Edge-Guided Multi-Scale YOLOv11n: An Advanced Framework for Accurate Ship Detection in Remote Sensing Imagery

Yan Shibo[1*], Liu Pan[2], Abudhahir Buhari[3]
Faculty of Engineering, Science and Technology, Kuala Lumpur University of Science and Technology,
Kuala Lumpur, Malaysia[1, 2, 3]
Department of Information Technology, Anhui Vocational College of Defense Technology, Lu'an, China[1]

***Abstract*—Ship detection in optical remote sensing imagery plays a vital role in maritime surveillance and environmental monitoring. However, existing deep learning models often struggle to generalize effectively in complex marine environments due to challenges such as noise interference, small object sizes, and diverse weather conditions. To address these issues, this study proposes an Edge-Guided Multi-Scale YOLO algorithm (YOLOv11n-EGM). The approach introduces multi-scale deep convolutional branches with varying kernel sizes to perform parallel feature extraction, enhancing the model's ability to detect objects of different scales. Additionally, the classic Sobel operator is incorporated for edge-aware feature extraction, improving the model's sensitivity to object boundaries. Finally, 1×1 convolutions are employed for feature fusion, reducing computational complexity. Experimental results on the ShipRSImageNet V1.0 dataset demonstrate that the improved model achieves notable gains in precision, recall, mAP@0.5, and mAP@0.5:0.95 compared to the baseline, highlighting its superior performance in challenging maritime scenarios. Qualitative analysis further shows that YOLOv11n-EGM can accurately detect both large and extremely small ships in cluttered scenes, with precise boundary localization. However, occasional misclassification in fine-grained categories (e.g., motorboat vs. hovercraft) highlights the challenge of small-instance recognition. Overall, the proposed method exhibits strong robustness and practical applicability in real-world maritime scenarios, offering a promising solution for edge-aware, multi-scale ship detection in remote sensing imagery.**

Keywords—Optical remote sensing imagery; ship detection; multi-scale deep convolution; edge-aware feature extraction

I. Introduction

Maritime transportation plays a crucial role in global trade, accounting for over 80% of total international trade volume. According to the International Maritime Organization (IMO), there are approximately 108,789 commercial vessels worldwide as of 2024. As the core carrier of marine economic activities, the operational efficiency of ships directly impacts the stability of international logistics systems [1]. Ship detection in optical remote sensing imagery has emerged as a critical research area due to its wide-ranging applications in maritime surveillance, port management, and marine traffic control. The task, commonly referred to as optical remote sensing ship detection, involves the automatic localization and classification of ship targets within high-resolution optical satellite images, and serves as a foundational technique for intelligent maritime monitoring systems [2].

However, the unique characteristics of ship targets in nadir-view remote sensing imagery—such as significant scale variation, complex backgrounds, and prominent edge structures—have posed substantial challenges to detection models. In recent years, models such as R-CNN, Fast R-CNN, and the YOLO series have gained attention for their effectiveness in visual target identification [2], [3], [4]. Among them, the YOLO series has demonstrated particular advantages in optical remote sensing applications due to its high detection speed and competitive accuracy [2].

Nevertheless, despite the general success of YOLO-based detectors in object detection tasks, their direct application to ship detection in optical remote sensing imagery—e.g., YOLOv11n—reveals several domain-specific limitations. These include difficulty in handling ships of drastically varying sizes, insufficient exploitation of edge and contour features critical for distinguishing ships from complex backgrounds, and degraded performance in fine-grained classification tasks such as differentiating 49 ship types and docks. YOLOv11n's baseline architecture lacks specialized modules to address the unique challenges posed by maritime scenes [2]. These limitations indicate a clear research gap in designing lightweight yet accurate detectors that can effectively capture multi-scale and edge-sensitive features in complex maritime scenes.

To address this gap, this study proposes an enhanced YOLOv11-based approach with the following key contributions:

A novel multi-scale deep convolutional module, EGM_Block, is introduced, employing parallel convolutional kernels of sizes 3, 5, and 7 to capture features at multiple scales.

1) A Sobel edge extraction module is incorporated to improve the model's sensitivity to edge features of ship targets.

2) An efficient feature fusion strategy is proposed to reduce computational complexity while maintaining high detection accuracy.

3) Extensive experiments are conducted on the publicly available ShipRSImageNet V1.0 dataset, which provides a comprehensive benchmark for evaluating ship detection performance in remote sensing imagery. The results validate the effectiveness and robustness of the proposed method.

The remainder of this study is organized as follows: Section II reviews related work; Section III details the proposed method; Section IV presents experimental results and analysis; Section V provides a discussion of the findings; and Section VI concludes the study and discusses future work.

II. Related Work

A. Traditional Methods for Ship Detection

Early approaches to ship detection in remote sensing imagery primarily relied on handcrafted features and classical machine learning algorithms. Techniques such as edge detection, texture analysis, and morphological operations were commonly employed to extract candidate regions, followed by classifiers like Support Vector Machines (SVM) or AdaBoost for final decision-making. While these methods offered a degree of interpretability, they were limited in their ability to handle variations in scale, orientation, and complex backgrounds, resulting in suboptimal performance in real-world maritime environments [2].

B. Deep Learning-Based Ship Detection

The emergence of deep learning has significantly advanced the field of object detection. Two-stage detectors, such as R-CNN, Faster R-CNN [3], and Mask R-CNN [4], introduced region proposal mechanisms that substantially improved detection accuracy, particularly for small and densely packed objects. After thorough experimental analysis and engineering-level testing, the two-stage model proved highly effective for ship detection in remote sensing imagery. However, in practical engineering applications, the two-stage model is difficult to apply due to real-time or resource-constrained environments, such as satellite-attached equipment or unmanned aerial vehicles [5].

To address the aforementioned challenges, researchers have developed single-stage detectors, such as SSD [6] and the YOLO ("You Only Look Once") series [7], etc. The single-stage detection architecture eliminates the step of region proposal and directly predicts bounding boxes and class probabilities. This design achieves a good balance between speed and accuracy. The YOLO series has been widely used in industry applications due to its real-time performance and continuous architectural improvements [8].

C. Evolution of the YOLO Series

With the joint efforts of researchers and engineers, the YOLO series has undergone multiple efficient iterations in a short period of time. Strictly speaking, each new version surpasses its predecessor in terms of accuracy, speed, and feature representation. YOLOv3 introduced multi-scale detection and residual connections [9], while YOLOv4 [10] and YOLOv5 [11] adopted techniques such as CSPNet, PANet, and data augmentation strategies. YOLOv5 has shown practical effectiveness in agricultural applications, including crop detection and maturity classification under real-world conditions [12]. Recent versions, including YOLOv7 and YOLOv8, have focused on optimizing the backbone network and introducing lightweight variants suitable for edge deployment [13].

YOLOv11 is one of the latest advancements in the YOLO object detection series. Not only has its backbone architecture been optimized, but its feature extraction modules have also been enhanced, resulting in improved detection accuracy and computational efficiency [14] [15]. Therefore, YOLOv11 generally performs better than previous versions in various general object detection tasks, and it is especially suitable for scenarios requiring real-time inference.

Nonetheless, when applied to ship detection in high-resolution optical remote sensing images, there are still some domain-specific challenges. Compared with conventional datasets, optical remote sensing images are captured from different viewing angles, resulting in large variations in ship scales, orientations, and arrangement densities. In particular, ships may appear in various environments such as ports, coastal areas, and open seas. These background conditions lead to significant changes in the overall appearance and spatial distribution of the images, which may reduce the effectiveness of standard detection procedures.

In addition, experimental results show that YOLO models may struggle to capture fine edge details and the contextual cues necessary to distinguish ships from background elements such as waves, docks, and infrastructure. Therefore, accurately identifying ships that are slender, small in size, and densely arranged in images remains a key challenge. Moreover, in real-world engineering scenarios where resources are limited and inference speed is critical—such as onboard or satellite systems—balancing detection accuracy and computational efficiency is also of great importance [16].

Considering the above-mentioned practical technical difficulties, there is a clear need for specialized architectural improvements that can address issues such as multi-scale representation, edge-aware feature learning, and stable performance in complex environments. The YOLOv11n-EGM framework proposed in this study aims to tackle these challenges by integrating an edge-guided multi-scale module and an adaptive loss strategy, thereby enhancing the reliability of ship detection in remote sensing imagery.

D. Challenges in Remote Sensing Ship Detection

Remote sensing imagery presents unique challenges for object detection algorithms. With the growing application of remote sensing in coastal environment monitoring, recent studies have extended its practical relevance to broader maritime scenarios [17]. In complex maritime environments, such as freezing seas, remote sensing systems may produce erroneous object detection results without proper surface discrimination mechanisms [18]. These images are captured from nearly vertical viewing angles, which often results in limited visible features of objects. Additionally, due to the large coverage area, ships frequently appear in various sizes and orientations within the same image, and are often situated in complex environmental conditions, such as cluttered port backgrounds, cloud cover, sea surface noise, and low contrast [19]. Moreover, ships typically exhibit distinct edge features that can be leveraged to improve detection accuracy; however, many existing models fail to fully utilize this information. Furthermore, for applications in real-time maritime monitoring systems, lightweight and computationally efficient models are essential [20] [21].

III. Methodology

This study introduces a lightweight ship detection framework specifically designed for optical remote sensing imagery—Edge-Guided Multi-Scale YOLO. The framework integrates multi-scale perception and edge contour enhancement mechanisms. Its architecture is based on an improved YOLOv11n backbone and is further enhanced by a custom-designed Edge-Guided Multi-Scale block (EGM-block), which improves the model's ability to capture structural and edge-related features across different scales.

A. Overall Network Architecture

The proposed method adopts YOLOv11n as the base detection framework and integrates the Edge-Guided Multi-Scale block (EGM-block) into the backbone network to enhance feature extraction capabilities. The backbone consists of multiple convolutional and residual layers, which progressively capture spatial and semantic representations from the input image. The EGM-Block is inserted at the fourth stage (P3/8) to enhance mid-level features in terms of edge awareness and multi-scale object perception. The C3k2 and C2PSA modules further refine deeper semantic features, while the SPPF module contributes to multi-scale contextual aggregation.

B. Edge-Guided Multi-Scale Block (EGM-Block)

As illustrated in Fig. 1, the proposed EGM-Block is composed of three functional components designed to enhance the model's ability to capture structural and scale-aware features in remote sensing imagery.

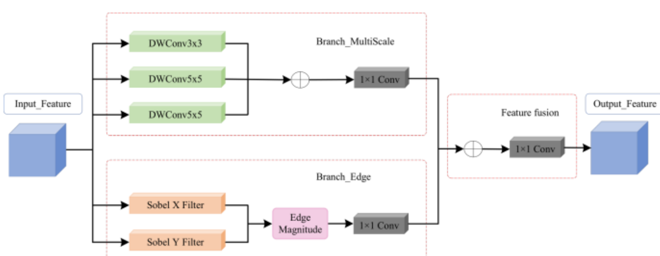

Fig. 1. Edge-Guided Multi-Scale Block.

1) Multi-scale depthwise convolutional branches: This component utilizes parallel depthwise separable convolutions with fixed kernel sizes of 3×3, 5×5, and 7×7 to extract features across multiple spatial resolutions. Such a design enables the model to effectively perceive ships of varying sizes, thereby improving its adaptability to scale diversity in optical remote sensing scenes.

2) Sobel-based edge enhancement branch: To reinforce contour sensitivity, fixed-weight Sobel filters are employed to extract directional edge information. Specifically, horizontal and vertical edge responses are obtained by convolving the input feature map with predefined Sobel kernels along the x and y axes. These directional responses are subsequently fused through a 1×1 convolution to generate an edge-enhanced feature representation. Here, $edge_x$ and $edge_y$ denote the horizontal and vertical edge responses, respectively, while edge represents the overall edge magnitude. A small constant varepsilon $= 1 \times 10^{-6}$ is added to ensure numerical stability during square root computation. The overall edge magnitude is then computed using the Euclidean norm, as formulated in Eq. (1):

$$edge = \sqrt{edge_x^2 + edge_y^2 + 1 \times 10^{-6}} \qquad (1)$$

3) Feature fusion module: The outputs from the multi-scale and edge branches are concatenated along the channel dimension and integrated using a 1×1 convolution to produce the final enhanced representation.

C. Attention Mechanism and Feature Enhancement

To improve the model's ability to focus on informative regions, the C2PSA (Parallel Spatial Attention) module—originally integrated into the YOLO11 architecture—is retained in the upper layers of the backbone. This attention mechanism operates by applying parallel spatial attention across feature maps, enabling the network to suppress irrelevant background noise while enhancing the response of target-relevant regions. Although not proposed in this work, the inclusion of C2PSA contributes to improved detection accuracy, particularly in complex remote sensing environments such as ports and coastal zones.

D. Loss Function and Training Strategy

To address the challenges of complex maritime environments, YOLOv11n-EGM incorporates a tailored loss weighting strategy. This approach assigns greater importance to overlapping ship instances, enhancing the model's responsiveness to occluded and densely packed targets, particularly in port areas. The edge-guided multi-scale enhancement mechanism further contributes to precise boundary localization. When combined with regression loss, it effectively reduces false detections caused by cluttered backgrounds.

The training process adheres to the standard YOLOv11n framework, utilizing the Stochastic Gradient Descent (SGD) optimizer with an initial learning rate of 0.01, adjusted dynamically through cosine annealing. While the core training methodology remains consistent, the integration of the EGM-Block and the customized loss scheme significantly improves the model's robustness and generalization capabilities. Detailed implementation settings, including data augmentation techniques and hardware configurations, are provided in Section IV.

E. Implementation Details

The proposed model is implemented using the PyTorch deep learning framework, leveraging its modular design and GPU acceleration capabilities. Input images are uniformly resized to 640×640 pixels to maintain consistency across training batches and to balance computational efficiency with spatial resolution. The model is trained with a batch size of 64, which allows for stable gradient updates while fully utilizing the memory capacity of high-performance GPUs.

To address the substantial scale variation of ship targets—ranging from small fishing boats to large cargo vessels—the architecture incorporates an Edge-Guided Multi-Scale (EGM) feature enhancement strategy. This mechanism enables the network to capture fine-grained details and contextual cues across different receptive fields, thereby improving detection accuracy for both small and large objects. In addition, the EGM-Block, embedded within the backbone, helps suppress irrelevant background features such as waves, docks, and coastal infrastructure, which often lead to false alarms in remote sensing imagery.

All experiments are conducted on the publicly available ShipRSImageNet V1.0 dataset, which provides a comprehensive benchmark for ship detection. The dataset includes 3,435 high-resolution images annotated with horizontal bounding boxes (HBB), oriented bounding boxes (OBB), and polygon masks. It is designed with a hierarchical classification structure across four task levels (Level 0 to Level 3), covering 50 distinct ship categories [5]. A detailed overview of the dataset and training configuration is presented in Section IV.

IV. Experimental Results and Analysis

To validate the effectiveness of the proposed YOLOv11n-EGM framework, a series of experiments is conducted on the ShipRSImageNet V1.0 dataset. This section presents the dataset characteristics, training configuration, evaluation metrics, and comparative results against baseline models. The experiments are designed to assess the model's performance across varying task complexities and ship categories.

A. Dataset Overview

This study utilizes the ShipRSImageNet V1.0 dataset, a large-scale, fine-grained benchmark designed for ship detection in high-resolution optical remote sensing imagery. The dataset comprises 3,435 images sourced from diverse satellite platforms and sensors, spanning various geographic regions and weather conditions. Each image measures approximately 930×930 pixels and contains ships of varying scales, orientations, and aspect ratios. A total of 17,573 ship instances are annotated using horizontal bounding boxes (HBB), oriented bounding boxes (OBB), and polygon masks, offering rich spatial and contextual information.

ShipRSImageNet V1.0 introduces a hierarchical classification structure across four levels (Level 0 to Level 3), encompassing 50 distinct ship types. The updated version includes a standardized test set to facilitate fair benchmarking and reproducibility in ship detection research [5].

B. Experimental Setup and Evaluation Metrics

All models are implemented using the PyTorch framework. Input images are resized to 640×640 pixels, with a batch size of 64. Training is conducted over 300 epochs on an NVIDIA RTX 4090 GPU with 24GB of memory. Data augmentation techniques such as random rotation and scale variation are applied to enhance generalization.

To assess detection performance, the following metrics are employed:

1) mAP@0.5 and mAP@0.5:0.95: Mean Average Precision at fixed and varying IoU thresholds.

2) HBB mAP: Detection accuracy using horizontal bounding boxes.

3) Precision, Recall, and F1 Score: Standard metrics for evaluating classification and localization quality.

Performance is further analyzed across different object sizes (small, medium, large) and classification levels.

C. Comparative Methods and Baseline Models

To rigorously assess the effectiveness of the proposed YOLOv11n-EGM architecture, a series of representative object detection frameworks are selected as baselines. These models encompass diverse backbone networks and detection paradigms. Comparative experiments are conducted across all four hierarchical levels (Level 0 to Level 3) defined in the ShipRSImageNet dataset, enabling evaluation from coarse-grained classification (ship vs. non-ship) to fine-grained ship type recognition.

D. Visual Results and Quantitative Analysis

To validate the effectiveness of the proposed YOLOv11n-EGM model, we conducted comparative experiments against several mainstream object detectors. All models were trained under identical conditions to ensure a fair and unbiased evaluation. The visual and quantitative results of these comparisons are summarized in Table I.

TABLE I. EXPERIMENTAL COMPARISON RESULTS OF DIFFERENT ALGORITHMS

Task Level	Method	$mAP_{50:95}$	mAP_{50}	mAP_{75}
Level 0: 1 ships and dock	Faster R-CNN	0.432	0.674	0.487
	Deformable DETR	0.505	0.763	0.596
	YOLOv9	0.447	0.662	0.523
	YOLOv10	0.594	0.760	0.656
	YOLOv11n-EGM	**0. 606**	**0. 775**	**0.691**
Level 1: 3 ship and dock	Faster R-CNN	0.372	0.566	0.421
	Deformable DETR	0.431	0.609	0.512
	YOLOv9	0.404	0.551	0.455
	YOLOv10	0.496	0.625	0.551
	YOLOv11n-EGM	**0.519**	**0.649**	**0.606**
Level 2: 24 ships and dock	Faster R-CNN	0.368	0.527	0.421
	Deformable DETR	0.509	0.635	0.583
	YOLOv9	0.399	0.497	0.464
	YOLOv10	0.505	0.608	0.557
	YOLOv11n-EGM	**0. 511**	**0. 618**	**0.565**
Level 3: 49 ships and dock	Faster R-CNN	0.381	0.529	0.438
	Deformable DETR	0.585	0.701	**0.760**
	YOLOv9	0.433	0.525	0.493
	YOLOv10	0.584	0.668	0.650
	YOLOv11n-EGM	**0.610**	**0.703**	0.656

1) Overall performance trends: YOLOv11n-EGM consistently demonstrates superior performance across all Task Levels and evaluation metrics. Notably, in Level 0 and Level 3 scenarios, its results in mAP(50:95) and mAP_75 significantly outperform other methods, validating the effectiveness of the proposed EGM-Block in enhancing multi-scale and edge-level feature representation. YOLOv10 ranks closely behind, indicating that the YOLO series achieves a commendable balance between model efficiency and detection accuracy. In contrast, Faster R-CNN and YOLOv9 exhibit noticeable performance degradation in more complex scenes (Level 2 and Level 3), suggesting limited adaptability to dense object distributions and occlusions.

Notably, we observed different comparative results of YOLOv11n-EGM across Level 0 to Level 3 of the ShipRSImageNet V1.0 dataset. This discrepancy is primarily due to the substantial variation in classification granularity across the hierarchical levels. Specifically, Level 0 distinguishes whether an object is a ship or not, involving only two classes. Level 1 further categorizes ship objects into four broad categories. Level 2 refines these categories into 24 subcategories, while Level 3 provides the most detailed classification, identifying 49 specific ship types and one dock class, resulting in a total of 50 categories. As the classification becomes increasingly fine-grained from Level 0 to Level 3, the complexity of the detection task escalates, placing greater demands on the network's feature extraction and discrimination capabilities. Consequently, the performance of the proposed algorithm varies across levels, reflecting its sensitivity to the granularity of the classification task.

To further illustrate the detection performance of the proposed YOLOv11n-EGM model under varying task complexities, Precision-Recall (PR) curves are plotted for Level 0 and Level 3 scenarios. These curves provide a comprehensive view of the trade-off between precision and recall across different confidence thresholds.

As shown in Fig. 2, YOLOv11n-EGM achieves a mean average precision (mAP) of 0.772 at an IoU threshold of 0.5. The model maintains high precision and recall in Level 0, where the task involves binary classification (ship vs. non-ship). This indicates strong performance in relatively simple detection scenarios with clear object boundaries and low category granularity. The category-wise analysis further shows that Dock achieves an average precision of 0.797, outperforming Ship at 0.747, likely due to the more consistent structural features of docks.

In contrast, Fig. 3 presents results from the more challenging Level 3 setting, where the model is required to distinguish among 49 ship types and docks, totaling 50 fine-grained categories. The overall mAP drops to 0.702, and the PR curves exhibit greater dispersion across classes. Despite this, the model preserves a balanced PR profile, demonstrating robustness in dense and cluttered maritime environments. The performance degradation reflects the increased difficulty posed by fine-grained classification, small object sizes, and complex backgrounds. These findings reinforce the importance of edge-aware feature extraction and multi-scale representation in enhancing detection accuracy under varying conditions.

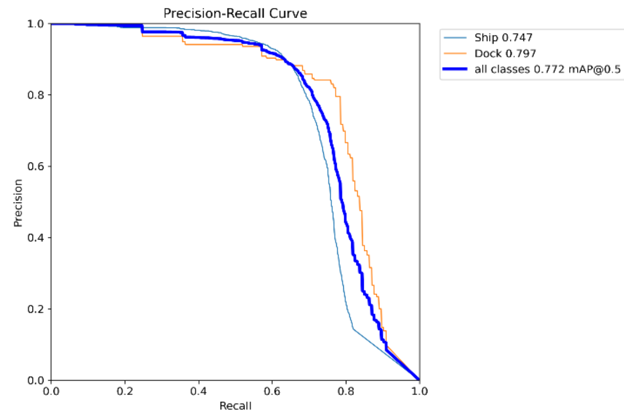

Fig. 2. Precision-Recall curve of YOLOv11n-EGM on Level 0 (lightweight scenario).

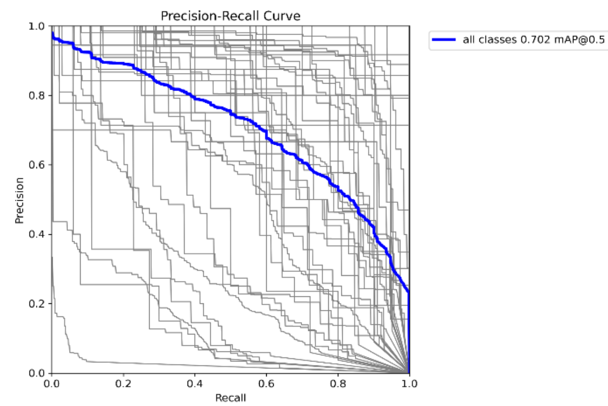

Fig. 3. Precision-Recall curve of YOLOv11n-EGM on Level 3 (complex scenario).

2) Performance variation across task levels: As the Task Level increases—corresponding to a higher number of ship and dock targets—all models experience a general decline in detection performance. However, YOLOv11n-EGM shows the least performance drop, highlighting its robustness in handling complex and cluttered environments. Specifically, in Level 3 (49 ship and dock), YOLOv11n-EGM achieves a mAP(50:95) of 0.610, surpassing both Deformable DETR (0.585) and YOLOv10 (0.584), thereby reinforcing its advantage in high-density scenarios.

To evaluate the classification accuracy across multiple ship categories, the normalized confusion matrix for Level 3 is presented in Fig. 4. This visualization highlights the model's ability to distinguish between fine-grained ship types under high-density conditions. The diagonal dominance in the matrix indicates strong classification performance, while off-diagonal elements reveal occasional misclassifications, which are primarily observed among visually similar ship classes.

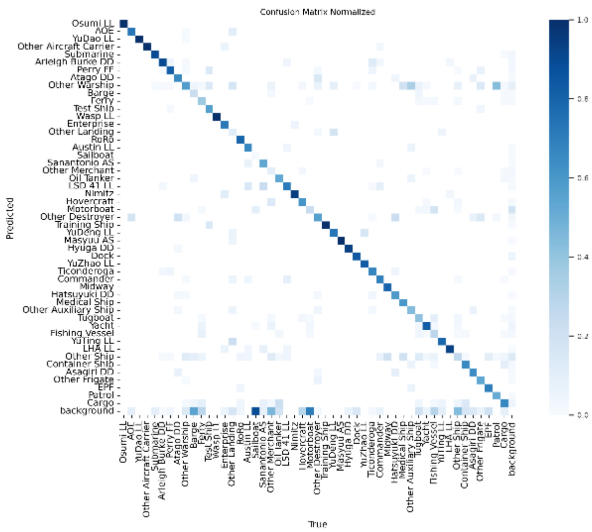

Fig. 4. Normalized confusion matrix of YOLOv11n-EGM on Level 3, illustrating multi-class classification accuracy.

The overall training dynamics of YOLOv11n-EGM are depicted in Fig. 5, which includes the evolution of key metrics such as loss, mAP, precision, and recall over 300 epochs. The smooth convergence and consistent improvement across metrics validate the stability and effectiveness of the training strategy.

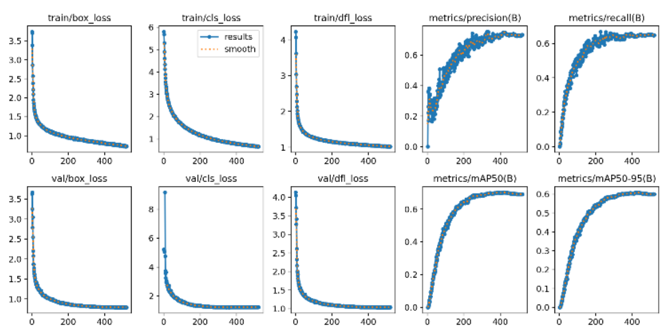

Fig. 5. Training performance of YOLOv11n-EGM over 300 epochs, including loss, mAP, precision, and recall curves.

To qualitatively assess the detection capability of YOLOv11n-EGM, Fig. 6 and Fig. 7 present sample predictions and corresponding ground truth annotations from the Level 3 validation set. The model accurately localizes and classifies multiple ship instances, even in densely packed scenes, demonstrating its practical applicability in real-world remote sensing tasks.

Notably, in several examples, both large vessels and extremely small ships appear within the same image and are successfully detected, with no significant omission. The bounding boxes closely align with the actual contours of the ships, indicating that the model effectively captures edge features and adapts to multi-scale object variations. These observations validate the effectiveness of the edge-guided multi-scale module and confirm that YOLOv11n-EGM achieves its intended improvements in fine-grained maritime detection.

However, we also observed a case where a very small ship instance (Motorboat) was misclassified as a Hovercraft. This confusion likely stems from the limited pixel-level detail available for extremely small targets, which poses a challenge for fine-grained classification.

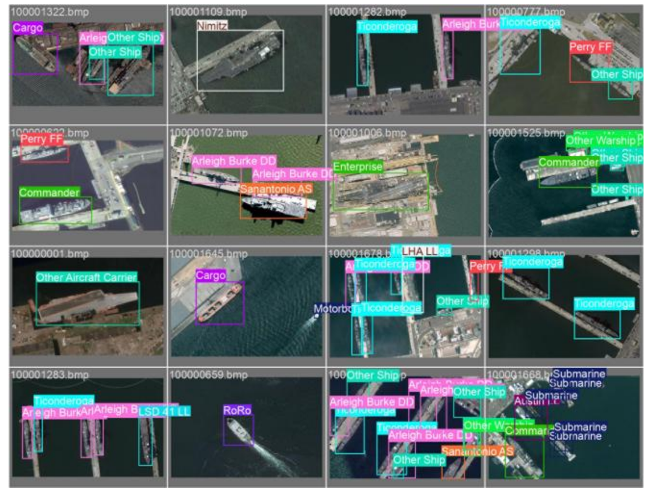

Fig. 6. Ground truth annotations on a Level 3 validation image.

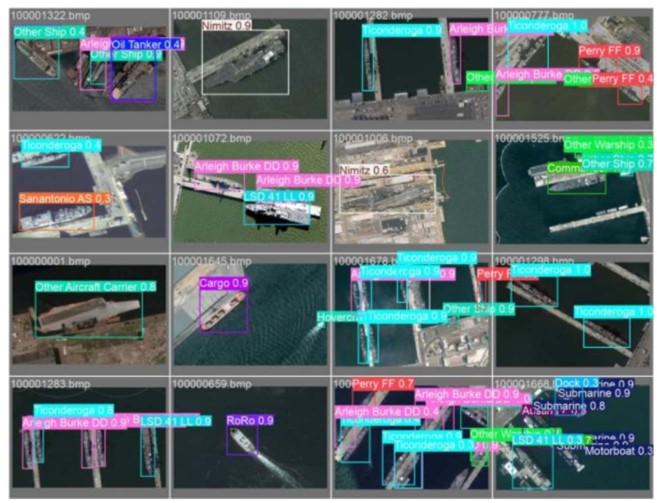

Fig. 7. YOLOv11n-EGM predictions on the same Level 3 validation image

V. DISCUSSION

The YOLOv11n-EGM model demonstrates strong performance in ship detection tasks, particularly in cluttered optical remote sensing scenes. However, we observed that very small vessels, while successfully detected, are sometimes misclassified as other ship types. This issue arises from insufficient fine-grained semantic representation, which limits the model's ability to distinguish subtle intra-class variations. Addressing this challenge may require enhancing class-specific feature discrimination or introducing adaptive refinement mechanisms. Overall, the results affirm the potential of lightweight architectures for real-time maritime monitoring, though further improvements are needed for precise recognition of small and visually ambiguous targets.

VI. Conclusion

This study presents Edge-Guided Multi-Scale YOLOv11n (YOLOv11n-EGM), an enhanced object detection framework tailored for ship detection in high-resolution remote sensing imagery. Built upon the YOLOv11n backbone, the proposed model integrates a novel EGM-Block, which simultaneously strengthens multi-scale feature extraction and edge-aware representation.

Extensive experiments were conducted on the ShipRSImageNet V1.0 dataset, which includes a hierarchical structure of four task levels (Level 0 to Level 3), encompassing a wide range of ship types and scene complexities. The results demonstrate that YOLOv11n-EGM consistently outperforms several state-of-the-art detectors across all evaluation metrics, particularly in challenging scenarios with dense and small-scale targets.

The proposed EGM-Block significantly improves the model's ability to detect fine-grained ship instances while maintaining robustness against background interference. Comparative studies further confirm the critical role of the multi-scale and edge-guided branches in enhancing detection accuracy and generalization.

A. Limitations

1) Sensitivity to small instances: Qualitative analysis reveals that the model struggles to detect extremely small ships (e.g., < 8 pixels in size), often resulting in lower recall rates. This limitation is particularly critical in high-resolution satellite imagery, where small vessels are common.

2) Spectral and weather dependency: The current model relies exclusively on optical imagery, making it vulnerable to adverse weather conditions such as dense cloud cover, fog, or nighttime scenarios. This spectral dependency limits its applicability in all-weather or low-visibility environments.

3) Computational overhead: The introduction of the edge-guided multi-scale module increases the model's complexity in terms of both parameter count and GFLOPs. This added computational burden may hinder real-time deployment on edge devices or UAVs with limited processing power.

B. Future Work

1) Enhancing small object detection: We aim to improve the detection of small ships by incorporating super-resolution pre-processing pipelines and exploring anchor-free dense prediction frameworks (e.g., keypoint-based detectors). These approaches may enhance the model's ability to localize and classify tiny instances more effectively.

2) SAR-Optical data fusion: To overcome the limitations of optical-only input, we plan to integrate Synthetic Aperture Radar (SAR) data with optical imagery. SAR's all-weather, day-and-night imaging capabilities can complement optical data, enabling more robust detection under challenging environmental conditions.

3) Model compression and acceleration: To reduce computational overhead, we will explore lightweight alternatives such as depthwise separable convolutions, Ghost modules, and neural architecture search (NAS). Additionally, knowledge distillation techniques will be investigated to transfer knowledge from the full model to a compact student model suitable for real-time inference.

References

[1] UNCTAD, Review of Maritime Transport 2024. United Nations, 2024. [Online]. Available: https://unctad.org/system/files/official-document/rmt2024_en.pdf

[2] T. Zhao *et al.*, "Ship Detection with Deep Learning in Optical Remote-Sensing Images: A Survey of Challenges and Advances," *Remote Sensing*, vol. 16, no. 7, Art. no. 7, Jan. 2024, doi: 10.3390/rs16071145.

[3] S. Wei, H. Chen, X. Zhu, and H. Zhang, "Ship detection in remote sensing image based on faster R-CNN with dilated convolution," presented at the 2020 39th Chinese control conference (CCC), IEEE, 2020, pp. 7148–7153.

[4] X. Nie, M. Duan, H. Ding, B. Hu, and E. K. Wong, "Attention mask R-CNN for ship detection and segmentation from remote sensing images," Ieee Access, vol. 8, pp. 9325–9334, 2020.

[5] Z. Zhang, L. Zhang, Y. Wang, P. Feng, and R. He, "ShipRSImageNet: A Large-Scale Fine-Grained Dataset for Ship Detection in High-Resolution Optical Remote Sensing Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 8458–8472, 2021, doi: 10.1109/JSTARS.2021.3104230.

[6] G. Wen *et al.*, "MS-SSD: multi-scale single shot detector for ship detection in remote sensing images," *Appl Intell*, vol. 53, no. 2, pp. 1586–1604, Jan. 2023, doi: 10.1007/s10489-022-03549-6.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.

[8] Q. Xu, Y. Li, and Z. Shi, "LMO-YOLO: A ship detection model for low-resolution optical satellite imagery," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 4117–4131, 2022.

[9] L. Chen, W. Shi, and D. Deng, "Improved YOLOv3 Based on Attention Mechanism for Fast and Accurate Ship Detection in Optical Remote Sensing Images," *Remote Sensing*, vol. 13, no. 4, Art. no. 4, Jan. 2021, doi: 10.3390/rs13040660.

[10] E. Yildirim and T. Kavzoglu, "Ship Detection in Optical Remote Sensing Images Using YOLOv4 and Tiny YOLOv4," in *Innovations in Smart Cities Applications Volume 5*, M. Ben Ahmed, A. A. Boudhir, İ. R. Karaș, V. Jain, and S. Mellouli, Eds., Cham: Springer International Publishing, 2022, pp. 913–924. doi: 10.1007/978-3-030-94191-8_74.

[11] X. M. Sun, Y. J. Zhang, H. Wang, and Y. X. Du, "Research on ship detection of optical remote sensing image based on Yolo V5," *J. Phys.: Conf. Ser.*, vol. 2215, no. 1, p. 012027, Feb. 2022, doi: 10.1088/1742-6596/2215/1/012027.

[12] V. Moya, A. Quito, A. Pilco, J. P. Vásconez, and C. Vargas, "Crop detection and maturity classification using a yolov5-based image analysis," *Emerg. Sci. J.*, vol. 8, no. 2, pp. 496–512, 2024.

[13] D. Chen, H. Zhao, Y. Li, Z. Zhang, and K. Zhang, "PMF-YOLOv8: Enhanced Ship Detection Model in Remote Sensing Images," Information Technology and Control, vol. 53, no. 4, Art. no. 4, Dec. 2024, doi: 10.5755/j01.itc.53.4.37003.

[14] L. He, Y. Zhou, L. Liu, W. Cao, and J. Ma, "Research on object detection and recognition in remote sensing images based on YOLOv11," Scientific Reports, vol. 15, no. 1, p. 14032, 2025.

[15] R. Khanam and M. Hussain, "YOLOv11: An Overview of the Key Architectural Enhancements," Oct. 23, 2024, *arXiv*: arXiv:2410.17725. doi: 10.48550/arXiv.2410.17725.

[16] W. Chen, B. Han, Z. Yang, and X. Gao, "MSSDet: Multi-Scale Ship-Detection Framework in Optical Remote-Sensing Images and New Benchmark," *Remote Sensing*, vol. 14, no. 21, Art. no. 21, Jan. 2022, doi: 10.3390/rs14215460.

[17] K. M. A. Aziz, "Quantitative monitoring of coastal Erosion and changes using remote sensing in a mediterranean Delta," *Civil Eng J*, vol. 10, no. 6, pp. 1842–1862, 2024.

[18] A. Nekrasov, A. Khachaturian, and C. Fidge, "Using semicircular sampling to increase sea water/ice discrimination altitude," *Emerging Science Journal*, vol. 8, no. 2, pp. 480–495, 2024.

[19] Q. Zhang, L. Wang, H. Meng, Z. Zhang, and C. Yang, "Ship Detection in Maritime Scenes under Adverse Weather Conditions," *Remote Sensing*, vol. 16, no. 9, Art. no. 9, Jan. 2024, doi: 10.3390/rs16091567.

[20] C. Lang, X. Yu, and X. Rong, "LSDNet: a lightweight ship detection network with improved YOLOv7," *Journal of Real-Time Image Processing*, vol. 21, no. 2, p. 60, 2024.

[21] P. Tian and Z. and Li, "YOLOv8-SP: ship object detection in optical remote sensing imagery," *Marine Geodesy*, vol. 0, no. 0, pp. 1–22, doi: 10.1080/01490419.2025.2493860.