DEVELOPING OBJECT DETECTION ALGORITHM FOR OPTOELECTRONIC SYSTEMS ON SURFACE VESSELS USING DEEP LEARNING MODELS

Minh Thuan Nguyen[1], *Van Nam Tran*[2],*Xuan Tung Truong*[3,*]

Abstract

Automatic detection of surface vessels is an important task in maritime surveillance and security. This paper proposes an improvement to the Ultralytics YOLOv8 model to achieve higher accuracy and faster processing speed when recognizing surface vessels under harsh lighting and weather conditions. The paper intergrates three main techniques: a new C3Plus block, a Position-wise Spatial Attention (PSA) mechanism, and a Convolutional Block Attention Module (CBAM) module to enhance the network's feature learning ability. Experiments on a diverse ship image dataset show that the improved model provides an increase in mAP of about 3–6% compared to the original YOLOv8 while maintaining a similar processing speed. In particular, in dark or noisy conditions, the CBAM and PSA improvements help reduce missing objects and improve the model's robustness.

Index terms

YOLOv8; CBAM; object detection; surface vessel detection; maritime vessels.

1. Introduction

Maritime surveillance plays a vital role in ensuring the safety, security, and operational effectiveness of modern naval forces. As surface vessels increasingly incorporate advanced sensors, high-precision weapon systems, and complex electronic countermeasures, the demands on optoelectronic tracking and recognition systems have grown substantially. Accurate and real-time vessel detection is essential not only for situational awareness but also for supporting autonomous navigation, early-warning systems, target tracking, and threat identification in both military and civilian maritime domains. However, the marine environment presents numerous challenges such as specular reflections on waves, rapidly changing illumination, atmospheric

[1] Naval Technical College

[2] Naval Academy

[3] Institute of Control Engineering, Le Quy Don Technical University

[*] Corresponding author, email: tungtx@lqdtu.edu.vn

DOI: 10.56651/lqdtu.jst.v14.n02.1114.ict

disturbances, and cluttered backgrounds. These factors significantly degrade the performance of classical computer-vision algorithms. Recent advances in deep learning, especially convolutional neural networks, have dramatically improved real-time object detection performance. Among them, the YOLO (You Only Look Once) family of models has demonstrated outstanding speed–accuracy trade-offs in many practical applications. YOLOv8, the latest redesign by Ultralytics, provides a lightweight yet high-performing architecture with strong generalization capabilities. Nevertheless, its performance remains limited in harsh maritime conditions, especially when detecting small, distant, or partially occluded vessels. Complex sea backgrounds can confuse feature extractors, while irregular vessel silhouettes and low-light scenarios often lead to missed or unstable detections. To address these challenges, recent research has focused on enhancing backbone structures, attention modules, and multi-scale feature fusion mechanisms. Works such as MobileViT-YOLO, YOLO-LPSS, and deformable-convolution-based YOLO variants have shown that improvements in contextual modeling and attention can significantly boost recognition capability in complex environments. Building upon these insights, the paper proposes an optimized variant of YOLOv8 tailored specifically for maritime applications. This paper introduces YOLOv8Plus, an improved architecture designed to enhance robustness and detection accuracy for surface vessels under diverse lighting and weather conditions. Three key innovations are integrated into the design:

- C3Plus, a refined multi-layer convolutional block for richer feature extraction;
- PSA, which selectively captures long-range dependencies without excessive computational cost; and
- CBAM Module, which strengthens both channel-wise and spatial attention to suppress background noise and highlight vessel-specific features.

These improvements aim to enhance discriminative capability while maintaining real-time inference speed, enabling deployment on embedded optoelectronic systems such as the NVIDIA Jetson AGX Orin. Extensive experiments conducted on both standard Visual Object Classes (VOC) and custom maritime datasets demonstrate that YOLOv8Plus achieves a 3–6% improvement in mean Average Precision (mAP) over the baseline YOLOv8 while keeping model size and inference time nearly unchanged. The results verify that incorporating advanced attention mechanisms and enhanced feature blocks substantially improves robustness in challenging maritime imaging scenarios. The remainder of this paper is organized as follows. Section 2 reviews recent studies on maritime object detection and improved YOLO architectures. Section 3 describes the fundamental components of YOLOv8 and the proposed enhancements. Section 4 details the methodology and model construction. Section 5 presents the experimental setup and evaluation metrics. Section 6 discusses performance results and comparisons. Finally, Section 7 concludes the paper and outlines future research directions. Overall, the proposed approach provides a balanced trade-off between detection accuracy and computational efficiency for maritime applications. The experimental analysis confirms the suitability of the improved architecture for real-time deployment in resource-constrained environments. These findings highlight the practical relevance of the proposed YOLOv8Plus model for robust ship detection under complex operating conditions.

2. Related work

Currently, modern surface vessels are equipped with high-technology weapon systems and advanced radar that enhance defensive capability and enable precise long-range strikes, posing major security challenges. Real-world conditions such as light reflections on water and complex backgrounds make ship detection difficult. In automatic maritime surveillance systems, accuracy and detection speed are especially important to ensure safety and operational effectiveness. The YOLOv8 convolutional neural network was introduced as a powerful architecture for real-time object detection, but in a highly dynamic marine environment (bright light, fog, waves), it still struggles in many different situations and has difficulty distinguishing between ships.

For example, Zhao *et al.* integrated MobileViT into YOLOv8 to increase contextual learning capability, achieving a 12.5% improvement in mAP50-95 compared with YOLOv7-tiny [1]. Shen *et al.* also proposed the YOLO-LPSS model for small vessels, achieving mAP 3–5% higher than YOLOv8n with only an additional 0.33 million parameters [2]. Meanwhile, Zhou *et al.* improved YOLOv8 using deformable convolution and the BiFormer block, increasing mAP50 by 3.7% and mAP50-95 by 6.7% under low-light conditions [3].

YOLOv8-Plus by Li *et al.* [4] introduced an additional output layer (TDLayer) and enhanced attention modules to detect small objects. In contrast, the YOLOv8Plus model in the paper particularly focuses on maritime applications, optimizing the C3Plus architecture and attention modules to improve recognition capabilities in sea environments. Other recent works also integrate attention mechanisms: YOLOv8-CBAM [5] integrates CBAM into YOLOv8 to improve accuracy for sheep detection; YOLOv8-LCNET [6] applies the PSA mechanism at the end of the Backbone to enhance global information; Zhang used SD-YOLO [7] in ship detection, improving the C3 block by combining coordinate attention and bottleneck (CB-C3). The YOLOv8Plus model proposed by the authors is designed to combine these advantages while adding specific improvements: for example, the C3Plus variant is an improved C3 block with a multi-layer convolution structure and two internal bottleneck plus blocks that allow deeper feature extraction in complex environments. Moreover, YOLOv8Plus also uses PSA to capture global information similar to proposals in YOLOv10, and uses CBAM to increase channel-wise and spatial attention weighting as in Woo [8].

Inspired by these achievements, the paper proposes an improved YOLOv8 architecture that leverages modern attention modules and expanded convolutional blocks to enhance ship recognition capability under diverse lighting and weather conditions in practical maritime environments.

3. Background

3.1. YOLOv8

Ultralytics' YOLOv8 is a state-of-the-art model in computer vision, released on January 10, 2023, designed to simultaneously optimize accuracy and speed for object detection and image analysis tasks. YOLOv8 inherits and advances the design principles of earlier YOLO versions, featuring a backbone, a neck, and an anchor-free split Ultralytics head. By removing the traditional anchor-based mechanism, the model improves generalization capability and simplifies the detection pipeline.

A key highlight of YOLOv8 is its ability to balance accuracy and computational efficiency, making it highly suitable for real-time applications. Beyond conventional object detection, YOLOv8 extends its functionality to a wide range of computer vision tasks, including instance segmentation, pose/keypoint estimation, oriented object detection, and image classification. Ultralytics provides multiple model variants from lightweight "n" and "s" versions optimized for resource constrained environments to more powerful "m," "l," and "x" versions that deliver higher performance allowing users to select configurations appropriate for their accuracy speed requirements.

Thanks to its architectural improvements and flexible design, YOLOv8 is considered one of the most robust and widely adopted models for practical computer vision applications, spanning object detection, segmentation, classification, and pose estimation. Thus, YOLOv8 is not merely a successor to previous YOLO releases but a unified, versatile, and efficient standard model suitable for both academic research and real-world deployment across diverse computer vision tasks.

3.2. C3Plus

The improved C3Plus module is based on the strengths of the C2f and C3 blocks. With flexible architecture, in shallow layers it requires only one BottleneckPlus with two consecutive convolutions, after which features are further enhanced, enriched, and refined through three convolution layers inside a CSP bottleneck of C3. This design helps the model better distinguish vessels from background noise (waves, sky clouds) without significantly increasing parameters due to maintaining the CSP structure. As a result, C3Plus not only enhances feature representation in the backbone but also refines features in the neck, especially for small or blurred objects. Figure 1 illustrates a C3Plus layer architecture.

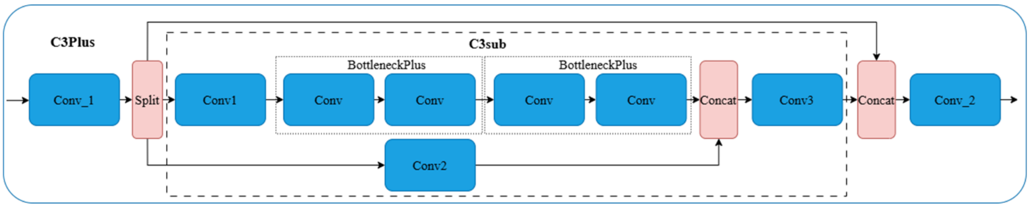

Fig. 1. C3Plus layer architecture.

3.3. Position-wise Spatial Attention

The PSA module is inserted at the output of the backbone after the C3Plus blocks. Its operation is as follows: after a 1 × 1 convolution, the output tensor is channel-wise split into two halves. The first half is passed through a Multi-Head Self-Attention block to acquire global relationships in this portion of features; the second half is forwarded through a standard branch. Afterwards, the results from the attention branch and the standard branch are concatenated and then forwarded to the subsequent FFN feature-enhancement layers. Thanks to PSA, the network can capture broad context and long-range dependencies in the image, which is very useful when vessels appear in complex noisy backgrounds. At the same time, PSA keeps computational cost acceptable by applying attention only to part of the features [9]. Assuming the input $X \in R^{BxCxHxW}$ and the number of spatial elements passing through the layer $N = H \times W$ are given, model parameters can be defined as follows:

$$Q, K, V = Split(qkv(X, W_{qkv}), dim = 2), Q, K, V \in R^{B\times num\_heads\times N\times head\_dim} \quad (1)$$

where, query (Q) represents important query information; key (K) evaluates similarity among features; value (V) carries the weighted transmitted values; W_{qkv} denotes convolutional weights; num_heads is the number of attention heads; $head\_dim$ is the dimension of each head; and qkv is the convolution layer computing the Q, K, V parameters. Through Q and K, the model determines the relevance between elements in the feature map. Figure 2 describes information flow through the PSA layer.

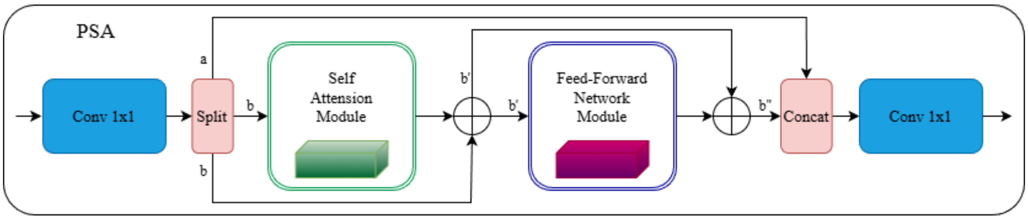

Fig. 2. PSA module.

3.4. Convolutional Block Attention Module

The general functional block diagram of [8] is shown in Figure 3. After the backbone and neck extract multi-level features, the paper applies the CBAM before feeding them into the YOLOv8 Head. CBAM first computes a channel attention map using avg-pooling and max-pooling of the channels, adjusting the importance of each feature channel. The channel attention map is computed as follows:

$$\begin{aligned} M_c(F) &= \sigma\left(MLP(AvgPool(F)) + MLP(MaxPool(F))\right) \\ &= \sigma\left(W_1\big(ReLU(W_0(F^c_{avg}))\big) + W_1\big(ReLU(W_0(F^c_{max}))\big)\right) \end{aligned} \quad (2)$$

where, σ is the sigmoid activation that normalizes output to [0, 1]; MLP is a multilayer perceptron with two fully connected layers, performs average pooling over the spatial dimension to produce a channel descriptor, performs max pooling [9] to produce another channel descriptor and the weights of the MLP. Next, CBAM creates a spatial attention map to weight spatial positions (indicating image regions likely containing vessels). This process helps the model highlight object-containing regions (e.g., bow, cabin) and reduce noise from the background (waves, cloud shadows). CBAM has been shown to improve performance in many detection and classification tasks, and in the authors' method it helps filter unnecessary information before predicting bounding boxes and labels. The spatial attention map output is computed as follows:

$$\begin{aligned} M_c(F) &= \sigma\big(f^{7\times 7}([AvgPool(F'); MaxPool(F')])\big) \\ &= \sigma\big(f^{7\times 7}([F^S_{avg}; F^S_{max}])\big) \end{aligned} \tag{3}$$

where, $f^{7\times 7}$ is a convolution with a 7 × 7 filter that enables the module to collect information from a wider spatial region, thereby improving the capacity to attend to important spatial features.

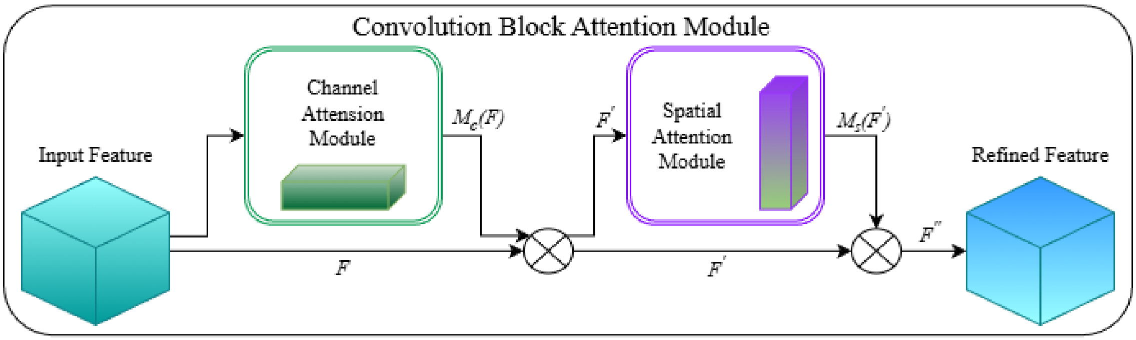

Fig. 3. Block function diagram of CBAM.

4. Methodology

The YOLOv8Plus model is developed with the objective of improving the accuracy, robustness, and real-time performance of surface-vessel detection in optoelectronic systems deployed at sea as shown in Figure 4.

Its design maintains the three-stage structure of modern YOLO architectures Backbone, Neck and Head, but integrates a series of enhanced feature extraction and attention mechanisms tailored to the characteristics of maritime imagery. The overall detection flow begins with input preprocessing, followed by hierarchical feature extraction using the C3Plus-enhanced Backbone, global context encoding via the PSA module, multi-scale feature aggregation in the Neck, and attention-refined prediction of bounding boxes and class labels in the detection Head. The architecture is designed to operate efficiently on embedded platforms such as the NVIDIA Jetson AGX Orin 64 GB, ensuring that accuracy improvements do not come at the expense of computational cost. At the core of the proposed model lies the C3Plus block, an improved feature-extraction module that enhances both the depth and expressiveness of representations learned from raw input images. Unlike the standard C2f/C3 modules in YOLOv8, C3Plus incorporates a BottleneckPlus structure that applies two sequential 3 × 3 convolutions inside each bottleneck. This deeper transformation path enables the network to capture complex maritime patterns such as ship hull curvature, overlapping vessel silhouettes, and reflections on water. The block preserves the Cross Stage Partial (CSP) architecture of YOLOv8 by splitting feature maps into two paths: a transformed path enriched with BottleneckPlus layers and a shortcut path that retains the original spatial information. This combination allows C3Plus to learn richer features while avoiding gradient degradation and excessive parameter growth. As a result, C3Plus significantly strengthens both low-level and mid-level feature extraction, benefiting detection across a wide range of vessel sizes and environmental conditions.

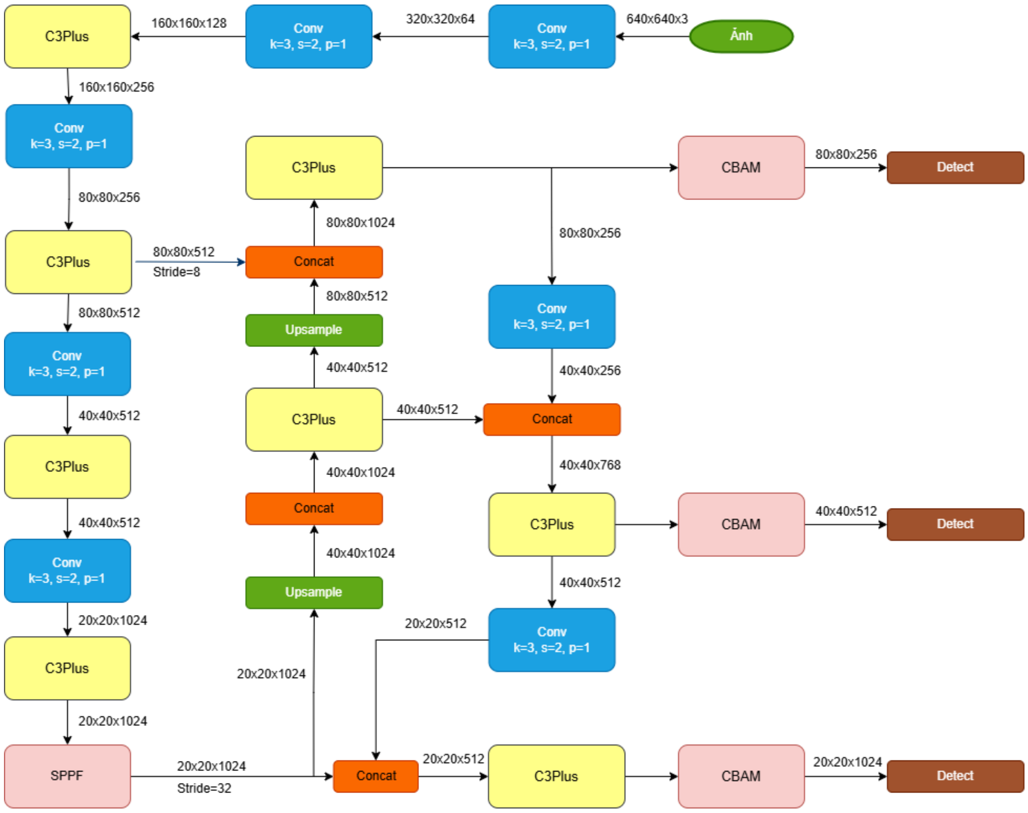

Fig. 4. YOLOv8Plus architecture.

To complement the improved convolutional blocks, the model incorporates the PSA module at the output of the Backbone. PSA introduces a mechanism for capturing long range spatial relationships a critical capability in maritime scenes where vessel targets often appear small, distant, or partially occluded within a large field of view. PSA partially applies multi-head self-attention by dividing the feature tensor into two channel-wise segments. One segment is processed through a lightweight self-attention branch to learn global dependencies, while the remaining segment is forwarded through a standard convolutional branch that preserves local details. The outputs of both branches are then concatenated and refined using feed-forward layers. This design balances the computational cost of attention with the need for enhanced contextual reasoning. In practice, PSA helps the model differentiate vessels from complex background patterns such as waves, sea–sky boundaries, and cloud shadows, especially in low-light or high-glare scenarios.

After feature extraction and global context encoding, the Neck aggregates multi-scale feature maps to strengthen the model's ability to detect vessels of varying sizes, including small or distant targets that often appear in maritime surveillance footage. Prior to generating final predictions, the model applies the CBAM to selectively enhance meaningful information and suppress irrelevant noise. CBAM operates in two sequential stages: channel attention and spatial attention. Channel attention adjusts the weighting of each feature channel based on global statistical descriptors, enabling the model to emphasize channels associated with vessel-specific characteristics while reducing interference from reflections and background textures. Spatial attention then focuses on highlighting regions that are likely to contain vessel structures by analyzing averaged and max-pooled spatial features. Incorporating CBAM in the detection Head ensures that bounding-box regression and classification are guided by the most informative spatial cues, leading to fewer missed detections and improved confidence scores in challenging sea conditions.

Collectively, the integration of C3Plus, PSA, and CBAM forms an enhanced YOLOv8Plus architecture specifically optimized for surface-vessel detection. C3Plus deepens the network's representational power, PSA expands its contextual understanding, and CBAM directs its attention toward the most relevant regions and channels. The synergy of these components allows YOLOv8Plus to achieve higher accuracy and stronger robustness while maintaining real-time inference performance a crucial requirement for deployment in maritime optoelectronic systems where rapid response and reliability are essential.

5. Experiment setup

In the paper, two datasets are used to train and evaluate the model: the VOC Dataset and the Ship Custom Dataset. The VOC dataset [10] is a standard benchmark in computer vision, developed from 2005 - 2012 by a research group associated with the PASCAL project of the European Union. With the goal of promoting progress in tasks such as object recognition, object detection, and segmentation, VOC provides approximately 21,503 object images, including 16,551 training/validation images and 4,952 test images, annotated with 20 common object classes such as humans, cars, animals (cats, dogs, birds), and household objects (chair, table, TV). Each image includes diverse annotation information such as bounding boxes, pixel-wise masks, and class labels. This dataset has become an important benchmark for evaluating and comparing the performance of modern deep-learning models within real-world research contexts.

The Ship Custom Dataset [11] is created by the authors with 10 typical surface-vessel classes (fishing vessel, coast guard vessel, warship, cargo ship, tugboat, passenger ship, fisheries surveillance vessel, submarine, sailboat, small boat). The experimental dataset includes 22,813 images with bounding-box annotations for various ship types, collected from online sources. The dataset is split into 20,004 training images, 1,965 validation images, and 844 test images. Images are preprocessed to 640 × 640 resolution and augmented with horizontal flips, slight rotations, brightness adjustments, and contrast changes to simulate diverse lighting conditions. The model is implemented on an NVIDIA RTX 4060 8 GB GPU using PyTorch. Each model (original and improved) is trained for 300 epochs using Stochastic Gradient Descent (SGD) with an initial learning rate of 0.01. Early stopping is applied if loss does not improve for eight epochs. Evaluation metrics include mAP50, mAP50-95, Precision, Recall, and F1-score on the validation set. Additionally, the model is deployed on an embedded Jetson AGX Orin 64 GB LPDDR5 256-bit device with up to 275 TOPS to ensure reliable benchmarking under realistic embedded deployment conditions scenarios. The model is converted to TensorRT, FP16, batch size = 1, to evaluate accuracy and FPS in real-time inference on video tests.

6. Results and discussion

The paper conducted experiments on YOLOv8n, YOLO11n and YOLOv8Plus using the VOC dataset and the Ship Custom dataset. Table 1 presents a detailed quantitative comparison of YOLOv8n, YOLO11n, and the proposed YOLOv8Plus on two datasets with distinct characteristics: the generic VOC dataset and the domain-specific Ship Custom Dataset. The results are analyzed in terms of detection accuracy, computational complexity, and inference efficiency in order to provide a comprehensive evaluation of the models.

On the VOC dataset, YOLOv8Plus consistently achieves the best detection accuracy across all major evaluation metrics. It attains the highest Precision (0.804) and Recall (0.737), indicating an improved balance between false positives and false negatives compared with YOLOv8n and YOLO11n. More importantly, YOLOv8Plus achieves the highest mAP50 (0.816) and mAP50-95 (0.616). The improvement in mAP50-95 is particularly significant, as this metric evaluates localization performance under stricter IoU thresholds, suggesting that YOLOv8Plus produces more accurate and tightly aligned bounding boxes. These results demonstrate that the architectural enhancements introduced in YOLOv8Plus effectively strengthen feature representation and localization precision without relying on a larger network.

On the Ship Custom Dataset, which is more challenging due to complex backgrounds, scale variations, and inter-class similarity among vessels, YOLOv8Plus again exhibits strong performance. It achieves the highest Precision (0.886), indicating a lower false-positive rate when detecting ships. Although YOLOv8n attains the highest Recall (0.804), YOLOv8Plus maintains competitive recall (0.786) while providing a better balance between Precision and Recall. In terms of overall detection accuracy, YOLOv8Plus delivers competitive mAP50 (0.859) and mAP50–95 (0.685), remaining close to or outperforming the other models under stricter evaluation criteria. These results suggest that YOLOv8Plus generalizes well to domain-specific maritime data and effectively captures discriminative features of surface vessels.

Beyond accuracy, Table 1 highlights that YOLOv8Plus achieves its performance gains with lower or comparable computational cost. YOLOv8Plus has the smallest number of parameters (approximately 2.76 M) and the smallest model size (5.8 MB) among the three models on both datasets. It also requires the lowest computational load, with 7.6 GFLOPs, indicating improved efficiency in feature extraction and inference. Despite this compact design, YOLOv8Plus maintains inference times comparable to YOLOv8n and YOLO11n and achieves real-time processing speeds of approximately 35 FPS. This demonstrates that the proposed modifications enhance accuracy without sacrificing real-time applicability.

Overall, the results in Table 1 demonstrate that YOLOv8Plus provides a favorable trade-off between detection accuracy and computational efficiency. Compared with YOLOv8n and YOLO11n, YOLOv8Plus delivers consistently higher or competitive accuracy on both generic and domain-specific datasets while reducing model complexity and maintaining real-time inference speed. These characteristics make YOLOv8Plus particularly suitable for practical deployment in resource-constrained environments and real-world applications, such as maritime surveillance and embedded vision systems.

Table 1. Model evaluation results on VOC dataset, ship custom dataset

METRICS	**VOC DATASET**			**SHIP CUSTOM DATASET**		
	YOLOv8n	**YOLO11n**	**YOLOv8Plus**	**YOLOv8n**	**YOLO11n**	**YOLOv8Plus**
Precision	0.801	0.803	**0.804**	0.860	0.882	**0.886**
Recall	0.720	0.734	**0.737**	**0.804**	0.800	0.786
mAP50	0.802	0.808	**0.816**	0.856	**0.865**	0.859
mAP50-95	0.597	0.610	**0.616**	0.689	**0.708**	0.685
Layers	**255**	328	321	**225**	295	321
Parameters	3,488,316	3,175,788	**2,760,818**	3,012,798	2,912,430	2,758,868
Model size (MB)	7.2	6.7	**5.8**	6.3	6.1	**5.8**
Gflops	8.2	7.7	**7.6**	8.2	7.7	**7.6**
Inference times (ms)	**1.1**	1.2	1.2	**1.5**	1.6	1.6
Frames per second	**37**	31	35	**37**	31	35

The normalized confusion matrix in Figure 5 shows that the YOLOv8Plus model achieves high classification performance for most ship classes, as indicated by the large values along the main diagonal, which reflect strong predictive accuracy during the training phase. Some classes exhibit confusion with the background or with other classes that share similar morphological characteristics, highlighting the challenges of distinguishing small targets and separating ship structural features under complex background conditions. Nevertheless, the overall misclassification rate remains low, demonstrating the effective feature extraction capability of the YOLOv8Plus model. These results confirm the stability and reliability of YOLOv8Plus when applied to a custom surface ship dataset.

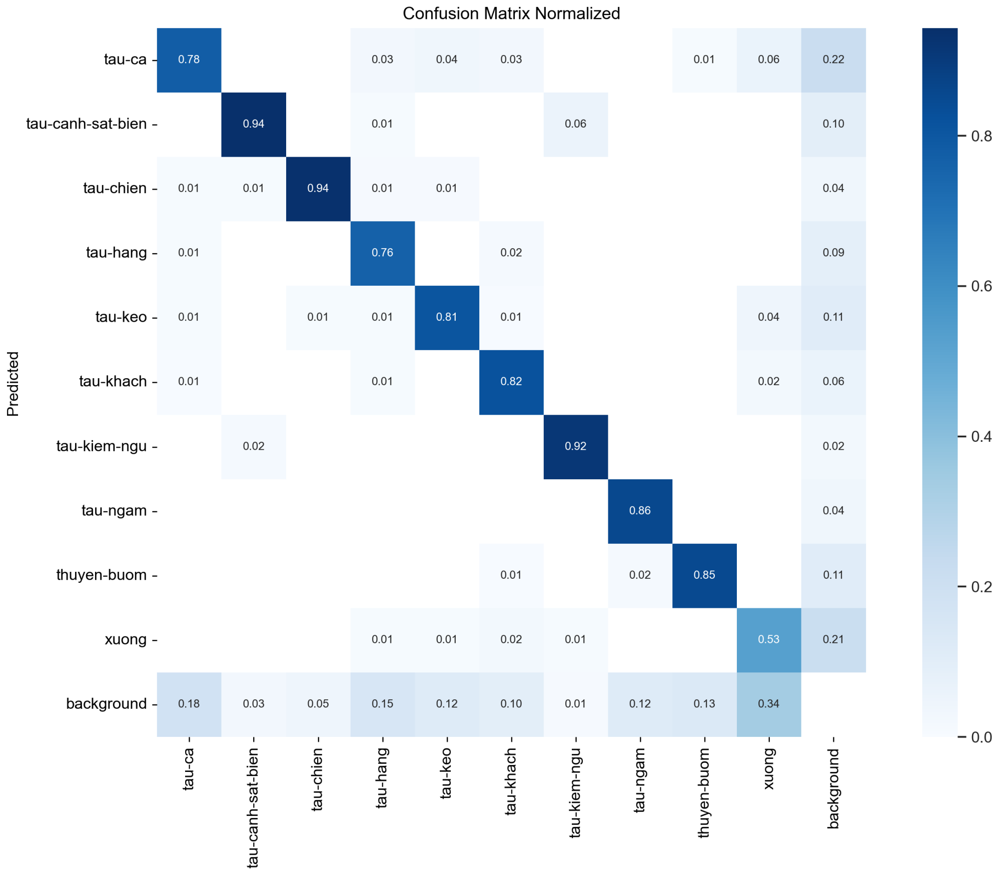

Fig. 5. Confusion matrix for training YOLOv8Plus on the ship custom dataset.

The Loss Function in Figure 6 shows a steady decreasing trend towards 0 over 300 epochs, indicating that the YOLOv8Plus model is learning ship features more efficiently from the Ship Custom Dataset. This steady decrease reflects the ongoing optimization of the Loss function. mAP: with an upward trend (mAP50 increases by more than 0.85, mAP50-95 increases by more than 0.69), the gradual increase in mAP over the epochs is a positive sign, indicating that the model is becoming more accurate.

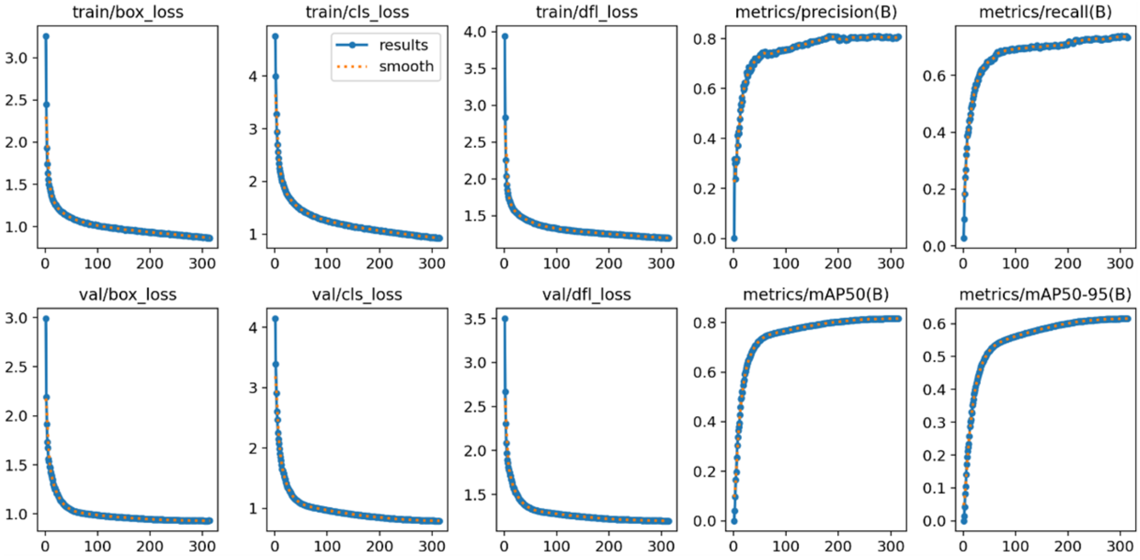

Fig. 6. Loss function and mAP plot for training YOLOv8Plus on ship custom dataset.

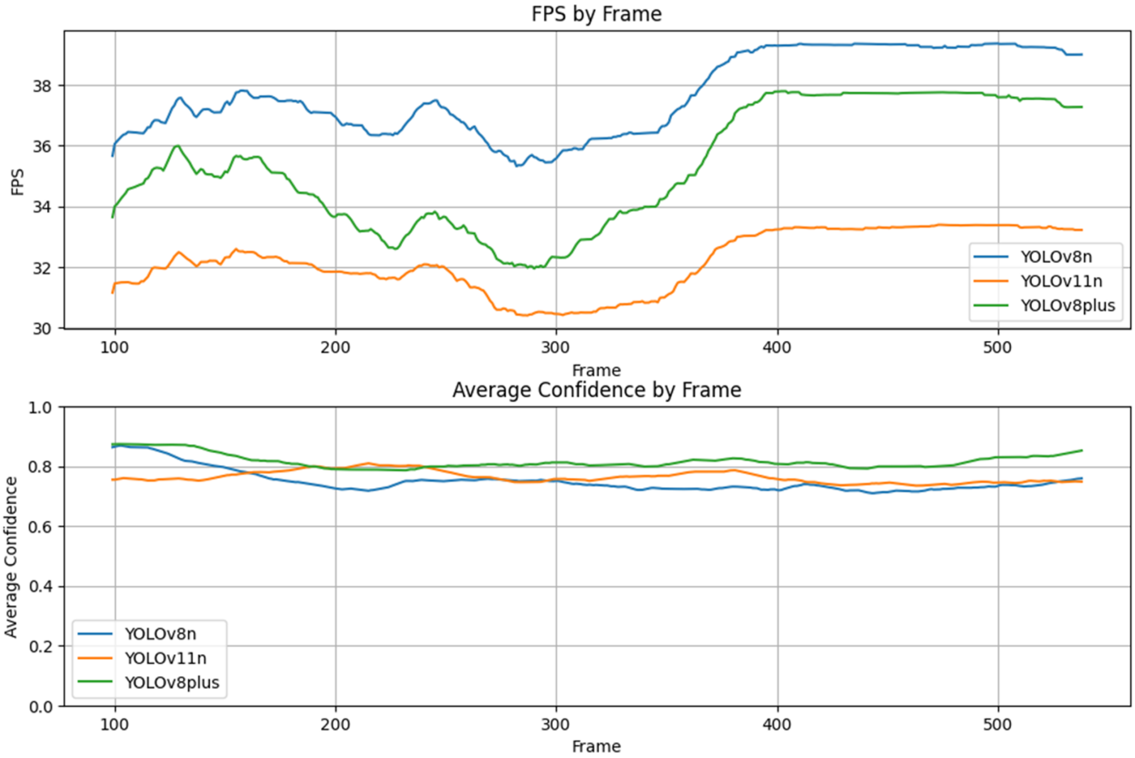

Fig. 7. Performance evaluation of 3 models YOLOv8n, YOLOv11n and YOLOv8plus based on FPS and confidence.

The analysis results, based on the "FPS by Frame" graph and "Average Confidence by Frame" in Figure 7, show the following differences between the three YOLO models: The YOLOv8Plus model has an average processing speed, ranging from 32-38 FPS. Although there is slight variation, the FPS of YOLOv8Plus generally remains high, closely matching YOLOv8n. The YOLOv8Plus model exhibits the highest average reliability and stability, ranging from 0.8-0.85. This indicates that YOLOv8Plus is the model capable of accurately and consistently identifying targets across frames, demonstrating a good balance between speed and accuracy. The YOLOv11n model showed the lowest processing speed among the three models, with FPS fluctuating mainly between 30.5-33.5 FPS. This implies that YOLOv11n may require more computational resources and have a more complex architecture, resulting in slower processing speeds compared to the other two models. The YOLOv11n model initially achieved a relatively high reliability score of approximately 0.78, but then showed a slight downward trend and stabilized at 0.74-0.83, indicating a fairly high but lower accuracy compared to YOLOv8Plus. The YOLOv8n model exhibited the highest and most stable processing speed throughout the evaluation, maintaining approximately 35.7-39.7 FPS. This suggests that YOLOv8n is the model with the fewest layers and the least complex architecture, and therefore the fastest processing capability. The YOLOv8n model recorded the lowest confidence level and showed a decreasing trend, fluctuating only between from 0.73 to 0.84.

7. Conclusion and future work

This study presents YOLOv8Plus, an enhanced deep-learning architecture designed to improve surface-vessel detection in complex maritime environments. By integrating three key components C3Plus for deeper feature representation, PSA for global context modeling, and CBAM for selective spatial channel attention the proposed model significantly strengthens its ability to extract robust features from optoelectronic imagery characterized by unstable illumination, wave-induced noise, and complex sea backgrounds. Experimental results on both the VOC dataset and a large-scale custom ship dataset demonstrate that YOLOv8Plus consistently outperforms the standard YOLOv8 and YOLO11n baselines, achieving a 3–6% increase in mAP while maintaining comparable model size, computational cost, and real-time processing speed. These improvements verify that the combination of enhanced convolutional blocks and attention mechanisms is effective for maritime detection tasks, and that the optimized model remains suitable for deployment on embedded systems such as the Jetson AGX Orin. The results also confirm the model's capability to detect vessels across different weather conditions, speeds, and viewing angles, thereby effectively supporting autonomous surveillance, threat monitoring, and situational awareness in naval and maritime security applications.

Although YOLOv8Plus achieves notable performance gains, several aspects warrant further investigation. Future work will focus on extending the model's capability to handle extreme maritime conditions such as strong glare during sunrise or sunset, heavy fog, and high sea states—scenarios in which vessel contours become highly indistinct. Another important direction is the development of lightweight or quantized variants specifically optimized for edge devices with limited power budgets, enabling integration into compact optoelectronic platforms. In addition, incorporating temporal information from video sequences may improve detection stability and reduce false alarms by leveraging motion cues and temporal continuity. Finally, exploring multimodal fusion with infrared, LiDAR, or radar sensors could further enhance detection reliability in low-visibility environments.

Through these future extensions, YOLOv8Plus offers a promising foundation for advanced maritime object-detection systems, contributing to the development of intelligent, real-time, and resilient surveillance technologies that support national security, autonomous navigation, and maritime domain awareness.

Acknowledgment

This research is funded by the Le Quy Don Technical University Research Fund under the grant number “2025.TXM.01".

References

[1] X. Zhao and Y. Song, “Improved Ship Detection with YOLOv8 enhanced with MobileViT and GSConv,” *Electronics*, Vol. 12, No. 22, p. 4666, 2023. DOI: 10.3390/electronics12224666

[2] L. Shen, T. Gao, and Q. Yin, “YOLO-LPSS: A lightweight and precise detection model for small sea ships,” *Journal of Marine Science and Engineering*, Vol. 13, No. 5, p. 925, 2025. DOI: 10.3390/jmse13050925

[3] L. Zhou, Y. Dong, B. Ma, Z. Yin, and F. Lu, “Object detection in low-light conditions based on DBS-YOLOv8,” *Cluster Computing*, Vol. 28, No. 1, 2024. DOI: 10.1007/s10586-024-04829-1

[4] H. Li and X. Pang, “YOLOv8-Plus: A small object detection model based on fine feature capture and enhanced attention convolution fusion,” *Academic Journal of Computing & Information Science*, Vol. 8, No. 3, 2025. DOI: 10.25236/AJCIS.2025.080316

[5] Q. Qin, X. Zhou, J. Gao, Z. Wang, A. Naer, L. Hai, S. Alatan, H. Zhang, and Z. Liu, “YOLOv8-CBAM: a study of sheep head identification in Ujumqin sheep,” *Frontiers in Veterinary Science*, Vol. 12, 2025. DOI: 10.3389/fvets.2025.1514212

[6] J. Nan, Y. Wang, K. Di, B. Xie, C. Zhao, B. Wang, S. Sun, X. Deng, H. Zhang, and R. Sheng, “YOLOv8-LCNET: An improved YOLOv8 automatic crater detection algorithm and application in the Chang’e-6 landing area,” *Sensors*, Vol. 25, No. 1, 2025. DOI: 10.3390/s25010243

[7] Y. Zhang, L.-Y. Hao, and Y. Li, “SD-YOLO: An attention mechanism guided YOLO network for ship detection,” in *2024 14th International Conference on Information Science and Technology (ICIST), China*, 2024, pp. 769–776. DOI: 10.1109/ICIST63249.2024.10805300

[8] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBCAM: Convolutional block attention module,” in *Proceedings of the European conference on computer vision (ECCV), Munich, Germany*, 2018, pp. 3–19. DOI: 10.48550/arXiv.1807.06521

[9] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, “YOLOv10: Real-time end-to-end object detection,” 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2405.14458

[10] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The PASCAL visual object classes challenge: A retrospective,” *International Journal of Computer Vision*, Vol. 111, No. 1, p. 98–136, 2015. DOI: 10.1007/s11263-014-0733-5

[11] M. T. Nguyen, “YOLOv8 object detection model dataset,” 22 April. 2025. [Online]. Available: https://universe.roboflow.com/thuannguyenminh/yolov8-ywsmu. [Accessed: 18 May. 2025].

Manuscript received 04-6-2025; Accepted 25-11-2025.

■

Minh Thuan Nguyen graduated with a Bachelor's degree in electrical and electronic engineering from the Naval Academy in 2018, where he gained a solid foundation in naval engineering principles and technology. He then pursued a Master's degree in control and automation engineering at Le Quy Don Technical University, graduating in 2025, specializing in advanced control theory and automated systems. He currently teaches and conducts research at the Naval Technical College. His research primarily focuses on weapon systems and modern control and automation techniques. In recent years, his research has increasingly concentrated on computer vision, machine learning, and deep learning, with particular emphasis on their application to intelligent sensing, target detection, and autonomous decision-making in maritime and defense systems. E-mail: junioweak2000@gmail.com.

Van Nam Tran graduated with a Bachelor's degree in electrical and electronic engineering from the Naval Academy in 2017, where he acquired a solid foundation in naval engineering principles and related technologies. He subsequently pursued a Master's degree in control and automation engineering at Le Quy Don Technical University, completing the program in 2025 with a focus on advanced control theory and automated systems. He is currently a lecturer at the Naval Academy, where he is actively involved in teaching and research activities. His research interests include aerodynamics, computer vision, and machine learning, with particular emphasis on their applications in engineering systems. He has participated in several academic and applied research projects in these areas. E-mail: namtrieuqchq@gmail.com.

Xuan Tung Truong received the Bachelor's degree in electrical and electronics engineering from the Institute of Control Engineering, Le Quy Don Technical University, Vietnam, in 2007, the Master's degree in computer engineering from the Embedded System Laboratory, Department of Computer Engineering and Information Technology, University of Ulsan, South Korea, in 2012, and the Doctor of Philosophy degree in robotics from the Faculty of Science, University of Brunei Darussalam, Brunei Darussalam, in 2017. He is currently a lecturer at the Institute of Control Engineering, Le Quy Don Technical University. His research interests include human–robot interaction, socially aware robot navigation, robotics, computer vision, and machine learning. E-mail: tungtx@lqdtu.edu.vn.

XÂY DỰNG THUẬT TOÁN PHÁT HIỆN ĐỐI TƯỢNG ỨNG DỤNG CHO HỆ QUANG ĐIỆN TỬ TRÊN TÀU MẶT NƯỚC SỬ DỤNG MÔ HÌNH HỌC SÂU

Nguyễn Minh Thuận, Trần Văn Nam, Trương Xuân Tùng

Tóm tắt

Phát hiện tự động đối tượng tàu thuyền trên mặt nước là nhiệm vụ quan trọng trong giám sát hàng hải và an ninh. Bài báo này đề xuất cải tiến cho mô hình YOLOv8 của Ultralytics để đạt độ chính xác cao hơn và tốc độ xử lý nhanh hơn khi nhận diện tàu biển dưới các điều kiện ánh sáng và thời tiết khắc nghiệt. Tác giả tích hợp ba kỹ thuật chính: khối C3Plus mới, cơ chế chú ý cục bộ PSA *(Position-wise Spatial Attention)* và module CBAM *(Convolutional Block Attention Module)* để nâng cao khả năng học đặc trưng của mạng. Thí nghiệm trên tập dữ liệu ảnh tàu biển đa dạng cho thấy mô hình cải tiến mang lại mAP tăng thêm khoảng 3–6% so với YOLOv8 gốc trong khi duy trì tốc độ xử lý tương đương. Đặc biệt, trong điều kiện tối hoặc nhiều nhiễu nền, cải tiến CBAM và PSA giúp giảm bỏ sót đối tượng và cải thiện độ bền của mô hình.

Từ khóa

YOLOv8; CBAM; phát hiện đối tượng; phát hiện tàu mặt nước; tàu biển.