Journal of Image Processing Theory and Applications (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/jipta.2026.090101
ISSN 2560-6239 Vol. 9 Num. 1

# *Underwater Sick Fish Detection Based on an Improved YOLOv11n Approach*

**Chong Zheng[1,a], Su Xu[1,b,*], Chuande Xu[1,c], Guangyu Du[1,d]**

*[1]School of Electronic Engineering, Jiangsu Ocean University, Lianyungang, Jiangsu, 222000, China*
*[a]1719275326@qq.com, [b]24354293@qq.com, [c]18662060690@163.com, [d]15751672846@163.com*
*[*]Corresponding author*

***Keywords:*** YOLOv11n; MobileNetV4; SlimNeck; Loss Function

***Abstract:*** To mitigate the issues of missed and false detections arising from scale variations, image blurring and target occlusion in underwater diseased fish detection, this study develops an enhanced YOLOv11n-based algorithm designated as YOLOv11-DFL. The backbone network incorporates the lightweight MobileNetV4, which combines depthwise separable convolution, lightweight attention modules, and multi-scale feature enhancement to balance feature extraction capability with significant reductions in parameters and computation. For the detection layer, the original P5 branch is removed and a P2 branch is added to enhance shallow detail capture, thereby reducing missed and false detections of small targets. The neck network integrates SlimNeck, which uses GSConv for channel splitting to cut redundant computation and VovGSCSP to strengthen multi-scale feature aggregation and gradient transfer, achieving lightweight design while boosting feature expression efficiency and small-target detection accuracy. Finally, the WIM loss function is adopted to improve model generalization and accelerate convergence. Experiments on the self-built sick fish dataset show that compared with the original YOLOv11n, the proposed method increases Precision, Recall, and mAP50 by 4.7%, 9.2%, and 7.0% respectively. Tests on underwater target datasets including RUOD, URPC2019, UDD, and DUO also demonstrate improved precision and recall, verifying the superior performance of YOLOv11-DFL in sick fish detection.

## 1. Introduction

In the global fisheries sector, submerged aquaculture holds a pivotal position as a highly efficient method of fish farming. With population growth and the rising demand for high-quality protein, aquaculture plays a vital role in meeting the increasing demand for aquatic products, contributing significantly to the stable growth of the global fisheries economy. However, behind the thriving development of submerged aquaculture, the issue of diseased fish has become a looming threat hanging over the heads of farmers, posing a serious challenge to the entire aquaculture industry. As farming scales expand and stocking densities increase, the prevalence of diseased fish has become increasingly pronounced. Traditional manual detection methods struggle to meet the demands, creating an urgent need for efficient and precise diseased fish detection technologies to safeguard the

healthy development of submerged aquaculture.

In early underwater aquaculture, detecting diseased fish primarily relied on manual experience, with farmers judging illness by visually inspecting fish appearance, behavior, and feeding patterns. While simple to implement, this method had significant limitations. With technological advancements, the rapid development of object detection algorithms in computer vision has revolutionized numerous industries. From identifying people and objects in security surveillance to detecting obstacles in autonomous driving, and extending to defect detection in industrial production, object detection algorithms have demonstrated formidable capabilities and vast application prospects.

Object detection technology is broadly classified into conventional methods and deep learning-based detection strategies. Traditional methods require a series of steps including image preprocessing, sliding windows, manual feature extraction and selection, classification, and post-processing. However, their reliance on manually designed features limits accuracy and generalization capabilities, while computational efficiency struggles to meet modern application demands. The successful application of AlexNet in 2012 accelerated the development of deep learning-based object detection, forming two mainstream approaches: Two-stage and One-stage methods. The Two-stage approach is represented by the R-CNN[1] series, such as Fast-RCNN[2]. These methods first locate candidate boxes for image features, then classify and recognize the regions within these boxes. While offering high detection accuracy, they suffer from significant computational overhead and poor real-time performance, failing to meet the requirements for certain underwater object detection tasks. One-stage methods, represented by the YOLO[3] series and SSD[4], directly produce object bounding boxes and categories from input to output. They achieve high detection accuracy, fast speed, and low computational complexity, In recent years, they have gained extensive research and application in underwater target detection. The YOLO series, after multiple iterations, has integrated lightweight structures and attention mechanisms, further balancing speed and accuracy. Related optimization methods continue to emerge. However, these methods lack a refined region proposal mechanism. In scenarios with dense multi-objects, occlusions, or complex backgrounds, they are prone to false negatives and false positives, limiting their applicability in complex environments.

In the field of underwater target detection, the YOLO series of algorithms has also gradually found application. Due to the unique characteristics of the underwater environment—such as light attenuation, image blurring, and complex backgrounds—underwater target detection faces numerous challenges. To address these, researchers worldwide have continuously pursued innovation and explored various solutions. Ji et al.[5], for instance, devised the Feature Boosting and Differential Pyramid Network (FBDPN), which enhances contextual relationships between features at different scales and preserves effective information. However, this approach suffers from high computational and parameter costs. Dai et al.[6] introduced the Gate-Controlled Cross-Domain Collaboration Network (GCC-Net), which enhances image visibility in low-contrast regions through real-time image enhancement and cross-domain feature interaction/fusion modules. However, excessive enhancement may cause loss of critical feature details, and false detections still occur in occluded target areas. Ma et al.[7] introduced a light noise model, embedding the Minimum Weighted Entropy Error (MWMEE) criterion into the YOLO network's loss function to optimize parameter training. They also employed a multi-error processing strategy to handle vector errors during information backpropagation, accelerating network convergence. Experiments validated the network's excellent detection performance. Zhou et al.[8] proposed the Underwater Object Detection Network (UODN), designing a Cross-Stage Multi-Branch Module (CSMB) and Large-Scale Kernel Spatial Pyramid Module (LSKP) and combining them as the network backbone to enhance feature extraction capabilities across various underwater object scales. Qu et al.[9] designed a lightweight and efficient partial convolution (LEPC) module that significantly reduces redundant computations while improving efficiency. They also integrated deep separable convolutions with varying expansion rates

into FasterNet, enhancing the model's ability to capture detailed features of small objects. Although these studies have made progress in improving small object detection algorithms, their effectiveness still needs to be enhanced. This paper aims to propose an improved YOLOv11-DFL object detection algorithm to achieve further enhancements in diseased fish detection performance.

## 2. YOLOv11 Overview and Its Improved Models

Since its debut in 2016, the YOLO series has revolutionized object detection with its "end-to-end real-time detection" approach, breaking through the efficiency bottleneck of traditional two-stage detection. Subsequent iterations—including YOLOv3's multi-scale detection, YOLOv5's engineering optimizations, and YOLOv8's modular architecture—have continuously tackled the challenge of balancing accuracy and speed. YOLOv11, developed by the Ultralytics team as an upgrade to YOLOv8, addresses shortcomings such as small object detection failures and insufficient robustness in complex scenarios. It preserves the series' core strengths while further enhancing performance.

YOLOv11 comprises five variants: YOLOv11n, YOLOv11s, YOLOv11m, YOLOv11l, and YOLOv11x. These models progressively increase in depth and width, with all versions adopting new architectures such as C3K2 blocks, SPPF, and C2PSA. They support multi-task capabilities including object detection, instance segmentation, image classification, and pose estimation. The network architecture diagram of YOLOv11 is shown in Figure 1.

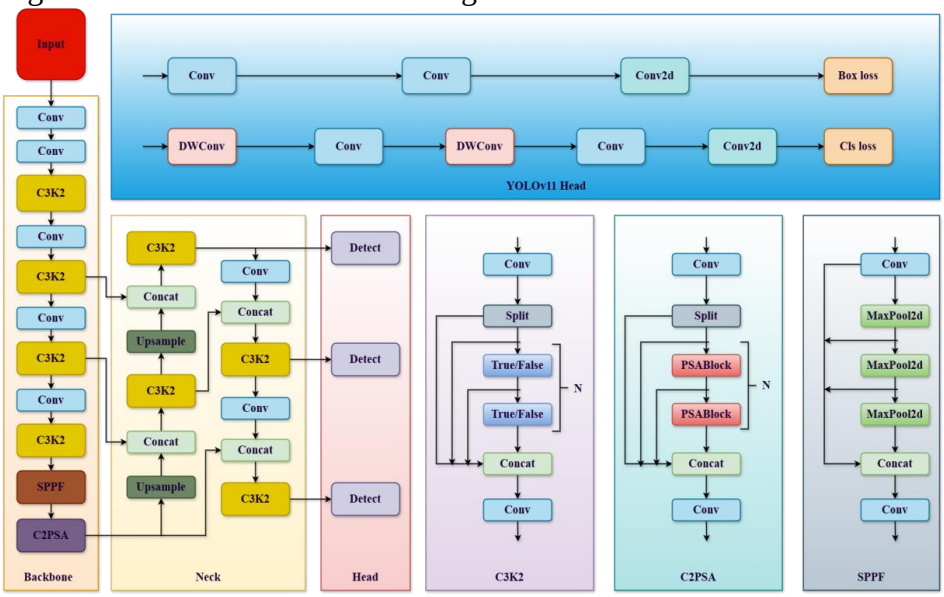


Figure 1: YOLOv11 Network Architecture Diagram.

YOLOv11 introduces targeted optimizations in trunk and neck network design: upgrading the original C2f module to C3K2. This module employs dual convolution kernels and a feature map segmentation mechanism in the bottleneck layer, achieving both refined feature representation and significantly enhanced feature extraction efficiency. Additionally, a new C2PSA module is introduced after the Spatial Pyramid Pooling (SPPF) module. This module, extended from C2f, integrates PSA attention mechanisms. By leveraging multi-head attention and feedforward networks (FFN), it enhances feature extraction capabilities—not only optimizing gradient flow and training efficiency through selective addition of residual blocks, but also mapping features to high-dimensional spaces via FFN to capture complex nonlinear relationships between features, further improving the model's efficiency in capturing effective features. Additionally, embedding two Depth-Wise Convolutions (DWConv) within the classification head effectively reduces computational redundancy, significantly improving overall computational efficiency.

The underwater diseased fish detection network proposed in this paper is based on the YOLOv11n model with enhancements, named YOLOv11-DFL. First, the lightweight neural network MobileNetV4 is introduced in the backbone section to reduce the number of parameters and computational load. The UIB module, as the core component of MobileNetV4, employs an inverted bottleneck structure combining "deep convolutions + pointwise convolutions." This approach preserves feature expression capabilities while reducing computational demands and supports dynamic channel adjustment for task adaptation. The ConvBN module combines convolution and batch normalization, handling fundamental feature extraction and data distribution standardization. This accelerates training convergence while stabilizing feature propagation and reducing overfitting risks. Next, the detection heads were replaced from P3, P4, P5 to P2, P3, P4 to align with lightweight positioning and accurately detect features of small-to-medium-sized objects. Subsequently, the SlimNeck module was introduced at the Neck layer, replacing C3K2 with VoVGSCSP and Conv with GSConv modules. This enhances feature propagation efficiency, prevents information loss, and balances speed with accuracy. Finally, the WIM loss function was adopted to replace YOLOv11's original loss function, optimizing training performance. The YOLOv11-DFL network architecture is illustrated in Figure 2.

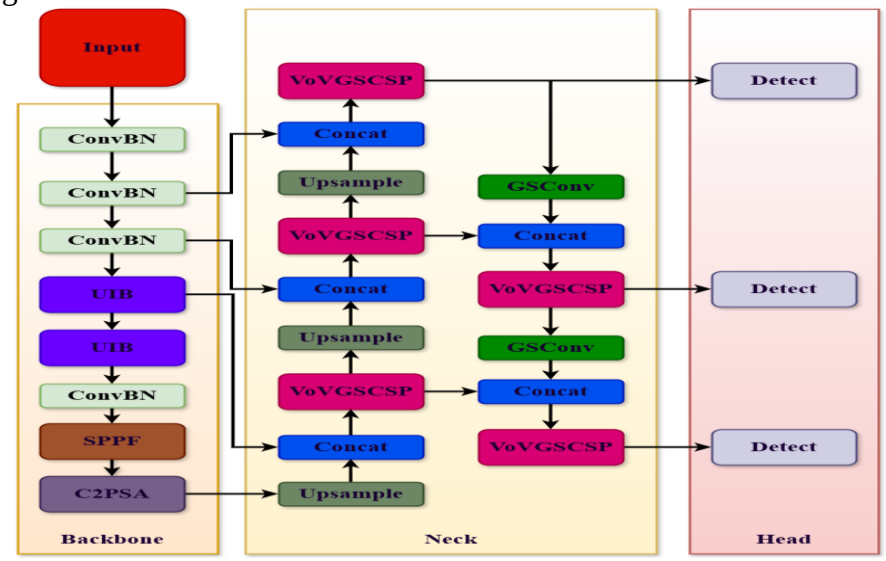


Figure 2: YOLOv11-DFL Network Architecture Diagram.

## 3. YOLOv11-DFL

### 3.1. MobileNetV4: A Lightweight Neural Network

MobileNetV4 is a neural network architecture designed for mobile devices, with the core objective of providing a versatile and efficient model solution for the mobile ecosystem. This architecture achieves lightweight model design by integrating the Universal Inverted Bottleneck (UIB) with the MobileMQA attention module[10]. The UIB module, serving as a unified and flexible structure, builds upon the inverted bottleneck (IB) block from MobileNetV2. It integrates variants including IB, ConvNext, Feedforward Network (FFN), and Extra Depth-wise (ExtraDW), making the architecture highly adaptable to precisely match diverse task requirements. The MobileMQA mechanism prioritizes simplicity over performance optimization, specifically tailored for mobile accelerators to deliver significant speed gains. This enables MobileNetV4 to substantially reduce parameter size and computational cost while maintaining excellent performance. Furthermore, through an innovative NAS approach, the model combines coarse-grained and fine-grained search. This dual optimization

significantly enhances network construction efficiency while achieving deep structural refinement, enabling the model to deliver optimal performance across diverse computing devices. The UIBblocks module structure is illustrated in Figure 3.

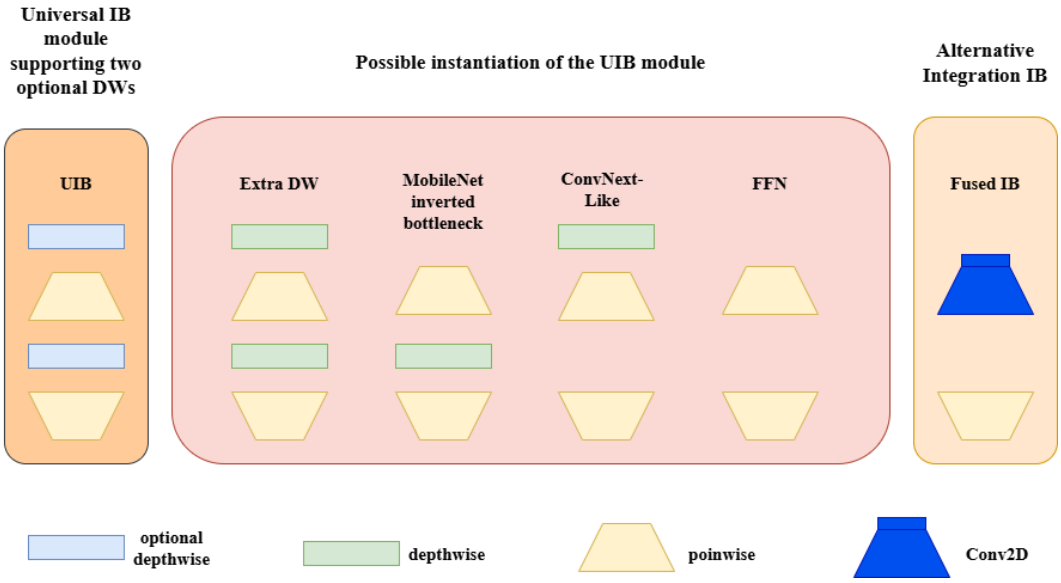


Figure 3: UIBblocks Diagram.

By adopting MobileNetV4 as the backbone network for MLL-YOLOv11n, this paper effectively reduces the model's parameter count and computational load, optimizing its lightweight architecture. This enhances the algorithm's deployment adaptability on mobile and embedded platforms.

### 3.2. Improved Detection Layer

Underwater inspection scenarios suffer from insufficient lighting and blurred images. Traditional detection heads based on YOLO series models not only struggle to accurately identify and locate aquatic organisms but also suffer from the drawback of excessive parameter complexity.

When processing 640×640 pixel input images, YOLOv11n employs three detection branches at resolutions of 20×20, 40×40, and 80×80, corresponding to the prediction tasks for large, medium, and small objects respectively. The 80×80 branch responsible for small object detection requires an 8x downsampling and feature fusion through the backbone network. This series of convolutional operations tends to cause significant loss of shallow-layer positional information. Furthermore, small objects inherently occupy a smaller pixel area in images, and their features become further weakened after multiple convolutions, ultimately increasing the risk of missed detections.

To address this issue, the proposed solution introduces a new 160×160 small object detection layer into the original architecture while removing the 20×20 large object detection layer. This adjustment broadens the model's scale coverage while effectively preserving precise spatial information from deep layers by fusing 4× downsampled backbone feature maps with other layer features. Simultaneously, eliminating detection layers with low sensitivity to small objects reduces both model parameters and computational complexity. This feature fusion approach efficiently combines low-resolution, high-semantic features with high-resolution, low-semantic features, significantly improving small object detection performance and reducing false negatives. Figure 4 illustrates the detection architecture comparison between YOLOv11-DFL and YOLOv11n.

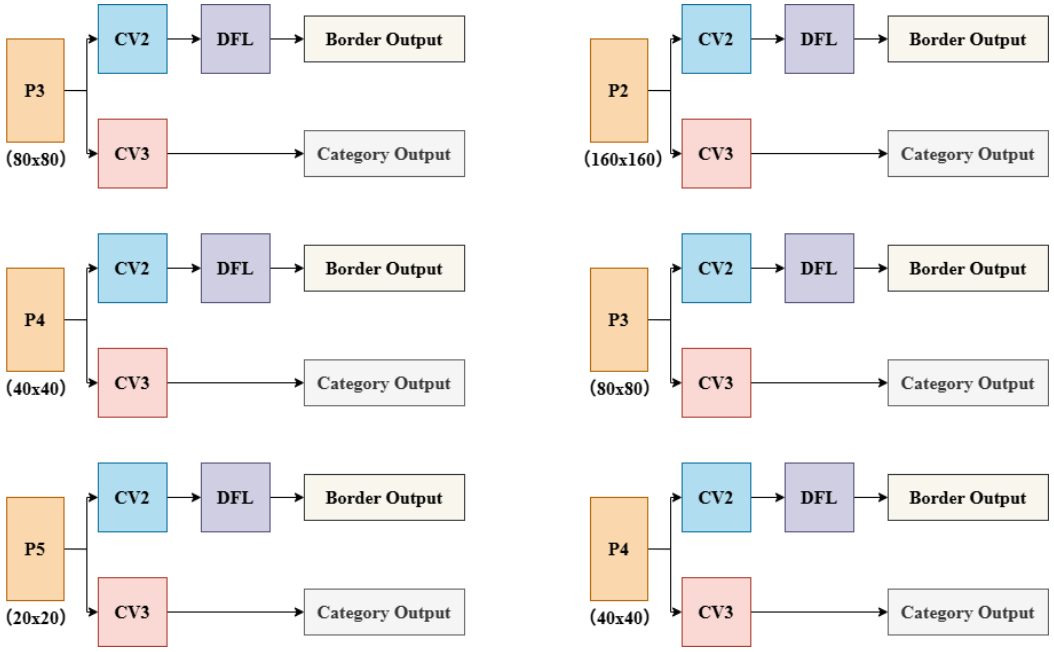


Figure 4: Detection architecture diagrams for YOLOv11-DFL (left) and YOLOv11n (right).

### 3.3. SlimNeck Feature Fusion Network

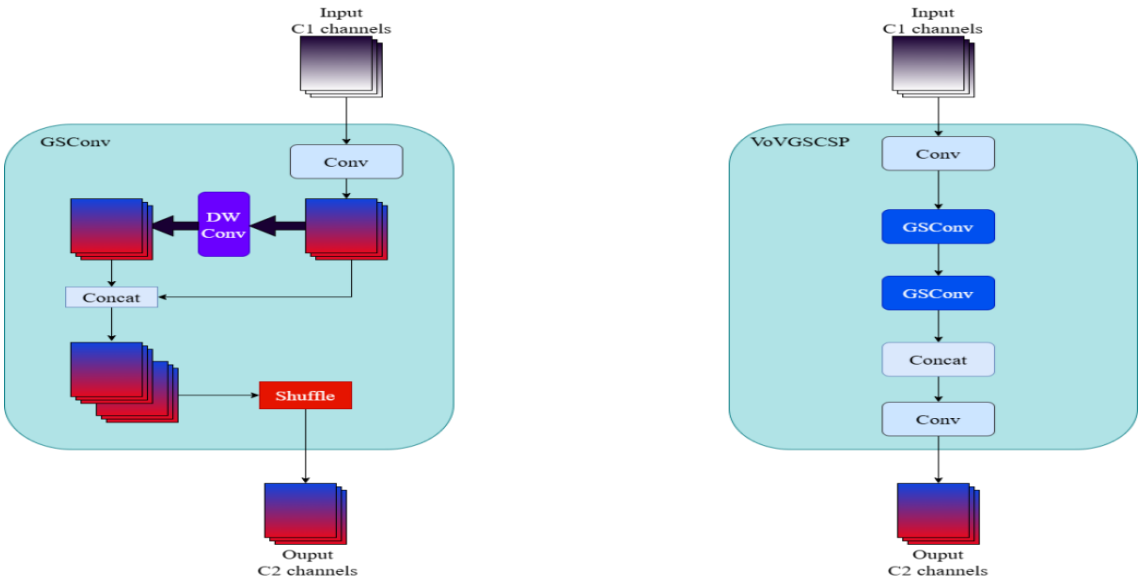


Figure 5: GSConv Module Structure Diagram (left) and VoVGSCSP Module Structure Diagram (right).

SlimNeck is a design paradigm aimed at achieving higher cost-effectiveness for detectors. Its network architecture integrates universal methods that enhance CNN learning efficiency, such as DensNet, VoVNet, and CSPNet. Its core approach involves constructing cross-level local network correlation modules through a single-stage aggregation strategy, thereby balancing accuracy and speed in object detection algorithms. This simultaneously addresses the accuracy degradation issue caused by lightweight models using Deep Separable Convolution (DSC) while reducing computational load. The structure's outstanding performance relies on two key modules: GSConv and

VoVGSCSP. GSConv first performs standard convolution (SC) downsampling on the input, then executes deep separable convolution (DSC), concatenates the results, and finally unifies the channel count via a shuffle operation. This approach makes the DSC output closely resemble SC results while preserving lightweight advantages and avoiding DSC's inherent drawbacks. The VoVGSCSP module optimizes the grid structure, leveraging cross-local network properties to effectively reduce computational load and network complexity. Given that integrating Slim-Neck into the YOLO series of lightweight networks achieves higher detection accuracy and speed with reduced latency and parameters, this paper replaces the C3k2 layer in the YOLOv11n neck network with the VoVGSCSP module and substitutes the Conv layer in the neck network with GSConv. The structural diagrams of the GSConv and VoVGSCSP modules are shown in Figure 5.

### 3.4. WIM Loss Function

The design and optimization of loss functions are central to enhancing the accuracy of detection models—scientifically selecting and combining loss functions enables efficient training of object detection models and strengthens their performance in complex scenarios. YOLOv11n defaults to CIoU Loss, which comprehensively measures the match between predicted and ground-truth bounding boxes by considering overlap area, center point distance, and aspect ratio, thereby effectively improving object localization capabilities. However, CIoU suffers from insufficient generalization and suboptimal performance in underwater object detection (e.g., fish school detection). To address this, this paper proposes the WIM composite loss function, which integrates WIoU-v3[11], Inner-IoU[12], and MPD-IoU[13]. This approach optimizes bounding box regression performance through multi-level geometric constraints.

WIoU-v3 evaluates anchor quality using a dynamic non-monotonic focusing mechanism: first defining outlier $\beta$, $\beta$ represents the ratio of $L_{IOU}^{*}$ to $\bar{L}_{IOU}$, where $L_{IOU}^{*}$ denotes the loss function between anchor boxes and target boxes, and $\bar{L}_{IOU}$ is the exponential moving average of $L_{IOU}^{*}$. Based on $\beta$, a non-monotonic focus factor $r$ is constructed with hyperparameters $\alpha$ and $\delta$. This factor dynamically adjusts loss weights during model training, enabling adaptive prioritization of samples with varying quality. The specific formula is as follows:

$$\begin{cases} L_{WIoU\text{-}v3}=r\cdot L_{WIoU\text{-}v1} \\ r=\frac{\beta}{\delta\alpha^{\beta-\delta}};\beta=\frac{L_{IoU}^{*}}{\bar{L}_{IoU}}\in[0,+\infty) \end{cases} \tag{1}$$

Inner-IoU optimizes IoU loss calculation by introducing auxiliary bounding boxes. Its core mechanism involves scaling down the dimensions of both ground truth and predicted boxes by a proportional factor s, then computing the intersection-union ratio based on these scaled boxes to enhance bounding box regression accuracy. Where $(x_c^{gt}, y_c^{gt})$ and $(x_c, y_c)$ represent the center coordinates of the ground truth and predicted boxes, respectively; $w^{gt}$ and $h^{gt}$, along with w and h, denote their widths and heights, respectively; inter denotes the overlapping area; and union denotes the total covered area. The final intersection-union ratio calculation is as follows:

$$b_l^{gt}=x_c^{gt}-\frac{w^{gt}\cdot s}{2},b_r^{gt}=x_c^{gt}+\frac{w^{gt}\cdot s}{2} \tag{2}$$

$$b_t^{gt}=y_c^{gt}-\frac{h^{gt}\cdot s}{2},b_b^{gt}=y_c^{gt}+\frac{h^{gt}\cdot s}{2} \tag{3}$$

$$b_l=x_c-\frac{w\cdot s}{2},b_r=x_c+\frac{w\cdot s}{2} \tag{4}$$

$$b_t=y_c-\frac{h\cdot s}{2},b_b=y_c+\frac{h\cdot s}{2} \tag{5}$$

$$\text{inter}=\max\left(\min\left(b_r^{gt},b_r\right)-\max\left(b_l^{gt},b_l\right)\right)\cdot \max\left(\min\left(b_b^{gt},b_b\right)-\max\left(b_l^{gt},b_t\right)\right) \tag{6}$$

$$\text{union}=\left(w^{gt}\cdot h^{gt}\right)\cdot s^2+(w\cdot h)\cdot s^2-\text{inte} \tag{7}$$

$$IoU_{inner}=\frac{\text{inter}}{\text{union}} \tag{8}$$

$$L_{Inner\text{-}IoU}=1-IoU_{inner} \tag{9}$$

MPD-IoU enhances regression accuracy and efficiency by minimizing the distance between the top-left and bottom-right corner vertices of the predicted box and the ground truth box. It simultaneously accounts for overlapping or non-overlapping areas, center point distance, and width-height deviation. Its streamlined computation process improves training effectiveness and convergence speed. Here, w and h denote the width and height of the ground truth bounding box, while$(x_1^{pred},y_1^{pred})$ and $(x_1^{gt},y_1^{gt})$ represent the coordinates of the top-left corner points of the predicted and ground truth boxes, respectively. $(x_2^{pred},y_2^{pred})$ and $(x_2^{gt},y_2^{gt})$ denote the coordinates of the bottom-right corner points of the two boxes. The specific formula is as follows:

$$L_{MPD\text{-}IoU}=1-MPD_{IOU} \tag{10}$$

$$MPD_{IOU}=\frac{\text{inter}}{\text{union}}+\frac{d_1^2}{w^2+h^2}+\frac{d_2^2}{w^2+h^2} \tag{11}$$

$$d_1^2=(x_1^{pred}-x_1^{gt})^2+(y_1^{pred}-y_1^{gt})^2 \tag{12}$$

$$d_2^2=(x_2^{pred}-x_2^{gt})^2+(y_2^{pred}-y_2^{gt})^2 \tag{13}$$

## 4. Experimental Design and Results Analysis

### 4.1. Dataset

This study's dataset was independently collected and constructed, centered on authentic underwater fish scenarios to ensure data authenticity and scene adaptability. The data collection platform was established as follows: a small-scale controlled tank experimental environment was built, featuring a custom-made fish tank measuring 3m × 1m × 1m and a custom-made underwater camera. Multiple common fish species were selected as research subjects, including carp, crucian carp, goldfish, koi, tilapia, black carp, grass carp, and silver carp, ensuring diverse coverage within the dataset. The filtered dataset comprises 824 validation images and 3,297 training images, totaling 4,121 images captured across various scenarios.

### 4.2. Experimental Environment

The experiment was conducted on a Windows 11 system with an Intel Core i7-13700 CPU and an NVIDIA GeForce RTX 4080 16GB GPU. PyTorch version 2.3.0, Python version 3.11.14, and CUDA version 12.1 were used. Specific experimental parameter settings are detailed in Table 1.

Table 1: Experimental Environment Parameters.

| Parameters | Settings |
|---|---|
| Epochs | 200 |
| Imgsz | 640 |
| Batch | 16 |
| Workers | 8 |

| | |
|---|---|
| Device | 0 |
| Optimizer | SGD |
| Close_mosaic | 10 |
| Resume | False |
| Project | Runs/Train |
| Name | Exp |
| Single_cls | False |
| Cache | False |

All datasets and models mentioned in the experiment were trained as specified in the table above. The training was set to 200 epochs, with input image dimensions of 640 ×640. The configuration used 8 threads and 16 batches, with SGD as the optimizer.

### 4.3. Evaluation Indicators

The experiment evaluates the performance of the proposed algorithm using metrics including precision (P), recall (R), mean average precision (mAP), parameters, computational complexity (GFLOPs), and model size. The calculation formulas are as follows:

$$P=\frac{TP}{TP+FP} \tag{14}$$

$$R=\frac{TP}{TP+FN} \tag{15}$$

$$AP=\int_0^1 P(R)Dr \tag{16}$$

$$mAP=\frac{1}{N}\sum_{i=1}^{N} AP(i) \tag{17}$$

TP denotes the count of true positive samples correctly identified by the model; FP stands for the count of true negative samples falsely classified as positive by the model; FN represents the count of true positive samples misjudged as negative by the model. N refers to the total number of categories within the dataset. AP is the average precision derived from the PR curve, and mAP denotes the mean value of average precision across all categories in the dataset. This evaluation index offers a holistic assessment of the model's detection performance.

### 4.4. Melting Experiment

YOLOv11-DFL improves upon YOLOv11n by modifying its backbone network, neck network, and detection layer. To evaluate the extent of algorithm performance optimization under different implementation sequences, a series of ablation experiments were designed. First, the lightweight neural network MobileNetV4 was incorporated into the backbone network, followed by the introduction of the Slim-Neck module into the neck network. The detection layer was further modified by removing the P5 detection head and adding a P2 detection head, culminating in the use of the WIM loss function. Model detection performance improvements were evaluated under identical experimental conditions, with results presented in Table 2. Here, Head-2 denotes the aforementioned detection layer modification; A, B, C, and D represent different improvement methods applied to the original YOLOv11n model; and √ indicates the addition of that specific model.

Table 2: Ablation Experiment Results.

| Method | MobileNetV4 | Head-2 | Slim-Neck | WIM | P/% | R/% | mAP50% | Params/M | FLOPS/G |
|---|---|---|---|---|---|---|---|---|---|
| YOLOv11n | | | | | 65.9 | 63.3 | 64.6 | 2.58 | 6.3 |
| A | √ | | | | 66.1 | 60.2 | 65.7 | 2.1 | 5.2 |
| B | √ | √ | | | 65.2 | 65.8 | 69.5 | 1.2 | 6.1 |
| C | √ | √ | √ | | 70.1 | 69.9 | 70.3 | 1.57 | 6.1 |
| D | √ | √ | √ | √ | 70.6 | 72.5 | 71.6 | 1.62 | 6.0 |

As shown in Table 2, first, by refining the feature network and optimizing the detection layer through approaches A and B, the model's parameter count was significantly reduced. Second, by improving the Conv structure and loss function via approaches C and D, detection accuracy and other performance metrics were markedly enhanced while maintaining substantially fewer parameters compared to the original model. Given the complex nature of underwater diseased fish detection environments, the improved model achieves high detection accuracy with reduced computational load and smaller memory footprint while maintaining real-time performance. This provides an efficient and feasible deployment solution for diseased fish detection.

Compared to the original YOLOv11n algorithm, the proposed YOLOv11-DFL object detection algorithm demonstrates superior performance in diseased fish identification tasks. Specifically, precision (P) improved by 4.7%, recall (R) increased by 9.2%, and mean average precision (mAP) rose by 7%. Meanwhile, the model parameter count decreased by 0.96 million, and computational load reduced by 0.3 gigaflops. The above data fully demonstrates the effectiveness of the implemented improvement strategy.

## 4.5. Comparative trial

To demonstrate the superior performance of the YOLOv11-DFL diseased fish detection algorithm proposed in this study, a series of comparative experiments were designed. These experiments contrasted YOLOv11-DFL with leading object detection algorithms in the field. These algorithms include YOLOv3-tiny, YOLOv5n, YOLOv6, YOLOv8n, YOLOv9s, YOLOv10n, and YOLOv13. All algorithms were tested on a custom diseased fish dataset, with experimental results presented in Table 3.

Table 3: Comparative Experimental Results.

| Model | P/% | R/% | mAP50% | mAP50-95% | Params/M | FLOPS/G |
|---|---|---|---|---|---|---|
| YOLOv3-tiny | 65.4 | 70.9 | 67.2 | 26.1 | 12.13 | 18.9 |
| YOLOv5n | 55.5 | 59.6 | 55.7 | 21.2 | 2.50 | 7.1 |
| YOLOv6 | 60.0 | 63.7 | 58.9 | 22.8 | 4.23 | 11.7 |
| YOLOv8n | 59.7 | 67.0 | 57.9 | 22.3 | 3.0 | 8.1 |
| YOLOv9s | 58.1 | 63.9 | 59.4 | 22.0 | 7.16 | 26.7 |
| Yolov10n | 43.0 | 62.6 | 48.6 | 19.4 | 2.69 | 8.2 |
| Yolov11n | 65.9 | 63.3 | 64.6 | 24.7 | 2.58 | 6.3 |
| Yolov13 | 54.4 | 57.8 | 56.0 | 21.2 | 2.44 | 6.2 |
| YOLOv11-DFL | 70.6 | 72.5 | 71.6 | 28.0 | 1.62 | 6.0 |

As shown in Table 3, our improved model—YOLOv11-DFL—outperforms other models across all metrics, including accuracy, recall, and mean average precision (mAP). Specifically, YOLOv11-DFL achieves 70.6% precision (P), 72.5% recall (R), and 71.6% mAP50, significantly surpassing other models. Moreover, its parameter count and computational complexity were reduced to 1.62 million and 6.0 GFLOPS, respectively, both lower than other models. Therefore, overall, YOLOv11-DFL achieves a favorable balance between accuracy, speed, and model size, demonstrating certain advantages in underwater diseased fish detection scenarios.

## 4.6. Generalization Experiment

To validate the effectiveness and generalization of the YOLOv11-UDLite algorithm for detecting diseased fish, comparative experiments were conducted on the RUOD, URPC2019, UDD, and DUO underwater object datasets. The experimental results are shown in Table 4. RUOD-YOLOv11n denotes results obtained using the original YOLOv11n model, while RUOD-Improved represents results from the modified YOLOv11-DFL model. The same applies to the other three groups.

Table 4: Generalization Experiment Results.

| Model | P/% | R/% | mAP50% | Params/M | FLOPS/G |
|---|---|---|---|---|---|
| RUOD-YOLOv11n | 78.4 | 65.4 | 73.7 | 2.58 | 9.4 |
| RUOD-Improvement | 85.3 | 76.6 | 84.1 | 1.62 | 6.1 |
| URPC2019- YOLOv11n | 69.7 | 59.1 | 63.8 | 2.58 | 9.4 |
| URPC2019-Improvemen | 71.9 | 66.7 | 72.3 | 1.62 | 6.3 |
| UDD- YOLOv11n | 61.0 | 51.4 | 54.7 | 2.58 | 9.4 |
| UDD-Improvemen | 70.9 | 62.6 | 68.1 | 1.62 | 6.3 |
| DUO- YOLOv11n | 79.5 | 69.0 | 77.6 | 2.58 | 9.4 |
| DUO-Improvemen | 83.5 | 74.8 | 82.9 | 1.62 | 6.3 |

As presented in Table 4, YOLOv11-DFL cuts down on model parameters and computational overhead while achieving comprehensive gains in Precision, Recall and mAP across four typical underwater target datasets, namely RUOD, URPC2019, UDD and DUO. This fully proves the remarkable validity of the improved strategies proposed in this study, thus confirming the advantages of the modified YOLOv11-DFL algorithm.

## 5. Visualization

To more intuitively assess the prediction accuracy of the improved model, four underwater images of diseased fish were selected. Predictions were generated using both the original YOLOv11n model and the improved YOLOv11-DFL model, with results shown in the figure 6. The results clearly demonstrate that the improved YOLOv11-DFL model exhibits excellent performance in detecting diseased fish underwater.

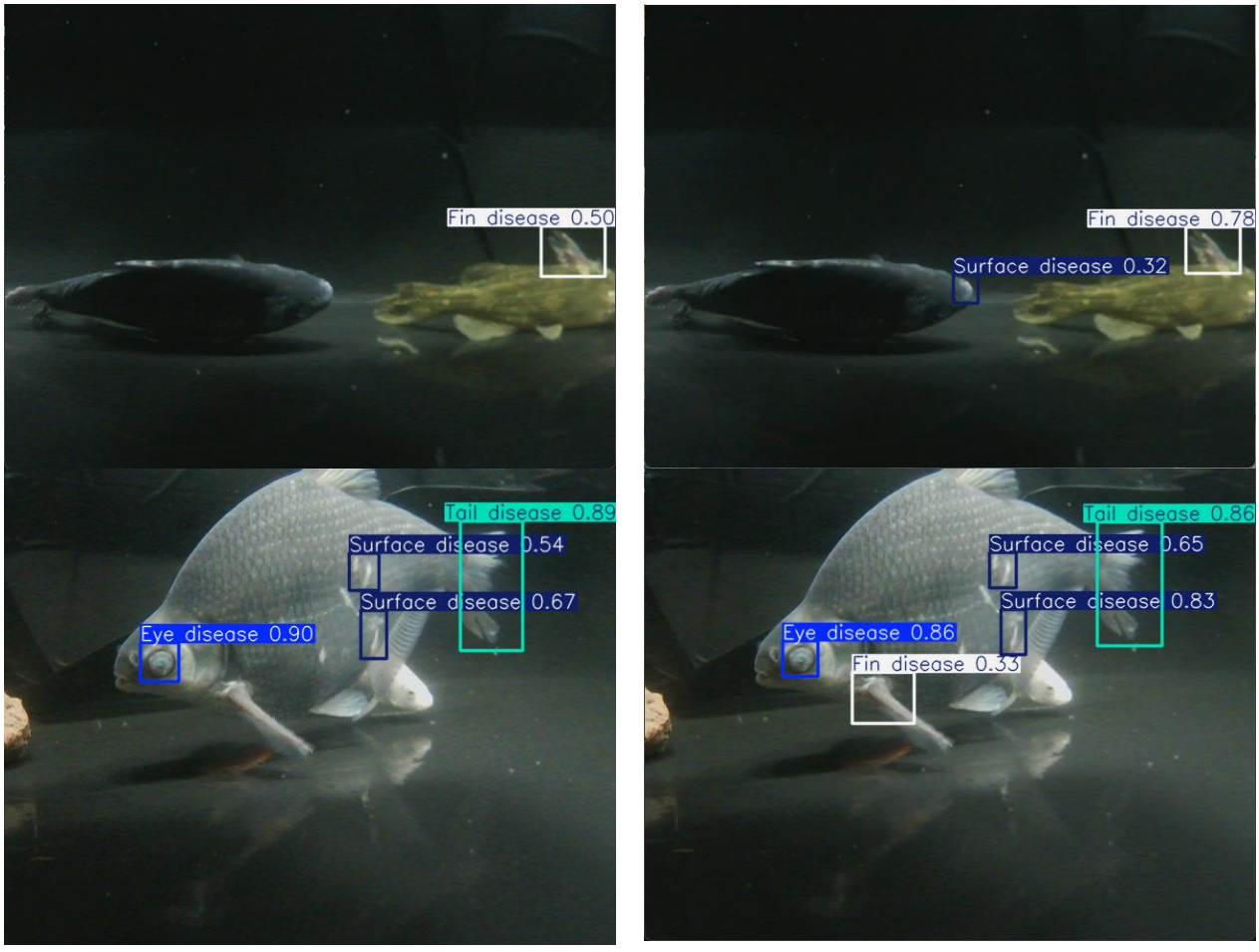


Figure 6: YOLOv11n model prediction (left) and YOLOv11-DFL model prediction (right).

## 6. Conclusion

In underwater diseased fish detection, the complex detection environment and significant interference often lead to reduced detection accuracy, with missed detections and false positives frequently occurring. Therefore, this paper proposes an improved diseased fish detection model based on YOLOv11n—YOLOv11-DF. This algorithm first incorporates the lightweight neural network MobileNetV4 into the backbone network, significantly reducing parameter count and computational load while maintaining stable feature extraction capabilities. Next, it removes the P5 detection head and adds the P2 detection head in the detection layer, greatly enhancing detection performance for small targets. Subsequently, the SlimNeck module is introduced into the neck network architecture, achieving a lightweight neck network while improving feature expression efficiency and detection accuracy for small targets. Finally, the WIM loss function is employed to balance training outcomes across diverse scenarios, enhancing model generalization and accelerating convergence. Experimental results demonstrate the proposed detection model's robust performance. Future work will focus on further optimizing the algorithm to strengthen model effectiveness and deploy it onto navigation devices for diseased fish detection.

## References

*[1] Girshick R, Donahua J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.*

*[2] Ren Sh Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with Region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.*
*[3] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. IEEE conference on computer vision and pattern recognition. 2016: 779-788.*
*[4] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]. European conference on computer vision. Cham: Springer, 2016: 21-37.*
*[5] Ji X, Chen Sh J, Hao L Y, et al. FBDPN: CNN-Transformer hybrid feature boosting and differential pyramid network for underwater object detection[J]. Expert Systems with Applications, 2024, 256: 124978.*
*[6] Dai L H, Liu H, Song P H, et al. A gated cross-domain collaborative network for underwater object detection[J]. Pattern Recognition, 2024, 149: 110222.*
*[7] Ma Hp, Zhang Y J, Sun Sh Y, et al. Weighted multi-error information entropy based you only look once network for underwater object detection[J]. Engineering Applications of Artificial Intelligence, 2024, 130: 107766.*
*[8] Zhou H, Kong M W, Yuan H X, et al. Real-time underwater object detection technology for complex underwater environments based on deep learning[J]. Ecological Informatics, 2024, 82: 102680.*
*[9] Qu Sh M, Cui C, Duan J L,et al. Underwater small target detection under YOLOv8-LA model[J]. Scientific Reports, 2024,14(1):16108.*
*[10] Qin D, Leichenr C, Delakirs M, et al. MobileNetV4-universal models for the mobile ecosystem [C]. Proceedings of the European Conference on Computer Vision, F, 2024.*
*[11] Tong Z J, Chen Y H, Xu Z W, et al. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism[J]. ArXiv preprint arXiv: 2301. 10051, 2023.*
*[12] Zhang H, Xu C, Zhang Sh J. Inner-IoU: More effective intersection over union loss with auxiliary bounding box[J]. ArXiv preprint arXiv: 2311. 02877, 2023.*
*[13] Ma S L, Xu Y. MPDIoU: A loss for efficient and accurate bounding box regression[J]. ArXiv preprint arXiv: 2307. 07662, 2023.*