China Instrument and Control Society
Instrumentation
INSTRUMENTATION, Vol. 11, No. 4, December 2024 https://doi.org/10.15878/j.instr.202400216

Article

Ship Detection Based on Improved SDD Algorithm

Hongcheng Chu[1], Tianhu Wang[1*], Qiannian Miao[1], Zeran Chen[2], Rong Wang[1], Wenjie Li[1], Tao Huang[3]

[1] School of Electrical Information Engineering, Jiangsu University of Technology, Changzhou, Jiangsu 213001, China

[2] Changzhou High-tech Technology Innovation and Entrepreneurship Service Center, Changzhou, Jiangsu 213001, China

[3] Technical Development Department, CRRC Nanjing Puzhen Rolling Stock Co., Ltd., Nanjing, Jiangsu 211800, China

[*] Corresponding author email: tianhu2003@126.com

Abstract: To address the issue of inadequate detection performance for small and medium-sized densely packed vessels in ship target detection, this paper proposes an improved Single Shot Multibox Detector (SSD) model to achieve more accurate detection. The algorithm redesigns the anchor boxes to fit the ship target detection dataset better and integrates the Squeeze-and-Excitation (SE) module into the Visual Geometry Group(VGG)network to enhance the channel features of the input feature maps. Additionally, the network's ability to perceive and represent important features is further enhanced by introducing the Convolutional Block Attention Module (CBAM), which is responsible for channel and spatial attention mechanisms. Finally, the feature pyramid module is employed to fuse six layers of features from the original network, thereby improving the SSD network's capability to detect small and occluded densely packed vessel targets. The experimental results show that the model's target recognition ability for fishing vessels improved from 58.07% to 65.87%; for patrol boats, the ability increased from 94.6% to 96.03%; and for inflatable boats, it rose from 72.08% to 74.93%. The overall mean Average Precision (mAP) also increased from the original model's 80.04% to 81.22%. Additionally, by clustering prior boxes, more suitable prior boxes for vessel detection were obtained, enhancing the model's perception capabilities for both large and small vessels.

Keywords: target detection; ship; feature fusion; SDD

Citation: Hongcheng Chu, Tianhu Wang, Qiannian Miao, Zeran Chen, Rong Wang, Wenjie Li, Tao Huang. "Ship detection based on improved SDD algorithm." Instrumentation 11, no.4 (December 2024). https://doi.org/10.15878/j.instr.202400216

1 Introduction

As a major maritime nation, China has consistently led the world in terms of shipping scale, vessel ownership, and port size[1]. In recent years, China has actively promoted and implemented the "Belt and Road" initiative, fostering trade cooperation with countries and regions, which has become a crucial strategic task for China's trade cooperation. However, this strategic task also presents significant challenges for the shipping industry, particularly in vessel target detection.

Currently, vessel detection research primarily relies on high-resolution Synthetic Aperture Radar (SAR) ocean vessel image datasets[2]. However, there is a lack of high-precision data equipment in routine vessel detection, and the applicable vessel datasets for conventional deep learning are limited. This leads to civilian vessel detection relying mainly on manual observation and traditional machine learning methods. Manual observation requires substantial manpower and time, and in complex meteorological and hydrological environments, manual observation is difficult to manage and prone to misjudgments and omissions.

Therefore, employing automated technology for vessel detection has become a trend. This approach can enhance detection efficiency and accuracy, reduce costs, and achieve all-weather monitoring and management. This is crucial for meeting the growing demands of China's shipping industry and the trade cooperation under

the "Belt and Road" initiative. Hence, establishing a robust vessel target detection system has become an urgent task to promote the development of the shipping industry and address future challenges.

In recent years, object detection algorithms in deep learning have mainly divided into two types: two-stage methods and one-stage methods. The two-stage methods include the R-CNN series, while the one-stage methods include the SSD (Single Shot MultiBox Detector) algorithm, which belongs to the one-stage techniques and can perform multi-box prediction. In 2014, R. Girshick et al. [3] first proposed Regions with CNN features (RCNN), applying deep learning to object detection, which laid the foundation for the two-stage detector method in object detection. Building on this, to address the issue of the RCNN model's abnormally slow performance during object detection tasks, K. He et al. proposed Spatial Pyramid Pooling Networks (SPPNet) [4]. The main contribution of SPPNet is the introduction of the Spatial Pyramid Pooling (SPP) layer, which allows the convolutional neural network to transform image features into a fixed length. Mapping the overall image features to a high-dimensional feature space enables the detection of any region of the image in that high-dimensional space, thereby avoiding the redundant computations that occur during selective search, significantly increasing the operational speed of the network for object detection. In 2015, R. Girshick introduced ROI pooling in Fast RCNN[5], which divides the image into regions and selects the maximum confidence prior boxes within the target regions through non-maximum suppression, thereby reducing the number of feature boxes required for detection. At the same time, ROI pooling maintains the extraction of feature maps of the same size from the network. S. Ren et al. proposed Faster RCNN [6], integrating feature extraction, region proposals, bounding box regression, and classification tasks into a single network, significantly improving the speed of object detection. In 2017, T. -Y. Lin et al. introduced Feature Pyramid Networks [7], addressing the poor performance of convolutional neural networks in image localization by proposing a top-down network architecture for feature fusion, which computes high-level semantic information at all scales, thus enhancing the model's detection performance across various scales.

Unlike RCNN, R. Joseph proposed the YOLO (You Only Look Once) [8] network, pioneering the concept of one-stage object detection. One-stage object detection adopts a different paradigm from two-stage detection: the entire image is divided into regions in advance, each responsible for detecting the objects that fall within it. The YOLO network model improves overall detection speed but reduces object localization accuracy. Following this, W. Liu et al. proposed the Single Shot MultiBox Detector (SSD) network [9]. Its main contribution is the introduction of small convolutional filters to predict object classes and offsets for bounding box positions. The primary distinction is that SSD detects objects of different scales across various feature layers of the network.

The main development direction of domestic research on vessel detection focuses on further improvements to existing deep learning-based object detection networks. Dong Quanshuai[10]proposed the RIRnet neural network for satellite vessel image recognition, which uses residual nesting as the basic structure to facilitate gradient propagation. The network employs the U-Net[11] architecture for semantic segmentation of satellite remote sensing images of vessels. Using a deconvolution structure for upsampling, the multi-scale convolution enhances the segmentation capabilities for target edge details and small objects. Zhang Yue[12] improved the YOLO model by replacing Mean Squared Error (MSE) with Generalized Intersection over Union (GIOU) to achieve more precise target localization. To address the issue of small object detection, the DSSD network was used as a submodule, and a non-maximum suppression algorithm was applied to detect occluded vessels, effectively solving the problem of object occlusion in images.

In this paper, we present several significant contributions to vessel target detection. We developed and annotated a dataset containing over 3000 images of vessels categorized into five primary types: cargo ships, cruise ships, fishing boats, inflatable boats, and fishery patrol boats. This dataset, explicitly tailored for vessel target detection, is a high-quality resource for advancing maritime surveillance research.

We employed the Single Shot Multibox Detector (SSD) algorithm to evaluate its effectiveness on the dataset. This classic object detection model provides a reliable baseline for assessing the impact of the subsequent enhancements introduced in the study.

Our work involves critical improvements to the SSD algorithm by incorporating feature fusion strategies and prior box clustering techniques. These modifications significantly boost detection accuracy, particularly for small, challenging targets commonly encountered in maritime environments.

Through extensive experimentation, we validated the effectiveness of these improvements, demonstrating that the enhanced SSD model delivers superior performance in classification and localization compared to the original version. These contributions advance the state-of-the-art in vessel target detection, offering valuable insights for future research and practical applications.

2 Network Training

The computer configuration is as follows: Windows 11 operating system, hardware configuration includes an R9 7945HX processor(AMD, Santa Clara, CA, USA), an NVIDIA RTX 3070 Ti graphics card (NVIDIA, Santa Clara, CA, USA), and 32GB of memory(Samsung,

Suwon, Gyeonggi-do, South Korea). The experimental code is written in Python and completed using the PyTorch framework. The libraries used in the research include CUDA 11.0 and CUDNN 11.0 for GPU computation, opencv-python 4.7.0.72 for image processing, tensorboard 2.12.0 for visualizing model results, and NumPy for matrix operations. Python 3.6 and PyTorch 1.9.0 were used for the research.

2.1 Vessel Target Detection Dataset

This study uses familiar vessel images as the data source, based on typical offshore and nearshore vessels found online. A total of 2, 627 vessel images were collected, including five significant categories of vessels: patrol boats (461 images), fishing boats (523 images), cruise ships (428 images), inflatable boats (477 images), and cargo ships (783 images). This paper uses the LabelImg tool to annotate the collected data, employing the Pascal VOC annotation format. The required images are opened in LabelImg, and the Create RectBox function is used to draw bounding boxes around the targets in the images. After saving, these annotations are stored as XML files. The operation interface is shown in Figure 1.

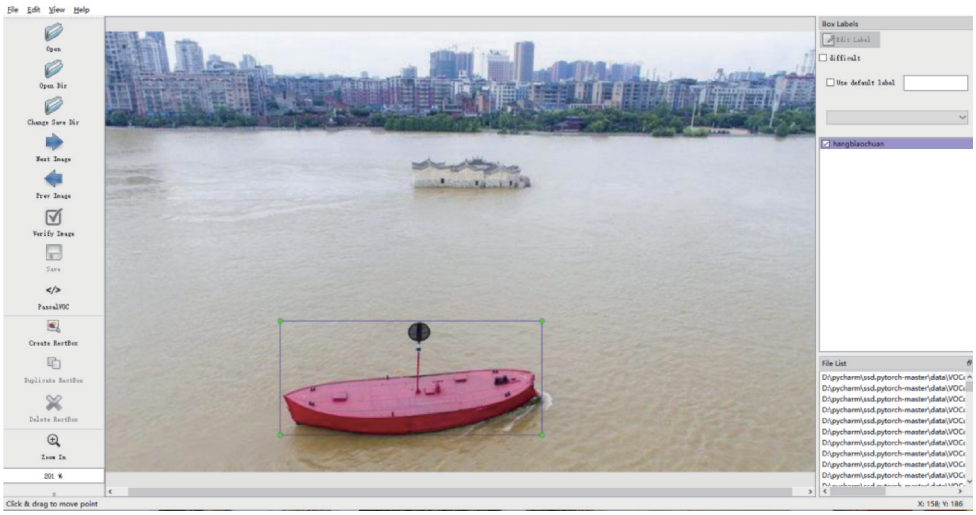

Fig.1 The operation interface of LabelImg

The JPEGImages folder contains the image information, while the Annotation folder holds the corresponding XML files. Due to specific differences in image formats for ship images found in online resources, which include ". JPEG," ". png," and ". jpg" formats, the varying pixel arrangements in these three formats can impact the learning of the network model. Therefore, this paper uses the Image function from the Pillow library to convert all images to a unified ". jpg" file format. The dataset is then divided using the train_test_split. py file, with a split ratio of 80% for the training set and 20% for the testing set. After obtaining the image indices for the training and testing sets, the index text files are placed in the ImageSets/Main folder. At this point, the preparation of the ship object detection dataset is complete.

2.2 Environment Configuration and Network Training

The experiment uses Single Shot MultiBox Detector 300(SSD300). During training, SSD primarily focuses on configuring parameters. The key parameters include the initial learning rate, momentum, batch size, weight decay, freezing, and optimizer type. The parameter settings for this experiment are detailed in Table 1.

In the "train. py" file, the overall network model is trained. According to the training results in Table 2, when using a threshold of 0.5, the model shows high accuracy and recall for detecting cruise ships and patrol boats, with AP (Average Precision) scores of 95.58% and 93.02%, respectively. However, the model's detection capability for fishing boats is relatively poor, with an AP of only 58.07%. The training results reveal that learning to detect fishing boats is more challenging. There are 140 actual fishing boat bounding boxes, but the model detected 1, 223 targets, which far exceeds the number of actual samples. This indicates many false detections for fishing boats, resulting in lower accuracy and recall.

Table 1 Network Parameter Settings

Parameter	Parameter Value
Learning Rate	0.002
Momentum	0.937
Weight Decay Coefficient Freezing	0.0005
Training Epochs	50
Non-Freezing Training Epochs	200
Freezing Batch Size Non-Freezing	16
Batch Size	8
Optimizer Type	SGD

The distribution of fishing boats in the image is relatively dense, with serious occlusion issues. Moreover,

Table 2 SSD Model Training Results

Type	AP@0.5	P	R	DETECTED	GROUND_TRUTH
Cruise Ship	95.58%	0.875	0.913	88	46
Patrol Boat	93.02%	0.913	0.875	188	72
Cargo Ship	84.90%	0.95	0.768	251	99
FishingBoat	58.07%	0.745	0.57	1223	140
InflatableBoat	72.08%	0.89	0.602	603	136

there are various fishing boats, making it difficult for the model to fit the features of the boats during training. As a result, the model struggles to extract image features during object detection, leading to poor detection performance. The model's prediction results are shown in Figure 2. The results indicate that when the detection image contains only a single boat, meaning the boat's features are apparent, the SSD algorithm performs quite accurately in detecting boats, with a precision close to 0.9-1.0. When the images of boats are dense, the SSD algorithm can still accurately detect boats that are not occluded. However, the performance of the SSD algorithm significantly declines for occluded boats. Additionally, the type of buoy vessel, the distance of the boats, and the kind of buoy will affect the detection performance of the algorithm. For this experimental dataset, the SSD algorithm favors using triangular buoys for buoy vessels. Regarding inflatable boats, the size ratio of the inflatable boat to the rower will influence the detection performance of the SSD algorithm. When the rower's proportion is too large, the model's ability to recognize inflatable boat types will experience a noticeable decline.

Fig.2 Ship Test Results of SSD

The experiment selected YOLO V5 as the optimal version for comparison. YOLO V5 excels in balancing model accuracy and speed. Compared to previous YOLO versions, YOLO V5 introduces several improvements, such as AutoAnchor for adaptive anchor box calculation, CIoU-based bounding box regression, and Mosaic data augmentation. These enhancements allow YOLO V5 to maintain high detection accuracy while achieving fast inference speeds. YOLO V5's model structure is also more lightweight, making it suitable for deployment on resource-constrained devices. As a result, it performs well in various real-world applications.

The training parameters for the YOLO V5 algorithm are shown in Table 3.

By running the train.py file, the overall training of the network model is executed, and TensorBoard is used to visualize the training process and results. The training results are shown in Table 4. The results include the AP values for boat object detection categories using a threshold of 0.5, along with the model's accuracy and recall rates for boat predictions. The model performs better in learning large vessels such as ships and patrol boats, achieving AP values of 92.1% and 94.6%,

Table 3 Network parameter setting

parameter	Parameter Value
Learning Rate	0.01
Momentum	0.937
Weight Decay Coefficient	0.0005
Epochs	300
Batch Size	16

respectively. However, the model's detection performance for fishing boats and inflatable boats is poorer, with AP values of 76.6% and 78.4%. This may be due to the high density of boats, leading to significant deviations during the model's learning process. Additionally, the training set contains large vessels such as cargo ships, which results in the model having poorer robustness to high thresholds during training, causing a decline in results.

The test.py file is run to execute the network model testing, and the test results are shown in Table 4. The results indicate that the model has poor perception ability for densely packed small boats but strong perception

Table 4 YOLO5 Model Training Results

Type	AP@0.5	P	R	DETECTED	GROUND_TRUTH
Cruise Ship	92.1%	0.905	0.907	86	46
Patrol Boat	94.6%	0.917	0.897	111	72
Cargo Ship	86.1%	0.847	0.813	246	99
FishingBoat	76.6%	0.758	0.725	433	140
InflatableBoat	78.4%	0.842	0.72	368	136

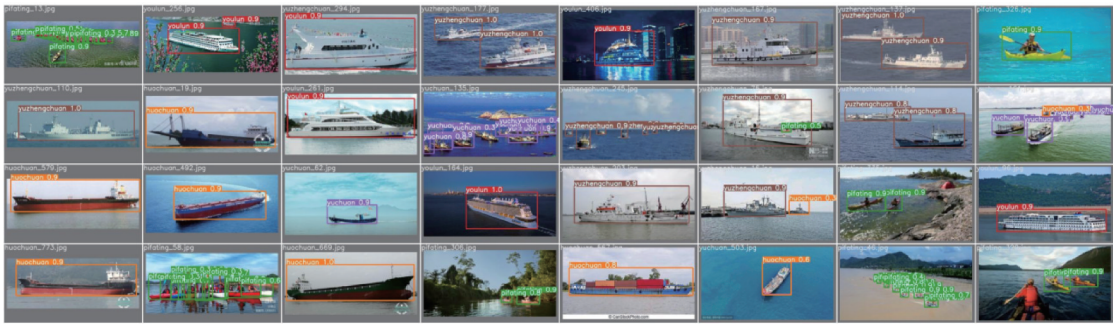

Fig.3 Ship Test Results of YOLO

ability for large vessels such as cargo ships. It is worth mentioning that for small cruise ships and patrol boats, which have similar appearances, their similarity in feature dimensions makes them challenging to distinguish. Therefore, the model's perception ability for both types is weak, leading to misclassification. For fishing boats and inflatable boats, there are instances of low confidence. Moreover, for boats with incomplete features, such as the inflatable boat images in the results, the network exhibits a poor ability to perceive local features due to the incomplete characteristics of the boat.

3 Improvements to the SSD Algorithm

Improvements to the SSD algorithm are proposed to address the poor performance in detecting small and medium-sized dense ships. First, regarding anchor boxes, the default settings are suitable for standard datasets but perform poorly for small vessels like fishing boats and inflatable boats. Therefore, K-means clustering, similar to the method used in YOLO, is applied to the dataset to obtain representative anchor box sizes of [21, 45, 88, 153, 207, 261, 315].

During the feature extraction, contextual feature information is considered for analyzing ship images. It is crucial to focus on features such as paddlers and oars for small vessels such as fishing boats and inflatable boats, which rely on manual paddling. Additionally, when ships overlap, inferring the global features of obscured ships from the explicit features of nearby visible ships can lead to more accurate detection of small vessels. To enhance the model's perception of global information, an attention mechanism is introduced.

Specifically, the SE (Squeeze-and-Excitation) module is incorporated into the VGG network to improve the model's ability to learn channel-wise weights. The SE module consists of three parts: Squeeze, Excitation, and Scale[13]. The Squeeze module compresses the spatial information of the input feature maps[14], the Excitation module combines the learned channel attention information with the input feature maps to assign weights to the channel features[15], and the Scale module performs element-wise multiplication of the calculated channel weights with the original feature maps. This enhancement improves the model's focus on different channels within ship images, thereby increasing the accuracy of ship detection.Figure 4 illustrates the SE module's workflow.

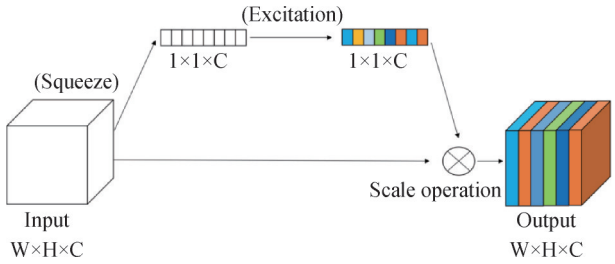

Fig.4 SE Module

For improvements in attention mechanisms, compared to the SE (Squeeze-and-Excitation) attention mechanism, which focuses on channel-wise weights, the CBAM (Convolutional Block Attention Module) divides the attention mechanism into two distinct modules: the channel attention module and the spatial attention module[16]. Each module captures attention in its respective domain—channel and spatial. The CBAM module captures channel and spatial attention through parallel average pooling and max pooling operations. The specific workflow of the CBAM module is illustrated in Figure 5.

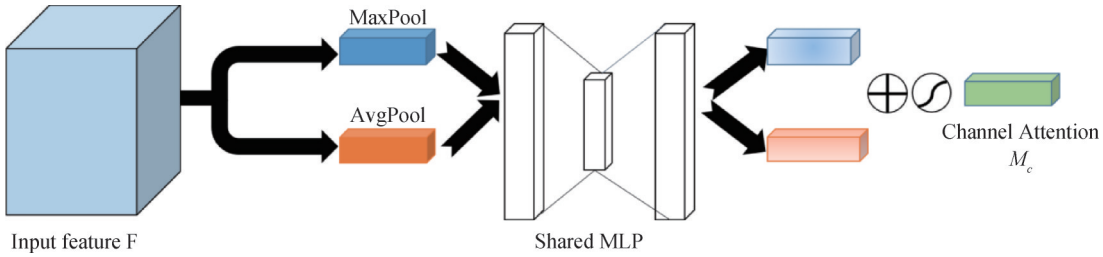

Fig.5 CBAM Module

Based on the representation of ship image features, recognizing hull characteristics becomes challenging when ships are densely packed. However, the feature attention mechanism improves the algorithm's ability to focus on small boats within dense clusters by allocating weights to channel features. Additionally, this attention mechanism helps the model learn more comprehensive contextual information, enhancing its inference capability. Therefore, this paper employs both SE (Squeeze-and-Excitation) and CBAM (Convolutional Block Attention Module) self-attention modules as the attention mechanisms for the algorithm model.

In the SSD algorithm model, six feature maps are connected. As the neural network deepens, semantic information becomes more affluent, but the feature maps become smaller, decreasing resolution. This makes deep feature maps prone to missing small boats. To address this issue, this paper introduces a Feature Pyramid Module to fuse features from six different levels, leveraging deep information to guide shallow information and thus enhancing the model's ability to detect small boats[17-18].

The operation of the Feature Pyramid Module is as follows: information from pooling layers and convolutional layers is combined through the attention mechanism to form six feature maps of different sizes, which are then used as final features input into the SSD network. Specifically, the first dimension feature map is fused with the 75×75 feature map from the VGG network and the second dimension feature map, and the first dimension feature map is also fused with the second dimension feature map. The first, third, and fourth dimension feature maps are fused. An attention mechanism module is introduced before these six feature maps are input into the SSD algorithm model. This attention mechanism module allows the model to utilize the fused features better and make more informed inferences based on global information, thereby improving the model's ability to handle complex samples and enhancing ship detection performance[19-20].

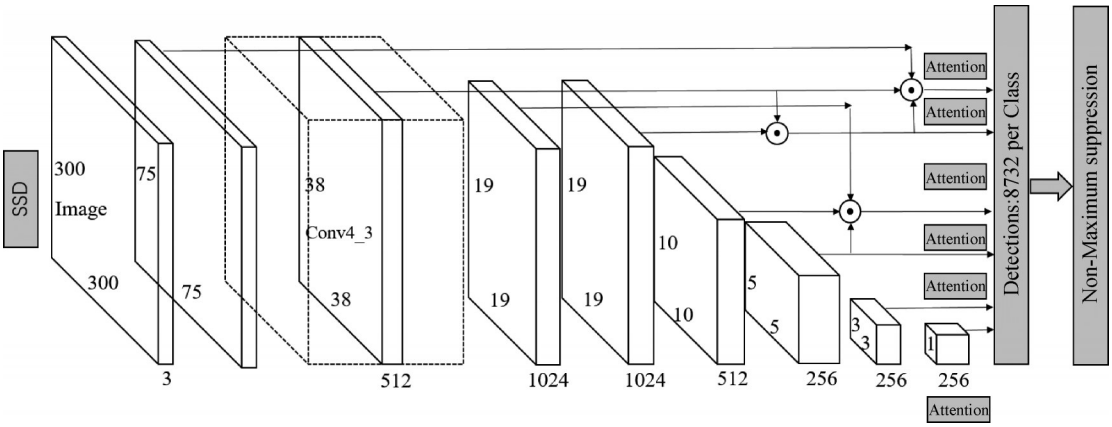

Fig.6 SSD Feature Pyramid Network

After implementing the SE attention mechanism and feature fusion, the network model achieved the final results for ship detection as shown in Table 5. The study indicates that introducing the feature fusion module significantly improved the model's performance in detecting dense ships by providing more accurate anchor box predictions. Without increasing the number of actual target boxes, the model successfully reduced the number of predicted boxes to a maximum of 314. This means the model reduced 531 detected targets on the test set while maintaining detection accuracy, thus improving the model's recall rate. Additionally, although the number of detection boxes was reduced in detecting fishing boats, the accuracy improved from 0.745 to 0.785. This indicates that the model has a more precise capability in detecting small boats, such as fishing boats.

4 Experimental Validation and Analysis

This section focuses on the experimental validation

Table 5 Results of the Proposed Model

Type	AP@0.5	P	R	Detected	Ground_Truth
Cruise Ship	**94.48%**	0.976	0.87	70(−18)	46
Patrol Boat	**96.03%**	0.913	0.875	157(−31)	72
Cargo Ship	82.42%	0.894	0.768	239(−12)	99
Fishing Boat	65.87%	0.785	0.521	**909(−314)**	140
Inflatable Boat	74.93%	0.854	0.603	458(−145)	136

and analysis of the YOLO algorithm model, the SSD algorithm model, and the improved SSD algorithm model in the context of ship target detection. It includes analysis of algorithm results and ablation experiments on the improved SSD algorithm.

4.1 Summary of Algorithm Results

Based on the experimental dataset, the comparison of ship target detection performance among the YOLO algorithm, SSD algorithm, and the improved SSD algorithm is shown in Table 6:

Table 6 Comparison of AP Values for YOLO and SSD

Type	Cruise Ship	Patrol Boat	Fishing Boat	Inflatable Boat	Cargo Ship	mAp@0.5
YOLO	92.1%	93.02%	**84.9%**	72.08%	84.9%	**85%**
SSD	95.58%	94.6%	58.07%	72.08%	84.9%	80.04%
SSD+(OURS)	94.48%	**96.03%**	65.87%	74.93%	82.42%	**81.22%**

The comparison of experimental results shows that the YOLO V5 model significantly outperforms the SSD network's 58.07% and the improved model in this paper's 65.97% in detecting small ships, achieving an accuracy of 84.9%. Therefore, the YOLO V5 model is capable of effectively handling the detection of small, dense ship targets with occlusions, such as densely packed fishing boats. By learning from the multi-scale feature fusion strategy of the YOLO model, the network in this paper enhances the SSD network's ability to detect small, occluded, and dense ship targets through feature fusion across six levels of the SSD network, improving the model's target recognition capability for fishing boats from 58.07% to 65.87%. Additionally, prior box clustering is used to obtain more suitable prior boxes for ship detection, thereby enhancing the model's perception of both large and small ships.

4.2 Ablation Experiments

In this section, ablation experiments are conducted to analyze the effectiveness and necessity of different key components of the proposed improved SSD model. The analysis focuses on the three main components of the model improvements: prior box selection, feature fusion, and the integration of attention mechanisms.

To evaluate the importance of each component in these improvements, four different variant models were designed:

SSD-FPN: A target detection model with self-attention mechanisms and feature pyramid fusion.

SSD-SE: A feature fusion target detection model with the addition of the SE attention mechanism module.

SSD-CBAM: A feature fusion target detection model incorporating the CBAM attention mechanism.

SSD-A: A target detection model with modified prior box information.

These models were compared with the completely improved model to analyze their performance and verify whether the various components complement each other and the practicality of feature fusion in the SSD network's target detection process.

The ablation experiments were conducted using the Ship dataset, which was collected independently. Model evaluation metrics include the average precision (AP) values for each category in the dataset and the overall mean average precision (mAP). The performance results of different models on the Ship dataset are presented in Table 5. These ablation experiment results help us understand the contribution of each component to the model's performance and provide deeper insights into the improved method.

Based on the table, the leftmost column shows the four variant models selected for the ablation experiments, representing different versions of the proposed SSD improvement model. The rightmost columns display the experimental results of these variant models on the ship dataset collected for this study, including the average precision (AP) values for each ship category and the overall mean average precision (mAP) at a threshold of 0.5.

The experimental results reveal the following key findings:

Improvement in Fishing Boat Detection through Feature Fusion: When self-attention mechanisms are not included and only feature fusion is used, the model significantly improves the detecting fishing boats while

Table 7 Ablation Experiment Results on the Ship Dataset

MODEL TYPE	Cruise Ship	Patrol Boat	Fishing Boat	Inflatable Boat	Cargo Ship	mAp@0.5
SSD	95.58%	94.6%	58.07%	72.08%	84.9%	80.04%
ssd-fpn	94.39%	**96.34%**	65.16%	71.41%	82.40%	80.73%
SSd-cbam	94.70%	93.97%	60.22%	71.02%	83.72%	80.24%
ssd-se	**96.80%**	97.86%	61.02%	69.80%	**86.42%**	81.36%
SSD-a	94.02%	95.62%	**66.22%**	70.14%	85.66%	**81.86%**
SSD+(OURS)	94.48%	96.03%	65.87%	**74.93%**	82.42%	81.22%

maintaining high accuracy for patrol boats. This indicates that feature fusion plays a crucial role in enhancing model performance.

Negative Impact of CBAM Attention Mechanism: It is noteworthy that the introduction of the CBAM attention mechanism leads to a decrease in overall model accuracy, particularly with a 4.9% drop in AP for fishing boats. This result is contrary to the intended benefit of the CBAM attention mechanism. Analysis reveals that CBAM, through the combination of max pooling and average pooling, emphasizes features of large ships while diminishing the representation of small boats, reducing overall performance in ship detection.

Positive Impact of SE Attention Mechanism: After incorporating the SE attention mechanism, the overall model accuracy improves, especially for fishing boats, which contrasts with the effect of the CBAM module. This indicates that the SE module, through channel-level attention allocation, enhances the model's perceptual capability.

Significant Improvement with Prior Box Adjustment: Adjusting the prior boxes results in a notable increase in the detection accuracy for small boats, improving from 58.07% to 66.22%. This further underscores the direct impact of prior box configuration on model performance.

Different Feature Extraction Strategies for Various Ship Types: The ablation experiment results suggest that different feature extraction strategies have varying preferences for enhancing detection capability for different ship types. This may be due to significant differences in feature space among different ship types. The performance improvement of various strategies for specific ship types may lead to a preference for the feature locations of these types, thus increasing the corresponding AP values.

In summary, the configuration of prior boxes is crucial for ship target detection models to enhance the model's perceptual ability. Additionally, different feature extraction strategies have varying effects on the detection capability for different ship types. Strengthening the model's functional space and fitting capability to improve the generalization ability of the feature space may help address these issues and achieve more comprehensive detection performance.

5 Conclusion and Outlook

This paper addresses the challenge of obtaining datasets for traditional ship target detection tasks by proposing using conventional ship images combined with deep learning object detection algorithms to complete ship target detection tasks. The work in this paper can be divided into two parts. First, we collected and annotated our ship target detection dataset, utilizing YOLO and SSD algorithms to perform ship target detection tasks, experimental verification and analysis of the algorithms, and ablation experiments on optimization methods to demonstrate their effectiveness.

Object detection, as one of the tasks closely aligned with practical needs in computer vision, is a direction that requires in-depth exploration. Although the ship network images created in this study have a certain level of validity, the trained object detection networks can meet some requirements in ship target detection. However, there are still shortcomings in the current work that need to be addressed.

The main areas for future work include the following. The collected ship target detection dataset still has certain drawbacks. For ship target detection tasks, the collected dataset still faces issues with sample quantity mismatch. The ship target dataset should be more targeted to enable the model to learn ship target features better, covering ship data from multiple angles and conditions. By broadening the typical applications of the dataset, the robustness of the model's operational scenarios can be enhanced, specifically including ship images in weather conditions such as heavy fog and heavy rain.

Analysis of the model training results indicates that target detection algorithms like YOLO and SSD still exhibit weak perception capabilities for dense ships and occluded ships during ship target detection tasks. Therefore, future designs could focus on developing network modules to enhance the ability to differentiate between occluded ships and dense ship areas. By introducing contrastive learning to increase feature differentiation among different types of ships, the model can be endowed with more substantial ship identification capabilities.

Author Contribution:

CHU Hong-cheng: Conceptualization, Methodology,

Software, Writing - original draft, Writing - review & editing. WANG Tian-hu*: Data curation, Formal analysis, Visualization, Writing - review & editing. MIAO Qian-nian: Investigation, Project administration, Supervision, Validation. CHEN Ze-ran: Resources, Funding acquisition, Writing - review & editing. WANG Rong: Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing. LI Wen-jie: Data curation, Formal analysis, Visualization, Writing - review & editing. HUANG Tao: Investigation, Project administration, Supervision, Validation.

Foundation Information:

This research was funded by Changzhou technology project (CZ20230025), Natural science foundation of jiangsu province (BK20150247).

Data Availability:

The authors declare that the main data supporting the findings of this study are available within the paper and its Supplementary Information files.

Conflicts of Interest:

The authors declare no competing interests.

Dates:

Received 06 August 2024; Accepted 05 December 2024; Published online 31 December 2024

References

[1] Li Jing, Xian Lin, Wang Haijiang. Research on Ship Detection Algorithm Based on YOLOv3 [J]. *Journal of Chengdu University of Information Technology* , **2023**, 38 (01): 37-43.

[2] Fu Gang. Research and Implementation of a River and Lake Surface Vessel Recognition Method Based on the SSD Object Detection Algorithm [D]. Advisors: Qian Xiaojun; Ruan Tongxiang. Nanjing Normal University, **2021**.

[3] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. **2014**: 580-587.

[4] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. *IEEE transactions on pattern analysis and machine intelligence* , **2015**, 37(9) :1904-1916.

[5] Girshick R. Fast r-cnn [C]//Proceedings of the IEEE international conference on computer vision.2015: 1440-1448.

[6] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. *Advances in neural information processing systems* , **2015**, 28.

[7] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. **2017**:2117-2125.

[8] Y . LeCun, Y . Bengio, and G. Hinton, "Deep learning, " nature, vol. 521, no. 7553, p. 436,**2015**.

[9] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October11-14, 2016, Proceedings, Part I 14. Springer International Publishing, **2016**: 21-37.

[10] Dong Quanshuai. Research on Satellite Vessel Image Recognition and Semantic Segmentation Methods [D]. Dalian University of Technology, School of Naval Architecture and Ocean Engineering, **2019**.

[11] Understanding the Effective Receptive Field in Deep Convolutional Neural Networks

[12] Zhang Yue. Research on Intelligent Detection of Marine Vessels Based on Improved YOLO Algorithm [D]. Shanghai Normal University, *Applied Statistics* , **2020**.

[13] Masci J, Meier U, Cireşan D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction[C]//Artificial Neural Networks and Machine Learning – ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June14-17, 2011, Proceedings, Part I21. Springer Berlin Heidelberg, **2011**: 52-59.

[14] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. *Communications of the ACM* , **2017**, 60(6): 84-90.

[15] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. *The journal of machine learning research* , **2014**, 15(1): 1929-1958.

[16] Wu H, Gu X. Max-pooling dropout for regularization of convolutional neural networks[C]//Neural Information Processing: 22nd International Conference, ICONIP2015, Istanbul, Turkey, November 9-12, 2015, Proceedings, Part I 22. Springer International Publishing, **2015**: 46-54.

[17] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. **2018**: 7132-7141.

[18] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). **2018**: 3-19.

[19] Dalal N,Triggs B. Histograms of oriented gradients for human detection [C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Ieee, **2005**, 1:886-893.

[20] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models[J]. *IEEE transactions on pattern analysis and machine intelligence* ,**2009**, 32(9) : 1627-1645.