



YOLO-FD: An accurate fish disease detection method based on multi-task learning



Xuefei Li^{a,b,c,d,1}, Shili Zhao^{a,b,c,d,2},
Chunlin Chen^{a,b,c,d,3}, Hongwu Cui^{e,4}, Daoliang Li^{a,b,c,d,5},
Ran Zhao^{a,b,c,d,*6}

^a National Innovation Center for Digital Fishery, China Agricultural University, Beijing 100083, China

^b Key Laboratory of Smart Farming Technologies for Aquatic Animal and Livestock, Ministry of Agriculture and Rural Affairs, China Agricultural University, Beijing 100083, China

^c Beijing Engineering and Technology Research Center for Internet of Things in Agriculture, China Agricultural University, Beijing 100083, China

^d College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

^e Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture and Rural Affairs, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China

ARTICLE INFO

Keywords:

Fish disease detection
Object detection
Semantic segmentation
Multi-task learning

ABSTRACT

Fish diseases often exhibit high risks of contagion, resulting in substantial economic losses. Accurate assessment of fish disease severity during diagnosis using deep learning poses a considerable challenge. Currently, deep learning models mainly focus on single tasks in fish disease detection, such as classification, object detection and segmentation. However, the accurate assessment of fish disease severity requires the integration of multiple dimensions of information, which is beyond the capabilities of traditional single-task methods. Therefore, this paper proposes YOLO-FD, a multi-task learning network specifically designed for simultaneous detection and segmentation. YOLO-FD extends the YOLOv8 backbone by integrating a novel semantic segmentation branch dedicated to precisely segmenting infected areas in diseased fish, while retaining the original object detection branch for identifying infected fish. Weight uncertainty and PCGrad are employed to balance the weights of different losses and to optimize conflicting gradients during the training process. With only a negligible increase in network parameters, YOLO-FD, tested on our constructed Nocardiosis fish dataset, achieves a detection accuracy of 94.2% mAP50 and gets mIOU of 79.4%, showcasing a 0.5% improvement over the baseline YOLOv8 and surpassing the state-of-the-art semantic segmentation network Deeplabv3plus by 4%. Notably, compared to the adapted multi-task network YOLOP, YOLO-FD demonstrates substantial improvements, displaying a 13.7% increase in mAP50-95 and a 15.1% boost in mIOU. On the VOC2012 segmentation dataset, the proposed method exhibits a 3.2% increase in mAP50 and a 2.2% rise in mAP50-95 compared to the baseline. Furthermore, results of the ablation experiment validate the effectiveness and generalization of the proposed multi-task learning approach. Source code is available at <https://github.com/fEIFE-Lee/YOLO-FD>.

1. Introduction

Fish plays a pivotal role in aquaculture, serving as a significant

protein source in the human diet. It boasts a wealth of phosphorus, calcium, beneficial fats, high-quality protein, and an array of essential nutrients crucial for human health (Lyubchenko et al., 2016).

* Corresponding author at: P. O. Box 121, China Agricultural University, 17 Tsinghua East Road, Beijing 100083, China.

E-mail addresses: cuihw@ysfri.ac.cn (C. Chen), dliangl@cau.edu.cn (D. Li), ran.zhao@cau.edu.cn (R. Zhao).

¹ ORCID: 0009-0001-0692-8702

² ORCID: 0000-0001-8462-5750

³ ORCID: 0009-0009-3369-8138

⁴ ORCID: 0000-0003-3356-1712

⁵ ORCID: 0000-0002-1842-419X

⁶ ORCID: 0000-0002-6945-8226

<https://doi.org/10.1016/j.eswa.2024.125085>

Received 19 December 2023; Received in revised form 21 June 2024; Accepted 11 August 2024

Available online 16 August 2024

0957-4174/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Nowadays, fish farming is gradually transforming into a new factory-based and intelligent intensive farming model (Manan et al., 2019). Although this transformation significantly enhances fish output efficiency, it introduces new complexities. Specifically, in high-density farming environment, fish diseases can swiftly propagate, resulting in collective infections and irreparable losses. Hence, timely identification and diagnosis of diseased fish to prevent widespread fish mortalities within the factory high-density farming holds immense practical significance (Yu et al., 2021).

Visible symptoms will manifest on the fish's body surface, accompanied by distinct abnormal behaviors after infection. The conventional approach for detecting fish infection primarily involves examining lesions on the skin, gills, eyes, or scales of the affected fish, alongside observing any abnormal behavior. However, this method is time-consuming, labor-intensive, and often fails to promptly identify diseased fish. Previous studies have reported traditional machine learning methods to identify fish diseases. Chakravorty et al. (2015) proposed an image processing-based method to diagnose the epidemic ulcerative syndrome (EUS), which causes massive mortality of freshwater fish. This method involves identifying and processing features from images of infected fish using principal component analysis to create feature vectors. Subsequently, these feature vectors were specifically classified using the Euclidean distance metric to discern the health condition of the fish. Waleed et al. (2019) employed various color spaces (RGB, Ycbcr, and XYZ) on the input image during the preprocessing stage, and then used Gaussian distribution to measure the probability of any infected area for segmentation.

With the advancement of deep learning, particularly in computer vision, an increasing number of models are being applied to detect fish anomalies. Zhao et al. (2022) proposed a high-precision, lightweight end-to-end dead fish detection model DM-YOLOv4. The model replaced the backbone network with MobileNetV3 and utilized depth-separable convolution to achieve a lightweight structure. Besides, the network incorporated deformable convolution to enhance detection accuracy, achieving an impressive 95.47 % accuracy in detecting dead fish. For real-time detection, Hu et al. (2021) introduced a noninvasive, rapid, low-cost procedure utilizing an underwater imaging system and the YOLOv3-Lite model to detect fish behavior in a mixed polyculture system. The proposed method produced a precision of 0.897 and 240 frames per second on a real dataset. Zhao et al. (2018) presented a method based on modified motion influence maps and RNN for detecting, locating, and identifying local abnormal fish behavior in intensive aquaculture environment. It achieves accuracy 98.91 %, 91.67 % and 89.89 % of detection, localization and recognition, respectively. Wang et al. (2022) proposed an improved YOLOv5 model, which overcomes the difficulties of masking, motion blur, and small target detection in juvenile cannibalism scenes. Specifically, the model improves the connection of the feature fusion network, and introduces a lightweight upsampling operator by adding an attention module to YOLOv5s, reaching 14 %, 12.6 % and 12.2 % increase of network speed, detection accuracy, mAP50 respectively compared to the baseline YOLOv5. Furthermore, fish health was also evaluated by quantitative analysis of specific abnormal behaviors of fish. Liao et al. (2011) extracted the average swimming speed of zebrafish based on trajectory analysis, and characterized the stress response of fish groups exposed to Cu²⁺ pollution through changes in swimming speed. Davidson et al. (2011) studied the health of rainbow trout under low water exchange rates by estimating swimming speed and side swimming volume.

Fish diseases are often reflected in multiple dimensions, such as visual surface infections and abnormal behavior. Previous machine learning methods and deep learning methods solely focused on analyzing a single characteristic dimension of diseased fish, resulting in the loss of valuable information from other dimensions. Developing multi-task of detection and segmentation algorithms can help aquaculture personnel better analyze the condition of the fish. Inspired by YOLOP (Wu et al., 2022), which can perform traffic object detection,

drivable area segmentation and lane detection simultaneously, making great use of effective information during driving. However, YOLOP merely adds the weighted losses from the three branches and fails to consider the gradient conflicts arising from multiple tasks. Therefore, this study proposed YOLO-FD, extending the YOLOv8 (Jocher et al., 2023) backbone by incorporating a novel semantic segmentation branch, and multi-task optimization strategies are used to automatically balance the weights of different losses and optimize the conflicting gradients during the training process. YOLO-FD can perform diseased fish object detection and infected areas segmentation simultaneously, and evaluate the degree of fish diseases from multiple perspectives of computer vision.

It is worth noting that the model we proposed differs fundamentally from the instance segmentation model. As illustrated in Fig. 1, instance segmentation involves object detection first, followed by semantic segmentation of each target instance within the bounding boxes. However, our model performs both object detection and semantic segmentation simultaneously. It assesses the severity of infection by statistically analyzing the pixel count within the bounding boxes' infected regions. Additionally, in contrast to instance segmentation, where the semantic segmentation for each target instance within bounding boxes is calculated separately, our approach only requires one-time computation for the entire image's semantics. Consequently, the proposed model gets a faster inference speed than instance segmentation.

The main contributions of this paper are as follows:

- (1) This paper proposed a multi-task learning network for simultaneous detection and segmentation of fish diseases. By combining the current state-of-the-art object detection model and adding only an additional semantic segmentation branch to its backbone, the proposed model accurately evaluated the degree of fish diseases from multiple informational perspectives.
- (2) For the optimization of multi-task learning, weight uncertainty and PCGrad were used to balance the weights of different losses and optimize conflicting gradients from multiple head during the training process.
- (3) We designed ablation experiments to validate the effectiveness of the multi-task scheme in both the *Nocardiosis* fish dataset and the public dataset VOC2012 segmentation. It was proved that joint end-to-end training for detection and segmentation tasks was sufficient. Furthermore, the multi-task scheme outperformed single-task scheme, yielding better overall results.
- (4) In the domain of fish disease detection, there is limited usage of multi-task learning methods. This paper introduces new perspectives and methodologies for subsequent research in this field.

2. Material and methods

2.1. Dataset acquisition

The experimental data was collected in the Langya Base of the Yellow Sea Fisheries Research Institute of the Chinese Academy of Fishery Sciences. In this study, 30 largemouth bass were raised and underwater videos of largemouth bass suffering from *Nocardiosis* were collected. The camera used was the Barlus SK2F-4PX10 underwater camera, which was placed at the bottom of the pool to record videos. It has a focal length of 3.6 mm and automatically adjusts for low light conditions. The videos were captured at a resolution of 1920 × 1080 with a frame rate of 25fps. Image frames were randomly extracted from videos spanning a week, and suitable pictures were selected as datasets for object detection and semantic segmentation. As shown in Fig. 2, in order to enhance the model's robustness, images captured from diverse scenarios of illumination, occlusion, and water reflections were gathered. Our study was approved by the China Agricultural University Experimental Animal Welfare and Animal Experiment Ethical Committee (Issue No. Aw51213202-5-1).



Fig. 1. Comparison between YOLOv8-seg (instance segmentation) and YOLO-FD (ours). (a) The infected area instance segmentation results of YOLOv8-seg with inference speed 45 ms on NVIDIA GeForce RTX 3060. (b) The Multi-task results of YOLO-FD with inference speed 29 ms on NVIDIA GeForce RTX 3060.

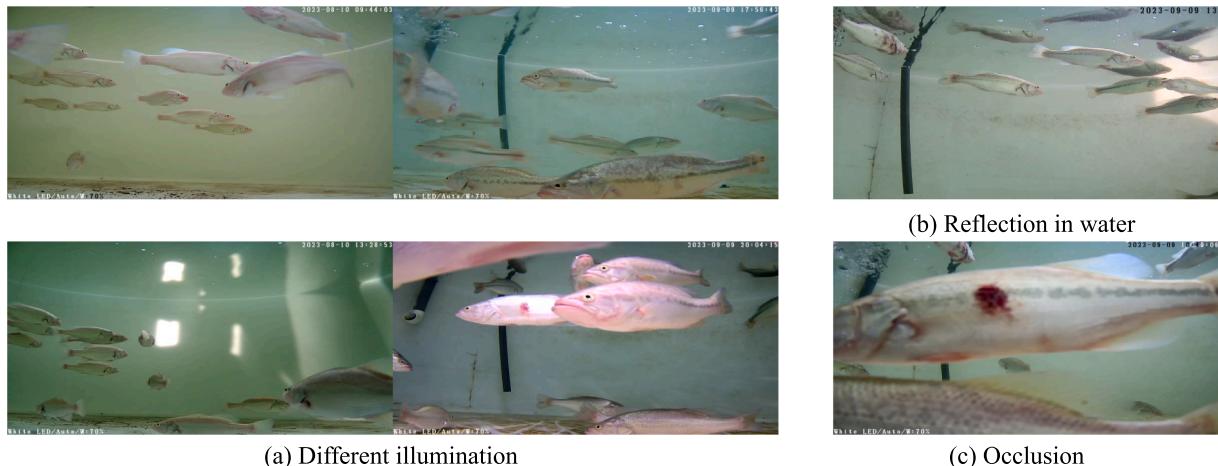


Fig. 2. Sample images of dataset.

Regard to data collection of *Nocardiosis* infected fish, the infected ones were gradually identified by visible and distinct characteristics of fish skin, eyes and behavior. Original dataset consists of 722 pictures containing *Nocardiosis* infected fish and 350 pictures of healthy fish, as a total of 1072 pictures, and then it was expanded through data augmentation, including applying mosaic, horizontal flipping, random shearing, and scaling to both object detection and semantic segmentation, and applying saturation adjustment to the object detection alone. The division ratio of training set and validation set was 8:2. LabelImg was used to annotate the object detection dataset and Labelme was used to annotate the semantic segmentation dataset. An annotation example of object detection and semantic segmentation is showed in Fig. 3.

2.2. Network architecture

The present study proposes a multitask learning network, YOLO-FD, which can perform diseased fish detection and infected area segmentation simultaneously, as shown in Fig. 4. Logically, the network can be

divided into four parts: Backbone, Neck, Detect head, and Segmentation head. The Backbone and Neck can be considered as a shared encoder, while the Detect head and Segmentation head can be seen as decoders, extracting the information they need. The recognition of diseased fish is performed at the image level in object detection, while semantic segmentation operates at the pixel level to identify infected areas. These segmented areas are then used to calculate the proportion of the fish disease infection. By combining the results of object detection and semantic segmentation, an accurate assessment of the disease severity is achieved.

2.2.1. YOLOv8

YOLOv8 is a cutting-edge, state-of-the-art model that builds on the success of previous YOLO versions and introduces new features and improvements to further improve performance and flexibility. It uses a series of convolutional layers to extract features from the input image and generate bounding boxes and confidence scores for the object. The architecture consists of several key components, including backbone,

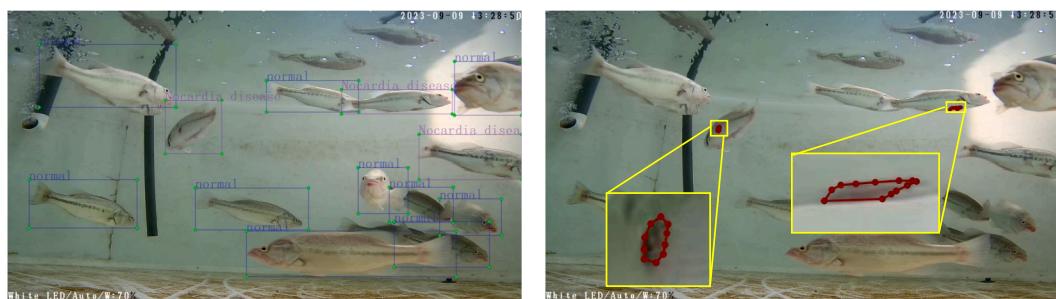


Fig. 3. Examples of object detection annotation on the left picture and semantic segmentation annotation and the enlarged annotation of infected areas on the right picture.

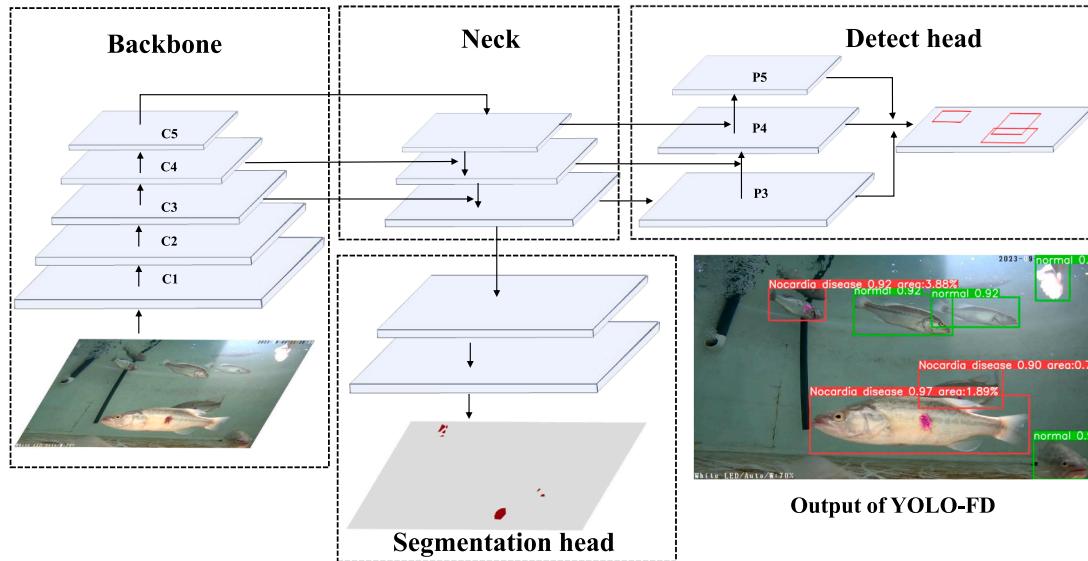


Fig. 4. The architecture of YOLO-FD.

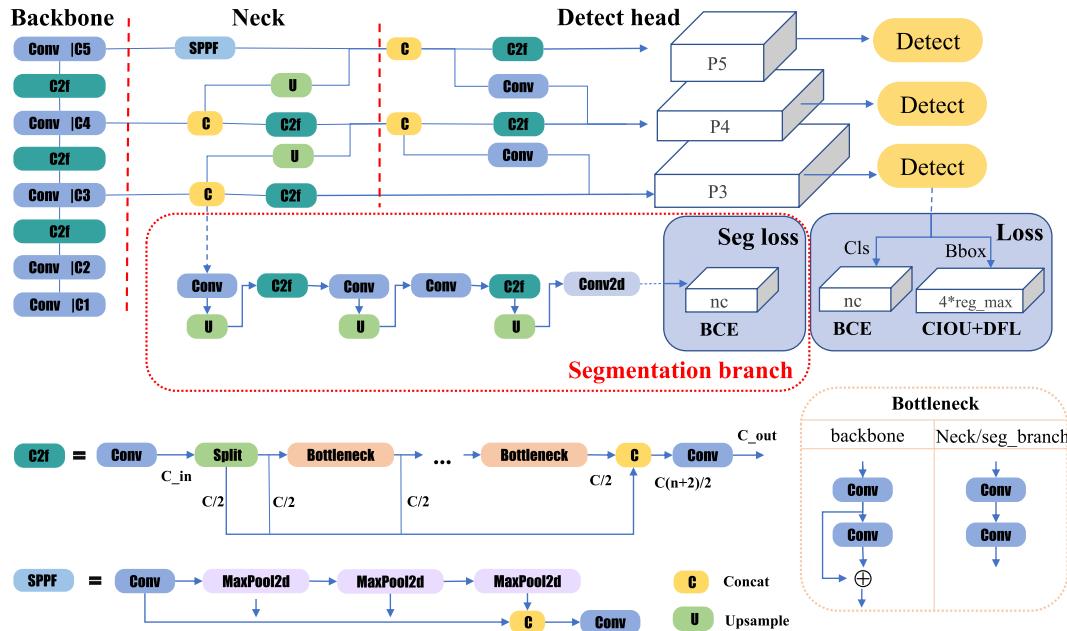


Fig. 5. The network architecture of YOLO-FD.

SPPF layer, C2f module, neck and detection module, as shown in Fig. 5. The backbone consists of convolutional layers and C2f modules, which are used to extract relevant features from the input image. The SPPF layer is applied to process features of various scales to achieve adaptive input size. Compared with SPP, SPPF improved speed by reducing calculation amount. The C2f module refers to design idea of ELAN in YOLOv7 (Wang et al., 2023) and is more lightweight than the C3 module in YOLOv5 (Jocher, 2020). Since the C2f structure has more residual connections, bringing a richer gradient flow which positively affects semantic segmentation in subsequent experiments as discussed in section 3.2. There are slight changes of the C2f modules in the backbone, neck, and segmentation branch of the network. Specifically, in the backbone, the bottleneck of the C2f module has residual connections to enrich the extracted features, but not in the neck and segmentation branch. In the neck, the structure of Feature Pyramid Network (FPN) allows the network to detect objects at multiple scales. In the detection

module, a decoupled detection head allows to run objectness, classification, and regression tasks independently, leading to an improvement of the overall accuracy of the model through handling respective tasks independently of each branch. Furthermore, the design of the detection head has been optimized for both speed and accuracy, as well as particular attention to the number of channels and kernel sizes at each layer to maximize performance. The detection head employs BCE loss for classification, and DFL loss and CIOU loss for regression.

According to different task requirements, YOLOv8 offers five versions: YOLOv8n, s, m, l, and x, differing in width and depth of the network. YOLOv8s, in particular, maintains detection accuracy with few network parameters, making it suitable for real-time detection. Therefore, in this study, YOLOv8s was selected as the foundation for the improved model. The images fed into the network undergo features extraction via the backbone, followed by the fusion of multi-scale features through the neck. The Detect head then outputs the classification

and bounding boxes of detected objects, which are used to calculate classification loss and bounding box regression loss.

2.2.2. Detect head

The Detect head consists of a bottom-up PAN (Path Aggregation Network) and decoupled detection heads. The structure of PAN merges information from different levels to enhance the receptive field, which aids in improving the network's understanding of the overall context information of the entire image. YOLOv8 departs from previous anchor-based methods and adopts an anchor-free approach, directly predicting offsets for all four sides of the bounding boxes, avoiding the generation of numerous redundant candidate boxes and introduction of prior knowledge, making it more suitable for targets of various scales and proportions. In terms of positive and negative sample assignment, a dynamic assignment strategy called Task-Aligned Assigner is employed. It calculates the IOU between anchors and Ground Truth Boxes using a task-based adaptive dynamic threshold to determine the allocation of positive and negative samples. Thus, Task-Aligned Assigner can adaptively adjust the allocation ratio of positive and negative samples in accordance with specific tasks and data distribution, and can achieve better detection performance.

2.2.3. Segmentation head

A simple segmentation head here is shown in Fig. 5 outlined by the red dashed box, and the head is composed of the fundamental building blocks of YOLOv8 for maintaining network consistency and lightweight design. FPN bottom-level feature maps of the network, with dimensions ($W/8, H/8, 384^*w$), serve as the input to the semantic segmentation branch, as depicted in Fig. 6. Besides, upsampling and convolution are symmetrically applied to effectively propagate and integrate information in the segmentation branch, while the C2f module is used to convert rough features into fine features, which enhances the model's understanding of image semantics. For different network widths, the segmentation branch is accordingly extended. In this paper, w holds a similar meaning as the original width w of YOLOv8, and for simplicity, we use YOLOv8s as the baseline, where w is set to 1. After three upsampling operations, the feature maps are restored to a size of ($W, H, 2$), where each pixel represents the probability of being an infected area in the input image. Besides, an additional SPPF module is unnecessary for the semantic segmentation branch to accommodate inputs at different scales due to the existed shared SPPF module in the neck network. Furthermore, nearest-neighbor interpolation is used in the upsampling layer instead of deconvolution, effectively reducing computational overhead while maintaining a certain level of accuracy.

In semantic segmentation, it does not have explicitly positive and negative samples as in object detection, but it involves pixel-level classification. The proposed model employs a class balancing, assigning lower weights to pixels of common classes but higher weights to pixels of less common classes, based on the sample quantities of pixel classes in the dataset. The class balancing strategy helps to address significant imbalance in the number of pixels among different classes, ensuring that each class has a reasonable impact on the loss function during training. Therefore, the class balancing approach is also adaptable to various semantic segmentation datasets.

2.3. Multi-task learning optimization

One of the main challenges in multi-task learning involves how to make multiple tasks achieve better performance when they cooperate with each other, aiming to realize a synergistic effect where one plus one is greater than two. This paper works out the challenge by utilizing weight uncertainty (Kendall et al., 2018) and PCGrad (Yu et al., 2020) to optimize multi-task learning. By balancing the magnitude of the losses and optimizing the gradient update directions, the model achieves better accuracy in both object detection and semantic segmentation tasks compared to single-task learning.

2.3.1. Weight uncertainty

The fusion of loss functions has been a major challenge in multi-task learning. To merge loss functions of both object detection head and semantic segmentation head, this paper employs a loss function fusion method using the homoscedastic uncertainty of each task to balance losses of multiple tasks. Next, we will derive a multi-task loss function based on maximizing the Gaussian likelihood with homoscedastic uncertainty. Let $f^W(x)$ be the model output with weights W when the input is x . The probability model is defined as Equation (1), in which a Gaussian function with mean given by the model output $f^W(x)$ is used to define the probability model for the regression task, where σ is the observation noise scalar:

$$p(y|f^W(x)) = \mathcal{N}(f^W(x), \sigma^2) \quad (1)$$

For the classification task in semantic segmentation, we typically squash the model's output divided by σ^2 through a softmax function and then sample from the resulting probability vector:

$$p(y|f^W(x), \sigma) = \text{Softmax}\left(\frac{1}{\sigma^2}f^W(x)\right) \quad (2)$$

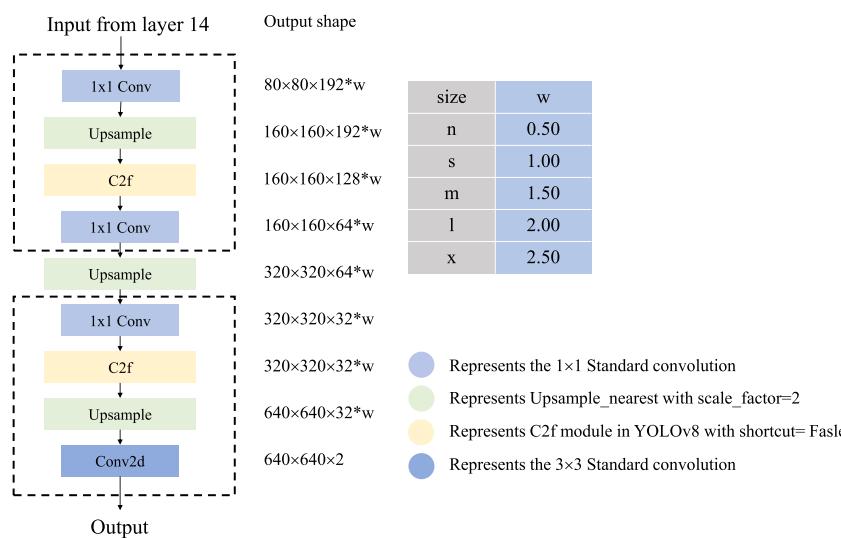


Fig. 6. The network composition of the proposed segmentation head, along with the output shapes at each layer and the network width scaling factor.

This scalar is fixed or learnable, and its size determines how uniform the discrete distribution is, which is related to its uncertainty.

When regard to multiple model outputs, we define the likelihood to factorize across the outputs, leading to the following multi-task likelihood:

$$p(y_1, \dots, y_K | f^W(x)) = p(y_1 | f^W(x)) \cdots p(y_K | f^W(x)) \quad (3)$$

Assuming that the output of the model consists of regression output y_1 of object detection task and classification output y_2 of semantic segmentation task, modeled by Gaussian likelihood and softmax likelihood mentioned above respectively, we get the following Equation (4):

$$\begin{aligned} p(y_1, y_2 | f^W(x)) &= p(y_1 | f^W(x)) \bullet p(y_2 | f^W(x)) \\ &= \mathcal{N}(f^W(x), \sigma_1^2) \bullet \text{Softmax}\left(\frac{1}{\sigma_2^2} f^W(x)\right) \end{aligned} \quad (4)$$

Equation (5) shows the minimization objective $\mathcal{L}(W, \sigma_1, \sigma_2)$ of our multi-output model. It denotes $\mathcal{L}_1(W) = \|y_1 - f^W(x)\|^2$ as the Euclidean loss for y_1 and $\mathcal{L}_2(W) = -\log \text{Softmax}(y_2, f^W(x))$ as the cross-entropy loss for y_2 , and finally the form of the fused loss function is obtained. As the noise parameter σ increases or decreases, the weight of $\mathcal{L}(W)$ decreases or increases correspondingly. and $\log \sigma$ limits the noise parameter to increase too much, so it is a regularization term to some extent.

$$\begin{aligned} \mathcal{L}(W, \sigma_1, \sigma_2) &= -\log p(y_1, y_2 = c | f^W(x)) \\ &= -\log \mathcal{N}(y_1; f^W(x), \sigma_1^2) \cdot \text{Softmax}(y_2 = c; f^W(x), \sigma_2^2) \\ &\approx \frac{1}{2\sigma_1^2} \|y_1 - f^W(x)\|^2 + \log \sigma_1 - \frac{1}{\sigma_2^2} \log \text{Softmax}(y_2, f^W(x)) + \log \sigma_2 \\ &= \frac{1}{2\sigma_1^2} \mathcal{L}_1(W) + \frac{1}{\sigma_2^2} \mathcal{L}_2(W) + \log \sigma_1 + \log \sigma_2 \end{aligned} \quad (5)$$

2.3.2. PCGrad

Another optimization issue in multi-task learning arises from conflicting gradients, which can notably degrade model performance under certain conditions. In this study, tasks like object detection and semantic segmentation may lack strong correlations intuitively, resulting in conflicting gradient update directions. Therefore, this paper employs the PCGrad to alleviate this situation.

We define ϕ_{ij} as the angle between the gradients g_i and g_j of two tasks. If $\cos \phi_{ij} < 0$, it is considered that there is a gradient conflict, and then PCGrad modifies the gradient of each task to minimize the negative impact on other task gradients. If not, it keeps the original gradient to continue updating the network. Assuming that the gradient of task T_i is g_i and the gradient of task T_j is g_j , (1) Firstly, determining whether there is a conflict between g_i and g_j by calculating the cosine similarity between them, where a negative value indicates the conflict, (2) If the cosine value is a negative number, we use the projection g'_i of g_i on the normal plane of g_j to replace g_i , as shown in Equation (6). If the gradient does not conflict, that is, the cosine value is non-negative, then the original gradient g_i remains unchanged. (3) PCGrad repeats the above operations for every two tasks in the multi-task.

$$g'_i = g_i - \frac{g_i \cdot g_j}{\|g_j\|^2} g_j \quad (6)$$

PCGrad and weight uncertainty optimize multi-task learning from

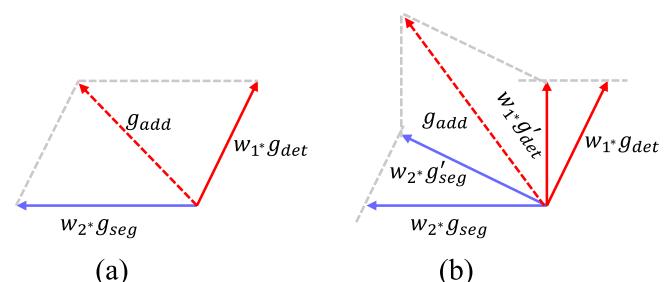


Fig. 7. Combining weight uncertainty with PCGrad.

different perspectives by considering gradient directions and loss function weights, respectively. Therefore, this study combines them to enhance overall performance of the network. As shown in Fig. 7, (a) Without PCGrad, the gradient's actual update direction g_{add} for weight uncertainty is the vector sum of the gradients for the detection task and the segmentation task, where w_1 and w_2 are learnable weights based on the uncertainty between tasks. (b) When combining PCGrad and weight uncertainty, the actual gradient update direction g_{add} is the vector sum of gradients from the detection task and the segmentation task after being projected onto each other's normal planes, where g_{det} represents the projection of g_{det} onto the normal plane of g_{seg} , while g'_{seg} represents the projection of g_{seg} onto the normal plane of g_{det} .

2.4. Loss function design

There are both detection and segmentation heads in the network, so the multi-task loss comprises two parts. The detection loss L_{det} is a weighted combination of classification loss and bounding box regression loss, as shown in Equation (7). The typical values are set as follows: $\alpha_1 = 0.5, \alpha_2 = 7.5, \alpha_3 = 1.5$.

$$L_{det} = \alpha_1 L_{cls} + \alpha_2 L_{iou} + \alpha_3 L_{dfl} \quad (7)$$

Where L_{cls} represents the classification loss, which is calculated by Binary Cross-Entropy loss. L_{iou} and L_{dfl} represent the bounding box regression loss. L_{iou} stands for *CloULoss*, which considers the distance, overlap rate, scale similarity, and aspect ratio between the predicted box and the ground truth box. L_{dfl} calculates the loss in the form of cross-entropy between the probabilities of the two nearest positions' offsets in the labels and the predictions, allowing the network to swiftly focus on the distribution around the target location.

Table 1
Experimental configuration.

Configuration	Parameter
CPU	13th Gen Intel Core i7-13700KF
GPU	NVIDIA GeForce RTX 3060
Operating system	Window10
Accelerated environment	CUDA 11.6, CUDNN 8.6.0
Development environment	Pycharm2023.3.2
Deep learning framework	Pytorch1.10.0

For the segmentation loss L_{seg} , Cross-Entropy Loss is utilized to minimize classification error between the model's pixel-level output and the target. L_{seg} is defined as follows:

$$L_{seg} = L_{ce} \quad (8)$$

In summary, by combining Equation (5), we derive the final form of the loss function as follows, where σ_1 and σ_2 are learnable parameters.

$$L_{all} = \frac{1}{2\sigma_1^2} L_{det} + \frac{1}{\sigma_2^2} L_{seg} + \log\sigma_1 + \log\sigma_2 \quad (9)$$

3. Experimental results and analysis

3.1. Experimental details

The experimental configuration is as shown in Table 1. A total of 1072 images containing both healthy and diseased fish were used for model training and validation. The optimizer used for the model is AdamW, with an initial learning rate of 0.00167, a momentum of 0.9, and a weight decay of 5e-4. The learnable parameters σ_1 and σ_2 for the loss functions in multi-task learning had initial values of 10.675 and 0.157, respectively, and weight decay was set to 1e-5. In the experiment, the batch size was set to 8, and end-to-end training was performed for 300 epochs without need for additional segmentation branch training.

3.2. Experimental results

In this section, our experimental results of object detection, semantic segmentation, and multi-task learning will be discussed. The model was trained and validated on the *Nocardiosis* fish dataset, using Precision, Recall, and Mean Average Precision as evaluation metrics for object detection. The semantic segmentation performance was evaluated by mean Intersection over Union. The inference speed includes preprocess time, model inference time, and postprocess time. Due to the limitations of the *Nocardiosis* fish dataset and for better validation of the model's generalization, additional experiment results have been supplemented in Appendix A.

3.2.1. Object detection result

Several advanced object detection models were selected and compared to the proposed model, such as Faster R-CNN, YOLOv5, YOLOX, and YOLOv8, as shown in Table 2. It can be seen our model surpasses Faster R-CNN, YOLOv5s, YOLOX-s, and YOLOv8s in Precision, mAP50, and mAP50-95. However, it shows lower speed than YOLOv5s and YOLOv8s due to its additional semantic segmentation head, which requires the input feature map to be upsampled three times to restore it to the original image size. This process consumes a significant amount of computational resources. Precision and Recall are typically trade-offs, where increasing Precision often leads to a decrease in Recall. In fish disease detection, we prioritize improving Precision to ensure accurate identification of diseased fish and to prevent false alarms.

Fig. 8 illustrates the comparison between Faster R-CNN, YOLOv5s, and YOLO-FD in detecting fish diseases. Due to the characteristic of multi-task information sharing, YOLO-FD can perform both holistic

semantic and pixel-level semantic detection of fish diseases, leading to more precise bounding box positions. As shown in Fig. 8, Faster R-CNN exhibits noticeable error detection and redundant detection boxes, while YOLOv5s shows both error detection and missed detection. Additionally, when comparing the first column on the right, YOLO-FD locates more accurate bounding box positioning in cases of occlusion, while Faster R-CNN detects redundant boxes, and YOLOv5s also exhibits error detection.

3.2.2. Segmentation result

Classic semantic segmentation networks were chosen for comparison with the proposed model, including U-Net, Deeplabv3, and Deeplabv3plus, and also lightweight versions of Deeplabv3 and Deeplabv3plus with MobileNet as the backbone to ensure that the number of parameters in the networks was in the same order of magnitude as YOLO-FD. As shown in Table 3, YOLO-FD shows an improvement of 11.3 % in mIOU compared to the U-Net commonly used for medical image segmentation. YOLO-FD also improves 4 % in mIOU even though compared to the deeplabv3plus_resnet50. Moreover, the inference speed of YOLO-FD is at least 5 times faster than other semantic segmentation models. Since the infected areas on the fish's body are not fixed and can vary in color, achieving accurate segmentation based solely on pixel semantics is challenging. Therefore, the overall mIOU may not be very high. However, the model's output is of size $S \times S$ in YOLO-FD, which differs from semantic segmentation models that restore the output to the original image size. This difference enhances local semantic features and helps address the challenge of accurately segmenting infected areas.

Fig. 9 shows the comparison among U-Net, deeplabv3plus_mobilenet, and YOLO-FD in segmenting infected areas in diseased fish. To offer a clearer view of the segmented infected areas, we cropped some images and resized them to their original size. The results demonstrate that YOLO-FD achieves more accurate and continuous segmentation. In addition, due to the shared information with object detection, YOLO-FD avoids errors like segmenting the infected areas of fish reflection on the water surface (as observed with deeplabv3plus_mobilenet) or incorrectly segmenting the eyes into infected areas (as observed with U-Net).

3.2.3. Multi-task result

Due to the limited application of multi-task networks in fish disease detection, we chose YOLOP for comparison. One of semantic segmentation branches of YOLOP was removed, leaving the drivable area semantic segmentation branch for segmenting the infected areas. The multi-task results are presented in Table 4, showcasing YOLO-FD outperforms YOLOP with a 13.7 % higher mAP50-95 and a 15.1 % higher mIOU. The increased FLOPs result from the c2f structure in YOLO-FD, which includes abundant residual connections. Despite this, our model can still maintain similar inference speeds and residual connections are proven to be beneficial for both object detection and semantic segmentation. Visual results can be seen in Fig. 10, where YOLO-FD demonstrates higher detection accuracy in object detection and generates fewer redundant boxes. In semantic segmentation, YOLO-FD excels at segmenting irregular infected areas of diseased fish, providing smoother segmentation results than YOLOP.

3.3. Ablation experiments

Three ablation experiments were conducted to validate the effectiveness and generalization of the proposed method by training and validating on the *Nocardiosis* fish dataset and VOC2012 segmentation dataset respectively. All evaluation metrics remained consistent with those mentioned above.

3.3.1. Joint training and alternating training

To verify the effects of different training strategies on network performance, we compared two different training methods: joint training

Table 2

Object detection results: comparison of YOLO-FD with current state-of-the-art object detection models.

Algorithms	Precision	Recall	mAP50	mAP50-95	params (M)	speed
Faster R-CNN	0.714	0.691	0.891	0.690	41.3	137.9ms
YOLOv5s	0.901	0.913	0.938	0.768	7.0	8.0ms
YOLOX-s	0.906	0.873	0.867	0.722	8.9	74.6ms
YOLOv8s	0.905	0.904	0.938	0.785	11.2	13.9ms
YOLO-FD	0.933	0.902	0.942	0.787	12.0	21.3ms

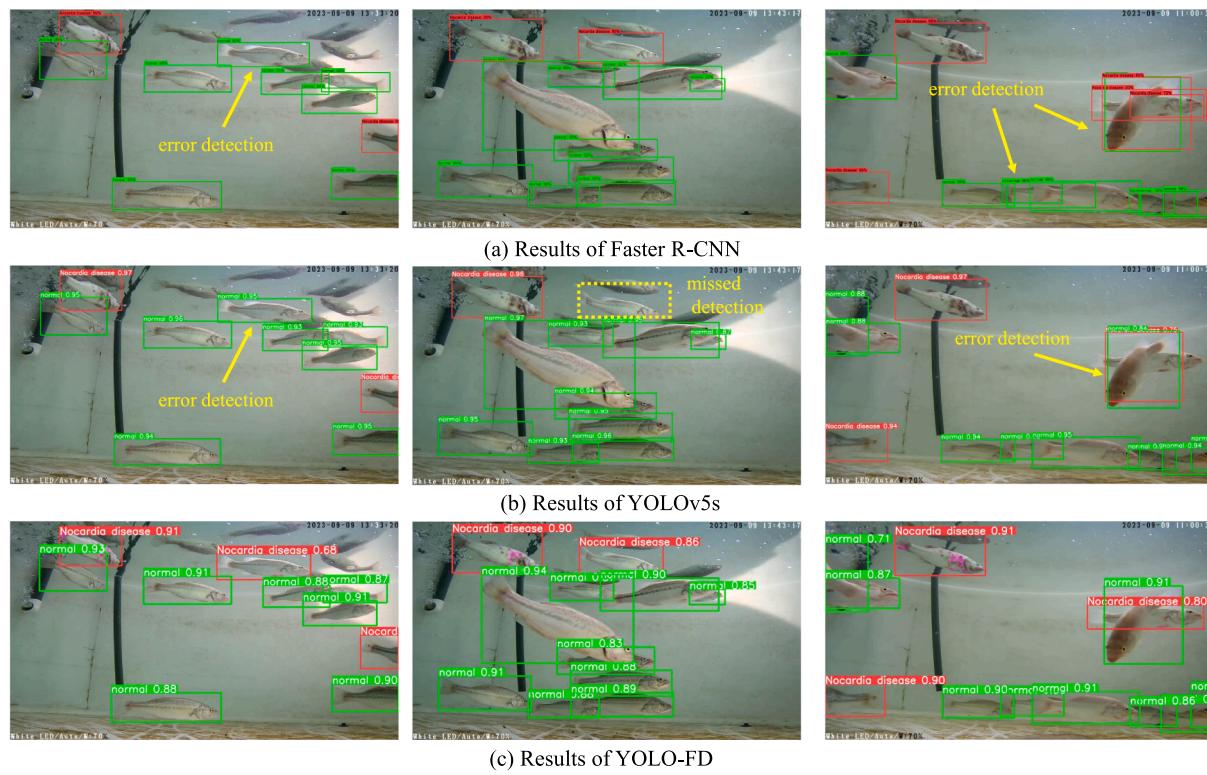


Fig. 8. Comparison of Faster R-CNN, YOLOv5s, and YOLO-FD in fish disease detection. Yellow arrows and dashed boxes in Faster R-CNN and YOLOv5s indicate error detections and missed detections, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3

Semantic segmentation results: Comparison of YOLO-FD with classic semantic segmentation networks.

Algorithms	mIOU	params(M)	FLOPs	speed
U-Net	0.667	4.3	40.1G	293 ms
deeplabv3_resnet50	0.704	39.6	51.1G	185 ms
deeplabv3_mobilenet	0.702	5.1	5.8G	101 ms
deeplabv3plus_resnet50	0.754	39.7	62.4G	198 ms
deeplabv3plus_mobilenet	0.735	5.2	16.8G	120 ms
YOLO-FD	0.794	12.0	52.8G	21.3 ms

and alternating training. In alternating training, Det first refers to performing detection before segmentation, whereas Seg first refers to performing segmentation before detection. The results are shown in Tables 5 and 6, our model performs well in both object detection and semantic segmentation by joint training, while only performs well in object detection but poorly in semantic segmentation through alternating training, resulting from the fact that our model is closer to the object detection task.

3.3.2. Multi-task and single-task

To verify the effectiveness of our multi-task scheme, performance of the multi-task scheme and that of the single-task scheme were compared in this section. Firstly, the proposed model was trained to perform both object detection task and segmentation task simultaneously. Secondly, the proposed model was trained to perform object detection task and segmentation task separately. Table 7 shows performance comparison between the multi-task scheme and the single-task scheme specific to their respective tasks. Table 8 also indicates that adopting a multi-task scheme does not compromise the accuracy of object detection and segmentation. On the contrary, it can maintain either the same or better performance compared to the single-task scheme. Specifically, by

sharing the same backbone, the multi-task model saves considerable inference time compared to performing each task individually.

We used Grad CAM (Selvaraju et al., 2017) to visualize the last layer of the shared backbone in the multi-task network. As shown in Fig. 11, in the Det only scheme, the model tends to focus on the overall fish or the center of the fish, with less attention to the infected areas. However, in the multi-task scheme, we find that the model pays more attention to the outer contour of the fish, which aids in precise localization of the bounding boxes. Moreover, the infected areas also receive significant attention, facilitating the identification and classification of diseased fish.

3.3.3. Multi-task optimization strategy

To verify the effects of different multi-task learning optimization strategies on model performance, we conducted comparative experiments on both *Nocardiosis* fish dataset and VOC2012 segmentation dataset. Due to computational resource limitations, the model was only trained and tested on the training and validation sets of the VOC2012 segmentation dataset, without loading pre-trained weights during the training process. As shown in Table 9, we used YOLOv8s as the baseline and integrated YOLO-FD with various multi-task learning optimization strategies including equal weighting, gradnorm (Chen et al., 2018), DWA (Liu et al., 2019), PCGrad, CAGrad (Liu et al., 2021), and weight uncertainty. It can be observed that YOLO-FD with equal weighting performs well in object detection but significantly sacrificed the accuracy of semantic segmentation. However, the proposed method, YOLO-FD with PCGrad and weight uncertainty, achieves a balanced performance in both aspects. It exhibits the highest object detection and semantic segmentation performance on the VOC2012 segmentation dataset compared to other methods. Meanwhile, on the *Nocardiosis* fish dataset, it maintains comparable mAP50 and mAP50-95 accuracy to YOLOv8s while achieving the highest mIOU.

Fig. 12 illustrates the mAP50 curve, mAP50-95 curve, and mIOU

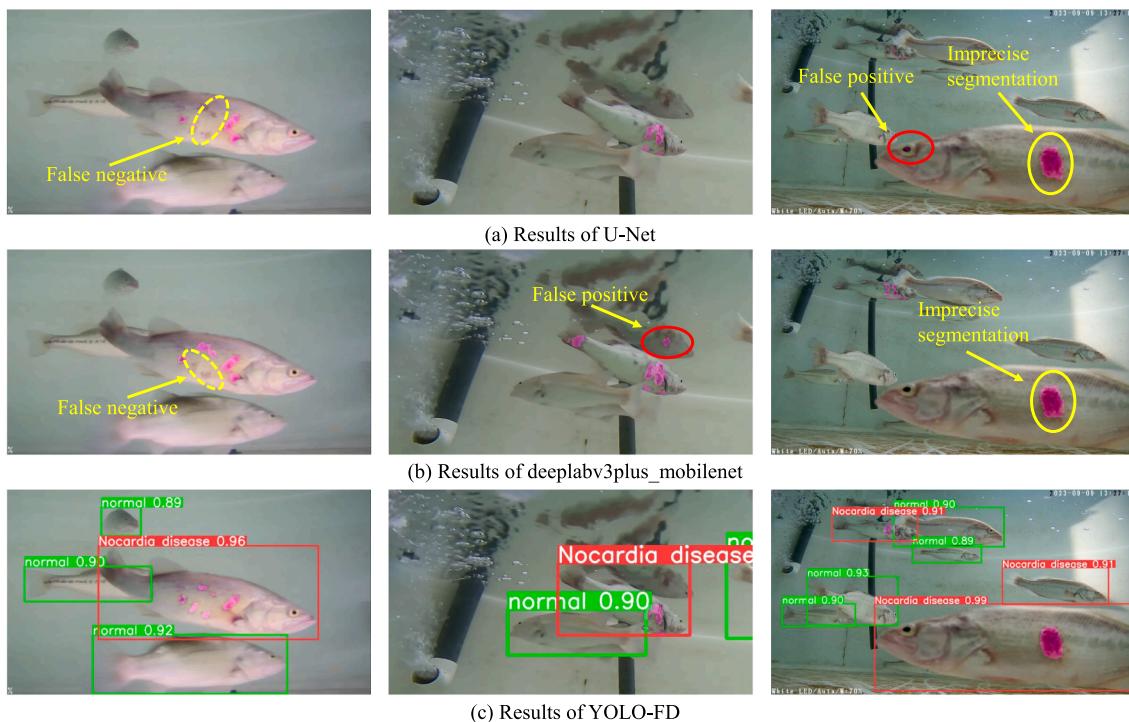


Fig. 9. Comparison of classic semantic segmentation networks and YOLO-FD in segmenting infected areas of diseased fish. Yellow dashed ellipses highlight false negatives, yellow solid ellipses indicate imprecise segmentation areas, and red solid ellipses mark false positives. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Multi-task results: comparison between YOLO-FD and YOLOP.

Algorithms	mAP50	mAP50-95	mIOU	params(M)	FLOPs	speed
YOLOP	0.939	0.650	0.643	7.9	15.4G	24.4ms
YOLO-FD	0.942	0.787	0.794	12.0	52.8G	21.3ms

curve throughout the training process on both datasets. It is visually evident that weight uncertainty, weight uncertainty combined with CAGrad, and weight uncertainty combined with PCGrad show more pronounced optimization effect on multi-task learning during training

Table 5
Comparison of different training strategies on *Nocardiosis* fish dataset.

Training method	Precision	Recall	mAP50	mAP50-95	mIOU
Det first	0.919	0.889	0.936	0.787	0.758
Seg first	0.912	0.916	0.938	0.790	0.761
Joint training	0.933	0.902	0.942	0.787	0.794

process compared to equal weighting, PCGrad, and DWA. Additionally, the combination of gradient optimization strategies with weight uncertainty slightly outperforms weight uncertainty solely, while weight uncertainty combined with PCGrad demonstrates superior performance

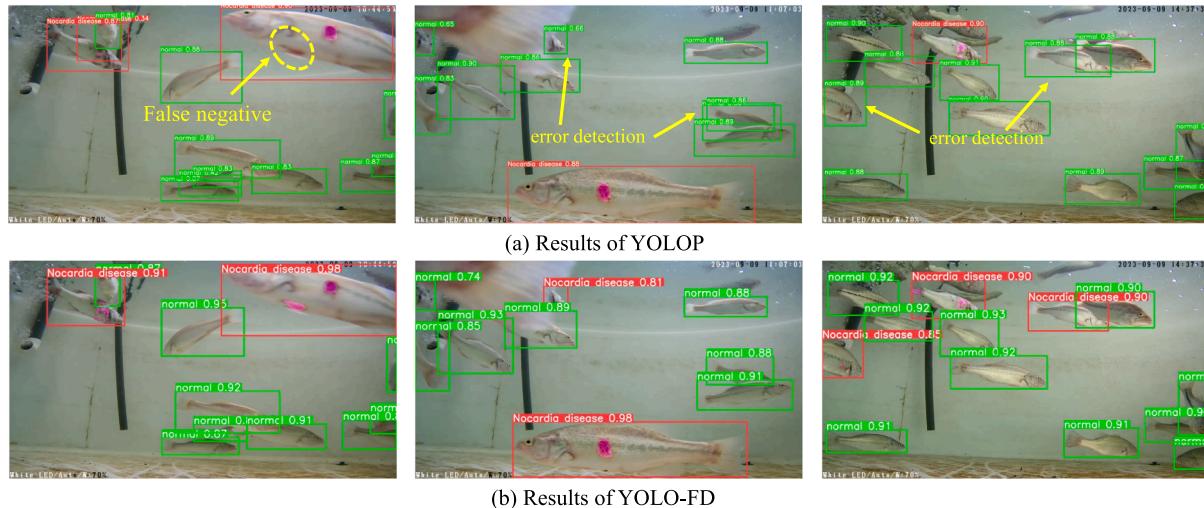


Fig. 10. Multi-task comparison in fish disease detection, yellow dashed ellipticals represent false negatives in semantic segmentation, and yellow arrows point to error detections. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6

Comparison of different training strategies on VOC2012 segmentation dataset.

Training method	Precision	Recall	mAP50	mAP50-95	mIOU
Det first	0.559	0.446	0.455	0.284	0.300
Seg first	0.552	0.431	0.448	0.282	0.301
Joint training	0.558	0.466	0.475	0.295	0.375

Table 7

Comparison of multi-task scheme and single-task scheme on Nocardiosis fish dataset.

Algorithms	Precision	Recall	mAP50	mAP50-95	mIOU	speed
Det only	0.905	0.903	0.937	0.785	—	13.9ms
Seg only	—	—	—	—	0.784	18.4ms
Multi-task	0.933	0.902	0.942	0.787	0.794	21.3ms

Table 8

Comparison of multi-task scheme and single-task scheme on VOC2012 segmentation dataset.

Algorithms	Precision	Recall	mAP50	mAP50-95	mIOU	speed
Det only	0.578	0.403	0.438	0.271	—	13.9ms
Seg only	—	—	—	—	0.375	18.4ms
Multi-task	0.558	0.466	0.475	0.295	0.375	21.3ms

in all metrics compared to other multi-task optimization strategies.

3.4. Real-world applicability

3.4.1. Practical application

This study used the aquaculture ponds at the National Innovation Center for Digital Fishery of China Agricultural University as a pilot site to design a Nocardia disease early warning system for fish. The system includes a graphical user interface (GUI), as shown in Fig. 13. By appropriately adjusting the position of the underwater cameras and configuring their network transmission parameters, an information collection platform was established. The underwater cameras transmit video data over the network, enabling real-time monitoring and early warning of fish diseases by obtaining video stream information online.

3.4.2. Economic benefits

Currently, the detection of fish diseases primarily relies on manual inspection. However, large fish farms cover extensive areas and often have poor lighting, making it difficult for workers to detect diseased fish from the water's surface. Bringing fish up for inspection disrupts their normal growth. Additionally, the early symptoms of fish diseases are often subtle and easily overlooked, leading to delayed detection and

Table 9

Comparison of different optimization strategies for multi-task learning.

Optimization Strategy	Nocardiosis fish dataset			VOC2012 segmentation dataset		
	mAP50	mAP50-95	mIOU	mAP50	mAP50-95	mIOU
YOLOv8s	0.938	0.784	—	0.443	0.273	—
equal	0.945	0.795	0.753	0.436	0.277	0.303
gradnorm	0.937	0.779	0.763	0.447	0.277	0.309
DWA	0.936	0.788	0.770	0.443	0.272	0.298
uncert.	0.945	0.784	0.780	0.467	0.286	0.374
cagrad+uncert.	0.943	0.786	0.787	0.470	0.287	0.358
pcgrad+uncert. (ours)	0.942	0.787	0.794	0.475	0.295	0.375

treatment. With rising labor costs, employing a large workforce for daily routine monitoring of fish diseases is impractical.

Take a medium-sized fish farm with 50,000 fish as an example. It typically requires at least three people to inspect all the fish tanks daily for signs of disease. Assuming a labor cost of approximately \$5 per hour and each worker spending about 4 h a day inspecting the ponds, the daily labor cost is around \$60, totaling about \$1,800 per month. Annually, the labor cost for disease inspection at such a fish farm would amount to \$21,600. This calculation does not include the potential losses from delayed disease detection, which can lead to widespread infection and significant fish mortality. Each delay in detecting an outbreak can result in substantial economic losses due to the need for extensive treatment and the loss of market-ready fish.

Implementing the automated fish disease monitoring technology developed in this study can significantly reduce the cost of manual inspections. The cost of automated monitoring equipment includes underwater cameras, network transmission devices, and data processing units. An underwater camera costs about \$100, and around 20 cameras are sufficient to cover a medium-sized fish farm, totaling \$2,000. Additional equipment, such as network transmission devices and data processing units, costs approximately \$2,500, bringing the total hardware cost to about \$4,500. This one-time investment is equivalent to just three months of labor costs. Therefore, this study is economically significant as the automated system not only reduces labor costs but also enhances early disease detection, reducing the risk of severe outbreaks and associated losses.

4. Conclusion

In this paper, we propose a simple but efficient network YOLO-FD that can simultaneously perform diseased fish detection and infected areas segmentation. Leveraging multi-task optimization strategies, the

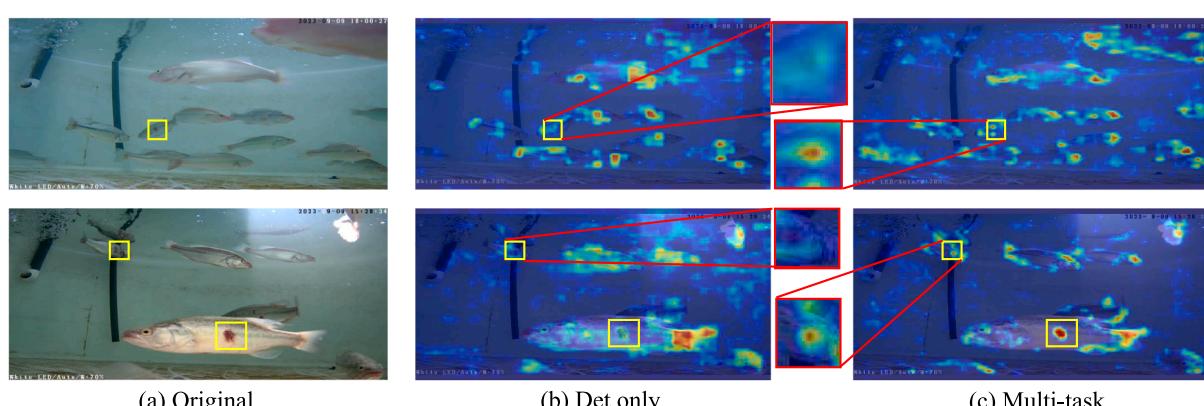


Fig. 11. Comparison between the heat maps of Det only and Multi-task. The yellow boxes indicate the infected areas, while the red boxes represent the enlarged infected areas. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

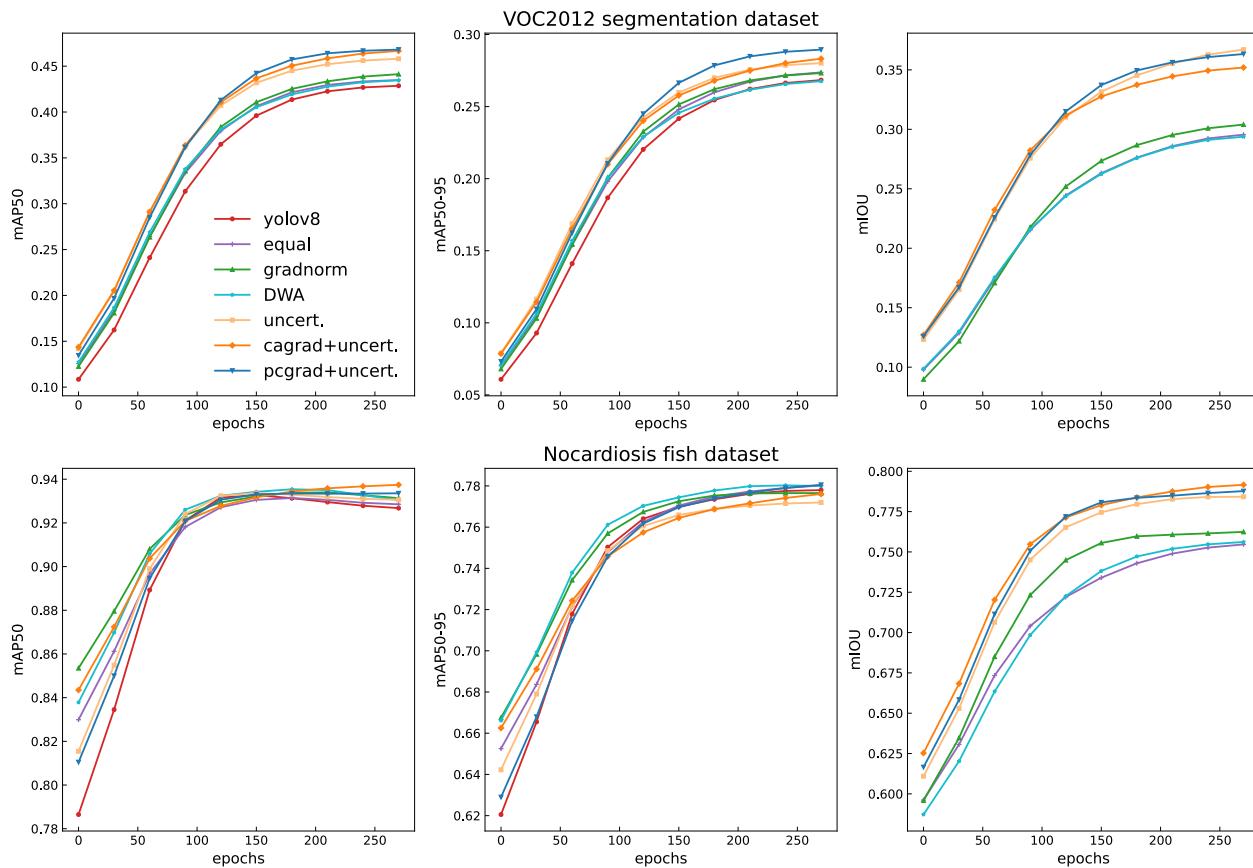


Fig. 12. The mAP50 curve, mAP50-95 curve, and mIoU curve for different multi-task learning optimization strategies on the VOC2012 segmentation dataset and the Nocardiosis fish dataset.



Fig. 13. Fish disease alert system.

model can automatically balance the weights of losses and optimize conflicting gradients, facilitating end-to-end training on various datasets without the need for adjusting hyperparameters in multi-task learning. On the *Nocardiosis* fish dataset, overall performance of the proposed model surpasses that of the state-of-the-art in both object detection and segmentation tasks. Additionally, we validate the effectiveness and generalization of the multi-task scheme in ablation experiments, showcasing that the added semantic segmentation branch promotes rather than compromises the performance of object detection. The multi-task model proposed in this paper is of significant importance in detection of diseased fish, which analyzes and assesses diseased fish from multiple dimensions, enabling accurate evaluation, and providing new ideas and methods for subsequent research.

Nevertheless, there are still many points for potential improvement identified in this study: (1) The dataset used in this paper is limited to largemouth bass with *Nocardiosis*. Future research endeavors will broaden the scope by incorporating different disease types and various

fish species to develop a more universal model for detecting fish disease. Furthermore, as the dataset expands, the annotated dataset from supervised learning will also increase linearly. To reduce the substantial cost of expert-annotated fish disease datasets, it is worth exploring the use of semi-supervised learning methods. (2) Subsequent studies will utilize the results of object detection for tracking purposes, enabling a deeper analysis of abnormal behaviors exhibited by diseased fish. (3) The multi-task optimization strategy used in this study has shown effectiveness, but there is room for further exploration. Future experiments will focus on more diverse multi-task optimization methods to further enhance the overall performance of the model.

CRediT authorship contribution statement

Xuefei Li: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Shili Zhao:** Conceptualization, Validation,

Formal analysis. **Chunlin Chen**: Validation, Formal analysis. **Hongwu Cui**: Resources. **Daoliang Li**: Supervision. **Ran Zhao**: Investigation, Resources, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Extensibility and generalization

A.1 Experimentation on tomato leaf diseases detect dataset

Our model performs excellently not only on fish disease datasets but also in detecting diseases on other datasets. we conducted tests on the tomato leaf diseases detect dataset ([Sylhet Agricultural University, 2024, Tomato Leaf Diseases Detect Dataset](#)) to better evaluate the model's performance. This dataset includes six disease labels along with a healthy label, namely bacterial spot, early blight, late blight, leaf mold, target spot, yellow leaf curl virus and healthy. As the dataset only contains YOLO-formatted data, the corresponding segmentation dataset was augmented accordingly, converting detection labels into pixel-wise labels with different colors representing different disease labels. Additionally, adjustments were made to detect entire leaves rather than localized lesions. The training weights for fish disease were used as pre-trained weights while other parameters remained consistent. The results are shown in [Fig. 14](#).

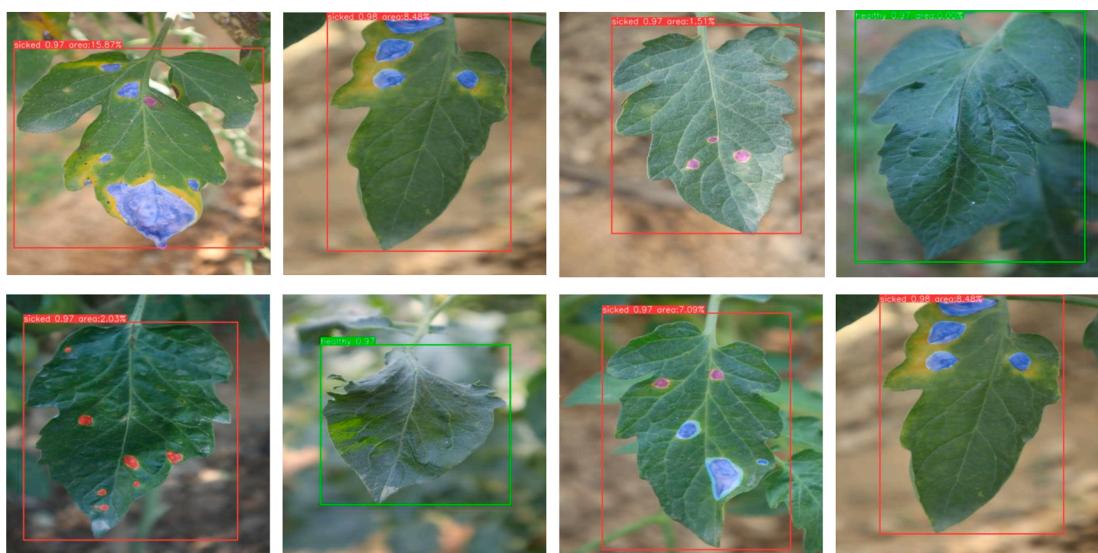


Fig. 14. Validation on the tomato leaf diseases detect dataset. For detection labels, there are only two types: “sicked” and “healthy”. For segmentation labels, there are six types: bacterial spot, early blight, late blight, leaf mold, target spot, and black spot, which are represented by the colors: magenta, blue, orange, purple, cyan, and orange-red, respectively.

A.2 VOC2012 segmentation results

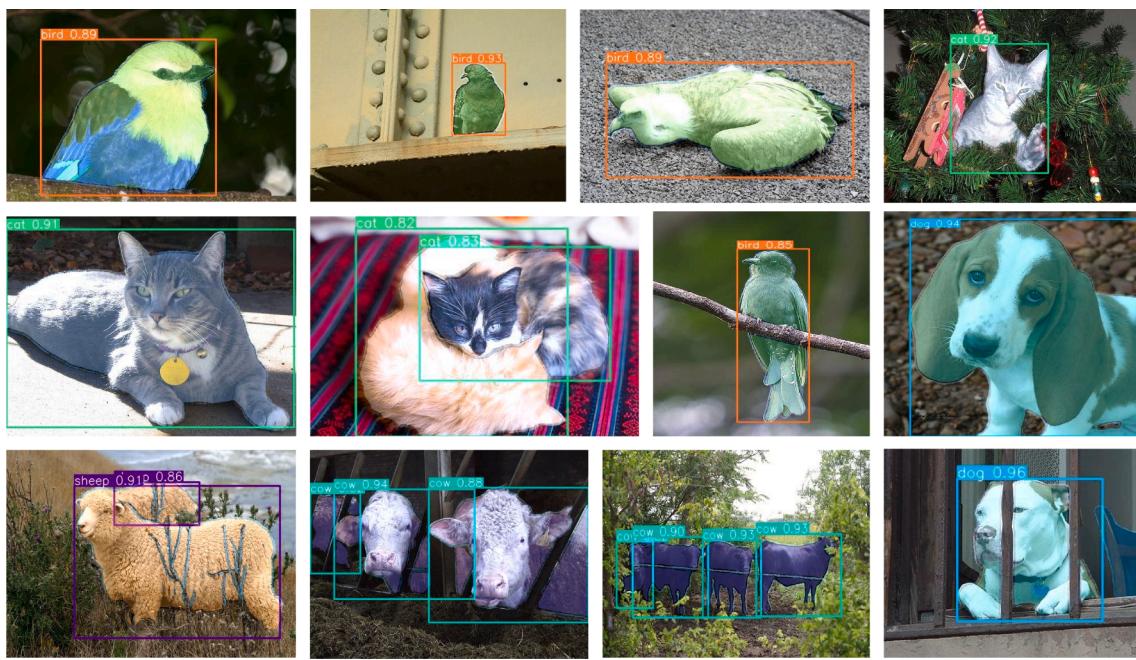
Similarly, in testing on the VOC2012 segmentation dataset, [yolov8s.pt](#) was utilized as the pre-trained weights while keeping other parameters consistent. Several images of animals were selected to showcase the results, as depicted in [Fig. 15](#).

Data availability

The authors do not have permission to share data.

Acknowledgments

This paper was supported by The National Natural Science Foundation of China (NO.32273188) and Key Research and Development Plan of the Ministry of Science and Technology (NO.2022YFD2001700).



VOC2012 segmentation dataset

Fig. 15. The detection and segmentation results on VOC2012 segmentation dataset. Although these are very similar to instance segmentations, but fundamentally different from instance segmentations.

Appendix B. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.eswa.2024.125085>.

References

- Chakravorty, H., Paul, R., & Das, P. (2015). Image processing technique to detect fish disease. *International Journal of Computer Science Security*, 9, 121–131.
- Chen, Z., Badrinarayanan, V., Lee, C.-Y., & Rabinovich, A. (2018). Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In *International conference on machine learning* (pp. 794–803). PMLR.
- Davidson, J., Good, C., Welsh, C., & Summerfelt, S. T. (2011). Abnormal swimming behavior and increased deformities in rainbow trout *Oncorhynchus mykiss* cultured in low exchange water recirculating aquaculture systems. *Aquacultural Engineering*, 45, 109–117.
- Hu, J., Zhao, D., Zhang, Y., Zhou, C., & Chen, W. (2021). Real-time nondestructive fish behavior detecting in mixed polyculture system using deep-learning and low-cost devices. *Expert Systems with Applications*, 178, Article 115051.
- Jocher, G. (2020). YOLOv5 by Ultralytics. url: <https://github.com/ultralytics/yolov5>.
- Jocher, G., Chaurasia, A., & Qiu, J. (2023). YOLO by Ultralytics. url: <https://github.com/ultralytics/ultralytics>.
- Kendall, A., Gal, Y., & Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7482–7491).
- Liao, Y., Xu, J., & Wang, W. (2011). A method of water quality assessment based on biomonitoring and multiclass support vector machine. *Procedia Environmental Sciences*, 10, 451–457.
- Liu, B., Liu, X., Jin, X., Stone, P., & Liu, Q. (2021). Conflict-averse gradient descent for multi-task learning. *Advances in Neural Information Processing Systems*, 34, 18878–18890.
- Liu, S., Johns, E., & Davison, A. J. (2019). End-to-end multi-task learning with attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1871–1880).
- Lyubchenko, V., Matarneh, R., Kobylin, O., & Lyashenko, V. (2016). Digital image processing techniques for detection and diagnosis of fish diseases. *International Journal of Advanced Research in Computer Science and Software Engineering*, 6(7), 79–83.
- Manan, H., Zhong, J. M. H., Kasan, N. A., Suratman, S., & Ikhwanuddin, M. J. A. (2019). Carbon dioxide flux from intensive aquaculture shrimp farming applying biofloc system of Setiu Terengganu. *Malaysia*, 509, 52–58.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618–626).
- Waleed, A., Medhat, H., Esmail, M., Osama, K., Samy, R., & Ghanim, T. M. (2019). Automatic recognition of fish diseases in fish farms. In *International Conference on Computer Engineering and Systems (ICCES)* (pp. 201–206). IEEE.
- Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2023). 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* (pp. 7464–7475).
- Wang, H., Zhang, S., Zhao, S., Lu, J., Wang, Y., Li, D., & Zhao, R. (2022). Fast detection of cannibalism behavior of juvenile fish based on deep learning. *Computers and Electronics in Agriculture*, 198, Article 107033.
- Wu, D., Liao, M.-W., Zhang, W.-T., Wang, X.-G., Bai, X., Cheng, W.-Q., & Liu, W.-Y. (2022). Yolop: You only look once for panoptic driving perception. *Machine Intelligence Research*, 19, 550–562.
- Yu, T., Kumar, S., Gupta, A., Levine, S., Hausman, K., & Finn, C. (2020). Gradient surgery for multi-task learning. *Advances in Neural Information Processing Systems*, 33, 5824–5836.
- Yu, X., Wang, Y., An, D., & Wei, Y. (2021). Identification methodology of special behaviors for fish school based on spatial behavior characteristics. *Computers and Electronics in Agriculture*, 185, Article 106169.
- Zhao, J., Bao, W., Zhang, F., Zhu, S., Liu, Y., Lu, H., Shen, M., & Ye, Z. (2018). Modified motion influence map and recurrent neural network-based monitoring of the local unusual behaviors for fish school in intensive aquaculture. *Aquaculture*, 493, 165–175.
- Zhao, S., Zhang, S., Lu, J., Wang, H., Feng, Y., Shi, C., Li, D., & Zhao, R. (2022). A lightweight dead fish detection method based on deformable convolution and YOLOv4. *Computers and Electronics in Agriculture*, 198.
- Sylhet Agricultural University. Tomato Leaf Diseases Detect Dataset. Roboflow Universe. <https://universe.roboflow.com/sylhet-agricultural-university/tomato-leaf-diseases-detect>. 2024.