious imbalance between the number of the positive and negative samples, and in general the number of negative samples is much higher than the positive samples. Among the samples of these bounding boxes, there are often some difficult negative samples, and the difficult negative samples are samples that are difficult to identify and distinguish with the model. Due to the imbalance of the positive and negative samples and the existence of the difficult samples, the detection accuracy of the training model is affected.

The OHEM is a mining algorithm for hard-to-detect samples, which is often used in image recognition and image detection tasks. In the two-stage object detection model, the OHEM generally acts between the bounding box generation module and the bounding box classification and the regression module. The OHEM algorithm usually selects difficult negative samples with complexity and high loss for training. Under limited conditions, the input positive and negative samples are cyclically sampled and the initial samples are redistributed and combined to form new samples. In this process, it will help the model to pay attention to those difficult negative samples and improve the detection effect of the model. On the other hand, this operation will ignore those simple samples, thereby changing the distribution of the entire positive and negative sample input.

3.5. Improvement of Bounding Box Mechanism

3.5.1. GIOU

The IOU is a very commonly used indicator in the object detection process, but the IOU is the concept of ratio; therefore, it does not pay attention to the area size and shape of the object and the bounding box. During the model training process of the object detection algorithm, a large number of bounding boxes will be formed and it is possible that the loss values of the bounding boxes are the same, but there are different IOU values. Therefore, the original IOU has two problems: (1) the prediction range of the bounding box and the

real object do not overlap, the value of the IOU is always 0, and it cannot be optimized; (2) the original IOU cannot discern the different alignments between the bounding boxes and the ground-truth objects.

In order to solve the above two problems, a few related research has proposed the GIOU algorithm. The GIOU also takes into account the non-overlapping areas of the bounding box and the real object, which can weaken the weakness of the original IOU. This paper replaces the IOU structure in the Faster RCNN algorithm with the GIOU, and the algorithm model can better optimize the bounding box during the training process, thereby improving the performance of the Faster RCNN model.

3.5.2. Soft-NMS

In the process of underwater biological object detection, the two-stage object detection algorithm will generate multiple bounding boxes for the same underwater bio-logical object in the image, and will generate confidence scores. There is overlap between the regions of these bounding boxes. In order to filter out qualified candidate prior boxes and eliminate redundant candidate priori boxes, the traditional NMS algorithm will select the bounding box A with the highest score. Then, it will calculate the IOU values of other bounding boxes that have significant overlap with A. If the IOU value exceeds the set threshold, it will be deleted. If the IOU value is lower than the set value it will be retained, and the bounding boxes that do not overlap will also be retained. However, according to the operation of the traditional NMS, the confidence score will be low and the bounding box with the IOU value greater than the set value will be deleted, which may not detect underwater objects.

The Soft-NMS algorithm does not simply remove those bounding boxes whose IOU is greater than the threshold, but only reduces the confidence in them. The Soft-NMS algorithm only needs to make simple modifications to the original NMS algorithm without adding additional parameters, and therefore it is easy to implement and can be easily applied to different object detection algorithms. The calculation formula of the Soft-NMS algorithm is shown below [40]:

$$S_i = \begin{cases} S_i & iou(M, b_i) < N_t \\ S_i(1 - iou(M, b_i)) & iou(M, b_i) \geq N_t \end{cases} \tag{4}$$

where M represents the candidate prior frame with high score, b_i is the candidate prior frame to be processed, N_t is the set IOU threshold, and the IOU function is a weight function for attenuation, which is used to attenuate the scores of the adjacent candidate prior frames overlapping with the candidate prior frame with the high score. The higher the degree of overlap, the greater the attenuation degree of the scores, and S_i is the score of the *ith* candidate prior frame in the S set.

In the improved Faster RCNN underwater object detection scheme, the original NMS algorithm is replaced by Soft-NMS to improve the detection performance of the model.

3.6. Multi-Scale Training

Multi-scale training can improve the performance of the network model. Generally, several parameters of different scales are set. After a specific number of iterations during training, the parameters can be randomly selected from the preset parameters. Selecting a certain scale as the standard of the input image size and then training can improve the robustness of the model to objects with different sizes. The resolution of the input image can affect the detection performance of the trained model. In the feature extraction network, a feature map with a size several times smaller than the original image is often generated, which makes it more difficult for the feature description of the small object to be captured by the detection network. Therefore, the robustness of the algorithm model can be improved to a certain extent by providing larger and richer images for training. In this paper, the multi-scale training method is used in the experiments of training the model. The length and width of the sample images in the training set are randomly changed during training. The image length varies from 350 to 700, and the image width varies from 250 to 600.

4. Experiments

The improved Faster RCNN underwater object detection is tested on a Linux system, and the development language is Python3.7. In the hardware configuration, the CPU is Xeon Gold 5218R, the memory is 32 G, the graphics card GPU is Tesla V100, the video memory is 16 G, the CUDA (Compute Unified Device Architecture, NVIDIA, Santa Clara City, CA, USA) version is 11.2, and the PyTorch (Facebook, Park, CA, USA) version is 1.10. The training, testing, and validation datasets used in the experiments are the underwater environment images collected through the Internet. The initial dataset is filtered, preprocessed, and data enhanced, and the number of images in the image dataset is expanded to 3500, and the resolution of the images is uniformly adjusted to 500 × 400.

4.1. Mosaic Data Enhancement and Related Preprocessing

4.1.1. Dataset Preprocessing

There were 2372 underwater environment image samples (for example, holothurian, echinus, scallop, starfish and waterweeds) collected initially in the experiments. For deep learning network model training, the number of samples in the experimental dataset is not sufficient. In order to overcome the problem of an insufficient number of dataset samples, this paper expands and optimizes the dataset samples by performing data enhancement techniques such as flipping, local trimming, color adjustment, Gaussian noise and salt and pepper noise on the images in the original dataset. The image flipping process is to flip the original image of the experimental data sample horizontally or vertically; the cropping operation is to locally cut the image, select the object position, and remove the irrelevant or disturbing parts. In order to enhance the robustness of the underwater object detection model and prevent overfitting, this paper actively introduces Gaussian noise and salt and pepper noise to some images during the data enhancement processing.

Since the color components of underwater imaging are different from those of land-based imaging, the underwater images are prone to imbalance in the main three colors. Therefore, color compensation is performed on the underwater images to adjust the image tones and to achieve better underwater image detail features. Since the color is attenuated when the light enters underwater, the red component is most obviously attenuated; therefore, the red component in the collected underwater environment images is small, while the information in the blue and green channels is relatively well preserved. Therefore, we mainly compensate the red color in the image to make the image closer to the real color. The data enhancement effect is shown in Figure 3. After the pre-processing and expansion of the dataset samples, the experimental dataset sample images are expanded from the original 2372 to 3500 and the resolution of the images is unified to 500 × 400.

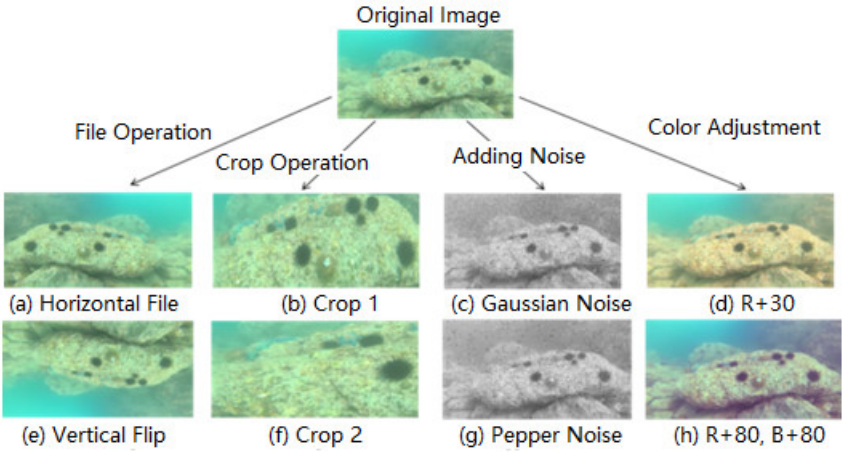
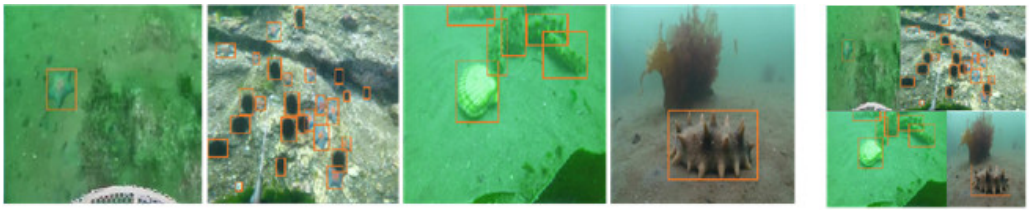


Figure 3. The effect of dataset preprocessing.

4.1.2. Mosaic Data Enhancement

The Mosaic data enhancement technique is a random selection of four images, and the selected images are stitched together by performing a random arrangement of layout, random zooming in and out, and random local trimming, as shown in Figure 4. This technique makes it possible to detect more types of sample datasets. In particular, for the random shrinking or zooming operation on the images, this data processing strategy is able to add more small volume objects, allowing a better robustness of the network. On the other hand, this method of stitching images by Mosaic enables the network to calculate the data of four images together at the time of detection, and therefore the batch processing image values of the training model can be set to be smaller, which can reduce the GPU resources.



(**a**) (**b**)

Figure 4. Mosaic data enhancement, (**a**) four randomly entered pictures (from left to right, they are respectively a starfish, water plants, a scallop, a sea cucumber), (**b**) data augmentation results.

4.2. Evaluation Indicators

The evaluation metrics of the object detection model firstly need to classify the recognition results into four categories according to the true labels; *TP* is the true example, *TN* is the true counter example, *FP* is the false positive example, and *FN* is the false counter example. The above four categories can be used to find the precision rate and recall rate of the model. The formula for calculating the precision rate is shown in Equation (5) [41], the formula for calculating the recall rate is specified in Equation (6) [41], and the formula for the *F*1 score is defined in Equation (7) [42].

$$\text{Pre}cision(P) = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Re}call(R) = \frac{TP}{TP + FN} \tag{6}$$

$$F1 = \frac{2PR}{P + R} \tag{7}$$

In Equation (5), *P* is a precision rate, and *R* is a recall rate. In order to balance the precision rate as well as the recall index, *mAP* (Mean Average Precision) is chosen as the comprehensive evaluation index of the model. The average of the *AP* values of all categories is the *mAP* of all categories of objects.

4.3. Comparative Experiments of Backbone Network Selection

In this paper, the Res2Net101 network module is selected as the feature extraction structure of the improved Faster RCNN in order to optimize its backbone network for better performance in the underwater object detection. In this section, many comparison experiments are conducted for the Faster RCNN (VGG16), the Faster RCNN (ResNet50), the Faster RCNN (ResNet101) and the Faster RCNN (Res2Net101), respectively.

The 3500 underwater environment images are divided into training data samples, validation data samples and test data samples, with a ratio of 7:2:1. After 300 epochs of iterative training and learning, the learning rate drops to 0.0001 and the parameters of the underwater object detection model converge. The improved Faster RCNN improves the performance in the underwater object detection by combining different feature extraction modules. The results of the comparative experiments are shown in Table 1.

Table 1. Performance of different feature extraction modules in the underwater object detection.

Methods	Avg-P(%)	mAP@0.5(%)
Faster RCNN (VGG16)	41.0	68.4
Faster RCNN (ResNet50)	41.2	68.1
Faster RCNN (ResNet101)	41.1	68.7
Faster RCNN (Res2Net101)	**41.5**	**69.1**

It can be known from Table 1 that the ResNet network with residual units has stable performance in underwater object detection. With the deepening of the network layers, the Faster RCNN (ResNet101) has better accuracy in underwater object detection than the original Faster RCNN. The recently appeared Res2Net network, which establishes a multi-level residual network connection inside the original residual module, can describe the multi-scale feature information in a finer area. The experimental result of the improved Faster RCNN (Res2Net101) is the best, with mAP@0.5 reaching 69.1%. Therefore, it can be known from Table 1 that the improved operation of the Faster RCNN backbone network is effective, and the improved Faster RCNN algorithm is more conducive to underwater object detection.

4.4. Comparison Experiments

In order to show the detection effect of the improved Faster RCNN model, other different object detection algorithms are also trained, tested and validated in this paper. Observe the performance comparison between the improved algorithm model in this paper and the other algorithms in the tasks of underwater object detection.

This paper selects the Fast RCNN, SSD, YOLOV3, the unimproved Faster RCNN and the improved Faster RCNN for comparison. According to the data enhancement preprocessing described in the previous section, the 3500 underwater environment images are divided into training data samples, verification data samples and test data samples, with a ratio of 7:2:1. After 500 epochs of iterative training and learning, the learning rate drops to 0.0001 and the parameters of the underwater object detection model converge. The experimental results of the Fast RCNN, SSD, YOLOV3, the Faster RCNN and the improved Faster RCNN are shown in Table 2.

Table 2. Performance of different models in underwater object detection.

Methods	Avg-P(%)	mAP@0.5(%)	*F*1
Fast RCNN	38.3	62.1	45.6
Faster RCNN	41.0	68.4	52.8
YOLOV3	42.1	70.1	54.7
SSD	42.0	70.4	54.5
Ours	**43**	**71.7**	**55.3**

As can be known from Table 2, in terms of underwater image detection the detection results of various algorithm models are compared. The improved Faster RCNN underwater object detection algorithm has relatively good *AP*, *mAP* and *F*1 scores, and these three indicators are higher than the original Faster RCNN algorithm; mAP@0.5(%) reaches 71.7%, which is 3.3% higher than the original algorithm, and the *F*1 score reaches 55.3%, which is 2.5% higher than the original algorithm.

As can be known from Figure 5, the improved Faster RCNN-based underwater object detection method has better performance than the Fast RCNN, the unimproved Faster RCNN, YOLOV3 and SSD. The improved method in this paper has higher accuracy when detecting multiple categories at the same time, and can reduce the occurrence of the missed detections and the false detections well. As can also be known from Figure 5, the method in this paper has also made progress in the detection of small underwater objects, and the small scallops in Figure 5 can also be accurately detected.

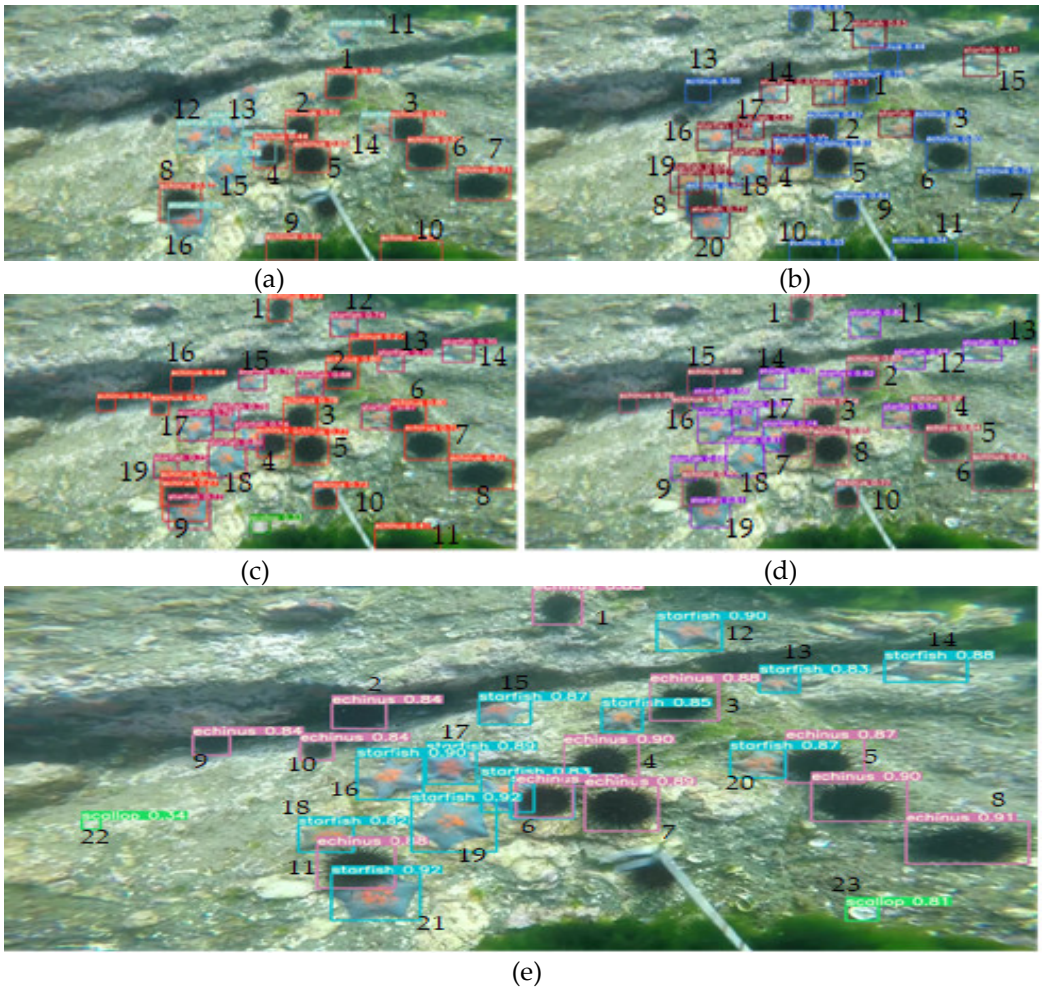


Figure 5. Results of different methods for underwater object detection. (**a**–**e**) are respectively the detection results based on Fast RCNN, Faster RCNN, YOLOV3, SSD, the improved Faster RCNN. The black creatures are sea cucumbers, the blue five-pointed creatures are starfishes, the green round creatures are scallops, and the black slander creatures are seaweeds in the images.

4.5. Ablation Experiments and Analysis

This section progressively demonstrates the effectiveness of each module through ablation experiments. Since our proposed improvement scheme has several modules to be combined, in order to determine whether the addition of modules is effective, this section verifies the rationality of the combination of methods through ablation experiments.

The ablation experiments take the Faster RCNN (Res2Net101) as the main structure, and then add the GIOU algorithm, the OHEM algorithm, the Soft-NMS algorithm, and multi-scale training to it one by one. Then, we train, test, and validate on the same underwater image dataset. The underwater environment images are divided into training data samples, validation data samples and test data samples, with a ratio of 7:2:1. The learning rate is initially set to 0.001. In order to better understand the performance of the model and make adjustments in a convenient and timely manner, pay attention to the performance changes of the model every three iterations. If the performance does not improve, adjust the learning rate to 90% of the original in the next training. Use the Adam [43–45] optimizer to optimize the model. After 500 epochs of iterative training and learning, the learning rate drops to 0.0001 and the parameters of the underwater object detection model converge. The ablation experiment results of the improved Faster RCNN underwater object detection model are shown in Table 3.

Table 3. Ablation experiment results of the improved Faster RCNN.

Methods	Avg-P(%)	mAP@0.5(%)
Faster RCNN (VGG16)	41.0	68.4
Faster RCNN (Res2Net101)	41.5	69.1
Faster RCNN (Res2Net101 + GIOU)	41.2	69.3
Faster RCNN (Res2Net101 + OHEM)	41.1	69.5
Faster RCNN (Res2Net101 + SoftNMS)	41.5	69.2
Faster RCNN (Res2Net101 + GIOU + OHEM + SoftNMS)	41.5	69.7
Faster RCNN (Res2Net101 + GIOU + OHEM + SoftNMS) + Multi-scale training	**43**	**71.7**

As can be known from Table 3, the Faster RCNN (Res2Net101) improves the accuracy of mAP@0.5 by 0.2% by replacing the original IOU algorithm with the GIOU algorithm. In order to overcome the problem of the unbalanced positive and negative samples, the OHEM technology is introduced. Through the experiments, it can be known that this operation increases mAP@0.5 by 0.4%. In order to optimize the bounding box mechanism of the model, the standard NMS is replaced with Soft-NMS and the model performance is slightly improved. Finally, by integrating various improved schemes, the improved Faster RCNN underwater object detection algorithm model is obtained. In the experiments, mAP@0.5 reaches 71.7%, which is 3.3% higher than the original Faster RCNN model. Therefore, the method is effective in underwater object detection.

The precision-recall curves are shown in Figure 6, and it can be known that the recall curve and the accuracy curve can quickly converge. The average accuracy of sea cucumber, sea urchin, scallop, starfish and aquatic plants are given in the curve of recall rate, and the average accuracy of all detection classes. It can be seen from in Figure 6 that sea urchins are in the recall curve. The area is the largest, so the accuracy of underwater target detection is also the highest, and then for the detection of aquatic plants, starfish, sea cucumbers and scallops the accuracy decreases in turn. In terms of detection speed, the improved method proposed in this paper is better than the original Faster RCNN in a single map. The film detection time is 0.005 s longer. After the 500th epochs Recall, Precision, mAP@0.5 and MAP@0.5:0.95 curves are stable, which shows that the proposed methods are effective in underwater object detection.

It can be known from Figure 7 that the proposed and improved method has higher accuracy in underwater object detection—the unimproved Faster RCNN cannot detect the starfish in the lower right corner of Figure 7, while the improved method in this paper can detect the starfish with different shapes in the lower right corner of Figure 7—and the improved model is more robust. At the same time, the improved model can detect small scallops in the upper right corner of Figure 7, and the performance of the small object detection is better than the original Faster RCNN.

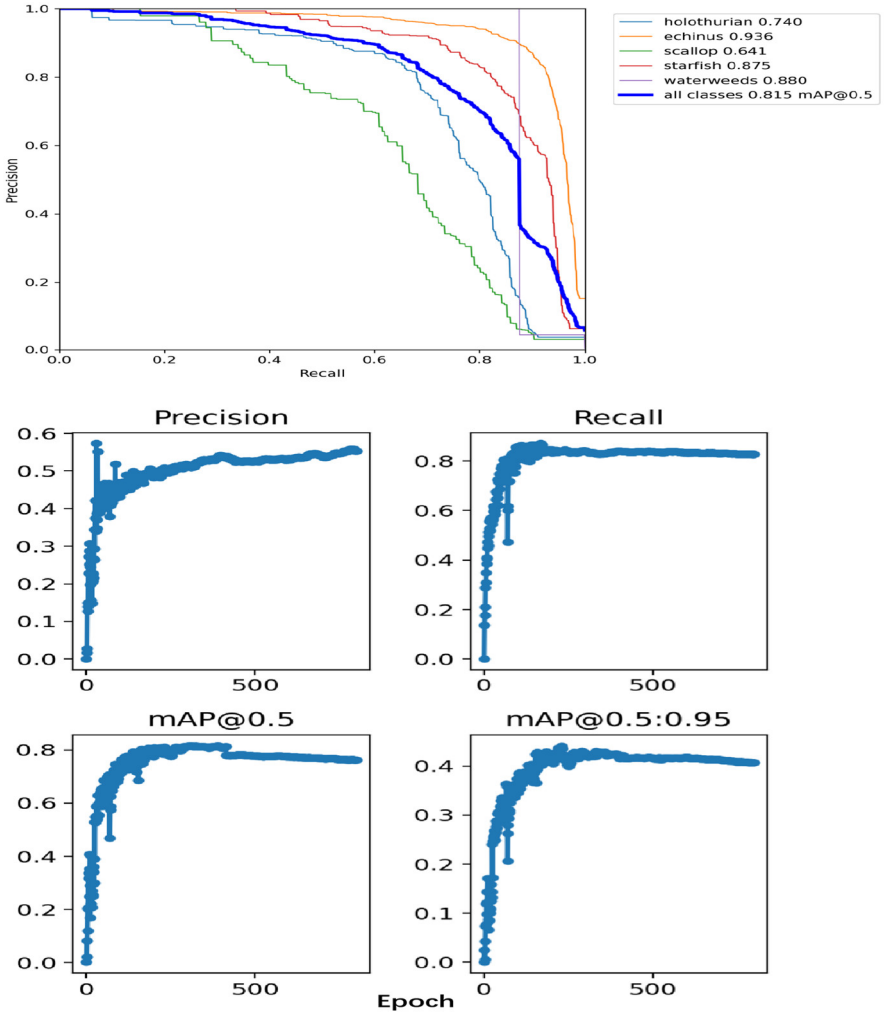


Figure 6. The precision-recall curves.

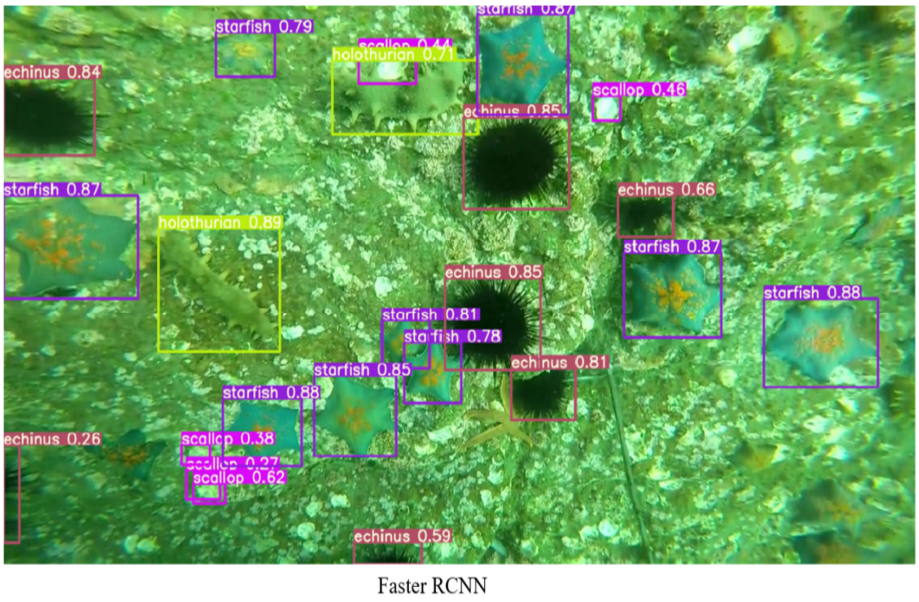


Figure 7. *Cont*.

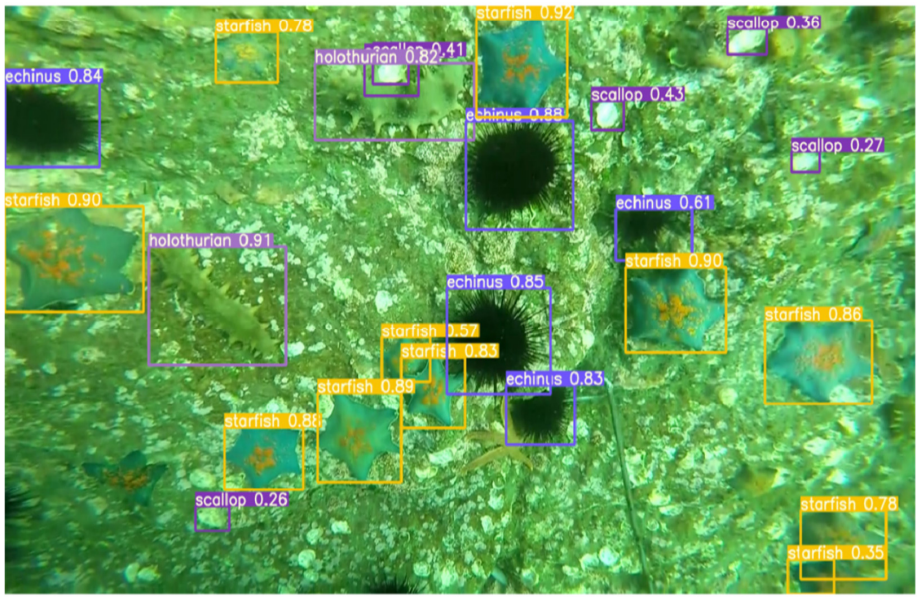


Figure 7. Underwater detection results between the Faster RCNN and our method.

5. Conclusions

The two-stage algorithm Faster RCNN is improved to make it suitable for underwater object detection tasks. We improve the backbone network of the Faster RCNN, optimize the imbalanced positive and negative samples, improve the regress mechanism of the bounding box, and also use multi-scale training on the training strategy. Through comparative experiments, it is verified that it is feasible to use the Res2Net101 network module as the feature extraction structure of the improved Faster RCNN. We also conduct ablation experiments based on the improved Faster RCNN underwater object detection algorithm, disassemble each improved part, and then conduct experiments one by one. It can be known from the experiments that based on the improved Faster RCNN model mAP@0.5 reaches 71.7% in the experiments, which is 3.3% higher than the original Faster RCNN model. The *F*1 score reaches 55.3%, a 2.5% improvement over the unimproved Faster RCNN model, which shows that the proposed methods are effective in underwater object detection.

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

Funding: This research received no external funding.

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

Abbreviations

RCNN	Regions with Convolutional Neural Network Feature
VGG16	Visual Geometry Group Network 16
Res2Net101	*Residual 2 Network 101*
OHEM	Online Hard Example Mining
GIOU	Generalized Intersection Over Union
Soft-NMS	Soft Non-Maximum Suppression
RPN	Region Proposal Network
ROI	Region of Interest
LSTM	Long Short-Term Memory
DTW	Dynamic Time Warping

References

1. Xu, X.; Zou, S.; Liu, J. Research on the promotion path of scientific and technological innovation ability of marine industry based on big data under the background of marine power strategy. In Proceedings of the 2021 International Conference on E-Commerce and E-Management (ICECEM), Dalian, China, 24–26 September 2021; pp. 356–360.
2. Gao, S.; Cheng, M.; Zhao, K.; Zhang, X.; Yang, M.; Torr, P.H.S. Res2net: A new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 652–662. [CrossRef] [PubMed]
3. Shrivastava, A.; Gupta, A.; Girshick, R. Training region-based object detectors with online hard example mining. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 761–769.
4. Grisham, M.P. Mosaic. In *A Guide to Sugarcane Diseases*; Rott, P., Bailey, R.A., Comstock, J.C., Croft, B.J., Eds.; La Librairie du Cirad: Montpellier, France, 2000.
5. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
6. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS–improving object detection with one line of code. In Proceedings of the IEEE International Conference on Computer Vision, Beijing, China, 17–20 September 2005; pp. 5561–5569.
7. Singh, B.; Najibi, M.; Davis, L.S. SNIPER: Efficient multi-scale training. In *Advances in Neural Information Processing Systems*; Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2018; Volume 31, pp. 9310–9320.
8. Li, W.; Logenthiran, T.; Phan, V.T.; Woo, W.L. A novel smart energy theft system (SETS) for IoT-based smart home. *IEEE Internet Things J.* **2019**, *6*, 5531–5539. [CrossRef]
9. Zeng, X.; Lin, S.; Liu, C. Multi-View Deep Learning Framework for Predicting Patient Expenditure in Healthcare. *IEEE Open J. Comput. Soc.* **2021**, *2*, 62–71. [CrossRef]
10. Kashyap, P.K.; Kumar, S.; Jaiswal, A.; Prasad, M.; Gandomi, A.H. Towards Precision Agriculture: IoT-enabled Intelligent Irrigation Systems Using Deep Learning Neural Network. *IEEE Sens. J.* **2021**, *21*, 17479–17491. [CrossRef]
11. Zhang, X.; Chen, M.; Zhan, X. Behavioral cloning for driverless cars using transfer learning. In Proceedings of the 2018 IEEE/ION Position, Location and Navigation Symposium (PLANS), Monterey, CA, USA, 23–26 April 2018; pp. 1069–1073.
12. Lin, Y.-Y.; Yang, J.-Y.; Kuo, C.-Y.; Huang, C.-Y.; Hsu, C.-Y.; Liu, C.-C.C. Use Empirical Mode Decomposition and Ensemble Deep Learning to Improve the Performance of Emotional Voice Recognition. In Proceedings of the 2020 IEEE 2nd International Workshop on System Biology and Biomedical Systems (SBBS), Taichung, Taiwan, 3–4 December 2020; pp. 1–4.
13. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
14. Mittal, S.; Srivastava, S.; Phani, .J.J. A Survey of Deep Learning Techniques for Underwater Image Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**. [CrossRef]
15. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
16. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of theEuropean Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Switzerland, 2016; pp. 21–37.
17. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
18. Deng, J.; Xuan, X.; Wang, W.; Li, Z.; Yao, H.; Wang, Z. A review of research on object detection based on deep learning. *J. Phys. Conf. Ser.* **2020**, *1684*, 012028. [CrossRef]
19. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [CrossRef]
20. Purkait, P.; Zhao, C.; Zach, C. SPP-Net: Deep absolute pose regression with synthetic views. *arXiv* **2017**, arXiv:1712.03452.
21. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
22. Roh, M.C.; Lee, J. Refining faster-RCNN for accurate object detection. In Proceedings of the 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), Nagoya, Japan, 8–12 May 2017; pp. 514–517.
23. Khasawneh, N.; Fraiwan, M.; Fraiwan, L. Detection of K-complexes in EEG waveform images using faster R-CNN and deep transfer learning. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 297. [CrossRef]
24. Iqbal, K.; Odetayo, M.; James, A.; Salam, R.A.; Talib, A.Z.H. Enhancing the low quality images using Unsupervised Colour Correction Method. In Proceedings of the 2010 IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10–13 October 2010; pp. 1703–1709.
25. Zhang, D.; Kopanas, G.; Desai, C.; Chai, S.; Piacentino, M. Unsupervised underwater fish detection fusing flow and objectiveness. In Proceedings of the 2016 IEEE Winter Applications of Computer Vision Workshops (WACVW), New York, NY, USA, 10 March 2016; pp. 1–7.
26. Yuan, F.; Huang, Y.-F.; Chen, X.; Cheng, E. A Biological Sensor System Using Computer Vision for Water Quality Monitoring. *IEEE Access* **2018**, *6*, 61535–61546. [CrossRef]

27. Wang, J.H.; Lee, S.K.; Lai, Y.C.; Lin, C.C.; Wang, T.Y.; Lin, Y.R.; Hsu, T.H.; Huang, C.W.; Chiang, C.P. Anomalous Behaviors Detection for Underwater Fish Using AI Techniques. *IEEE Access* **2020**, *8*, 1–11. [CrossRef]
28. Phillips, J.J. ROI: The search for best practices. *Train. Dev.* **1996**, *50*, 42–48.
29. Jang, E.; Gu, S.; Poole, B. Categorical reparameterization with gumbel-softmax. *arXiv* **2016**, arXiv:1611.01144.
30. Xu, Q.; Zhang, X.; Cheng, R.; Song, Y.; Wang, N. Occlusion Problem-Oriented Adversarial Faster-RCNN Scheme. *IEEE Access* **2019**, *7*, 170362–170373. [CrossRef]
31. Hahn, G.; Lutz, S.M.; Laha, N.; Lange, C. A framework to efficiently smooth L1 penalties for linear regression. *bioRxiv* **2020**, 1–35.
32. Qassim, H.; Verma, A.; Feinzimer, D. Compressed residual-VGG16 CNN model for big data places image recognition. In Proceedings of the 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 8–10 January 2018; pp. 169–175.
33. Theckedath, D.; Sedamkar, R.R. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Comput. Sci.* **2020**, *1*, 1–7. [CrossRef]
34. Lin, S.L. Application Combining VMD and ResNet101 in Intelligent Diagnosis of Motor Faults. *Sensors* **2021**, *21*, 6065. [CrossRef] [PubMed]
35. Cheng, J.; Tian, S.; Yu, L.; Lu, H.; Lv, X. Fully convolutional attention network for biomedical image segmentation. *Artif. Intell. Med.* **2020**, *107*, 101899. [CrossRef]
36. Arthur, D.; Vassilvitskii, S. K-means++: The advantages of careful seeding. In Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007, New Orleans, LA, USA, 77–9 January 2007; pp. 1027–1035.
37. Krishna, K.; Murty, M.N. Genetic K-means algorithm. *IEEE Trans. Syst. Man Cybern. Part B Cybernetics* **1999**, *29*, 433–439. [CrossRef]
38. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
39. Zhou, W.; Chen, Y.; Liu, C.; Yu, L. GFNet: Gate Fusion Network with Res2Net for Detecting Salient Objects in RGB-D Images. *IEEE Signal Process. Lett.* **2020**, *27*, 800–804.
40. Kaiyan, Z.; Xiang, L.; Weibo, S. Underwater object detection using transfer learning with deep learning. In Proceedings of the CIPAE 2020: 2020 International Conference on Computers, Information Processing and Advanced Education, Ottawa, ON, Canada, 16–18 October 2020; pp. 157–160.
41. Albahli, S.; Nida, N.; Irtaza, A.; Yousaf, M.H.; Mahmood, M.T. Melanoma Lesion Detection and Segmentation Using YOLOv4-DarkNet and Active Contour. *IEEE Access* **2020**, *8*, 198403–198414. [CrossRef]
42. He, M.X.; Hao, P.; Xin, Y.Z. A robust method for wheatear detection using UAV in natural scenes. *IEEE Access* **2020**, *8*, 189043–189053. [CrossRef]
43. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
44. Wang, Y.; Liu, J.; Yu, S.; Wang, K.; Han, Z.; Tang, Y. Underwater Object Detection based on YOLO-v3 network. In Proceedings of the 2021 IEEE International Conference on Unmanned Systems (ICUS), Beijing, China, 22–24 October 2021; pp. 571–575.
45. Mathias, A.; Dhanalakshmi, S.; Kumar, R. Occlusion aware underwater object tracking using hybrid adaptive deep SORT-YOLOv3 approach. *Multimed. Tools Appl.* **2022**, *81*, 44109–44121. [CrossRef]