International Core Journal of Engineering Volume 10 Issue 12, 2024
ISSN: 2414-1895 DOI: 10.6919/ICJE.202412_10(12).0009

A Lightweight Ship Target Detection Method based on Improved YOLOv8

Shiquan Yuan, Jianming Zheng

Department of Jilin Institute of Chemical Technology University, Jilin 132000, China

Abstract

In the realm of modern maritime and inland shipping management, the demands for sophisticated oversight are continually escalating. With the seamless fusion of deep learning and computer vision, ship target detection has emerged as a pivotal detection method within shipping management. To address the challenges posed by varying ship sizes and high quantities in complex scenarios, where traditional computer vision methods suffer from inadequate detection accuracy and sluggish recognition speeds, this paper introduces a lightweight ship target detection approach, CL-YOLO, based on an improved YOLOv8 model. Initially, we integrate the Coordinate Attention mechanism with the C2f module, decoupling and optimizing the channel attention mechanism. By individually fusing features across two distinct spatial dimensions, we devise an efficient one-dimensional feature encoding strategy that alleviates the model's computational burden. Furthermore, to enhance the model's detection efficiency, we incorporate the Large Kernel Attention (LKA) mechanism into the Backbone, significantly boosting detection accuracy without altering the parameter count. Upon experimenting with the optimized model on the SeaShips dataset, we observed an 1.1% increase in mean Average Precision (mAP) compared to the unoptimized YOLOv8, achieving a remarkable 97.5%. This underscores CL-YOLO's superior accuracy in detecting inland ships.

Keywords

YOLOv8; Object Detection and Recognition; Attention Mechanism; Ships.

1. Introduction

Ship target detection holds immense importance in enhancing maritime safety, boosting shipping efficiency, safeguarding the marine environment, and fostering international maritime cooperation. Firstly, precise ship detection technology can markedly decrease the incidence of maritime accidents, particularly in congested harbors and inland waterways. By enabling real-time monitoring of ship movements, target detection technology substantially mitigates the risks of collisions and groundings. Secondly, automatic ship identification technology optimizes resource allocation and enhances safety management through traffic statistics and identity verification, thereby augmenting the operational efficiency of ports. Furthermore, ship target detection technology plays a pivotal role in monitoring illegal activities, including illegal fishing, sand mining, and waste dumping, aiding in the enforcement of international regulations and the preservation of marine ecosystems.

As artificial intelligence technology rapidly advances, ship target detection methods grounded in deep learning, notably the implementation of the YOLO model, have emerged as a focal area of research and application. This sophisticated technological means facilitates automatic identification and localization of ships, thereby markedly enhancing monitoring efficiency and accuracy. It offers robust technical support across diverse domains, including maritime traffic management, maritime rescue operations, environmental conservation, and international maritime cooperation.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202412_10(12).0009

Target detection methods leveraging deep learning technology have come to occupy an increasingly prominent role in contemporary industrial applications, thanks to their exceptional ability in feature extraction and multi-task processing. These methods can be broadly classified into two primary categories: The first category encompasses two-stage detection algorithms, exemplified by Fast R-CNN [1] and Faster R-CNN [2]. While these algorithms excel in detection accuracy, they tend to have a slower detection process. The second category comprises single-stage detection algorithms, such as YOLO [3], SSD [4], and RetinaNet [5]. These algorithms rely on a single network to directly predict the location and class probabilities of targets, offering a notable increase in detection speed compared to two-stage algorithms. In scenarios involving ship detection and recognition, single-stage algorithms are particularly advantageous.

In recent years, numerous scholars have conducted extensive research on target detection for ship identification tasks [6-11]. Notably, Gao [12] and his team proposed an innovative method that integrates a dynamic denoising module with YOLOv4 to construct a novel feature pyramid network, thereby enhancing the model's representation capabilities. Huang Wanli [13], inspired by the attention mechanism and dilated convolution, designed three modules and seamlessly embedded them within the YOLOv7 network structure, ultimately boosting the efficiency of the detection algorithm. Yin [14] and his colleagues leveraged two enhanced versions of YOLOv4 and YOLOv5s networks to achieve real-time and precise end-to-end detection. Woo and his team [15] introduced ConNextV2, which can be seamlessly integrated into YOLO, significantly elevating the network's performance across various recognition benchmarks. Wang et al. [16] aimed to improve the detection capabilities for small targets at sea and enhance the semantic information of high-order features in the model by incorporating the SimAM attention mechanism and Transformer structure into the YOLOv5 algorithm. Zhang et al. [17] utilized the YOLOv5 model in conjunction with a dark channel dehazing algorithm, SE attention mechanism module, and an improved non-maximum suppression technique to facilitate efficient ship identification in foggy conditions, thereby enhancing recognition capabilities in foggy and rainy weather. However, the model's ability to identify small targets still leaves room for improvement. Zhang et al. [18] significantly bolstered the YOLOv5 model's capacity to identify and locate small fishing boats by integrating the Convolutional Block Attention Module (CBAM) and Bidirectional Feature Pyramid Network (Bi-FPN) structure. Despite these advancements, with the increasing demand for multi-scale performance of the model, there is still potential for further optimization. These models are lightweight, easy to deploy, and exhibit robust performance in complex environments.

Despite advancements in target detection technology, several challenges remain. Firstly, it is difficult for models to accurately locate and identify targets in complex backgrounds. Secondly, balancing the computational cost and performance of models poses a challenge. To address these issues, this paper introduces an improved detection algorithm called "CL-YOLO," based on YOLOv8, which is designed to achieve dual optimization in terms of model lightweightness and detection performance.

2. Methodology

To tackle the challenge of YOLOv8's suboptimal performance in detecting ships amidst complex near-shore backgrounds, a high density of small targets, and substantial scale variations, this paper introduces an enhanced river ship detection and recognition approach, named CL-YOLO, which builds upon the YOLOv8 model.

2.1 Introduction to YOLOv8 Basics

The YOLOv8 model is structured around two core components: the detection network and the classification network. Its detailed architecture encompasses four vital parts: Input, Backbone, Neck, and Head, as depicted in Figure 1. In comparison to its predecessors, YOLOv8 showcases notable advancements and refinements in several respects. Firstly, the Backbone and Neck sections have been innovatively enriched with the C2f module, which amplifies the gradient flow. Secondly, for loss computation, YOLOv8 utilizes the Task Aligned Assigner approach for positive sample allocation

ISSN: 2414-1895 DOI: 10.6919/ICJE.202412_10(12).0009

and incorporates the Distribution Focal Loss, facilitating the network's ability to swiftly concentrate on the distribution of target positions and their adjacent zones.

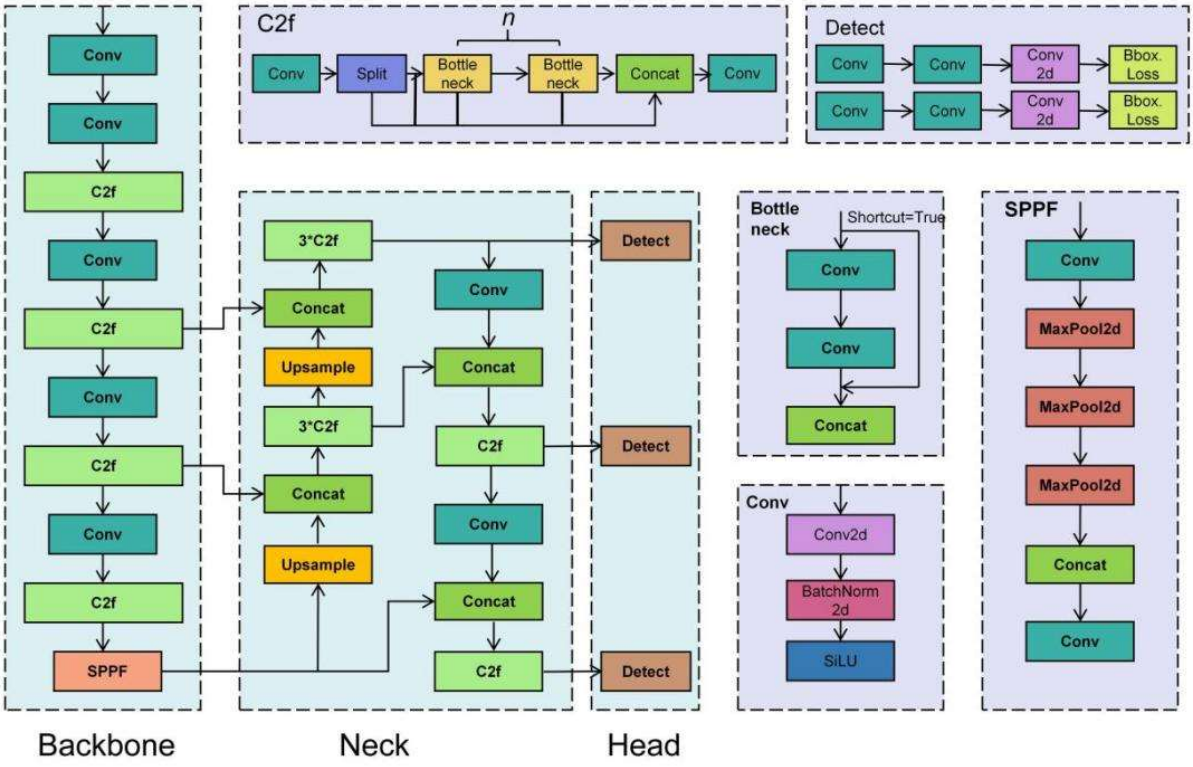

Figure 1. YOLOv8 Algorithm Framework

2.2 The YOLOv8 Algorithm's Improvement

Addressing the aforementioned issues, this paper introduces a ship detection algorithm named CL-YOLO, which integrates two functional modules, LKA and CA, based on YOLOv8. The structure of CL-YOLO is illustrated in Figure 2.

In refining the model, a C2f-CA module is crafted by seamlessly integrating the Coordinate Attention mechanism with the C2f module. The essence of this module is to decouple and refine the channel attention mechanism, employing a strategy that aggregates features across two distinct spatial dimensions for efficient one-dimensional feature encoding. The design concept behind the C2f-CA module focuses on preserving the precise spatial localization of features while adeptly capturing long-range dependencies along a single spatial dimension. This mechanism allows the model to effectively harness inter-channel correlations without excessively burdening computational resources. The computational demands and extensive parameters associated with the attention mechanism are mitigated by the LKA module, which not only guarantees comprehensive global information extraction but also overcomes its inherent limitations. Consequently, the model demonstrates robust performance in identifying and locating target objects, especially in environments characterized by complex backgrounds.

In conclusion, the refined YOLO network demonstrates a notable enhancement in performance for ship detection tasks, achieving an improvement in mAP@0.5 from 96.4% to 97.5%. This advancement underscores the YOLO model's superior efficacy in ship detection and enhances its applicability in real-world scenarios.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202412_10(12).0009

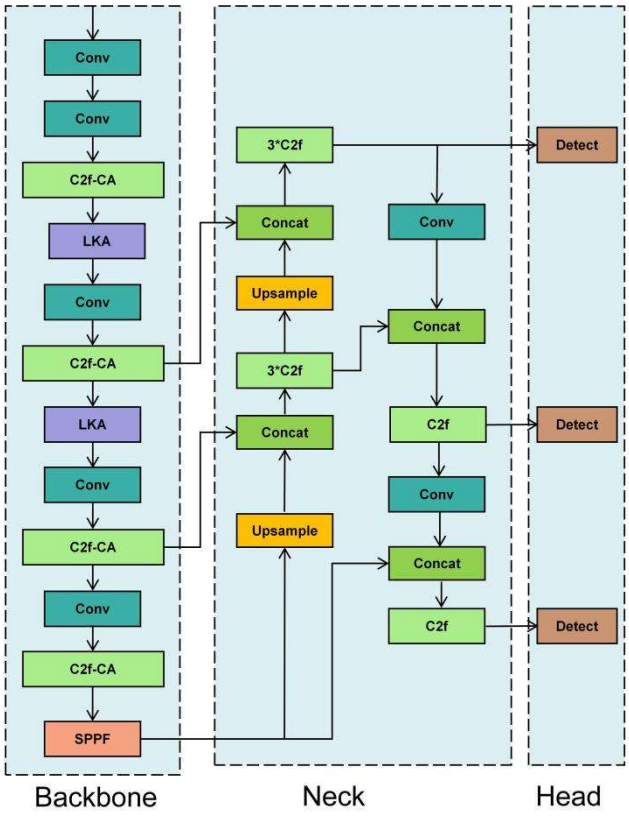

Figure 2. CL-YOLO Improvement Modules

2.2.1 CA Attention Mechanism

The Coordinate Attention mechanism and the C2f module are combined to form the C2f-CA module. The core idea behind the CA attention mechanism is to introduce coordinate information, enabling the model to better understand the relationships between different positions. The structure of the CA attention mechanism is illustrated in Figure 3. In Figure 3, C represents the number of channels in the input feature map, H and W denote the height and width of the input feature map respectively, r is the reduction ratio, and X Avg Pool and Y Avg Pool refer to one-dimensional horizontal global pooling and one-dimensional vertical global pooling, respectively.

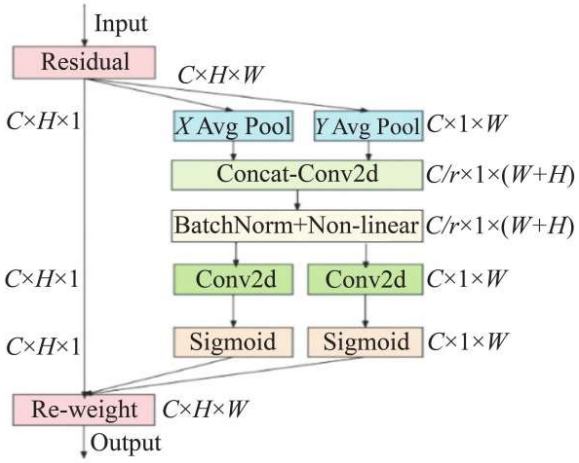

Figure 3. CA Attention Mechanism

The CA attention mechanism initially applies two global average pooling operations to the input feature map: one along the width and the other along the height. This results in two feature maps that

ISSN: 2414-1895 DOI: 10.6919/ICJE.202412_10(12).0009

individually capture the global characteristics in the width and height dimensions. The aforementioned process is mathematically expressed in Equation (1).

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i,j) \tag{1}$$

in equation (1):$z_c$denotes the global features;$x_c(i,j)$ is the input of the original features.

And then, through channel concatenation and convolutional smoothing, it generates a representation with a global receptive field and more precise information, as shown in Equation (2).

$$f = \delta\Big(F_1([z^h, z^w])\Big) \tag{2}$$

In Equation (2): F1 represents the feature map after batch normalization processing; f is the feature map obtained through the Sigmoid activation function; δ is a non-linear activation function;zh and zw are the aggregated feature maps.

It is then transformed through batch normalization and non-linear mapping into tensors with the same number of channels as the input data, as shown in Equations (3) and (4).

$$\mathrm{g^h} = \delta\Big(F_h(f^h)\Big) \tag{3}$$

$$\mathrm{g^w} = \delta\big(F_w(f^w)\big) \tag{4}$$

In Equations (4) and (5): f^h and f^w represent two independent tensors obtained by splitting the feature map f along the spatial dimensions; F_h and F_w denote the tensors that transform f^h and f^w respectively to have the same number of channels as the input; $\mathrm{g^h}$ and $\mathrm{g^w}$ are the attention weights obtained after the aforementioned calculations for the input feature map in the height and width directions, respectively;δ is the Sigmoid function. Finally, by performing weighted multiplication on the original feature map, the final feature map with attention weights in both the width and height directions is obtained. The final output of Coordinate Attention is given by Equation (5).

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{5}$$

In Equation (5): $y_c(i,j)$ represents the finally output feature; $x_c(i,j)$ denotes the original feature; $g_c^h(i)\ and\ g_c^w(j)$ indicate the features that have been processed through aggregation to restore the original number of channels in the height and width dimensions, respectively.

2.2.2 LKA Attention Mechanism

In recent times, the self-attention mechanism has garnered extensive application within the realm of computer vision. Nevertheless, it encounters several intrinsic challenges when dealing with image data. Notably, the self-attention mechanism frequently treats images as one-dimensional sequences, which runs counter to their inherent two-dimensional structure, thereby yielding a less intuitive information representation. Additionally, for images of high resolution, the self-attention mechanism's substantial computational complexity poses a significant resource consumption burden.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202412_10(12).0009

Most crucially, despite its robust adaptability in the spatial dimension, the self-attention mechanism tends to neglect the informational disparities present in the channel dimension of images.

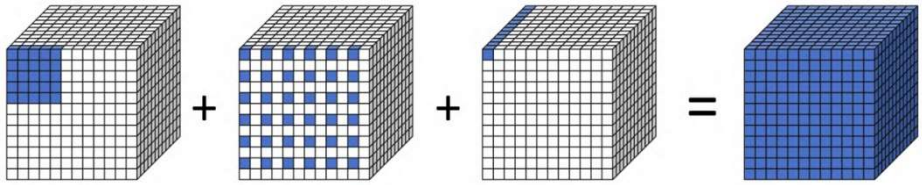

Figure 4. LKA Attention Mechanism

To overcome these limitations of self-attention mechanisms, researchers have proposed attention mechanisms based on large kernel convolution, with the LKA (Large Kernel Attention module) being a typical representative. As shown in Figure 4, the LKA module combines three types of convolutions: regular convolution uses a 5x5 or 7x7 size convolution kernel to capture local features; Voids convolution, also known as dilated convolution, increases the receptive field while maintaining the same number of parameters. The size of the convolution kernel can be set according to the context, and the specific value is omitted here. The spacing is set to 3 to enhance the ability to capture long-distance information; And the channel convolution 1x1 convolution kernel is used for re-calibrating features in the channel dimension, which makes up for the lack of channel adaptability in the self-attention mechanism.

The design of the LKA module cleverly integrates the advantages of convolution and the Transformer, overcoming the shortcomings of traditional convolution in modeling long-distance dependencies while also compensating for the Transformer's deficiencies in capturing local information and adapting to the channel dimension.

Therefore, to overcome the shortcomings of traditional convolution, this paper adopts large kernel attention convolution to extract image features with a global receptive field, making the extracted image features have long-range dependency compared to those obtained through traditional convolution.

3. Experiments

3.1 Experimental Platform

The experimental environment is configured as follows: The operating system of the experimental platform is Windows 10 (Professional Edition), with an Intel(R) Core(TM) i7-13700KF CPU @ 3.4 GHz and 64 GiB of RAM. The graphics card is an Nvidia GeForce RTX 3090 with 24 GiB of VRAM. The development language is Python 3.9, the deep learning framework is PyTorch 2.0.1, the CUDA version is 12.1, and the cuDNN version is 8.9.

3.2 Evaluation Metrics

The experiment will evaluate the model's performance based on parameters (Params), FLOPS, precision, recall, and mean average precision (mAP), with mAP@50 being used for the average precision. Precision indicates the model's false detection rate when detecting targets, while recall represents the miss rate. mAP@0.5 is calculated as the average of the sum of average precisions (AP) for all labels divided by the total number of categories. A higher mAP@0.5 value signifies better detection accuracy for the model. The formulas for P (Precision), R (Recall), and mAP are as follows:

$$P = \frac{TP}{TP + FP} \tag{6}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202412_10(12).0009

$$R = \frac{TP}{TP + FN} \tag{7}$$

$$AP = \int_0^1 PRdr \tag{8}$$

$$mAP = \frac{1}{n}\sum_{i=0}^{n} AP_i \tag{9}$$

Where P stands for precision, R for recall, TP for the number of true positives (correctly classified as positive), FN for the number of false negatives (incorrectly classified as negative), FP for the number of false positives (incorrectly classified as positive), and n for the total number of categories.

3.3 SeaShips Dataset

In this experiment, we utilized the publicly available SeaShips dataset, a large-scale dataset of ships provided by the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing at Wuhan University. Among its 7,000 published images, the dataset covers six common types of ships. The dataset was divided into training, validation, and test sets at a ratio of 7:2:1.

3.4 Ablation Experiment

To verify the performance of each component of the CL-YOLO model, a series of ablation experiments were designed for comparison. The results are shown in Table 1. Here, A represents the backbone network of the improved model using LKA, and B represents the addition of the C2f-CA attention mechanism to the backbone network. Experiment 1 is the original YOLOv8 network without any improvements. From Experiment 2, it can be seen that after using C2f-CA, the model's parameter count and FLOPS are reduced, precision is improved, recall is decreased, and mAP@50 remains unchanged. Experiment 3 shows that the model's parameter count remains the same, while precision, recall, and mAP@50 performance indicators have significant improvements. Experiment 4 demonstrates that with the adoption of both C2f-CA and LKA, all indicators show significant improvements compared to the original network. By directly comparing Experiments 1 and 4, we can conclude that the proposed network can reduce computational costs while improving various performance indicators, proving that our CL-YOLO model has higher performance than YOLOv8. Figure 5 shows a comparison of detection effects between YOLOv8 and CL-YOLO.

Table 1. Ablation experiment results

Number	A	B	Params/M	FLOPS/G	Precision/%	Recall/%	mAP@50/%
1	×	×	20.8	52.7	95.1	91.6	96.4
2	×	√	17.6	37.6	95.2	91.4	96.6
3	√	×	20.6	52.1	95.5	92.5	96.9
4	√	√	**17.7**	**35.4**	**96.2**	**92.2**	**97.5**

ISSN: 2414-1895 DOI: 10.6919/ICJE.202412_10(12).0009

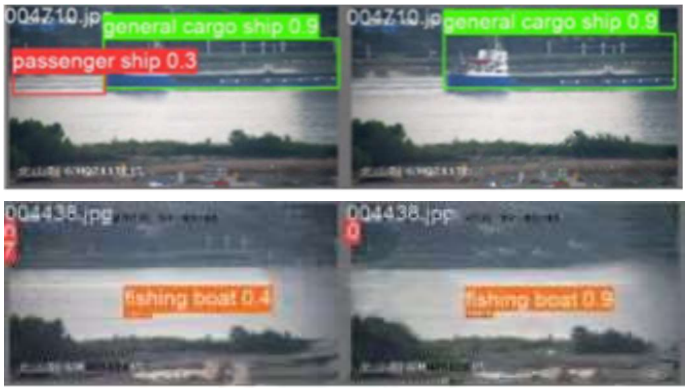

Figure 5. Comparison of detection effects between YOLOv8 and CL-YOLO.

3.5 Comparative Experiment

To verify the effectiveness of CL-YOLO in improving ship recognition performance for object detection, a comparative experiment was conducted under the same conditions between the improved algorithm and other commonly used object detection models.

Table 2. Comparison of Experimental Results

Model	Precision/%	Recall/%	mAP@50/%
YOLOv3	84.7	80.7	89.1
YOLOv5	89.8	78.8	87.9
YOLOv7	94.5	86.4	94.3
YOLOv8	95.1	91.6	96.4
Faster R-CNN	87.2	80.7	89.1
SSD	86.3	82.6	88.2
RT-DETR	94.2	85.9	93.2
CL-YOLO	**96.2**	**92.2**	**97.5**

As shown in Table 2, our proposed CL-YOLO object detection algorithm achieves a 1.1% improvement in mAP compared to YOLOv8, reaching 97.5%. Additionally, CL-YOLO demonstrates varying degrees of precision improvement when compared to six other mainstream object detection algorithms. This experiment proves the superiority of CL-YOLO for ship target detection.

4. Conclusion

To address the challenges in complex scenarios where ships vary greatly in size, type, and quantity, and traditional computer vision methods lack sufficient detection accuracy and have slow recognition speeds for ships, this paper proposes a lightweight ship target detection method based on an improved YOLOv8: CL-YOLO. Experimental analysis on the SeaShips public dataset demonstrates that CL-YOLO can better detect ships against complex backgrounds. Furthermore, CL-YOLO reduces computational load, alleviates the burden on the model, and significantly improves accuracy.

References

[1] Girshick R. Fast r-cnn[C].Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[2] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems, 28, 91-99.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202412_10(12).0009

[3] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[4] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C].Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[5] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C].Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[6] ZHANG S M, SUN Y W, FAN W, et al. Research progress in the application of deep learning methods for marine fishery production: a review[J]. Journal of Dalian Ocean University, 2022, 37(4): 683-695.

[7] WANG S X, ZHANG S M, ZHU W B, et al. Application of an electronic monitoring system for video target detection in tuna longline fishing based on YOLOR5 deep learning model [J]. Journal of Dalian Ocean University, 2021, 36(5): 842-850.

[8] TU W, YU H, ZHANG P, et al. Farmed fish detection by improved Yolov8 based on channel non-degradation attention with weighted spatial pyramid pooling attention network attention [J]. Journal of Dalian Ocean University, 2023, 38(4): 717-725.

[9] YUAN H C, TAO L. Detection and identification of fish in electronic monitoring data of commercial fishing vessels based on improved Yolov8 [J]. Journal of Dalian Ocean University, 2023, 38(3): 533-542.

[10] MEI H B, HUANG A Z, YAO A H C. A lightweight and efficient object detection method based on improved yolov7 model [J]. Journal of Dalian Ocean University, 2023, 38(6): 1032-1043.

[11] WU Z G, CHEN M. Lightweight detection method based on improved yolov7 model [J]. Journal of Dalian Ocean University, 2023, 38(1): 129-139.

[12] GAO Y,WU Z,REN M,et al. Improved YOLOv4 based on attention mechanism for ship detection in SAR images[J].IEEE Access,2022,10:23785-23797.

[13] Huang. Research on Fall Detection Algorithm Based on Improved YOLOv7 [D]. Wuhan: Jianghan University, 2023.

[14] YIN Y,LEI L,LIANG M,et al. Research on fall detection algorithm for the elderly living alone based on YOLO[C] / / 2021 IEEE International Conference on Emergency Science and Information Technology (ICESIT). Chongqing:IEEE,2021:403-408.

[15] Woo S, Debnath S, Hu R, et al. Convnext v2: Codesigning and scaling convnets with masked autoencoders[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 16133-16142.

[16] WANG Wenliang, LI Yanxiang, ZHANG Yifan,et al.MPANet-YOLOv5:Complex sea target detection by multipath aggregation network[J]. Journal of Hunan University(Natural Science Edition),2022, 49(10):69-76.

[17] ZHANG X P, XU Z Y, QU S, et al. Recognition algorithm of marine ship based on improved YOLOv5 deep learning [J]. Journal of Dalian Ocean University, 2022, 37(5): 866-872.

[18] ZHANG D C, LI H T, LI X, et al. Optimization of Yolov5 fish vessel target detection based on CBAM and BiFPN [J]. Fishery Modernization, 2022, 49(3): 71-80.