

Research Article

Research on Automatic Navigational Buoy Recognition Based on YOLOv11

Yaoming Wei¹ , Weijia Lu¹ , Chengdong Chu^{2,*} , Xiaoming Jia³ 

¹Merchant Marine College, Shanghai Maritime University, Shanghai, China

²Shanghai Maritime Pilots' Association, Shanghai, China

³DongYing Maritime Safety Administration, DongYing, China

Abstract

With the rapid development of intelligent ships, deep learning-based automatic identification of maritime buoys has emerged as a critical research direction for maritime intelligence. To address the challenge of balancing real-time performance and detection accuracy in complex inland waterway environments, this paper proposes an automatic identification method based on the YOLOv11 object detection algorithm. Specifically, by integrating advanced C3K2 and C2PSA modules, the model's capability for feature extraction and global information perception in cluttered backgrounds is significantly enhanced. To mitigate the scarcity of data samples, data augmentation techniques-including rotation and Gaussian noise elimination-were applied to construct buoy dataset, which consists of 913 high-quality annotated images. Furthermore, an incremental learning strategy with multi-stage iterative training was introduced to improve the model's generalization across diverse scenarios. Experimental results demonstrate that while maintaining high-efficiency real-time response, the proposed model achieves a mAP of 93%. This performance outperforms traditional algorithms such as Cascade-RCNN and SSD, as well as previous versions like YOLOv8, providing robust technical support for safe collision avoidance and waterway situational awareness in intelligent shipping.

Keywords

Intelligent Ships, Buoy Identification, YOLOv11, Deep Learning, Artificial Intelligence

1. Introduction

1.1. Background and Motivation

With the rapid advancement of intelligent and autonomous ships, perception capability has become a fundamental requirement for automatic collision avoidance and navigational decision-making. Accurate detection and recognition of navigational aids are essential for maritime situational awareness and safety; however, complex sea states, variable illumination,

and the diverse visual characteristics of navigational marks make robust and high-precision recognition challenging in real-world environments. Although deep learning-based computer vision methods have demonstrated strong real-time detection performance and have been increasingly applied to maritime perception, existing studies mainly focus on generic targets such as vessels and obstacles. Navigational buoys, as critical infrastructure for waterway safety, possess unique visual signatures that are often obscured by waves or distorted by

*Correspondence: Chengdong Chu (pilotccd@163.com)

Received: 24 January 2026; **Accepted:** 4 February 2026; **Published:** 14 February 2026



Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

light reflection. Therefore, developing a specialized detection framework that can maintain high precision under dynamic and resource-constrained maritime conditions is of paramount importance for the next generation of intelligent maritime transportation systems. This study aims to bridge this gap by leveraging the latest YOLOv11 architecture to achieve a more reliable and efficient buoy recognition system.

1.2. Literature Review

Current deep learning-based object detection models can generally be classified into two categories: single-stage and two-stage detectors. Single-stage detection algorithms, such as SSD [1], RetinaNet [2], and the YOLO series [3], directly perform object localization and classification in a unified framework, offering superior inference speed. In contrast, two-stage algorithms, including Fast R-CNN [4], Faster R-CNN [5], and Mask R-CNN [6], first generate region proposals and then conduct refined classification, which often limits their real-time applicability on moving vessels. In the maritime domain, existing target recognition studies have predominantly focused on ship detection in inland waterways, employing methods such as LiDAR-based recognition [7] and YOLO-based visual frameworks [8]. For example, Sun Haitao optimized the anchor generation of YOLOv4 using K-Means

and introduced a BiFPN module to enhance feature extraction [9]. Xiao Zheng and Wang Jiye improved unmanned surface vessel recognition by incorporating multi-scale attention mechanisms into YOLOv5 [10]. Similarly, Yang Zhiyuan and Luo Liang et al. enhanced performance by introducing lightweight asymmetric detection heads and a Slim-FPN structure into YOLOv8 [11].

Despite these advancements, research specifically targeting navigational mark recognition remains limited. Wang Ning and Zhi Min pointed out that while the YOLO series offers notable advantages in computational efficiency and deployment, models still face challenges on resource-constrained devices. In comparison, SSD demonstrates strong generalization but its shallow layers limit deep semantic feature extraction, while RetinaNet suffers from high complexity [12]. Although Ni Hanjie and Chu Xiumin et al. attempted to apply improved Faster R-CNN models for navigational mark detection [13, 14], current research still exhibits significant gaps, particularly in addressing the influence of flashing characteristics and precise recognition under long-distance observation. Consequently, this study adopts the YOLO framework, specifically focusing on the systematic investigation of distinct buoy categories to provide a more comprehensive solution for maritime navigation.

2. Research Method

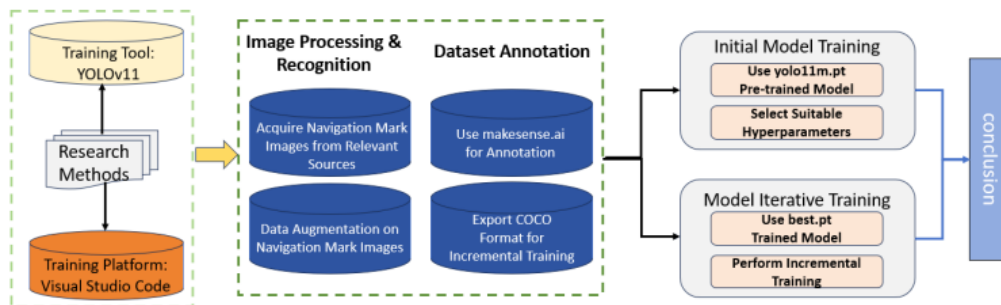


Figure 1. Research method.

2.1. Introduction of YOLOv11

In terms of model architecture, YOLOv11 introduces several key optimizations to enhance its feature representation and processing efficiency. Specifically, the backbone incorporates C3K2 modules, which utilize specialized kernel structures to boost feature extraction depth. To improve the prioritization of critical spatial information, a C2PSA module is

strategically positioned following the SPPF block; this integration of the PSA mechanism allows the network to suppress background clutter and focus on salient targets, such as maritime buoys. Additionally, the detection head leverages depth-wise separable convolutions, effectively minimizing computational redundancy while maintaining high inference speed. These advancements make YOLOv11 highly suitable for demanding real-time maritime perception tasks. The comprehensive structure of the network is depicted in Figure 2 [15].

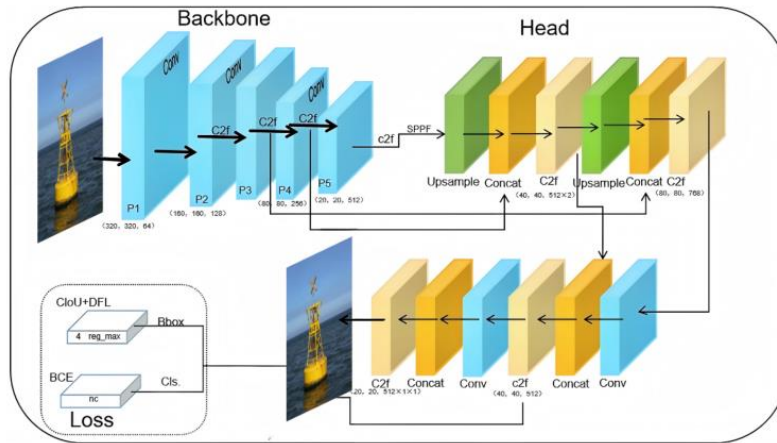


Figure 2. The overall architecture of YOLOv11.

2.2. Automatic Navigational Mark Recognition

2.2.1. Dataset Preparation and Preprocessing

After data collection, the dataset undergoes data cleaning and preprocessing. The primary objective of data cleaning is to remove noisy samples, including low-quality images and incorrectly annotated data. Image processing techniques [16], such as edge detection and morphological operations [17], are employed to ensure that navigational marks can be clearly distinguished in the images. To further improve dataset quality, image standardization is applied by adjusting parameters such as image size and brightness, which contributes to enhancing the model's recognition performance under varying environmental

conditions.

Dataset preprocessing also includes data augmentation, which is a key strategy for improving model robustness. Common augmentation techniques, including rotation, cropping, and color transformation, are used to increase sample diversity and enable the model to better adapt to variations in viewpoint, illumination, and environmental conditions. For instance, random rotation and scaling allow the model to learn features of navigational marks at different orientations and distances, while color transformations improve adaptability to diverse lighting conditions. These augmentation strategies enhance the model's generalization capability and effectively reduce the risk of overfitting. Figure 3 illustrates examples of navigational mark images after data augmentation.

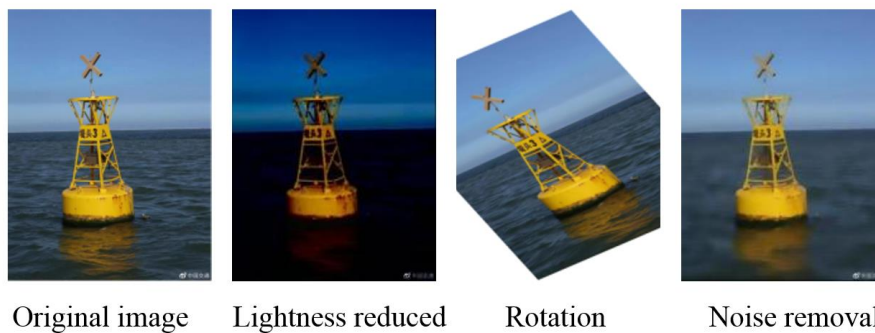


Figure 3. Navigational mark images after data augmentation.

The dataset construction involves collecting navigational mark images under diverse conditions to ensure comprehensive coverage of different illumination, weather, and sea states, including daytime and nighttime scenes as well as clear, foggy, and rainy weather and varying wave conditions. Images are precisely annotated using a labeling tool to define the category and location of each navigational mark, and the annotations

are exported in COCO or YOLO formats to support subsequent training. Data augmentation is applied to expand the dataset and improve model generalization, while preprocessing steps such as normalization and resizing standardize the input data and enhance training efficiency. Figure 4 shows the annotation interface of the makesense ai platform.

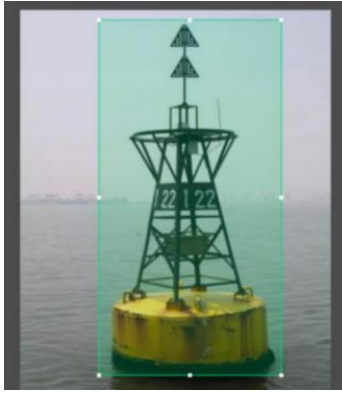


Figure 4. Annotation interface.

2.2.2. Training the Initial Model

Table 1 lists the training environment.

Table 1. Training environment.

Operating System	Windows11
Graphics Card	NVIDIA RTX 4070
CPU	INTEL i5-12800HX
Memory	16GB
Visual Studio Code	1.96.4
Python	3.9.21
Pytorch	2.5.1+cu124
Ultralytics	8.3.43

YOLOv11 was selected as the baseline model due to its balanced performance in accuracy and real-time inference, making it suitable for automatic navigational mark recognition. Data augmentation techniques, including rotation, translation, scaling, cropping, and color transformation, were applied to enhance model generalization under varying environmental and illumination conditions. Model performance and training efficiency were regulated through appropriate hyper-parameter settings, including epochs, batch size, patience, number of workers, optimizer, momentum, and learning rate. The detailed hyper-parameter configurations are listed in Table 2.

Table 2. Hyper-parameter configurations.

Hyper-parameter	Parameter value
epoch	50
batch size	-1 (auto)
patience	10

Hyper-parameter	Parameter value
workers	6
optimizer	SGD
momentum	0.9
lr	0.001

2.2.3. Iterative Model Training

Incremental learning is adopted during model training to efficiently handle environmental variations and the introduction of new navigational mark categories. Unlike conventional deep learning approaches that require retraining the entire network, incremental learning enables YOLOv11 to incorporate new samples and classes with low computational cost while preserving performance on previously learned categories. Given the diversity of navigational marks, this strategy allows progressive category expansion and improves model generalization.

Experimental results show that mini-batch incremental training with a small number of newly added samples per iteration enables the YOLOv11 model to rapidly adapt to changing marine environments, achieving a recognition accuracy exceeding 90%. In addition, incremental learning significantly reduces computational overhead by selectively updating model parameters, which is particularly beneficial in resource-constrained scenarios. Model optimization is further achieved through iterative fine-tuning guided by validation performance, including early stopping and targeted data augmentation for underperforming classes. During the training phase, to bridge the performance gap between initial results and the final model, a multi-stage incremental learning strategy was applied. The model was initially trained with a batch size of 16 for 100 epochs, achieving a baseline accuracy of 90%. Subsequently, by increasing the batch size to 32 and reducing the learning rate for fine-tuning, the model refined its detection of small-scale buoy targets, ultimately elevating the accuracy to 93%. The hyper-parameter settings for incremental learning are listed in Table 3.

Table 3. Incremental hyper-parameter configurations.

Hyper-parameter	Parameter value
epoch	50
batch size	-1 (auto)
patience	20
workers	6
optimizer	SGD
momentum	0.9

Hyper-parameter	Parameter value
lr	0.001

3. Results

In studies on YOLOv11-based automatic navigational mark recognition, evaluating model performance is a critical factor in determining its practical applicability. Accordingly, various loss functions and evaluation metrics are analyzed. The `box_loss` reflects the positional discrepancy between predicted and ground-truth bounding boxes and is used to optimize bounding box regression accuracy. Lower values indicate smaller localization errors and more accurate target positioning.

The `cls_loss` measures the model's accuracy in target cate-

gory classification, with lower values indicating reduced classification errors and improved classification performance. The `df1_loss`, an enhanced form of classification loss, focuses on improving the recognition of hard samples and rare categories; lower values suggest superior performance in these challenging detection tasks.

Precision reflects the proportion of correctly predicted positive samples among all positive predictions, indicating the reliability of the model's detection. Higher precision corresponds to fewer false positives. Recall measures the proportion of actual targets correctly detected by the model, representing detection coverage; higher recall indicates fewer missed detection. The mAP provides a comprehensive evaluation of localization and classification performance across different IoU thresholds and is widely used in object detection tasks. Higher mAP values indicate better overall detection performance. Figure 5 presents the performance curves obtained after 50 training epochs.

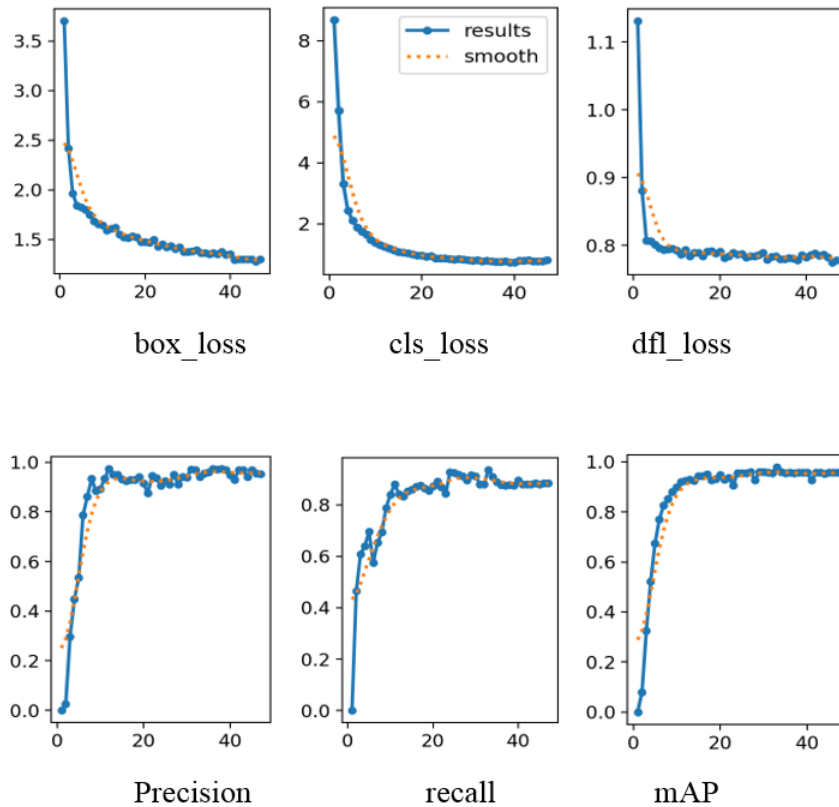


Figure 5. Data analysis chart.

The main evaluation metrics rely on precision, recall and the mAP curve. In the following formulas, (1) is the calculation formula for precision, (2) is the calculation formula for recall, and (3) and (4) are the calculation formulas for mAP, (5) (6) (7) The three formulas respectively introduce `box_loss`, `cls_loss` and `df1_loss`.

$$\text{precision} = \frac{TP}{TP + FP} \quad (1)$$

In the formula, TP represents the number of samples correctly predicted as positive, while FP represents the number of samples wrongly predicted as positive.

$$\text{recall} = \frac{TP}{TP + FN} \quad (2)$$

In the formula, FN represents the number of positive samples wrongly predicted as negative.

$$AP = \frac{1}{N} \sum_{i=1}^N \text{Precision}(i) \quad (3)$$

In the formula, N is the number of categories, and $\text{Precision}(i)$ is the precision rate of the i -th category.

$$mAP = \frac{1}{K} \sum_{k=1}^K AP_k \quad (4)$$

In the formula, K is the number of categories, and AP_k is the average precision of category k .

$$L_{box} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (5)$$

In the formula, IoU represents Intersection over Union, the ratio of the overlap area between the predicted box and the ground truth box. ρ^2 represents the square of the Euclidean distance between the centers of the predicted box and the ground truth box. c represents the diagonal length of the smallest closed rectangle that can cover both the predicted box and the ground truth box. v measures the consistency of the aspect ratio. α represents weight parameter.

$$L_{cls} = - \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1-y_i) \log(1-\hat{y}_i)] \quad (6)$$

In the formula, y_i represents the true label (either 0 or 1). $\hat{y}_i$ represents the probability predicted by the model for this category (after applying the Sigmoid activation function).

$$L_{DFL}(P_i, P_{i+1}) = -(y_{i+1} - y) \log(P_i) + (y - y_i) \log(P_{i+1}) \quad (7)$$

In the formula, y represents the actual coordinate offset. y_i and y_{i+1} represent two discrete integer coordinates immediately adjacent to the actual value y . P_i and P_{i+1} represent the probability values predicted by the model at the points y_i and y_{i+1} .

The final recognition performance of the model is strongly influenced by the training dataset. In current navigational mark recognition tasks, training samples are often dominated by common mark types, while certain categories, such as special marks, have relatively few examples. This imbalance can lead to poor learning and increased misclassification for underrepresented classes. To improve model generalization, it is essential to expand the dataset to include navigational marks captured under diverse marine conditions. In this study, a dataset of 913 images covering all navigational mark categories was constructed, including variations in illumination, weather, and sea states, as well as augmented images. The model was trained for 50 epochs, and the experimental data are summarized in Table 4, with comparative results shown in Table 5 [17].

Table 4. Experimental data table.

No.	Beacon Name	Number of Images	Successfully Identified	Identification Accuracy Rate
1	Port Hand Mark	89	83	93.2%
2	Starboard Hand Mark	72	66	91.7%
3	Preferred Channel Port Hand Mark	98	92	93.9%
4	Preferred Channel Starboard Hand Mark	78	72	92.3%
5	North Cardinal Mark	82	75	91.5%
6	South Cardinal Mark	105	98	93.3%
7	East Cardinal Mark	93	86	92.5%
8	West Cardinal Mark	79	73	92.4%
9	Isolated Danger Mark	65	59	90.8%
10	Safe Water Mark	82	79	96.3%
11	Special Mark	70	67	95.7%
12	recall			91%

No.	Beacon Name	Number of Images	Successfully Identified	Identification Accuracy Rate
13	mAP			93%

Table 5. Comparative results.

Algorithm	mAP/%
Cascade-RCNN	92.11
Faster-RCNN	91.36
SSD	89.02
YOLOv11	93

The results show that YOLOv11 achieves over 90% accuracy for all navigational mark categories. Safe water marks and special marks exhibit the highest accuracies of 96.3% and 95.7%, respectively, due to their distinctive appearances. In contrast, isolated danger marks have fewer samples and a relatively lower accuracy of 90.8%. The accuracies of lateral marks and cardinal marks are similar. The overall mAP reaches 93%, outperforming other algorithms. Additionally, YOLOv11's fast inference capability provides reliable support for real-time navigational mark monitoring at sea.

4. Discussion

Existing automatic navigational mark recognition systems are primarily based on the YOLOv11 model. Although promising recognition results have been achieved, there remain several aspects that require further improvement and optimization. A detailed analysis of the model architecture and data processing pipeline can help identify potential enhancement strategies, thereby improving the effectiveness and applicability of the system in real-world deployments.

In YOLOv11-based navigational mark recognition, the selection of training strategies and the implementation of learning rate scheduling have a direct impact on model performance and accuracy. An appropriately configured learning rate determines both convergence speed and final generalization capability. An excessively high learning rate may cause training instability or divergence, whereas an overly low learning rate can lead to prolonged training time and convergence to local optima. Therefore, adopting dynamic learning rate adjustment strategies during training is essential for achieving stable and efficient optimization.

Incorporating multi-scale feature fusion can further enhance the model's ability to recognize navigational marks of varying sizes. The YOLOv11 framework employs a feature

pyramid network to adaptively weight feature channels, enabling better handling of scale variations and improved understanding of complex background scenes. This capability provides more reliable decision support for navigational mark recognition in diverse maritime environments.

In practical applications, navigational marks are often observed under complex and dynamic environmental conditions. Sensor fusion techniques, such as integrating radar or LiDAR with visual camera data, can significantly improve recognition accuracy and robustness. By combining high-precision ranging information with visual perception, the system can not only enhance target localization accuracy but also estimate the distance between vessels and navigational marks, thereby improving navigational safety. Such multimodal fusion approaches strengthen interference resistance and ensure stable detection and tracking performance under varying weather and illumination conditions.

5. Conclusions

This study demonstrates that YOLOv11-based automatic navigational mark recognition can effectively address the limitations of traditional manual methods in complex maritime environments. By classifying navigational marks and analyzing their visual features, the proposed approach achieves over 90% accuracy, with particularly strong performance on distinct mark types such as safe water and special marks. Data augmentation further improves robustness under varying weather and lighting conditions, while adaptive learning rate strategies enhance training stability and convergence. However, real-world deployment may still be affected by dynamic sea conditions and image noise, necessitating scenario-adaptive testing and parameter optimization to ensure reliable performance in practical navigation systems.

Future work will optimize the YOLOv11 architecture and feature extraction, expand datasets, and explore diversified data augmentation to improve recognition accuracy and robustness in complex marine environments. For practical deployment, edge computing should be adopted to reduce latency and enable real-time processing on floating marks or vessels. Night-time recognition remains a key challenge due to low visibility and indistinct appearance, requiring targeted enhancement. Additionally, integrating stereo vision or LiDAR ranging can improve distance measurement and overall system performance, supporting maritime safety management and ocean monitoring.

Abbreviations

YOLO	You Only Look Once
SSD	Single Shot MultiBox Detector
Fast R-CNN	Fast Region-based Convolutional Network
Faster R-CNN	Faster Region-based Convolutional Network
Mask R-CNN	Mask Region-based Convolutional Network
PSA	Point-wise Spatial Attention
box_loss	Bounding Box Loss
cls_loss	Classification Loss
dfl_loss	Distribution Focal Loss
mAP	Mean Average Precision
IoU	Intersection over Union
TP	True Positive
FP	False Positive
FN	False Negative

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector [C] Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
- [2] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection [C] Proceedings of the IEEE international conference on computer vision. 2017: 2980–2988. <https://doi.org/10.48550/arXiv.1708.02002>
- [3] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection [C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779–788.
- [4] Girshick R. Fast r-cnn [C]. Proceedings of the IEEE international conference on computer vision. 2015: 1440–1448.
- [5] Ren S. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. arXiv preprint arXiv: 1506.01497, 2015. <https://doi.org/10.1109/TPAMI.2016.2577031>
- [6] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision (ICCV). New York: IEEE Press, 2017: 2980–2988. <https://doi.org/10.48550/arXiv.1703.06870>
- [7] WANG Guihuai, XIE Shuo, LIU Chenguang, et al. Obstacle recognition method for inland unmanned ship based on LiDAR [J]. Optical Technique, 2018, 44(05): 602–608. <https://doi.org/10.13741/j.cnki.11-1879/o4.2018.05.015>
- [8] XU Yanwei, LI Jun, DONG Yuanfang, et al. Survey of YOLO series object detection algorithms [J]. Journal of Frontiers of Computer Science and Technology, 2024, 18(09): 2221–2238. <https://doi.org/10.3778/j.issn.1673-9418.2402044>
- [9] SUN Haitao. Research on ship target detection and recognition method based on improved YOLO network [D]. Yanshan University, 2023. <https://doi.org/10.27440/d.cnki.gysdu.2023.001616>
- [10] XIAO Zheng, WANG Jiye, XIA Yeliang. Ship recognition method for inland waterway based on MSAM-YOLOv5 [J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2023, 51(05): 67–73+118. <https://doi.org/10.13245/j.hust.230511>
- [11] YANG Zhiyuan, LUO Liang, WU Tianyang, et al. Lightweight optical remote sensing image ship target detection algorithm based on improved YOLOv8 [J]. Computer Engineering and Applications, 2024, 60(16): 248–257.
- [12] WANG Ning, ZHI Min. Survey of single-stage general object detection algorithms under deep learning [J/OL]. Journal of Frontiers of Computer Science and Technology, 1–32 [2025-04-19].
- [13] NI Hanjie. Research on association rules of inland waterway navigation mark accidents and image recognition algorithms [D]. Wuhan University of Technology, 2021. <https://doi.org/10.27381/d.cnki.gwlg.2021.001226>
- [14] NI Hanjie, CHU Xiumin, ZHANG Binpeng, et al. Inland waterway navigation mark detection algorithm based on improved Cascade-RCNN [J]. Navigation of China, 2022, 45(03): 99–105.
- [15] WANG Guoguo, GONG Xiaoyu, YUAN Fating. Lightweight algorithm for insulator defect detection based on YOLO11s [J/OL]. Journal of Chongqing Technology and Business University (Natural Science Edition), 1–10 [2025-02-07].
- [16] Han Q, Li SH, Hou, MY, et al. A novel OL-mapping operator-based edge detection approach. NEUROCOMPUTING. Volume 618. <https://doi.org/10.1016/j.neucom.2024.129078>
- [17] Li ZX, Li Y, Zhang K. Adaptive feature extraction of underwater target signal based on mathematical morphology feature enhancement. JOURNAL OF VIBRATION AND CONTROL. <https://doi.org/10.1177/10775463241263923>