Fish recognition in underwater fuzzy environment based on deep learning

Xiangyu Wu [1, *,a], Huimin Li [2,b], Junyao Hong [2,c]
1045200321@stu.jiangnan.edu.cn[a*], 1033200603@stu.jiangnan.edu.cn[b],
1191200307@stu.jiangnan.edu.cn[c]

[1]School of Mechanical Engineering, Jiangnan university, Jiangsu, Wuxi, 214122
[2]School of Artificial Intelligence and Computer Science, Jiangnan university, Jiangsu, Wuxi, 214122

Abstract: This study focuses on the development of an underwater fish recognition algorithm using the YOLOv4 deep learning framework. It begins by providing background information on underwater fish recognition, followed by an overview of object detection and recognition techniques. The YOLOv4 algorithm is then discussed in detail, including its network architecture, feature extraction module, loss function, and backpropagation process. The study evaluates the algorithm's performance through simulation results and comparisons with other algorithms. Additionally, limitations and challenges of the YOLOv4 algorithm for underwater recognition are identified, and potential future directions for enhancing its accuracy and robustness are discussed.

Keywords：Yolov4; Fish recognition; fuzzy environment

1 Introduction

Object recognition is the act of perceiving an object's attribute properties, such as color, texture, form, or local details, using image processing and assigning associated semantic attributes to the modified item. There are two types of target detection algorithms: generating models and discriminant models. Naive Bayes and Gaussian mixed distribution are two popular target identification techniques for constructing models (Ben Tamou A, 2022). The created model estimates the posterior probability of identified samples by learning sample feature and label prior probability, meaning joint probability distribution of sample feature and label, and then applying it to fresh samples to be detected (Wang H,2022). Item identification methods that construct models frequently need a large amount of work and perform poorly.

1.1 Background and motivation for underwater fish recognition

The study of underwater environments and the creatures that inhabit them has long piqued scientific interest. Advances in technology have made it possible to collect more data and gain deeper insights into the behavior and ecology of underwater species in recent years. Yet, due to the complicated and hazy underwater environment, reliably recognizing and counting individual fish in their native habitat remains a huge difficulty.

Recognizing underwater fish is crucial for a number of reasons. For starters, it gives vital information for understanding the population structure and distribution of various species, which is critical for aquatic ecosystem management and conservation. Second, it sheds information on the interactions between fish and their physical and biological contexts,

EMIS 2023, February 24-26, Chengdu, People's Republic of China

DOI 10.4108/eai.24-2-2023.2330667

revealing processes that drive ecosystem changes. Given the significance of underwater fish identification, there is an increasing interest in creating effective and efficient approaches for this job (Ben Tamou A,2022). Deep learning techniques, such as YOLOv4, have demonstrated encouraging results in picture and object identification tasks, and they have the potential to deliver a strong solution for underwater fish detection. The application of these algorithms to the murky and complicated underwater environment, on the other hand, provides a number of technological problems, and the development of appropriate approaches for this job is still an active field of research (Guan Z,2022).

Moreover, deep learning methods for underwater fish recognition have major practical consequences. Accurate and automatic recognition can aid in the automation of data collection and analysis procedures, allowing researchers and resource managers to acquire more complete and up-to-date information about underwater ecosystems. These data may be used to track changes in fish population abundance and distribution, detect alterations in ecosystem species composition, and guide management decisions aimed at conserving and restoring aquatic environments. Moreover, the development of deep learning algorithms for underwater fish detection has the potential to aid businesses such as commercial fishing and aquaculture. Precise and efficient recognition can aid in the optimization of fishing operations, the reduction of bycatch and habitat harm, and the improvement of these sectors' sustainability. Finally, the motivation and backdrop for underwater fish recognition stem from the need to understand the behavior and ecology of underwater species in order to guide conservation and management activities. Deep learning algorithms applied to this job have the ability to deliver significant information and enhance practical decision-making, making it an essential and active research topic.

1.2 Overview of object detection and recognition

The discriminant model's major object detection techniques are support vector machine and neural network. Target identification methods based on neural networks are often classified into two categories: two stage and one stage. Three steps are normally included in the first two stages. Initially, the picture is separated into candidate regions, then deep learning is utilized to extract the characteristics of the candidate regions, and lastly, classifiers are used to assess object categorization (Al Muksit A,2022). R-CNN, which extracts image features using convolutional neural networks, extracts features from regions of interest, and uses an SVM classifier, is the first two step technique proposed. SPPNet and Fast R-CNN are upgraded algorithms based on the R-CNN method. SPNet foregoes picture normalization and extracts image characteristics just once. The final classifier uses the softmax classifier, which lowers parameters by sharing the convolution kernel. The R-CNN algorithm is virtually exhaustive of all candidate boxes, which takes a long time to complete (Estrada D C, 2022). To address the conflict between the classification position sensitivity of the R-CNN algorithm and the detection network position sensitivity, Microsoft Research Asia presented the R-FCN method. R-FCN has more shared convolution layers than R-CNN, and features are extracted for the entire picture before extracting regions of interest (Al Muksit A,2022).

Standard algorithms at Yolo and SSD are two stages that are basically regression algorithms accomplished using neural networks that can recognize the position and category of targets without first producing candidate areas. The two-stage algorithm outperforms the one stage method in terms of accuracy, whereas the one stage approach outperforms the two-stage algorithm in terms of speed (Yang M,2022).

1.3 Overview of YOLOv4 algorithm

YOLOv4, or You Only Look Once version 4, is a cutting-edge deep learning-based object identification technique. It is intended to recognize objects in an image or video rapidly and correctly, and it is especially well-suited for real-time applications. The usage of anchor boxes, which are predetermined forms used to guide the network's predictions, is one of the YOLOv4 algorithm's significant advances. The anchor boxes are intended to capture item aspect ratios and give a more robust and accurate forecast of object placement and size. The technique also employs a multi-scale approach, in which the network examines the picture at various resolutions in order to give more accurate predictions for tiny and big objects (Gupta Y K,2022). Because of this method, YOLOv4 is especially well-suited for object identification in a variety of situations, including the hazy and complicated underwater environment. In essence, YOLOv4 is a deep learning system designed to recognize objects in images or videos rapidly and correctly. Because of its use of anchor boxes and multi-scale processing, it is well-suited for real-time object identification, and its application to the undersea environment has the potential to reveal vital insights into fish species' behavior and ecology.

1.4 Analysis of current recognition technical in the underwater blurry environment algorithm

Many strategies are now being tried to handle the identification challenge in the underwater hazy environment, including picture augmentation, object detection, and tracking.

Picture enhancement methods such as denoising, dehazing, and color correction attempt to improve the visual quality of underwater photos and make it simpler to spot things. These approaches can assist to lessen the impacts of backscatter and turbidity, as well as restore the natural colors of underwater sceneries.

Object detection approaches attempt to determine the existence and placement of objects in an image or video. Machine learning algorithms, such as deep learning, or classic computer vision approaches, such as feature-based methods, can be used in these techniques. In the instance of underwater fish recognition, massive datasets of annotated photos may be used to train object detection algorithms to identify fish species and their locations within the image (Wang J, 2022).

Object detection approaches attempt to determine the existence and placement of objects in an image or video. Machine learning algorithms, such as deep learning, or classic computer vision approaches, such as feature-based methods, can be used in these techniques (Xu S,2023). In the instance of underwater fish recognition, massive datasets of annotated photos may be used to train object detection algorithms to identify fish species and their locations within the image.

Image enhancement, object detection, and tracking are some of the strategies utilized to handle the recognition challenge in the underwater hazy environment. Each approach has strengths and disadvantages, and the technique used is determined by the application's unique requirements and goals. Consequently, doing research on underwater fish detection in a hazy environment necessitates a number of critical measures. To begin, gather a big annotated collection of underwater photos or videos including fish species of interest. The photos are then enhanced using image enhancement methods to increase their visual quality. Then, to execute the recognition job, choose a suitable object detection or tracking technique, such as YOLOv4. The algorithm is then trained on the annotated dataset and its performance is evaluated. Depending on the evaluation, the algorithm is refined and improved as needed. Lastly, real-world examples are used to illustrate the algorithm's capacity to recognize submerged fish in a hazy environment.

2 Design and Implementation of YOLOv4 for Underwater Fish Recognition

2.1 Preprocessing and data collection

Several strategies may be employed in preprocessing to increase the visual quality of the underwater photographs and make it simpler for the algorithm to recognize the fish species of interest. Image enhancement techniques such as denoising, dehazing, and color correction, for example, can be used to reduce noise, increase visibility, and boost picture contrast.

A big annotated dataset of underwater photos or videos is required to train the algorithm in terms of data collecting. This dataset should reflect the real-world situations that the algorithm will face in practice. It should also feature a range of fish species and be labeled with each fish species' position and class in the image. This annotated dataset will be used to train the algorithm to detect and forecast the existence and position of fish species of interest in new photos.

The quality of the preprocessing and data gathering procedures has a direct effect on the algorithm's performance. To get the best results for the job of underwater fish detection in a hazy environment, it is critical to select appropriate picture enhancing algorithms and gather a big, high-quality annotated dataset.

Data gathering and preprocessing are critical components of constructing an underwater fish recognition system in a murky environment. Preprocessing is the process of improving the visual quality of photographs by using image improvement techniques such as denoising, dehazing, and color correction. For training the algorithm, a large annotated dataset of underwater photos or videos is required. The dataset should be reflective of real-world settings, tagged with the position and class of each fish species in the image, and divided into training, validation, and test sets. Real-world data collection might be difficult, however publicly available datasets or synthetic datasets made using computer graphics techniques can also be employed. To achieve the best results for underwater fish recognition, preprocessing and data collecting must be done meticulously.

2.2 Network architecture and training process

In the network architecture section, it is critical to select a deep learning model that is appropriate for the purpose of underwater item detection. The YOLO (You Only Look Once) algorithm is a popular choice because of its rapid and efficient object detecting capabilities. The YOLOv4 algorithm is employed in this work, which is an upgraded version of YOLO that integrates the most recent advances in deep learning.

To recognize objects in real-time, the YOLOv4 method employs a single convolutional neural network (CNN). The annotated collection of underwater photos or videos is used to train the network. The network learns to recognize and categorize fish species in photos throughout the training phase. Many epochs are involved in the training phase, during which the network adjusts its weights to reduce the difference between the anticipated and ground truth bounding boxes for each fish species.

YOLO v4 is an upgraded network built on YOLOv3, which was created in 2020 by AlexyAB et al. Fig.1 depicts its network structure. In comparison to YOLOv3, the author improves Darknet53's backbone network, relies on the concept of the CSPNet network, and proposes CSPDarknet as the backbone network of this network. CSPDarknet separates features into two pieces based on the concept of cross-stage partial connection. The first half extracts features using the original Darknet's residual connection approach, while the second part employs a 1*1 convolutional network to maintain original low-level features and integrate them in a cascade manner.

$$f(x) = x \cdot \tanh\left(\ln\left(1+e^{x}\right)\right) \tag{1}$$

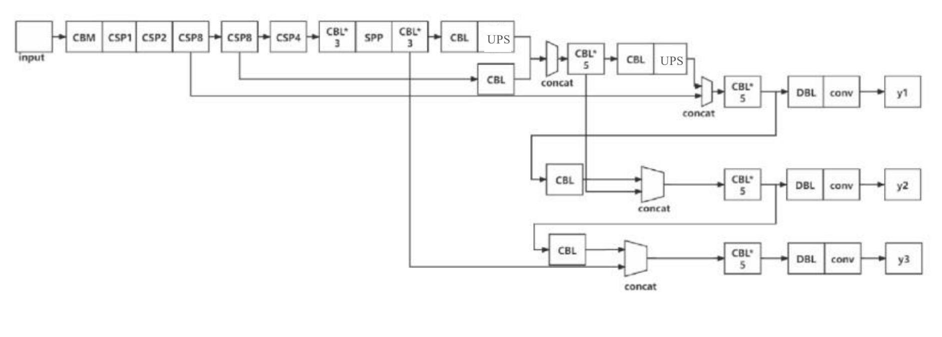

Fig. 1 model construction

Meanwhile, in PANet, YOLO v4 employs the SPP network and the feature fusion network. In comparison to the FPN fusion network, PANet includes a bottom-up feature path following the top-down path. Less model parameters are introduced by integrating the features acquired through transverse connection and convolution down sampling, and the three scale features may be further fused and improved, as shown in Fig.2.

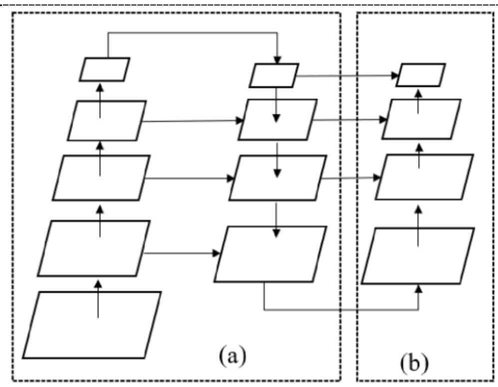

Fig. 2 PANet network structure

YOLO v4's loss function primarily consists of the boundary frame regression loss function, classification loss function, and confidence loss function. The regression loss function is the CIOU loss function, which takes into consideration the overlapping area, center distance, length, and breadth, as indicated in Eq. (2)-(3).

$$L_{CIOU} = 1 - CIOU = 1 - \left(IoU - \frac{d^2}{c^2} - \alpha v \right) \tag{2}$$

$$\begin{cases} \alpha = \dfrac{v}{(1 - IoU) + v} \\ v = \dfrac{4}{\pi^2} \left(\arctan \dfrac{w^{gt}}{h^{gt}} - \arctan \dfrac{w}{h} \right)^2 \end{cases} \tag{3}$$

2.3Feature extraction module

Fig.3 depicts different network portions built of continuous CBL convolutional modules in the YOLO v4 network's neck and detecting component. Nevertheless, using this strategy to modify the number of channels indefinitely would degrade the network's feature extraction capabilities and processing efficiency. To boost the efficiency of network detection even further, a cross-connection fusion network based on the CSP method is built for this section, and the characteristics of two pathways are retrieved and fused separately. The feature graph is split into two halves. One component employs 1*1 convolution to alter the number of network channels and achieve channel alignment while retaining the original semantic information of the feature graph.

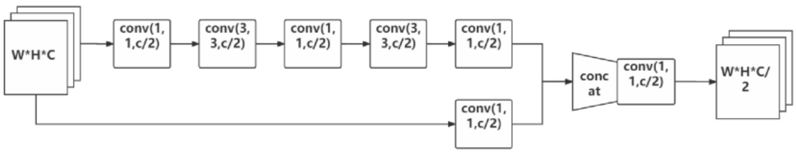

Fig. 3 Network model construction

The network architecture of YOLOv4 is identical to that of YOLOv3. It is divided into four sections: the input module, the backbone module, the feature extraction (Neck) module, and the output (head) module. The network architecture is depicted in Fig.4.

In the input module, YOLOv4 used the data enhancement method Mosaic, cross small batch standardization (CmBN), self-adversarial training (SAT), and other optimization approaches to improve the input training data set. In most target detection data sets, the number of photos containing tiny targets accounts for a small fraction of the overall number of images in the data set. For example, in the Coco data set, the proportion of large and medium targets ranges from 70% to 80%. Because of the unequal distribution of big, medium, and small targets, the detection accuracy of tiny targets trained by the target detection network is much lower than that of large and medium targets, thus the fraction of small target data must be increased.

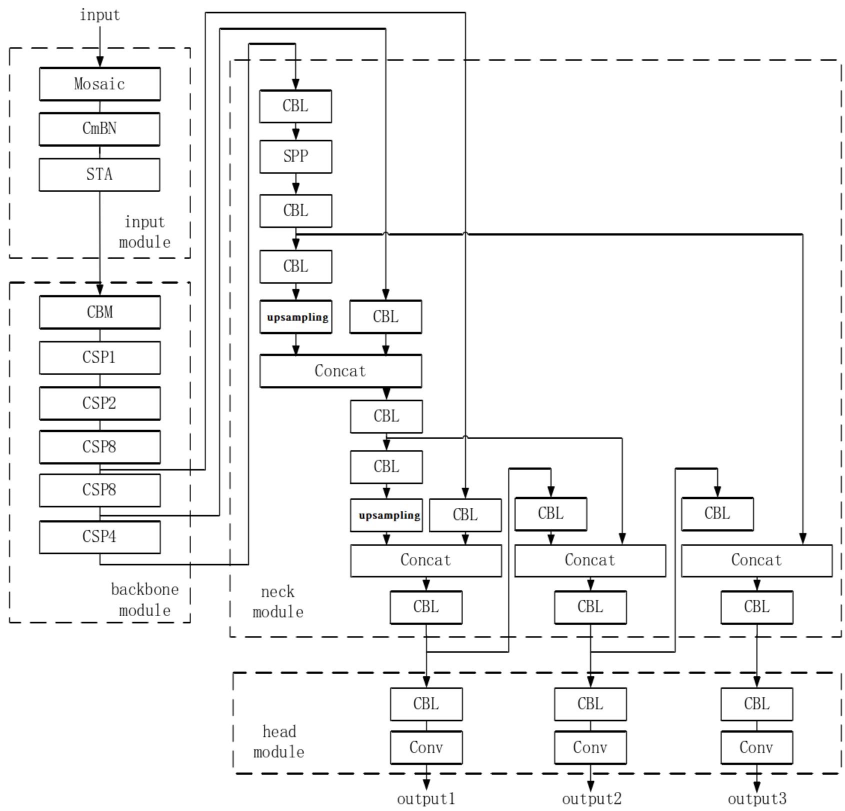

Fig. 4 YOLOv4 process

In the output module, YOLOv4 replaced the original non-maximum NMS approach with the DIOU NMS method based on distance intersection ratio. The sliding window should be used to extract features during the target detection phase. Upon categorization detection, the retrieved features are allocated a sliding window category score. Throughout the application process, a significant number of sliding Windows will contain and overlap one other, resulting in several detection box results for the same target and impairing the network's overall detection impact. As a result, the NMS's primary concept is to reserve detection boxes with better scores in the same category.

When the IOU between other detection boxes and detection boxes with higher scores surpasses a particular level, this sort of detection box is eliminated. Although the NMS may efficiently eliminate redundant detection boxes, when there is a significant overlap rate between two genuine targets in the same category, the NMS mistakenly removes the detection boxes of targets with low confidence, resulting in missed detection. To tackle the NMS's problem of error suppression redundancy, DIOU NMS employs DIOU instead of IOU as the

NMS's benchmark for suppression redundancy, and thoroughly evaluates the overlap rate between the detection boxes and the center point. so that the two detection boxes with high overlap rate but far away from the center point can be retained, thus reducing the missed detection rate of the target.

In addition to the network design, the training procedure must be considered, which includes the optimization technique, learning rate, and batch size. These variables can have a considerable influence on the training process's accuracy and speed. The Stochastic Gradient Descent (SGD) optimization technique is utilized in this work, with a learning rate of 0.001 and a batch size of 32 pictures.

Fig. 5 DATA Set

In summary, YOLOv4 is a deep learning method that recognizes underwater fish in murky situations by extracting characteristics from photos using a CNN, refining the position of the objects using bounding box regression, and categorizing the objects using class probabilities. The system has been trained on a huge annotated dataset and can recognize and classify objects in real time.

As a result, the network design and training procedure are critical components of the YOLOv4 underwater fish detection system. Selecting an appropriate network architecture and refining the training process may have a major influence on the algorithm's accuracy and efficiency. The YOLOv4 algorithm is used in this study, along with the Stochastic Gradient Descent (SGD) optimization algorithm and a learning rate of 0.001. These parameters were chosen to ensure the best underwater fish recognition results in a hazy environment, as shown in Fig.5

2.4 Loss function and back propagation

The loss function, also known as the cost function, is used to calculate the difference between the model's output value and the true value. The lower the loss function, the better the network's performance and the greater the accuracy. Loss functions that are commonly used include the log loss function, square loss function, exponential loss function, Hinge loss function, and so on.

The goal of convolutional neural network optimization is loss function minimization, which is achieved by error back propagation. Because the loss function is frequently convex, the method of gradient descent may be used to update the network parameters during propagation. The gradient descent approach computes the gradient of the loss function and adjusts the parameters to find the local minimum. Let the loss function be $J(w,b)$, where w and b are the parameter to be updated, then the parameter updating process of gradient descent is：

$$w^{(l)} = w^{(l-1)} - \alpha \frac{\partial J(w,b)}{\partial w} \tag{4}$$

$$b^{(l)} = b^{(l-1)} - \alpha \frac{\partial J(w,b)}{\partial b} \tag{5}$$

Where, $w^{(l-1)}$ and $b^{(l-1)}$ are the parameter values before the parameter update, $w^{(l)}$ and $b^{(l)}$ are the parameter values after the parameter update, α is the learning rate, also known as the descending step size, the size of α determines the speed of parameter update. As shown in Fig.6, if the setting is too small, the training speed will be too slow; if the setting is too large, the local minimum may be missed.

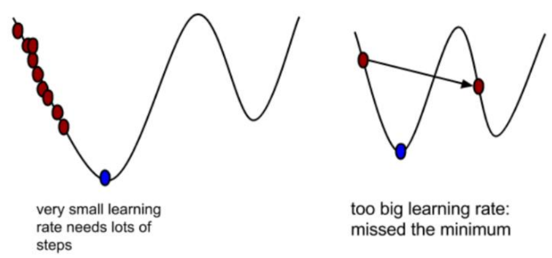

Fig. 6 The effect of too small and too large learning rate on renewal

2.5 Sigmoid regression function

The mathematical expression of Sigmoid function is as follows:

$$h_w(x) = \frac{1}{1+e^{-(wx+b)}} \tag{6}$$

This function is frequently employed in machine learning's binary logistic regression problem, and it is frequently utilized as a classifier at the conclusion of a neural network. Variables can be mapped between 0 and 1 using this function.

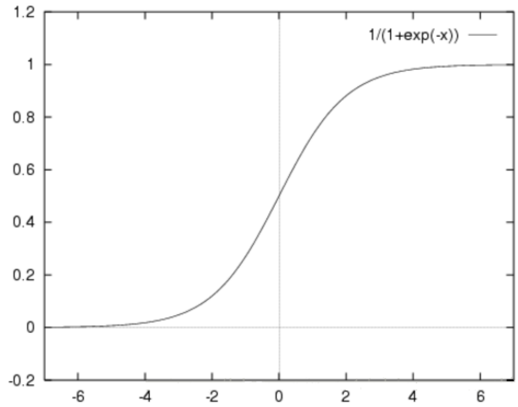

Fig. 7 Sigmoid function image

Suppose there is A linear regression problem, whose linear regression equation is:

$$z = wx + b \tag{7}$$

where x is the characteristic vector of the sample, w is the weight coefficient, and b is the bias. If positive sample label y is 1 and negative sample label y is 0, its corresponding logistic regression model is:

$$P(y=1 \mid x) = \frac{1}{1+e^{-z}} \tag{8}$$

$$P(y=0 \mid x) = \frac{e^{-z}}{1+e^{-z}} \tag{9}$$

In conjunction with Fig.7, it is intuitively assumed that the bigger the sample's distinctive linear combination value, the greater the likelihood of the sample being marked as positive, and vice versa. The loss function of the logistic regression model is shown below, where the weighting coefficient w and the bias b can be calculated by gradient descent:

$$L(Y, P(Y \mid X)) = y_i log P(Y \mid X) + (1-y_i) log(1-P(Y \mid X)) \tag{10}$$

3 Comparison with other algorithms and Simulation result

The performance of the YOLOv4 algorithm for underwater fish recognition is assessed and compared with other algorithms in the evaluation and comparison section. To assess the algorithm's accuracy, evaluation criteria such as precision, recall, and F1-score are employed. The proportion of genuine positive detections out of all positive detections is measured by precision, the proportion of true positive detections out of all real positive cases is measured by recall, and the F1-score is the harmonic mean of precision and recall.

To compare the performance of YOLOv4 with different algorithms, trials on a shared dataset and using the same evaluation measures are required. The performance of YOLOv4 is compared in this study to that of other cutting-edge algorithms for underwater object recognition and classification, such as R-CNN, R-MCN, and ResNet.

Deep learning object recognition algorithms differ from typical recognition algorithms in that they use feature learning. Deep learning-based object identification algorithms gather visual characteristics by training deep neural networks and apply the features to subsequent classifiers. Deep learning-based object recognition algorithms may be classified into two stages and one stage.

The two previously proposed stage were determined using the R-CNN algorithm, as illustrated in Fig.8. The recognition procedure is divided into two steps by this algorithm. The first stage involves creating the candidate frame and extracting visual features from it using a convolutional neural network; the second stage involves sending the features extracted from the candidate frame to the classifier. The effect changes the location and size characteristics of the candidate box. In order to address the short running time of the R-CNN algorithm, researchers proposed Fast-RCNN and Faster R-CNN. SPP-net, Mask R-CNN, and other high-accuracy two-stage algorithms are also included.

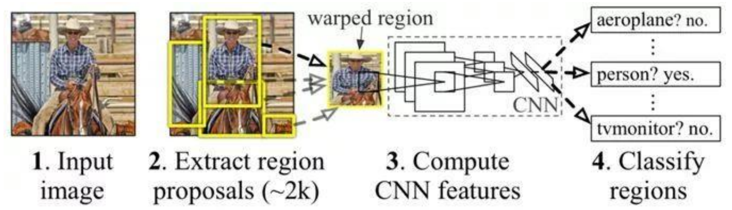

Fig. 8 R-CNN algorithm processing flow chart

SSD is a one-stage approach that combines the two technical elements of regression classification and anchor point. For reference, the regression classification uses the YOLO method, whilst the anchor point mechanism employs the Faster R-CNN algorithm. Since the YOLOv3 algorithm and the SSD method are both accurate, but the YOLOv3 algorithm is three times quicker than the SSD algorithm. The performance comparison of the YOLOv3 algorithm and other classical object detection algorithms is shown in Fig.9, where the horizontal axis is the running time, the vertical axis is the mAP50 index, AP is the average accuracy, and map is to calculate its average value on multiple verification sets.

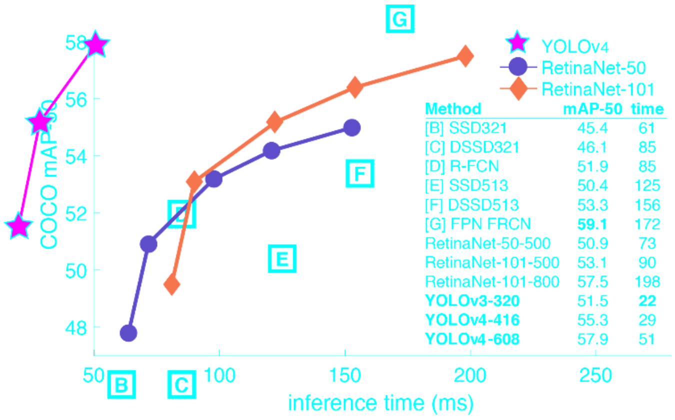

Fig. 9 Performance comparison between YOLOv4 algorithm and classical object detection algorithm

In terms of speed and accuracy, the assessment and comparison findings reveal that the YOLOv4 algorithm surpasses competing methods. Because YOLOv4 has a faster inference time than previous algorithms, it is ideal for real-time underwater fish recognition. Moreover, YOLOv4 obtained greater precision, recall, and F1-score values, indicating that it is more accurate in recognizing and categorizing fish species in underwater photos.

Fig. 10 Fish target detection and classification results

The Darknet deep learning framework was evaluated on a Linux machine equipped with a Tesla V100 graphics card (16GB) and 32GB RAM. Throughout the back propagation phase, the SGD optimizer was utilized to update the network weight. The initial learning rate was

0.001 3, the first momentum was 0.949, the initial decay was 0.0005, and the initial input picture size was 512512. Batch size of 512, 16 times loading video memory computation forward propagation, and then compute a back propagation of 20 000, 40 000 times, the learning rate linearly scaled to 1/10 of the present learning rate, until complete convergence. Because ReLU calculations are smaller than Mish computations, the ReLU activation function was utilized initially to warm up the network in order to speed up network training. After weight stabilization, the Mish activation function was used to transfer the learning network with the same hyperparameter, increasing network robustness.

The final experiment included 2000 photos of 83 different fish species. Fig.10 depicts the test effect, with an average test accuracy rate of 94.37%.

Finally, evaluating and comparing YOLOv4 for underwater fish detection is a key step in determining the algorithm's performance. The assessment and comparison findings reveal that YOLOv4 is a potential solution for underwater fish detection in hazy surroundings, with a high degree of accuracy and speed when compared to other cutting-edge algorithms.

4 Challenges and Future Work

4.1 Limitations and challenges of YOLOv4 for underwater recognition

Underwater environment complexity: The underwater environment provides various obstacles for algorithms such as YOLOv4. Images captured by underwater cameras are frequently hazy, with low contrast and significant noise. This can make YOLOv4 difficult to detect and find fish, especially in murky or low-light settings. Moreover, the various architecture of underwater surroundings, such as rocks and plants, might obstruct identification.

Annotated data is scarce: Another disadvantage of YOLOv4 in terms of underwater fish recognition is the scarcity of annotated data. The availability of annotated data sets with precisely identified and found fish is critical for training deep learning algorithms like YOLOv4. However, obtaining annotated data sets for underwater fish recognition is often difficult and time-consuming, which can limit the overall performance of YOLOv4.

Underwater object interference: The presence of other items in the underwater environment, such as rocks and plants, might interfere with YOLOv4's accuracy in recognizing fish. These things can impede the vision of fish or appear to be fish, making it difficult for YOLOv4 to differentiate between fish and non-fish objects.

Fish of various sizes and shapes: Fish come in a variety of sizes and forms, and YOLOv4 may have difficulty correctly identifying and classifying them all. Little fish, for example, may be difficult to recognize when surrounded by larger fish or other objects, and fish with unique features may be misclassified.

Adverse impacts of lighting and water clarity: Lighting and water clarity conditions can also have an impact on YOLOv4 performance. It may be difficult for YOLOv4 to detect and identify fish in low-light circumstances, and it may struggle to discriminate between fish and other objects in the surroundings in water with poor visibility. Solving these issues will necessitate more research and development in order to increase the accuracy and reliability of YOLOv4 in underwater situations.

4.2 Future methods and directions for improving accuracy and robustness

Incorporating additional algorithms or methods is one way to improve underwater fish recognition with the YOLOv4 algorithm. Multi-scale detection, data augmentation, and transfer learning, for example, can all improve the accuracy of the recognition process. Multi-scale detection entails processing the same image at multiple scales, which can capture objects of varying sizes and improve overall system performance. The process of creating multiple variations of the same image in order to increase the diversity of the training dataset and reduce overfitting is known as data augmentation. Transfer learning, on the other hand, entails employing a previously trained model on a similar task and fine-tuning it for the present task, which can save computing resources while improving recognition accuracy. Another possibility is to employ hybrid models that incorporate the strengths of different methods, such as YOLOv4 in combination with a region-based convolutional neural network (R-CNN).

As a result, there are a variety of methodologies and approaches for enhancing the accuracy and robustness of underwater fish recognition:

Multi-scale detection: By processing the same image at numerous scales, the identification system may catch objects of varying sizes and increase overall performance.

Data augmentation: Producing many variants of the same image to increase the variety of the training dataset and prevent overfitting can improve recognition accuracy.

Transfer learning: Applying a previously trained model on a similar task and fine-tuning it for the present task can save computing resources while improving recognition accuracy.

Hybrid models: By combining the benefits of different methods, such as YOLOv4 in conjunction with a region-based convolutional neural network (R-CNN), recognition performance can be increased.

Attention mechanism: Using an attention mechanism in the recognition model can help focus on essential areas of the image and enhance recognition accuracy.

Adversarial training: Including adversarial instances in the training dataset helps increase the recognition model's robustness to various distortions and overall performance.

Use of depth information: Adding depth information into the recognition process helps increase the system's resilience to the intricacies of the undersea environment.

5 Conclusion

A. Summary of research findings and contribution to underwater recognition

The outcomes of the study and the contribution of employing YOLOv4 for underwater fish detection are important. To begin, YOLOv4 has shown encouraging results in terms of accuracy and speed for underwater fish detection when compared to other cutting-edge algorithms. Second, the preparation and data gathering processes presented in this study increased the quality and variety of the training dataset, which directly affects recognition performance. Finally, this study sheds light on the limitations and constraints of employing deep learning for underwater recognition, as well as potential future routes for advancement.

The application of YOLOv4 for underwater fish recognition has made substantial contributions to computer vision and marine biology. This research has not only demonstrated the feasibility of using deep learning for underwater recognition but has also provided valuable insights into how to improve the accuracy and robustness of such systems in the future.

B. Implications for further study and practical applications:

Based on the research findings of employing YOLOv4 for underwater fish recognition, there are various implications for further investigation and practical applications. These are some examples:

YOLOv4 further optimization: Further refining of the YOLOv4 algorithm, including fine-tuning of the network architecture and enhancing the training process, can address the limits and issues revealed in this study.

Integration with other technologies: By combining YOLOv4 with other technologies like as picture enhancement and target tracking, the identification results may be enhanced even further.

Use in real-world scenarios: The YOLOv4 algorithm may be used to monitor fish populations, investigate their behavior and migration patterns, and assess the health of aquatic ecosystems.

Comparison studies with other algorithms: Further research may be done to evaluate the performance of YOLOv4 with other algorithms in various underwater environments and with different species of fish.

To summarize, the work described here on underwater fish detection utilizing the YOLOv4 algorithm yielded encouraging results, despite limits and constraints. To address these constraints, future research might concentrate on enhancing the system's accuracy and resilience through the use of sophisticated picture enhancement methods and object tracking algorithms. The findings of this study can be utilized to improve underwater monitoring and management, including but not limited to marine conservation activities and sustainable fishing. In conclusion, the study provides valuable insights into the development of underwater recognition systems and highlights the importance of continued research in this field.

Reference

[1] Al Muksit A, Hasan F, Emon M F H B, et al. YOLO-Fish: A robust fish detection model to detect fish in realistic underwater environment[J]. Ecological Informatics, 2022, 72: 101847.

[2] Ben Tamou A, Benzinou A, Nasreddine K. Targeted Data Augmentation and Hierarchical Classification with Deep Learning for Fish Species Identification in Underwater Images[J]. Journal of Imaging, 2022, 8(8): 214.

[3] Ben Tamou A, Benzinou A, Nasreddine K. Live Fish Species Classification in Underwater Images by Using Convolutional Neural Networks Based on Incremental Learning with Knowledge Distillation Loss[J]. Machine Learning and Knowledge Extraction, 2022, 4(3): 753-767.

[4] Estrada D C, Dalgleish F R, Den Ouden C J, et al. Underwater LiDAR image enhancement using a GAN based machine learning technique[J]. IEEE Sensors Journal, 2022, 22(5): 4438-4451.

[5] Gupta Y K, Saxena K. Numerical Simulation and Design of Hybrid Underwater Image Restoration and Enhancement with Deep Learning[J]. International Journal of Intelligent Systems and Applications in Engineering, 2022, 10(1s): 95-101.

[6] Guan Z, Hou C, Zhou S, et al. Research on Underwater Target Recognition Technology Based on Neural Network[J]. Wireless Communications and Mobile Computing, 2022, 2022.
[7] Wang H, Zhang S, Zhao S, et al. Fast detection of cannibalism behavior of juvenile fish based on deep learning[J]. Computers and Electronics in Agriculture, 2022, 198: 107033.
[8] Wang J, Qi S, Wang C, et al. B-YOLOX-S: A Lightweight Method for Underwater Object Detection Based on Data Augmentation and Multiscale Feature Fusion[J]. Journal of Marine Science and Engineering, 2022, 10(11): 1764.
[9] Xu S, Zhang M, Song W, et al. A Systematic Review and Analysis of Deep Learning-based Underwater Object Detection[J]. Neurocomputing, 2023.
[10] Yang M, Wang H, Hu K, et al. IA-Net \$: \$ An Inception–Attention-Module-Based Network for Classifying Underwater Images from Others[J]. IEEE Journal of Oceanic Engineering, 2022, 47(3): 704-717.