Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 9440212, 11 pages
https://doi.org/10.1155/2021/9440212

Research Article

Ship Target Detection Algorithm Based on Improved YOLOv3 for Maritime Image

Dehai Chen, Shiru Sun, Zhijun Lei, Heng Shao, and Yuzhao Wang

School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou 341000, China

Correspondence should be addressed to Dehai Chen; 9120010041@jxust.edu.cn

Received 24 June 2021; Accepted 7 August 2021; Published 6 September 2021

Academic Editor: Xinqiang Chen

Accurate identification of ships is the key technology of intelligent transportation in water. At the same time, it also provides a judgment basis for water traffic safety control. This paper proposed a detection method of ships in water based on improved You Only Look Once version 3 (YOLOv3), which is called Feature Attention, Feature Enhancement YOLOv3 (AE-YOLOv3). The feature attention module was constructed by introducing the attention mechanism, which was embedded in Darknet-53 for feature recalibration, which improved the feature extraction ability of the model in the complex navigable background. For the problem of insufficient semantic information of low-level features in the feature fusion process, a feature enhancement module was constructed and applied to the feature fusion part to enhance the receptive field size of the corresponding feature layer and the correlation degree of feature extraction network. Experiments were carried out on the public SeaShips dataset. Experiments show that the detection accuracy is as high as 98.72%, which is better than that of other mainstream ship identification models, fully verifying the superiority of this method in the detection of waterborne traffic ships.

1. Introduction

With the development of shipping industry in the world, the number of ships is increasing, the ships are developing in the direction of large scale and high speed, and the navigation safety is becoming more and more important. This objectively promotes the development of ship navigation and automated driving technology [1]. In order to improve the efficiency, reliability, and safety of ship navigation, shipping industry is now gradually developing towards the direction of intelligent and full automation of ship autopilot, automatic avoidance, and automatic docking and leaving dock, making intelligent ships gradually become a new research direction [2–4]. With the increase of water traffic flow density [5], the navigation environment is becoming more and more complex. The target detection of surface ships is one of the key technologies of intelligent navigation. It is of great significance to accurately identify ships in water areas. This is an important part of water traffic safety monitoring and management, which provides a basis for water traffic management and control [6–8].

In the process of ship navigation, the ship target detection can help maritime participants take measures to avoid potential water traffic accidents [9]. Before Convolutional Neural Network is applied to ship target detection, traditional ship detection algorithms mainly include region selection, combined feature extraction, and background texture modeling [10]. In recent years, with the rapid development of computer vision, deep learning algorithms have achieved good results in the field of target detection and image processing [11]. Ren et al. proposed an improved Faster R-CNN (Faster Region-based Convolutional Neural Network) algorithm for ship target detection, and the accuracy of target detection is improved [12, 13]. When Faster R-CNN performs target detection, the feature maps extracted by features are used to generate region proposals using Region Proposal Networks (RPN), and 300 region proposals are generated for each image, while each region of interest (RoI) is made to generate a fixed size feature map by RoI pooling layer and is finally discriminated by fully connected network, which greatly increases the calculation time and the complexity of the network. Guo et al. proposed a DepthFire Single-Shot MultiBox Detector (DF-SSD) framework, which improves detection accuracy [14], but SSD algorithm predicts output through multilayer convolution, and the semantic information is lost too much, which is not friendly to the detection of small targets. Huang et al. did a comparative experiment on trajectory compression and visualization in different water areas, applied to maritime intelligent traffic management and collision avoidance [15]. This provides a theoretical basis for the design of the ship detection and tracking system.

Ship-YOLOv3 algorithm was proposed by Huang et al. It improved the detection accuracy by changing the YOLOv3 network structure, reducing some convolution operations, and increasing the jump connection mechanism [16]. Redmon et al. proposed YOLO algorithm for target detection [17], which leads the learning boom. However, when applying the YOLO algorithm to ship detection, the existing research often strives to improve the overall detection accuracy, while ignoring the very small targets that actually exist [18]. What is more, YOLO algorithm only predicts a set of category probability values. It brings many difficulties to researchers. With the increase in the density of water traffic flow, there are more and more types of ships that need to be identified on the waterway. Under the complex navigation background, the detection accuracy is relatively low for extremely small-sized ships, and the generalization ability of the model is poor. In this context, an algorithm that can detect all ships in time is particularly important [19, 20].

Through the analysis of the existing ship target detection algorithms, we find that YOLOv3 has greater advantages than YOLOv1 and YOLOv2. The details are as follows: YOLOv3 provides three types of suggestion boxes for large, medium, and small objects. 13×13 is used to detect large objects, 26×26 is used to detect medium objects, and 52×52 is used to detect small objects. In addition, 1×1 convolution replaces the fully connected layer, and the multiscale feature map is convolved with the 1×1 convolution to obtain the detected feature map of the corresponding size, which increases the nonlinear characteristics of the network and achieves feature information integration while the current feature map size remains unchanged. Because only the box with greater confidence is detected, the calculation time is faster, and the detection of small targets can also be friendly. YOLOv4 is improved on the basis of YOLOv3, and the network performance has been improved. However, in actual experiments, we considered that there is only one GPU card; YOLOv3 was selected for this work. Compared with R-CNN series and SSD algorithm, it has higher accuracy and faster calculation time. Based on the requirements of detection accuracy and speed, we chose YOLOv3 algorithm to design the ship detection frame, as it accurately locates the position of ships in maritime images.

Our primary academic contributions can be summarized as follows: (1) We proposed the AE-YOLOv3 water ship detection algorithm to realize the end-to-end detection of ship objects. At the same time, aiming at the problem of too little data in the ship dataset, the dataset was expanded through the technical means of data enhancement to avoid the occurrence of overfitting. (2) We built a feature attention module based on the attention mechanism, embedded the feature extraction network of AE-YOLOv3 to recalibrate the feature channel, and enhanced the spatial connection of the feature map [21]. (3) Then, we built a feature enhancement module based on the idea of multiscale feature fusion to improve the feature fusion part and used dilated convolution to enhance the receptive field of network layer, which solved the problem of the feature disappearance caused by too deep network [22]. (4) We generated boundary box and accurately predicted ship position, and the ship dataset was trained on a complete frame and compared with mainstream ship detection algorithms to verify the effectiveness of the AE-YOLOv3 algorithm, which provides a theoretical basis for water traffic safety.

2. Data Description

A new large-scale dataset of ships was proposed by Shao et al. It is called SeaShips and is designed for training and evaluating ship object detection algorithms [23]. The SeaShips dataset consists of 7000 images. All of the images were taken from approximately 7000 real-world video clips obtained by surveillance cameras in a deployed shoreline video surveillance system. They are carefully selected to cover all possible imaging variations, such as different scales, hull sections, lighting, viewpoints, backgrounds, and occlusion. All images are annotated with ship type labels and high-precision border boxes. In practice, the SeaShips dataset is expected to advance the research and application in ship detection. It contains 6 ship classes: ore carrier, bulk cargo carrier, general cargo ship, container ship, fishing boat, and passenger ship. The number of images in each ship category is shown in Table 1. Among them, the mixed type represents six classes of mixed categories with ships occluding each other in the image.

In order to improve the generalization of the framework, we use data enhancement technology to generate more ship images by applying generic enhancements [24]. The dataset was randomly scaled and rotated to increase to 14000. During the training process, 90% of the ship images were randomly selected as the training dataset, and the remaining 10% were used as the test dataset. Six types of ships are shown in Figure 1.

3. Methodology

The AE-YOLOv3 algorithm does not need to generate a region of interest in advance; instead, it directly trains the network in a regression way. By using the K-means algorithm to cluster the bounding boxes of the training samples, 3 groups of predefined bounding boxes were preset on 3 scale sizes, and the subsequent positioning prediction would be based on these 9 bounding boxes. Firstly, feature extraction was carried out on the original 416×416 input image through the feature extraction network, and then the feature vectors were fed into the Feature Pyramid Networks (FPN) structure to generate 3 grid regions on the scale, which were 13×13, 26×26, and 52×52, respectively. Each grid region predicted three bounding boxes. A total of

Table 1: Number of images in each ship category.

Ship category	Images	Percentage
Ore carrier	1141	0.1630
Bulk cargo carrier	1129	0.1613
Container ship	814	0.1163
General cargo ship	1188	0.1697
Fishing boat	1258	0.1797
Passenger ship	705	0.1007
Mixed type	765	0.1093
Total	7000	1

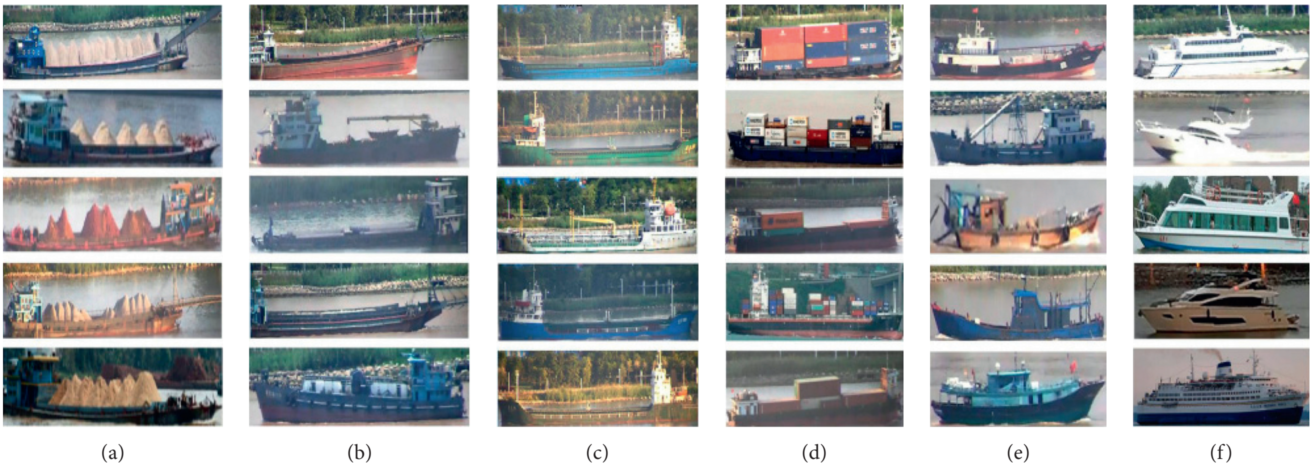


Figure 1: Example images of six ship types: (a) ore carrier; (b) bulk cargo carrier; (c) general cargo ship; (d) container ship; (e) fishing boat; (f) passenger ship.

$(52 \times 52 + 26 \times 26 + 13 \times 13) \times 3 = 10647$ boundary boxes were generated. A vector N is predicted in each bounding box, and the composition of vector N is shown in the following equation:

$$N = \left(t_x + t_y + t_w + t_h\right) + N_0 + \left(N_1 + N_2 + \cdots + N_n\right). \tag{1}$$

The first 4 elements in the vector N represent the 4 coordinates related to the bounding box. The calculation formulas for the distance from the center of the final predicted result's bounding box to the upper left corner of the feature map are shown in (2) and (3), and the calculation formulas for the length and width of the predicted bounding box are shown in (4) and (5):

$$b_x = \delta\left(t_x\right) + C_x, \tag{2}$$

$$b_y = \delta\left(t_y\right) + C_y, \tag{3}$$

$$b_w = p_w \times e^{t_w}, \tag{4}$$

$$b_h = p_h \times e^{t_h}. \tag{5}$$

Among them, δ represents Sigmoid function, C_x and C_y represent the offset of the grid to which the bounding box belongs relative to the upper left corner of the picture, p_h and p_w represent the length and width of the predefined bounding, b_x and b_y represent the distance from the center of the final predicted result's bounding box to the upper left corner of the image, and b_h and b_w represent the length and width of the predicted bounding box. N_0 represents the probability value of the object in the prediction box. The remaining n values in the vector N represent the scores of the predicted object belonging to one of the n categories, which are obtained by the Sigmoid function. Finally, nonmaximum suppression is performed on the generated prediction frame to obtain the final prediction result. The overall detection process of the AE-YOLOv3 algorithm is shown in Figure 2. It consists of three parts: Darknet-53 feature extractor, multiscale feature fusion, and multiscale detection branch structure.

The target detection process of AE-YOLOv3 is as follows: In the first step, feature extraction of training data is carried out through Darknet-53 embedded in feature attention module. In the second step, the spatial correlation degree of the target features is enhanced by the improved feature fusion part, and the extraction capability of small target feature information is enhanced. The third step is the generation of bounding box and ship position prediction, and the calculation of the boundary box is shown in (2)–(5). Then, the final ship mark box is obtained after applying the nonmaximum suppression (NMS) algorithm.

In Figure 2, Res-FAn represents a residual structure with n Res-FA modules, and Res-FA represents the residual structure of the embedded feature attention module.

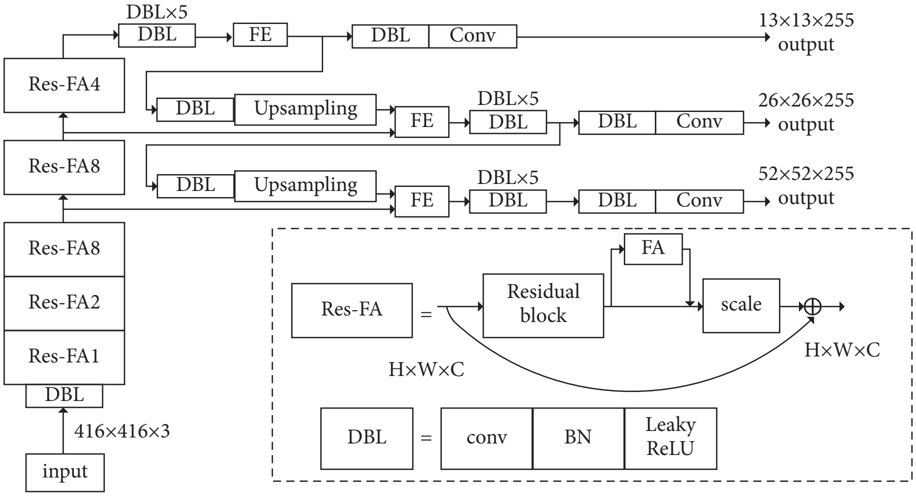


Figure 2: The algorithm network structure of AE-YOLOv3.

Darknetconv2d_BN_Leaky (DBL) consists of convolution, Batch Normalization (BN) [25], and Leaky Rectified Linear Unit (ReLU) [26].

Due to the complex background and irregular size of ship images, the recognition difficulty of the YOLOv3 algorithm increases. At the same time, aiming at the defect of YOLOv3 algorithm for small target recognition, considering the recognition accuracy and speed issues, this paper designed AE-YOLOv3 based on multiscale feature attention (FA) module and feature enhancement (FE) module. The FA module is embedded in the Darknet-53 network for weight redistribution of feature channel. Because the Darknet-53 network uses a large number of residual connections, the structure needs to be adjusted when the feature attention module is added. At the same time, the feature enhancement module is designed to embed the feature fusion part to enhance the ability to extract the feature information of small ship targets, and the FE module solves the problem of high-level feature maps becoming smaller due to convolution operation.

The specific network design process is as follows: ① In the Darknet-53 network, the designed FA module is embedded in the adjusted residual structure, and the weight of the feature channel relationship in the feature extraction network is redistributed, which strengthens the spatial connection of ship targets at different scales. ② Then, the calculated weight is weighted to the original ship image feature map through the scale operation, which strengthens the ability to extract features of multiscale targets, and finally outputs the corresponding feature map. ③ The three feature maps of Darknet-53 network are output as the input of feature fusion part; then, through the feature enhancement module, the detection performance of small target ships is enhanced.

3.1. Darknet-53 Structure. The Darknet-53 feature extractor uses a series of convolutional layers of 1×1 and 3×3, and each convolutional layer is connected to two neural network structural units: Batch Normalization and Leaky ReLU; two convolutional layers constituted a residual convolution group. There were five residual convolution groups in Darknet-53, and the residual convolution group adopted a jump-layer connection method, forming the residual block. After multiple convolution operations, the image of $416 \times 416 \times 3$ can be output to the image of $13 \times 13 \times 1024$. While improving the calculation speed of the algorithm, the complexity of the network is reduced, and the occurrence of overfitting is avoided.

3.2. Feature Attention (FA) Module. Like the human visual system, the attention mechanism consciously grabs the most useful information from enormous target information, can learn useful features, and suppresses useless features. By adding different weights to the feature channels transmitted by the neural network, the network will pay attention to the channel with larger weights for parameter update. In the process of forward propagation, the important feature channels occupy a larger proportion, and the final detection output image is also more prominent to show the focus of the network. Wang proposed a Convolutional Neural Network using attention mechanism through the study of feature networks [27]. With the deepening of the network, the attention module will learn adaptively to extract useful information from the images. With the attention mechanism being used by more and more researchers, it has been verified that it has a positive effect on the improvement of network performance.

When the YOLOv3 algorithm is used for ship detection, researchers usually focus on improving the overall accuracy but ignore the extremely small ships in the water area, which has certain safety risks in practical applications. Nowadays, more and more cargo is transported on the water, and the ship inertia is relatively large. Therefore, improving the detection accuracy of extremely small ships in the distance can be judged in advance, to avoid collision accidents. For the above problems, this paper constructs a feature attention (FA) module to recalibrate the feature channel to improve the feature extraction ability of the network. The structure is shown in Figure 3, where FC means fully connected, ReLU means Rectified Linear Unit.

The specific steps of the feature attention module are as follows: Firstly, convolution of 1×1 and 3×3 is added in front of global average pooling to realize cross-channel information integration, which enhances the spatial connection of ship images at different scales and the ability to extract multiscale features. Then, through global average pooling, the global spatial information of the feature map is transformed into a one-dimensional vector for summation, and the global information of the feature map is obtained. Global average pooling is a special pooling proposed by Lin et al. [28]. It is commonly used to aggregate spatial information, perform average pooling of the entire feature map, and finally get a value. The specific formula is shown as follows:

$$G_c = \frac{1}{H \times W} \sum\nolimits_{i=1}^{H} \sum_{j=1}^{W} U_c(i, j). \tag{6}$$

Among them, G_c is the vector sum after global average pooling of the feature map, H and W are the width and height of the input feature map, and $U_c(i, j)$ is the value of the c-th channel U_c at (i, j); then, the size of the feature map changes from $C \times H \times W$ to $C \times 1 \times 1$.

In order to strengthen the spatial connection between feature channels and obtain the weights of feature maps of different channels, the following processing needs to be performed on G_c:

$$X_c = \sigma(W_2 \delta(W_1 G_c)). \tag{7}$$

Among them,X_c is the weight of the corresponding feature map, $W_1 \in R^{(C/r)\times C}$ and $W_2 \in R^{C\times(C/r)}$ are the weight matrices of the fully connected layer, $\delta$is the ReLU function, and σ is the Sigmoid function.

Finally, the feature channel is recalibrated, and then the weight (X_c) is multiplied by the input feature map (U_c). The specific formula is shown as follows:

$$G_c = U_c \otimes X_c, \tag{8}$$

where U_c is the output matrix of the c-th channel after weight calibration.

3.3. Feature Enhancement (FE) Module. When the YOLOv3 algorithm is used for ship detection, the feature fusion part adopts the top-down fusion method. Due to the layered convolution, the high-level feature layer is greatly reduced in sensitivity to the small target feature information of the input image, and the learning ability is insufficient. Even if the original network will integrate the high-level feature layer with strong semantic information and the low-level feature layer, the problem of insufficient detection capabilities of the network for small targets cannot be avoided. For the above problems, a feature enhancement (FE) module is designed to directly act on the feature fusion part of the YOLOv3 algorithm, which increases the receptive field of the convolutional layer and the semantic information of the output feature layer.

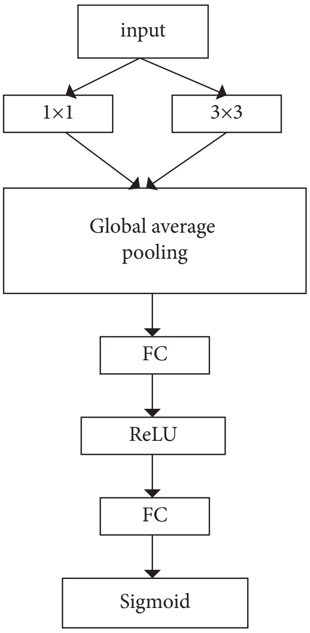


Figure 3: Feature attention module structure.

This paper uses dilated convolution that can increase the receptive field of the convolution layer, but not the amount of training parameters, by using different sizes of traditional convolution and different steps of dilated convolution to construct feature enhancement (FE) module; it is embedded in the feature fusion process. In the channel latitude, the feature correlation between networks is enhanced, thereby improving the performance of small target detection. The structure is shown in Figure 4.

Branch1, branch2, and branch3 are the three branch structures of the FE module. After the input ship image passes through the residual module of Darknet-53, it will output 5 feature maps, among which the output feature maps 3, 4, and 5 will be the input of the next module (see Figure 2). Since the feature fusion part adapts the top-down fusion method, after the FE module is embedded in the feature fusion part, the input of branch1 and branch2 is the high-level output feature map, and the input of branch3 is the adjacent low-level output feature map. After the concat operation, the feature fusion goal is achieved. The calculation formulas of the three multiscale output feature maps after feature fusion are shown as follows:

$$X_i = \begin{cases} \text{concat}[b_1(x_i), b_2(x_i)], & i = 5, \\ \text{concat}[b_1(x_i), b_2(x_i), b_3(x_{i+1})], & i = 3, 4. \end{cases} \tag{9}$$

Among them, $b_1(x_i), b_2(x_i), b_3(x_{i+1})$ represent the corresponding convolution combination of branch1, branch2, and branch3; x_i and x_{i+1} represent the output feature map of Darknet-53; and concat[] is the feature connection operation. Branch1 uses a traditional convolution of 1×1 and a dilated convolution with the size of 3×3 and the step size of 2; branch2 uses a traditional convolution of 3×3 and a dilated convolution with the size of 3×3 and the step size of 3; and branch3 uses deconvolution to operate, and to a certain extent it solves the problem that the high-

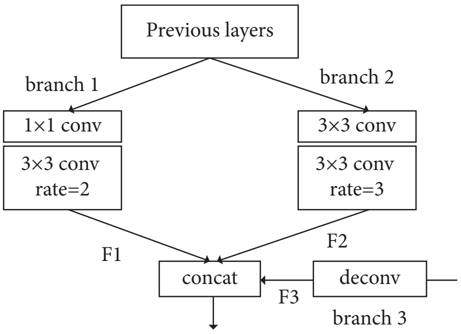


Figure 4: Feature enhancement module structure.

level feature map becomes smaller due to the convolution operation. F_1 is the calculation result of branch1, and the calculation formula is shown in (10); F_2 is the result of convolution with different size and step, and the calculation formulas of F_2 is shown in (11); F_3 is the result of deconvolution, and the specific calculation formula is shown in (12).

$$F_1 = F^{H\times W\times C} \otimes b^{1\times 1\times (C/2)} \otimes \left(A^{3\times 3\times (C/2)}\right)^2, \quad (10)$$

$$F_2 = F^{H\times W\times C} \otimes b^{3\times 3\times (C/2)} \otimes \left(A^{3\times 3\times (C/2)}\right)^3, \quad (11)$$

$$F_3 = \frac{i_s + 2p - k}{s} + 1. \quad (12)$$

Among them, $\otimes$ is the convolution operation, $F^{H\times W\times C}$ is the input feature map of the current layer of the feature enhancement module, b is the traditional convolution, A is the dilated convolution, $i_s$is the size of the input feature map in the next layer of the feature module, k is the size of the convolution kernel, s is the step size, and p is the boundary padding (0 in this paper). Finally, F_1, F_2, and F_3 are concatenated.

4. Experiments

The proposed algorithm is applied to ship target detection, which has been described in detail in the above sections. The experiment in this paper was carried out on Google Colaboratory. The GPU version is NVIDIA Tesla P100, which contains 16 GB RAM, and the simulation platform is PyTorch framework on Python (version 3.7).

In order to verify the superiority of the AE-YOLOv3 algorithm in this paper, we conduct comparative experiments in the horizontal and vertical directions based on the same experimental environment, and mean average precision (mAP) is used as the evaluation index, including precision and recall. The calculation formulas for precision and recall are shown as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (13)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (14)$$

where TP is the number of ships detected by the AE-YOLOv3, FP is the wrong detection of the ship target, and FN is the number of ship targets not detected. A larger mAP means that the algorithm's target detection accuracy is better. The mAP calculation formula is shown as follows:

$$\text{mAP} = \int_0^1 P(R)d(R), \quad (15)$$

where P is precision, R is recall, and the area of the P-R curve (see Figure 5) is the size of mAP.

Comparison experiment is conducted in the same experimental environment. The algorithm proposed in this paper, Faster Region-based Convolutional Neural Network (Faster R-CNN) [29], Single-Shot Detection (SSD) [30], and You Only Look Once version 3 (YOLOv3) [31] are used for ship target detection experiments. When the model stops converging, the training is terminated. After the training is completed, the weight file is called and tested based on Faster R-CNN, SSD, YOLOv3, and AE-YOLOv3. Frames per second (FPS) of the algorithm are tested at the same time, representing the number of pictures that can be processed per second. The test results are shown in Table 2.

According to Table 2, the detection accuracy of the water transportation ship detection algorithm proposed in this paper is higher than the mainstream detection algorithm, and it has achieved 98.72%. The improvement of accuracy is mainly due to the excellent network design of YOLOv3 algorithm and the addition of feature attention module and feature enhancement performance of feature fusion part. However, it is inferior to the original YOLOv3 algorithm in FPS; the main reason is that the FE module uses deconvolution to increase the resolution of the high-level feature map, which increases the amount of calculation and causes the model to slow down. Taking into account the requirements of detection accuracy and speed, the algorithm proposed in this paper can meet the requirements of ship target detection. Aiming at the reduction of FPS, future research will solve the problem of computation and information redundancy in the network, while ensuring accuracy, and improve the detection accuracy and speed of the model.

In order to better reflect the superiority of the model in this paper, recall rates and accuracy of the four algorithms were calculated respectively, and the corresponding P-R curve is drawn, as shown in Figure 5. Low recall means that there are few ship targets in the maritime image, and precision represents the detection accuracy of the algorithm. Experiments show that the algorithm proposed in this paper has a significant improvement in accuracy and recall and correctly detects more ship objects, which proves its superiority in ship detection.

Figure 6 shows the mAP comparison chart of the AE-YOLOv3 algorithm and the current three mainstream algorithms, and it also covers the various categories of the SeaShips dataset. In this paper, the score threshold is designed to be 0.6; that is to say, the detected ship target coincides with the boundary box to 60%, and the detection is regarded as successful. Among them, the average detection accuracy of AE-YOLOv3 for six types of ships is as high as

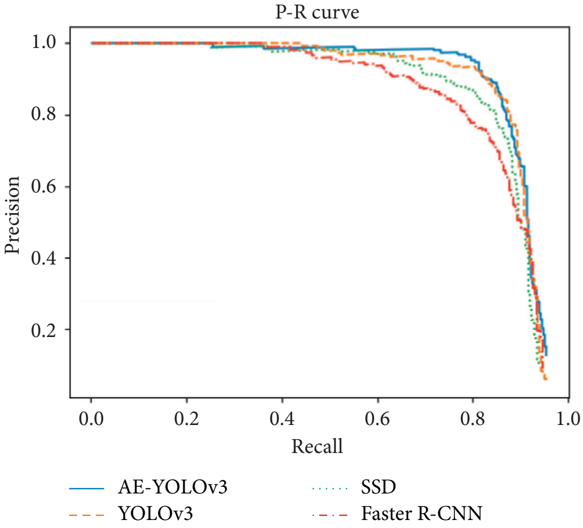


Figure 5: P-R curve comparison chart.

98.72%, which is 2.83% higher than the original YOLOv3 algorithm, 3.27% higher than the SSD algorithm, and 6.89% higher than the Faster R-CNN. For each category, the AE-YOLOv3 algorithm has a significant increase in the AP value of the small target fishing boat compared to the other three algorithms, indicating that adding a feature attention (FA) module to perform feature channel calibration can improve the detection of small targets. Another notable change is the ore carrier category; by observation of SeaShips dataset, it can be found that the image target containing ore carrier has a high mixing rate with the background, and the background has a great disturbance. However, the AP value of the ore carrier category of AE-YOLOv3 algorithm is 12%, 3%, and 1% higher than the other three. The main reason is that the embedded FE module can strengthen the feature information of the high-level feature layer, which increases the ability of the model to grasp the target information and avoids the error identification.

It can be seen from Table 3 that precision and recall have a significant improvement compared with the other three algorithms, indicating that the number of missed targets of the AE-YOLOv3 algorithm is significantly reduced (the ship image is judged as a nonship image) and its performance is significantly improved. Good target detector performance can provide reference for maritime personnel and avoid the danger of ship collision caused by the negligence of maritime personnel [32]. More specifically, applying the AE-YOLOv3 algorithm to the video tracking system will issue an early warning when a collision is about to occur, which greatly reduces the probability of marine accidents and promotes the development of intelligent shipping.

The visual inspection comparison is shown in Figure 7. It shows the detection results of different algorithms using the same training strategy. Among them, the first line is the detection result of AE-YOLOv3 algorithm, the second line is the detection result of the original YOLOv3 algorithm, the third line is the detection result of SSD algorithm, and the fourth line is the detection result of Faster R-CNN algorithm. According to Figure 7, compared with the other three algorithms, AE-YOLOv3 has a significant improvement. The number of ship categories detected by AE-YOLOv3 is significantly more than that of the other three. In particular, the detection effect of small targets is significantly improved, and the recognition rate of occluded targets is also higher, while the original YOLOv3, SSD, and Faster R-CNN have many missed targets. It can be seen from the comparison in the first column that AE-YOLOv3 has good results for the detection of multiple targets and small targets, while the original YOLOv3, SSD, and Faster R-CNN all have missed detection and low accuracy; it can be seen from the comparison between the second column and the third column that AE-YOLOv3 also has a good effect on the detection of occluded targets.

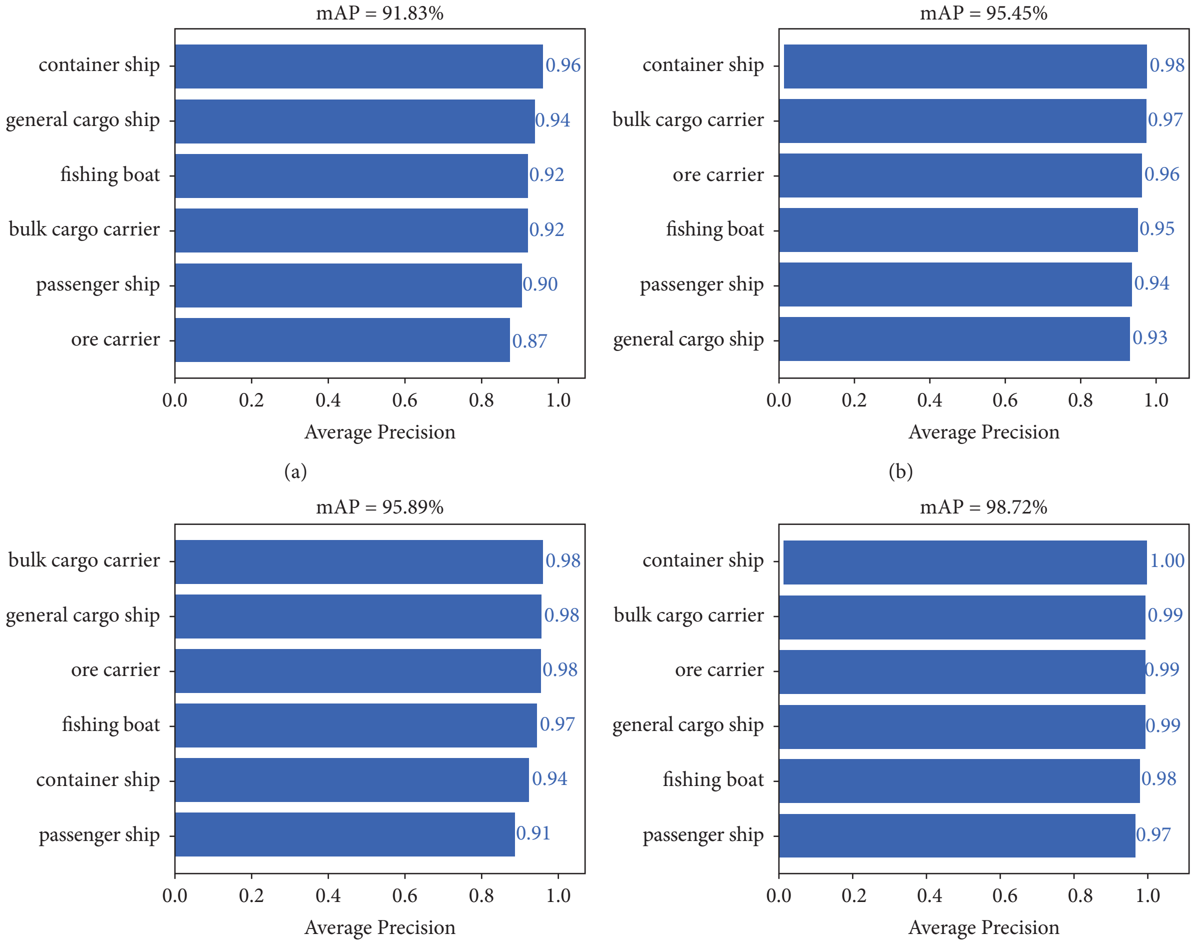
Table 2: Comparison of SeaShips dataset test results.

Detection algorithm	mAP (%)	FPS
Faster R-CNN	91.83	7
SSD	95.45	30
YOLOv3	95.89	35
AE-YOLOv3	98.72	32

From what has been discussed above, AE-YOLOv3 in this paper has a good detection effect for small targets, target occlusion, and incomplete target information and can circle the ship target with a suitable bounding box, especially for small fishing boats. The effect is significantly improved; it benefits from the design of feature attention module and feature enhancement module. Through the feature channel calibration of the feature extraction network, the spatial connection is strengthened. At the same time, the FE module of the feature fusion part strengthens the correlation and resolution between the high-level networks and enhances the feature extraction ability. An excellent target detector is extremely important. The excellent performance of AE-YOLOv3 can provide a reference basis for maritime affairs

Table 3: Ship detection performance for different algorithms.

Algorithm	Recall (%)	Precision (%)
Faster R-CNN	88.235	89.615
SSD	91.938	92.995
YOLOv3	93.145	94.628
AE-YOLOv3	97.568	97.388



(a)

(b)

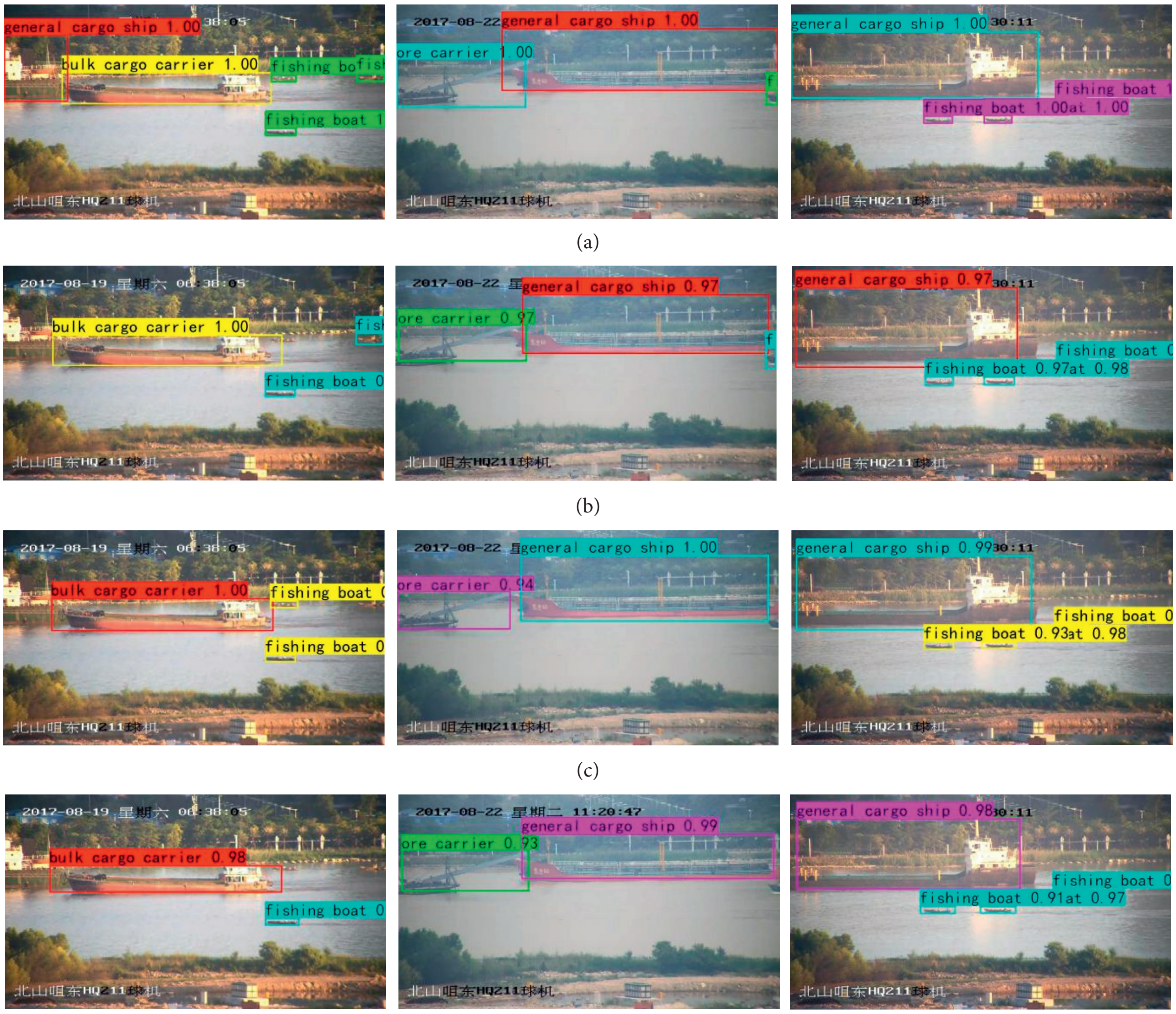
(c)

(d)

Figure 6: Comparison of AP curves of different algorithms: (a) Faster R-CNN's mAP; (b) SSD's mAP; (c) YOLOv3's mAP; (d) AE-YOLOv3's mAP.

bureaus and ship design, as well as providing a judgment basis for maritime related personnel to control ship navigation in advance to avoid water traffic accidents, so as to achieve the effect of safe navigation.

The use of video surveillance technology for automatic monitoring of ships plays an important role in marine safety and maritime transportation, fishery management, ship traffic monitoring, and so on [33–35]. Ship tracking is based entirely on the results of ship detection. If the detected target is wrong, then subsequent target tracking based on this will also make an error. For target detection algorithms, it is currently difficult to achieve very high accuracy, and occlusion has always been a difficult problem to solve. The improved algorithm in this paper solves this problem well, so it can provide a theoretical basis for the subsequent tracking of ships on the water.



(a)

(b)

(c)

(d)

Figure 7: The detection results of four different algorithms on SeaShips test set: (a) AE-YOLOv3; (b) YOLOv3; (c) SSD; (d) Faster R-CNN.

5. Conclusions

Ship target detection and tracking are very important to the development of intelligent shipping, which requires ships to overcome all difficulties in the complex navigable environment, so as to avoid the occurrence of water traffic accidents. We proposed an algorithm model based on YOLOV3 to detect ships in maritime images. AE-YOLOv3 was implemented in three steps. The first step is to construct a feature attention module by introducing an attention mechanism and embed it in Darknet-53 for feature recalibration. The second step is to build a feature enhancement module and apply it to feature fusion to enhance the receptive field size of the corresponding feature layer and the relevance of the feature extraction network. The third step is to output multiscale feature map by predicting the branch structure to obtain the best detection frame.

Although our method has achieved excellent performance in ship target detection, there are still some limitations in this work; we can do some research to improve the performance of the algorithm in the future. First, we considered that the ships in the SeaShips dataset are all in a horizontal position, and this work was carried out on the SeaShips dataset; therefore, it is necessary to collect multiangle ship images as training dataset in the future. Second, maritime images were collected in a good environment (without disturbance from storm, rain, and snow), so detecting ship objects at complex environment will be an important research in the future. Last but not least, the image of the surveillance video will be blurred due to vibration when the ship is sailing, so testing the performance of our model under the vibration background will be a good exploration.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (61463020 and 61763015).

References

[1] R. W. Liu, J. Nie, S. Garg, Z. Xiong, Y. Zhang, and M. S. Hossain, "Data-driven trajectory quality improvement for promoting intelligent vessel traffic services in 6G-enabled maritime IoT systems," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5374–5385, 2021.

[2] Y. Gui, X. Li, and L. Xue, "A multilayer fusion light-head detector for SAR ship detection," *Sensors (Basel, Switzerland)*, vol. 19, no. 5, 2019.

[3] B. Wu, Y. H. Tang, X. P. Yan, and C. G. Soares, "Bayesian network modelling for safety management of electric vehicles transported in RoPax ships," *Reliability Engineering & System Safety*, vol. 209, Article ID 107466, 2021.

[4] J. M. Mou, P. F. Chen, Y. X. He et al., "Vessel traffic safety in busy waterways: A case study of accidents in western shenzhen port," *Accident Analysis & Prevention*, vol. 123, pp. 461–468, 2019.

[5] Y. Yang and X. Chen, "Research on the statistical method of ship flow based on deep learning and virtual detection line," in *Proceedings of the IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, pp. 280–285, Chongqing, China, May 2019.

[6] R. G. Wright, "Intelligent autonomous ship navigation using multi-sensor modalities," *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, vol. 13, no. 3, pp. 503–510, 2019.

[7] B. Wu, T. T. Cheng, T. Leung, and Y. Wang, "Fuzzy logic based dynamic decision-making system for intelligent navigation strategy within inland traffic separation schemes," *Ocean Engineering*, vol. 197, 2020.

[8] B. Liu, S. Z. Wang, Z. X. Xie, J. Zhao, and M. Li, "Ship recognition and tracking system for intelligent ship based on deep learning framework," *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, vol. 13, no. 4, pp. 699–705, 2019.

[9] Z. Shao, L. Wang, Z. Wang, W. Du, and W. Wu, "Saliency-aware convolution neural network for ship detection in surveillance video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 3, pp. 781–794, 2020.

[10] Z. Liu, J. Hu, L. Weng, and Y. Yang, "Rotated region based CNN for ship detection," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 900–904, Beijing, China, September 2017.

[11] L. Deng and D. Yu, "Deep learning: Methods and applications," *Foundations and Trends in Signal Processing*, vol. 7, no. 3, pp. 197–387, 2014.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[13] L. Qi, B. Li, L. Chen et al., "Ship target detection algorithm based on improved faster R-CNN," *Electronics*, vol. 8, no. 9, 2019.

[14] H. W. Guo, H. Y. Bai, Y. X. Zhou, and W. Li, "DF-SSD: A deep convolutional neural network-based embedded lightweight object detection framework for remote sensing imagery," *Journal of Applied Remote Sensing*, vol. 14, no. 1, 2020.

[15] Y. Huang, Y. Li, Z. Zhang, and R. W. Liu, "GPU-accelerated compression and visualization of large-scale vessel trajectories in maritime IoT industries," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 10794–10812, 2020.

[16] H. Huang, D. Sun, R. Wang, C. Zhu, and B. Liu, "Ship target detection based on improved YOLO network," *Mathematical Problems in Engineering*, vol. 2020, Article ID 6402149, 10 pages, 2020.

[17] J. Redmon, S. Divvala, R. Girshick et al., "You only look once: Unified, real-time object detection," *Computer Vision and Pattern Recognition*, pp. 779–788, IEEE Computer Society, Boston, MA, USA, 2015.

[18] F. Fukun Bi, B. Bocheng Zhu, and L. Mingming Bian, "A visual search inspired computational model for ship detection in optical satellite images," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 4, pp. 749–753, 2012.

[19] X. Nie, M. Yang, and R. W. Liu, "Deep neural network-based robust ship detection under different weather conditions," in *Proceedigns of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 47–52, Auckland, New Zealand, October 2019.

[20] X. Chen, L. Qi, Y. Yang et al., "Video-based detection infrastructure enhancement for automated ship recognition and behavior analysis," *Journal of Advanced Transportation*, vol. 2020, Article ID 7194342, 12 pages, 2020.

[21] 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.

[22] T. Zhang, X. Zhang, J. Shi, and S. Wei, "Depthwise separable convolution neural network for high-speed SAR ship detection," *Remote Sensing*, vol. 11, no. 21, 2019.

[23] 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.

[24] J. Jiang, X. Fu, R. Qin, X. Wang, and Z. Ma, "High-speed lightweight ship detection algorithm based on YOLOv4 for three-channels RGB SAR image," *Remote Sensing*, vol. 13, no. 10, 2021.

[25] S. Loffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning, PMLR*, pp. 448–456, Lille, France, July 2015.

[26] A. K. Dubey and V. Jain, "Comparative study of convolution neural network's relu and leaky-relu Activation functions," *Lecture Notes in Electrical Engineering*, vol. 553, pp. 873–880, 2019.

[27] F. Wang, M. Jiang, Q. Chen et al., "Residual attention network for image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3156–3164, Honolulu, HI, USA, July 2017.

[28] M. Lin, Q. Chen, and S. Yan, "Network in network," *Computer Vision and Pattern Recognition*, Springer, Berlin, Germany, 2014.

[29] Y. Ren, C. Zhu, and S. Xiao, "Small object detection in optical remote sensing images via modified faster R-CNN," *Applied Sciences*, vol. 8, no. 5, 2018.

[30] G. Yu, H. Fan, H. Zhou, T. Wu, and H. Zhu, "Vehicle target detection method based on improved SSD model," *Journal of Artificial Intelligence*, vol. 2, no. 3, pp. 125–135, 2020.

[31] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *Computer Vision and Pattern Recognition*,

Springer, Berlin, Germany, 2018, https://arxiv.org/abs/1804.02767.

[32] W. He, S. Xie, X. Liu et al., "A novel image recognition algorithm of target identification for unmanned surface vehicles based on deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 4, pp. 4437–4447, 2019.

[33] H. Li, L. Chen, F. Li, and M. Huang, "Ship detection and tracking method for satellite video based on multiscale saliency and surrounding contrast analysis," *Journal of Applied Remote Sensing*, vol. 13, no. 2, 2019.

[34] X. Chen, Z. Li, Y. Yang, L. Qi, and R. Ke, "High-resolution vehicle trajectory extraction and denoising from Aerial videos," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 5, pp. 3190–3202, 2021.

[35] L. Zhao and G. Shi, "A trajectory clustering method based on Douglas-Peucker compression and density for marine traffic pattern recognition," *Ocean Engineering*, vol. 172, pp. 456–467, 2019.