water

MDPI

*Article*

# CME-YOLOv5: An Efficient Object Detection Network for Densely Spaced Fish and Small Targets

**Jianyuan Li [1,2], Chunna Liu [1,*], Xiaochun Lu [2] and Bilang Wu [1]**

[1] State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100048, China; lijianyuan22@163.com (J.L.); wubl@iwhr.com (B.W.)
[2] College of Hydraulic and Environmental Engineering, China Three Gorges University, Yichang 443002, China; luxiaochun1014@163.com
* Correspondence: liucn@iwhr.com; Tel.: +86-158-1104-3560

**Abstract:** Fish are indicative species with a relatively balanced ecosystem. Underwater target fish detection is of great significance to fishery resource investigations. Traditional investigation methods cannot meet the increasing requirements of environmental protection and investigation, and the existing target detection technology has few studies on the dynamic identification of underwater fish and small targets. To reduce environmental disturbances and solve the problems of many fish, dense, mutual occlusion and difficult detection of small targets, an improved CME-YOLOv5 network is proposed to detect fish in dense groups and small targets. First, the coordinate attention (CA) mechanism and cross-stage partial networks with 3 convolutions (C3) structure are fused into the C3CA module to replace the C3 module of the backbone in you only look once (YOLOv5) to improve the extraction of target feature information and detection accuracy. Second, the three detection layers are expanded to four, which enhances the model's ability to capture information in different dimensions and improves detection performance. Finally, the efficient intersection over union (*EIOU*) loss function is used instead of the generalized intersection over union (GIOU) loss function to optimize the convergence rate and location accuracy. Based on the actual image data and a small number of datasets obtained online, the experimental results showed that the mean average precision (mAP@0.50) of the proposed algorithm reached 94.9%, which is 4.4 percentage points higher than that of the YOLOv5 algorithm, and the number of fish and small target detection performances was 24.6% higher. The results show that our proposed algorithm exhibits good detection performance when applied to densely spaced fish and small targets and can be used as an alternative or supplemental method for fishery resource investigation.

**Keywords:** densely spaced fish; small targets; CME-YOLOv5; attention mechanism; multiscale; loss function



**Citation:** Li, J.; Liu, C.; Lu, X.; Wu, B. CME-YOLOv5: An Efficient Object Detection Network for Densely Spaced Fish and Small Targets. *Water* **2022**, *14*, 2412. https://doi.org/10.3390/w14152412

Academic Editors: Yizi Shang, Yongping Wei, Ling Shang, Akiyuki Kawasaki and Yuchuan Wang

Received: 6 July 2022
Accepted: 1 August 2022
Published: 3 August 2022



## 1. Introduction

Changes in fish stocks can directly reflect the status of river ecosystems. The General Committee on fisheries in the Mediterranean counted the fishing indicators of fisheries in the Mediterranean region from 1970 to 2017 and found that the ecosystem structure had changed due to overexploitation, and the catch had been declining since 2010 [1]. Therefore, regular or irregular fishery resource studies and assessments are needed to confirm the status of ecosystems. It is important for responsible fishing and environmental protection. Early investigations of fishery resources used electric fishing, artificial spray-net fishing, ground cage, and prick-net methods [2–5] to collect fish samples. However, there are several problems with traditional fishery resource survey methods: 1. their excessive dependence on manual operations makes them time-consuming and laborious; 2. they result in greater disturbance of fish and aquatic ecosystems; and 3. small target fish are easily missed [6]. Traditional research methods have difficulty meeting increasing

requirements for environmental protection and monitoring. Therefore, it is necessary to use the latest technology to implement fishery resource studies.

In recent years, in the era of artificial intelligence, computer information technology has developed rapidly, and there have been advances in computer vision [7]. As its core field, object detection technology has made a major breakthrough [8]. An increasing number of deep-learning object detection methods have been applied to underwater object detection. Pei Qianqian et al. [9] applied the deep learning object detection algorithm YOLOv3 [10] in an engineering fishway and conducted real-time detection of passing fish in the fishway, but the model is too single, the water quality of the fishway will change with the season, and the target detection effect is poor when the background fluctuation is relatively large. Youssef et al. [11] proposed an object detection algorithm to improve the clarity of water images. The multiscale Retinex (MSR) algorithm [12] was used to enhance the blurred water image or video in the system to increase its clarity, and then the YOLOv3 object detection algorithm was used to identify the enhanced image. The detection accuracy was significantly improved. However, the MSR algorithm is an image enhancement algorithm based on a physical model, which has a slow processing speed and a correspondingly slow image recognition speed. Fan Weiya [13] improved the Faster R-CNN [14] algorithm by increasing the number of anchors in the RPN network, changing the deformation convolution and adding three single fully connected channels. The detection accuracy was improved by 6.23% to 92.44% compared to that of the original model algorithm, improving the applicability of the fish object detection algorithm. However, the Faster R-CNN algorithm is a two-stage target detection algorithm, and the processing speed of the algorithm is slow. If the number of anchor points and channels is added, the calculating model parameters will be aggravated, which will lead to a significant decrease in the model processing speed. Yao et al. [15] used the object detection algorithm of YOLOv4 [16] for underwater target recognition, replaced the upsampling module in the original model with a deconvolution module, removed the SPP layer, and added depth detachable convolution to reduce network computations. The results showed that compared with the original YOLOv4, the improved mAP reached 75.34, which was nearly 12% higher than YOLOv4. Qiang et al. [17] proposed an improved SSD [18] algorithm based on ResNet instead of VGG and proposed depth-separable deformation convolution, which improved fish detection accuracy and speed in complex water environments. Wu Rui et al. [19] improved the YOLOv5 model [20] by introducing the convolutional attention mechanism module, and the results showed that the improved method greatly improved the identification accuracy and speed of benthic organisms in coral reefs. In summary, the above method studies are based on static individual identification or the detection of conventional targets that are only applicable to general scenes. There are few studies on the dynamic identification of fish and small targets, and the identification of dense underwater fish and small targets still has a high rate of missed detection and error. There are challenges of mutual occlusion and shadows cast by densely spaced underwater fish [21], and small object detection has always been one of the key difficulties in the object detection field [22,23], with the problems of less effective image features information, fuzziness and other difficulties. The needs of fishery research cannot be met due to the limited ability of current target detection technology to overcome the problems of fish occlusion and the difficulties with small target detection. This situation needs to be improved by advancing image recognition.

To address the above issues, in this paper, we propose a CME-YOLOv5 algorithm. Based on the YOLOv5 object detection algorithm, the CA attention mechanism is improved, the detection layers are expanded to 4 according to the characteristics of small objects, and the GIOU loss is replaced with *EIOU* loss. The problems of poor underwater detection of dense fish swarms, small target positioning, and few pixels and low accuracy are solved, and the accuracy of the algorithm for dense fish swarms and small targets while ensuring real-time performance effectively improved. This technology can become an alternative or supplementary method for fishery resource studies.

## 2. Efficient Object Detection Network Design

### *2.1. YOLOv5 Object Detection Model*

The YOLOv5 algorithm framework is divided into four parts: the first part is the input layer, and the input size is 640 × 640's three-channel image; the second part is the backbone network, which uses the Darknet-53 network framework as a model to extract the image features; the third part is the neck module, which is located between backbone and the last output layer, includes spatial pyramid pooling-Fast (SPPF) using the maximum pooling method and a path aggregation network (PANet) under an instance segmentation framework, and repeatedly features the fusion and extraction of the shallow and deep information in the three feature layers to make full use of the context information; the fourth part predicts and decodes the three generated 20 × 20, 40 × 40, 80 × 80 feature maps (YOLO Head) and directly obtains the position of the prediction box in the image and class of each object.

In the YOLOv5 model, there are four models with different network depths and widths. According to the cost-performance ratio in Table 1, YOLOv5 l with small calculation parameters, high accuracy and high speed was selected as a basis for improvement and experimentation.

**Table 1.** Parameters of different YOLOv5 models [20].

| Model | Size (Pixels) | $mAP_{val}^{0.5:0.95}$ | $mAP_{val}^{0.5}$ | Speed CPU b1 (ms) | Speed V100 b1 (ms) | Speed V100 b32 (ms) | Params (M) | FLOPs @640 (B) |
|---|---|---|---|---|---|---|---|---|
| YOLOv5s | 640 | 37.4 | 56.8 | 98 | 6.4 | 0.9 | 7.2 | 16.5 |
| YOLOv5m | 640 | 45.4 | 64.1 | 224 | 8.2 | 1.7 | 21.2 | 49 |
| YOLOv5l | 640 | 49 | 67.3 | 430 | 10.1 | 2.7 | 46.5 | 109.1 |
| YOLOv5x | 640 | 50.7 | 68.9 | 766 | 12.1 | 4.8 | 86.7 | 205.7 |

### *2.2. Improved CME-YOLOv5 Recognition Method*

YOLOv5 has the characteristics of fast detection, high efficiency and flexibility in target recognition. Underwater images generally have the problems of low contrast, blurring, color deviation, and obscuration, which lead to poor quality and difficulty with the detection and identification of fish in dense schools. This affects the multiobject recognition of underwater schools of fish and is disadvantageous for small object fish detection. Therefore, it is necessary to further improve the YOLOv5 network to improve detection accuracy and network performance. In this paper, a dense fish school and small object recognition algorithm for the CME-YOLOv5 network are proposed; the model is shown in Figure 1. The innovative features of the model are as follows: (1) The C3 structure converged attention mechanism coordinate attention (CA) in the YOLOv5 network is used to form the C3CA structure instead of the C3 structure in the backbone extraction network to increase the model's attention to key information, reduce the interference of invalid object information, and enhance the feature expression ability of small objects in the detection network by focusing on essential information from extensive amounts of available information. (2) The number of detection layers in the YOLOv5 detection module (YOLO Head) is expanded from 3 to 4 to better capture global information and rich context information, improve the ability of the model to capture different dimensions information, and improve the detection performance of the YOLOv5 network to multiscale objects, to extract features better in dense schools of fish and improve the ability of the model to deal with small object detection. (3) The *EIOU* loss function, which considers overlapping area, center point distance, length, width and side length true difference, and adds focal loss to solve the sample imbalance problem in bounding box regression, is used instead of the GIOU loss function, thus addressing the problems of the slow convergence of the GIOU loss function in the horizontal and vertical direction and its inability to optimize the case when the predicted bounding box and ground-truth bounding box do not intersect.

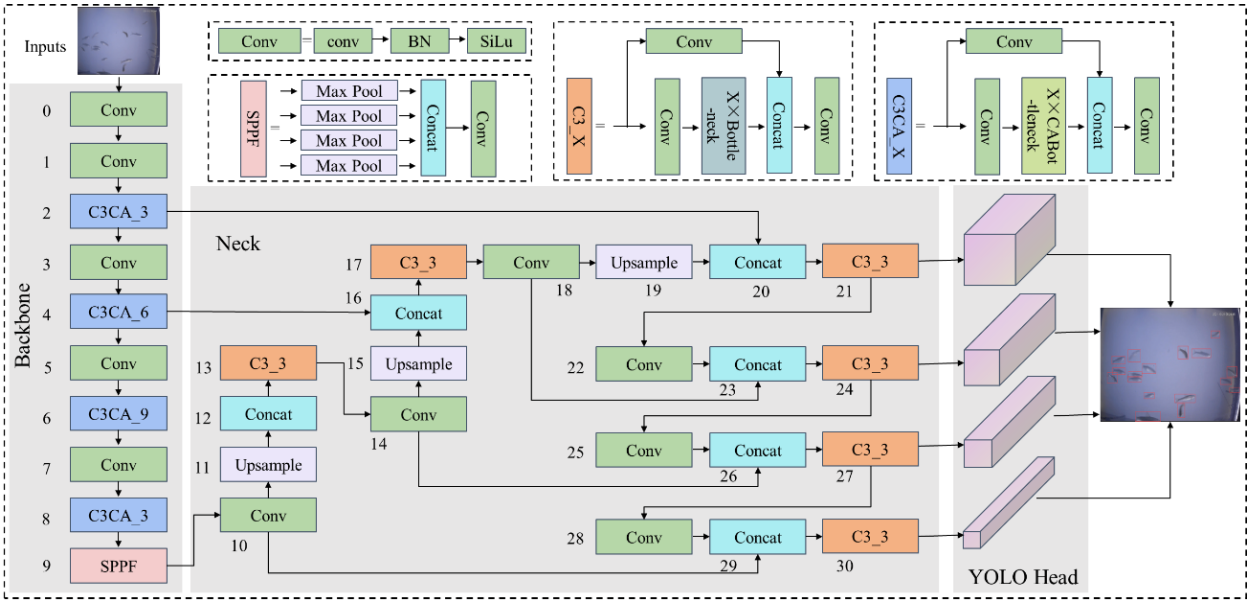


**Figure 1.** This is a figure that showing CME-YOLOv5 network structure. The numbers 1 to 30 in the figure represent the layers of the CME-YOLOV5 model.

### 2.2.1. CA Attention Mechanism Module

Coordinate attention (CA) [24] (as shown in Figure 2) is a lightweight and efficient mechanism in the channel and X and Y spatial directions, through which channel attention is decomposed into two different spatial directions for aggregating features in a one-dimensional feature coding process. It captures long-term dependencies in one space, retains precise location information in the other, and forms a pair of direction-aware and position-sensitive feature maps so that these feature maps can be used complementary to enhance the representation of effective information.

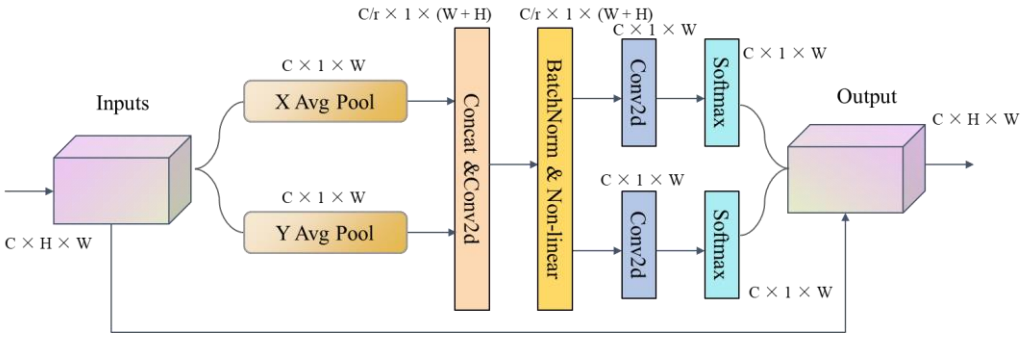


**Figure 2.** Schematic diagram of the CA mechanism module.

In CA, full average pooling of the input feature maps in the height and width directions is first carried out to obtain the feature maps ($Z_C^h(h)$, $Z_C^w(w)$). Then, the feature maps are split in the height and width directions together to obtain the feature maps after convolution, batch normalization and nonlinear sigmoid activation, where $\sigma$ is a *sigmoid* function. Next, the feature map $F$ is convolved with the original height and width to obtain feature graphs $F_h$ and $F_w$, respectively, with the same number of channels as the original. After the *sigmoid* activation function, the attention weight in height and width and the attention weight in the width direction of the feature map ($g^h$, $g^w$) are obtained. Finally, the feature map with attention weight in the height and width direction is obtained by multiplicative weighting calculation on the original feature map, which can enhance the important information and help the model locate and identify the target more accurately.

2.2.2. Multiscale Detection Layer

Three detection layers of network feature maps, 20 × 20, 40 × 40, and 80 × 80, are obtained after the initial YOLOv5 network structure passes through the backbone network and the neck enhancement module, which are used to detect large, medium and small objects, respectively. For conventional detection, it may be possible to achieve the desired effect, but for dense groups of fish with individuals of different sizes, there are often omissions or poor detection accuracy, especially for small objects. Therefore, to detect individuals in dense underwater fish schools, a detection layer that can detect smaller objects is added based on the original three detection layers. The model is shown in Figure 1. The convolution is carried out at the 17th convolutional layer, which originally needs to be downsampled and then upsampled, and the feature concatenation is carried out at the 20th and 2nd layers so that the network actively learns, adaptively fuses features and concatenates the process information, thus increasing the sensing field. Then, the 160 × 160 network feature map is obtained through convolution. YOLO Head1 is a new detection layer introduced in our method. The second detection layer (YOLO Head2) of 80 × 80 is obtained through convolution after feature concatenation of layer 23 and layer 18. The third detection layer (YOLO Head3) of 40 × 40 is obtained through convolution after feature concatenation of layer 26 and layer 14. The 29th layer and the 10th layer are concatenated with features, and then the 4th detection layer (YOLO Head4) of 20 × 20 is obtained through convolution.

2.2.3. Optimized Loss Function

The loss function calculates the difference between the forward calculation result and the real value of each iteration of the neural network and evaluates the difference between the predicted value and the real value of the model. Generally, the better the loss function, the better the model's performance. At present, object detection regression loss functions include *IOU* [25], GIOU [26], DIOU [27], CIOU [28] and *EIOU* loss [29]. The original YOLOv5 loss function is GIOU loss. GIOU loss uses closure as a penalty term, which may lead to the problem of nonconvergence of the results in the model training process. Therefore, we use *EIOU* loss to calculate regression loss.

$$L_{EIOU} = L_{IOU} + L_{dis} + L_{asp} = 1 - IOU + \frac{\rho^2(b, b^2)}{c^2} + \frac{\rho^2(w, w^2)}{c_w^2} + \frac{\rho^2(h, h^2)}{c_h^2}, \quad (1)$$

$$\rho^2(b, b^{gt}) = \sqrt{\left(b_x - b_x^{gt}\right)^2 + \left(b_y - b_y^{gt}\right)^2}, \quad (2)$$

In the formula, intersection over union (*IOU*) is the intersection and union ratio between the predicted bounding box and ground-truth bounding box; Equation (2) represents the Euclidean distance between the center point of the predicted bounding box and ground-truth bounding box; $b$ is the center point of the predicted bounding box; $b^{gt}$ is the center point of the ground-truth bounding box; $w$ is the width of the predicted bounding box; $w^{gt}$ is the width of the ground-truth bounding box; $h$ is the height of the predicted bounding box; $h^{gt}$ is the height of the ground-truth bounding box; $C$ is the diagonal distance of the minimum closure region that can contain both the predicted bounding box and ground-truth bounding box.

The *EIOU* loss function consists of *IOU* loss $L_{IOU}$, centre distance loss $L_{dis}$ and side length loss $L_{asp}$, which can optimize the convergence speed and positioning accuracy and reduce the likelihood of inaccurate regression results.

## 3. Dataset

The experimental dataset in this analysis consisted of 1500 pictures, of which 65% were images of underwater fish and small target fish collected from 4 coexisting hydropower stations and 1 fish breeding station in Xinjiang and Tibet. To enhance the robustness of the training results and improve the detection effect of the model, 35% were datasets (labeled

fish in the wild) provided by NOAA, which included images of large numbers of fish and small target fish. LabelImg was used to label the datasets one by one. The labeling requirements were as follows: (1) in the dense fish group, fish visibility of more than 1/5 should be labeled; and (2) in the small target image, to prevent overfitting and reduce misidentification, the image pixel can be labeled if it does not reach the lost frame rate. After the annotation was complete, scripts were used to convert it into files required for YOLOv5 training, and the datasets were randomly divided into training sets and validation sets at a ratio of 8:2.

## 4. Experimental Protocols and Evaluation Measures

### *4.1. Experimental Platform and Protocols*

This experiment was implemented with the Windows 10 operating system, Intel TM i7-11800 h CPU processor, GeForce RTX3080 GPU graphics card, 16 GB video memory, CUDA11.1 for training acceleration and the PyTorch 1.9 deep learning framework for training. The image input size was 640 × 640, the initial learning rate was 0.01, the final learning rate was 0.1, the SGD optimization model was used, and the training batch size was 8. The specific model parameter configuration is shown in Table 2.

**Table 2.** Model parameter configuration table.

| Parameters | Configuration | Parameters | Configuration |
|---|---|---|---|
| operating system | Windows10 | initial learning rate | 0.01 |
| CPU | i7-11800H | final learning rate | 0.1 |
| GPU | GeForce RTX3080 | optimizer | SGD |
| CUDA | 11.1 | optimizer momentum | 0.937 |
| image-size | 640 × 640 | batch size | 8 |

### *4.2. Model Evaluation Measures*

To verify the detection and recognition ability of our proposed model for images of densely spaced underwater fish and small objects, precision was adopted to estimate the correct proportion of all objects predicted by the model. Recall that the model predicts the correct proportion of objects among all real objects. The average accuracy (mAP), the area under the P-R curve, measures the performance of the model.

$$Precision = \frac{TP}{TP + FP}, \tag{3}$$

$$Recall = \frac{TP}{TP + FN}, \tag{4}$$

$$AP = \int_0^1 P(r)dr, \tag{5}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i, \tag{6}$$

$TP$, $TN$, $FP$, and $FN$ are abbreviations for true positive, true negative, false positive, and false negative, respectively. Positive and negative represent the predicted results of the model. If the $IOU$ value is greater than the threshold (set to 0.5), the prediction is positive; if the $IOU$ value is less than the threshold, the prediction is negative. True and false indicate whether the predicted result is the same as the real result; if the results are the same, the assessment is set to true; and if they are different, the assessment is set to false, as shown in Table 3 below:

**Table 3.** Division of positive and negative samples.

| Predicted Value / Real Value | Positive | Negative |
|---|---|---|
| Positive | True Positive (*TP*) | False Negative (*FN*) |
| Negative | False Positive (*FP*) | True Negative (*TN*) |

## 5. Results and Discussion

### *5.1. Results Analysis*

#### 5.1.1. C3CA Ablation Experiment

To assess the efficiency of the C3 structure, we conducted a test replacing the C3 structure with C3CA at different locations and used mAP as an evaluation index. According to the C3CA ablation experiment, it can be seen from Table 4 that the replacement method of Framework 1 exhibited the largest accuracy improvement, which is 1.5 and 1.1% higher than those of Framework 2 and Framework 3, respectively. As a result, the first method was adopted.

**Table 4.** C3CA ablation experiment.

| Framework | Backbone | Neck | mAP |
|---|---|---|---|
| Framework 1 | √ | | 93.2 |
| Framework 2 | | √ | 91.7 |
| Framework 3 | √ | √ | 92.1 |

#### 5.1.2. CME-YOLOv5 Ablation Experiment

To assess the effectiveness and progressiveness of the algorithm proposed in this paper, 8 groups of ablation experiments were conducted with the same verification set to evaluate the images of different improvement schemes on the detection performance of the model. The accuracy, recall, average accuracy, average detection time and model loss value of each model were used to evaluate the impact of different modules on the YOLOv5 target detection algorithm under the same experimental conditions. The lower the model loss value, the better the regression of the model. The objective evaluation index results are shown in Table 5 and the map is shown in Figure 3. According to the data in Table 5, after the CA attention mechanism was fused with the C3 structure, compared with the initial YOLOv5 model, mAP@0.50 increased by 2.7 percentage points, and the detection time increased by 6.1 ms. Although the detection time increased, the model detection accuracy effectively improved, indicating that this method can improve the extraction ability of target feature information, suppress the interference of invalid feature information, and maximize the utilization of feature information. After expanding the 3 detection layers of YOLOv5 to 4, map_0.5 increased by 1.6 percentage points, the model detection accuracy improved, the small target detection performance increased, and the method was able to detect objects on different scales. Replacing GIOU loss with *EIOU* loss lowered the training loss value of the model, reduced the average detection time, and slightly improved mAP@0.50, indicating that *EIOU* can optimize the convergence speed and positioning accuracy and reduce the phenomenon of nonconvergence of regression results. The final results showed that each enhanced method introduced in this paper exhibited a different performance improvement over YOLOv5. The proposed algorithm mAP reached 94.9%, which was 4.4 percentage points higher than that of YOLOv5 compared with mAP@0.50. The proposed algorithm was inferior to YOLOv5 in average detection speed, and the detection time of a single image increased by 8.5 ms. However, the algorithm introduced in this paper can meet the requirements of real-time detection, and the detection accuracy is greatly improved.

**Table 5.** This is a table that evaluates the impact of different improvement schemes on model detection performance: Models 1–3 are single module improvement experiments, Model 1 integrates the CA attention mechanism, Model 2 expands the detection layer, Model 3 uses the *EIOU* loss function, and Models 4–6 are the improvement experiments of the two modules. Model 4 is the fusion CA attention mechanism and the extended detection layer, Model 5 is the fusion CA at-tention mechanism and the use of the *EIOU* loss function, and Model 6 is the extended detection layer and the use of the *EIOU* loss function.

| Order Number | Model | CA | Multiscale Detection Layer | *EIOU* | Precision (%) | Recall (%) | mAP@0.50 (%) | Average Detection Time (s) | Model Training Loss |
|---|---|---|---|---|---|---|---|---|---|
| 0 | YOLOv5 | | | | 83.9 | 84.7 | 90.5 | 14.3 | 0.0240 |
| 1 | Model 1 | √ | | | 89.7 | 87.0 | 93.2 | 20.4 | 0.0305 |
| 2 | Model 2 | | √ | | 86.4 | 86.1 | 92.1 | 17.0 | 0.0308 |
| 3 | Model 3 | | | √ | 87.8 | 84.4 | 91.1 | 14.1 | 0.0209 |
| 4 | Model 4 | √ | √ | | 89.8 | 90.4 | 94.3 | 23.2 | 0.0354 |
| 5 | Model 5 | √ | | √ | 90.1 | 86.5 | 93.8 | 20.4 | 0.0257 |
| 6 | Model 6 | | √ | √ | 85.5 | 88.2 | 93.0 | 16.9 | 0.0275 |
| 7 | CEM-YOLOv5 | √ | √ | √ | 92.3 | 88.1 | 94.9 | 22.8 | 0.0316 |

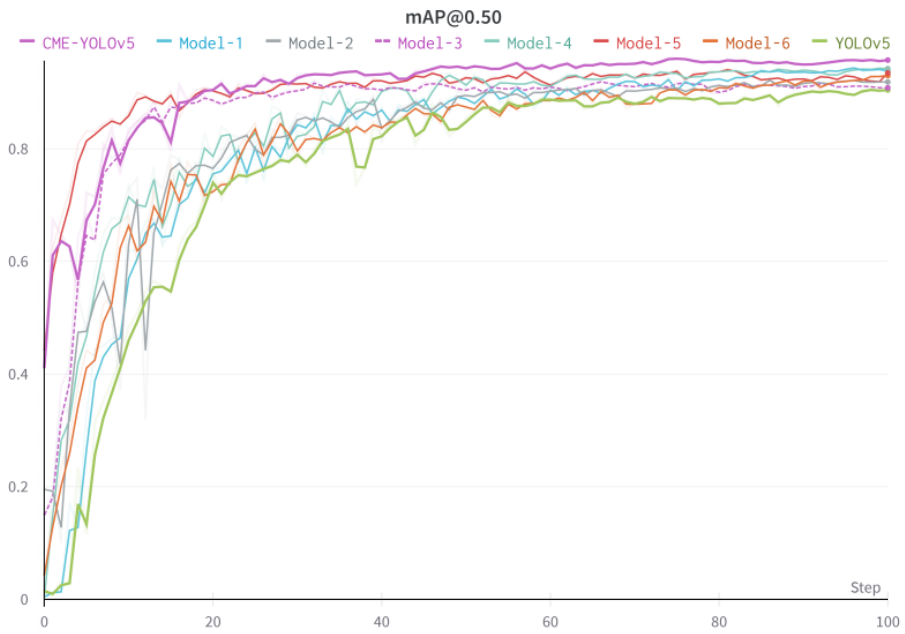


**Figure 3.** mAP@0.50 of the ablation experiment.

### 5.1.3. Comparison of Experimental Results

To assess the detection performance of our algorithm in representative experiments of the detection of more difficult images, the results of our method were compared with those of YOLOv5. As seen in Figure 4, our algorithm greatly reduced the missing detection of dense fish schools, improved the detection accuracy for small target fish with few pixels and a lack of feature information, and reduced the extraction of useless image feature information. The performance was better than that of the initial YOLOv5 model. Figure 4A1 shows the detection results of a photograph taken near a hydropower station. The image has challenges, such as low contrast, blurred vision and occlusion. The YOLOv5 algorithm exhibited serious missed detection on the right side of the figure. However, our algorithm effectively detected small objects without feature information. As seen in Figure 4A2, YOLOv5 failed to detect the abnormal angle of small target fish in the lower right corner with fewer pixels. Our algorithm detected small objects with blurred vision and a lack of

pixels on the upper right of the image (Figure 4A3). Figure 4B1 shows a small target fish school. The detection performance of our algorithm was significantly higher than that of YOLOv5, indicating that the model improves small target detection. In Figure 4B2,B3, there are a large number of occluded objects in the image. Our algorithm effectively detected target fish occluded by other fish, which demonstrated the ability of the method to detect occluded and highly overlapping objects.

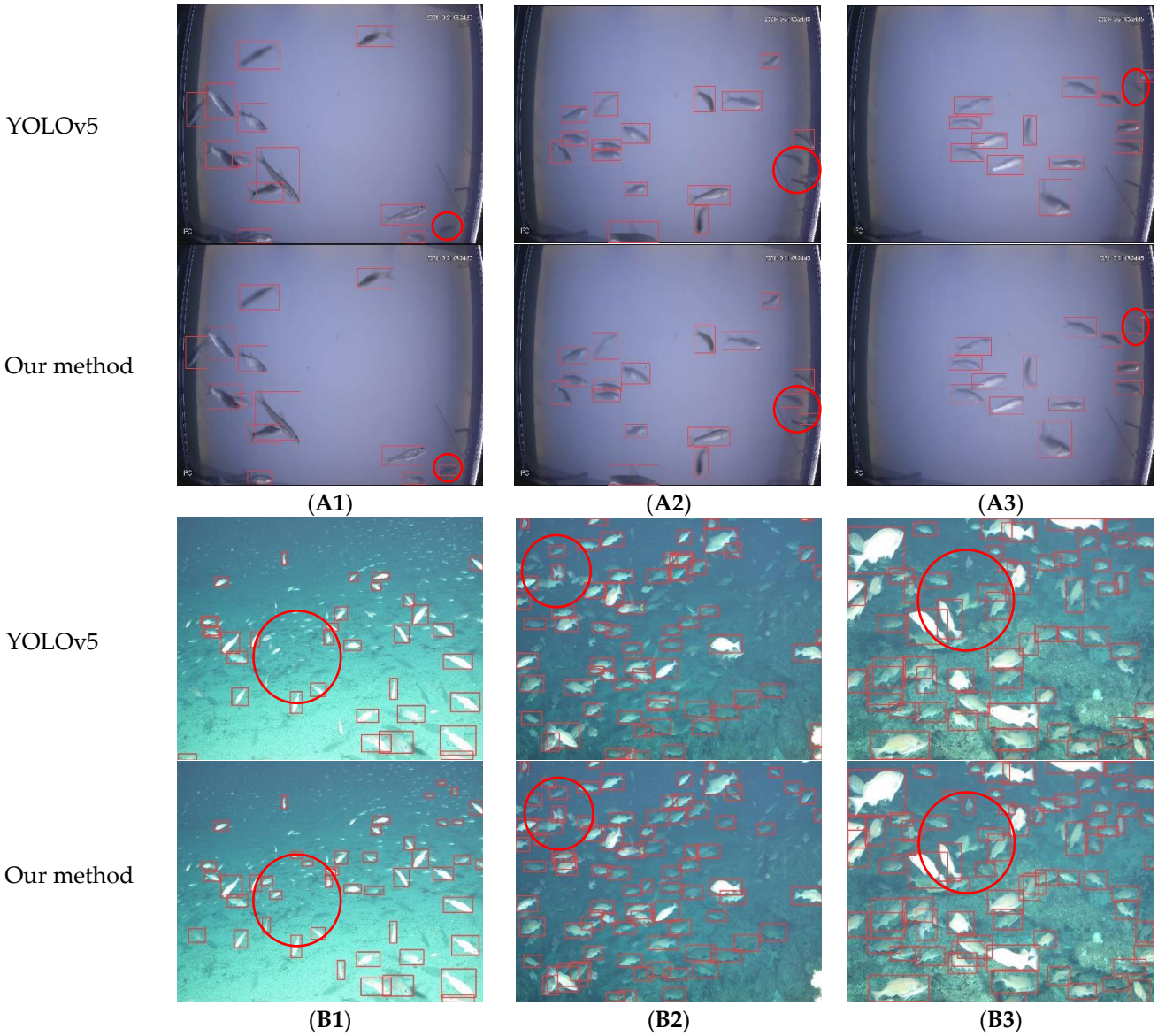


**Figure 4.** This is a figure showing the prediction results of the model: (**A1**) detection result of an occluded target in a power station image; (**A2**) detection result of a small target in a power station image; (**A3**) detection result of a fuzzy target in a power station image; (**B1**) small target fish image detection results; (**B2**) densely spaced fish and small target detection results; and (**B3**) fish and small target detection results under exposure.

Table 6 shows the number of fish detected by our algorithm and YOLOv5. According to the table, the total number of objects detected by our algorithm was 49 more than that detected by YOLOv5, and the detection ratio increased by 24.6%. Our improved algorithm had better detection performance for densely spaced fish and small objects.

**Table 6.** Model prediction.

| Model | Picture 1 | Picture 2 | Picture 3 | Picture 4 | Picture 5 | Picture 6 | Total Number |
|---|---|---|---|---|---|---|---|
| YOLOv5 | 15 | 12 | 13 | 33 | 68 | 58 | 199 |
| CEM-YOLOv5 | 17 | 13 | 14 | 47 | 83 | 74 | 248 |
| Quantity ratio | 113.3% | 108.3% | 107.7% | 142.4% | 122.1% | 127.6% | 124.6% |

### *5.2. Discussion*

To verify the effectiveness and progressiveness of the CIM-YOLOv5 algorithm proposed in this paper for densely spaced and small target fish, the same dataset was used to compare its performance to that of the SSD, Faster R-CNN, YOLOv4 and YOLOv5 target detection algorithms. As seen in the data in Table 7, compared with the SSD, Faster R-CNN, YOLOv4 and YOLOv5 detection algorithms, the accuracy of CME-YOLOv5 achieved the optimal level. mAP@0.50 was 18.4, 15.3, 10.0 and 4.4 percentage points higher than SSD, Faster R-CNN, YOLOv4 and YOLOv5, respectively. The algorithm proposed in this paper uses C3CA instead of the C3 module based on YOLOv5, expands the detection layer from 3 to 4, and replaces the *EIOU* loss function, which can allow the model to achieve better detection performance, focus more attention on key information areas, and improve its ability to detect small objects. However, it is worth noting that adding the CA attention mechanism and expanding the detection layer led to an increase in the number of model parameters; compared with the original YOLOv5 algorithm, the computation of the model also increased, resulting in an increase in the average detection time.

**Table 7.** Model prediction.

| Model | mAP@0.50 (%) | Average Detection Time (ms) |
|---|---|---|
| SSD | 76.5 | 36.5 |
| Faster R-CNN | 79.6 | 61.5 |
| YOLOv4 | 89.2 | 30.7 |
| YOLOV5 | 90.5 | 14.3 |
| CME-YOLOv5 | 94.9 | 22.8 |

The model proposed in this paper is only certified in small target detection of underwater fish, but it does not affect the application of the model to small target scenes in other academic/industrial fields or datasets, such as UAV aerial photography and dense crowds. In the future, the application of computer vision technology to actual scenes is the trend and focus of current research, but many models currently focus more on improving accuracy and their detection speed will be limited. In fact, many model structures will have some redundant modules, which will lead to more useless calculations when the network is transmitted forwards/backwards, and this will not increase our accuracy. At present, the distillation and pruning of the network may improve these problems. In the future, we plan to develop a model with better performance and dynamic high-speed detection of targets.

## 6. Conclusions

To address problems such as many fish, density, mutual occlusion and small targets with little effective information and fuzziness, in this paper, we propose a method for densely spaced fish and small target recognition based on YOLOv5. Compared with other models, it has stronger advantages in various indicators.

First, aiming at the problems of poor positioning and less effective information in underwater target detection, this paper proposes that first, the attention mechanisms Ca and C3 structure are fused to increase the ability of the network to extract key information; second, aiming at the problem of large number and intensive detection tasks, the 3 detection

layers of YOLOv5 were expanded to 4. Finally, *EIOU* loss was replaced by GIOU loss to optimize convergence speed and reduce inaccurate regression results.

The experimental results showed that the improved algorithm proposed in this paper had different effects on different indicators; mAP@0.50 reached 94.9%, which had better accuracy. The number of image detections reached 248, which was 49 more than that of YOLOv5, and the detection effect was 24.6 percent higher. Target detection performance improved. In summary, our proposed algorithm had higher accuracy and detection performance for densely spaced fish and small objects and is more suitable for underwater fishery resource studies.

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

**Funding:** This research was funded by the China Institute of Water Resources and Hydropower Research "Five Talents" special project (No.SD0145B032021) and Youth Programme of the National Natural Science Foundation of China (No.51809291).

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Some of the datasets in this study are openly available from NOAA (https://swfscdata.nmfs.noaa.gov/labeled-fishes-in-the-wild/, (accessed on 6 July 2022)).

**Acknowledgments:** Thanks for the help from Chunna Liu and Xiaochun Lu all the time.

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

## References

1. Dimarchopoulou, D.; Keramidas, I.; Sylaios, G.; Tsikliras, A. Ecotrophic Effects of Fishing across the Mediterranean Sea. *Water* **2021**, *13*, 482. [CrossRef]
2. Qi, X. Hanjiang upstream of fishery resources survey. *Chin. J. Fish. Res.* **2022**, *44*, 21–32.
3. Li, K.; Shen, Z.; Chen, Y.; Ye, H. Preliminary study on fish diversity and Protection Countermeasures in Banduo section of the Yellow River. *Chin. J. Hydroecol.* **2012**, *33*, 104–107.
4. Liu, K.; Yu, C.; Yu, N.; Zhang, P.; Jiang, Q.; Niu, W. Current situation and protection analysis of juvenile fish resources in spring and autumn in Zhoushan coastal waters. *Fish. Res.* **2021**, *43*, 121–132.
5. Kelson, S.J.; Hogan, Z.; Jerde, C.L.; Chandra, S.; Ngor, P.B.; Koning, A. Fishing Methods Matter: Comparing the Community and Trait Composition of the Dai (Bagnet) and Gillnet Fisheries in the Tonle Sap River in Southeast Asia. *Water* **2021**, *13*, 1904. [CrossRef]
6. Wang, Y.; Wang, Y.; Lin, C.; Wei, Y.; Ma, W.; Wu, L.; Liu, D.; Shi, X. Review on monitoring methods of fish passage effect. *Chin. J. Ecol.* **2019**, *38*, 586–593.
7. Wang, L.; Chen, Y.; Tang, L.; Fan, R.; Yao, Y. Object-based convolutional neural networks for cloud and snow detection in high-resolution multispectral imagers. *Water* **2018**, *10*, 1666. [CrossRef]
8. Gadamsetty, S.; Ch, R.; Ch, A.; Iwendi, C.; Gadekallu, T.R. Hash-based deep learning approach for remote sensing satellite imagery detection. *Water* **2022**, *14*, 707. [CrossRef]
9. Pei, Q.; Peng, S.; Liu, Y. Real-time fish detection in fishway based on deep learning. *Inf. Commun.* **2019**, *2*, 67–69.
10. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
11. Wageeh, Y.; Mohamed, H.E.-D.; Fadl, A.; Anas, O.; ElMasry, N.; Nabil, A.; Atia, A. YOLO fish detection with Euclidean tracking in fish farms. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 5–12. [CrossRef]
12. Petro, A.B.; Sbert, C.; Morel, J.M. Multiscale retinex. *Image Proces. Line* **2014**, *4*, 71–88. [CrossRef]
13. Fan, W. Ocean Fish Image Recognition And Application Based on Deep Learning. Master's Thesis, Chongqing Normal University, Chongqing, China, 2019.
14. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
15. Yao, L. Underwater Target Recognition Based on Improved YOLOv4 Neural Network. *Electronics* **2021**, *10*, 1634.
16. Bochkovskiy, A.; Wang, C.; Liao, H.Y.M. Yolov4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2004**, arXiv:2004.10934.
17. Qiang, W.; He, Y.; Guo, Y.; Li, B.; He, L. Exploring underwater target detection algorithm based on improved SSD. *Xibei Gongye Daxue Xuebao/J. Northwest. Polytech. Univ.* **2020**, *38*, 747–754. [CrossRef]

18. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; Berg, A. SSD: Single Shot MultiBox Detector. In Proceedings of the Computer Vision—ECCV 2016, 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016.
19. Wu, R.; Bi, X.J. Benthic biometrics identification of coral reefs based on improved YOLOv5. *J. Harbin Eng. Univ.* **2022**, *4*, 580–586.
20. YOLOv5 in PyTorch. Available online: https://github.com/ultralytics/yolov5 (accessed on 6 July 2022).
21. Feng, Z.; Xie, Z.; Bao, Z.; Chen, K. Real-time dense small target detection algorithm for uav based on improved YOLOv5. *J. Aviation.* 1-15[2022-08-02]. Available online: http://kns.cnki.net/kcms/detail/11.1929.V.20220509.2316.010.html (accessed on 6 July 2022).
22. Dai, Y.; Zhao, X.; Li, L.; Liu, W.; Chu, X. Infrared dim small target detection algorithm in complex background based on improved yolov5 Infrared technology. *Infrared Technol.* **2022**, *44*, 504.
23. Wang, G.; Ding, H.; Yang, Z.; Li, B.; Wang, Y.; Bao, L. TRC-YOLO: A real-time detection method for lightweight targets based on mobile devices. *IET Comput. Vision* **2022**, *16*, 126–142. [CrossRef]
24. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. *arXiv* **2021**, arXiv:2103.02907.
25. Yu, J.; Jiang, Y.; Wang, Z.; Cao, Z.; Huang, T. Unitbox: An advanced object detection network. In Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016; pp. 516–520.
26. Rezatofighi, H.; Tsoi, N.; Gwak, J.Y.; 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.
27. Zheng, Z.; Wang, P.; Liu, W.; Li, J. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *arXiv* **2019**, arXiv:1911.08287. [CrossRef]
28. Zheng, Z.; Wang, P.; Ren, D.; Liu, W.; Ye, R.; Hu, Q.; Zuo, W. Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation. *IEEE Trans. Cybern.* **2021**, *52*, 8574–8586. [CrossRef] [PubMed]
29. Zhang, Y.F.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and Efficient *IOU* Loss for Accurate Bounding Box Regression. *Neurocomputing* **2022**, *506*, 146–157. [CrossRef]