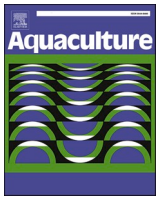
Aquaculture 582 (2024) 740558

Contents lists available at ScienceDirect

Aquaculture

journal homepage: www.elsevier.com/locate/aquaculture



Rapid detection of fish with SVC symptoms based on machine vision combined with a NAM-YOLO v7 hybrid model

Yaoyi Cai [a], Zekai Yao [a], Haibo Jiang [a], Wei Qin [b], Jun Xiao [b], Xiuxiang Huang [a], Jiaji Pan [a,b,*], Hao Feng [b,*]

[a] College of Engineering and Design, Hunan Normal University, Changsha, Hunan 410083, PR China
[b] State Key Laboratory of Developmental Biology of Freshwater Fish, College of Life Science, Hunan Normal University, Changsha, Hunan 410083, PR China

ARTICLE INFO

Keywords:
SVC detection
Convolutional neural network
Machine vision
NAM-YOLO v7
Artificial intelligence

ABSTRACT

Surveillance of fish disease is essential to contain the spreading epidemics. Current surveillance mostly relies on molecular biology testing methods, which often requires complex procedures and trained operators. These techniques may hardly to be used in many small aquaculture farms. In this paper, a method based on machine vision combined with a target detection network is proposed. A convolutional neural network (CNN) was developed to detect fish infected with SVCV by analyzing the images. The datasets for the CNN that is implemented from a YOLO v7 deep learning algorithm are information extracted from images containing fish populations. An Auto-MSRCR algorithm was used to adaptively enhance the images to minimize manual intervention. We introduced a novel NAM Attention mechanism integrated with the ELAN module in the original YOLO v7 deep learning algorithm. Both channel attention and spatial attention modules were utilized to suppress the insignificant features in the datasets to achieve a more precise and efficient detection method. Also, a loss function MPDIoU was incorporated to avoid missing detection by this vision-based methodology. After training of the model, the detection testing results show that the NAM-YOLO v7 network achieves over 95% prediction accuracy and 93.8% recall, which is superior than other state-of-art YOLO series models. Also, the time for detection for each image only takes 0.18 s illustrating the integrated modules improved the computation efficiency. This novel technique is of great potential to be applied in small farms for rapid and early detection. It can be a vital supplementary tool in developing more effective surveillance strategy by combined with other established methods.

1. Introduction

Spring viremia of carp (SVC) is an acute, serious infectious disease that threatens cyprinids and some non-cyprinids fishes (Ahne et al., 2002). The direct pathogen of SVC is a bullet shaped RNA virus (spring viremia of carp virus, SVCV). SVCV can infect many host species including common carp (*Cyprinus carpio*), grass carp (*Ctenopharyngodon idella*), koi (*Cyprinus carpio* koi), silver carp (*Hypophthalmichthys molitrix*), goldfish (*Carassius auratus*), rainbow trout (*Oncorhynchus mykiss*) and many others (Dixon and Longshaw, 2005; Embregts et al., 2017; Emmenegger et al., 2016; Goodwin, 2009). The SVC outbreaks usually occur in the spring when the water temperatures are 11–17 °C, with very high morbidity and mortality (Shao and Zhao, 2017). Outbreaks of SVC have been reported in Western Europe, North America and East Asian countries, resulting in significant losses to aquaculture (Padhi and Verghese, 2008). It is listed as a notifiable disease by the OIE. Since there are no effective vaccines or drugs to control the spread of SVC currently, the feasible approach to contain SVC epidemics is early detection and diagnosis (Fouad et al., 2019). As a result, many countries have regular surveillance strategies in place to prevent outbreaks.

There are several molecular biology methods for diagnosing SVCV. Polymerase chain reaction (PCR) or RT-PCR methods target a specific glycoprotein gene or nucleoprotein gene that can detect many SVCV subtypes (Shimahara et al., 2016). RT-qPCR provides quantitative results and is more sensitive than PCR (Shao et al., 2016) and is widely used in pathogen testing. Loop-mediated isothermal amplification technology (LAMP) (Shivappa et al., 2008) is a thermostatic nucleic acid amplification technology that designs primers specific to the regions of

* Corresponding authors at: State Key Laboratory of Developmental Biology of Freshwater Fish, College of Life Science, Hunan Normal University, Changsha, Hunan 410083, PR China; College of Engineering and Design, Hunan Normal University, Changsha, Hunan 410083, PR China.
E-mail addresses: pan.jiaji@hunnu.edu.cn (J. Pan), fenghao@hunnu.edu.cn (H. Feng).

https://doi.org/10.1016/j.aquaculture.2024.740558
Received 12 September 2023; Received in revised form 2 January 2024; Accepted 5 January 2024
Available online 6 January 2024

the target gene and uses strand-substitution DNA polymerase to amplify under thermostatic conditions. It is easier to use and suitable for rapid detection applications. In addition to these molecular biology techniques, histological analysis can be used to identify the pathological changes in organs and tissues of SVCV-infected fish (Pan et al., 2023). However, the extraction of bio-specimen for diagnosis using these techniques relies on professional operators and delicate equipment. The diagnosis procedures are normally labor-intensive and time-consuming. Qualified personnel and instruments are insufficient in many small aquaculture farms. Effective preventive real-time detection technologies are in urgent need for these ponds or aquaculture farms. In fact, many SVCV infected fish have symptoms such as hemorrhage, slow swimming, and imbalance movement. The symptoms of the infected fish could possibly be utilized for rapid identification and detection of SVC among fish populations.

In addition to the molecular biology techniques, machine vision technologies are rapidly advancing in aquaculture due to its fast and accurate advantages. The automatic fish classification was realized with 3D reconstructions established from texture and shape features extracted from collected images (Spampinato et al., 2010). Thus, the unusual fish behaviors can be detected by clustering the fish trajectories. Huang et al. extracted 66 types of features consisting of color, shape and texture features of the fish images(Huang et al., 2012). The fish recognition or classification was improved by a balanced-guaranteed optimization tree method. The optical flow, entropy and statistical methods were combined to detect the motion of fish. The dispersion behaviors of the fish groups can be monitored(Zhao et al., 2016). The abnormal behaviors of fish schools can also be monitored using the features in the images by Harris angle detection and Lucas-Kanade optical flow after segmentation (Yu et al., 2021). However, these detection techniques identify abnormal fish behaviour by manually extracting features based on collected data, leading to major limitations of their models in practical applications. The small target in the water (Zhu et al., 2023a, 2023b), blurring of the images (Oreifej et al., 2011) are also very critical problems compromising the detection precision, asking for more sophisticated approaches.

In recent years, deep learning models have been widely used for adaptive discrimination of fish status and target detection of abnormal individuals during freshwater fish farming. A local abnormal behaviour detection method was proposed based on corrected kinematic image maps and recurrent neural networks with high accuracy performance (Zhao et al., 2018). Convolutional neural networks (CNN) and machine vision were adopted to detect and evaluate fish appetites (Zhou et al., 2019). This approach does not require manual design of the feature-assisted model for identification so as to reduce great labor efforts. A method using CNNs and spatio-temporal image fusion has been proposed. It is applied to efficiently identify and classify different behaviour states of fish populations. Therefore, a better fish feeding identification is achieved by fish behaviour recognition (Han et al., 2020). Deep learning models have also been combined with graph theory to aid accurate breeding selection by classifying and recognizing key parts of target fish and then matching them to behaviour templates (Wang et al., 2020). However, these studies have focused on discriminating fish behaviour and have not been able to achieve precise location and tracking of abnormal fish individuals.

Multi-class targets can be quickly, accurately and adaptively discriminated and tracked by deep-learning target detection models. Different individuals in a population are able to achieve their category determination and precise localization based on these target detection models. As a result, target detection models based on deep learning techniques have also begun to be applied to achieve accurate identification of fish individuals in the aquaculture process. A method that combines YOLOv3 and MobileNetv1 networks to achieve real-time fish tracking and population monitoring by processing fish images acquired in real freshwater fish farms was proposed (Cai et al., 2020). The improved YOLOv3-Lite and MobileNetV2 were integrated to build a low-cost, lightweight network to achieve automatic detection of fish behaviour (Hu et al., 2021). In addition, an underwater camera unit combined with the YOLO v5m model was attempted to be used for the precise location and identification of diseased fish individuals during aquaculture (Wang et al., 2023a, 2023b). Other target detection models have the potential to enable the detection of abnormal fish, such as YOLO v5s, YOLO v6, YOLOR and YOLO v7 (Pan and Li, 2022; Bist et al., 2023; Song et al., 2023; Wang et al., 2023a, 2023b; Yu et al., 2023). However, most of these methods are based on underwater cameras to acquire images of fish while they are in motion, thus enabling the detection and tracking of anomalous individuals based on depth target detection networks.These methods do not achieve a balance between accuracy and model size, and fail to meet the need for rapid detection while maintaining high accuracy. The defects of narrow capture field, low recognition accuracy and high cost of image acquisition equipment make it difficult to be applied in the freshwater fish farms.

In this study, we adopt deep learning technology in SVC detection based on the improved YOLO v7 algorithm. A CNN model called NAM-YOLO-v7 was established. For accurate identification and detection of fish with SVC symptoms in freshwater fish farming, a method combining machine vision techniques with the NAM-YOLO v7 hybrid model is proposed. First, we created datasets for training the model by obtaining images of fish populations with some SVCV-infected fish in the experimental environment. Second, the collected images are enhanced using the automatic multi-scale color reduction Retinex (Auto-MSRCR) algorithm. Third, a new target detection network NAM-YOLO v7 is designed and applied for the exact localization of fish with SVC symptoms in images. It combines a Normalization-based Attention Module (NAM) with the YOLO v7 network (Liu et al., 2021). Based on the proposed NAM-YOLO v7 network, it is possible to achieve more precise localization of SVCV-infected fish and increase their detection rate. Our suggested technology detects fish with SVC symptoms during freshwater fish farming in a low-cost and rapid approach, which may serve as an supplementary and alternative strategy in the future virus surveillance programs.

2. Materials and methods

2.1. Fish and virus

Adult wild-type zebrafish were obtained from the Chinese Zebrafish Resource Center and raised in accordance with the protocol (Zhou, 2024). The ages of the fish were between 3 and 4 months. Adult zebrafish (length ~ 30 mm) collected from the breeding house were used for this experiment. SVCV kept in our lab was cultivated in epithelioma papulosum cyprini (EPC) cells at 26 °C with 2% fetal bovine serum. Zebrafish were infected with SVCV by intraperitoneally injection. After 1–2 days of SVCV injection, these zebrafish demonstrated symptoms including abdominal and gill hemorrhages, slow and imbalanced swimming with a bloated abdomen. The new generations of SVCV were cultivated and extracted from the SVCV infected fish tissues. The new generations of SVCV were cultivated and extracted from the SVCV infected fish tissues. We determined the SVCV genome copy numbers per unit volume of viral fluid. The process included the following steps: constructing plasmid standards, grinding SVCV-infected fish, centrifuging the supernatant viral liquid, purifying the viral liquid to extract RNA, and reverse transcription to obtain cDNA. Finally, absolute fluorescence quantification was used to obtain the copy number of SVCV genome in the viral liquid as 0.74–1.39 $\times$ 10^6 copy/uL.

2.2. Overview of the full methodological process

The overall flow of the proposed method for detecting fish with SVC symptoms is shown in Fig. 1. A high-resolution camera is fixed above the aquarium tank, capturing real-time images of the fish as they navigate through the water. The real-time captured images are fed into a

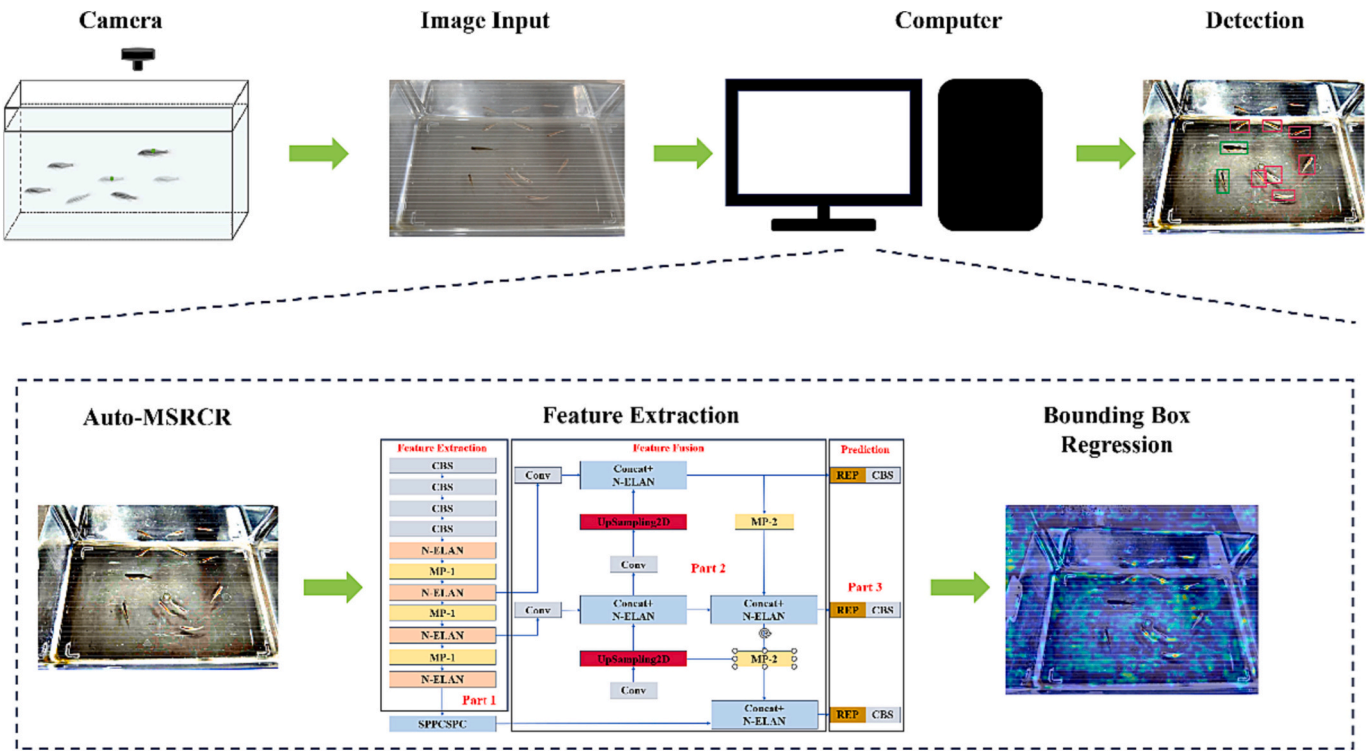


Fig. 1. Overview of the detection procedure using the machine vision technology combined with a NAM-YOLO v7 hybrid model. The real-time captured images are fed into a computer to implement image enhancement (Auto-MSRCR), feature extraction, category determination and region regression to obtain the precise location of the SVCV-infected fish.

computer to implement image enhancement, feature extraction, category determination and region regression to obtain the precise location of the SVCV-infected fish. In this scenario, the real-time captured images are augmented using the Auto-MSRCR method to improve the detection saliency of the fish targets. Fish with SVC symptoms will be distinguished from normal fish based on the proposed NAM-YOLO v7 network using rectangular boxes to achieve localization and tagging. Thereafter, the amount and position of fish with SVC symptoms can be determined.

2.3. Data acquisition

Zebrafish were challenged with 10 μl SVCV suspension as the detection targets. The healthy zebrafish were placed in an aquarium tank with introducing SVCV challenged fish. Videos of the zebrafish were acquired by a 48-megapixel camera fixed on top of the tank in 1920 × 1080 resolution at a 60 fps frame rate. Videos of fish were recorded and saved on a computer for three days in a row. The image frames were extracted from the videos as datasets for the SVCV infected fish detection analysis by OpenCV (a software to implement this machine vision technology). 1814 images were taken out of the recorded video at predetermined intervals, and a detection dataset is obtained. These images were used for the following analysis. All fish with SVC symptoms in the images were identified and localized. SVCV-infected and healthy fish on the images were labelled separately using the LaelImg software. Finally, the created dataset contains 1814 screened images and coordinate information for different categories of fish.

2.4. Image enhancement method based on auto-MSRCR

In order to improve the prominence and contrast of the fish with SVC symptoms in the images, the collected images were enhanced by the Auto-MSRCR algorithm. The algorithm is based on the assumption that the relationship between the acquired image and the real image is shown in Eq. 1.

$$S(x,y) = R(x,y)L(x,y) \tag{1}$$

where x,y are the horizontal and vertical coordinates of the pixel point respectively; $S(x,y)$ denotes the value of the pixel point of the image captured by the camera at coordinates (x,y); $R(x,y)$ denotes the value of the reflected light intensity of the object at coordinates (x,y) and $L(x,y)$ represents the ambient light irradiation component of the scene at coordinates (x,y). The Auto-MSRCR algorithm can then be used to get the pixel values of the restored real image at coordinates (x,y) as shown in Eq. 2:

$$R_{Auto-MSRCR}(x,y) = C(x,y)\sum_{n=1}^{N}\omega_n\{InS(x,y) - In[S(x,y)*G_n(x,y)]\} \tag{2}$$

where $C(x,y)$ represents the recovery coefficient of the pixel at coordinates (x,y) on the image. N represents the number of channels, which usually takes the value of 3 for RGB images. ω_n is the weighting factor for the *nth* scale. $G_n(x,y)$ is the Gaussian filter function on the *nth* scale at coordinates (x,y). In Eq. 2, the Gaussian filter function $G(x,y)$ is defined as follows:

$$G(x,y) = kexp\left(-\frac{x^2+y^2}{\sigma^2}\right) \tag{3}$$

where σ is the scaling factor in the Gaussian function and k is the normalization factor that ensuring $\iint G(x,y)dxdy = 1$. Meanwhile, the formula for the coefficient $C(x,y)$ is shown as follows:

$$C(x,y) = \beta\left\{log\left[\alpha S(x,y) - log\left[\sum_{n=1}^{N} S_n(x,y)\right]\right]\right\} \tag{4}$$

where α and β are individually customised constants indicating the degree of gain and the strength of the controlled nonlinearity, respectively. Thus, based on the Auto-MSRCR algorithm, the captured images of fish

can be enhanced and the effects of scattered light and uneven illumination on the water surface can be eliminated as much as possible.

2.5. Development of the proposed NAM-YOLO v7 network

We propose a novel NAM-YOLO v7 network with a normalization-based attention mechanism for localization and tracking of SVCV-infected fish. A graphical illustration of the developed network is shown in Fig. 2. The enhanced images based on the Auto-MSRCR algorithm were fed into the NAM-YOLO v7 network to achieve accurate discrimination and localization of fish with SVC symptoms. In our proposed NAM-YOLO v7 network, the NAM attention module is introduced thereby enhancing the detection accuracy. At the same time, the original ELAN structure in the network is optimized thereby increasing its detection speed.

2.5.1. Structure of the proposed NAM-YOLO v7 network

The proposed NAM-YOLO v7 network model consists of a feature extraction layer, a feature fusion layer and a prediction layer, as shown in Fig. 2. As shown in Part 1 of Fig. 2, the feature extraction layer consists of four general convolution (CBS) modules, three general convolution-maximum pooling (MP) modules, and four efficient layer aggregation modules (ELAN) fused with NAM Attention. In the general convolution module, 2D convolution, batch specification and SiLU activation functions are stacked for reducing the size of the image feature map and extracting features. This series of convolutional operations is referred to as CBS. The efficient layer aggregation network incorporating NAM Attention (N-ELAN) module consists of the ELAN structure combined with the NAM Attention mechanism. The ELAN module is an efficient network structure that enables the network to learn more features and be more robust by controlling the shortest and longest gradient paths. The MP module consists of the maximum pooling layer and the CBS module. In the MP module, the CBS module is used to adjust the number of channels as well as to implement feature fusion. MP modules can be divided into two different types, MP-1 and MP-2, depending on the number of branches. The MP-1 module consists of 2 branches and the MP-2 module consists of 3 branches.

As shown in Part 2 of Fig. 2, the feature fusion layer consists of one spatial pyramid pooling cross stage partial (SPPCSPC) module, four N-ELAN modules, two 2D upsampling modules, four 2D convolution modules, and two MP modules. The SPPCSPC module is a Cross Stage Partial Network with a Spatial Pyramid Pooling (SPP) block that obtains multiscale area features by employing the max pool and concatenation procedures to overcome the issue of duplicating gradient information. Three different sizes of maximum pooling layers as well as shortcut branches are included in the SPPCSPC module. Based on this structure, four receptive fields exist in the network and are used to match targets of different sizes to be detected, thus better enabling the detection of targets of different sizes at different resolutions. In addition, the number of channels is expanded to twice the number of input images in the feature fusion layer, thus reducing feature loss.

As shown in Part 3 of Fig. 2, the final prediction layer is responsible for giving the location information of the target to be detected as well as the category information. The original YOLO v7 network suffers from the defect that the optimization process of the loss function is constrained by the aspect ratio of the target to be detected, which leads to a decrease in its detection accuracy. Therefore, in our proposed NAM-YOLO v7 network, the MPDIoU loss function is used instead of the CIoU loss function to improve the simultaneous detection accuracy for targets with different aspect ratios. Finally, the prediction of the exact position of the target to be detected is achieved by filtering the target localization region using MPDIoU border loss calculation and non-maximum suppression (NMS) methods. Based on the improved loss calculation function, fish with SVC symptoms of different body morphologies and sizes can be successfully identified and localized by the model, thus reducing the interference of different morphologies in their swimming process on the detection results.

2.5.2. The structure of the N-ELAN module

As shown in Fig. 3, the N-ELAN module is one of the core innovations to better enable the localization and tracking of fish with SVC symptoms in different morphologies during movement. The designed N-ELAN

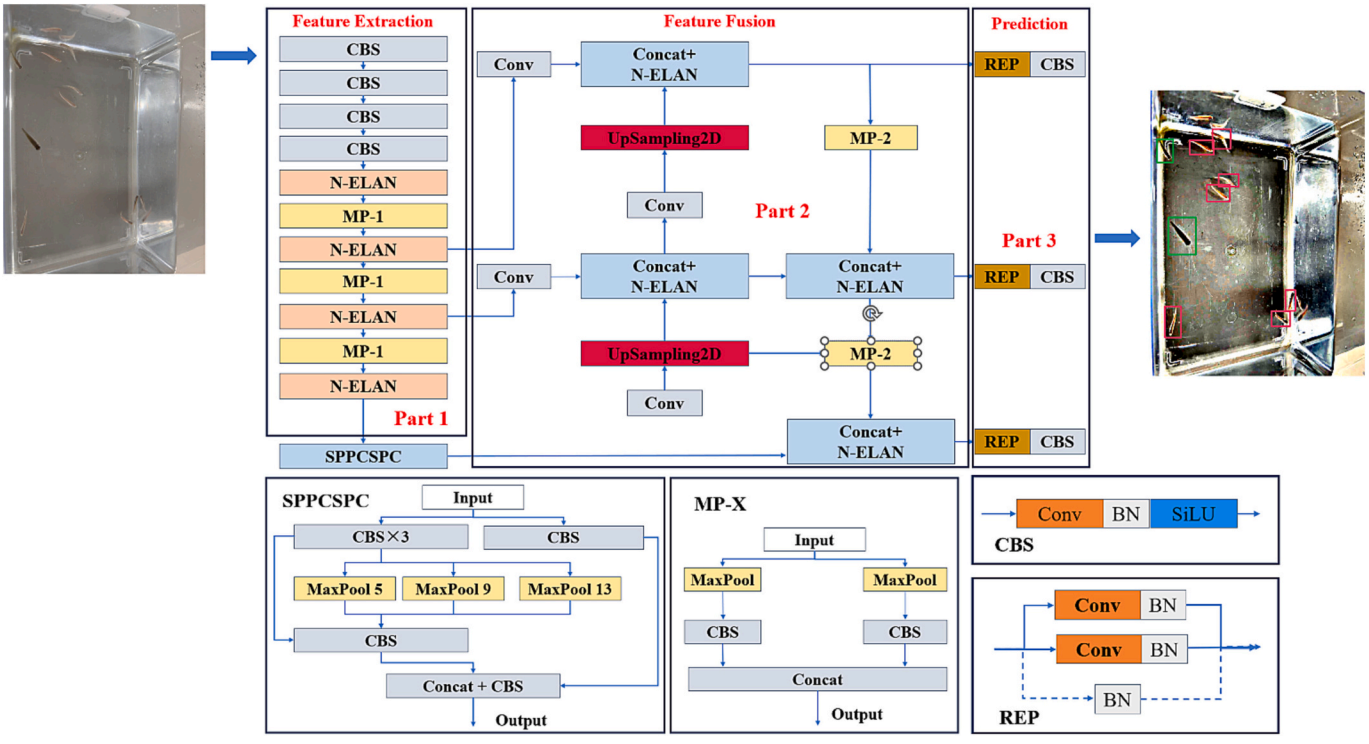


Fig. 2. The specific structure of the NAM-YOLO v7 CNN. (a) Part 1: feature extraction layer; (b) Part 2: feature fusion layer; (c) Part 3: prediction output layer.

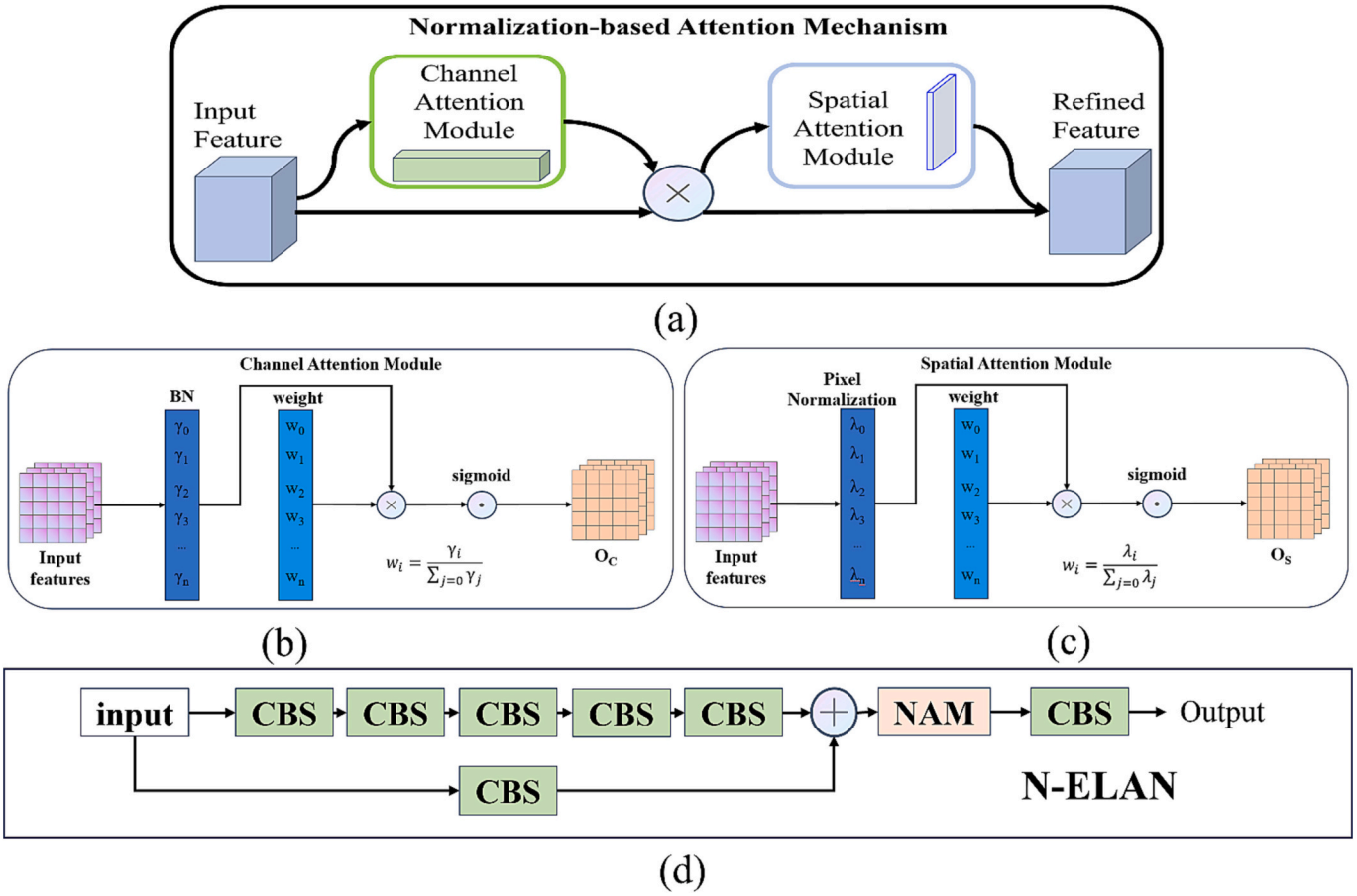


Fig. 3. Illustration of the N-ELAN structure. (a) Overall structure of the NAM module; (b) channel attention submodule; (c) spatial attention submodule (d) integration of N-ELAN.

module achieves feature extraction and enhancement of targets at different scales by introducing the NAM. Mutual information between features of different dimensions can be extracted based on the adopted N-ELAN module while suppressing unimportant features. As a result, more and deeper image feature information can be captured by the NAM-YOLO v7 network embedded in the N-ELAM module, thus enabling the localization and capture of fish with SVC symptoms in different postures during swimming.

NAM is a redesigned attention mechanism based on a convolutional block attention module (CBAM). The overall structure of the NAM is shown in Part a of Fig. 3. As shown in Part a of Fig. 3, the channel attention module and the spatial attention module are two submodules of the NAM-based feature augmentation module. The specific structure of the channel attention mechanism submodule is shown in Part b of Fig. 3. In the channel attention sub-module, the batch normalization (BN) layer is used to implement a normalized redistribution of input features, as shown in Eq. 5:

$$B_{out2} = BN(B_{in}) = \gamma \frac{B_{in} - \mu_{B_c}}{\sqrt{\sigma_{B_c}^2 + \nu}} + \beta_c \tag{5}$$

where μ_{B_c} and σ_{B_c} are the mean and standard deviation of the input features, respectively. γ and β_c are affine transformation parameters, representing scale and displacement, respectively. B_{in} and B_{out1} represent the input eigenvalues and the output eigenvalues after channel enhancement respectively. ν represents the residual value. After the features extracted from the image are fed into the attention submodule, the output features are shown in Eq. 6:

$$M_C = sigmoid\left(W_\gamma B_{out1}\right) \tag{6}$$

where $W_\gamma = \gamma_i / \sum_{j=0}\gamma_i$ is the output weight of the network.

The specific structure of the spatial attention mechanism submodule is shown in Part c of Fig. 3. In the spatial attention submodule, the pixel normalization layer is used to achieve a balance of weights between different pixel points and for the enhancement of different features. Thus, the output of the image extracted features after this layer is shown in Eq. 7:

$$B_{out2} = BN_S(B_{in}) = \lambda \frac{B_{in} - \mu_{B_s}}{\sqrt{\sigma_{B_s}^2 + \varepsilon}} + \beta_s \tag{7}$$

where μ_{B_s} and σ_{B_s} are the mean and standard deviation of all pixels of the input respectively. λ and β_s are affine transformation parameters in the pixel normalization layer, representing scale and displacement, respectively. B_{in} and B_{out2} represent the input eigenvalues and the output eigenvalues after channel enhancement respectively. ε represents the residual value. Finally, the output of the extracted feature maps in the image based on the spatial attention submodule is shown in Eq. 8:

$$M_s = sigmoid\left(W_\gamma B_{out2}\right) \tag{8}$$

where $W_\lambda = \lambda_i / \sum_{j=0}\lambda_i$ is the weight value of the output pixel.

Finally, NAM is combined with the ELAN network in YOLO v7 and forms the improved N-ELAN network, as shown in Part d of Fig. 3. The NAM module is inserted into the original ELAN network to achieve the weighting of the feature maps extracted from the image in both channel and pixel dimensions. Based on the weighted features, the model can

better achieve the position discrimination of the fish with SVC symptoms during the movement and improve its tracking accuracy.

2.5.3. Calculation of the loss function

The loss function was used to calculate the gap between the predicted area and the real area of the fish with SVC symptoms based on the proposed NAM-YOLO v7 network implementation. During the network training process, calculations based on the loss function can bring the output as near to the true value as feasible, leading to more accurate predictions. However, the loss function in the original YOLO v7 network is affected by the aspect ratio of the detection frame during the computation process, causing difficulties in its optimization. Due to the large differences between the morphology and position of different individuals during the swimming process of the fish, it is difficult for the existing YOLO v7 network to achieve precise positioning, which affects the accuracy of tracking. At the same time, there may be a certain degree of overlapping of bit positions between fish in different postures, causing them to be difficult to detect well by the existing YOLO v7 network (Ma and Xu, 2023). Therefore, the loss function CIoU used in the original YOLO v7 network is replaced by the new loss function MPDIoU based on the minimum point distance.

The computational procedure of MPDIoU is as follows: for detection frames A and B, assume that (x_1^A, y_1^A) and (x_2^A, y_2^A) denote the upper-left and lower-right corner point coordinates of A, respectively. (x_1^B, y_1^B) and (x_2^B, y_2^B) denote the upper-left and lower-right corner point coordinates of B, respectively. The parameters w and h are denoted as the width and height of the input image. The square of the distance between the prediction box and the gound truth box at upper-left corner point is calculated as follows: $d_1^2 = (x_1^B - x_1^A)^2 + (y_1^B - y_1^A)^2$ and the square of the distance between the prediction box and the gound truth box at the upper-left corner point is calculated as follows: $d_2^2 = (x_2^B - x_2^A)^2 + (y_2^B - y_2^A)^2$.

Finally, the MPDIoU is obtained by

$$MPDIoU = \frac{A \cap B}{A \cup B} - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \tag{9}$$

To make each bounding box $B_{prd} = \left[x^{prd}, y^{prd}, w^{prd}, h^{prd}\right]^T$ predicted by the model converge to its corresponding true labelled bounding box $B_{gt} = \left[x^{gt}, y^{gt}, w^{gt}, h^{gt}\right]^T$, the minimal loss function is established as follows:

$$\mathscr{L} = \min_{\Theta} \sum_{B_{gt} \in B_{gt}} \mathscr{L}\left(B_{gt}, B_{pred} | \Theta\right) \tag{10}$$

where B_{gt} denotes the set of real labelled bounding boxes. The parameter Θ represents the regression parameter of the deep model. The loss function is defined as $L_{MPDIoU} = 1 - MPDIoU$. All the parameters of the loss function of the regression of the bounding box can be determined by four coordinates as follows:

$$|C| = \left(max\left(x_2^{gt}, x_2^{prd}\right)\right) - min\left(x_1^{gt}, x_2^{prd}\right) \times \left(max\left(y_2^{gt}, y_2^{prd}\right)\right) - min\left(y_1^{gt}, y_1^{prd}\right) \tag{11}$$

where $|C|$ denotes the area of the smallest outer rectangle covering B_{gt} and B_{prd}.

The center coordinates of the real label bounding box are calculated as $x_c^{gt} = \left(x_1^{gt} + x_2^{gt}\right)/2$ and $y_c^{gt} = \left(y_1^{gt} + y_2^{gt}\right)/2$. The center coordinates of the predicted label bounding box are calculated as $x_c^{prd} = \left(x_1^{prd} + x_2^{prd}\right)/2$ and $y_c^{prd} = \left(y_1^{prd} + y_2^{prd}\right)/2$, where w_{gt} and h_{gt} represent the width and height of the real label bounding box. w_{prd} and h_{prd} represent the width and height of the predicted label bounding box. In addition, w_{prd} and h_{prd} are calculated as $w_{prd} = x_2^{prd} - x_1^{prd}$ and $h_{prd} = y_2^{prd} - y_1^{prd}$, respectively. The coordinates of the upper left and lower right corner points can be used to determine all of the factors taken into account in the existing loss functions. These factors include the non-overlapping area, the distance from the center point, and the deviation of the width and height, demonstrating how the algorithm could streamline the calculation process and take many factors into account.

3. Results

3.1. Training configuration of the NAM-YOLO v7 model

The SVCV infection experiment and protocols were inspected and approved by the Ethics Committee of Hunan Normal University (No. 2023–047). From Section 2.3, a total of 1814 images were used for model training and testing of fish with SVC symptoms. All the acquired images are divided into a training set and a test set. Seven healthy fish and two fish injected with SVCV were included in each image.

The key parameters and environment settings during training of the NAM-YOLO v7 network are shown in Table 1. The training set of the model consists of 1450 images and the testing set consists of 364 images. The images in the training set were fed into the NAM-YOLO v7 model for training after mosaic data enhancement (Zhou et al., 2023). In the training of the NAM-YOLO v7 model, the batch size is set to 16. The maximum number of iterations was set to 200. The loss function is MPDIoU combined with BCEWithLogits. The resolution of the input image for the NAM-YOLO v7 network is 1080 × 1080. The values of IOU_thres and Conf_thres are set to 0.65 and 0.001, respectively. Two NVIDIA GTX 3090 GPUs were used in the training process, under the environment of CUDA 11.0 and cuDNN 8.0.3. The code implementation of our proposed method was realized by Python 3.8.

3.2. Enhanced image based on Auto-MSRCR

The captured images of the fish while swimming are enhanced based on the automatic MSRCR algorithm, which is shown in Fig. 4. As shown in Fig. 4, the visibility of different individuals in the fish school in the image is improved. In aquatic environments, changes in light conditions may result in a decrease in the contrast between a fish target exhibiting SVC symptoms and its surroundings. Based on the automatic adjustment of the enhancement factor in the Auto-MSRCR algorithm, the acquired images are better adapted to different lighting conditions. The enhanced contrast of the images reduce the effects of illumination. In addition, fish with SVC symptoms can be further enhanced in terms of color significance based on the Auto-MSRCR algorithm, thus improving the accuracy of their detection.

In particular, when the images show aggregations of healthy and virus-infected individuals in fish populations, their similar colors make it difficult to distinguish between the two types of targets. The different categories of targets can be strengthened based on the hierarchical reinforcement automatically achieved by the adopted Auto-MSRCR algorithm, thus improving the precision detection of SVC-infected fish in the fish population.

Table 1
The hyperparameters and the training details.

Implementation details	Parameter name	Selected value
Training	Batch size	16
	Max-epochs	200
	Loss_function	MPDIoU + BCEWithLogits
	Input_size	1080 × 1080
	Label_smoothing	True
	IOU_thres	0.65
	Conf_thres	0.001
Environment	GPU	2 Nvidia RTX 3090 GPUs
	Platform	Python 3.8
	Implementation tools	PyTorch 1.7.1, CUDN 11.0, cuDNN 8.0.3

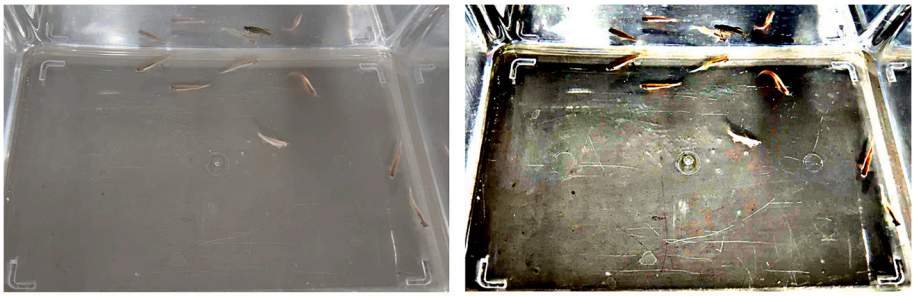


Fig. 4. Enhanced image based on the Auto-MSRCR algorithm: (a) original image; (b) image after data enhancement.

3.3. Experimental and prediction results

3.3.1. Evaluation indicator

The accuracy of the detection for fish infected with SVC symptoms will be evaluated based on several indicators, such as the precision (P), recall (R), F1 score, and mean average precision (mAP). Precision is the probability of being identified as a diseased fish in the diseased samples. It represents the accuracy of the prediction result of a positive sample. Recall is the probability of being determined as a diseased fish by the model among all diseased fish. It represents the overall prediction precision. The F1 score is introduced to achieve a balance between accuracy and recall. AP stands for the average precision for normal fish or fish with SVC symptoms. mAP represents the mean of the AP values of normal fish and fish with SVC symptoms by integration. The precision, recall, F1 score and mAP are calculated as follows:

$$precision = \frac{TP}{TP + FP} \tag{12}$$

$$recall = \frac{TP}{TP + FN} \tag{13}$$

$$F1\_Score = \frac{2 \times precision \times recall}{precision + recall} \tag{14}$$

$$mAP = \frac{1}{N}\sum_{i=0}^{N-1}\left(\int_0^1 P_i(R)dR\right) \tag{15}$$

where true positive (TP) means that the sample is classified as positive and correctly categorized. The false negative (FN) means that the sample is classified as negative but incorrectly categorized. The false positive (FP) means that the healthy fish sample is classified as positive. $P_i(R)$ is the precision value at which the target of class i is recognized. N denotes the number of categories of targets to be detected.

3.3.2. Experimental results

Due to the large amount of data needed for the training of the detection model, a 5-fold cross-validation approach is used to further enable the full utilization of the training set and the selection of the model with the best performance. In 5-fold cross-validation, 1450 images in the training set are divided into five smaller image sets, where each smaller dataset contains 290 independent images. Subsequently, the detection model for fish with SVC symptoms was trained using a combination of four smaller datasets as a training dataset. Meanwhile, the remaining one smaller dataset will be tested after the network completes every ten training epochs to obtain its prediction accuracy and loss function value. Based on this cross-validation approach, a detection network with optimal performance was obtained by setting a certain range of hyperparameter values. The remaining 364 images were used as an independent test set for the trained model. The set of optimal key hyperparameter values was obtained by setting a certain range of hyperparameter values and sampling tests based on a fixed step size. To prevent overfitting of the training data, early termination of model training is allowed if the mAP values on the validation dataset do not improve. After 10 rounds of elapsed time, if the mAP on the validation set is not improved or even affected, the training is terminated early.

The evaluation metrics of the proposed NAM-YOLO v7 network after training based on the training sets are shown in Fig. 5. The value of the F1-Score is between 0 and 1, with closer to 1 indicating a better balance of precision and recall in the model. A smaller loss value means that the model is more accurate in localizing SVCV-infected fish. The plot of precision and recall shows the model performance under different precisions and recalls, which are used to filter the optimal model. As shown in Fig. 5, we obtained the optimal target detection model after 200 iterations based on a 5-fold cross-validation process. The detection accuracy of fish with SVC symptoms was 97.3%, the recall was 93.8% and its mAP was 94.6% based on the optimal network obtained from training. According to the above results, our proposed model achieves high detection accuracy, while its detection results can cover all the fish with SVC symptoms as much as possible.

To further validate the performance of the proposed NAM-YOLO v7 model, we compare it with other target detection models based on an independent validation set, including YOLOv5s, YOLOv6, YOLOX, YOLOR and YOLO v7. These models used for comparison were all based on the same training set using a 5-fold cross-validation method to complete the training. These models are tested for their performance based on an independent test set, using precision, recall and detection time as evaluation metrics. The results of the performance evaluation of the obtained models are shown in Table 2. As shown in Table 2, our proposed NAM-YOLO v7 network applied to detect SVCV-infected fish in the images acquired by the fish population achieved the best performance with 97.3% prediction accuracy and 93.8% recall. The time to complete the detection of an image based on the NAM-YOLO v7 model is only higher than that of the YOLO v6 network, but its performance is much better. The above results show that the proposed NAM-YOLO v7 network applied to the detection of fish with SVC symptoms in the fish population has optimal performance and good detection speed.

An ablation experiment was performed to fully demonstrate the usefulness of our proposed enhancement strategy in the NAM-YOLO v7 network (Wang et al., 2019). The ablation experiment is a set of experiments in which components of the NAM-YOLO v7 are removed to measure the impact of these components on the performance of the network (Zhu et al., 2023a, 2023b). The results of the ablation experiments are shown in the last three rows of Table 2. As shown in Table 2, after adding the NAM module, the accuracy increased by 1.7%, the recall increased slightly by 0.3%, and the detection time per each image decreased. The initial YOLO v7 network improves its accuracy by 1.7% and its recall by 0.3% after the addition of the NAM module. This is mainly because the original YOLO v7 network has been enhanced by the addition of the NAM module to better extract the deep image features of the SVC-infected fish for its discrimination. After replacing the original loss function with MPDIoU in the network, the accuracy is increased by 1.1% and the recall is increased by 0.7%. This result suggests that the

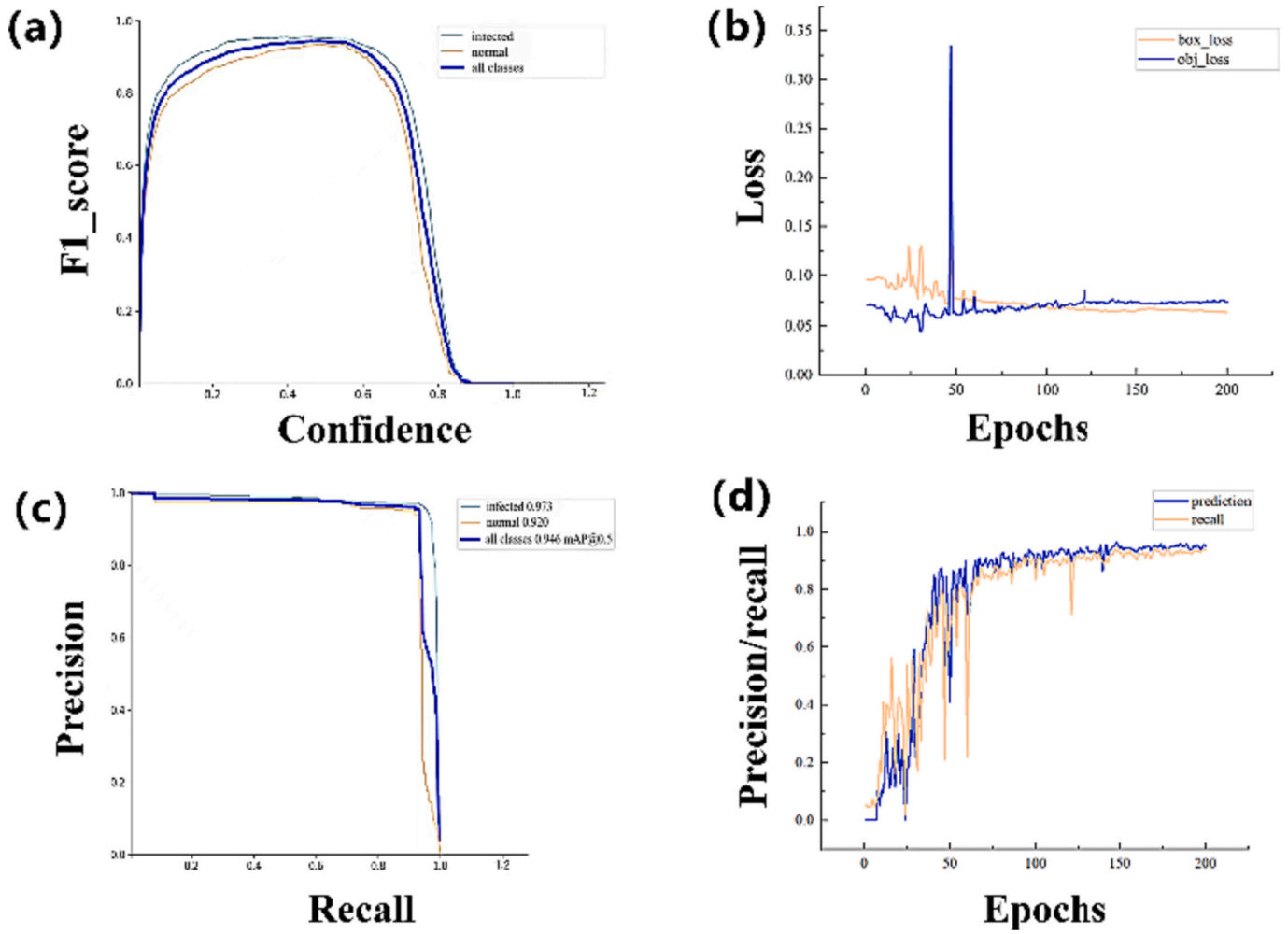


Fig. 5. The training results of our proposed NAM-YOLO v7 network: (a) the results of the indicator F1 score during the training process; (b) the value of the loss function during the training process; (c) the relationship between accuracy and recall during the training process; (d) variation curve of the ratio of accuracy to recall during the training process.

Table 2
Performance comparison between different object detection models.

Detection models	Precision (%)	Recall (%)	Inference per image (ms)
YOLO V5s network	91.4	90.3	1243
YOLO V6 network	93.2	91.7	**138.8**
YOLOX network	94.2	92.5	317
YOLOR network	91.9	92.1	1531
YOLO v7 network	94.5	92.8	218
YOLO v7 + NAM Attention	96.2	93.1	184
YOLO v7 + NAM Attention + MPDIoU (NAM-YOLO v7)	**97.3**	**93.8**	172

application of MPDIoU as a loss function in the proposed NAM-YOLO v7 network can enhance the recognition ability of the model when fish with SVC symptoms overlap with healthy fish. Compared with the original YOLO v7 network, the proposed NAM-YOLO v7 network achieves a 2.8% improvement in accuracy and a 1% improvement in the recall rate for the detection of fish infected with SVCV. These results amply demonstrate the effectiveness of these improvement strategies that we have proposed.

Finally, based on the proposed NAM-YOLO v7 network combined with machine vision techniques, fish with SVC symptoms were accurately identified and localized (Fig. 6). SVCV-infected fish showing symptoms were labelled in the red rectangular box, while other healthy fish were labelled in the blue rectangular box. Although the differences in shape and size between healthy fish and fish with SVC symptoms were small, the posture while swimming and the color of the hemorrhage demonstrated difference between the two types of fish. The images of two different types of fish while swimming are captured by machine vision methods so that these differences are converted into features on the image thus captured and extracted by our proposed NAM-YOLO v7 network. Based on the captured features, fish with SVC symptoms can be discriminated and precisely located. Although the scattering of light on the water surface and fish feces introduces a certain degree of interference to the detection process, more discriminative features can be adaptively extracted by the NAM module in the network and reduce the effect of these interference to a certain extent.

Gradient-weighted class activation mapping (Grad-CAM) was used (Zhu et al., 2023b) to explore the mechanism by which our proposed NAM-YOLO v7 network enables the detection of fish with SVC symptoms. The output of the last layer of the network during the detection of fish with SVC symptoms based on the NAM-YOLO v7 network in the experiment shown in Fig. 6 was used to reveal how it works, as shown in Fig. 7. From the results presented in Fig. 7, different categories of fish have the highest weight of features in their torso part while swimming, and the extracted features are basically concentrated on the body and

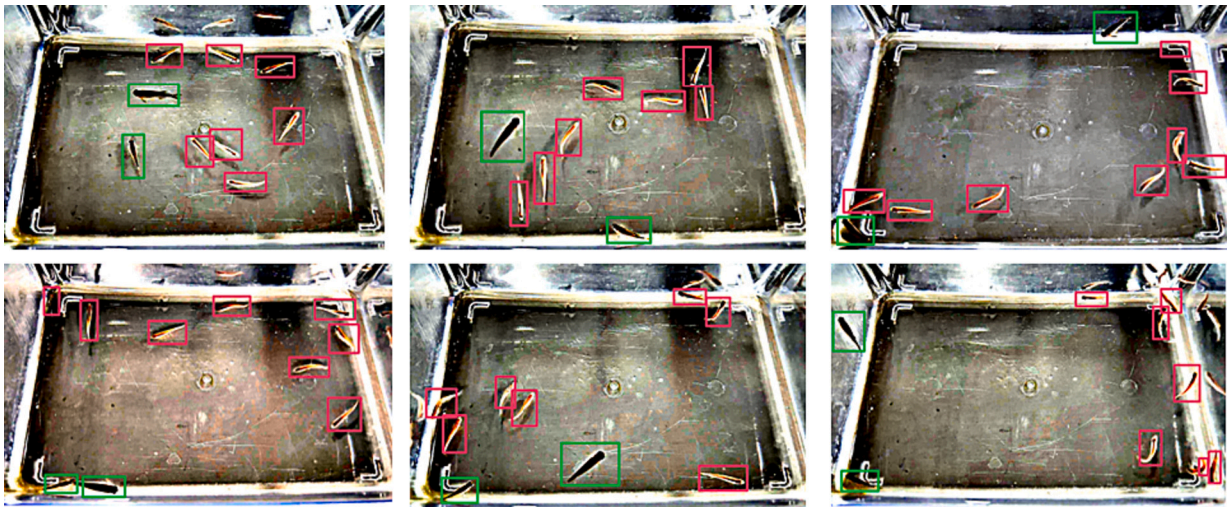


Fig. 6. Detection of target fish with SVC symptoms using the NAM-YOLO v7 network. The red anchor boxes indicate the fish with SVC symptoms detected by the CNN. The green anchor boxes indicate the fish without SVC symptoms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

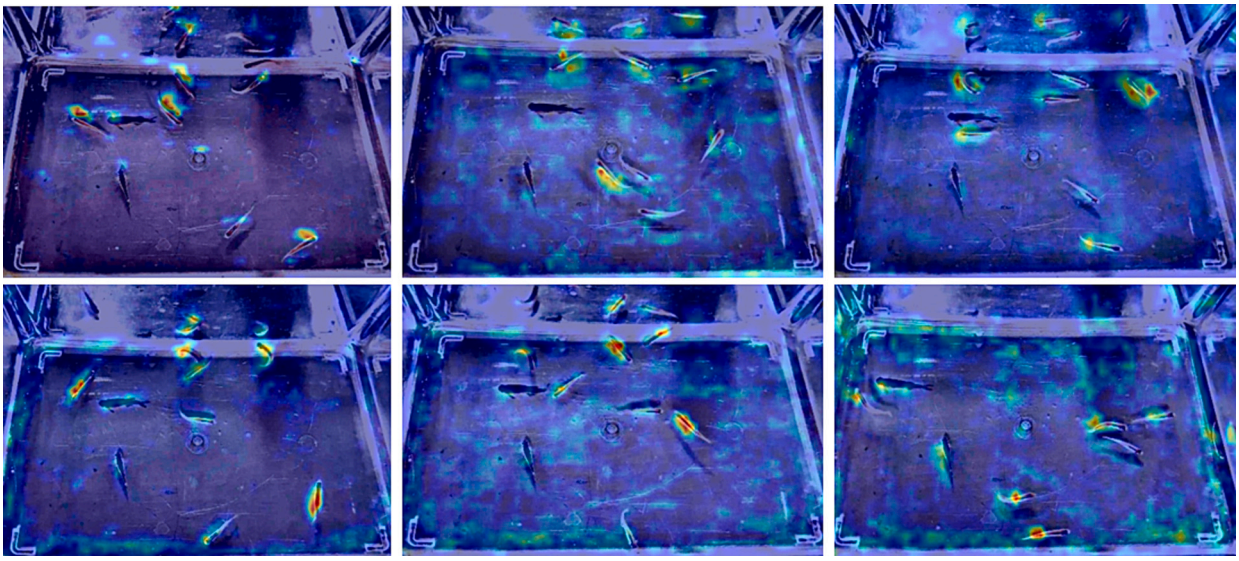


Fig. 7. Grad-CAM plot derived from the prediction layer of NAM-YOLO v7 network. The darker the color, the greater the value and the more important the feature information indicated.

head of the fish. Combining the results in Fig. 6 and Fig. 7, it can be seen that the NAM-YOLO v7 network based on our training can effectively extract the key features of the fish during movement to achieve discrimination and tracking of the fish with SVC symptoms. These results demonstrate the modified modules in our model are indispensable to focus on the essential part in the images, which allows for the realization of precisely detecting the specific SVCV-infected fish.

Finally, the trained model was deployed and integrated into a monitoring system for the fish farming process, enabling online monitoring and tracking of fish with SVC symptoms during the farming process, as shown in Fig. 8. The designed monitoring software was developed based on the PyQt5 platform, which enables the counting and tracking of fish with SVC symptoms in the aquaculture farms. The number of fish with SVC symptoms in the picture is displayed at the bottom of the software. In the future, the application and dissemination of the developed system in the actual freshwater fish farming process should be feasible. It will be very convenient to carry out a surveillance plan with many devices equipped with video capture and online data sharing. A data analysis center can be established to detect the infected fish with the further optimized neural network model.

4. Discussion

The possible outbreaks of SVC are very concerning. Hence, SVC is listed as one of the OIE notifiable diseases. Most existing molecular biology testing methods for SVC surveillance depend on professional operators and biofacilities (Shivappa et al., 2008; Clouthier et al., 2021). The procedures for these tests, such as PCR/qPCR, are normally time and labor-intensive, compared to vision-based detection techniques. Molecular biology testing methods with high precision are indispensable for inspecting at essential transmission routes such as importing sites of fish outside of the country. However, small farms may suffer from insufficient trained persons or facilities to perform biological tests. To lay a foundation for a more convenient and cost-effective preventive surveillance system, this work aims to develop a convenient method for rapid detection using novel approaches. The images are utilized to

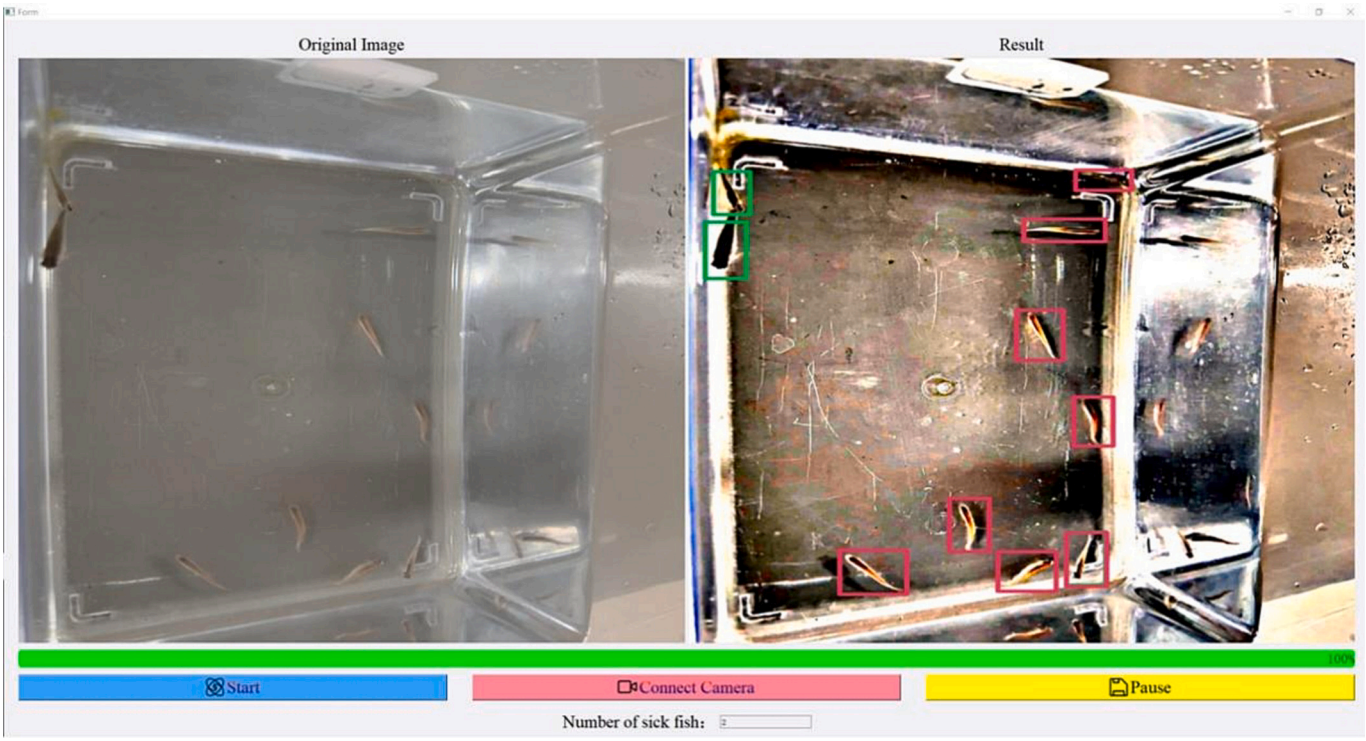


Fig. 8. SVC detection and counting system developed using PyQt. On the left is the image captured by the camera in real time, and on the right is the image that is the result of the detection. The number of fish with SVC symptoms in the picture is displayed at the bottom of the software.

detect the SVCV-infected fish after enhancement as datasets utilized by an improved YOLO v7 network.

A good dataset is the preliminary requirement to develop a high-precision target detection neural network model (Zhao et al., 2021). Due to the limitation of currently available datasets in detecting fish with SVC symptoms, this study experimentally collected the data by recording SVCV-infected and healthy zebrafish. The images for the neural network were then obtained from the video by frame-splitting and filtering. The labeling of the dataset is also critical to provide key information such as the location and category of the target in the images. Despite the difficulty brought by the overlapping or occlusion of fish during data acquisition, as many targets as possible were labelled in order to provide sufficient training data for the network. A total of 1814 images were extracted and over 10,000 targets were identified. Underwater videos or image recordings can be affected by a variety of factors, such as water refraction, scattering, and color attenuation, leading to image distortion, blurring, and color distortion (Zhao et al., 2021). Thus, Auto-MSRCR is developed for data enhancement to improve image quality. In addition, to prevent the network from learning irrelevant features, such as fish orientation, we introduce mosaic data enhancement, including updown and left-right flipping, to ensure that the detection by the model is less compromised by the collected lower quality images. The data enhancement lays a foundation for the feasible and efficient application of our neural network in precisely detecting the SVC symptoms. Also, the data enhancement algorithm can be migrated or adopted in detecting other fish diseases by machine vision technologies, which can benefit related studies and applications in the similar fields.

Several excellent models have been proposed and applied to fish aquaculture. The YOLO series models were adopted in real-time fish monitoring (Hu et al., 2021; Wang et al., 2023). In this study, a target detection network based on YOLO v7 was introduced as a base model and used to implement the detection process of fish with SVC symptoms. At the same time, NAM modules are added to the YOLO v7 network and combined to form an N-ELAN structure thus enabling the enhancement of sensitive features in the image (Zhao et al., 2021; Jalal et al., 2020; Hu et al., 2018). The weights of different channels in the image can be adjusted based on the NAM module according to the importance of different channel information to enable useful features to be further extracted, thus improving the detection accuracy of fish with SVC symptoms. In addition, the introduction of the NAM module can further suppress some meaningless image features, thus reducing the computation of the model and decreasing the detection time of a single image. By replacing the original CIOU loss function in YOLO v7 with the MPDIoU loss function, the detection error due to the difference in the aspect ratio of the individuals in the fish population was eliminated to some extent. At the same time, the MPDIoU loss function was used to reduce the detection error of the model due to the overlapping of individuals when the fish were aggregated. This is because the MPDIoU function solves the problem better by considering the minimum distance in the calculation process. Therefore, the same aspect ratio is not restricted during the optimization process. The high precision and high recall rate of our model demonstrated the effectiveness compared to other established models (Table 2). In the aquaculture farms, the detection error must be exaggerated by more complicated environment. Hence, the integrated module and updated loss function will be inevitable to guarantee the detection accuracy and facilitate the following realistic applications.

Based on the above improvement strategies, our proposed NAM-YOLO v7 network achieves the best performance compared to other target detection networks (e.g., YOLO v5s, YOLO v6, YOLO X, YOLO R, and YOLO v7) (Hou et al., 2021; Li et al., 2023). Based on our designed NAM-YOLO v7 network, the accuracy and recall of achieving the detection of fish with SVC symptoms while swimming are 97.3% and 93.8%, respectively. Its detection time for a single image is only 172 ms. Meanwhile, an ablation experiment shows that our proposed improved strategy effectively enhances the performance of the target detection network applied in the detection of fish with SVC symptoms. Experiments using Grad-Cam showed that the model succeeded in focusing its attention highly on diseased fish, demonstrating its accuracy and

robustness in identifying fish with SVC symptoms. These results fully demonstrate that our proposed method of machine vision combined with a trained NAM-YOLO v7 network can effectively achieve online detection of fish with SVC symptoms during freshwater fish farming. This technique is essential for the rapid detection and isolation of potentially infected fish so that appropriate measures can be taken to contain or cut off transmission.

Although we realized the identification of fish with SVC symptoms, our proposed method suffered from some limitations as a preliminary investigation. The datasets collected in this study do not cover the SVCV infection of different types, ages and sizes of fish. More datasets need to be collected to make the model more widely applicable. Second, the dataset for the neural network detection method is based on the image format extracted from the video stream. Some features of fish behaviors may be limited. In addition, the labeling of images before the training of the object detection model is very time-consuming. Weakly supervised or unsupervised models may be adopted to achieve reliable detection under the circumstances of limited data labeling (Xie et al., 2021; Bar et al., 2022).

Surveillance is of great significance in warning and containing the epidemic diseases. In this study, we developed a CNN-based vision technique for SVC detection. The neural network is established based on YOLOv7 network. By integrating with NAM Attention mechanism and MPDIoU, the detection precision and computation cost are improved than other state-of-art detection algorithms in SVC detection. The target loss in detection is minimized. The ablation experiment demonstrates the enhancement of these modules. The output heat map of the NAM-YOLOv7 network by Grad-Cam illustrate the CNN is focusing on the expected fish targets instead of other insignificant area. Overall, the experimental testing datasets demonstrate that this NAM-YOLOv7 network achieved the goal of rapid and accurate detection of fish with SVC symptom, which could provide a very efficient and cost-effective method to serve a broader disease surveillance plan in the future.

CRediT authorship contribution statement

Yaoyi Cai: Writing – original draft, Methodology, Conceptualization. **Zekai Yao:** Writing – original draft, Software, Methodology. **Haibo Jiang:** Visualization, Software. **Wei Qin:** Investigation. **Jun Xiao:** Investigation. **Xiuxiang Huang:** Investigation. **Jiaji Pan:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization. **Hao Feng:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (31920103016, U21A20268, 32173010), the fellowship of China Postdoctoral Science Foundation (No. 2022M711128), Hunan Provincial Science and Technology Department (2022JJ40276, 2022JJ30383). National College Student Innovation and Entrepreneurship Training Program (S202310542044X, 2023227, 2022123). We thank the anonymous reviewers for their comments and questions.

References

Ahne, W., Bjorklund, H.V., Essbauer, S., Fijan, N., et al., 2002. Spring viremia of carp (SVC). Dis. Aquat. Org. 52 (3), 261–272.

Bar, A., Wang, X., Kantorov, V., et al., 2022. Detreg: Unsupervised pretraining with region priors for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14605–14615.

Bist, R.B., Subedi, S., Yang, X., et al., 2023. A novel YOLOv6 object detector for monitoring piling behavior of cage-free laying hens. AgriEngineering. 5 (2), 905–923.

Cai, K., Miao, X., Wang, W., et al., 2020. A modified YOLOv3 model for fish detection based on MobileNetv1 as backbone. Aquac. Eng. 91, 102117.

Clouthier, S.C., McClure, C., Schroeder, T., et al., 2021. Bayesian latent class model estimates of diagnostic accuracy for three test methods designed to detect spring viremia of carp virus. Prev. Vet. Med. 190, 105338.

Dixon, P.F., Longshaw, C.B., 2005. Assessment of commercial test kits for identification of spring viraemia of carp virus. Dis. Aquat. Org. 67 (1–2), 25–29.

Embregts, C.W.E., Rigaudeau, D., Vesely, T., et al., 2017. Intramuscular DNA vaccination of juvenile carp against spring viremia of carp virus induces full protection and establishes a virus-specific B and T cell response. Front. Immunol. 8, 1340.

Emmenegger, E.J., Sanders, G.E., Conway, C.M., et al., 2016. Experimental infection of six north American fish species with the North Carolina strain of spring viremia of carp virus. Aquaculture 450, 273–282.

Fouad, A.M., Soliman, H., Abdallah, E.S.H., et al., 2019. In-vitro inhibition of spring viremia of carp virus replication by RNA interference targeting the RNA-dependent RNA polymerase gene. J. Virol. Methods 263, 14–19.

Goodwin, A.E., 2009. Spring viremia of carp virus (SVCV): global status of outbreaks, diagnosis, surveillance, and research. Isr. J. Aquacult-Bamid. 61 (3), 180–187.

Han, F., Zhu, J., Liu, B., et al., 2020. Fish shoals behavior detection based on convolutional neural network and spatiotemporal information. IEEE Access 8, 126907–126926.

Hou, Q., Zhou, D., Feng, J., et al., 2021. Coordinate attention for efficient mobile network design. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13713–13722.

Hu, J., Shen, L., Sun, G., et al., 2018. Squeeze-and-excitation networks. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 7132–7141.

Hu, J., Zhao, D., Zhang, Y., et al., 2021. Real-time nondestructive fish behavior detecting in mixed polyculture system using deep-learning and low-cost devices. Expert Syst. Appl. 178, 115051.

Huang, P.X., Boom, B.J., Fisher, R.B., 2012. In Underwater live fish recognition using a balance-guaranteed optimized tree. In: Computer Vision–ACCV 2012: 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5-9, 2012, Revised Selected Papers, Part I 11. Springer, pp. 422–433.

Jalal, A., Salman, A., Mian, A., et al., 2020. Fish detection and species classification in underwater environments using deep learning with temporal information. Ecol. Inform. 57, 101088.

Li, J., Lian, Z., Wu, Z., et al., 2023. Artificial intelligence-based method for the rapid detection of fish parasites (Ichthyophthirius multifiliis, Gyrodactylus kobayashii, and Argulus japonicus). Aquaculture 563, 738790.

Liu, Y., Shao, Z., Teng, Y., 2021. Hoffmann, N., 2021. NAM: Normalization-based attention module. arXiv preprint arXiv, p. 12419.

Ma, S., Xu, Y., 2023. MPDIoU: A Loss for Efficient and Accurate Bounding Box Regression. arXiv preprint arXiv, p. 07662.

Oreifej, O., Shu, G., Pace, T., et al., 2011. A two-stage reconstruction approach for seeing through water. IEEE CVPR 2011, 1153–1160.

Padhi, A., Verghese, B., 2008. Detecting positively selected codons in the glycoprotein of spring viraemia of carp virus (SVCV) isolates from the USA and China. J. Fish Dis. 31 (10), 785–791.

Pan, Z., Li, D.L., 2022. EPSA-YOLO-V5s: a novel method for detecting the survival rate of rapeseed in a plant factory based on multiple guarantee mechanisms. Comput. Electron. Agric. 193, 106714.

Pan, J., Qin, W., Chen, Z., et al., 2023. Analyzing the transmission of SVCV among zebrafish populations by compartment mathematical modeling. Aquaculture 564, 739076.

Shao, L., Zhao, J.R., 2017. Isolation of a highly pathogenic spring viraemia of carp virus strain from grass carp (Ctenopharyngodon idella) in late summer, China, 2016. Virus Res. 238, 183–192.

Shao, L., Zhao, J.R., Zhang, H.Q., 2016. Spring viraemia of carp virus enters grass carp ovary cells via clathrin-mediated endocytosis and macropinocytosis. J. Gen. Virol. 97, 2824–2836.

Shimahara, Y., Kurita, J., Nishioka, T., et al., 2016. Development of an improved RT-PCR for specific detection of spring viraemia of carp virus. J. Fish Dis. 39 (3), 269–275.

Shivappa, R.B., Savan, R., Kono, T., et al., 2008. Detection of spring viraemia of carp virus (SVCV) by loop-mediated isothermal amplification (LAMP) in koi carp, *Cyprinus carpio* L. J. Fish Dis. 31 (4), 249–258.

Song, C., Zhang, F., Li, J., et al., 2023. Detection of maize tassels for UAV remote sensing image with an improved YOLOX model. J. Integr. Agric. 22 (6), 1671–1683.

Spampinato, C., Giordano, D., Di Salvo, R., et al., 2010. In Automatic fish classification for underwater species behavior understanding. In: Proceedings of the first ACM International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams, pp. 45–50.

Wang, C., Dong, S., Zhao, X., et al., 2019. GAN: deep learning semisupervised salient object detection in the fog of IoT. IEEE T. Inf. 16 (4), 2667–2676.

Wang, J.-H., Lee, S.-K., Lai, Y.-C., et al., 2020. Anomalous behaviors detection for underwater fish using ai techniques. IEEE Access 8, 224372–224382.

Wang, Z., Liu, H., Zhang, G., et al., 2023a. Diseased fish detection in the underwater environment using an improved YOLOV5 network for intensive aquaculture. Fishes 8 (3), 169.

Wang, Y., Wu, H., Hua, X., et al., 2023b. Biological characters identification for hard clam larva based on the improved YOLOX-s. Comput. Electron. Agric. 212, 108103.

Xie, E., Ding, J., Wang, W., et al., 2021. Detco: unsupervised contrastive learning for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8392–8401.
Yu, X., Wang, Y., An, D., et al., 2021. Identification methodology of special behaviors for fish school based on spatial behavior characteristics. Comput. Electron. Agric. 185, 106169.
Yu, G.Y., Cai, R.L., Su, J.P., et al., 2023. U-YOLOv7: a network for underwater organism detection. Ecol. Inform. 75, 102108.
Zhao, J., Gu, Z., Shi, M., et al., 2016. Spatial behavioral characteristics and statistics-based kinetic energy modeling in special behaviors detection of a shoal of fish in a recirculating aquaculture system. Comput. Electron. Agric. 127, 271–280.
Zhao, J., Bao, W., Zhang, F., et al., 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, Z., Liu, Y., Sun, X., et al., 2021. Zhou, composited FishNet: fish detection and species recognition from low-quality underwater videos. IEEE T. Image Proc. 30, 4719–4734.
Zhou, Y., 2024. A YOLO-NL object detector for real-time detection. Expert Syst. Appl. 238, 122256.
Zhou, C., Xu, D., Chen, L., et al., 2019. Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision. Aquaculture 507, 457–465.
Zhu, Q., Ma, K., Wang, Z., et al., 2023a. YOLOv7-CSAW for maritime target detection. Front. Neurorobot. 17.
Zhu, S., Zhang, Y., Feng, Y., et al., 2023b. GW-net: An efficient grad-CAM consistency neural network with weakening of random erasing features for semi-supervised person re-identification. Image Vis. Comput. 137, 104790.