Ship Remote Sensing Target Recognition Based on YOLOV5

Check for updates

Ning Hao, Yunwei Li, Yusen Ma, and Xinan Zhang

Abstract Ship remote sensing target recognition is a critical task in various maritime applications, including surveillance, navigation assistance, and disaster management. However, traditional methods face challenges in detecting and recognizing ships in complex maritime environments, which include various types of ships, sea conditions, and environmental factors. In recent years, deep learning-based object detection algorithms have shown promising results in detecting and recognizing ships in remote sensing images. In this paper, we propose a ship remote sensing target recognition method based on the YOLOV5 algorithm. Our approach uses a deep convolutional neural network to extract high-level features from remote sensing images and detect and classify ships. The proposed method uses anchor-based object detection to identify ship locations and a multi-scale feature fusion strategy to capture different ship sizes and orientations. We also introduce a new ship dataset, which includes various ship types and sea conditions, to evaluate the performance of our proposed method. Experimental results show that our method outperforms other common ship detection algorithms in terms of detection accuracy. Our method can significantly contribute to improving ship detection and recognition in real-world maritime applications, especially in complex scenarios.

Keywords YOLOV5 · Remote sensing image · Deep learning

N. Hao
Beijing Institute of Spacecraft Environment Engineering, Haidian District Space City, Beijing 100048, China

Y. Li · Y. Ma (✉)
College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China
e-mail: mayusen@hrbeu.edu.cn

Modeling and Emulation in E-Government National Engineering Laboratory, Harbin Engineering University, Harbin 150001, China

X. Zhang
School of Physics, Mathematics and Computing, University of Western Australia, Perth, WA 6009, Australia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
S. Li (ed.), *Computational and Experimental Simulations in Engineering*, Mechanisms and Machine Science 146, https://doi.org/10.1007/978-3-031-44947-5_44

551

1 Introduction

Remote sensing is a non-contact and long-distance detection technology that utilizes sensors or remote sensors to detect the electromagnetic radiation and reflection characteristics of objects. Through information transmission, processing, interpretation, and analysis, the geometric features and physical properties of objects can be obtained. With the continuous development of remote sensing technology, the detailed information contained in high-resolution remote sensing images has become increasingly abundant, making remote sensing one of the most effective ways to collect water surface object information. Identifying, classifying, and extracting information from ship images in remote sensing has become a hot research topic among scholars. Studying the detection methods of ship targets in remote sensing images is of great significance in both civilian and military fields.

However, extracting information from remote sensing images is particularly difficult due to the complexity of the remote sensing images themselves. Compared with natural image detection, remote sensing targets have the characteristics of strong uncertainty, large scale differences, and dense distribution of targets. The size of targets in remote sensing images is often different due to different capture heights. Many targets are small in size and closely arranged, so they are easy to be ignored. In addition, the complex background of remote sensing images is also likely to interfere with recognition. All of these present challenges for target detection and recognition.

At present, scholars have done a lot of research work on object detection in remote sensing images and proposed many solutions. In the early days, for the problem of ship target recognition, the ship target detection method based on support vector machine [1] was usually used to detect ships, but its technical performance was relatively poor. In 2011, Xia et al. [2] proposed an uncertain ship target extraction algorithm based on the dynamic fusion model of multi-features and variance features of optical remote sensing images, which further improved the ship recognition rate of remote sensing images. In 2016, a new ship detection method SVDNet [3] based on convolutional neural network and singular value solution compensation algorithm greatly improved the speed of ship detection. In 2018, reference [4] proposed a new method for offshore ship detection based on Mask R-CNN, which enhanced the robustness of offshore ship detection. In 2020, the application of dilated convolution on Faster R-CNN [5] improved the ability to extract ship features. At the same time, Li and Cai [6] proposed an improved remote sensing image ship detection algorithm based on Yolo V3, which uses window sliding segmentation technology to cut a large image into several small images, which improve the performance of small target recognition. In 2021, Reference [7] improves the detection performance of ship targets and the ability to recognize rotating targets by introducing Feature Pyramid Network (FPN) and Rotating Region Proposal Network (RRPN). In 2022, reference [8] proposed a ship recognition method for weakly supervised ship detection by separating objects from the background through an attention mechanism, which further improved the accuracy of ship detection.

But there are still challenges in ship target detection in remote sensing images, such as complex scenes, ships appearing in arbitrary directions, dense distribution of ships, large scale variations, significant appearance changes, target scale changes, and class imbalance. To address these issues, we propose an improved ship detection method for remote sensing images based on the YOLOv5 algorithm. We enhance the detection performance by adding two additional convolutional layers after the SPP (Spatial Pyramid Pooling) layer. Specifically, we adopt the SPPFCSPC [9] architecture, which stands for SPP-Fully Connected-Spatial Pyramid Convolutional layers, known to perform well in various object detection tasks. Furthermore, we integrate Transform Prediction Heads (TPH) into the YOLOv5 network model. TPH is a novel approach for object detection, where it replaces part of the convolutional layers in the YOLOv5 network model. TPH consists of a multi-head attention layer and a fully connected layer, which can capture local information and exploit the potential of feature representations using an attention mechanism. By incorporating TPH into the YOLOv5 network model, our proposed method can better detect small targets in remote sensing images.

2   Target Detection in Remote Sensing

Accurately detecting small targets on the sea surface and small objects in space, is crucial for preventing accidents, such as detecting enemy ships before they enter our territorial waters and avoiding disasters through advanced planning. However, target detection faces challenges in accurately detecting small targets, especially in high-resolution images. For instance, in a 1024×1024 image that contains numerous small targets, the detection difficulty is significantly increased.

The main difficulties in designing small sample target detection algorithms in current detection algorithms are the excessively large sampling rates, excessively large receptive fields, and conflicts between semantic space and space. Assuming the length of a small object is 10×10, the sampling rate under convolutional conditions in general target recognition detection is about 16%, and small objects may not even occupy a single pixel on the feature map, which further increases the difficulty of the target measurement system. Moreover, in convolutional networks, the receptive field of feature points is relatively larger than the downsampling rate, which means that small objects occupy fewer features and may contain features from surrounding areas, making it challenging to detect small targets. Most detection algorithms currently use a top-down approach, where deep and shallow feature maps may have conflicts in semantics and space due to poor balancing.

To optimize small object detection, the following methods can be considered: using SPP-Fully Connected-Spatial Pyramid Convolutional (SPPFCSPC) instead of SPPF, and integrating Transform Prediction Heads (TPH) into the YOLOv5 network model.

2.1 SPPFCSPC

The SPPFCSPC architecture consists of a traditional Spatial Pyramid Pooling (SPP) layer followed by two fully connected layers and two convolutional layers. The fully connected layers are utilized to reduce the dimensionality of the feature maps generated by the SPP layer, while the convolutional layers extract more complex and discriminative features from the reduced feature maps. This architecture effectively captures multi-scale features, thereby improving the accuracy of object detection. The structure diagram of the SPPF layer is depicted in Fig. 1, and the structure diagram of the SPPFCSPC layer is illustrated in Fig. 2.

By incorporating the SPPFCSPC architecture into our proposed method, we are able to handle input images of varying sizes and achieve superior detection performance. Our experimental results on our ship dataset demonstrate that our method outperforms the standard YOLOv5 model with only the SPP layer in terms of detection accuracy and computational efficiency.
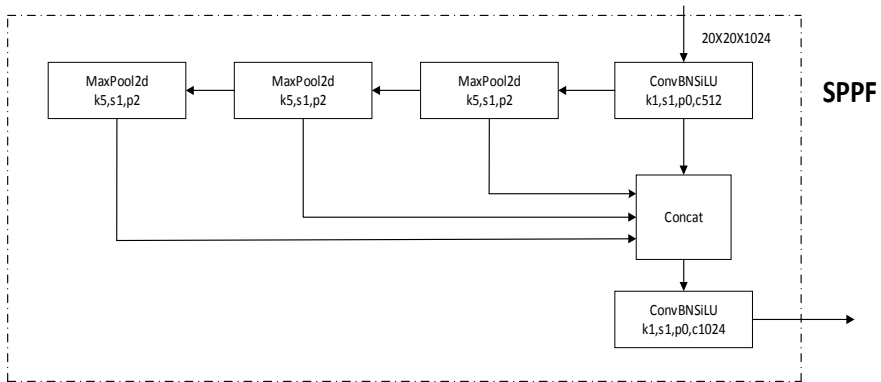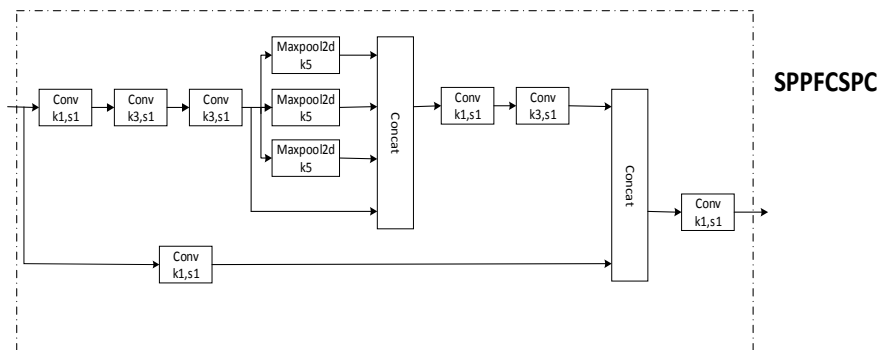

Fig. 1 SPPF structure diagram

Fig. 2 SPPFCSPC structure diagram

In conclusion, utilizing the SPPFCSPC architecture as opposed to the standard SPP layer is a straightforward yet effective approach to enhance the detection performance of the YOLOv5-based ship remote sensing target recognition method. Our proposed method has potential applications in various maritime scenarios, including surveillance, navigation assistance, and disaster management.

2.2 Transform Prediction Heads

The TPH-YOLOv5 approach [10] involves replacing certain convolutional blocks and CSP bottleneck blocks in the YOLOv5 network model with Transform encoder blocks. The Transform encoder block is composed of two sub-layers: a multi-head attention layer and a fully connected layer (MLP), which are connected using a residual network. This allows the Transform encoder block to capture local information and leverage attention mechanisms to extract potential feature representations.

Specifically, the multi-head attention layer learns to attend to different parts of the input feature map, enabling it to capture fine-grained information. The MLP layer then transforms the attended feature map into a higher-dimensional space, facilitating complex feature interactions. By combining these two layers in a residual block, the TPH-YOLOv5 approach can capture both local and global information, resulting in more informative feature representations.

Overall, the integration of the Transform encoder block in TPH-YOLOv5 enhances the feature representation capabilities of the network, improving its ability to detect small objects. The updated YOLOv5 structure diagram is shown in Fig. 3.

3 Experiments

3.1 Dataset

In order to evaluate the performance of the ship remote sensing target recognition method proposed in this paper under the conditions of multi-scale and different ship orientations, we collected a batch of satellite visible light imaging datasets of ships. This dataset includes remote sensing images of various ships captured from different angles and positions. There are a total of 1000 pictures, all of which are high-resolution pictures with a resolution of 1024×1024. The picture size ranges from 140 to 350 kb, and 80% of the pictures are larger than 300 kb. Each picture contains at least 1 and at most 21 ship data objects making it suitable for training and testing ship detection and recognition algorithms. To facilitate the training process, we use the VOC annotation method to annotate the data, so that ships can be accurately and
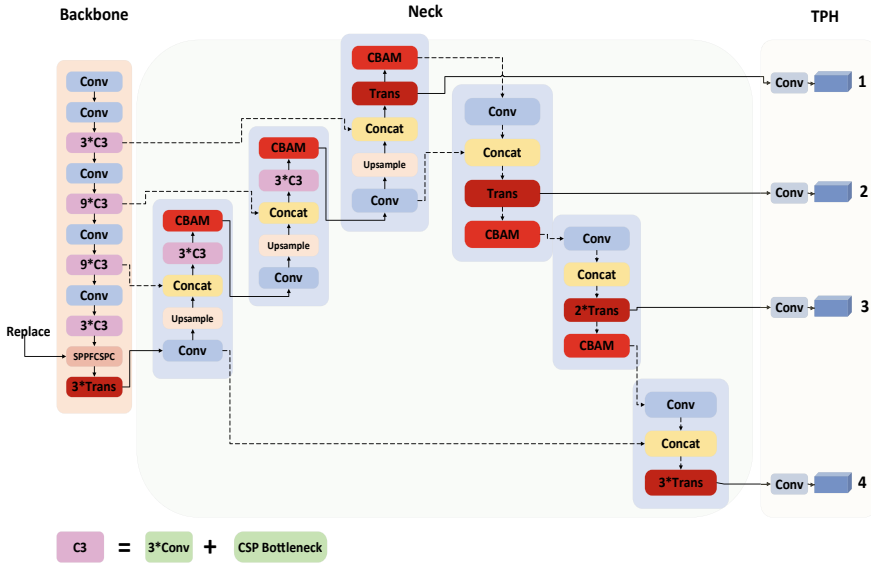
Fig. 3 Improved YOLOV5 structure diagram

efficiently annotated in each image. We randomly split the dataset into two parts: training set and validation set with a ratio of 9:1.

This dataset provides a comprehensive benchmark for us to conduct comparative experiments and to evaluate the robustness of the algorithm under various environmental conditions. We can ensure that the performance of our proposed method is reliable and accurate, helping to develop effective ship detection and recognition algorithms in real marine applications.

3.2 Performance Comparison

To evaluate the effectiveness of our proposed ship remote sensing object recognition method, we compare it with several advanced ship detection algorithms, including SSD [10], YOLOV3 [11], YOLOV4 [12] and the method in YOLOV5 [13]. Evaluations were performed using standard metric Average Precision (mAP) and precision (P). The threshold IoU of mean Average PrecisionIoU is from 0.5 to 0.95, with a step size of 0.05. Table 1 summarizes the experimental results obtained under the same training data. Results show that our proposed method outperforms other methods in terms of AP, indicating its superior performance in accurately detecting and identifying ships in remote sensing images. Figure 4 shows some examples of detection results obtained by different methods, where we mark successfully detected ships in each image with solid red lines. These examples illustrate the effectiveness

Table 1 The average performance of the proposed ship object detection method and SSD, YOLOv3, YOLOv4 and YOLOv5 on the remote sensing image dataset

Method	P	mAP_0.5:0.95
SSD	0.703	0.298
YOLOV3	0.687	0.290
YOLOV4	0.695	0.223
YOLOV5	0.769	0.368
OURS	0.802	0.398

of our proposed method in detecting ships of different sizes, orientations, and sea conditions. Comprehensive evaluations of our proposed method against advanced algorithms and visualization examples of detection results demonstrate its superior performance in ship remote sensing object recognition. Our method can accurately detect and identify ships in remote sensing images, making it a promising solution for real-world maritime applications.

Figure 4 shows the comparison of the detection results of SSD and our proposed method.

Figure 5 shows the comparison of the detection results of YOLOV4 and our proposed method.

Figure 6 shows the comparison of the detection results of YOLOV5 and our proposed method.

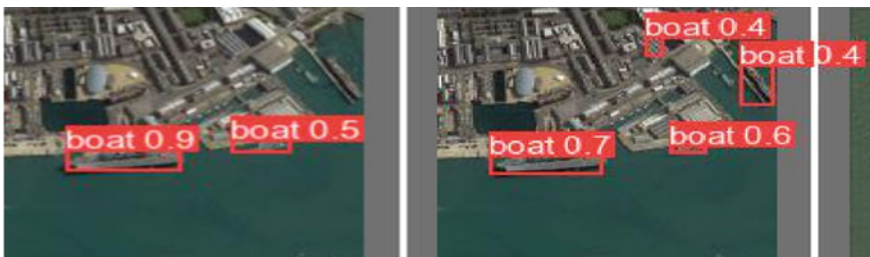**Fig. 4** The detection results of SSD (left) and our method (right)

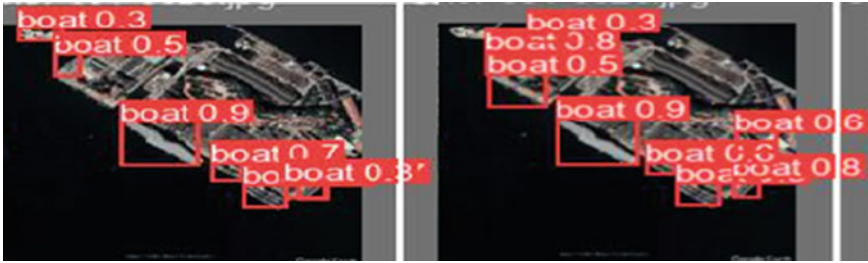**Fig. 5** The detection results of YOLOV4 (left) and our method (right)

Fig. 6 The detection results of YOLOV5 (left) and our method (right)

4 Conclusions

In conclusion, we have proposed a ship remote sensing target recognition approach based on the YOLOv5 object detection framework. The proposed approach replaces the SPPF module with the SPPFCSPC module and integrates Transform Prediction Heads (TPH) into the YOLOv5 network model to improve the accuracy of ship detection in remote sensing imagery. The SPPFCSPC module enhances the feature maps by using a combination of spatial pyramid pooling (SPP) and cross-stage partial connections (CSPC). This leads to an improvement in feature representation and a reduction in computational cost, making the approach more efficient for real-world applications. The integration of TPH into the YOLOv5 network model improves the localization accuracy of ships in remote sensing images by predicting affine transformation parameters for each bounding box. This helps to mitigate the impact of ship rotation and perspective changes on detection accuracy. The experimental results on the ship detection dataset show that our proposed method has obvious accuracy advantages compared with common ship recognition methods. In conclusion, the proposed ship remote sensing object recognition method based on YOLOv5 with SPPFCSPC and TPH modules provides a promising solution for accurate and efficient ship detection in remote sensing images. The method has potential applications in various fields such as maritime surveillance, navigation, and environmental monitoring.

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

References

1. Yi, L., Shoushi, X.: A method for ship target recognition in remote sensing images based on support vector machines. Comput. Simul. **2006**(06), 180–183 (2006)
2. Xia, Y., Wan, S., Yue, L.: A novel algorithm for ship detection based on dynamic fusion model of multi-feature and support vector machine. In: Proceedings of the 2011 Sixth International Conference on Image and Graphics, Hefei, China, pp. 521–526 (2011). https://doi.org/10.1109/ICIG.2011.147
3. Zou, Z., Shi, Z.: Ship detection in spaceborne optical image with SVD networks. IEEE Trans. Geosci. Remote Sens. **54**(10), 5832–5845 (2016). https://doi.org/10.1109/TGRS.2016.2572736
4. Nie, S., Jiang, Z., Zhang, H., Cai, B., Yao, Y.: Inshore ship detection based on mask R-CNN. In: IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, pp. 693–696 (2018). https://doi.org/10.1109/IGARSS.2018.8519123
5. Wei, S., Chen, H., Zhu, X., Zhang, H.: Ship detection in remote sensing image based on faster R-CNN with dilated convolution. In: Proceedings of the 2020 39th Chinese control conference (CCC), Shenyang, China, pp. 7148–7153 (2020). https://doi.org/10.23919/CCC50068.2020.9189467
6. Li, X., Cai, K.: Method research on ship detection in remote sensing image based on Yolo algorithm. In: Proceedings of the 2020 International Conference on Information Science, Parallel and Distributed Systems (ISPDS), Xi'an, China, pp. 104–108 (2020). https://doi.org/10.1109/ISPDS51347.2020.00029
7. Zhang, T., Zhang, X., Ke, X.: Quad-FPN: a novel quad feature pyramid network for SAR ship detection. Remote Sens. **13**(14), 2771 (2021). https://doi.org/10.3390/rs13142771
8. Yang, Y., Pan, Z., Hu, Y., Ding, C.: PistonNet: object separating from background by attention for weakly supervised ship detection. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. **15**, 5190–5202 (2022). https://doi.org/10.1109/JSTARS.2022.3184637
9. Zhu, X., Lyu, S., Wang, X., et al.: TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2778–2788 (2021)
10. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: Ssd: single shot multibox detector. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, pp. 21-37. Springer, New York (2016)
11. Cui, H., Yang, Y., Liu, M., Shi, T., Qi, Q.: Ship detection: an improved YOLOv3 method. In: OCEANS 2019—Marseille, Marseille, France, pp. 1–4 (2019). https://doi.org/10.1109/OCEANSE.2019.8867209
12. Zhou, L.Q., Piao, J.C.: A lightweight YOLOv4 based SAR image ship detection. In: Proceedings of the 2021 IEEE 4th International Conference on Computer and Communication Engineering Technology (CCET), Beijing, China, pp. 28–31 (2021). https://doi.org/10.1109/CCET52649.2021.9544265
13. Fu, Q., Chen, J., Yang, W., Zheng, S.: Nearshore ship detection on SAR image based on Yolov5. In: Proceedings of the 2021 2nd China International SAR Symposium (CISS), Shanghai, China, pp. 1–4 (2021). https://doi.org/10.23919/CISS51089.2021.9652233