Marine ship detection and classification based on YOLOv5 model

Xuan Zhang1*, Mingzhong Yan1, Daqi Zhu2 and Yang Guan3

1 Logistics Engineering College, Shanghai Maritime University, Shanghai, 201306, China
2 Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai, 200093, China
3 Shanghai branch of Southwest Electronics and Telecommunication Technology Research Institute, Shanghai, 200434, China

*Corresponding author’s e-mail: 1799149500@qq.com

Abstract. An improved deep learning neural model YOLOv5-DN based on YOLOv5 is proposed for marine ship detection and classification in the area of harbours and heavy traffic waterways. The CSP-DarkNet module in YOLOv5 is replaced by CSP-DenseNet to promote the accuracy of target detection and classification in the proposed model. Sample marine ships in the data set are divided into six classes: ore carriers, general cargo ships, bulk cargo ships, container ships, passenger ships, and fishing ships to meet the detection needs in the areas of ports and waterways. The data set are grouped into a training set, testing set, and validating set by the proportion of 6:2:2. Experiments show that the improved model has better average accuracy, from 62.2% to 71.6%.

1. Introduction
With the rapid development of global economic activities, maritime traffic management is facing more challenges than ever[1]. In order to promote the efficiency of ship traffic service (VTS), intelligent target detection and classification technique is playing an important role in the maritime field[2].

The traditional target detection algorithms are mostly based on the sliding window model, which is not targeted to the selection of candidate regions, and is easy to produce window redundancy, and has disadvantages such as poor robustness and poor applicability[3]. In recent years, with the rapid development of artificial intelligence, deep learning has been widely used in target detection[4], which has higher detection accuracy and speed than traditional target detection algorithms. Deep learning neural network models can be divided into two types: one-stage and two-stage. Common one-stage models mainly include SSD, YOLO, etc. Yanpeng Wang, etc.[5] based on SSD model and using transfer learning technology, proposed a recognition algorithm for inland river ship targets based on single-pass multi-frame detector. Mingwei Sheng, etc.[6] proposed a ship detection algorithm based on the improved YOLOv3 model in view of the poor detection ability of the traditional target detection algorithm to ship targets under the changeable marine environment conditions. Nan Li, etc.[7] proposed an improved YOLOv5 model to detect SAR images of ships in complex scenes.

R-CNN[8] and Faster R-CNN are typical two-stage models, Xin Zhang, etc. [9] improved the R-CNN algorithm by applying soft-NMS to improve the accuracy of ship images detection. Jianghong Zhao, etc.[10] improved the Faster R-CNN and detected ship targets in remote sensing images by...
replacing the VGG-16 network with ResNet. The two-stage model divides target detection into two steps, that is, candidate regions are generated in the image first, and then targets in the candidate regions are positioned and classified. However, the one-stage model skips the step of candidate regions generation and greatly improves the speed of target detection compared with the two-stage model[11].

In this paper, YOLOv5-DN model is constructed by replacing CSP-Darknet structure with CSP-DenseNet structure in YOLOv5, so as to improve the detection and classification ability of multi-type ship targets in harbors and channels.

2. YOLOv5 network structure

As shown in figure 1, the structure of YOLOv5 can be divided into 4 parts: INPUT, Backbone, Neck, and Prediction. After the image is input in YOLOv5, the size is firstly automatically adjusted, and then the image is divided into S×S grid cells to predict whether there are objects to be detected in the box within a certain height (h) and width (w) with each grid as the center, namely confidence. Each cell will predict M borders and generate M confidence values, and then predict whether the object to be detected is included or not, if it does, the class of all N classes which it belongs to will be detected, and N probability vectors are given. Then, the Non-Maximum Suppression algorithm [12](NMS) will be used to screen out the box with the maximum score of the same category in a certain region and remove the redundant target box.

![Figure 1. YOLOv5 network structure.](image)

2.1. The CSP-DarkNet in YOLOv5

The structure of CSP-Darknet in YOLOv5 is shown in figure 2. The input image enters Part1 after the Base layer for convolution operation (which can be omitted). The other part enters Part2 for convolution and then enters Dark Block module for feature map sampling and then enters Transition layer. The feature map with the same size is obtained after operation respectively, and finally enters Transition layer. Due to the shallow depth of the CSP-Darknet, there is a lot of room for improvement in accuracy. In this experiment, CSP-Darknet is improved to achieve higher accuracy.
3. The improvement of Backbone

3.1. CSP-DenseNet

The one-stage CSP-DenseNet structure is shown in Figure 3, in which each stage includes a local Dense module and a local Transition layer. In the local Dense module, the feature map of the base layer is divided into two parts through the channel \( x_0 = [x'_0, x''_0] \), where \( x''_0 \) is directly connected to the Transition layer at the end of the stage, \( x'_0 \) is connected to the Transition layer through the whole Dense module. The output of the Dense layer \([x'_0, x_1, \cdots, x_k]\) will go through a Transition layer, and the output, \( x_T \), will be spliced with \( x'_0 \), \( x_U \) is then output through another Transition layer. The parameters of forward propagation equation and CSP-DenseNet are updated as follows:

\[
\begin{align*}
    x_k &= w_k \ast [x'_0, x_1, \cdots, x_{k-1}] + b_k \quad (1) \\
    x_T &= w_T \ast [x'_0, x_1, \cdots, x_k] + b_T \quad (2) \\
    x_U &= w_U \ast [x'_0, x_T] + b_U \quad (3) \\
    w'_k &= f(w_k, g''_0, g_1, g_2, \cdots, g_{k-1}) \quad (4) \\
    w'_T &= f(w_T, g''_0, g_1, g_2, \cdots, g_{k-1}) \quad (5) \\
    w''_T &= f(w''_U, g'_0, g_1, \cdots, g_T) \quad (6)
\end{align*}
\]

*"represents convolution operation, \( b_k \) represents bias, \([x_0, x_1, \cdots]\) represents the concat operation, \( w_i \) represents the weight, \( x_i \) is the output of the \( i \)th dense layer. The \( f \) is the weight updating function, and \( g_i \) represents the gradient transmitted back to the \( i \)th dense layer.

It can be seen that the structure integrates the gradient information from the Dense layer and the feature map that doesn't go through the Dense layer, \( x'_0 \) separately. In this way, repeated gradient information will not be included in the weight update. Compared with Dense-Net, CSP-DenseNet not only retains the characteristics of Dense-Net, but also reduces the number of parameters and computation, making detection faster without affecting the detection accuracy. Moreover, CSP-DenseNet has more feature reuse than CSP-DarkNet, making CSP-DenseNet structure more sensitive to features. Meanwhile, CSP-DenseNet is less likely to lose gradient information during network training than CSP-DarkNet, and gradient information transmission is more efficient. Therefore, this paper replaced the CSP-DarkNet structure in YOLOv5 with the CSP-DenseNet structure to form the YOLOV5-DN model, which has higher detection accuracy than the YOLOv5 model.
4. Experimental results and analysis

4.1. Experimental environment
The experimental model in this paper is based on the Framework of Pytorch1.7.1 and the platform is PyCharm Community Edition 2020.2.3. The model training is completed in the experimental environment of GeForce RTX 2080GPU and CUDA10.1. The operating system is Windows.x64. Using Python for model development.

4.2. Data set and experimental parameters
In this experiment, ships are classified into ore carrier, general cargo ship, bulk cargo ship, container ship, passenger ship and fishing ship according to their appearance and use. 7000 images of various ships are collected, and the categories of images are shown in table 1.

| Classes          | Quantity (pieces) | Ratio |
|------------------|-------------------|-------|
| ore carrier      | 1044              | 0.149 |
| general cargo ship | 1083            | 0.154 |
| bulk cargo ship  | 1136              | 0.162 |
| container ship   | 1159              | 0.165 |
| passenger ship   | 877               | 0.125 |
| fishing ship     | 1058              | 0.151 |
| mixed            | 643               | 0.091 |

The ratio of training set, test set and verification set in the experimental data set is set as 6:2:2. After adjustment, the batch-size in the neural network is eventually set to 8 and epochs to 300.

4.3. Experimental results and analysis
Mean Average Precision(mAP) is one of the important criteria to measure the target detection by the model. For multi-target detection, each target should have an AP value first, and then take the weighted average, namely mAP. AP is the area enclosed by the X-axis and Y-axis plots using Recall(R) and Precision(P) respectively.

YOLOv5 and YOLOV5-DN are used for training respectively, and AP values of various classes and mAP@0.5 indexes are compared, as shown in figure 4 and figure 5.

![Figure 4. Training results of YOLOv5](image)
Figure 5. Training results of YOLOv5-DN.

Statistics and comparison are shown in table 2.

| Models       | ore carrier | general cargo ship | bulk cargo ship | container ship | passenger ship | fishing ship | mAP@0.5     |
|--------------|-------------|--------------------|-----------------|----------------|----------------|--------------|-------------|
| YOLOv5       | 0.828       | 0.539              | 0.664           | 0.845          | 0.218          | 0.637        | 0.622       |
| YOLOv5-DN    | 0.880       | 0.725              | 0.687           | 0.931          | 0.426          | 0.644        | 0.716       |
| Upgrade rate | 6.28%       | 34.50%             | 3.46%           | 10.18%         | 95.41%         | 1.10%        | 15.11%      |

Through comparison, it can be found that the detection accuracy of the YOLOv5-DN model for each class of ship targets is higher than that of YOLOv5. The accuracy of fishing ships remains basically stable. The accuracy of container ships increased from 0.845 to 0.931, and the accuracy of passenger ships increased from 0.218 to 0.246, the upgrade rate is as high as 95.41%. Besides, mAP@0.5 increases from 0.622 to 0.716.

Subsequently, the YOLOv5 and YOLOv5-DN models are used to detect the images in verification set respectively. The results of the detection comparison are shown in figure 6 and figure 7.
Figure 6. Detection results of YOLOv5.

Figure 7. Detection results of YOLOv5-DN.

Statistical results are shown in table 3:

| Images | Classes of ships | Detection results | Images | Classes of ships | Detection results |
|--------|------------------|-------------------|--------|------------------|------------------|
| a1     | ore carrier      | Not detected      | b1     | ore carrier      | 0.76             |
| a2     | general cargo ship | Not detected     | b2     | general cargo ship | 0.80             |
| a3     | bulk cargo carrier | Not detected    | b3     | bulk cargo carrier | 0.67             |
| a4     | container ship   | Error detection   | b4     | container ship   | 0.70             |
| a5     | passenger ship   | 0.57              | b5     | passenger ship   | 0.72             |
| a6     | bulk cargo carrier | Prediction boxes overlap | b6     | bulk cargo carrier | Prediction boxes separation |

Table 3. Comparison of detection results between the two models.
By comparison, it can be seen that the YOLOv5-DN model can detect ore carriers, general cargo ships and other classes more accurately when they are not detected in YOLOv5. In addition, the improved YOLOV5-DN model can detect container ships and other ships more accurately when they are mistakenly detected in YOLOv5. Meanwhile, when the prediction boxes are overlapped in YOLOv5, YOLOv5-DN model can better separate the prediction boxes and improve the confidence. Therefore, the YOLOv5-DN model proposed in this paper is superior to YOLOv5 in overall accuracy of target detection.

5. Conclusion
In terms of the accuracy of detecting various classes of ships, the YOLOv5-DN model has an effective improvement compared to YOLOv5. Among them, the general cargo ships and passenger ships have a better improvement effect than other classes, and the detection accuracy of container ships and ore carriers exceeds 0.87. Although the detection accuracy upgrade rate of passenger ships is as high as 95.41%, the effect is still not good. The main reason is that the number of images of passenger ships in the data set is relatively small, and the ship targets are different sizes. In order to further improve the accuracy of the model's detection effect of this class, the follow-up should mainly focus on establishing sufficient data sets for this class of ship target, and further optimizing the network structure and strengthening learning and training. At the same time, try to improve the model's target detection ability under complex backgrounds, such as sea fog and insufficient light.

This project is supported by the National Natural Science Foundation of China (62033009,U1706224) and the Science and Technology Innovation Action Plan of Shanghai (206510712300,18DZ2253100).

References
[1] Liu, B., Wang, S.Z. (2019) Ship tracking recognition based on Darknet network and YOLOv3 algorithm. Journal of Computer Applications, 39: 1663-1668.
[2] Liu, S.Z. (2021) Target edge detection method of ship remote sensing image under noise interference. Ship Science And Technology, 43: 190-192.
[3] Xing, Y.C., Li, D.J. (2021) Remote sensing image target detection based on YOLOv5. JiangXi Science, 39: 725-732.
[4] Zhao, X.B., Qin, Y.Y. (2021) A review of object detection algorithms based on deep learning. Automobile Applied Technology, 46: 207-209.
[5] Wang, Y.P., Yang, Y. (2019) Single shot multibox detector for ships detection in inland waterway. Journal of Harbin Engineering University, 40: 1258-1262.
[6] Sheng, M.W., Li, J. (2021) Ship target detection algorithm based on the improved YOLOv3. Navigation And Control, 20: 95-109.
[7] Li, N., Ye, X.D. (2021) A ship detection method for SAR images in complex scene based on Improved YOLOv5. Journal of Signal Processing, 16: 1-16.
[8] Wang, L.L., Zhang, X. (2019) Identification of sewage treatment plants in SAR images based on Faster R-CNN. Journal of Computer Applications, 39: 50-54.
[9] Zhang, X., Guo, F.L. (2020) Detection and recognition of ships at sea based on R-CNN algorithm. Application Research of Computers, 37: 314-315+319.
[10] Zhao, J.H., Zhang, X.G. (2020), Ship targets recognition in remote sensing images based on deep learning. Science of Surveying and Mapping, 45: 110-116.
[11] Fan, J.L., Tian, S.B., Huang, K. (2020) Multi-scale object detection algorithm for aircraft carrier surface based on Faster R-CNN. Journal of Systems Engineering and Electronics, 16: 1-10.
[12] Chen, P.C., Qi, B.R. (2021) Research on the application of an improved non-maximum suppression algorithm in YOLOv3 model. Digital Technology and Application, 39: 121-123.