Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Computer Vision-Based Detection Model for Wild Chinese Giant Salamanders in Complex Environments

junyi chen

chenjunyi1224@163.com

Hubei Minzu University

xinyao Yang

Hubei Minzu University https://orcid.org/0009-0002-7796-4679

Nanqing Sun

Hubei Minzu University

Didi Lu

Henan University of Technology

Hang Gao

Hubei Minzu University

da Qiu

Hubei Minzu University

fang Tian

Hubei Minzu University

Zhangnian Qin

Hubei Xianfeng Zhongjian River Giant Salamander National Nature Reserve Management Center

Xingmin Tuo

Hubei Xianfeng Zhongjian River Giant Salamander National Nature Reserve Management Center

Lirong Chen

Hubei Xianfeng Zhongjian River Giant Salamander National Nature Reserve Management Center

Lei He

Hubei Xianfeng Zhongjian River Giant Salamander National Nature Reserve Management Center

Article

Keywords: Chinese giant salamander, object detection, YOLO v11, Deep learning

Posted Date: September 10th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7517311/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

A Computer Vision-Based Detection Model for Wild Chinese Giant Salamanders in Complex Environments

**********[1] **********[1] **********[3] **********[1*] **********[1]
********** **********[1]

**********[2] **********[2] **********[2] **********[2]

(*1.* **********)
2. **********,**********
3 **********)

Abstract: To achieve rapid and accurate identification of wild Chinese giant salamanders in complex field environments, we propose an improved recognition model based on YOLO v 11n. This model incorporates an Efficient Multi-scale Attention Module (EMA) into the Bac kbone layer, replaces the Complete Intersection over Union (CIoU) loss function with Wise-IoU (WIoU) loss, and introduces Lightweight Adaptive Extraction of Convolutions (LAE) in to the Head layer. Ablation and comparative experiments demonstrate that the improved mo del achieves recall, precision, F1 score, and frame rate of 94.85%, 95.39%, 95.12%, and 77. 20 f/s, respectively. The model occupies 11.56 MB of memory and performs 8.65×10^9 float ing-point operations. Compared to the baseline YOLO v11n, the recall, precision, F1 score, and frame rate are 5.70, 6.13, 5.92 percentage points higher and 27.1 fps faster, respectivel y. The proposed YOLO v11n-EWL model demonstrates significant improvements in stability, recognition speed, and accuracy. The improved model meets the real-time detection require ments for wild Chinese giant salamanders in natural habitats and can sustain long-term outd oor operation. Based on this, an all-weather image recognition and behavior detection syste m for wild salamanders was developed, successfully detecting over 20 instances of wild Chi nese giant salamanders. This system provides a stable and reliable monitoring tool for adva ncing the conservation of wild salamanders and their habitats.

Key words: Chinese giant salamander; object detection; YOLO v11; Deep learning

0. Introduction

The existence of wild Chinese giant sa lamanders, classified as a national Grade II protected species, serves as an important in dicator of the ecological well-being in their habitat. Consequently, developing reliable an d effective multi-faceted techniques and tool s for detecting giant salamanders is crucial for their conservation and population recove ry.

The Hubei Xianfeng Zhongjian River Giant Salamander National Nature Reserve i s situated in the towns of Gaoleshang and Zhongbao within Xianfeng County, forming part of the Zhongjian River Basin, a signifi cant tributary of the Qingjiang River in the Yangtze River system. The giant salamand er serves as the reserve's emblematic specie s. This amphibian exhibits a nocturnal lifest yle, primarily inhabiting river segments and

Received: Revised:
Funding: **********
Author Profile: **********
Corresponding author: **********

smaller tributaries. The habitat is marked by intricate and low-light conditions, presenting three main obstacles for conventional detection techniques: the darkness at night, the complicated environmental backdrop, and other elements that hinder visual identification, often resulting in missed sightings; current monitoring approaches suffer from considerable delays, failing to deliver real-time, quantitative insights into the salamanders' presence, locations, timings, and behaviors; and there is a notable absence of automated systems that can reliably function in challenging field conditions while providing structured data promptly. Consequently, there is an urgent need for the creation of a comprehensive observation system that can effectively and consistently detect and identify wild giant salamanders in real-time, even in complex lighting situations, while also enabling quantitative data analysis, which is crucial for the conservation and study of these creatures in protected areas.

In recent times, the application of deep learning has surged in intricate settings. Hu and colleagues enhanced the YOLOv3 framework by refining both the pooling block and the loss function, leading to improved accuracy and performance in analyzing fish behavior. Xu and his team utilized the faster R-CNN model to study fish behavior in varying ammonia levels. Meanwhile, Wang and associates introduced small target detection frames for multi-scale fusion, optimizing the YOLOv5 model to better identify abnormal behaviors in zebra snapper. HUSSAM and his team integrated color multi-scale Retinex enhancement with the YOLO algorithm, resulting in improved fish detection accuracy. DU and colleagues developed a ResNet50-LSTM algorithm to effectively and reliably identify fish breeding behaviors in complex environments. Tu and his team incorporated a large detection head into the YOLOv8 model and employed the ECAM attention mechanism to minimize background interference, allowing for a sharper focus on key features of individual fish and enhancing the detection of indistinct fish. These investigations highlight the promising applications of deep learning, although it has yet to be utilized in research concerning giant salamanders.

To address the challenges related to the detection and inadequate safeguarding of giant salamanders and their habitats, we concentrate on the development of an intelligent, all-weather monitoring system and the promotion of sustainable practices for these creatures in the Hubei Xianfeng Zhongjian River National Nature Reserve. We introduce an innovative target recognition model, YOLO v11-EWL, and establish a comprehensive system for the intelligent recognition and behavioral monitoring of giant salamanders in various weather conditions.

1. Equipment and dataset construction

1.1 Experimental equipment

The developed imaging device operates without physical contact to ensure the safety of the giant salamander's habitat. It features a positioning system for identifying the salamanders and an imaging system dedicated to photographing them. This device is designed for prolonged outdoor use, enabling continuous, unattended operation for 24 hours, as illustrated in Fig. 1.



Fig. 1 Schematic diagram of image acquisition device

1. Camera for image capture 2. Antenna for positioning 3. Switch device 4. Power supply unit 5. Gateway interface 6. Motherboard for positioning module 7. Extension cable for positioning antenna 8. Environmental sensor

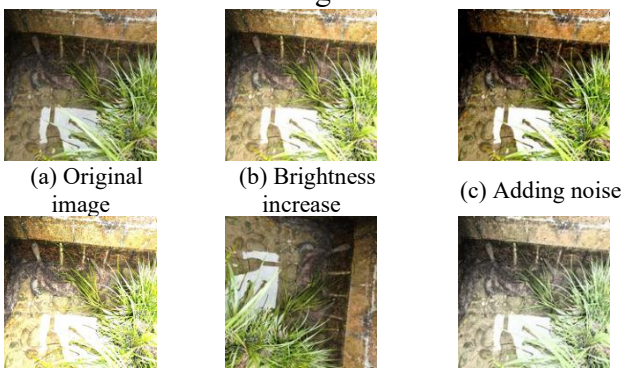
The power supply module converts 220 V AC into 12V DC, making it suitable for various devices, including environmental sensors, RFID systems, and cameras. These components communicate with the server via a gateway, allowing for centralized management through a web interface. The RFID positioning system on the front end monitors the signal strength of tagged salamanders in real time. When it detects a specific Received Signal Strength Indicator (RSSI) value, it activates the camera to capture synchronized images of multiple targets. These images are sent to the server through a switch, where they are linked with RFID positioning data and environmental sensor details to create a structured dataset. Later, this model can be utilized on the server to identify wild salamanders that do not have RFID tags.

1.2 Data Set Construction

1.2.1 Data acquisition

Images of giant salamanders were gathered between December 2024 and February 2025 at the Hubei Xianfeng Zhongjian River National Nature Reserve dedicated to these creatures.

The initial dataset utilized in the study comprises 1,020 images, which is insufficient for effective model training. To allow the model to thoroughly grasp the features of the giant salamander, an additional 530 images were sourced from online resources. The dataset was further augmented through techniques such as rotation, sharpening, brightness adjustment, and noise addition, resulting in a total of 6,952 image samples. This not only enhances the dataset but also simulates various weather conditions. Examples of the data augmentation effects can be seen in Fig. 2.


(a) Original image (b) Brightness increase (c) Adding noise

(d) Sharpening (e) Rotation (f) HSV filtering

Fig. 2 Data enhancement effect diagram

1.2.2 Data labeling

The data set utilizes X-Any Labeling software for its labeling process, with the resulting TXT file detailing the dimensions of

the bounding box for giant salamanders. This data set is categorized into three parts: training, testing, and validation, following a distribution of 70%, 20%, and 10%, which corresponds to 4866, 1390, and 695 samples, respectively. An example of the manual labeling of giant salamanders is illustrated in Fig. 3.

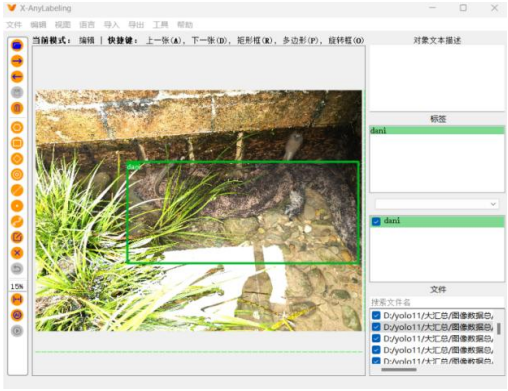


Fig. 3 Artificial labeling map of giant salamanders

2 Construction of giant salamander detection model

2.1 YOLO v11 model

The YOLO v11 architecture is composed of three main components: the Backbone, Neck, and Head layers. The Backbone utilizes the C3k2 module to improve feature extraction through the Bottleneck Block Spatial Pyramid Pooling Fast (SPPF) and Cross stage partial with pyramid squeeze attention (C2PSA) modules. The Neck layer is responsible for merging and enhancing features, while the Head layer generates the final output. Despite its strengths, the C2PSA module struggles with effectively integrating multi-scale features in the complex environments of giant salamander reserves. Additionally, the CIoU (Complete Intersection over Union) loss function exhibits instability when anchor frame quality is inconsistent, and the base model's memory usage reaches 59.65 MB, leading to increased computational costs. The Neck layer continues to focus on feature fusion and enhancement, with the Head layer delivering the ultimate detection results.

2.2 Improvement of YOLO v11n-EWL model

An enhanced recognition model is introduced, building upon the YOLO v11n framework. This model incorporates an Efficient Multi-Scale Attention (EMA) module into the Backbone layer, substitutes the CIoU loss function with a Wise-IoU (WIoU) loss function, and integrates a Lightweight Adaptive Extraction of Convolutions (LAE) into the Head layer to boost accuracy. The combination of these three strategies leads to the development of the YOLO v11n-EWL model. The architecture of this model is illustrated in Fig. 4.

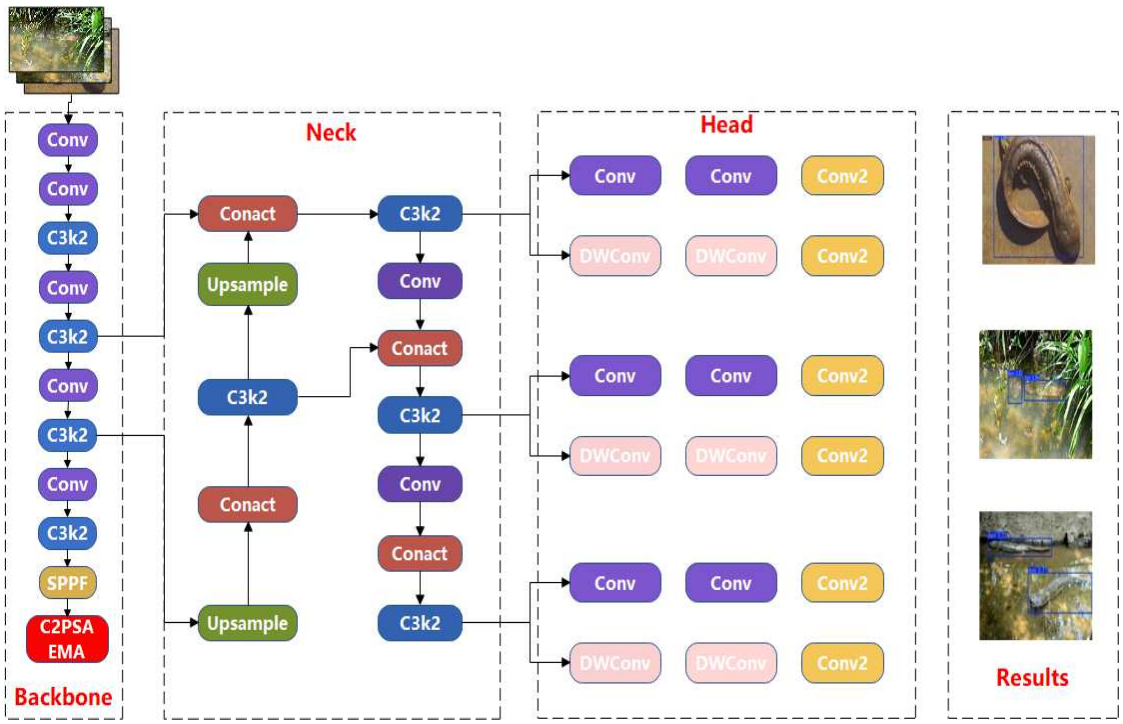


Fig.4 YOLO v11n-EWL Network Architecture Diagram

2.2.1 Efficient Multiscale Attention Module EMA

The protected area's environment is intricate and fluctuating, causing the salamander's coloration to blend seamlessly with its surroundings. Numerous environmental disruptions can lead to missed detections. To enable the model to grasp the unique traits of the salamander during training and optimization, facilitating swift identification in such a complex setting, this research incorporates the EMA high-efficiency multiscale attention module in place of the standard attention mechanism within the C2PSA of the backbone network. This enhancement aims to bolster the multiscale features of the YOLO v11n model. The design of the EMA attention module is illustrated in Fig. 5.

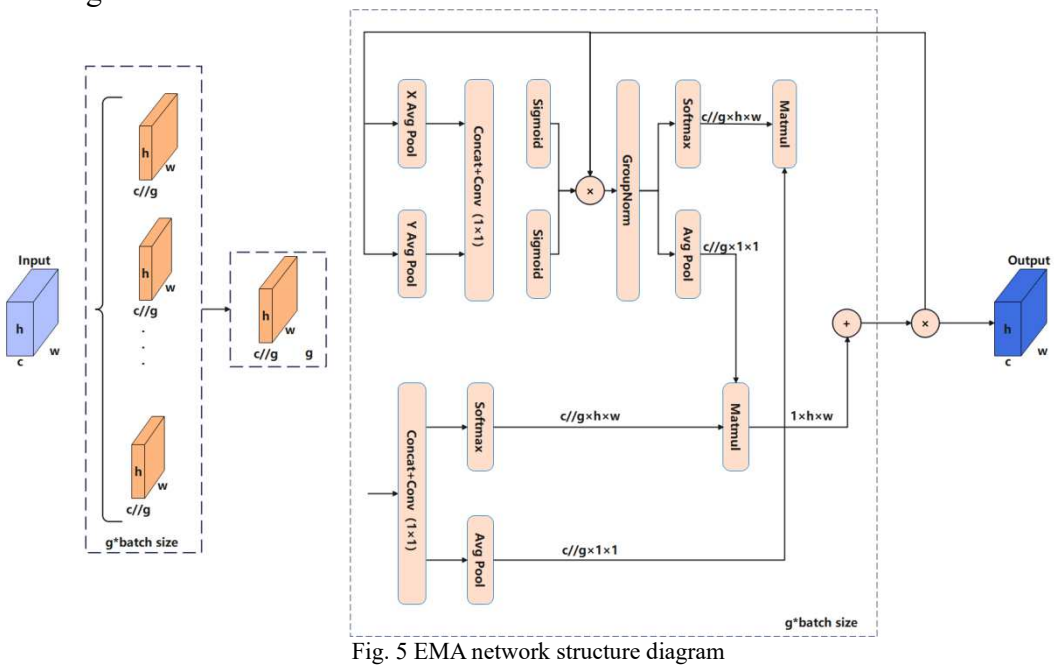


Fig. 5 EMA network structure diagram

EMA[14] aims to retain multiscale characteristics while preserving the dimensions of the channels. It reformats certain channels into batch dimensions and organizes the channel dimensions into various sub-features, preventing a decrease in channel dimensionality to improve the extraction of features.

$X \in R^{C \cdot H \cdot W} X = [X_0, X_i, \ldots, X_{G-1}] \, X \in R^{C//G \cdot H \cdot W}$ The process unfolds as follows: initially, EMA takes any given input and partitions it into G sub-features based on the channel dimension, meaning that the input channel C is divided into G groups. Subsequently, attention weight descriptors for the grouped feature maps are derived through three distinct pathways. The first two pathways utilize a 1×1 branch that incorporates a 1×1 convolution operation, while the third pathway employs a 3×3 branch featuring a 3×3 convolution operation. Global spatial information is then encoded from the outputs of both the 1×1 and 3×3 branches using two-dimensional global average pooling. The output feature maps for each group are created by combining the two generated spatial attention weight values. Ultimately, global context information is extracted by analyzing pixel-level pairwise relationships through the Sigmoid activation function, utilizing X AvgPool for horizontal global pooling and Y AvgPool for vertical global pooling.

The formula for the 2D global average pooling operation is

$$Z_C = \frac{1}{H \cdot W} \sum_{j}^{H} \sum_{i}^{W} X_C(i,j) \tag{1}$$

WhereZ_C - output value of the Cth channel after pooling

$X_C(i,j)$ --Input value of the Cth channel at the position of width i height j

2.2.2 WIOU loss function

The intricate backdrop of the salamander images captured in the outdoor environment of the protected area leads to a significant class imbalance between the salamander bodies and their surroundings. During model training, there are considerably fewer high-quality anchor frames compared to those with substantial errors. The presence of low-quality samples generates a large gradient, which can negatively influence the model's optimization trajectory. The structure is illustrated in Fig. 6.

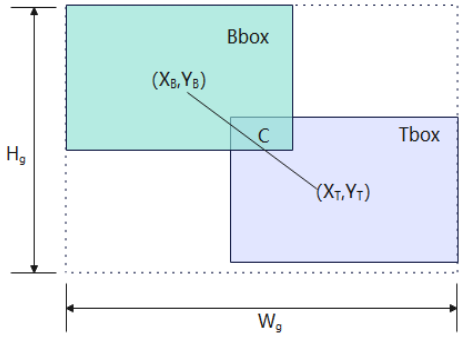


Fig 6. Iou structure diagram

In order to address the issues mentioned, this research introduces WIoU as a substitute for the initial loss function. WIoU[15] features a dynamic non-monotonic focusing approach that assesses the quality of anchor frames, along with a strategy for dynamically modifying gradient gains. The calculation for WIoUv1 is as follows:

$$L_{WIoUv1} = R_{WIoU} L_{IoU} \tag{2}$$

where

$$R_{WIoU} = exp\left(\frac{(x - x_0)^2 + (y - y_0)^2}{(W_I^2 + H_I^2)^*}\right) \tag{3}$$

Where L_{WIoUv1} - weighted IoU loss function version 1

R_{WIoU}--Based on the centroid distance weighting factor

$$L_{IoU} = 1 - \text{IoU} \tag{4}$$

whereL_{IoU} - standardized IoU loss

W_I $H_I R_{WIoU}$* indicates the division from the computational graph, which stops it from contributing to gradient calculations. As a result, it acts as a constant during forward computation, remaining a fixed value and not engaging in

backpropagation, thereby avoiding the generation of gradients that could hinder convergence.

To avoid the inclusion of subpar salamander image samples that generate significant detrimental gradients, the outlier degree β is established to assess the quality of the anchor frame, which is determined using the formula below.

$$\beta = \frac{L_{IoU}^{*}}{L_{IoU}} \tag{5}$$

Where L_{IoU}^{*} - modified IoU loss

In the case of anchor frames that are of poor quality and contain significant outliers, a reduced gradient gain is allocated. The construction of WIoUv3 incorporates these outliers along with WIoUv1, determined by the following calculation:

$$L_{WIoUv3} = rL_{WIoUv1} \tag{6}$$

Where L_{WIoUv3} - weighted IoU loss function version 3

2.2.3 Lightweight Adaptive Extraction Convolution Module LAE

To enhance the speed of model recognition and minimize memory usage, a lightweight adaptive extraction convolution method is implemented. The LAE module serves as an effective feature extraction component, enhancing the model's adaptability to intricate scenes by dynamically modifying convolution kernel parameters and selecting features adaptively, all while decreasing computational demands. The configuration of LAE is illustrated in Fig. 7.

In the section on Lightweight extraction, the input image undergoes convolution in groups, specifically with groups set to c/16. This indicates that the input channels are segmented into c/16 distinct groups, enhancing the efficiency of feature extraction. Moving to the Dimension mapping section, the convolved features are reorganized (R), meaning the outcomes of the group convolution are rearranged into various spatial dimensions for effective feature representation. The Adaptive extraction (AE) segment involves applying adaptive average pooling to the feature map, which consolidates feature information across height and width. This pooling process reduces the spatial dimensions of the inputs, allowing the model to concentrate more on global features. A 1x1 convolution is then applied to the pooled feature map, further compressing and adjusting the number of channels. Subsequently, the feature map is rearranged and subjected to Softmax normalization, enabling adaptive adjustments to the weights of each channel, thereby enhancing the network's expressive capabilities. Finally, the combined feature maps are added to the input features through a residual connection, ensuring a seamless flow of information and producing the final feature maps with dimensions of h/2×w/2×c.

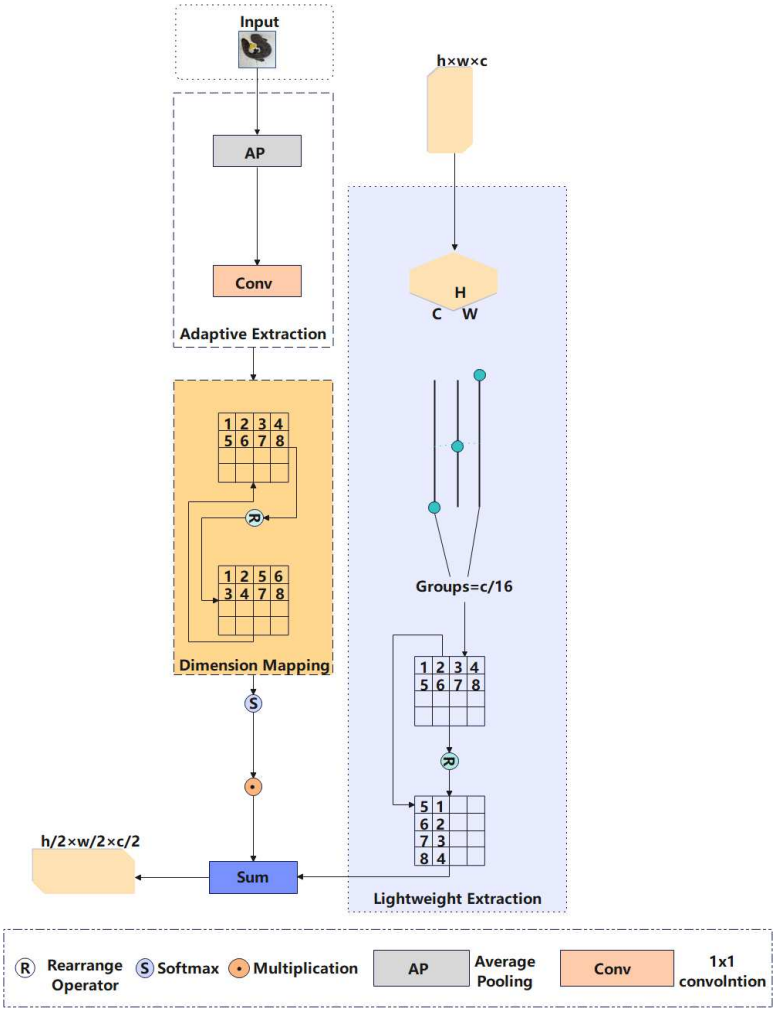


Fig. 7 LAE network structure diagram

3. Experimental results and discussion

3.1 Experimental environment and evaluation index

The study utilizes the Autodl computing platform equipped with an NVIDIA GeForce RTX 4090 GPU featuring 24GB of VRAM, alongside the Pytorch 2.3.0 deep learning framework, Python version 3.12.3, and Conda 24.4.0. In the model's training phase, a total of 500 epochs were conducted, with an initial learning rate established at 0.001 and momentum configured at 0.937. The evaluation of the model's detection capabilities was carried out through metrics such as Precision (P), Recall (R), F1 score, floating-point operations, memory consumption, and frame rate.

3.2 Performance Comparison Experiments of Different Attention Mechanisms

To assess the impact of the EMA attention module on the YOLO v11n detection model for salamander images, both coordinate attention (CA) and the convolutional block attention module (CBAM) were implemented to enhance the model. Performance comparison tests were carried out using the same dataset of salamander images within a consistent experimental setup. The findings are presented in Table 1.

The YOLO v11n+EMA model outperforms YOLO v11n+CA by 7.05, 4.53, 5.80 percentage points in recall, precision, and F1 score, along with a frame rate increase of 1.30 f/s. Additionally, it surpasses YOLO v11n+CBAM by 3.26, 1.00, 2.13 percentage points in the same metrics and achieves a

3.20 f/s higher frame rate. These findings highlight the substantial impact of the EMA attention module on enhancing the performance of the YOLO v11n model.

Tab. 1 Results of comparative tests between different attention modules

Model	Recall rate /%	Precision rate /%	F1 Score /%	Frame rate /(f s^{-1})
YOLO v11n+CA	85.14	86.73	85.92	62.20
YOLO v11n+CBAM	88.93	90.26	89.59	60.30
YOLO v11n+EMA	92.19	91.26	91.26	63.50

3.3 Results of ablation experiments

To assess the practicality of the three enhanced techniques, ablation tests were conducted on an identical dataset within the same configuration settings, utilizing YOLO v11n as the foundational network. This involved integrating the EMA Attention Mechanism Module into the Backbone Layer, substituting the CIoU loss function with the WIoU loss function, and incorporating the Lightweight Adaptive Extractive Convolutional (LAE) in the Head Layer. The results of the validation are presented in Table 2, where \√\ signifies the adoption of a particular enhancement and \×\ denotes its exclusion. According to Table 2, the modified network YOLO v11n-EWL demonstrates improvements in recall, precision, F1 score, and frame rate by 5.12, 5.07, 5.10 percentage points, and 22.60 frames per second (f/s), respectively, while also decreasing the model's memory usage by 48.67 MB compared to the version that only uses the WIoU loss function. When compared to the model that solely employs the LAE, it shows enhancements in recall, precision, F1 score, and frame rate by 3.32, 3.80, 3.57 percentage points, and 8.90 f/s, respectively, with a reduction in memory usage of 7.29 MB. Additionally, it outperforms the model with only the EMA by 2.66, 4.13, 3.40 percentage points, and 15.10 f/s, while also lowering the memory footprint by 67.07 MB. Compared to the model that includes both EMA and LAE, it achieves improvements of 3.59 percentage points, 2.14 percentage points, and 1.60 f/s, respectively. The YOLO v11n-EWL model, incorporating all three enhancements, surpasses the base YOLO v11n network by 5.70, 6.13, 5.92 percentage points, and 27.10 f/s, while also reducing memory consumption by 48.09 MB. The enhancements lead to a significant decrease in memory usage by 94.85%, 95.39%, 95.12%, and 77.20%, resulting in a total memory consumption of 11.56 MB. The improvements stem from the EMA module's ability to better capture pixel-level relationships, the WIoU's filtering of low-quality anchor frames, and the LAE module's reduction of floating-point operations, collectively enhancing the model's performance, robustness, and generalization capabilities, ultimately improving salamander detection in complex environments.

In the task of recognizing salamanders through images, simply adding a new module or loss function does not necessarily enhance the model's performance; it could even lead to a decline in effectiveness. This suggests that relying on a solitary enhancement strategy is insufficient for optimizing the model. A holistic approach that takes various factors into account is essential for achieving better performance.

Tab. 2 Results of ablation experiments

EMA	LAE	WIoU	Recall rate/ %	Precision rate / %	F1 Score / %	Frame rate / (f s^{-1})	Floating-point operations	Model Memory Usage/MB
×	×	×	89.15	89.26	89.20	50.10	1.21×10^{10}	59.65
×	×	√	89.73	90.32	90.02	54.60	1.56×10^{10}	60.23
×	√	×	91.53	91.59	91.55	68.30	1.04×10^{10}	18.85
√	×	×	92.19	91.26	91.72	62.10	1.23×10^{11}	78.63
√	√	×	91.26	93.25	92.24	75.60	1.13×10^{10}	15.63
√	√	√	94.85	95.39	95.12	77.20	8.65×10^{9}	11.56

Please be aware that in the table, a \ √ \ indicates the implementation of the improvement, while a \×\ signifies that the improvement has not been implemented.

3.4 Comparison of results of different models

To assess the effectiveness of various models for this task in an unbiased manner, the following were evaluated using the dataset presented in this study: YOLO v8n, YOLO v10n, YOLO v11n, Deformable DETR, SSD, and the enhanced version YOLO v11n-EWL. The outcomes of salamander detection for each model are detailed in Table 3.

The enhanced model presented in this study demonstrates performance metrics of 94.85% in recall, 95.39% in precision, 95.12% in F1 score, and a frame rate of 77.20 f/s. These results indicate improvements of 6.29, 6.05, 6.18 percentage points, and 27.84 f/s over YOLO v10n-EWL; 9.12 points compared to YOLO v8n-EWL, with respective increases of 8.43, 8.78 percentage points, and 26.51 f/s; 5.70, 6.13, 5.92 percentage points, and 27.1 f/s over YOLO v11n; 4.84, 4.13, 4.49 percentage points, and 38.00 f/s compared to SSD; and 8.56, 6.14, 6.14, 6.18 percentage points, and 27.84 f/s against Deformable DETR. Overall, the model excels in identifying giant salamanders, meeting the requirements for real-time detection and proving suitable for outdoor applications.

Tab.3 Comparison of different model performances

Model	Recall rate/ %	Precision rate / %	F1 score / %	Frame rate / (f s^{-1})	Floating point operations	Model memory usage / MB
YOLO v8n-EWL	85.73	86.96	86.34	50.69	1.19×10^{11}	49.60
Deformable DETR	86.29	89.25	87.74	43.24	1.37×10^{10}	58.63
SSD	90.01	91.26	90.63	39.20	2.65×10^{11}	83.50
YOLO v10n-EWL	88.56	89.34	88.94	49.36	2.64×10^{11}	55.40
YOLO v11n	89.15	89.26	89.20	50.10	1.21×10^{10}	59.65
YOLO v11n-EWL	94.85	95.39	95.12	77.20	8.65×10^{9}	11.56

3.5 Visualization and analysis of comparison results

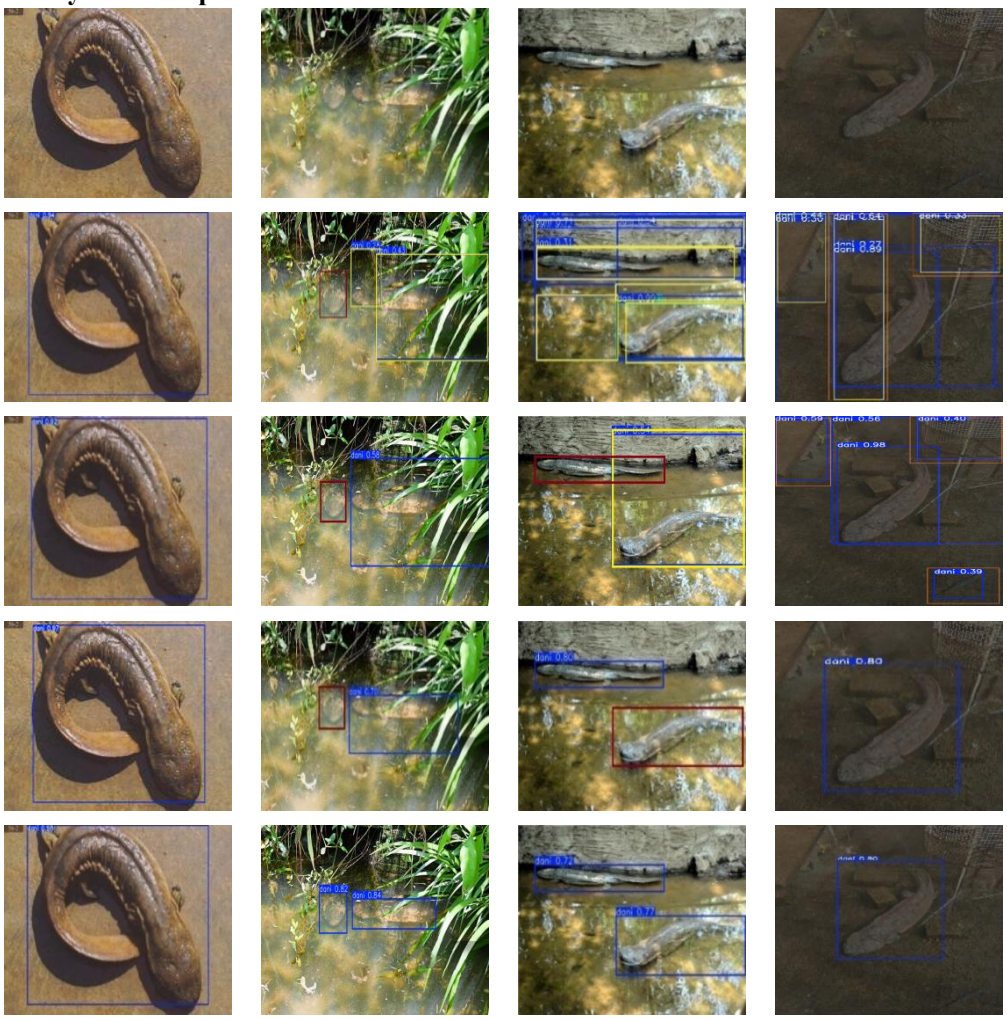


Fig. 8 Comparison of model running results before and after improvement

Figure 8 illustrates that the YOLO v8 n-EWL model struggles with accurately re cognizing individual salamander images in unobstructed conditions, often resulting in missed detections and subpar anchor frame s in intricate settings, as well as in scenar ios with multiple salamanders in simpler c ontexts. The recognition accuracy shows a gradual enhancement with the YOLO v10n -EWL and YOLO v11n models; however, missed detections persist in complex envir onments and when multiple salamanders ar e present. In low-light conditions, both Y OLO v10n-EWL and YOLO v8n-EWL ex hibit a marginally improved recognition ac curacy, yet they still encounter numerous misdetections and inferior anchor frames. Conversely, the YOLO v11n-EWL demonst rates a high detection accuracy for single salamander images even in complex enviro nments, with no missed detections or low-quality anchor frames when identifying mu ltiple salamanders. In low-light settings, it maintains high recognition accuracy withou t any misdetections or issues related to lo w-quality anchor frames. Overall, the enha nced model excels in detection tasks, sho wcasing excellent adaptability and robustne ss, while effectively analyzing salamander features, thereby confirming the viability o

f the proposed improvements.

3.6 Salamander Detection System Design and Application

3.6.1 System Implementation

The web interface is primarily designed for system operation and maintenance staff, facilitating the holistic administration of the platform across various scenarios and devices. The management backend is responsible for the complete oversight and upkeep of the entire system, encompassing functions such as configuring equipment parameters, monitoring devices, managing data, conducting site and salamander inspections, and handling historical data, as illustrated in Fig. 9.

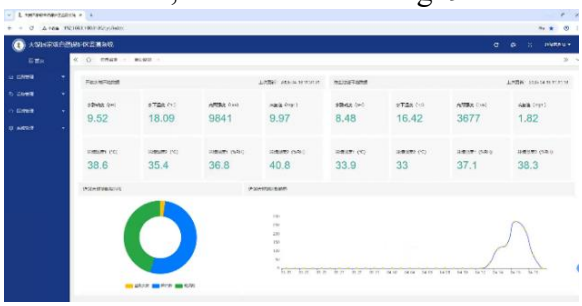


Fig. 9 Web front-end interface

3.6.2 System Application

To assess the effectiveness of the giant salamander detection system, detection equipment was deployed at the Hubei Xianfeng Zhongjian River Giant Salamander National Nature Reserve from March to June 2025. Given the rarity of giant salamanders and the unpredictability of their appearances, daily continuous monitoring was implemented in the field to test the system's reliability and the algorithm's performance. During these observations, the salamanders were counted concurrently with their detection. Data on the quantity and occurrence of giant salamanders, along with the timing of their sightings, was recorded, and environmental data was gathered to promptly evaluate the condition of the salamander population within the reserve.

To assess the system's transmission speed, images are stored on the server once the system is activated, and the video feed is fed into the salamander recognition platform. Each image is approximately 1.8 MB, with a transfer rate of around 1.68 MB/s, while the video stream operates at about 3.5 MB/s, ensuring real-time salamander detection. The trained model for salamander images, which were previously captured by a camera activated by RSSI signals, is utilized for identification through a platform developed with PyQt5. This setup allows for continuous 24-hour monitoring of the outdoor camera's video stream to recognize salamanders, enabling their documentation when they appear in the camera's field of view. The interface displaying the identified salamanders is illustrated in Figure 10, and Figure 11 features some of the wild salamanders photographed after the system reached stability.

By integrating enhancements from EMA, WIoU, and LAE, the model demonstrates consistent performance when faced with challenges like low lighting, obstructions, and variations in target size, achieving significantly lower rates of false and missed detections compared to other models. It effectively addresses critical issues such as nighttime low visibility and intricate backgrounds in conservation areas. In practical applications, the model retains the high efficacy observed in laboratory settings, successfully mitigating the high false negative rates associated with conventional techniques, and ensuring reliable identification of giant salamanders. The system's recall, precision, F1 score, and frame rate are recorded at 94.85%, 95.39%, 95.12%, and 77.

20 frames per second, respectively. Its adaptability is notable; with data re-labeling and model fine-tuning, it can be utilized for the automatic monitoring of other rare and endangered species within protected regions. While the system showed commendable performance during the testing phase, further evaluation and optimization are necessary to enhance detection capabilities under extreme conditions, such as heavy rain and flooding, which can lead to murky waters, strong currents, and significant surface reflections. In conclusion, the system has proven effective and resilient in the Hubei Xianfeng Zhongjian River Giant Salamander National Nature Reserve, though challenges remain in extreme environmental conditions.

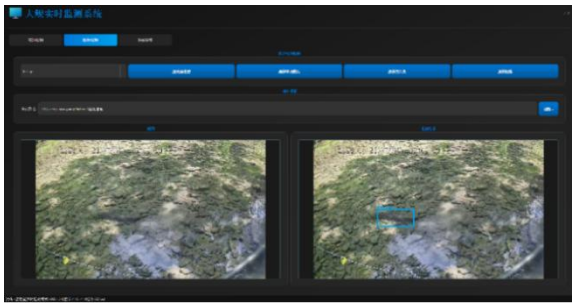
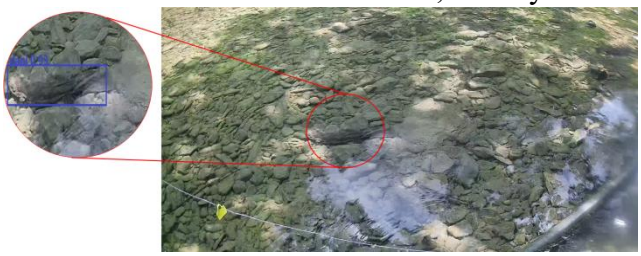


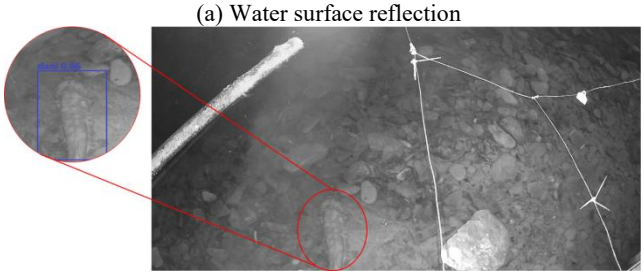
Fig 10 Giant salamander identification platform



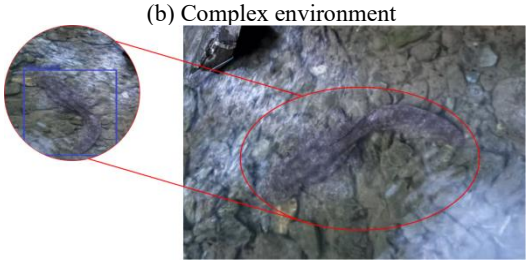
(a) Water surface reflection



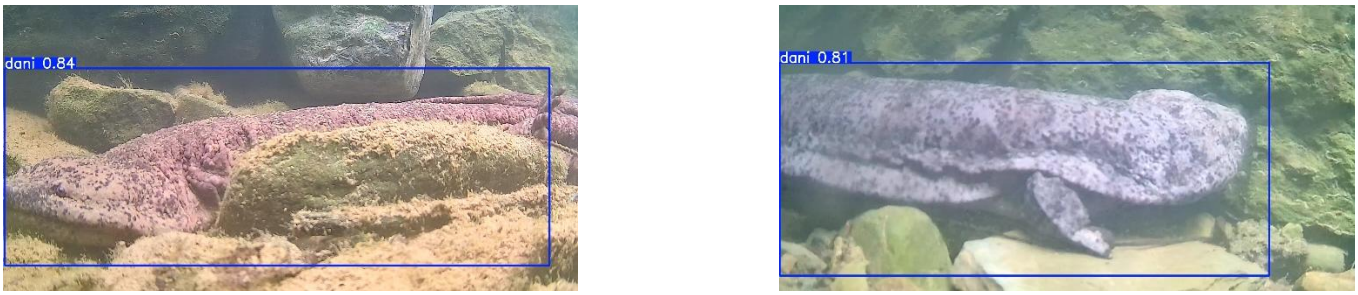
(b) Complex environment



(d) Low light at night



(e) Rainy day environment



(f) Underwater environment

Fig 11 Wild giant salamanders taken in various environments by the equipment

4. Conclusion

A new model, named YOLO v11n-EWL, has been developed for identifying and detecting giant salamanders. This model incorporates the EMA module, modifies the loss function, and includes the lightweight LAE module, resulting in improved recall, accuracy, F1 score, and frame rate compared to the original YOLO v11n model, achieving scores of 94.85%, 95.39%, 95.12%, and 77.20 frames per second, respectively. Additionally, the memory footprin

t of this model is 11.56 MB. It demonstrates consistent performance and notable robustness, maintaining the same memory usage of 11.56 MB.

In identical experimental settings, the suggested model is evaluated against the YOLO v10n-EWL and YOLO v11n models. When compared to YOLO v10n-EWL, the new model shows improvements in recall, precision, F1 score, and frame rate by 6.29, 6.05, 6.18 percentage points, and 27.84 frames per second, respectively. In relation to YOLO v11n, these metrics are enhanced by 5.70, 6.13, 5.92 percentage points, and 27.1 frames per second, all surpassing the performance of the baseline YOLO v11n model. This model demonstrates the ability to swiftly identify salamanders in various scenarios, including single and multiple salamanders, as well as in low-light and complex environments.

Creating automated systems for identifying and detecting salamanders, along with a monitoring setup for the Zhongjian River Salamander Reserve, which is capable of capturing images of salamanders outdoors over extended periods and allows for data analysis in real-time. This technology is valuable for recognizing salamanders in intricate habitats, and the identification platform addresses the need for effective salamander monitoring. By utilizing environmental data, we can gain deeper insights into the habitats of the giant salamander, guide future reserve development, formulate enhanced conservation strategies for the giant salamander, and aid in the species' recovery efforts.

Methods

Data acquisition and preprocessing

Image samples of the Chinese giant salamander were collected from the Zhongjian River Giant Salamander National Nature Reserve in Xianfeng, Hubei Province between December 2024 and February 2025. To overcome the challenge of scarcity of wild giant salamander samples, the initial collection of 1020 original images was expanded by supplementing 530 images from public network resources.To further enhance the diversity of the dataset and the generalization ability of the model, we use a variety of data enhancement techniques to enrich the dataset. Rotation, sharpening, brightness adjustment, adding noise, and HSV color space filtering are included to simulate different weather, lighting conditions, and environmental disturbances. After enhancement, the total size of the dataset reaches 6952 images, which effectively improves the robustness of the model in complex environments. All images are manually annotated using X-AnyLabeling software, and the annotation information is saved in a TXT file in YOLO format, containing the location information of the target bounding box. The final dataset is randomly divided into training set (4866 pieces), validation set (695 pieces), and test set (1391 pieces) according to a ratio of 7:2:1 to ensure the fairness and reproducibility of the experiment.

Integration of efficient multi-scale attention modules (EMA).

In order to enhance the model's ability to characterize multi-scale features, especially in complex scenes where the body color of giant salamander is mixed with the environmental background, we use the Efficient Multi-scale Attention Module (EMA) to replace the original attention mechanism in

the Backbone network. Without reducing the channel dimension, the EMA module groups the channels and extracts the attention weights through 1x1 and 3x3 volume integral branches respectively, and then captures the pixel-lev·el relationship and global context information through two-dimensional global average pooling and Sigmoid activation function, so as to significantly improve the model's focus ability on key features.

Application of the Wise-IoU (WIoU) loss function

Aiming at the problem of uneven distribution of anchor frame mass and harmful gradient in low-quality samples in field images, we replace the original CIoU loss function with the loss function of dynamic non-monotonic focusing mechanism. A small gradient gain is allocated to the low-quality anchor frame, which guides the model to optimize towards the high-quality sample more stably, improving the convergence speed and stability of training.

Introduction of Lightweight Adaptive Extraction Convolution (LAE).

In order to reduce the complexity of the model, improve the inference speed, and make it more suitable for deployment on edge devices with limited computing resources, we have introduced a lightweight adaptive extraction convolutional LAE module at the Head layer. Through group convolution, dimensional rearrangement, adaptive pooling and 1x1 convolutional operations, LAE dynamically adjusts convolutional kernel parameters and realizes adaptive feature selection, significantly reducing the amount of floating-point operations and model parameters, while maintaining strong feature extraction capabilities.

Experimental setup and evaluation indicators

All experiments were conducted on the Autodl computing power platform using an NVIDIA GeForce RTX 4090 GPU (24GB of VRAM). The deep learning framework is PyTorch 2.3.0, and the programming language is Python 3.12.3. A total of 500 epochs were trained on the model, with an initial learning rate of 0.001 and a momentum of 0.937. To comprehensively evaluate model performance, we use the following evaluation metrics: Precision (P), Recall (R), F1 Score, Frames per Second (FPS), model memory footprint, and floating-point operation.

Data availability

Data will be made available on request.

Code availability

The algorithms used for data loading, reconstructions, and digital signal processing are standard and are outlined in detail in the Methods. Python scripts can be provided by the corresponding authors upon request.

Acknowledgments

This work is supported by the China University Industry-University-Research Innovation Fund for Intelligent Internet of Things Innovation Education Special Funding Project Plan (2024WA064) and the Hubei Provincial Natural Science Foundation Program Youth Project (2020CFB337).

References

[1] WANG Cheng, JIANG Wansheng, ZHOU Qiang, et al. Preliminary investigation into the activity p atterns and displacement distances of the chinese giant Preliminary investigation into the activity patt erns and displacement distances of the chinese giant salamander (*Andrias davidianus*)[J].

[2] DENG Zhiyong, XlE Yixing, LIU Yi,et al. River health assessment of Zhangjiajie Giant Salamande r Na tional Nature Reserve in Hunan Province[J]. Water Resources & Hydropower Information,2024, 45(3):111-117,129. (in Chinese)

[3] YUAN Yuding.Problems and measures existing in the ecological environment of the Qingjiang Rive r Basin[J].Modern Science,2023(1):32-41,94. (in Chinese) Modern Science,2023(1):32-41,94.

[4] Shanghai Ocean University. A neural network-based method and system for detecting fish diseases: CN202211623709.8[P]. 2023-03-14.

[5] ZHANG Zheng,LU Xiang,HU Qingsong. Lightweight model for river crab detection based on imag e enhancement and improved YOLO v5s [J]. Transactions of the Chinese Society for Agricultural M achinery,2024,55(11):124-131,374.

[6] KE Ao,WANG Yucong,HU Boyu,et al. Review on image-based wildlife detection and recognition[J]. Computer Systems & Applications,2024,33(1):22- 36. (in Chinese)

[7] LI Daoliang,LI Wanchao,DU Zhuangzhuang. Classification method of feeding intensity of sea bass based on self-attention-DSC-CNN6 and multi-modal Transactions of the Chinese Society for Agricult ural Machinery, 2025,56(1):16-24. (in Chinese)

[8] Hu, J.; Zhao, D.; Zhang, Y.; Zhou, C.; Chen, W. Real-time nondestructive fish behavior detecting i n mixed polyculture system using deep-learning and low-cost devices. -Expert Syst. Appl. 2021, 178, 115051. [CrossRef]

[9] Xu, W.; Zhu, Z.; Ge, F.; Han, Z.; Li, J. Analysis of behavior trajectory based on deep learning in ammonia environment for fish.Sensors 2020, 20, 4425. [CrossRe]

[10] Wang, H.; Zhang, S.; Zhao, S.; Wang, Q.; Li, D.; Zhao, R. Real-time detection and tracking of fis h abnormal behavior based on improved YOLOV5 and SiamRPN ++. Comput. Electron. Agric. 202 2, 192, 106512. [CrossRef]

[11] HUSSAM E M,ALI F,OMAR A,et al. MSR-YOLO:method to enhance fish detection and tracking i n fish farms[J] Procedia Computer Science,2020,. 170:539-546.

[12] DU L, LU Z, LI D. Broodstock breeding behavior recognition based on Resnet50-LSTM with CB AM attention mechanism[J]. Computers and Electronics in Agriculture,2022,202:107404.

[13] Tu, W.; Yu, H.; Zhang, P.; Wei, S.; Zhang, X.; Yang, Z.; Wu, J.; Lin, Y.; Hu, Z. Farmed fish det ection by improved YOLOv8 based on channel non-degradation with spatially coordinated attention. J. Dalian Ocean. Univ. 2023, 38, 717.

[14] Qing S, Qiu Z, Wang W, et al. Improved YOLO-FastestV2 wheat spike detection model based on a multi-stage attention mechanism with a LightFPN detection head[J].Frontiers in Plant Science,2024,1 5:1411510.

[15] Dong X, Zhang L, Xu C, et al. Detection of pine wilt disease infected pine trees using YOLO v5 optimized by attention mechanisms and loss functions[J]. Ecological Indicators,2024,168:112764.

[16] Pan W, Chen J, Lv B, et al. Lightweight marine biodetection model based on improved YOLO v1 0[J].Alexandria Engineering Journal,2025,119:379-390.

[17] Ai J Q, et al. An SAR target classification algorithm based on the central coordinate attention mod ule[J].IEEE Sensors Journal.2023,2,24:1941-1952.

[18] Lv M, SU W H. YOLO V5-CBAM-C3TR: An optimized model based on transformer module and attention mechanism for apple leaf disease detection[J].Frontiers in Plant Science, 2024, 14: 1323301.

[19] Wang Xiaojuan, et al. AgeDETR: attention-guided efficient DETR for space target detection[J].Rem ote Sensing,2024,9,18:3452.

[20] Fu Z, et al. DLSW-YOLO v8n: a novel small maritime search and rescue object detection framew ork for UAV images with deformable large kernel net[J].Drones Drones ,2024,8(7):310.

[21] YU, Zhongwen, et al. Lsm-yolo: compact and effective roi detector for medical detection[J]. arXiv prepr int, arXiv:2408.14087,2024 .

[22] YUAN Hhongchun,LI Chunqiao. research on FDC -YOLO v8 underwater biological object detection methodImproved by deformable convolution[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(11):140-146. (in Chinese)

[23] LUO Bin,LI Jiachao,ZHOU Yanan, et al. Lightweight rice blast spores detection method based on i mproved YOLO v8[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(11):32 -38.

[24] YANG Duanli,QI Junlin,CHEN Hun, et al. Individual behavioral identification and differential analy sis of free-range laying hens based on improved YOLO v8n model[J].Transactions of the Chinese S ociety for Agricultural Machinery,2024,55(11):112-123. (in Chinese)

[25] XU Haozhe,WANG Yuanxing. Algorithm for aerial image detection of unmanned aerial vehicles bas ed on improved YOLO v10n[J]. Computer Science and Application,2025,15(1):230-238.

[26] Chen R, Zhou H, et al. YOLO-CE: an underwater low-visibility environment target detection algorit hm based on YOLO 11[J]. The Journal of Supercomputing, 2025, 81(5): 1-24.

[27] TIAN Fang,YANG Xinyao,SUN Nanqing,et al. Salamander Detection Model Based on Improved Y OLO v11n-EWL[J/OL]Transactions of the Chinese Society for Agricultural Machinery, 1-10[2025-08-12].https://link.cnki.net/urlid/11.1964.s.20250619.1059.002.

[28] LI Zonggang,SONG Qiufan,DU Yajiang,et a1. Research on robot riveting defect detection method b ased on improved DETR[J]. Journal of Railway Science and Engineering,2024,21(4):1690-1700.

[29] HU Yan,YUAN Zihao,TU Xiaoguang,et a1. improved SSD object detection algorithm based on cont rastive learning[J].Infrared Technology,2024,46(5). 548-555. (in Chinese)

[30] ZHANG Xiuzai,QIU Ye,SHEN Tao. Small pedestrian target detection in subway scene based on imp roved SSD algorithm[J]. Development,2025,62(2):397-407.(in Chinese)