



# Ship Detection and Recognition Based on Improved YOLOv7

Wei Wu<sup>1</sup>, Xiulai Li<sup>2</sup>, Zhuhua Hu<sup>1</sup> and Xiaozhang Liu<sup>3,\*</sup>

<sup>1</sup>School of Information and Communication Engineering, Hainan University, Haikou, 570228, China

<sup>2</sup>School of Cyberspace Security, Hainan University, Haikou, 570228, China

<sup>3</sup>School of Computer Science and Technology, Hainan University, Haikou, 570228, China

\*Corresponding Author: Xiaozhang Liu. Email: lxzh@hainanu.edu.cn

Received: 24 February 2023; Accepted: 18 April 2023; Published: 09 June 2023

**Abstract:** In this paper, an advanced YOLOv7 model is proposed to tackle the challenges associated with ship detection and recognition tasks, such as the irregular shapes and varying sizes of ships. The improved model replaces the fixed anchor boxes utilized in conventional YOLOv7 models with a set of more suitable anchor boxes specifically designed based on the size distribution of ships in the dataset. This paper also introduces a novel multi-scale feature fusion module, which comprises Path Aggregation Network (PAN) modules, enabling the efficient capture of ship features across different scales. Furthermore, data preprocessing is enhanced through the application of data augmentation techniques, including random rotation, scaling, and cropping, which serve to bolster data diversity and robustness. The distribution of positive and negative samples in the dataset is balanced using random sampling, ensuring a more accurate representation of real-world scenarios. Comprehensive experimental results demonstrate that the proposed method significantly outperforms existing state-of-the-art approaches in terms of both detection accuracy and robustness, highlighting the potential of the improved YOLOv7 model for practical applications in the maritime domain.

**Keywords:** Ship position prediction; target detection; YOLOv7; data augmentation techniques

## 1 Introduction

Ship recognition is a technology that analyzes image features, such as color, shape, and texture, and has numerous applications in the field of intelligent transportation [1–3]. Due to the development of inland waterway transport modes and the increasing density of inland waterway traffic, automatic ship identification and tracking systems face significant challenges [4,5]. Traditional ship identification methods are no longer sufficient to cope with the complexity of the traffic environment. For instance, these methods cannot identify vessels without information on sailing time, direction, or speed. Additionally, existing vessel identification methods require extensive preprocessing of image data, leading to large data volumes and storage space requirements.



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The most commonly used ship recognition technologies in the field of intelligent transportation include Automatic Identification System (AIS) based on computer vision [6,7], Global Positioning System (GPS) based on Light Detection and Ranging (LiDAR) [8,9], and Human-Aided Inertial System (HAIS) based on Electronic Chart Display and Information System (ECDIS) [10,11]. Although these ship recognition methods based on communication and navigation equipment have their advantages, they also have significant limitations in traffic-intensive waters such as ports because they cannot obtain visual images of ships. Currently, research on deep learning-based ship recognition methods has made progress through automatic extraction of ship image features and continuous learning and training, achieving ship recognition in maritime traffic videos and images [12–15]. Deep learning-based ship recognition methods mainly adopt two approaches: introducing deep learning algorithms into ship recognition and fully utilizing image information through image pre-processing. Recently, deep learning-based object detection methods, such as You Only Look Once (YOLO), have achieved remarkable success in ship detection and recognition. However, YOLOv7 models have limitations in adapting to diverse ship shapes and sizes. Therefore, in this paper, we propose an improved YOLOv7 model that addresses these challenges and achieves better performance in ship detection and recognition.

This paper makes significant contributions in several areas. First, the YOLOv7 model structure is improved by replacing anchor boxes with more suitable ones for ship detection and adding a multi-scale feature fusion module to better capture ship features of different scales. Second, data preprocessing techniques such as data augmentation and sample balancing are adopted to increase data diversity, robustness, training efficiency, and model generalization. Third, extensive experiments on benchmark datasets demonstrate the proposed method's effectiveness in achieving state-of-the-art performance in terms of both detection accuracy and efficiency. Overall, the paper provides a valuable contribution to the field of ship detection and recognition.

The paper is structured as follows: Section 2 reviews related works on target detection. Section 3 explains the improvements made to the YOLOv7 model. Section 4 outlines the experiment methodology. Section 5 presents the results and discussion. Finally, Section 6 provides a summary of the paper.

## 2 Related Works

Target detection algorithms represented by deep learning can be divided into two main categories: two-stage target detection algorithms and one-stage target detection algorithms. Implementing two-stage target detection involves two processes, extracting the object region and then performing CNN classification on the region to identify it. One-stage target detection requires only one extraction of features to achieve target detection. In contrast, one-stage target detection allows for fast detection requirements but may be slightly less accurate than dual-stage target detection.

### 2.1 Two-Stage Target Detection

The two-stage target detection framework generally consists of two stages: candidate region extraction and detection. In 2014, Girshick et al. [16] at UC Berkeley proposed the Region-based Convolutional Neural Network (R-CNN) algorithm, which significantly outperformed the contemporary OverFeat algorithm [17]. R-CNN employs the Selective Search (SS) algorithm [18] to select candidate frames; each frame is then fed separately into the Convolutional Neural Network (CNN) to extract features. Finally, the bounding box is predicted based on regression and Support Vector Machine (SVM) classification. In 2015, He et al. proposed the Spatial Pyramid Pooling Network (SPP-Net) algorithm [19], which introduces a spatial pyramid pooling layer between the convolutional layer

and the fully connected layer. This replaces the cropping and scaling operations in R-CNN, ensuring consistent image sizes. SPP-Net accelerates computation and reduces computational costs. In the same year, Girshick [20] from Microsoft Research proposed the Fast R-CNN algorithm. By taking the whole image as input and passing it through the CNN network, Fast R-CNN draws on the SPP-Net concept and efficiently addresses the R-CNN algorithm's need to crop and scale image regions to the same size through the pooling layer structure of Region of Interest (ROI) Pooling. Subsequently, Ren et al. [21] proposed the Faster R-CNN algorithm, accompanied by the Region Proposal Networks (RPN) [22]. This reduced computational overhead and resolved the delay in generating positive and negative sample candidate frames for the Fast R-CNN algorithm. In 2017, He et al. made another breakthrough in the field of target detection by proposing the Mask R-CNN [23]. The Mask R-CNN algorithm replaces the ROI Pooling layer with ROI Align and adds a branching Fully Convolutional Network (FCN) layer for semantic segmentation based on border recognition.

## 2.2 One-Stage Target Detection

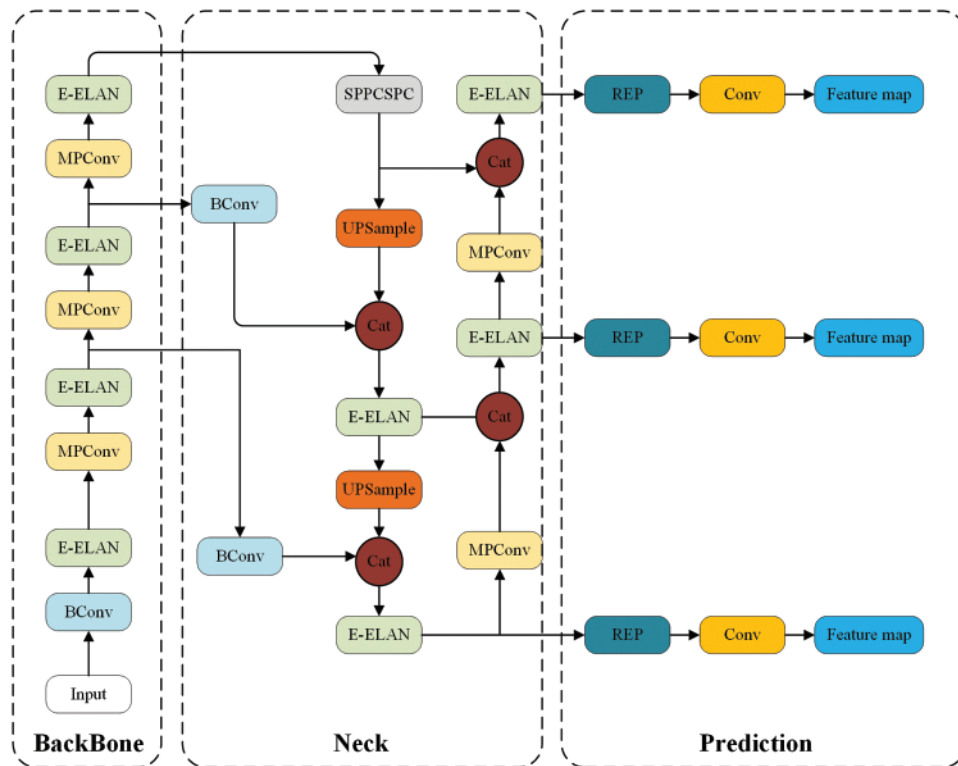
Two-stage target detection methods are often unsuitable for practical application scenarios due to the need for initial extraction of target candidate regions followed by detection. This results in higher model complexity and increased computational effort to meet system detection speed requirements. Consequently, one-stage target detection methods, such as Single Shot MultiBox Detector (SSD) and the YOLO series, are widely used in the industry for their fast computing speed, lightweight design, and suitability for deployment. YOLO [24] treats target detection as a regression problem, dividing the image into an  $S \times S$  grid and predicting the detection frame information and class probability of the object within each grid. Since then, one-stage detection algorithms, such as YOLOv2 [25], YOLOv3 [26], YOLOv4 [27], YOLOv5, YOLOv6 [28], and YOLOv7 [29], have been proposed and gained significant attention. The YOLOv3 network is one of the more mature and classical target detection networks, improving on various aspects of YOLO and combining the advantages of many network structures, such as ResNet and Feature Pyramid Network (FPN). YOLOv4 provides a systematic analysis of data pre-processing, detection network design, and prediction network processes. Based on these analyses, it designs an efficient target detector suitable for a single graphics card. YOLOv5 offers four different sizes of target detectors to meet the needs of various applications. Single Shot Detector (SSD) [30] series is also representative of one-stage target detection methods. Detection of objects at different scales is achieved by using feature maps of different layers to detect objects of varying scales. Objects of different sizes are detected at different resolutions, with high-resolution feature maps detecting small-scale objects and low-resolution feature maps detecting large-scale objects. In addition, researchers have made numerous improvements to the SSD [31–34] foundation.

## 3 Algorithm Description

### 3.1 YOLOv7

The earliest YOLO algorithm (YOLOv1) was released in 2015 and achieved one-stage detection for the first time, ensuring not only higher recognition accuracy but also faster operation than previous two-stage methods. The YOLOv7 network consists of four modules, namely input, backbone, head and prediction (see Fig. 1). In the real world, the input images are often of different sizes and not fixed. When performing target detection, they need to be resized to a fixed size. The function of the input module is to scale the image to meet the size requirements of the Backbone, which consists of several BConv convolutional layers, Efficient Local Attention Network (E-ELAN) convolutional layers and Mixed Precision Convolutional (MPConv) convolutional layers. The E-ELAN convolutional layer

maintains the original ELAN design architecture and learns more diverse features by guiding the computational blocks of different feature groups to improve the learning ability of the network without destroying the original gradient path. The MPCConv convolutional layer adds a Maxpool layer to the BConv layer, forming two branches, the upper branch halves the image aspect through Maxpool and halves the image channel through the BConv layer. The lower branch uses the first BConv layer to halve the image channels and the second BConv layer to halve the image aspect, and finally the Cat operation is used to fuse the features extracted from the upper and lower branches to improve the feature extraction capability of the network. The Head module consists of a Path Aggregation Feature Pyramid Network (PAFPN) structure, which makes it easier to pass information from the bottom to the top by introducing bottom-up paths, thus achieving efficient fusion of features at different levels. The Prediction module uses the Reparameterized Convolutional Block (REP) structure to adjust the number of image channels for the three different scales of P3, P4 and P5 features output from PAFPN, and finally  $1 \times 1$  convolution for confidence, category and anchor frame prediction.



**Figure 1:** YOLOv7 network structure

### 3.2 Improved YOLOv7

Firstly, we have made improvements to the YOLOv7 model architecture. Specifically, we have taken the following measures: Replacing Anchor boxes. In the traditional YOLOv7 model, the Anchor boxes are fixed and cannot adapt to targets of different sizes and shapes. In order to better adapt to the varying sizes and irregular shapes of ships in ship detection tasks, we designed a set of Anchor boxes that are more suitable for ship detection based on the size distribution of ships in the dataset, thereby improving detection accuracy. Adding a Multi-Scale Feature Fusion Module. To better capture ship

features at different scales, we added a Multi-Scale Feature Fusion Module. Specifically, we added the PAN (Path Aggregation Network) module to the YOLOv7 model to fuse feature maps from different levels, thereby improving the accuracy and robustness of ship detection.

Secondly, we have made improvements to data preprocessing. Specifically, we have used the following methods: Data Augmentation. To address issues such as complex lighting conditions and image noise in maritime environments, we have used data augmentation techniques, including random rotation, random scaling, random cropping, and other methods to increase the diversity and robustness of the data. Through data augmentation, we can improve the model's generalization ability without increasing the labeled data. Dataset Balancing. To address the problem of imbalanced positive and negative samples in the dataset, we balanced the dataset. Specifically, we used random sampling to reduce the size of the dataset while ensuring that the ratio of positive and negative samples was balanced, thereby improving training efficiency and the model's generalization ability.

#### 4 Experiment

To evaluate the ship recognition performance of the improved YOLOv7 algorithm, this experiment was trained and tested on the SeaShips dataset.

**Introduction to the dataset.** The Seaships dataset [35] was collected to train and evaluate the capability of the target detection algorithm in ship detection. 7000  $1920 \times 1080$  images were collected in SeaShips. Each image in the SeaShips dataset is annotated with the exact ship label and bounding box, as shown in Fig. 2. The SeaShips dataset was created from images taken by an on-site video surveillance system deployed around Hengqin Island, Zhuhai, China. The images selected for the dataset images cover different features including different ship types, hull sections, scale, viewing angles, lighting and different levels of occlusion in various complex environments. Different degrees of obscuration in complex environments.

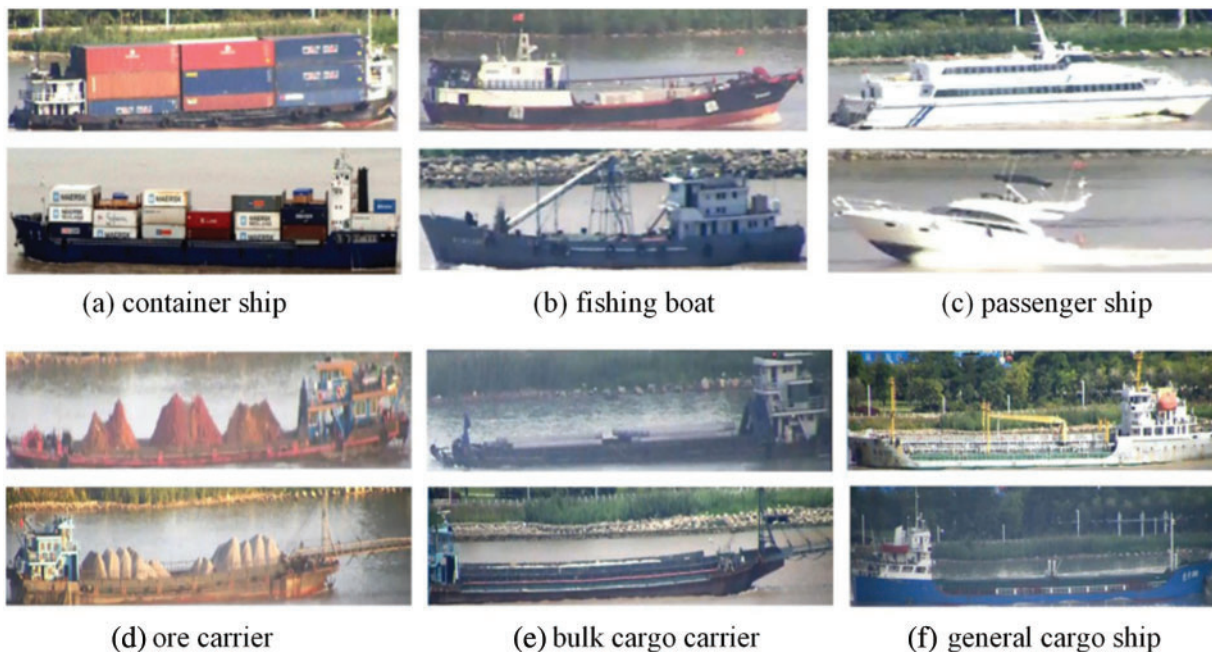


Figure 2: Different categories of Seaships

**Experimental equipment.** Parameters of the model training device used in the experiments: OS, Windows 10; GPU, NVIDIA GeForce RTX 3080 Ti; CPU, 8× Xeon E5-2686 v4; deep learning framework, Torch 1.9.0+CUDA 11.1.

**Experimental parameter settings.** Due to the limitations of the experimental equipment, we scaled the length and width of the input image to 1/2 of the original size, i.e.,  $960 \times 540$  pixels. The optimizer used SGD, the learning rate was set to 0.01, the momentum was 0.9, the weight decay was 0.0005, the batch size was 16, and 200 epochs were trained, with 10 training epochs and 1 test epoch alternating.

**Assessment methods.** Evaluation metrics are essential for assessing the performance of a model. In this paper, Precision, Recall, and mAP (mean average precision of the entire class) have been chosen as evaluation metrics. All three metrics have a range of values from 0 to 1, where the closer the value is to 1, the better the detection accuracy and model performance. The experiments employed mAP@0.5 (mean Average Precision, with IoU threshold greater than 0.5) as the evaluation metric to compute the mean average precision for all classes. We chose to use precision instead of accuracy because, in many cases, precision is a more suitable metric for evaluating machine learning models, particularly when working with imbalanced datasets where the number of positive and negative samples is unequal. Accuracy can be deceptive in such cases, as it may appear high solely due to the abundance of negative samples, while the model's ability to correctly identify positive samples may be poor.

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

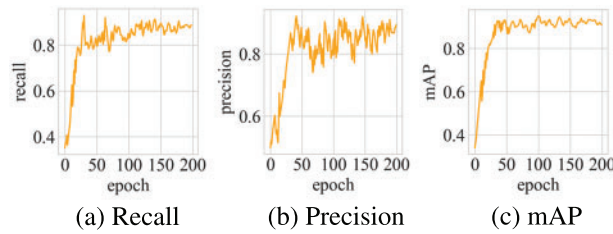
$$Recall = \frac{TP}{TP + FN} \quad (2)$$

In the above equation, TP indicates the number of predictions that agree with the true class; FP indicates the number of samples where the true class is negative but the prediction is positive; and FN indicates the number of samples where the true class is positive but the prediction is negative. The full class average precision is obtained by taking a weighted average of the average correctness of all class tests, and its value can also be expressed as the area enclosed by Precision and Recall as the two axes of a right-angle coordinate system respectively.

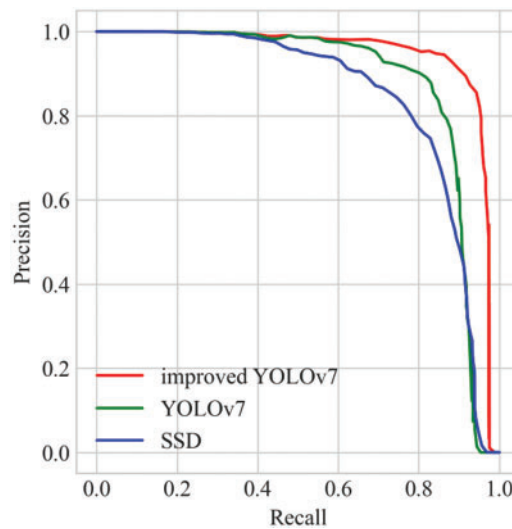
## 5 Result and Discussion

During the training phase of the model, the changes in Recall can be observed from Fig. 3a, while the changes in Precision can be seen in Fig. 3b, and the changes in mAP can be derived from Fig. 3c. After 100 epochs of training, the magnitudes of changes in Precision and Recall gradually decreased. However, due to certain labeling errors and uneven data distribution in the experimental dataset, Precision and Recall fluctuated greatly during the training process. The changes in mAP gradually decreased after about 50 iterations of the model. After 200 iterations of the improved YOLOv7 algorithm, the final mAP of the model was maintained at 90.15%.

Fig. 4 presents the precision-recall (P-R) curve plots for each model in the field of ship identification. It is evident that the area enclosed by the P-R curve formed by the improved YOLOv7 model is significantly larger than the other two models. It is well-known that the area enclosed by the P-R curve and the coordinate axis is the mAP value, and a higher mAP value indicates better detection performance of the model. Therefore, the results demonstrate that the improved YOLOv7 outperforms the other models in the field of ship identification detection.

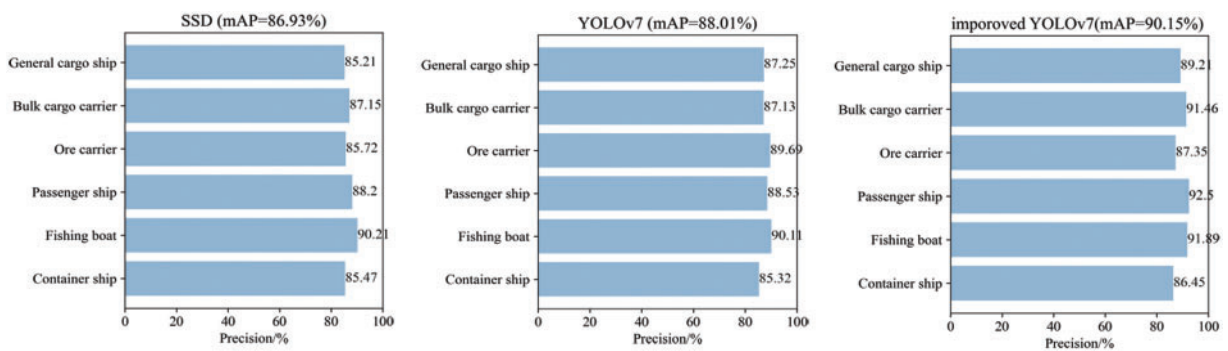


**Figure 3:** Improved YOLOv7 performance evaluation metrics



**Figure 4:** P-R curve comparison

Fig. 5 displays the category accuracy and mAP values of the three models for six different categories of boats. As shown in the figure, the detection accuracy of the improved YOLOv7 algorithm reaches 90.15%, with a particularly high mAP for small fishing boats at 91.89%, which is significantly better than the SSD and YOLOv7 models. This result highlights the advantages of the improved YOLOv7 model in inland waterway ship identification detection, which can meet the detection needs for intelligent ship navigation.



**Figure 5:** mAP for each ship category under different algorithms

Fig. 6 showcases some of the ship detection results obtained by the improved YOLOv7 algorithm. The results demonstrate that the algorithm is capable of accurately detecting different types of ships, regardless of their size. In Figs. 6b and 6d, we can see that the algorithm can identify both large and small ships with high confidence. Moreover, the improved YOLOv7 algorithm is capable of detecting multiple targets in a single image and accurately identifying each target category separately, as demonstrated in Fig. 6f. These results showcase the effectiveness and robustness of the proposed algorithm in detecting ships in complex environments.

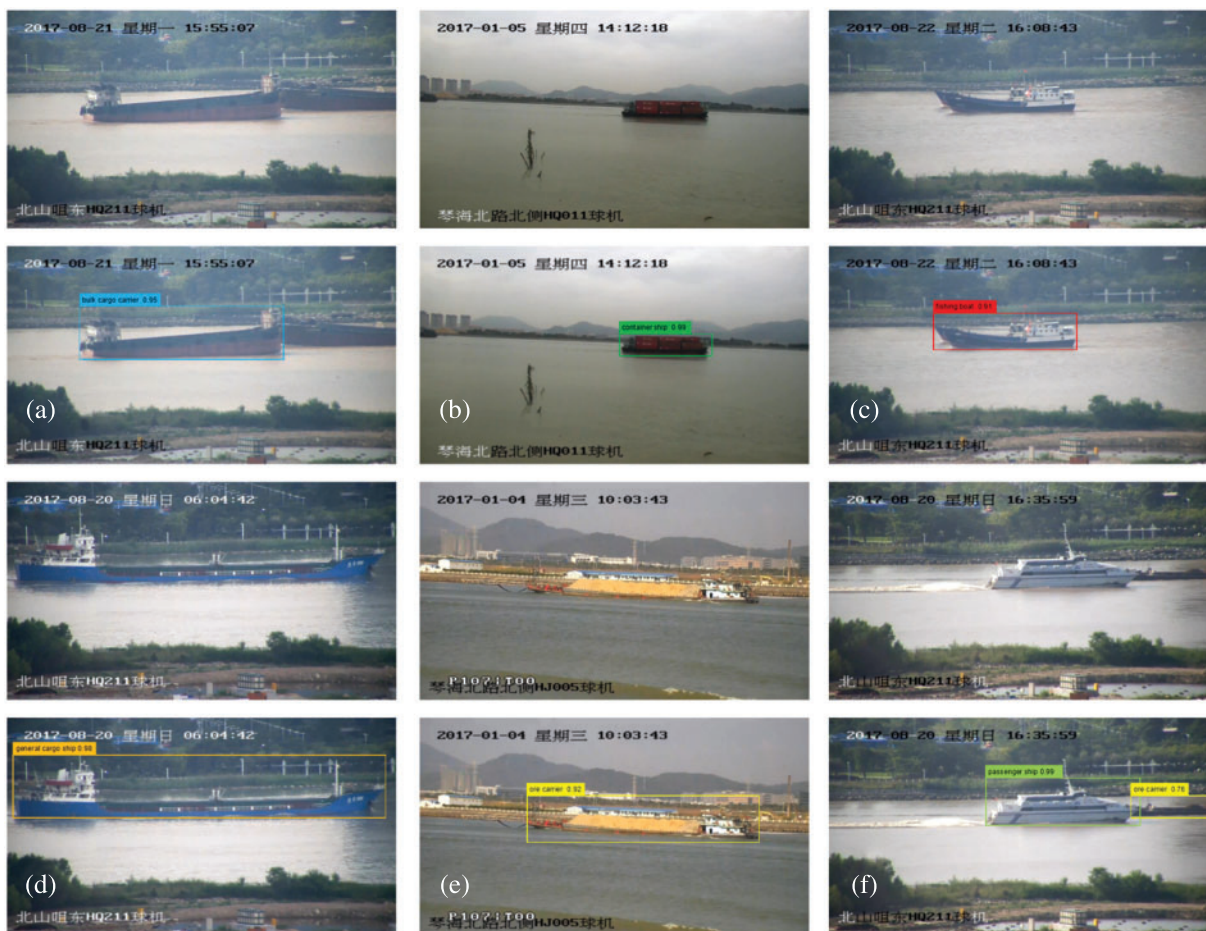


Figure 6: Ships detection results by improved YOLOv7

## 6 Conclusion

In summary, the proposed improved YOLOv7 model with specifically designed anchor boxes, a novel multi-scale feature fusion module, and enhanced data preprocessing with data augmentation and balanced sampling, has shown significant improvements in ship detection and recognition tasks. The experimental results demonstrate that the improved YOLOv7 model outperforms existing state-of-the-art approaches in terms of both detection accuracy and robustness, with a final mAP of 90.15%. The model has demonstrated high accuracy and mAP values for small fishing boats, making it a suitable option for inland waterway ship identification detection. Additionally, the algorithm can accurately



detect different types of ships of varying sizes, even in complex environments, and can identify multiple targets in a single image with high confidence. These findings highlight the potential of the improved YOLOv7 model for practical applications in the maritime domain.

**Funding Statement:** This work is supported by the Key R & D Project of Hainan Province (Grant No. ZDYF2022GXJS348, ZDYF2022SHFZ039).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] Y. Ren, J. Yang, Q. Zhang and Z. Guo, "Ship recognition based on Hu invariant moments and convolutional neural network for video surveillance," *Multimedia Tools and Applications*, vol. 80, pp. 1343–1373, 2021.
- [2] Z. Liu, L. Yuan, L. Weng and Y. Yang, "A high resolution optical satellite image dataset for ship recognition and some new baselines," in *Proc. ICPRAM*, Porto, Portugal, pp. 324–331, 2017.
- [3] D. Zhang, J. Zhan, L. Tan, Y. Gao and R. Župan, "Comparison of two deep learning methods for ship target recognition with optical remotely sensed data," *Neural Computing and Applications*, vol. 33, pp. 4639–4649, 2021.
- [4] N. MacAulay, "Molecular mechanisms of brain water transport," *Nature Reviews Neuroscience*, vol. 22, no. 6, pp. 326–344, 2021.
- [5] W. Xu, Y. Xing, J. Liu, H. Wu, Y. Cui *et al.*, "Efficient water transport and solar steam generation via radially, hierarchically structured aerogels," *ACS Nano*, vol. 13, no. 7, pp. 7930–7938, 2019.
- [6] M. Fournier, R. Casey Hilliard, S. Rezaee and R. Pelot, "Past, present, and future of the satellite-based automatic identification system: Areas of applications (2004–2016)," *WMU Journal of Maritime Affairs*, vol. 17, pp. 311–345, 2018.
- [7] D. Yang, L. Wu, S. Wang, H. Jia and K. X. Li, "How big data enriches maritime research—a critical review of automatic identification system (AIS) data applications," *Transport Reviews*, vol. 39, no. 6, pp. 755–773, 2019.
- [8] B. Hu, X. Liu, Q. Jing, H. Lyu and Y. Yin, "Estimation of berthing state of maritime autonomous surface ships based on 3D LiDAR," *Ocean Engineering*, vol. 251, pp. 111131, 2022.
- [9] J. Patoliya, H. Mewada, M. Hassaballah, M. A. Khan and S. Kadry, "A robust autonomous navigation and mapping system based on GPS and LiDAR data for unconstrained environment," *Earth Science Informatics*, vol. 15, pp. 2703–2715, 2022.
- [10] Y. Hao, P. Zheng and Z. Han, "Automatic generation of water route based on AIS big data and ECDIS," *Arabian Journal of Geosciences*, vol. 14, pp. 1–8, 2021.
- [11] Q. Wang, M. Zhong, G. Shi, J. Zhao and C. Bai, "Route planning and tracking for ships based on the ECDIS platform," *IEEE Access*, vol. 9, pp. 71754–71762, 2021.
- [12] M. Leclerc, R. Tharmarasa, M. C. Florea, A. C. Boury-Brisset, T. Kirubarajan *et al.*, "Ship classification using deep learning techniques for maritime target tracking," in *Proc. FUSION*, Salamanca, Spain, pp. 737–744, 2018.
- [13] Z. Chen, D. Chen, Y. Zhang, X. Cheng, M. Zhang *et al.*, "Deep learning for autonomous ship-oriented small ship detection," *Safety Science*, vol. 130, pp. 104812, 2020.
- [14] Y. Ren, X. Li and H. Xu, "A deep learning model to extract ship size from sentinel-1 SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021.
- [15] T. Zhang, X. Zhang, X. Ke, C. Liu, X. Xu *et al.*, "HOG-ShipCLSNet: A novel deep learning network with hog feature fusion for SAR ship classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–22, 2021.
- [16] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. CVPR*, Columbus, Ohio, USA, pp. 580–587, 2014.

- [17] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus *et al.*, “OverFeat: Integrated recognition, localization and detection using convolutional networks,” in *Proc. ICLR*, Scottsdale, Arizona, USA, 2013.
- [18] J. R. Uijlings, K. E. A. Van De Sande, T. Gevers and A. W. M. Smeulders, “Selective search for object recognition,” *International Journal of Computer Vision*, vol. 104, pp. 154–171, 2013.
- [19] K. He, X. Zhang, S. Ren and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.
- [20] R. Girshick, “Fast r-cnn,” in *Proc. ICCV*, Santiago, Chile, pp. 1440–1448, 2015.
- [21] S. Ren, K. He, R. Girshick and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *Advances in Neural Information Processing Systems*, vol. 28, pp. 1137–1149, 2015.
- [22] Q. Fan, W. Zhuo, C. Tang and Y. Tai, “Few-shot object detection with attention-RPN and multi-relation detector,” in *Proc. CVPR*, Long Beach, CA, USA, pp. 4013–4022, 2020.
- [23] K. He, G. Gkioxari, P. Dollár and R. Girshick, “Mask r-cnn,” in *Proc. ICCV*, Venice, Italy, pp. 2961–2969, 2017.
- [24] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 779–788, 2016.
- [25] J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp. 7263–7271, 2017.
- [26] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” arXiv preprint arXiv:1804.02767, 2018.
- [27] A. Bochkovskiy, C. Wang and H. Y. Mark Liao, “Yolov4: Optimal speed and accuracy of object detection,” arXiv preprint arXiv:2004.10934, 2020.
- [28] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng *et al.*, “YOLOv6: A single-stage object detection framework for industrial applications,” arXiv preprint arXiv:2209.02976, 2022.
- [29] C. Wang, A. Bochkovskiy and H. Y. Mark Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” arXiv preprint arXiv:2207.02696, 2022.
- [30] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed *et al.*, “SSD: Single shot multibox detector,” in *Computer Vision–ECCV 2016: 14th European Conf.*, Amsterdam, The Netherlands, Springer, October 11–14, 2016, Proceedings, Part I 14, pp. 21–37, 2016.
- [31] P. Zhou, B. Ni, C. Geng, J. Hu and Y. Xu, “Scale-transferrable object detection,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 528–537, 2018.
- [32] N. Dvornik, K. Shmelkov, J. Mairal and C. Schmid, “Blitznet: A real-time deep network for scene understanding,” in *Proc. of the IEEE Int. Conf. on Computer Vision*, Venice, Italy, pp. 4154–4162, 2017.
- [33] Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen *et al.*, “M2det: A single-shot object detector based on multi-level feature pyramid network,” in *Proc. of the AAAI Conf. on Artificial Intelligence*, Honolulu, HI, USA, vol. 33, pp. 9259–9266, 2019.
- [34] T. Kong, F. Sun, C. Tan, H. Liu and W. Huang, “Deep feature pyramid reconfiguration for object detection,” in *Proc. of the European Conf. on Computer Vision (ECCV)*, Munich, Germany, pp. 169–185, 2018.
- [35] Z. Shao, W. Wu, Z. Wang, W. Du and C. Li, “Seaships: A large-scale precisely annotated dataset for ship detection,” *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2593–2604, 2018.