YOLO-Based Ship Detection in Remote Sensing: A Practical Perspective

Shibo Yan[1, 2, a, *], Pan Liu[2, b] and Abudhahir Buhari[2, c]

[1]Department of Information Technology, Anhui Vocational College of Defense Technology, Luan 237000, China

[2]Faculty of Engineering, Science and Technology, Kuala Lumpur University of Science and Technology, Kuala Lumpur, Malaysia

[a]35311186@acdt.edu.cn, [b]242924784@s.iukl.edu.my, [c]abudhahir@iukl.edu.my

[*] Corresponding author

Abstract

Ship detection in remote sensing imagery is essential for maritime surveillance, port operations, and coastal safety. This study presents a practical evaluation of YOLO-based models tailored for real-time deployment in civilian maritime environments. The experiments utilize a curated version of the ShipRSImageNet_V1 dataset, from which all military vessels were removed to focus exclusively on civilian ships. This refinement ensures the model's applicability to non-military monitoring tasks and enhances its relevance for port and commercial vessel management. To support scalable deployment, the proposed framework integrates unmanned aerial vehicles (UAVs) for data acquisition with a collaborative edge-cloud inference strategy. Lightweight models operate onboard UAVs, while centralized GPU resources handle complex computations, enabling efficient and responsive maritime monitoring. This architecture aligns with emerging trends in low-altitude economy and supports rapid decision-making in dynamic coastal scenarios. Comparative results demonstrate that YOLOv8n, YOLOv11n, and YOLOv12n achieve a favorable balance between detection accuracy and computational efficiency. Future research will focus on improving model robustness, expanding domain-specific datasets, and incorporating multi-modal data to enhance performance in diverse operational conditions.

Keywords

Ship Detection; Remote Sensing Imagery; YOLO Object Detection; Civilian Maritime Monitoring.

1. INTRODUCTION

Efficient monitoring of maritime activities in port areas is essential for ensuring operational safety, optimizing logistics, and supporting intelligent infrastructure management. With the increasing availability of high-resolution optical remote sensing imagery, automated ship detection has become a viable solution for enhancing situational awareness in coastal environments [1]. Among various object detection frameworks, the YOLO (You Only Look Once) family of models has gained significant attention due to its balance between speed and accuracy, making it suitable for real-time applications [2-3].

While many existing studies focus on ship detection in open sea or military contexts, practical port surveillance requires models that can handle complex backgrounds, densely packed vessels, and diverse civilian ship types. In this study, we evaluate multiple versions of the YOLO algorithm—namely YOLOv8n, YOLOv11n, and YOLOv12n—for ship detection in optical remote sensing images, with a particular emphasis on civilian maritime scenarios [4]. The experiments are conducted on a curated subset of the ShipRSImageNet_V1 dataset, from which all military vessels have been removed to focus exclusively on civilian ships, ensuring relevance to non-military monitoring tasks [5].

Considering the practical requirements of real-world deployment, this study is designed to be compatible with edge-cloud collaborative inference architectures. Unmanned aerial vehicles (UAVs) are used for data acquisition, where lightweight models perform initial detection onboard, and more complex computations are offloaded to centralized GPU servers [6]. This setup supports scalable, efficient, and responsive maritime monitoring, aligning with emerging trends in low-altitude economic applications.

Rather than introducing new model architectures, this study focuses on a comparative analysis of existing YOLO variants to assess their practicality and adaptability for port-related ship detection. The findings aim to support the development of lightweight and deployable ship detection systems that integrate deep learning with remote sensing technologies for intelligent port management.

2. RELATED WORK

2.1. Ship Detection in Remote Sensing Imagery

Ship detection has long been a critical task in maritime surveillance, environmental monitoring, and port management. Traditional approaches often relied on handcrafted features and classical image processing techniques, such as edge detection, texture analysis, and thresholding. While these methods offered interpretability and low computational cost, their performance was highly sensitive to background complexity, lighting conditions, and object scale.

With the advancement of machine learning, especially convolutional neural networks (CNNs), ship detection has seen significant improvements in both accuracy and robustness. Early deep learning-based methods typically employed two-stage detectors like Faster R-CNN, which achieved high precision but suffered from slower inference speeds. In contrast, single-stage detectors such as SSD and YOLO emerged as efficient alternatives, offering real-time performance with competitive accuracy [7].

Recent studies have focused on enhancing detection performance in complex coastal environments, where ships often appear in cluttered backgrounds, vary in size, and exhibit occlusion. Techniques such as multi-scale feature fusion, attention mechanisms, and rotation-invariant detection have been proposed to address these challenges. However, many of these methods are computationally intensive and may not be suitable for real-time deployment in port surveillance systems.

2.2. YOLO-Based Detection in Remote Sensing Applications

The YOLO (You Only Look Once) series has become one of the most widely adopted object detection frameworks in remote sensing due to its speed, simplicity, and end-to-end design. From YOLOv1 to the latest lightweight variants, the architecture has evolved to incorporate deeper backbones, improved feature aggregation, and better anchor box strategies, making it increasingly effective for detecting small and densely packed objects in aerial and satellite imagery.

In the context of remote sensing, YOLO has been applied to a variety of tasks, including vehicle detection, building extraction, and ship recognition. Its ability to process large-scale images efficiently makes it particularly suitable for applications where real-time or near-real-time analysis is required, such as UAV-based monitoring and smart port systems.

Several studies have benchmarked YOLO against other detection models on datasets like HRSC2016, DIOR, and DOTA, demonstrating its competitive performance in terms of both accuracy and inference speed [8]. However, most existing work either focuses on military ship detection or general object detection in remote sensing, with limited emphasis on practical port surveillance scenarios involving civilian vessels.

This study aims to bridge that gap by evaluating YOLO-based models in the context of port monitoring, using publicly available datasets that approximate real-world conditions. Rather than proposing new architectures, we focus on assessing the practicality of existing YOLO variants for deployment in intelligent port management systems.

3. METHODOLOGY

3.1. Model Selection

To evaluate the practicality of YOLO-based ship detection in port surveillance scenarios, we selected three lightweight variants of the YOLO architecture: YOLOv8n, YOLOv11n, and YOLOv12n. These models are specifically designed for lightweight and efficient inference, making them highly suitable for deployment on edge-side monitoring systems within intelligent port environments.

YOLOv8n serves as a baseline model, featuring a streamlined backbone and simplified detection head. YOLOv11n introduces enhancements in feature aggregation and spatial representation, aiming to improve detection accuracy in cluttered environments [3]. YOLOv12n further incorporates multi-scale fusion and edge-aware mechanisms, which are beneficial for identifying small and densely packed vessels. Despite architectural differences, all three models maintain a consistent input-output format, allowing for fair comparison under identical training conditions.

These models are not only computationally efficient but also adaptable to various deployment environments. In practical port surveillance systems, lightweight YOLO variants can be deployed on edge devices such as UAVs for real-time ship detection, while more powerful versions may operate on cloud servers at port command centers. This flexibility supports a collaborative edge-cloud architecture, enabling responsive and scalable monitoring solutions [9].

3.2. Dataset Description

The primary dataset used in this study is a curated subset of ShipRSImageNet_V1, refined to focus exclusively on civilian vessels. All military ship categories were deliberately removed, resulting in a dataset that better reflects non-military port surveillance scenarios. The selected subset includes 1,576 optical remote sensing images, annotated with 6,758 ship instances across 14 civilian ship categories, which have been reindexed from the original category IDs (37–50) to a new range of 0–13.

To facilitate model training and evaluation, the dataset was re-partitioned into three subsets: 945 images for training, 315 images for validation, and 316 images for testing. Each image is annotated using horizontal bounding boxes (HBB), which are compatible with standard YOLO training pipelines. This dataset configuration ensures a realistic and focused benchmark for assessing ship detection performance in civilian maritime environments [5].

Figure 1 shows the distribution of ship instances across the 14 civilian categories, providing insight into the dataset's class balance.

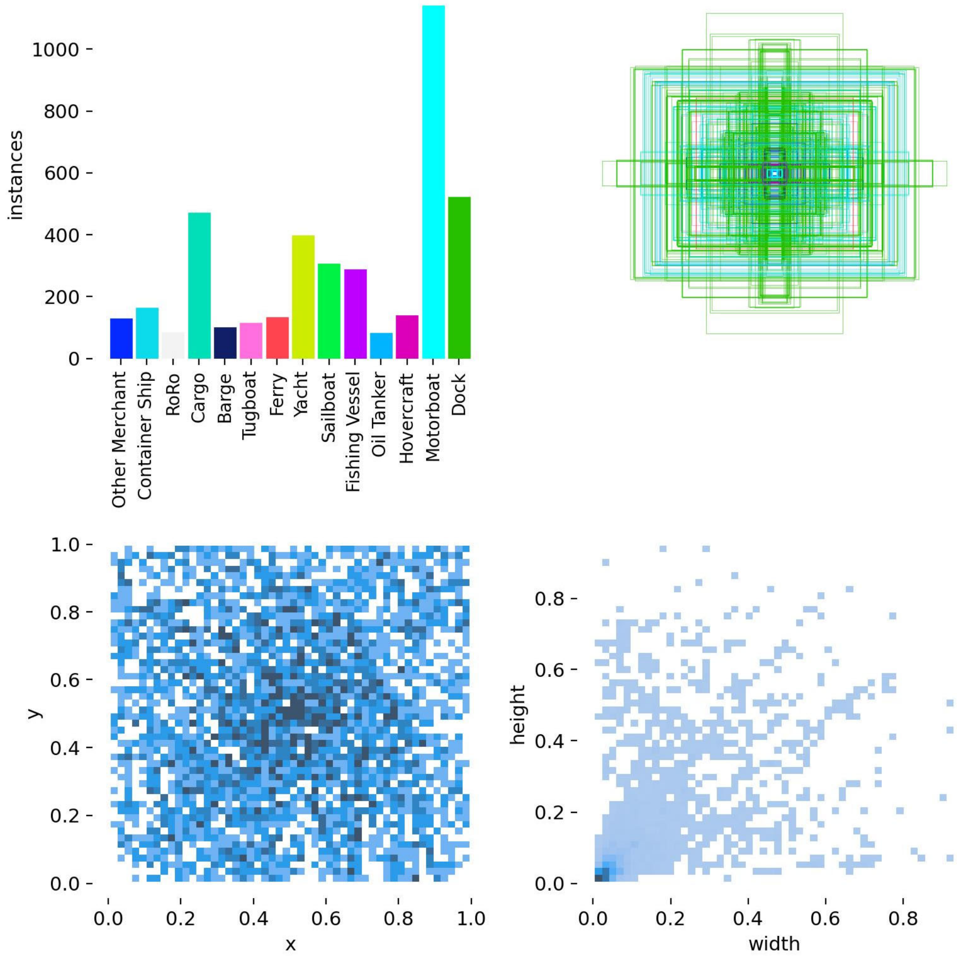

Figure 1. Distribution of civilian ship categories in the curated ShipRSImageNet_V1 subset

3.3. Data Preprocessing

All images were resized to 640×640 pixels to match the input requirements of the selected YOLO models. Data augmentation techniques such as random horizontal flipping, scaling, and brightness adjustment were applied to improve model generalization. The datasets were split into training and testing subsets using an 80/20 ratio. Annotation files were converted into YOLO format, with each line representing a bounding box and its corresponding class label.

3.4. Training Configuration

Each model was trained using the same hyperparameter settings to ensure consistency in evaluation. The training was conducted for 300 epochs with a batch size of 64 and an initial learning rate of 0.001. All experiments were performed on a workstation equipped with an NVIDIA RTX 4090 GPU.

During training, we monitored key performance metrics including mean Average Precision (mAP), precision, recall, and inference speed to assess model effectiveness. The mAP is computed as the mean of the Average Precision (AP) across all object classes, defined as:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \qquad (1)$$

Where N is the number of object classes, and APiAPi represents the Average Precision for class ii, calculated from the precision-recall curve. In this study, we report both **mAP@0.5** and

mAP@0.5:0.95, where the latter averages AP over IoU thresholds ranging from 0.5 to 0.95 in steps of 0.05, providing a more comprehensive evaluation of detection performance.

4. EXPERIMENTS AND RESULTS

4.1. Model Performance on ShipRSImageNet_V1 (Civilian Subset)

To evaluate the effectiveness of lightweight YOLO models in civilian ship detection tasks, we conducted experiments using the curated civilian subset of the ShipRSImageNet_V1 dataset described in Section 3.2. All models were trained under identical conditions for 300 epochs. The evaluation focused on key detection metrics, including mean Average Precision at IoU thresholds of 0.5 (mAP@0.5) and 0.5:0.95 (mAP@50:95), as well as precision and recall. The results are summarized in Table 1.

Table 1. Performance Comparison of YOLOv8n, YOLOv11n, and YOLOv12n on ShipRSImageNet_V1 (Civilian Subset)

Metric	YOLOv8n	YOLOv11n	YOLOv12n
$\boldsymbol{mAP}_{50:95}$	0.340	0.409	0.393
$\boldsymbol{mAP}_{50}$	0.462	0.528	0.516
Precision	0.647	0.677	0.775
Recall	0.468	0.527	0.518

The results indicate that YOLOv11n achieves the highest overall detection accuracy, particularly in terms of mAP@0.5 and mAP@50:95. YOLOv12n demonstrates superior precision, suggesting strong confidence in its predictions. YOLOv8n, while slightly lower in accuracy, still performs competitively and may offer advantages in speed and resource efficiency, which are further illustrated in the visual analysis section.

To further illustrate the performance of YOLOv11n, we present a summary of its training and evaluation metrics, including precision, recall, and mean Average Precision (mAP) curves. These metrics provide a comprehensive view of the model's detection capabilities throughout the training process, as shown in Figure 2.

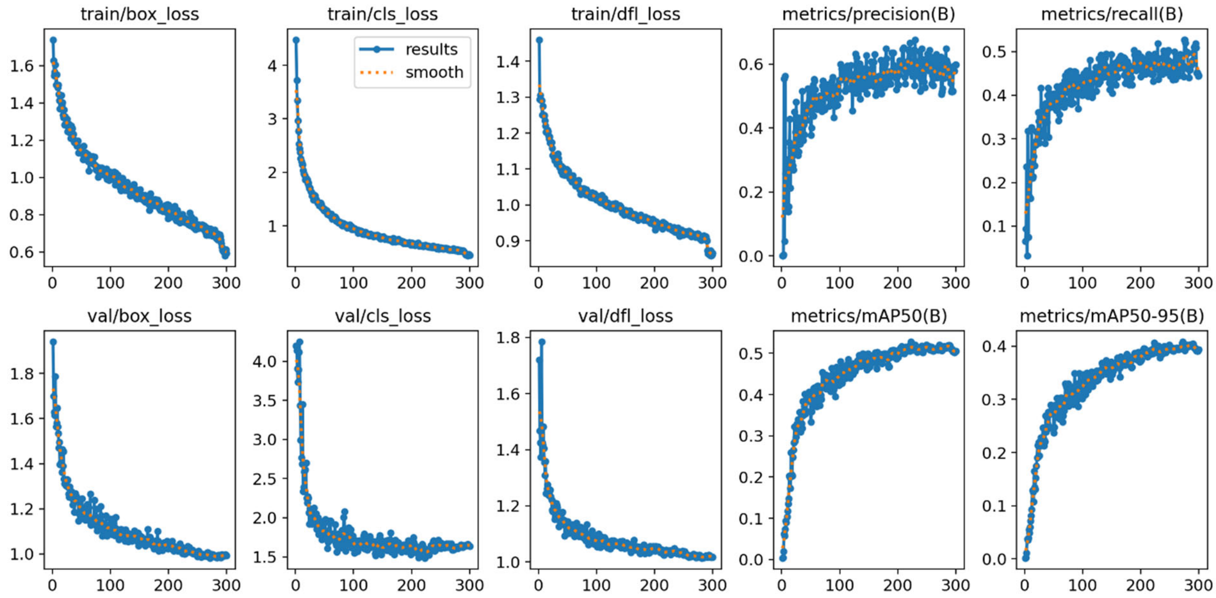

Figure 2. Detection performance overview of YOLOv11n

4.2. Detection Visualizatio

To qualitatively assess detection performance, we visualized the outputs of each model on representative images from the test set. As shown in Figure 5, both models demonstrate reasonable localization and confidence in detecting civilian ships. YOLOv11n shows slightly better performance in cluttered scenes.

To qualitatively assess the detection performance of YOLOv11n, we visualized its predictions on representative test samples. As shown in Figure 3, the model demonstrates accurate localization and reasonable confidence in identifying civilian ships, even in moderately cluttered scenes.

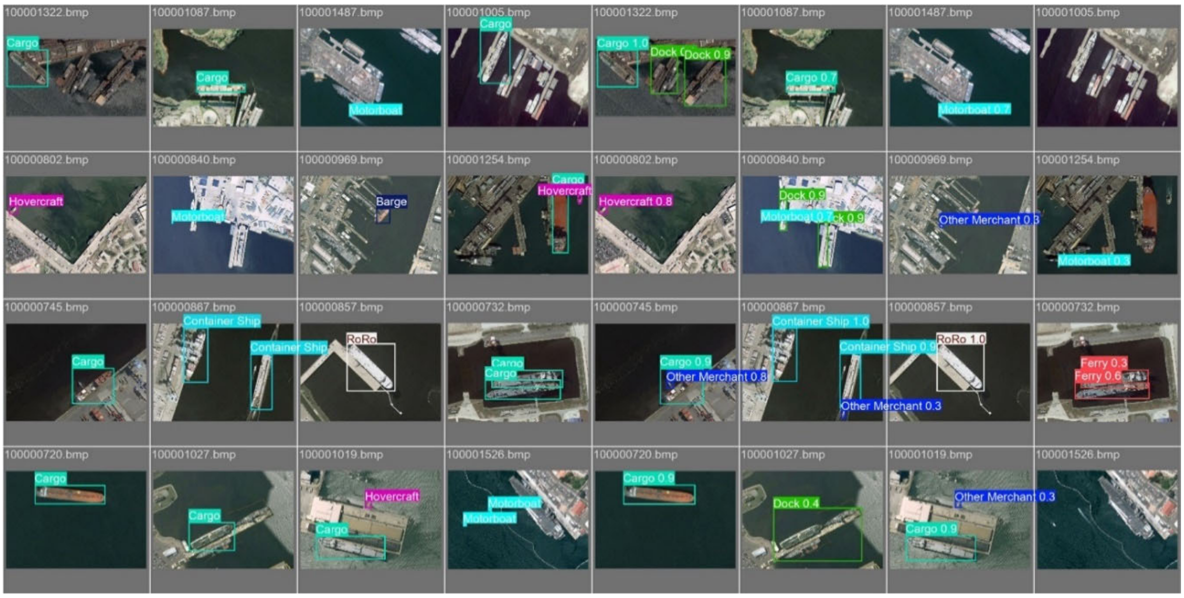

Figure 3. YOLOv11n detection results on a test image: ground truth vs. prediction

4.3. Port Scenario Analysis

In real-world port environments, ship detection systems must operate under diverse conditions, including occlusion, background clutter, and variable lighting. The curated ShipRSImageNet_V1 subset provides a realistic approximation of such challenges. Our experiments demonstrate that both YOLOv8n and YOLOv11n are capable of handling these conditions effectively.

Furthermore, the lightweight nature of these models enables deployment on edge devices such as UAVs. When combined with cloud-based processing at port command centers, this edge-cloud collaborative architecture supports scalable, responsive, and intelligent ship detection systems aligned with the development of low-altitude economic applications.

To further evaluate the model’s robustness in realistic port environments, we analyzed the classification performance using a normalized confusion matrix. As shown in Figure 4, YOLOv11n maintains consistent accuracy across ship categories, even under conditions of occlusion and background clutter. This supports its applicability in edge-cloud collaborative systems for port surveillance.

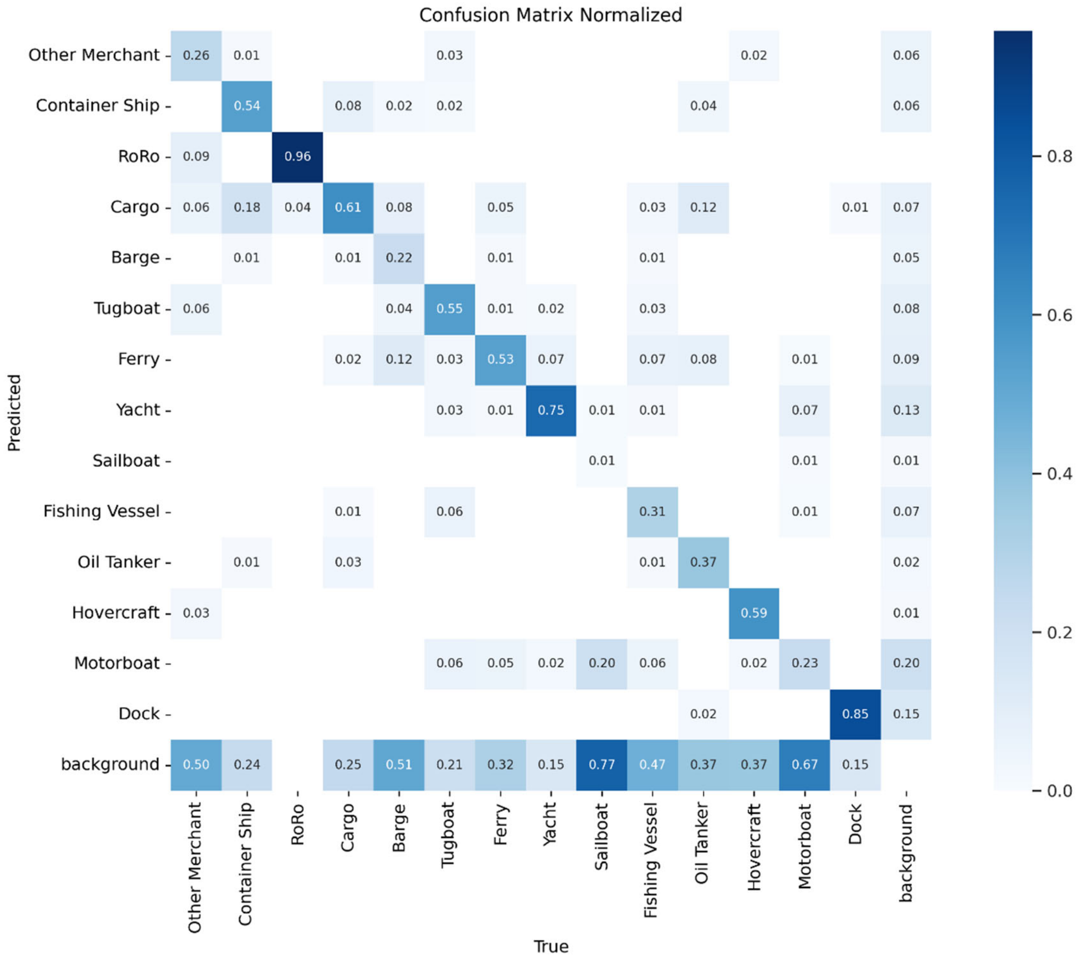

Figure 4. Normalized confusion matrix of YOLOv11n

5. DISCUSSION

5.1. Feasibility of Model Deployment

The experimental results on the ShipRSImageNet_V1 civilian subset confirm the practical feasibility of deploying lightweight YOLO models for ship detection in remote sensing. This dataset captures diverse maritime conditions, making it suitable for evaluating real-world applicability.

In operational scenarios, a collaborative edge-cloud architecture can support efficient deployment. UAVs equipped with lightweight YOLO variants (e.g., YOLOv8n, YOLOv11n, YOLOv12n) can perform real-time inference at the edge, while port command centers handle further analysis using high-performance computing. This setup enhances responsiveness and aligns with low-altitude economic development, leveraging UAV flexibility and 5G-enabled data transmission for intelligent port monitoring.

5.2. Limitations

Despite promising results, several challenges remain:

- Dataset Coverage: While ShipRSImageNet_V1 offers realistic scenes, it may not fully reflect dynamic port conditions such as varying weather and lighting.
- Model Sensitivity: Lightweight models may underperform in detecting small or densely packed ships, especially in cluttered backgrounds.

5.3. Future Work

To address the limitations identified above, future research can explore the following directions:

- Dedicated Dataset Construction: Develop a specialized port surveillance dataset by collecting diverse UAV imagery, covering scenarios such as nighttime operations, foggy conditions, and multi-scale ship distributions.
- Model Enhancement: Integrate advanced modules such as attention mechanisms, edge-aware components (e.g., EGM-Block), or Transformer-based backbones to improve detection accuracy and robustness.
- Adaptive Deployment Strategies: Investigate dynamic switching mechanisms between edge and cloud models based on computational load, image complexity, and network conditions.
- Multimodal Fusion: Combine visual data with auxiliary sources such as AIS (Automatic Identification System) or radar signals to enhance detection reliability and reduce false positives.

REFERENCES

[1] Zhao, T., Wang, Y., Li, Z., Gao, Y., Chen, C., Feng, H., & Zhao, Z. (2024). Ship detection with deep learning in optical remote-sensing images: A survey of challenges and advances. *Remote Sensing*, *16*(7), 1145.

[2] Hussain, M. (2023). YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. *Machines*, *11*(7), 677.

[3] Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*.

[4] Tian, Y., Wang, X., Zhu, S., Xu, F., & Liu, J. (2023). LMSD-Net: a lightweight and high-performance ship detection network for optical remote sensing images. *Remote Sensing*, *15*(17), 4358.

[5] Zhang, Z., Zhang, L., Wang, Y., Feng, P., & He, R. (2021). 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*, *14*, 8458-8472.

[6] Yuan, Y., Gao, S., Zhang, Z., Wang, W., Xu, Z., & Liu, Z. (2024). Edge-cloud collaborative UAV object detection: Edge-embedded lightweight algorithm design and task offloading using fuzzy neural network. *IEEE Transactions on Cloud Computing*, *12*(1), 306-318.

[7] Lu, X., Li, B., Yue, Y., Li, Q., & Yan, J. (2019). Grid r-cnn. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7363-7372).

[8] Qing, Y., Liu, W., Feng, L., & Gao, W. (2021). Improved Yolo network for free-angle remote sensing target detection. *Remote Sensing*, *13*(11), 2171.

[9] Min, L., Dou, F., Zhang, Y., Shao, D., Li, L., & Wang, B. (2025). CM-YOLO: Context modulated representation learning for ship detection. *IEEE Transactions on Geoscience and Remote Sensing*.