WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 7, Number 1, Year 2025
DOI: https://doi.org/10.62051/ijcsit.v7n1.06
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Underwater Object Detection Using YOLOv8 Enhanced with Region-based Feature Aggregation Attention

Rui Yang *

School of Computer Science, Yangtze University, Jingzhou 434023, China
*Corresponding Author: Rui Yang

ABSTRACT

With the increasing demand for high-nutritional-value marine products such as sea cucumbers, scallops, and starfish, efficient underwater object detection has become critical for intelligent marine industry applications. Traditional manual sorting methods are inefficient, labor-intensive, and error-prone, making them unsuitable for large-scale industrial needs. Deep learning-based object detection, particularly the YOLO family of algorithms, has shown great potential in addressing these challenges. However, existing models still struggle with low image clarity, color distortion, and occlusion common in underwater environments. In this study, we propose an enhanced YOLOv8n model that integrates a Region-based Feature Aggregation (RFA) attention mechanism to improve feature representation in underwater scenarios. The URPC2020 dataset was preprocessed and adapted for YOLOv8 training, and extensive experiments were conducted. Results demonstrate that the proposed model achieves improvements of 5.6%, 7.6%, 6.7%, and 20.2% in precision, recall, mAP50, and mAP50–95, respectively, compared to the baseline YOLOv8n. Furthermore, the proposed approach outperforms state-of-the-art detectors including YOLOv9s and YOLOv10s while maintaining lightweight architecture. An integrated underwater detection system was also developed with real-time image/video processing and graphical interface support, meeting the practical needs of the marine industry.

KEYWORDS

Underwater object detection; YOLOv8; Attention mechanism; RFA; Deep learning; Marine industry

1. INTRODUCTION

The development and utilization of marine resources are becoming increasingly critical in the context of industrial upgrading and global food security.

Traditional seafood processing, which relies on manual identification and sorting, suffers from low efficiency and high error rates, limiting its scalability. Deep learning-based computer vision methods, particularly object detection frameworks such as R-CNN [1] and YOLO [2-5], provide promising solutions for automated underwater detection.

However, underwater environments present unique challenges, including light scattering, low contrast, color distortion, and frequent occlusions [9]. These factors degrade image quality and reduce the robustness of detection algorithms.

YOLO algorithms, known for their single-stage and real-time detection capabilities [2–5], are widely adopted in complex applications. Nevertheless, while attention mechanisms and multi-scale feature fusion have been introduced to improve detection accuracy [6-8], existing methods often increase computational complexity, which is problematic in embedded or real-time scenarios.

To address these challenges, this study introduces an enhanced YOLOv8n model [5] that integrates the Region-based Feature Aggregation (RFA) attention mechanism, which strengthens feature extraction in challenging underwater conditions without significantly increasing computational costs.

The contributions of this paper are as follows:

(1) Development of a preprocessing pipeline to adapt the URPC2020 dataset annotations for YOLOv8 training.

(2) Integration of the RFA attention mechanism into YOLOv8n, improving feature representation in underwater environments.

(3) Comprehensive experimental evaluation, including ablation studies, comparison with alternative attention modules [6-8], and benchmarking against state-of-the-art detectors [2-5].

(4) Development of a complete underwater object detection system with real-time image/video processing and graphical user interface support for practical deployment.

2. RELATED WORK

Early object detection approaches relied on hand-crafted features and classical vision techniques such as Haar cascades, HOG, and DPM.

With the advent of deep learning, two-stage detectors such as R-CNN and Faster R-CNN [1] improved detection accuracy but were computationally expensive. Single-stage detectors, notably the YOLO family [2–5] and SSD, offered real-time performance while balancing accuracy and speed.

For underwater detection, researchers have adapted YOLO variants [2–5] by enhancing multi-scale feature fusion, improving robustness to blurred and noisy images, and incorporating attention mechanisms such as CBAM [6] and SENet [7].

For instance, BiFPN-based feature fusion, contextual transformers, and pooling redesign strategies have been proposed. Despite these improvements, challenges remain regarding model efficiency, generalization, and robustness in low-contrast and occluded scenes [9].

Our work addresses these challenges by embedding the RFA attention module, conceptually related to feature aggregation approaches such as DFANet [8], into YOLOv8n [5], achieving improved accuracy while retaining computational efficiency.

3. METHODOLOGY

3.1. Dataset and Preprocessing

The URPC2020 dataset [10], collected in real underwater environments, contains images of sea urchins, sea cucumbers, starfish, and scallops. The dataset was divided into 5,542 training images, 800 validation images, and 1,200 test images.

Since the dataset annotations were provided in XML format incompatible with YOLOv8, we developed Python scripts to convert annotations into YOLOv8 text format. This preprocessing ensured compatibility and facilitated efficient training.

3.2. YOLOv8 Baseline

YOLOv8 is a state-of-the-art object detector featuring an anchor-free decoupled head, CSPDarknet-based backbone, and PAN-FPN-inspired neck structure.

The YOLOv8n variant was selected as the baseline due to its lightweight architecture and suitability for real-time underwater deployment, as shown in Figure 1.

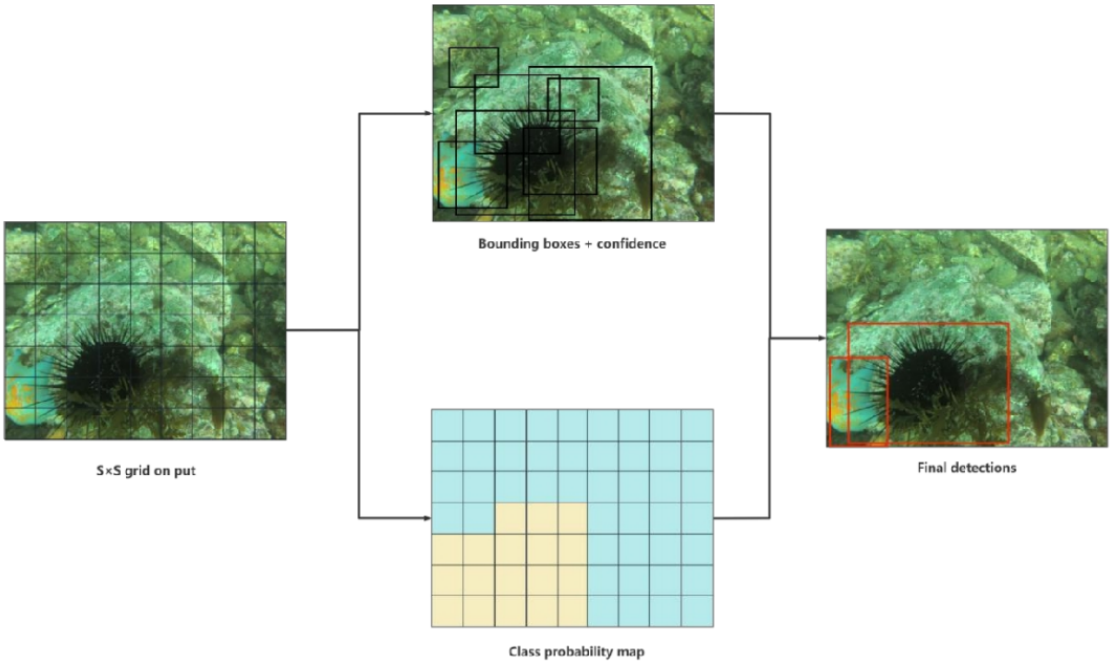

Figure 1. Schematic diagram of YOLO algorithm detection idea

3.3. Integration of RFA Attention

To enhance feature representation in underwater scenes, we integrated the RFA attention mechanism into the YOLOv8n backbone.

The RFA module aggregates features within dynamically defined regions, assigning adaptive weights to emphasize informative local features while suppressing irrelevant background noise.

This allows the model to better capture structural details in low-contrast underwater imagery.

3.4. Training and Implementation

Experiments were conducted using an NVIDIA RTX 3080 Ti GPU, with CUDA 11.8 and PyTorch. The training was performed with a batch size of 32, input resolution of 640×640, and 100 epochs.

Transfer learning was applied by initializing the model with pretrained YOLOv8n weights. Early stopping was employed to prevent overfitting, and the best-performing model weights were saved for evaluation.

References are cited in the text just by square brackets [1]. (If square brackets are not available, slashes may be used instead, e.g. /2/.) Two or more references at a time may be put in one set of brackets [3, 4]. The references are to be numbered in the order in which they are cited in the text and are to be listed at the end of the contribution under the heading References, see our example below.

4. EXPERIMENTAL RESULTS

The performance curves, confusion matrix, and normalized confusion matrix of the original and improved models visualize the differences in the metrics, providing visual verification of the effectiveness of the algorithm.

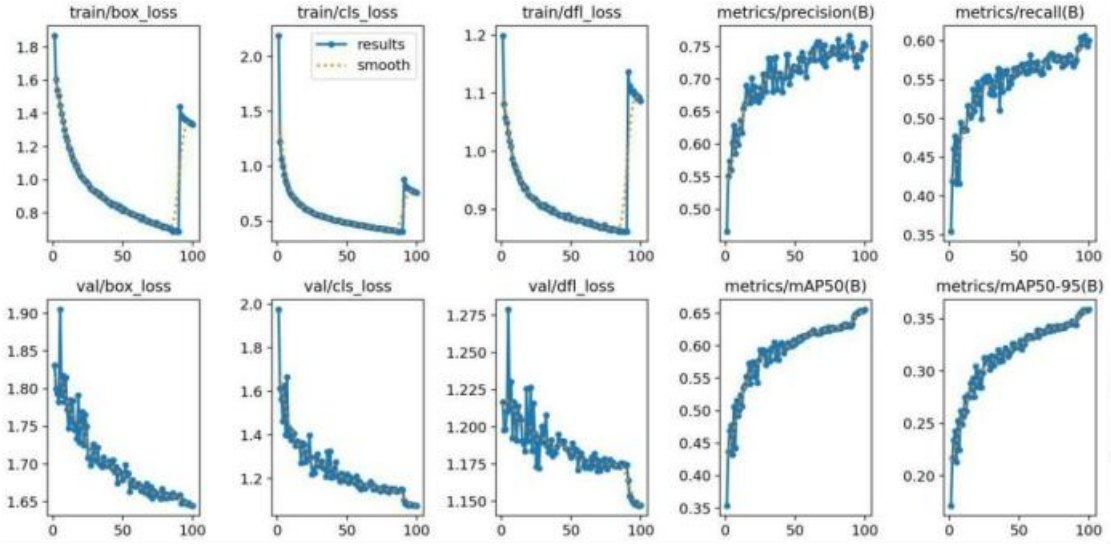

Figure 2. Performance curveof the original model

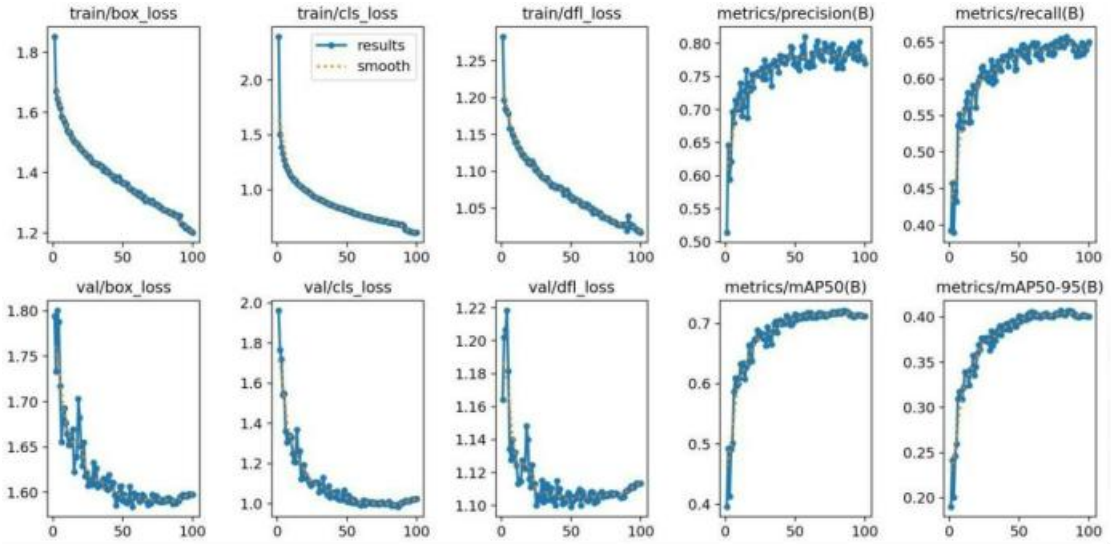

Figure 3. Performance curveof the improved model

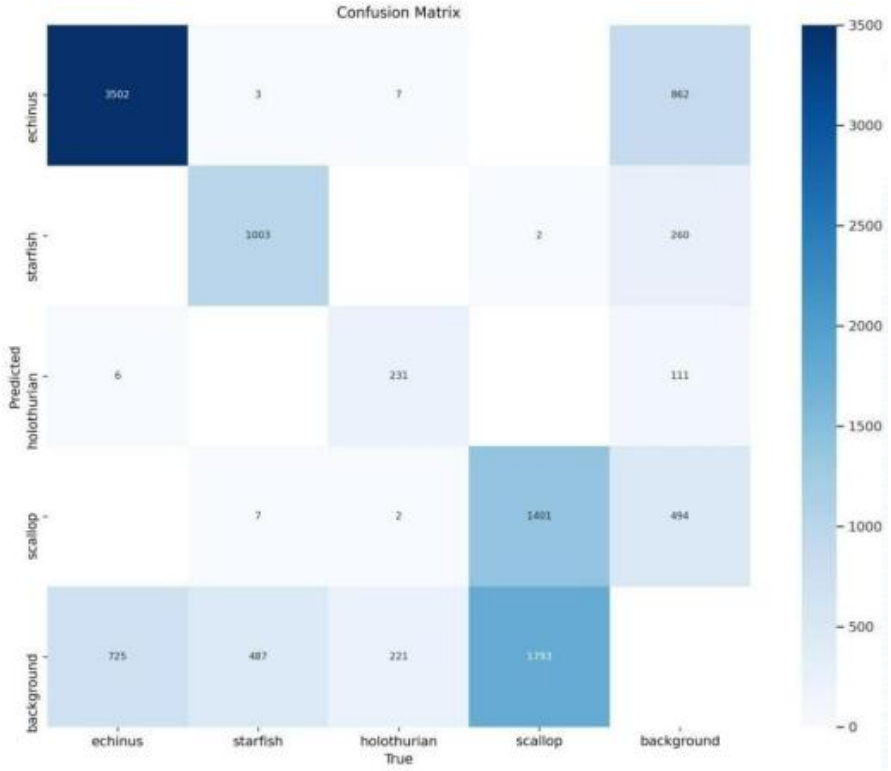

Figure 4. Confusion matrixof the original model

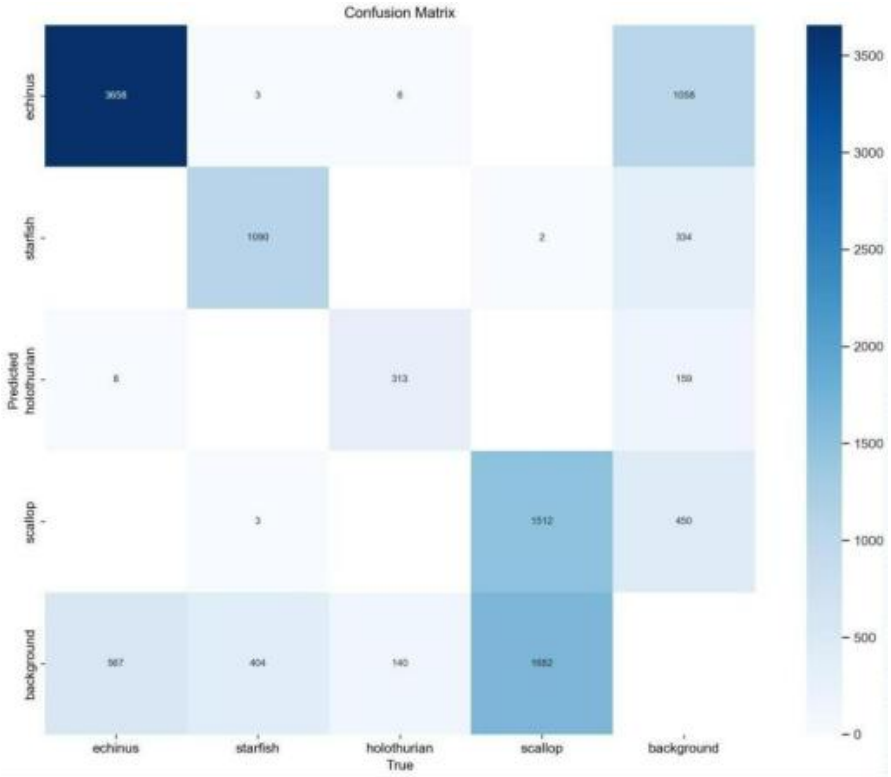

Figure 5. Confusion matrixof the improved model

From the training loss curves (top row of subplots), the original model exhibits significant fluctuations during the later training stages (epochs 50–100), with abnormal jumps in both the overall training loss and the box regression loss near epoch 100, indicating instability in the optimization process. Such oscillations may be attributed to gradient instability or overfitting. In contrast, the improved

model shows a smoother downward trend in the training loss curves, with consistently lower values and reduced oscillations in the later stages, suggesting that the RFA module enhances feature representation stability and thereby improves model convergence.

The validation loss curves (third subplot from the left in the bottom row) further confirm the improvement in generalization: while the baseline model's val/box_loss and val/dfl_loss display high-frequency oscillations after approximately 50 epochs, the improved model achieves smoother curves with lower final convergence values, demonstrating superior stability and generalization performance.

5. CONCLUSION

This study proposed an enhanced YOLOv8n model incorporating the RFA attention mechanism for robust underwater object detection. Through dataset preprocessing, model integration, and systematic experiments, the proposed approach demonstrated significant improvements in detection accuracy while maintaining lightweight architecture. The developed detection system further bridges the gap between algorithmic research and real-world deployment.

Future work will explore combining image enhancement techniques with detection frameworks and extending the approach to multi-class and multi-scene marine applications.

REFERENCES

[1] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149.

[2] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788.

[3] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[4] Jocher, G., et al. (2022). YOLOv5: Implementation details and updates. GitHub repository: https://github.com/ultralytics/yolov5

[5] Ultralytics (2023). YOLOv8: Next-generation real-time object detection. GitHub repository: https://github.com/ultralytics/ultralytics

[6] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. Proceedings of the European Conference on Computer Vision (ECCV), 3–19.

[7] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-Excitation Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7132–7141.

[8] Li, H., Xiong, P., Fan, H., & Sun, J. (2019). DFANet: Deep feature aggregation for real-time semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 9522–9531.

[9] Li, X., et al. (2020). Object detection in underwater environments: A survey. Pattern Recognition Letters, 135, 148–156.

[10] URPC2020 Dataset: Underwater Robot Picking Contest. Official competition dataset, 2020.