


Article

DDEYOLOv9: Network for Detecting and Counting Abnormal Fish Behaviors in Complex Water Environments

Yinjia Li ^{1,2,3,4}, Zeyuan Hu ^{1,2,3,4,*} , Yixi Zhang ^{1,2,3,4}, Jihang Liu ^{1,2,3,4}, Wan Tu ^{1,2,3,4} and Hong Yu ^{1,2,3,4}

¹ College of Information Engineering, Dalian Ocean University, Dalian 116023, China; liyjee@126.com (Y.L.); lazyxx@163.com (Y.Z.); ljh1365590957@outlook.com (J.L.); wan.tu@outlook.com (W.T.); yuhong@dloou.edu.cn (H.Y.)

² Dalian Key Laboratory of Smart Fishery, Dalian 116023, China

³ Key Laboratory of Environment Controlled Aquaculture, Dalian Ocean University, Ministry of Education, Dalian 116023, China

⁴ Key Laboratory of Marine Information Technology of Liaoning Province, Dalian 116023, China

* Correspondence: huzy@dloou.edu.cn

Abstract: Accurately detecting and counting abnormal fish behaviors in aquaculture is essential. Timely detection allows farmers to take swift action to protect fish health and prevent economic losses. This paper proposes an enhanced high-precision detection algorithm based on YOLOv9, named DDEYOLOv9, to facilitate the detection and counting of abnormal fish behavior in industrial aquaculture environments. To address the lack of publicly available datasets on abnormal behavior in fish, we created the “Abnormal Behavior Dataset of *Takifugu rubripes*”, which includes five categories of fish behaviors. The detection algorithm was further enhanced in several key aspects. Firstly, the DRNELAN4 feature extraction module was introduced to replace the original RepNCSPPELAN4 module. This change improves the model’s detection accuracy for high-density and occluded fish in complex water environments while reducing the computational cost. Secondly, the proposed DCNv4-Dyhead detection head enhances the model’s multi-scale feature learning capability, effectively recognizes various abnormal fish behaviors, and improves the computational speed. Lastly, to address the issue of sample imbalance in the abnormal fish behavior dataset, we propose EMA-SlideLoss, which enhances the model’s focus on hard samples, thereby improving the model’s robustness. The experimental results demonstrate that the DDEYOLOv9 model achieves high *Precision*, *Recall*, and *mean Average Precision (mAP)* on the “Abnormal Behavior Dataset of *Takifugu rubripes*”, with values of 91.7%, 90.4%, and 94.1%, respectively. Compared to the YOLOv9 model, these metrics are improved by 5.4%, 5.5%, and 5.4%, respectively. The model also achieves a running speed of 119 frames per second (*FPS*), which is 45 *FPS* faster than YOLOv9. Experimental results show that the DDEYOLOv9 algorithm can accurately and efficiently identify and quantify abnormal fish behaviors in specific complex environments.

Keywords: aquaculture; fish behavior; YOLOv9; fish target detection; abnormal behavior monitoring

Key Contribution: The proposed DDEYOLOv9 model achieves the detection and counting of various abnormal behavior fish in industrial aquaculture environments. Its excellent detection accuracy and real-time performance contribute to enabling precision aquaculture and reducing aquaculture costs for farmers.



Citation: Li, Y.; Hu, Z.; Zhang, Y.; Liu, J.; Tu, W.; Yu, H. DDEYOLOv9: Network for Detecting and Counting Abnormal Fish Behaviors in Complex Water Environments. *Fishes* **2024**, *9*, 242. <https://doi.org/10.3390/fishes9060242>

Academic Editors: Xinjun Chen and Bin Liu

Received: 17 May 2024

Revised: 16 June 2024

Accepted: 19 June 2024

Published: 20 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Fish are an essential source of food for humans, and the aquaculture industry plays a vital role in the global agricultural economy. With the growth of population and economic development, the demand for fish is increasing, and the scale of fish farming is also becoming larger and larger, which brings new opportunities and challenges to fish farming [1]. Due to the high-density cultivation of fish in industrial aquaculture environments, coupled

with the complexity and uncontrollability of underwater environments, fish growth may be affected by various abnormal conditions, such as diseases, pollution, parasites, etc., leading to abnormal behaviors. This can result in a decrease in the quantity and quality of fish, causing significant losses to the aquaculture industry [2]. Therefore, it is crucial to quickly and accurately detect abnormal fish behavior to mitigate these risks and ensure the sustainable development of aquaculture.

For fish farmers, timely monitoring of the health status of fish populations is particularly important. Monitoring and counting instances of abnormal fish behavior can accurately assess the health status of a fish school in a breeding pool. A higher frequency of abnormal behavior should be given more attention and analyzed further to identify potential causes and implement appropriate measures in a timely manner. However, traditional methods for detecting abnormal fish behavior often rely on manual inspection or personal subjective experience, which can be time-consuming, expensive, and prone to errors. Without an automated detection system, determining the cause of death may require manual sampling from the farm, visually checking for abnormal symptoms, or waiting for dead organisms to float to the water surface. These manual methods are time-consuming, subjective, lack consistency, have a high error rate, and are difficult to quantify. Moreover, traditional manual contact detection methods can cause stress, injury, and even death in sensitive aquatic animals, severely impacting their growth and health. Therefore, there is an urgent need for technology that can automatically detect abnormal fish behavior to promote the healthy and sustainable development of the aquaculture industry.

Computer vision technology has advanced rapidly in recent years, offering a non-destructive and rapid method for detecting fish behavior [3]. The use of computer vision for fish behavior analysis has become a prominent research area. For instance, Yu et al. [4] utilized the Harris corner detection method to extract feature points of specific behaviors. They then employed the Lucas–Kanade optical flow method to determine the speed of carp in the sub-image, ultimately assessing whether the swimming speed of the fish school was abnormal. However, this approach relies on traditional computer vision techniques, necessitating the manual design of feature extraction algorithms and exhibiting certain limitations. When confronted with large and diverse datasets and complex recognition tasks, these methods often struggle to achieve optimal performance and accuracy, making them unsuitable for practical applications. Therefore, continuous exploration of more advanced and effective technologies is necessary to meet these challenges.

In recent years, deep learning models have gained popularity in the field of aquaculture due to their efficient feature representation capabilities [5]. These models, with their multi-layer learning networks, can extract semantic information from the pixel level. They are capable of performing end-to-end detection of instances of semantic objects, such as fish body targets, without the need for explicitly defined features. This characteristic makes them well suited for detecting fish behavior through image analysis. For instance, Gupta et al. [6] designed a convolutional neural network (CNN) based on two-dimensional images to identify fish with wounds on the surface. Their model achieved high recognition accuracy in classifying abnormal and normal fish. Similarly, Yu et al. [7] used an improved YOLOv4 model to detect four common fish skin diseases and their static features. These studies primarily focused on detecting specific appearance characteristics of fish bodies, overlooking the information related to fish health status contained in their behavior. To identify the phenomenon of dead fish, marked by fish turning over, Zhao et al. [8] utilized the lightweight MobileNetV3 network as the feature extraction backbone to reduce the number of parameters. They improved dead fish detection accuracy by incorporating deformable convolution into the YOLOv4 model. However, a limitation of this method is that it can only detect fish after they have died, and it cannot predict the health status of the fish before their death. Wang et al. [9] proposed an improved YOLOv5 model for detecting and tracking abnormal fish behavior. By modifying the path aggregation network, the model achieves the detection of small targets within fish schools, significantly enhancing its effectiveness in detecting abnormal fish behavior in ideal environments. However, further

improvement is needed to enhance the detection accuracy in complex industrial aquaculture settings. Subsequently, Wang et al. [10] utilized the multi-head attention mechanism in the YOLOv5 backbone network, improved the feature fusion performance based on the BiFPN concept, and employed the up-sampling operator Carafe to replace traditional up-sampling methods. These enhancements overcome challenges such as small target sizes, severe occlusions, and blurred targets in fish school images in real environments. Cai et al. [11] implemented the detection of SVC-infected fish schools using an improved YOLOv7 model. However, this application is relatively limited, as it is only applicable to detecting SVC infections and exhibits poor generalization, making it challenging to address the complex and diverse health conditions of fish. On the other hand, Wang et al. [12] employed the enhanced YOLOX-S algorithm for abnormal fish detection in complex aquatic environments. They improved the detection accuracy by incorporating coordinate attention. However, this enhancement increased the network's weight and parameters, reducing the detection speed and failing to meet real-time system requirements. Furthermore, all these studies can only differentiate between normal and abnormal fish schools, lacking the capability to specifically detect and analyze various abnormal behaviors.

Based on the above analysis, the current research on fish abnormal behavior detection based on deep learning faces several challenges: (1) In complex aquatic environments, the detection performance for high-density fish schools with occlusions, blurriness, and fish body deformation is poor, leading to feature loss and missed detections. (2) There is limited analysis of fish behavior characteristics, and the models' detection capabilities are relatively singular, only able to differentiate between normal and abnormal behaviors. This limitation makes it challenging to detect and differentiate between various abnormal behaviors. (3) Balancing between accuracy and model size is difficult, as the current models struggle to maintain high precision while meeting the need for fast detection.

In view of the above studies and existing problems, this paper takes the *Takifugu rubripes* in an industrial aquaculture environment as the research object and proposes a model based on improved YOLOv9, called DDEYOLOv9, to realize the real-time detection and counting of a variety of abnormal behavior fish in the fish school. The main contributions of this paper are as follows:

1. This study collected and created a dataset for recognizing abnormal fish behavior, called the "Abnormal Behavior Dataset of *Takifugu rubripes*". This dataset comprises 4000 annotated images of 50 *Takifugu rubripes*. This dataset fills a gap in resources for related research fields, providing valuable data support for researchers. By thoroughly analyzing this dataset, we can more accurately identify abnormal fish behavior, thereby providing strong support for the conservation of aquatic organisms and the maintenance of ecological balance.
2. This study designed the DRNELAN4 module to enhance the receptive field, improve the network's perception of global features, enable the model to better capture contextual information of input data, and alleviate issues such as image turbidity and occlusion in complex underwater environments for fish imagery.
3. The DCNV4-Dyhead detection head proposed in this paper effectively enhances the adaptability to scale transformation and shape change of detected fish, improves the perception ability and detection accuracy of the model, and enables the model to accurately detect various abnormal behaviors of fish through images.
4. By dynamically adjusting the weight and optimization strategy of easy samples and hard samples, the proposed EMA-SlideLoss loss function enables the model to pay more attention to fish with abnormal behaviors that are difficult to identify and fewer in number and alleviates the problem of sample imbalance in the dataset.

This article is divided into four sections. Section 2 describes the construction of the dataset and introduces the model used in this experiment. Section 3 presents the experimental results, and analyzes and discusses the experiments and their results. Lastly, Section 4 consists of the conclusions and future prospects of this research.

2. Materials and Methods

2.1. Data Acquisition and Annotation

2.1.1. Prepare the Required Materials

Due to its high nutritional value, protein content, and distinctive taste, *Takifugu rubripes* is highly favored by many, making it a valuable cultured fish with significant economic and culinary importance [13]. However, the high density and intensive aquaculture mode of *Takifugu rubripes* puts forward strict requirements for water quality management. The higher the breeding density, the greater the risk of water quality deterioration, which provides favorable conditions for the spread of diseases in the fish population, making the prevention and control of diseases an important link that cannot be ignored in the breeding process. Farmers must adopt more elaborate and efficient breeding strategies to ensure the stability of the breeding environment and the health of fish, so as to ensure the sustainable development of the aquaculture industry. Therefore, this study used the highly valuable *Takifugu rubripes* as the research object. The juvenile *Takifugu rubripes* were provided by the Dalian Tianzheng Industry Daheishi Aquaculture Workshop, with a body length of 8–9 cm and an average weight of 10 g, totaling 50 individuals.

The experimental platform consists of two parts (as shown in Figure 1). The first part is a breeding tank with water filtration and oxygen supply functions. The tank has a diameter of 60 cm, a height of 60 cm, and a water depth of 40 cm. Prior to the experiment, the experimental fish were temporarily raised in a circulating aquaculture tank for 8 weeks, maintaining a water temperature of 15–22 °C, pH of 6.5–6.9, and dissolved oxygen above 6.0 mg/L. The second part is the information collection system. A network camera was used for recording, with a video resolution of 1920 × 1080 and a frame rate of 60 fps. The camera was positioned above the breeding tank, 50 cm from the water surface, to capture the fish school. The video of the fish shoal was recorded and saved continuously on the computer.

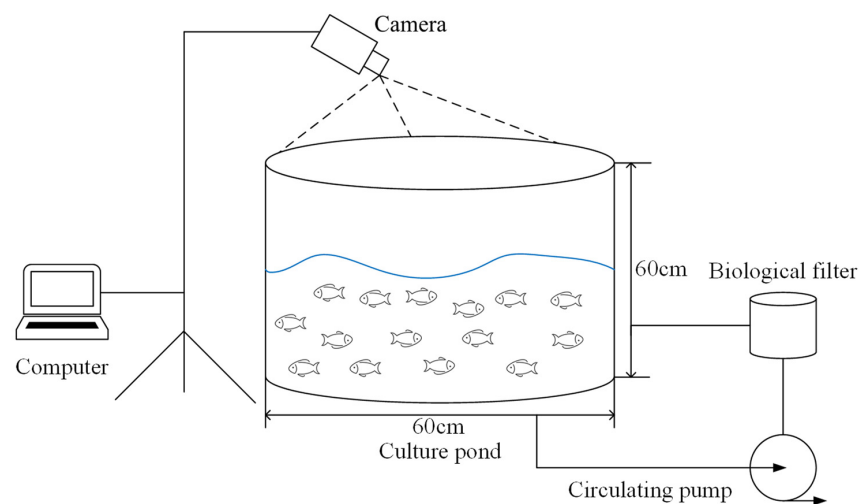


Figure 1. Image acquisition.

2.1.2. Data Acquisition

Adverse changes in aquatic environments can directly lead to abnormal physiological activities in fish [14,15]. Fish adjust their behavior according to minor changes in the ecosystem. Juvenile fish are relatively weak in vitality during the initial stage, making them susceptible to the influences of two key factors: the aquaculture water environment and dissolved oxygen levels [16,17]. We collected data on abnormal fish behavior in four abnormal environments: weakly acidic (pH abnormal), low temperature (below 15 °C), high temperature (above 25 °C), and hypoxia (low dissolved oxygen). Through observation, we found that, in the environment with abnormal pH levels, *Takifugu rubripes* exhibit “convulsion” behavior. In low-temperature environments, they display a “head down and

tail up” behavior. In high-temperature environments, some fish exhibit “rollover” behavior, while, in hypoxic environments, some fish show a “head up and tail down” behavior. This is illustrated in Figure 2.

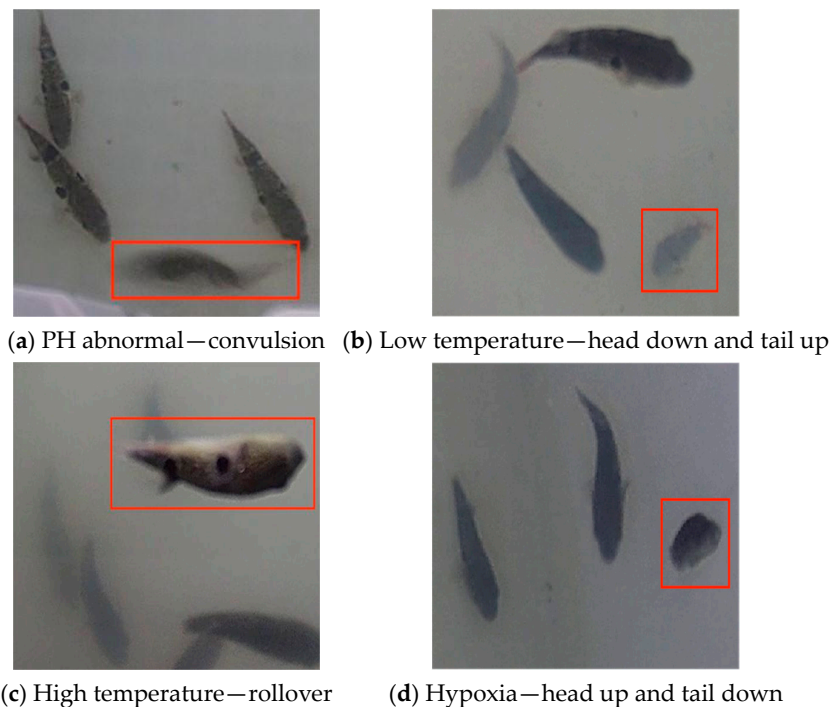


Figure 2. Abnormal behavior of *Takifugu rubripes* (framed fish with abnormal behavior).

The data collection equipment used in the experiment consisted of a laptop computer and a network camera. Continuous videos were selected and extracted frame by frame. After data cleaning and manual screening, a dataset comprising 1000 images for each of the four abnormal environments (“PH abnormal”, “low temperature”, “high temperature”, and “hypoxia”) was created for the detection of fish abnormal behavior.

2.1.3. Data Annotation and Dataset Construction

In this study, manual annotation was performed using LabelImg. The labels for the abnormal environments were as follows: “Normal-Fish” and “PH abnormal-Fish” for the pH abnormal environment, “Normal-Fish” and “Low temperature-Fish” for the low-temperature environment, “Normal-Fish” and “High temperature-Fish” for the high-temperature environment, and “Normal-Fish” and “Hypoxia-Fish” for the hypoxic environment. The dataset of 4000 images collected was randomly assigned at the ratio of 8:1:1 to construct a training set, a validation set, and a test set, named “Abnormal Behavior Dataset of *Takifugu rubripes*”. The sample distribution of the dataset is shown in Figure 3, and the training set and the test set cannot come from the same video sequence. This data will be used in the training of the model to identify abnormal fish behavior.

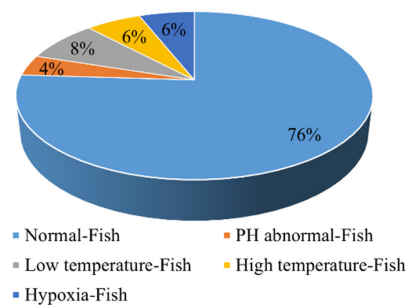


Figure 3. Sample distribution of the abnormal behavior dataset of *Takifugu rubripes*.

SSD [20]. These methods have advantages such as high detection rates and low memory usage [21,22]. The difference between these two types of models lies in their approach. Two-stage object detection methods first use a Region Proposal Network (RPN) to generate candidate regions, which are then classified and regressed upon. On the other hand, one-stage object detection methods directly classify and regress on the entire image to achieve object detection. This approach ensures both efficient processing speeds and high object detection accuracy, making it suitable for real-time monitoring tasks in industrial aquaculture facilities. As a representative of one-level detectors, YOLO series models have been favored by a large number of researchers for their excellent recognition accuracy and speed and are the mainstream choice for real-time detection projects.

The YOLOv9 [23] model represents the latest advancement in the YOLO series, with its main contributions being (1) the introduction of Programmable Gradient Information (PGI) to address the various changes required for deep network detection of multiple targets. PGI can provide complete input information for the target task to calculate the objective function, thereby obtaining reliable gradient information to update network weights. (2) A new lightweight network architecture based on gradient path planning, called the Generalized Efficient Layer Aggregation Network (GELAN), was designed to demonstrate the effectiveness of PGI. It integrates two network modules, the Cross Stage Partial Network (CSPNet) and the Effective Long-Range Aggregation Network (ELAN) [24], to generalize the capabilities of ELAN, supporting any computational block. A new feature extraction module, RepNCSPPELAN4, was proposed based on the GELAN architecture. The GELAN architecture takes into account lightweight, inference speed, and accuracy, contributing to the improved accuracy and robustness in object detection. In summary, in this study, the YOLOv9 model is used to realize the fish school abnormal behavior detection task, and the model is improved to achieve higher recognition accuracy.

The YOLOv9 algorithm provides five pre-trained network models: YOLOv9-T, YOLOv9-S, YOLOv9-M, YOLOv9-C, and YOLOv9-E. In this study, YOLOv9-E with higher accuracy is selected as the basic model to meet the needs of the network for high detection accuracy.

2.2.3. DRNELAN4 Model

Takifugu rubripes are highly cultured fish, and there are always residual bait and fish excreta in industrial aquaculture environments, which causes water turbidity. So that the collected image will change with the change in water quality, illumination, and fish status, resulting in fish features that are not obvious, there are different degrees of superposition, deformation, and occlusion. The original YOLOv9 model cannot extract clear features and has low detection accuracy. To solve this problem, this study uses Dilated Reparam Block to improve the feature extraction module RepNCSPPELAN4 in the YOLOv9 model and proposes the DRNELAN4 module to improve the model performance and reduce the inference cost.

In complex aquatic environments, the features of densely packed fish in images are often blurry, making traditional convolutional neural networks (CNNs) perform poorly in feature extraction. However, for CNNs of the same depth, using large kernel convolutions has a larger receptive field than using small kernel convolutions. Large receptive fields that do not rely on deep stacking have stronger feature extraction capabilities. Therefore, to improve the accuracy of fish detection in schools, the use of large kernel convolutions is particularly crucial. They can more effectively capture global features in images, thereby enhancing the overall detection performance.

However, the large kernel convolution requires more training parameters, while the extended convolution combines the advantages of a large receptive field and a small number of parameters. The Dilated Reparam Block uses a combination of a non-dilated small kernel and multiple dilated small kernel layers to enhance a non-dilated large kernel convolutional layer. Its hyperparameters include the size of the large kernel K , the size of the parallel convolutional layers k , and the dilation rates r . As shown in Figure 5, the case involves four parallel layers, where $K = 9$, $r = (1, 2, 3, 4)$, and $k = (5, 3, 3, 3)$. Firstly,

each batch normalization (BN) is merged into the preceding convolutional layer. Each layer with a dilation rate $r > 1$ is transformed into function 1, and all resulting kernels are added together using appropriate zero padding. As the value of K increases, the number of dilated layers used also increases accordingly. This not only makes the kernel size larger but also increases the dilation rate. This modification allows the Dilated Reparam Block to improve the network’s spatial information capturing capabilities while keeping the number of trainable parameters and computational efficiency constant. It enables the network to have a broader receptive field without increasing the model depth, thus achieving a favorable trade-off between precision and parameters.

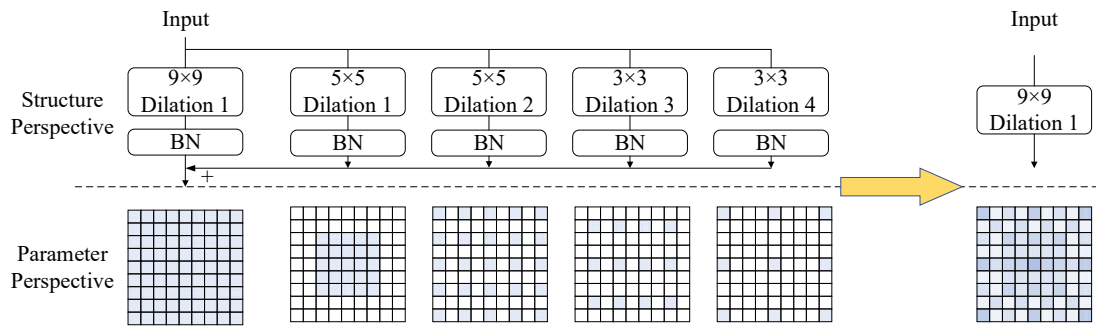


Figure 5. Dilated Reparam Block. A dilated small kernel conv layer is used to augment the non-dilated large kernel conv layer. From a parametric point of view, this dilated layer is equivalent to a non-dilated conv layer with a larger sparse kernel, so that the whole block can be equivalently transformed into a single large kernel conv.

When the features of fish schools in an image are not clear or are occluded by other fish, the RepNCSPPELAN4 module of YOLOv9 may not capture the complex details of the fish well. Additionally, this module is limited to perceiving features within a fixed range, making it difficult to comprehensively capture key information about the targets. This can result in reduced detection accuracy in complex environments and increase the likelihood of false positives. Aiming at the above problems, this study uses Dilated Reparam Block to improve the RepNCSPPELAN4 module in the YOLOv9 model and proposes the DRNELAN4 module, as shown in Figure 6. The Dilated Reparam Block replaces the RepConvN in RepNCSPPELAN4, which helps the model capture larger receptive fields when processing images, thereby enhancing the model’s ability to capture global features. This allows the model to obtain more global contextual information, aiding in more precise localization of fish targets in complex environments. It improves detection accuracy while reducing the model’s computational complexity.

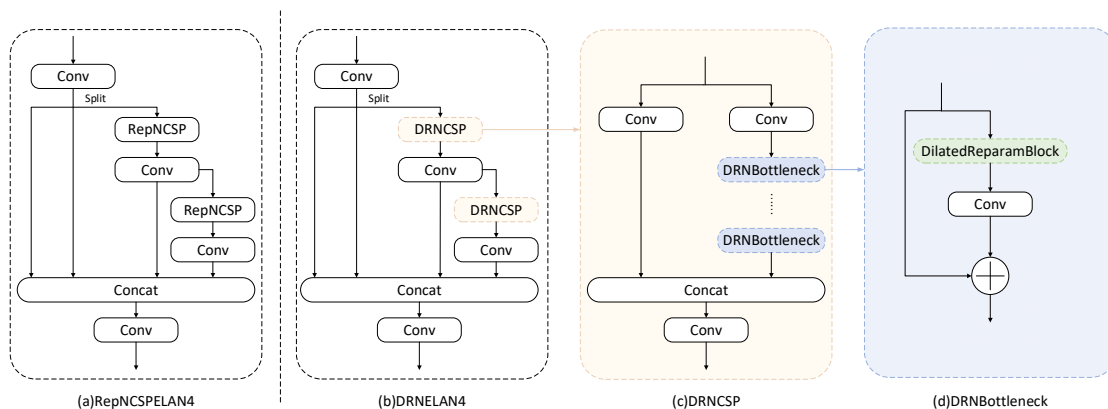


Figure 6. Comparison of improved DRNELAN4 and RepNCSPPELAN4 modules.

2.2.4. DCNv4-Dyhead Model

In different abnormal environments, *Takifugu rubripes* showed different abnormal behaviors, including “convulsion”, “head down and tail up”, “rollover”, “head up and tail down”, etc. In cases where there are many types of similar abnormal behaviors, the original YOLOv9 model may not achieve the desired detection results. This is because different abnormal behaviors in the image have different shape ratios and appear at different rotations and positions, which places higher demands on the model’s ability to learn multi-scale features. To better integrate the diversity of feature scales brought by the varying scales of different abnormal fish behaviors, and to capture the inherent spatial relationships between different scales and shapes, this study proposes the DCNv4-Dyhead detection head. It replaces the original detection head of the YOLOv9 model, enabling the model to effectively identify multiple similar abnormal fish behaviors.

The Dynamic Head (Dyhead) [25] method combines the multi-head self-attention mechanism of the scale-aware feature layer with the spatial perception of spatial positions and the task perception of the output channels. This integration significantly enhances the representation capability of the object detection head. The core components of the Dyhead framework include scale-aware attention, spatial attention, and channel attention. They respectively focus on features at different scales, spatial position information, and channel information. By overlaying these three types of attention, more comprehensive feature capture and more accurate detection results can be obtained. The calculation formula is given in Equation (1).

$$W(F) = \pi_C(\pi_S(\pi_L(F) \cdot F) \cdot F) \cdot F \quad (1)$$

The attention function is represented by the symbol W . The feature tensor $F \in R^{L \times S \times C}$ is a three-dimensional tensor, where L represents the hierarchy of the feature map, S represents the width–height product of the feature map, and C represents the number of channels in the feature map. The scale-aware attention module $\pi_L(\cdot)$, the spatial-aware attention module $\pi_S(\cdot)$, and the task-aware attention module $\pi_C(\cdot)$ are three different attention functions that operate on dimensions L , S , and C , respectively.

The formula for calculating the scale-aware attention π_L is shown in Equation (2). It can dynamically integrate features based on the semantic importance of different scales.

$$\pi_L(F) \cdot F = \sigma \left(f \left(\frac{1}{SC} \sum_{s,c} F \right) \right) \cdot F \quad (2)$$

In this process, the linear function $f(\cdot)$ is approximated using a 1×1 convolution operation, with $\sigma(x) = \max\left(0, \min\left(1, \frac{x+1}{2}\right)\right)$ serving as the activation function for this approximation.

Based on the feature fusion, the spatial perception attention module π_S can focus on discriminative regions that are consistent between spatial positions and feature layers. It first uses Deformable ConvNets v2 (DCNv2) [26] to sparsify the attention learning and then aggregates features across layers at the same spatial position. However, DCNv2 introduces additional overhead when sampling non-nearby positions, leading to a slower convergence speed. This study improved the deformable convolution part of the spatial perception attention module π_S by replacing the Deformable ConvNets v2 (DCNv2) module with the Deformable Convolution v4 (DCNv4) [27] module and proposed the DCNv4-Dyhead detection head.

DCNv4 is an efficient dynamic sparse operator that uses adaptive aggregation windows and dynamic aggregation weights with an unbounded value range. It removes softmax normalization in spatial aggregation, enhancing the dynamism and expressiveness of spatial aggregation (as shown in Figure 7). It also optimizes memory access to minimize redundant operations, improving speed. DCNv4 significantly accelerates model convergence and greatly improves processing speed.

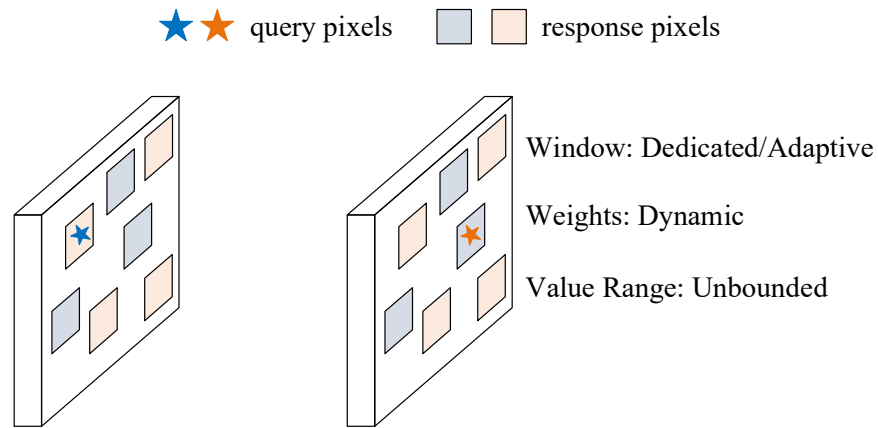


Figure 7. The core operation of spatial aggregation of query pixels at different locations in the same channel in DCNv4. DCNv4 combines DCNv3’s use of dynamic weights to aggregate spatial features and convolution’s flexible unbounded values for aggregate weights.

The improved π_S (Improved- π_S) is calculated as shown in Equation (3):

$$\pi_S(F) \cdot F = \frac{1}{L} \sum_{l=1}^L \sum_{g=1}^G \sum_{k=1}^K w_{l,g} \cdot F(l; p_k + \Delta p_{gk}; c) \cdot m_{gk} \quad (3)$$

where K is the number of sparse sampling positions, and G is the total number of aggregation groups. For the g -th group, $w_g \in R^{C \times C'}$ and $m_{gk} \in R$ represent the position-independent projection weights for that group, where $C' = C/G$ represents the dimensions of that group. $m_{gk} \in R$ represents the modulation scalar for the k -th sampling point in the g -th group, which is normalized along the K dimension using the softmax function. Δp_{gk} is the offset corresponding to the grid sampling position p_k of the g -th group. Both are learnt from the input features at the median level of F .

To achieve joint learning and generalize different representations of objects, a task-aware attention π_C (Task-aware Attention) is deployed at the end, as shown in Equation (4). It uses dynamic on-off functions to support different tasks:

$$\pi_C(F) \cdot F = \max(\alpha^1(F) \cdot F_c + \beta^1(F), \alpha^2(F) \cdot F_c + \beta^2(F)) \quad (4)$$

Here, F_c is the feature slice of the c -channel, and $[\alpha^1, \alpha^2, \beta^1, \beta^2]^T = \theta(\cdot)$ is a hyper-function that learns to control the activation threshold. Applying these three attention mechanisms sequentially can stack them multiple times to form Dyhead blocks. The DCNv4-Dyhead structure proposed in this study is shown in Figure 8.

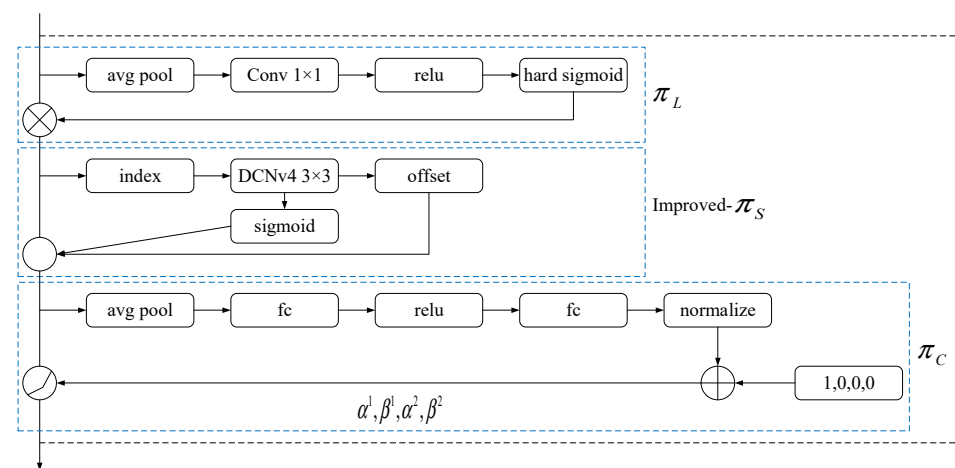


Figure 8. Structure of DCNv4-Dyhead.

The working principle of DCNv4-Dyhead is shown in Figure 9. After passing through the scale-aware attention module, the feature map becomes more sensitive to the multi-scale differences of the foreground fish targets. After passing through the improved spatial perception attention module in this study, the dynamic and sparse nature of the DCNv4 module enables the feature map to focus on discriminative spatial positions of the foreground objects, adapting to the scale variations of various similar fish abnormal behaviors and improving the model’s processing speed. Finally, the feature maps are reshaped based on the requirements of different downstream tasks through the task-aware attention module, forming different activations.

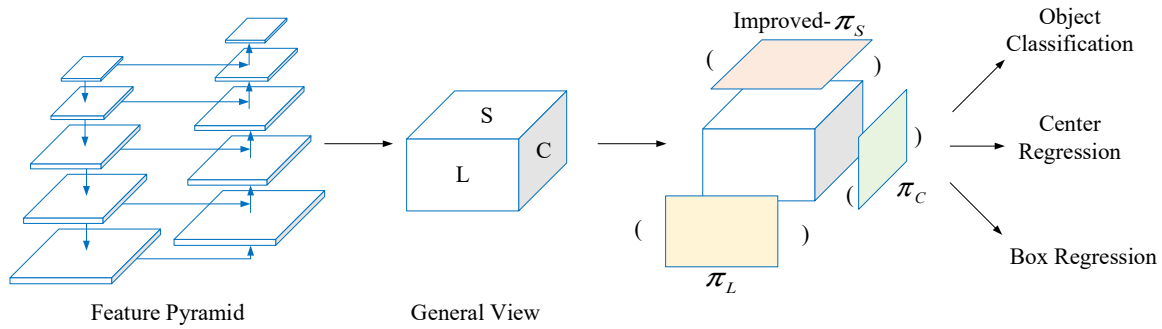


Figure 9. An illustration of the DCNv4-Dyhead approach.

2.2.5. EMA-SlideLoss

The purpose of this study is to quickly and accurately detect a small number of abnormal behaviors in fish in aquaculture environments, thereby effectively preventing the large-scale occurrence of diseased fish in aquaculture ponds. Therefore, to better simulate real aquaculture conditions, the dataset of fish abnormal behaviors contains significantly fewer samples of fish exhibiting harder-to-identify abnormal behaviors compared to the number of samples showing normal behaviors. This imbalance in sample data directly reduces the model’s detection performance in multiple target categories, affecting its accuracy and reliability in practical applications. The original loss function of the YOLOv9 model does not consider sampling difficulties and calculates the loss uniformly on all samples. This results in some low-confidence detection frames being retained while some high-confidence detection frames are suppressed, leading to the poor detection performance of abnormal behavior in fish. To address the limitation of sample imbalance, this study proposed a new loss function, Exponential Moving Average Sample Weighting Function (EMA-SlideLoss), based on SlideLoss. By assigning higher weights to hard samples, EMA-SlideLoss helps the model learn more difficult features.

SlideLoss uses adaptive parameters to solve the sample imbalance problem. The weight of the loss function for positive and negative samples is adjusted according to the IoU value between the predicted frame and the real target frame, so that the model pays more attention to the samples that are difficult to detect, so as to improve the overall detection accuracy. The implementation principle is shown in Equation (5):

$$f(x) = \begin{cases} 1 & x \leq \mu - 0.1 \\ e^{(1-\mu)} & \mu - 0.1 < x < \mu \\ e^{(1-x)} & x \geq \mu \end{cases} \quad (5)$$

Among them, μ is the average IoU value of all bounding boxes. Samples with an IoU value greater than μ are considered easy samples, while those with an IoU value lower than μ are considered hard samples. Due to the model’s poor ability to recognize hard samples, SlideLoss amplifies the relative loss of hard samples, while emphasizing samples that are misclassified, enabling the model to effectively utilize the limited instance features of hard samples.

SlideLoss considers the temporal information of targets in real aquaculture environments. However, due to limitations in the propagation speed of water, the behavior of fish in images may exhibit motion blur, deformation, and other phenomena. SlideLoss cannot achieve a truly smooth “slide” in adjusting the mean μ . Therefore, this study adopts the idea of Exponential Moving Average (EMA) to weigh the values of the time series, proposing a new loss function called EMA-SlideLoss. Specifically, based on SlideLoss, EMA-SlideLoss dynamically adjusts the value of μ ($auto_mu$) using the exponential moving average and adaptively adjusts the loss according to the dynamically adjusted $auto_mu$. This helps improve the robustness and training effectiveness of the model. The dynamically adjusted function is shown in Equation (6):

$$EMA_t = d \cdot \left(1 - \exp\left(-\frac{t}{t_{all}}\right) \right) \cdot EMA_{t-1} + (1 - d) \cdot auto_mu \quad (6)$$

Here, d represents the decay coefficient, t represents the iteration number, and t_{all} represents the total number of training iterations. When $t = t_{all}$, the decay factor will increase to $(e - 1)/e$. This means that, in the early stages of training, the model considers historical data more, while, in the later stages of training, the model considers the current observations more. Additionally, we gradually reduce the weight assigned to hard samples to prevent the excessive interference caused by these challenging instances throughout the entire training process. EMA-SlideLoss provides a smooth mechanism to balance the influence of historical data and current observations on the loss. The calculation of the final loss function is shown in Equation (7):

$$f(x) = \begin{cases} 1 & x \leq EMA_t - 0.1 \\ e^{1-EMA_t} & EMA_t - 0.1 < x < EMA_t \\ e^{1-x} & x \geq EMA_t \end{cases} \quad (7)$$

2.3. Experimental Platform and Model Training Parameters

2.3.1. Experiment Platform and Training Hyperparameters

The experiment of this study uses a consistent experimental environment: for the Windows 10 operating system, the detailed software and hardware environment configuration is shown in Table 1.

Table 1. Experimental platform.

Platform	Version
CPU	Intel(R) Core(TM) i7-12700, 2.1 GHz
GPU	GeForce RTX 3070 Ti
CUDA/CUDNN	V 11.3.1/V 8.2.1
Python	V 3.8
Pytorch	V 1.10.0

The experimental hyperparameters are as follows: an initial learning rate of 0.01, input image size of 640×640 , 200 epochs, and a batch size of 8.

2.3.2. Evaluation Criteria

The aim of this study is to establish a fish abnormal behavior detection model that balances detection accuracy and speed. *Precision* (P) represents the percentage of fish correctly predicted in the predicted results, while *Recall* (R) represents the percentage of fish correctly predicted out of the total number of fish. The *mean Average Precision* (mAP) is calculated based on the *Precision-Recall* (PR) curve, which is composed of *Precision* and *Recall*. mAP can comprehensively evaluate the model’s detection ability for targets of different sizes and shapes, providing a more objective reflection of the model’s accuracy.

Therefore, this study uses three metrics to evaluate the accuracy of the model: *Precision*, *Recall*, and *mAP*. The equations are as follows:

$$Precision = \frac{TP}{FP + TP} \quad (8)$$

$$Recall = \frac{TP}{FN + TP} \quad (9)$$

$$AP = \int_0^1 P(R)R \quad (10)$$

$$mAP = \sum_{i=1}^k \frac{AP}{k} \quad (11)$$

FPS (*Frames Per Second*) refers to the number of frames detected per second. In the practical application of aquaculture, it is essential to detect the abnormal status of fish in real time. *FPS*, as one of the performance evaluation indicators, can reflect the model's advantage in processing speed. This study uses *FPS* to evaluate the real-time performance of the model.

2.3.3. Experimental Design

This study conducted three sets of experiments: (1) By comparing the original YOLOv9 model with the improved DDEYOLOv9 model on the “Abnormal Behavior Dataset of *Takifugu rubripes*”, the effectiveness of this model in detecting fish datasets with multiple abnormal behaviors in complex environments was verified. (2) Conducting ablation experiments on the “Abnormal Behavior Dataset of *Takifugu rubripes*” to verify the effectiveness of the proposed DRNELAN4 module, DCNv4-Dyhead detection head, and EMA-SlideLoss loss function and to demonstrate the rationality of their fusion. (3) Comparing it with advanced underwater target detection models to verify the superiority of the DDEYOLOv9 algorithm.

3. Results and Discussion

3.1. Comparison Experiment before and after Model Improvement

This study conducted comparative experiments on the original model and the improved DDEYOLOv9 model on the “Abnormal Behavior Dataset of *Takifugu rubripes*” to verify the effectiveness of the improved DDEYOLOv9 model. The training results of YOLOv9 and DDEYOLOv9 are shown in Figures 10 and 11. Figure 10a–c show the performance comparison of the two models in detecting all types of fish behaviors, including *Precision* (*P*), *Recall* (*R*), and *mAP*. Figure 11a–c provide a detailed comparison of the detection accuracy of each behavior category by the two models.

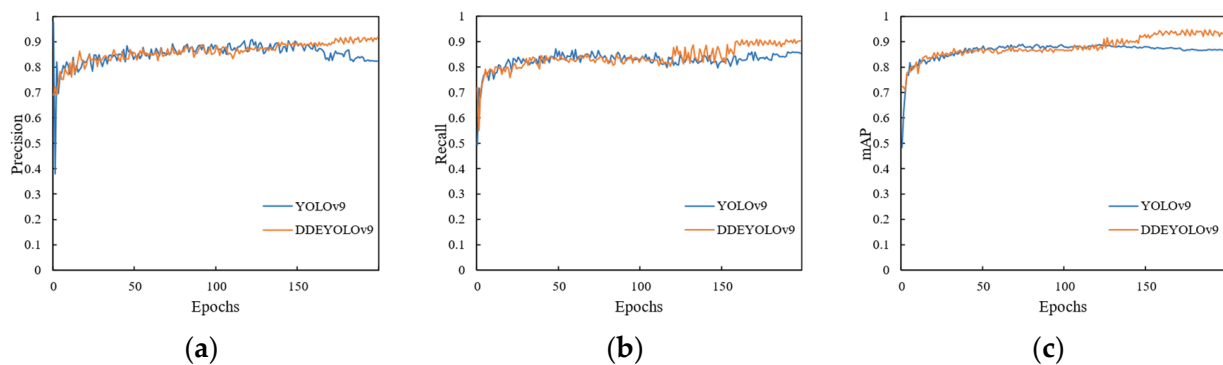
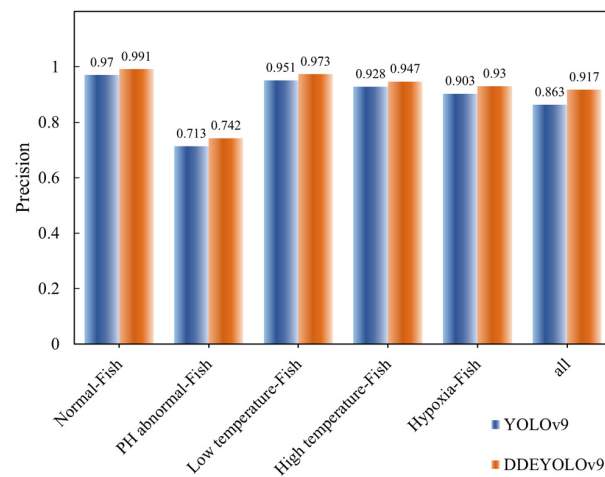
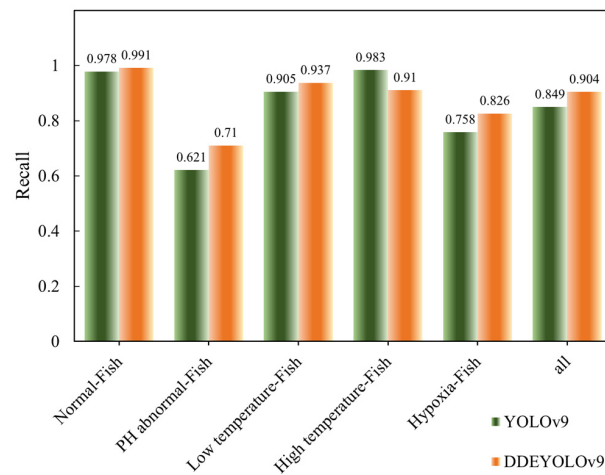


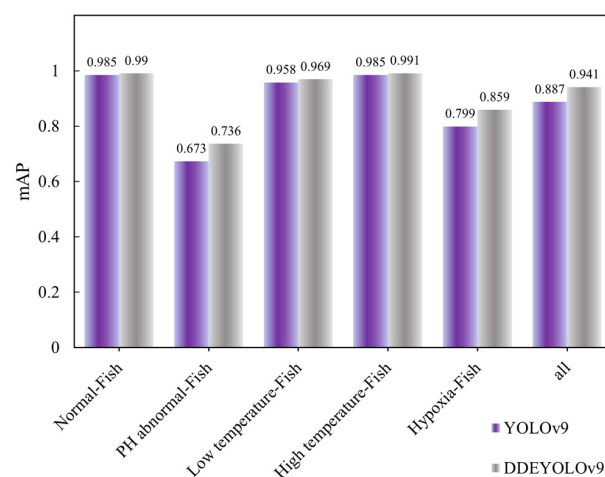
Figure 10. Comparison of the learning curves of the training dataset before and after improvement. (a) shows the *Epochs – Precision* curves of YOLOv9 and DDEYOLOv9 models. (b) shows the curve of *Epochs – Recall*; (c) shows the plot of the *Epochs – mAP*.



(a)



(b)



(c)

Figure 11. Comparison of accuracy before and after improvement. (a) shows the bar graph of *Precision* comparison for the six behavioral categories of the shoal; (b) shows the *Recall* comparison bar charts for the six behaviors; (c) presents the *mAP* versus bar charts for the five behaviors.

From Figure 10a–c, it can be seen that, after 50 epochs, the *mAP* curves of both models tend to stabilize. The *P*, *R*, and *mAP* curves of DDEYOLOv9 are higher than those of the original YOLOv9 model. This is because the improved DDEYOLOv9 model includes specific enhancements such as dynamic tuning and better feature extraction mechanisms, which allow it to exploit additional information and improve performance after the initial learning phase. Although both models use the same learning hyperparameters, the improved DDEYOLOv9 can benefit more from the learning rate adjustment over time, allowing it to escape from local minima and achieve better performance. From Figure 11a–c, it can be seen that, in the detection of all fish behavior categories, the *P*, *R*, and *mAP* of DDEYOLOv9 reach 91.7%, 90.4%, and 94.1%, respectively, which are 5.4%, 5.5%, and 5.4% higher than those of YOLOv9, validating the effectiveness of the improvement techniques in this study.

Figure 11a–c demonstrate that, on four challenging hard sample categories (“PH abnormal-Fish”, “Low temperature-Fish”, “High temperature-Fish”, and “Hypoxia-Fish”), DDEYOLOv9 achieves significantly improved detection accuracy compared to YOLOv9, with a *mAP* increase of 6.3%, 1.1%, 0.6%, and 6%, respectively. The experimental results demonstrate that the proposed method effectively addresses the issue of sample imbalance in the fish behavior dataset. By adopting the EMA-SlideLoss loss function, the DDEYOLOv9 algorithm successfully mitigates the challenges posed by sample imbalance, effectively improving the performance of the model in handling imbalanced datasets, validating its feasibility and effectiveness in practical applications.

3.2. Ablation Experiments

The ablation experiments aimed to verify the optimization effects of each enhancement module. In this study, we conducted an ablation analysis on DDEYOLOv9, incorporating specific enhancement functions into the YOLOv9 model, namely DRNELAN4, DCNv4-Dyhead, and EMA-SlideLoss. The experimental results, as shown in Table 2, demonstrate that each module contributes to varying degrees of improvement in the accuracy of DDEYOLOv9. Figure 12 visually presents the detection results of fish abnormal behaviors in different abnormal environments in the dataset.

Complex underwater environments and fish density are key factors affecting the detection of aquatic targets. When the fish density is too high, there can be occlusion and overlap between fish, which makes it difficult for the target to be recognized, thereby reducing the performance of the detection algorithm. Additionally, unclear images of fish groups exhibit limited target information, which also increases the difficulty of fish detection. The experimental results using the DRNELAN4 feature extraction module, as shown in Model 1, improved the YOLOv9 model’s *P* from 86.3% to 88.6%, *R* from 84.9% to 86.4%, and *mAP* from 88.7% to 89.6%. The results indicate that the DRNELAN4 module effectively enhances the model’s receptive field, improves the network’s perception of global features, and helps the model better capture contextual information of the input data in complex, blurry underwater images. Additionally, the Dilated Reparam Block reparameterizes the convolutional layers to improve the performance without the need for additional inference costs, achieving a good balance between accuracy and parameters.

Table 2. Ablation experimental results.

Model	DRBGELAN	DCNv4-Dyhead	EMA-SlideLoss	Precision P/%	Recall R/%	Mean Average Precision mAP/%	Frames per Second FPS/ $f \cdot s^{-1}$
YOLOv9				86.3	84.9	88.7	74
Model 1	✓			88.6	86.4	89.6	103
Model 2		✓		89.4	86.8	90.2	86
Model 3			✓	90.2	89.8	91.5	74
Model 4	✓	✓		90.6	87.9	91.9	119
Model 5	✓		✓	91.4	90.1	92.5	103
Model 6		✓	✓	90.8	90.3	91.8	86
DDEYOLOv9	✓	✓	✓	91.7	90.4	94.1	119

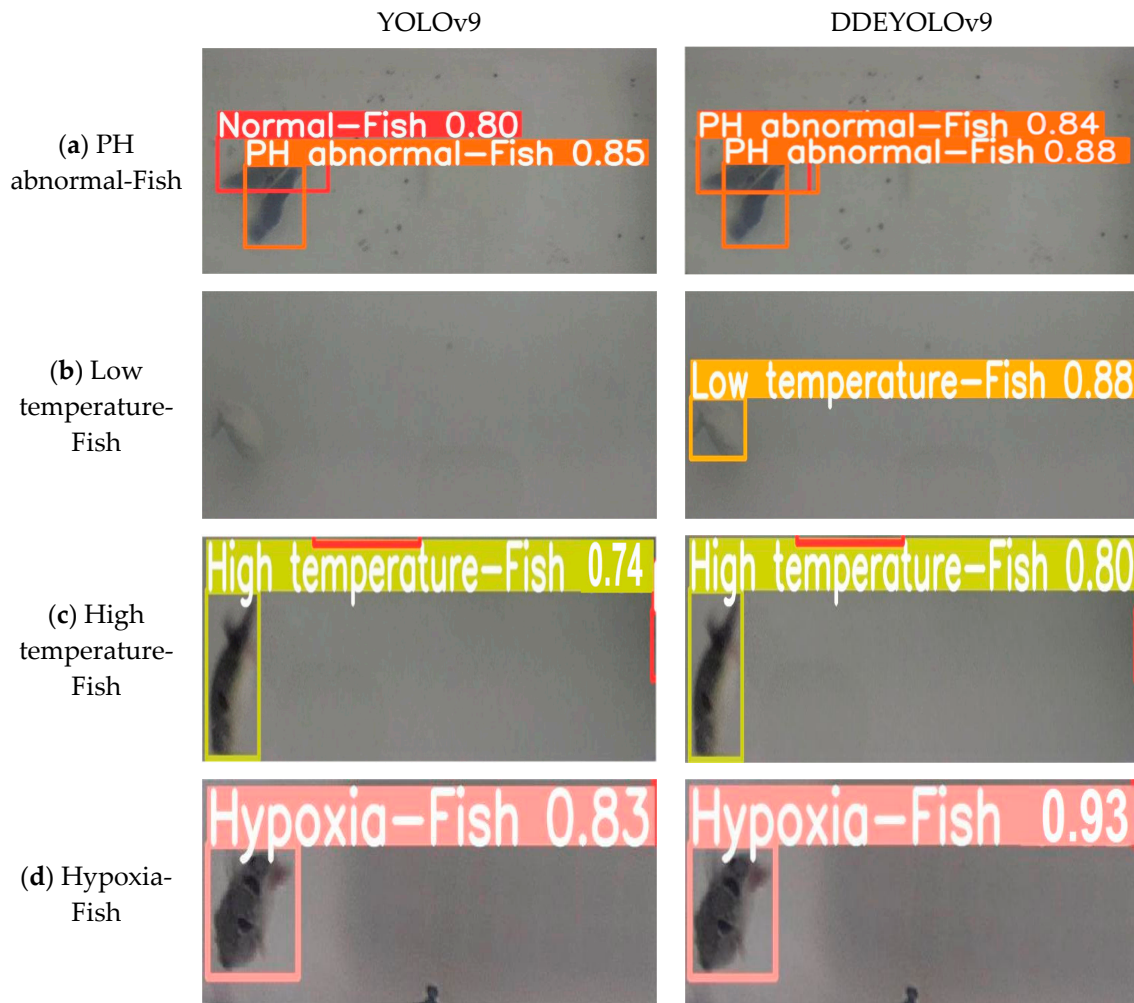


Figure 12. Renderings of the detection of abnormal behaviors of fish in different abnormal environments ((a) YOLOv9 has false detection, and (b) YOLOv9 has missed detection).

The complex and diverse behavioral characteristics of fish groups increase the difficulty of fish state classification. After replacing the original detection head of the YOLOv9 model with the DCNv4-Dyhead detection head proposed in this study, the experimental results, as shown in Model 2, demonstrate improvements in the model's P by 3.1%, R by 1.9%, and mAP by 1.5%. The Dyhead improves the model's adaptability to changes in the size and shape of the detected fish by introducing advanced feature extraction and attention mechanisms. This enhancement allows the model to effectively identify various abnormal behaviors in fish groups. Additionally, the DCNv4-Dyhead module proposed in this study utilizes the DCNv4 module to accelerate the convergence and computation speed of the algorithm.

Replacing the original loss function in the YOLOv9 model with EMA-SlideLoss, the experimental results, as shown in Model 3, demonstrate improvements in P , R , and mAP by 3.9%, 4.9%, and 2.8%, respectively, across all categories in the dataset. In the challenging hard sample categories of "PH abnormal-Fish", "Low temperature-Fish", "High temperature-Fish", and "Hypoxia-Fish", where detection is more difficult, the mAP of DDEYOLOv9 significantly exceeds that of YOLOv9 by 6.3%, 1.1%, 0.6%, and 6%, respectively (as illustrated in Figure 11). It is verified that, for the problem of sample imbalance in the fish abnormal behavior dataset, EMA-SlideLoss can adaptively adjust the weights of positive and negative samples, so that the model pays more attention to the samples that are difficult to classify, which is robust in underwater target detection and effectively

improves the overall detection accuracy. At the same time, the computational complexity of the model is not increased.

In the experiments of Model 4 to Model 6, we respectively incorporated two out of the three enhancement modules. The results indicate that the detection P , R , and mAP have all shown varying degrees of improvement compared to YOLOv9, validating the necessity of each enhancement module and the rationale behind their integration. The DDEYOLOv9 model, which combines all three enhancement modules, exhibits the best detection performance, suggesting a strong coupling between the three enhancement modules. Compared to the original YOLOv9, DDEYOLOv9 demonstrates significant performance enhancement.

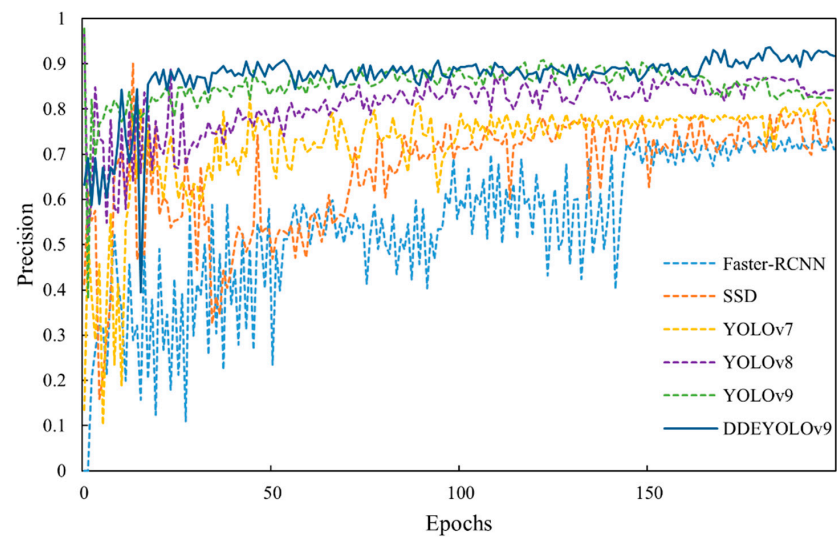
3.3. Model Comparison Experiment

In order to further verify the detection performance of the proposed model, it is compared with several mainstream underwater object detection algorithms, including Faster-RCNN [18], SSD [20], YOLOv7 [28], YOLOv8, and the baseline model YOLOv9 [23]. The same training method was used to train the network model on the “Abnormal Behavior Dataset of *Takifugu rubripes*”. Using P , R , mAP , and FPS as the main evaluation indicators for the experimental comparison, the experimental results are shown in Table 3. The performance comparison plots of different model detections are shown in Figure 13.

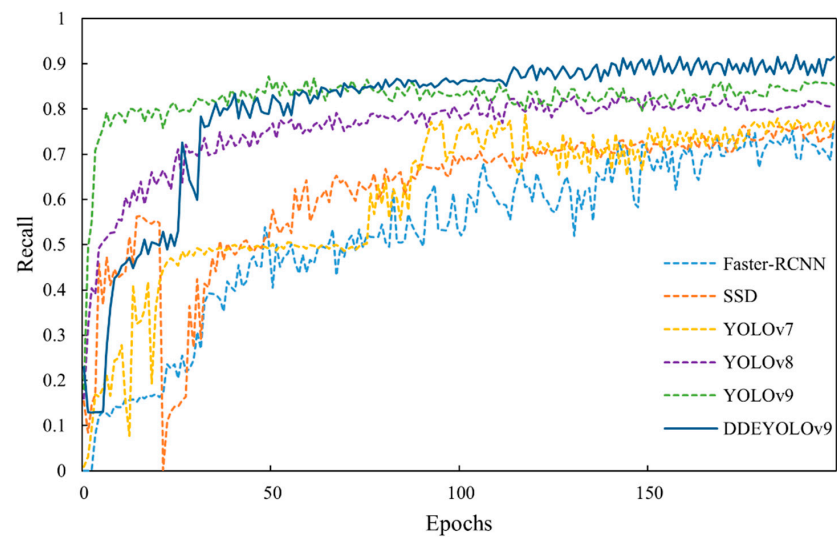
Table 3. Model comparison of experimental results.

Model	Precision P/%	Recall R/%	Mean Average Precision mAP/%	Frames per Second FPS/f·s ⁻¹
Faster-RCNN	73.6	76.8	77.1	32
SSD	77.4	77.2	79	45
YOLOv7	80.3	79.6	82.1	62
YOLOv8	86.5	79.7	85.7	66
YOLOv9	86.3	84.9	88.7	74
DDEYOLOv9	91.7	90.4	94.1	119

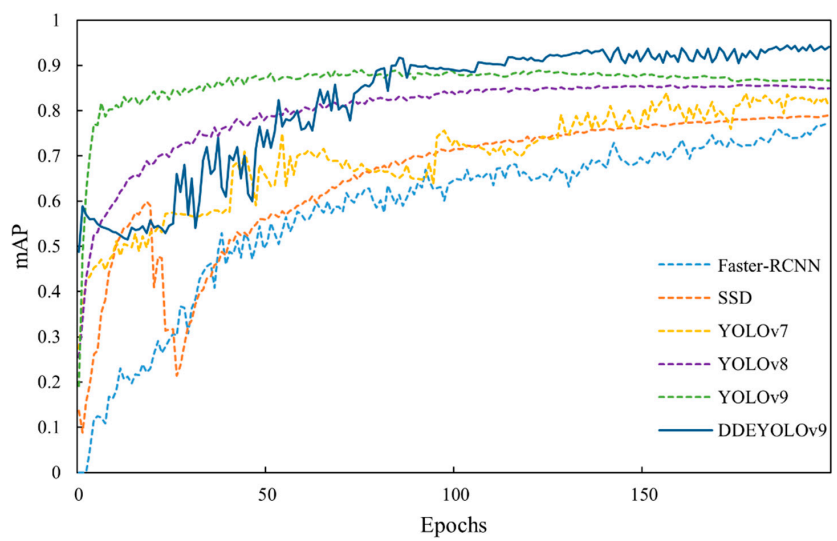
The analysis in Table 3 shows that the proposed algorithm is superior to other models in the detection of abnormal behavior of *Takifugu rubripes* in P , R , and mAP , and the detection speed is the fastest, reaching 91.7%, 90.4%, and 94.1%, which can realize the real-time detection of abnormal behavior of fish in real breeding environments. For other single-stage comparative models, they perform well in detecting normal-sized targets but struggle to accurately recognize small target fish schools with insufficient feature information in complex environments. Additionally, when detecting fish exhibiting multiple behaviors, issues such as overlap and occlusion lead to decreased model performance, making it challenging to accurately identify various fish features. Moreover, in non-balanced sample data, the number of easy samples is huge, controlling the variation of the loss, which causes the model to only learn the characteristics of easy samples while ignoring the learning of difficult samples, thereby affecting the overall detection performance of the model. The detection speed of DDEYOLOv9 is improved by 87FPS compared to Faster-RCNN. This is because Faster-RCNN requires generating candidate regions and then classifying these regions after refining their positions, which increases the computational complexity and time cost. Through comparative experiments, it has been verified that the DDEYOLOv9 algorithm has faster processing speed and better detection capability for abnormal fish behaviors in complex environments.



(a)



(b)



(c)

Figure 13. Performance comparisons. (a–c) show the *Epochs – Precision*, *Epochs – Recall*, and *Epochs – mAP* curves of the six models respectively.

4. Conclusions

To address the issue of poor identification of abnormal fish behaviors in specific complex aquatic environments, this study proposes a high-precision real-time detection algorithm based on YOLOv9, named DDEYOLOv9, and the “Abnormal Behavior Dataset of *Takifugu rubripes*”, which contains 4000 images, abnormal behavior of fish in four abnormal situations and one normal behavior, was established for this research task. To achieve better performance, this study first addressed the challenges posed by complex aquatic environments, such as turbid water, high-density fish aggregations, and mutual occlusion, by designing the DRNELAN4 feature extraction module. This module expands the model’s receptive field and enhances its ability to perceive global features, aiding in learning the contextual information of fish targets and compensating for the feature loss caused by unclear or occluded fish bodies. Secondly, to tackle the difficulty in detecting and classifying multiple fish behaviors with similarities, the study designed the DCNv4-Dyhead detection head. This head seamlessly integrates the multi-head self-attention mechanism of the scale-aware feature layer, spatial perception of spatial positions, and task perception of output channels to adapt to fish behavior features at different scales. This minimizes redundant operations and achieves the comprehensive fusion of multi-scale information. Lastly, to address the issue of imbalance between normal and abnormal behavior data samples in the dataset, the study designed the EMA-SlideLoss loss function. By adaptively learning the threshold parameters of positive and negative samples and assigning higher weights to hard samples, the model focuses more on learning abnormal behaviors.

To verify the effectiveness of the above improvements, three kinds of experiments were designed in this study. In the comparative experiments, the DDEYOLOv9 model achieved P , R , and mAP of 91.7%, 90.4%, and 94.1%, respectively, which were 5.4%, 5.5%, and 5.4% higher than those of the original YOLOv9 model. The detection speed reached 119 FPS, which was 45 FPS higher than that of YOLOv9. Both the detection accuracy and speed were significantly higher than those of other mainstream object detection algorithms in the comparative experiments, demonstrating the superiority of this model. Secondly, in the ablation experiment, we evaluated the contribution of each component of the improvement module to the overall performance one by one to ensure its effectiveness and practicality in the whole improved model. The experimental results showed that the proposed improved module achieved significant improvement in P , R , and mAP , which proved that they had good integration.

This study surpasses the traditional manual recognition method in detecting the abnormal behavior of fish in a specific complex breeding environment and can provide valuable technical support for the automation and intelligence of fish abnormal behavior recognition and counting. The DDEYOLOv9 model has the potential to be applied in the aquaculture industry, enabling the early detection of abnormal fish behaviors in complex aquatic environments while reducing farming costs. This can help prevent fish diseases, thereby improving aquaculture quality and reducing losses. Additionally, it offers beneficial decision support for disease warning in the aquaculture industry. In future work, we will further expand the scale of the dataset to improve the model’s generalization ability and apply it to a wider range of fields and scenarios.

Author Contributions: Conceptualization, Y.L.; methodology, Y.L.; resources, Y.L. and W.T.; data curation, Y.Z. and J.L.; writing—original draft preparation, Y.L. and Z.H.; writing—review and editing, Y.L., Z.H. and H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Liaoning Province Science and Technology Plan Joint Fund (2023-BSBA-001), the Basic Scientific Research Project of the Liaoning Provincial Department of Education (JYTQN2023132), the Key Projects of the Educational Department of Liaoning Province (LJKZ0729), the Key R&D Projects in Liaoning Province (2023JH26/10200015), and the Key Laboratory of Environment Controlled Aquaculture (Dalian Ocean University) at the Ministry of Education (202313).

Institutional Review Board Statement: The study was approved by the Ethics Review Committee of Dalian Smart Fisheries Key Laboratory, Approval Code: 2024003, Approval Date: 12 March 2024.

Data Availability Statement: Data are contained within the article. The “Abnormal Behavior Dataset of *Takifugu rubripes*” is publicly available and can be downloaded from the following link: <https://doi.org/10.6084/m9.figshare.26038312>, accessed on 14 June 2024.

Acknowledgments: We greatly appreciate the careful and precise reviews by the anonymous reviewers and editors.

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

References

1. Yang, X.; Zhang, S.; Liu, J.; Gao, Q.; Dong, S.; Zhou, C. Deep learning for smart fish farming: Applications, opportunities and challenges. *Rev. Aquac.* **2021**, *13*, 66–90. [CrossRef]
2. Zhao, S.; Zhang, S.; Liu, J.; Wang, H.; Zhu, J.; Li, D.; Zhao, R. Application of machine learning in intelligent fish aquaculture: A review. *Aquaculture* **2021**, *540*, 736724. [CrossRef]
3. Zhang, L.; Li, B.; Sun, X.; Hong, Q.; Duan, Q. Intelligent fish feeding based on machine vision: A review. *Biosyst. Eng.* **2023**, *231*, 133–164. [CrossRef]
4. Yu, X.; Wang, Y.; An, D.; Wei, Y. Identification methodology of special behaviors for fish school based on spatial behavior characteristics. *Comput. Electron. Agric.* **2021**, *185*, 106169. [CrossRef]
5. Kaur, G.; Adhikari, N.; Krishnapriya, S.; Wawale, S.G.; Malik, R.Q.; Zamani, A.S.; Perez-Falcon, J.; Osei-Owusu, J. Recent advancements in deep learning frameworks for precision fish farming opportunities, challenges, and applications. *J. Food Qual.* **2023**, *2023*, 4399512. [CrossRef]
6. Gupta, A.; Bringsdal, E.; Knausgård, K.M.; Goodwin, M. Accurate wound and lice detection in Atlantic salmon fish using a convolutional neural network. *Fishes* **2022**, *7*, 345. [CrossRef]
7. Yu, G.; Zhang, J.; Chen, A.; Wan, R. Detection and identification of fish skin health status referring to four common diseases based on improved yolov4 model. *Fishes* **2023**, *8*, 186. [CrossRef]
8. Zhao, S.; Zhang, S.; Lu, J.; Wang, H.; Feng, Y.; Shi, C.; Li, D.; Zhao, R. A lightweight dead fish detection method based on deformable convolution and YOLOV4. *Comput. Electron. Agric.* **2022**, *198*, 107098. [CrossRef]
9. Wang, H.; Zhang, S.; Zhao, S.; Wang, Q.; Li, D.; Zhao, R. Real-time detection and tracking of fish abnormal behavior based on improved YOLOV5 and SiamRPN++. *Comput. Electron. Agric.* **2022**, *192*, 106512. [CrossRef]
10. Wang, H.; Zhang, S.; Zhao, S.; Lu, J.; Wang, Y.; Li, D.; Zhao, R. Fast detection of cannibalism behavior of juvenile fish based on deep learning. *Comput. Electron. Agric.* **2022**, *198*, 107033. [CrossRef]
11. Cai, Y.; Yao, Z.; Jiang, H.; Qin, W.; Xiao, J.; Huang, X.; Pan, J.; Feng, H. Rapid detection of fish with SVC symptoms based on machine vision combined with a NAM-YOLO v7 hybrid model. *Aquaculture* **2024**, *582*, 740558. [CrossRef]
12. Wang, Z.; Zhang, X.; Su, Y.; Li, W.; Yin, X.; Li, Z.; Ying, Y.; Wang, J.; Wu, J.; Miao, F.; et al. Abnormal Behavior Monitoring Method of *Larimichthys crocea* in Recirculating Aquaculture System Based on Computer Vision. *Sensors* **2023**, *23*, 2835. [CrossRef] [PubMed]
13. Hou, H.; Zhang, Y.; Ma, Z.; Wang, X.; Su, P.; Wang, H.; Liu, Y. Life cycle assessment of tiger puffer (*Takifugu rubripes*) farming: A case study in Dalian, China. *Sci. Total Environ.* **2022**, *823*, 153522. [CrossRef] [PubMed]
14. Liao, Z.; Cui, X.; Luo, X.; Ma, Q.; Wei, Y.; Liang, M.; Xu, H. Exposure of farmed fish to petroleum hydrocarbon pollution and the recovery process: A simulation experiment with tiger puffer *Takifugu rubripes*. *Sci. Total Environ.* **2024**, *913*, 169743. [CrossRef] [PubMed]
15. Islam, S.I.; Ahammad, F.; Mohammed, H. Cutting-edge technologies for detecting and controlling fish diseases: Current status, outlook, and challenges. *J. World Aquac. Soc.* **2024**, *55*, 13051. [CrossRef]
16. Cheng, S.; Zhao, K.; Zhang, D. Abnormal water quality monitoring based on visual sensing of three-dimensional motion behavior of fish. *Symmetry* **2019**, *11*, 1179. [CrossRef]
17. Bao, Y.J.; Ji, C.Y.; Zhang, B.; Gu, J.L. Representation of freshwater aquaculture fish behavior in low dissolved oxygen condition based on 3D computer vision. *Mod. Phys. Lett. B* **2018**, *32*, 1840090. [CrossRef]
18. 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.* **2016**, *39*, 1137–1149. [CrossRef] [PubMed]
19. 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. [CrossRef]
20. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37. [CrossRef]
21. Krichen, M. Convolutional neural networks: A survey. *Computers* **2023**, *12*, 151. [CrossRef]
22. Xie, X.; Cheng, G.; Li, Q.; Miao, S.; Li, K.; Han, J. Fewer is more: Efficient object detection in large aerial images. *Sci. China Inf. Sci.* **2024**, *67*, 112106. [CrossRef]

23. Wang, C.; Yeh, I.; Liao, H.M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *arXiv* **2024**, arXiv:2402.13616.
24. Wang, C.; Liao, H.M.; Yeh, I. Designing network design strategies through gradient path analysis. *arXiv* **2022**, arXiv:2211.04800. [[CrossRef](#)]
25. Dai, X.; Chen, Y.; Xiao, B.; Chen, D.; Liu, M.; Yuan, L.; Zhang, L. Dynamic head: Unifying object detection heads with attentions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 7369–7378. [[CrossRef](#)]
26. Zhu, X.; Hu, H.; Lin, S.; Dai, J. Deformable convnets v2: More deformable, better results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9308–9316. [[CrossRef](#)]
27. Xiong, Y.; Li, Z.; Chen, Y.; Wang, F.; Zhu, X.; Luo, J.; Wang, W.; Lu, T.; Li, H.; Qiao, Y.; et al. Efficient Deformable ConvNets: Rethinking Dynamic and Sparse Operator for Vision Applications. *arXiv* **2024**, arXiv:2401.06197. [[CrossRef](#)]
28. Wang, C.; Bochkovskiy, A.; Liao, H.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.