



An Efficient Enhanced-YOLOv5 Algorithm for Multi-scale Ship Detection

Jun Li[✉], Guangyu Li[✉], Haobo Jiang[✉], Weili Guo[✉], and Chen Gong

Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China
{jun_li, guangyu.li2017, jiang.hao.bo, wlguo, chen.gong}@njust.edu.cn

Abstract. Ship detection has gained considerable attentions from industry and academia. However, due to the diverse range of ship types and complex marine environments, multi-scale ship detection suffers from great challenges such as low detection accuracy and so on. To solve the above issues, we propose an efficient enhanced-YOLOv5 algorithm for multi-scale ship detection. Specifically, to dynamically extract two-dimensional features, we design a MetaAconC-inspired adaptive spatial-channel attention module for reducing the impact of complex marine environments on large-scale ships. In addition, we construct a gradient-refined bounding box regression module to enhance the sensitivity of loss function gradient and strengthen the feature learning ability, which can relieve the issue of uneven horizontal and vertical features in small-scale ships. Finally, a Taylor expansion-based classification module is established which increases the feedback contribution of gradient by adjusting the first polynomial coefficient vertically, and improves the detection performance of the model on few sample ship objects. Extensive experimental results confirm the effectiveness of the proposed method.

Keywords: Multi-scale Ship Detection · Improved YOLOv5 Network · Attention Module

1 Introduction

Ship detection is a critical aspect of maritime supervision and plays an essential role in intelligent maritime applications such as sea area monitoring, port management, and safe navigation [24]. In recent years, various methods such as foreground segmentation [4, 16, 25], background subtraction [3, 22], and horizon detection [12, 24] have been widely explored and have made considerable progress. However, traditional ship detection methods often lack robustness and may have limited applicability in the presence of complex noise interference.

Supported by the National Science Fund of China under Grant 62006119.

J. Li and G. Li — Equal contributions.

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024
B. Luo et al. (Eds.): ICONIP 2023, LNCS 14452, pp. 252–263, 2024.
https://doi.org/10.1007/978-981-99-8076-5_18

Meanwhile, owing to the development of deep learning in object detection, deep learning-based object detectors have achieved significant advancements. For example, Faster r-cnn [17] is a classic two-stage detection method that employs a region proposal network to generate detection boxes directly. SSD [10] enhances the detection accuracy of multi-scale objects by conducting object detection on multiple feature layers. CenterNet [5] is a detection method that detects the center point and size of an object without anchor. The YOLO [1, 13–15, 23] series are classic single-stage object detection methods that extract multi-scale features via a backbone network and a feature pyramid network, while introducing an anchor frame mechanism to enhance the model's robustness.

Inspired by these deep learning-based detection methods above, there is a growing research efforts towards deep learning-based ship detection. Region proposal network-based methods [7, 9] and regression-based methods [2, 19] have made certain progress. However, various issues, such as false detection and missed detection, persist in ship detection due to factors like the influence of background noise on the sea surface, the uneven distribution of horizontal and vertical features of ships, and the different sizes of ships.

To relieve the issues above, we propose a novel efficient enhanced-YOLOv5 algorithm for multi-scale ship detection. Specifically, in order to mitigate the issue that complex marine environments disrupting large-scale ships, we propose a MetaAconC-inspired dynamic spatial-channel attention module that extracts two-dimensional features, mitigating the environmental impact on large-scale ships. Aiming at the problem of uneven horizontal and vertical features of small-scale ships, we design a gradient-refined bounding box regression module to increase the gradient sensitivity, enhancing the learning ability of the algorithm on small-scale ship features. In order to relieve the challenge that sensitivity of the cross entropy function to class imbalance, we establish a Taylor expansion-based classification module, by adjusting the first polynomial coefficient vertically to increase the contribution feedback of the gradient, improving the detection performance of the model on few sample ship objects. To summarize, our main contributions are as follows:

- We propose a novel efficient enhanced-YOLOv5 algorithm for multi-scale ship detection, where a MetaAconC-inspired dynamic spatial-channel attention module is designed to mitigate the influence of complex marine environments on large-scale ships.
- To mitigate the problem of uneven horizontal and vertical features of small-scale ships, we design an effective gradient-refined bounding box regression module to enhance the learning ability of the algorithm on small-scale ship features.
- To further relieve the challenge that sensitivity to class imbalance, we also construct a Taylor expansion-based classification module to increase feedback contribution and improve the detection performance on few sample ships.

2 Method

2.1 Overall Framework

The structure of an efficient enhanced-YOLOv5 algorithm is shown in Fig. 1. The algorithm comprises several components: a backbone network that extracts features from three different scales. The MetaAconC-inspired dynamic spatial-channel attention module which located in the three feature processing channels behind the backbone network to focus on the feature refinement of multi-scale ships, and a feature pyramid network for feature enhancement. Finally, the detection heads generate the final predictions, and our proposed modules, namely the gradient-refined bounding box regression module and the Taylor expansion-based classification module improve accuracy through gradient calculations and backpropagation during training.

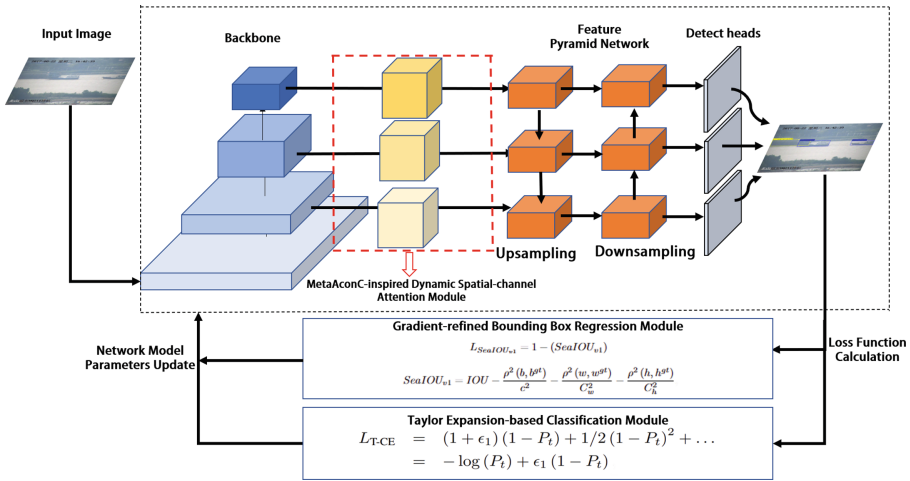


Fig. 1. The pipeline of an efficient enhanced-YOLOv5 algorithm framework.

2.2 MetaAconC-Inspired Dynamic Spatial-Channel Attention Module

Due to the large span of large-scale ships in the image, its learned feature distribution tends to be largely split, which may potentially confuse the object semantics, thereby presenting limited detection accuracy. Especially in the complex marine environments, the semantic information of ships is easily polluted by background noise, which makes it difficult to learn. To mitigate the influence of complex marine environments on large-scale ships, we propose a MetaAconC-inspired dynamic spatial-channel attention module as shown in Fig. 2. In detail,

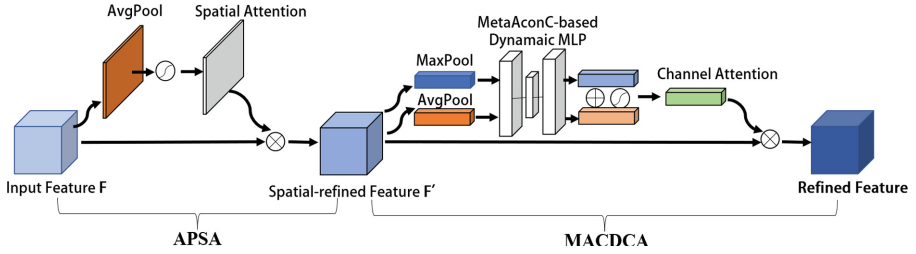


Fig. 2. The overview of the MetaAconC-inspired dynamic spatial-channel attention module. APSA denotes the average pooling-based spatial attention module. MACDCA denotes the MetaAconC-inspired dynamic channel attention module.

the average pooling-based spatial attention module obtains the intra-channel relationship of input features. Secondly, the MetaAconC-inspired dynamic channel attention module dynamically summarize the spatial relationships of features. As such, our module effectively learns the multi-dimensional features information of ships and the impact of complex marine environments on the noise of large ships is mitigated.

Average Pooling-Based Spatial Attention Module. The module integrates ship characteristic information between different channels, and further eliminates the negative impact of complex marine environment on large-scale ships through the similar semantic characteristics of background noise in channel dimensions. After obtaining the feature $F \in R^{H \times W \times C}$ through the CSPDarkNet53 backbone network from the input image, we input F into the average pooling-based spatial attention module to obtain global information by utilizing global averaging pooling of channel dimensions, followed by sigmoid function to produce spatial-refined attention weight $\in R^{H \times W \times 1}$, which is then multiplied with the input feature F to obtain spatial-refined feature F' , which is fed into the next module.

MetaAconC-Inspired Dynamic Channel Attention Module. Since background noise is not invariable in spatial dimension, and a variety of unnecessary noise will be formed in complex marine environment, we designed the module to dynamically adjust attention mode, better learn ship characteristics, and effectively reduce the interference of dynamic background noise. This module conducts global average pooling and maximum pooling of spatial dimensions to F' , and add the results through a two-layer neural network based on the MetaACON function [11] and sigmoid activation function to obtain channel-refined attention weight $\in R^{1 \times 1 \times C}$. Finally, we multiply this weight with feature F' to obtain refined feature. The smooth maximum function has been utilized to expand the Maxout function, resulting in the Acon series activation functions. The MetaAcon function allows the adaptive activation of neurons through the modification

of a parameter, denoted by γ , which is defined as follows:

$$f_{(x)} = (p_1 - p_2)x \cdot \sigma(\gamma(p_1 - p_2)x) + p_2x, \quad (1)$$

where x represents the input, and σ is the sigmoid function. p_1 and p_2 are two channel-wise learnable parameters. The channel-wise parameter γ dynamically adjusts the activation of neurons through convolution operations, controlling whether they should be activated or not. The formula for γ is given by:

$$\gamma = \sigma W_1 W_2 \sum_{h=1}^H \sum_{w=1}^W x_{c,h,w}, \quad (2)$$

where W_1 and W_2 represent two 1×1 convolution layers.

2.3 Gradient-Refined Bounding Box Regression Module

The CIOU loss [27] is a widely used bounding box regression loss, which plays a crucial role in the YOLOv5 algorithm. However, CIOU loss has two main drawbacks in correspondence learning. (i) First, the current approach only takes into account the aspect ratio of the bounding box, without considering the actual height and width of the object. Ships are not all regular rectangles, and the aspect ratio of different ship types varies greatly. For example, the shape of the fishing boat is very slender, small in height but large in width. However, in order to better accommodate tourists, ships such as passenger ships and cruise ships are very tall compared to their width. As a consequence, the differences in aspect ratios of ships can hinder the accurate fitting of ships with varying shapes especially small-scale ships, leading to misidentification and missed detections. (ii) Second, the loss function gradient remains constant, which renders the model insensitive to fitting multi-scale objects, making small-scale ship detection more challenging.

To mitigate the issue (i), we divide the aspect ratio into height and width, and calculate them respectively [26]. In this way, the fitting direction of the regression module is closer to the shape of the ship. The width-height loss directly minimizes the width-height difference between the target box and the bounding box so that the model can better fit the ships with different shapes, which is defined as follows:

$$L_{SeaIOU_{v1}} = 1 - (SeaIOU_{v1}), \quad (3)$$

where $SeaIOU_{v1}$ is defined as:

$$SeaIOU_{v1} = IOU - \frac{\rho^2(b, b^{gt})}{c^2} - \frac{\rho^2(w, w^{gt})}{C_w^2} - \frac{\rho^2(h, h^{gt})}{C_h^2}, \quad (4)$$

where b and b^{gt} represent the center points of the bounding box and target box, respectively. $\rho(\cdot)$ represents the Euclidean distance. c represents the area of the smallest enclosing box that covers both boxes. C_w and C_h are the width and height of the minimum circumscribed frame that covers both boxes.

To mitigate the issue (ii), we establish a gradient-refined bounding box regression module that increases the gradient sensitivity of the loss function. Specifically, we modify the invariance of the gradient by applying a logarithmic function. The absolute gradient value decreases with the increase of the overlap, which is more favorable for bounding box regression. As such, when the distance between the boxes is far away, its gradient absolute value is larger, which is more conducive to the detection of small-scale ships. This approach enhances the contribution of small-scale ships to the feature learning ability of the model. The formula for the modified loss function is defined as:

$$L_{SeaIOU_{v_2}} = \alpha \cdot \ln \alpha - \alpha \cdot \ln(\beta + (SeaIOU_{v_1})), \tag{5}$$

where α and β represent parameters that control the gradient sensitivity of the loss function.

2.4 Taylor Expansion-Based Classification Module

The cross entropy loss is a popular classification loss, which plays a crucial role in the YOLOv5 algorithm, which is defined as:

$$L_{CE} = -\log(P_t) = \sum_{j=1}^{\infty} 1/j (1 - P_t)^j = (1 - P_t) + 1/2(1 - P_t)^2 \dots, \tag{6}$$

where P_t is the model’s prediction probability of the ground-truth class.

However, it is sensitive to class imbalance. The cross-entropy loss assumes that the classes are balanced, which may result in the model becoming biased towards the majority class and failing to capture the features of the minority class. Specifically, In the training process, it back-propagates each type of ship according to the same contribution, making the model more inclined to learn the ship object with a large number of samples. However, the learning efficiency of the ship object with a few sample is very low, which greatly limits the detection performance of ships with few samples. In the application of ship detection, the sample number of ships is very uneven. Some ship types are very common, during training, more samples can be provided for the model to learn features and improve the detection performance. However, some ship types are not as common as the above ships, and the number of their samples is very small. It is difficult for the ship detection model to get enough learning samples in the training stage, so it is difficult to learn the characteristics of ships with few samples. Expanding datasets is a feasible approach, but it costs a lot. Therefore, it is necessary to optimize the training strategy.

To mitigate the issue, we establish a Taylor expansion-based classification module which presents the loss function as a linear combination of polynomial function. We get its gradient formula based on the cross entropy loss function, which is shown as:

$$-\frac{dL_{CE}}{dP_t} = \sum_{j=1}^{\infty} (1 - P_t)^{j-1} = 1 + (1 - P_t) + (1 - P_t)^2 \dots \tag{7}$$

From the above formula, it can be seen that the first term of the cross entropy loss function is the largest, which is 1. The subsequent terms are smaller and smaller, which means that the first term contributes the most to the gradient gain. By adjusting the first polynomial coefficient vertically [8], we increase the feedback contribution of cross-entropy gradient. This module further strengthens the fitting ability and alleviates the sensitivity to class imbalance, which is defined as:

$$L_{T-CE} = (1 + \epsilon_1)(1 - P_t) + 1/2(1 - P_t)^2 + \dots = -\log(P_t) + \epsilon_1(1 - P_t), \quad (8)$$

where ϵ_1 represents the parameter we adjusted in the first polynomial coefficient.

In this way, the sensitivity of the classification module to the number of samples is improved, the problem of low gradient gain of few sample ships is alleviated, and the detection performance of the model for few sample ship templates is enhanced

3 Experiments

3.1 Experimental Settings

Dataset. In this paper, we evaluate the performance of the proposed method on the SeaShips dataset [20], a well-known large-scale and precisely annotated maritime surveillance dataset released by Wuhan University in 2018. The dataset collected by the coastal land-based camera in Hengqin, Zhuhai, including 6 types of ships with different sizes, contains 31,455 images, 7,000 of which are publicly available. We divide the pictures according to the official scale. The training set and the validation set are 1750, and the remaining 3500 are used as the test set. The detection difficulties include ship size change, complex background interference and so on. In this dataset, the size of the fishing boat object is small, the sample size of the passenger ship is small, thus the detection accuracy of the algorithm for them is one of the main indicators to verify the performance of the model to small-scale ship object and few sample ship object.

Evaluation Indicators. We adopt evaluation indicators of COCO dataset, including $mAP_{0.5}$, $AP_{0.5}$, $mAP_{0.75}$, and $AP_{0.75}$. AP (Average Precision) is the area enclosed by the X-axis and Y-axis plots using *Recall* and *Precision* respectively. $AP_{0.5}$ and $AP_{0.75}$ are AP s at IoU threshold of 0.5 and 0.75, respectively. For multi-object detection, each object would have an AP value first, and then take the weighted average to obtain mAP (Mean Average Precision).

Implementation Details. For our experiments, one GeForce RTX 2080ti GPU card is used, and the CUDA version is 10.0. The cuDNN version is 7.5.1, and the PyTorch version is 1.2.0. All models are trained for 300 epochs with batch size of 4, an initial learning rate of 1e-2, which is then reduced to a minimum of 1e-4 using a cosine annealing algorithm. We utilize the *sgd* optimizer with

momentum 0.937 and weight decay 5e-4. All models are deployed according to the above Settings. YOLOv5 network is the original network of our method. We set $\alpha = 5$, $\beta = 4$ and $\epsilon_1 = 1$. In order to demonstrate the efficacy of the proposed method, we conduct an experimental comparison with the other conventional object detection methods on the Seaships dataset.

Table 1. Detection results on the Seaships dataset. It shows $mAP_{0.5}$ and $AP_{0.5}$ in each class. The bold number has the highest score in each column.

Model	$mAP_{0.5}$	Bulk cargo carrier	Container ship	Fishing boat	General cargo ship	Ore carrier	Passenger ship
Faster r-cnn ¹	0.949	0.958	0.994	0.906	0.966	0.950	0.917
Faster r-cnn ²	0.946	0.927	0.990	0.917	0.969	0.938	0.933
SSD300 ²	0.935	0.949	0.987	0.888	0.962	0.930	0.893
SSD300 ³	0.891	0.918	0.967	0.809	0.925	0.898	0.831
YOLOv3	0.941	0.952	0.983	0.923	0.968	0.943	0.878
YOLOv4	0.921	0.901	0.975	0.901	0.937	0.918	0.894
Shao	0.874	0.876	0.903	0.783	0.917	0.881	0.886
YOLOv5	0.952	0.953	0.988	0.940	0.974	0.935	0.922
Ours	0.966	0.961	0.991	0.956	0.982	0.951	0.952

[1] ResNet50 [6] is selected as the backbone network.

[2] VGG16 [21] is selected as the backbone network.

[3] MobileNetv2 [18] is selected as the backbone network.

Table 2. Detection results on the Seaships dataset. It shows $mAP_{0.75}$ and $AP_{0.75}$ in each class. The bold number has the highest score in each column.

Model	$mAP_{0.75}$	Bulk cargo carrier	Container ship	Fishing boat	General cargo ship	Ore carrier	Passenger ship
Faster r-cnn ¹	0.658	0.608	0.806	0.553	0.767	0.576	0.636
Faster r-cnn ²	0.650	0.537	0.852	0.582	0.731	0.569	0.629
SSD300 ²	0.673	0.704	0.903	0.509	0.789	0.609	0.525
SSD300 ³	0.491	0.499	0.745	0.286	0.594	0.445	0.373
YOLOv3	0.631	0.612	0.860	0.474	0.727	0.647	0.470
YOLOv4	0.506	0.487	0.673	0.360	0.579	0.472	0.467
YOLOv5	0.762	0.769	0.920	0.669	0.850	0.721	0.646
Ours	0.785	0.788	0.940	0.660	0.848	0.706	0.767

[1] ResNet50 [6] is selected as the backbone network.

[2] VGG16 [21] is selected as the backbone network.

[3] MobileNetv2 [18] is selected as the backbone network.

3.2 Quantitative Analysis

As shown in Table 1, we conduct an experimental comparison of $mAP_{0.5}$ and $AP_{0.5}$ with the other eight classical object detection methods on the Seaships dataset. The proposed method achieves a high $mAP_{0.5}$ of 96.6%, with the 3 ship classes having the highest AP values. In particular, for passenger ship with a smaller sample, $AP_{0.5}$ reaches 95.2%, an improvement of 3% over the original network. In addition, For small-scale fishing boat, $AP_{0.5}$ reaches 95.6%, an increase of 1.6% over the original network. Compared to Faster r-cnn [17] with various backbone networks, our proposed method alleviates the interference of complex environment by adding the proposed attention module, with $mAP_{0.5}$

increasing by 1.7% and 1.9%. Compared to SSD [10] with various backbone networks, our proposed method further enhance multi-scale features, with $mAP_{0.5}$ increasing by 3.1% and 7.5%. Specifically, for fishing boat, $AP_{0.5}$ increases by 6.8% and 14.7%. Compared to the YOLO series networks [1, 15], our proposed method improves the feature description power of the model for multi-scale ships and achieves higher detection accuracy. Compared to Shao [19], our proposed method increases $mAP_{0.5}$ by 9.2% by reducing the complex environment interference and sample imbalance sensitivity with the proposed regression and classification module. Particularly for fishing boat and container ship, $AP_{0.5}$ increases by 17.3% and 8.8%.

In order to further verify the performance of our proposed model more strictly, we experimentally compare $mAP_{0.75}$ and $AP_{0.75}$ with five other classical object detection methods on the Seaships dataset. Table 2 presents the performance of different methods on Seaships, our proposed method also achieves the highest detection performance of 78.5%, an improvement of 2.3% over the original network. It's worth noting that passenger ship with fewer samples, $AP_{0.75}$ reaches 76.7%, an improvement of 12.1% over the original network. In conclusion, our proposed method is more effective than other classical methods for improving the accuracy of multi-scale ship detection.

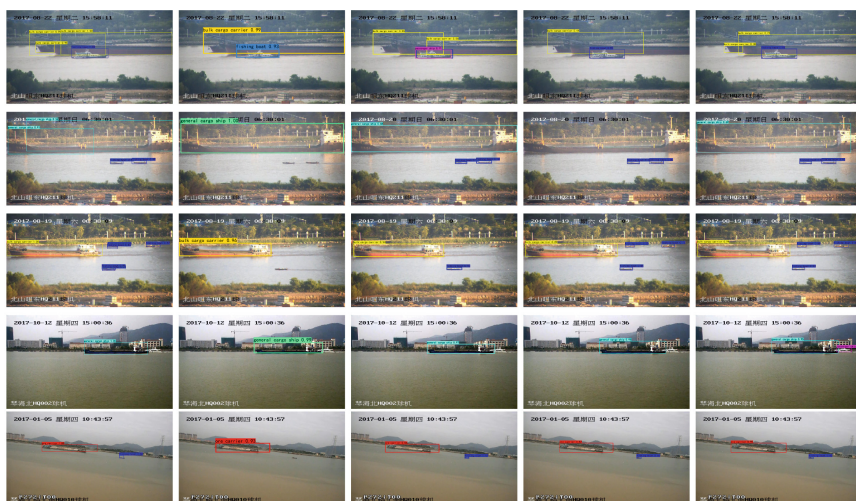
Table 3. Ablation experimental results of module on seaships Dataset.

Model	$mAP_{0.5}$	Bulk cargo carrier	Container ship	Fishing boat	General cargo ship	Ore carrier	Passenger ship
YOLOv5	0.952	0.953	0.988	0.940	0.974	0.935	0.922
+ <i>b</i>	0.954	0.944	0.988	0.948	0.977	0.943	0.921
+ <i>b</i> + <i>c</i>	0.963	0.958	0.993	0.956	0.976	0.948	0.950
+ <i>a</i> + <i>b</i> + <i>c</i> (ours)	0.966	0.961	0.991	0.956	0.982	0.951	0.952

3.3 Ablation Studies

Table 3 displays the effect of the three proposed modules on the performance of the method. To ensure fair comparison, we use the same experimental setup for all the methods. *a* represents the MetaAconC-inspired dynamic spatial-channel attention module, *b* represents the gradient-refined bounding box regression module and *c* represents the Taylor expansion-based classification module.

The original network YOLOv5 achieves the $mAP_{0.5}$ of 95.2%. After *b* is added, The method improves small-scale ship detection performance by increasing gradient sensitivity, resulting in an $AP_{0.5}$ increase of 0.8% for fishing boat. Then the method enhances accuracy further by adding *c* to reduce class imbalance sensitivity, yielding an overall $mAP_{0.5}$ improvement of 0.9%, and $AP_{0.5}$ improved by 2.9% for passenger ships with fewer samples. After adding *a*, by focusing on the extraction of ship characteristics, the influence of complex Marine environment is weakened. Our method combined with the proposed attention module raises the $mAP_{0.5}$ to 96.6%, 1.4% higher than the original network YOLOv5. Experimental results show that our modules significantly improve ship detection performance across different sizes and ship types.



(a) Faster r-cnn (b) SSD300 (c) YOLOv4 (d) YOLOv5 (e) Ours

Fig. 3. Qualitative comparison of different methods on Seaships.

3.4 Qualitative Analysis

Figure 3 illustrates the ship detection performance of our proposed method and the other classical methods under various complex conditions. From the first line, it can be seen that in the occlusion case, Faste r-cnn gets a duplicate bounding boxes due to the region proposal network. SSD300, YOLOv4 and YOLOv5 all miss the bulk cargo carrier that is hiding from each other. And from the fourth line, except our method, the other object detectors do not detect the obscured passenger ship. As can be seen from the second and third lines, when multi-scale ships exist at the same time, Faste r-cnn also produces redundant detection boxes. SSD and YOLOv4 fail to detect small fishing ships. Our original network YOLOv5 can not handle the detection of multi-scale ships well, resulting in the detection of small ships, while missing the detection of large ships across the whole map. By adding the proposed attention module, our proposed method alleviates the problem of semantic information fragmentation of large ships and detects these ships well. As can be seen from the fifth line, for the small ship object scenario, the position of the detection box of Faste r-cnn is offset and SSD failes to detect the small-scale fishing ship. It can be concluded that our proposed method effortlessly handles these situations with ease.

4 Conclusion

In this paper, we have proposed an efficient enhanced-YOLOv5 algorithm for multi-scale ship detection. Our approach consists of three components, specifically a metaAconC-inspired dynamic spatial-channel attention module has been

designed to reduce the impact of complex marine environments on large-scale ships. Also, We have mitigated the issue of uneven horizontal and vertical features of small-scale ships by constructing a gradient-refined bounding box regression module. Moreover, we have proposed a Taylor expansion-based classification module to alleviate the sensitivity to class imbalance and improve the detection performance to few sample ships. The experimental results demonstrate the effectiveness of our proposed method. In future work, our model should further improve its ability to detect small-scale ships in complex marine environments.

Acknowledgement. This work is supported by the National Science Fund of China under Grant 62006119.

References

1. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: Yolov4: Optimal speed and accuracy of object detection. arXiv preprint [arXiv:2004.10934](https://arxiv.org/abs/2004.10934) (2020)
2. Chen, J., Xie, F., Lu, Y., Jiang, Z.: Finding arbitrary-oriented ships from remote sensing images using corner detection. *IEEE Geosci. Remote Sens. Lett.* **17**(10), 1712–1716 (2019)
3. Chen, Z., Yang, J., Kang, Z.: Moving ship detection algorithm based on gaussian mixture model. In: 2018 3rd International Conference on Modelling, Simulation and Applied Mathematics (MSAM 2018), pp. 197–201. Atlantis Press (2018)
4. Dong, C., Feng, J., Tian, L., Zheng, B.: Rapid ship detection based on gradient texture features and multilayer perceptron. *Infrared Laser Eng.* **48**(10), 1026004–1026004 (2019)
5. Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., Tian, Q.: CenterNet: keypoint triplets for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6569–6578 (2019)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
7. Kim, K., Hong, S., Choi, B., Kim, E.: Probabilistic ship detection and classification using deep learning. *Appl. Sci.* **8**(6), 936 (2018)
8. Leng, Z., Tan, M., Liu, C., Cubuk, E.D., Shi, X., Cheng, S., Anguelov, D.: PolyLoss: a polynomial expansion perspective of classification loss functions. arXiv preprint [arXiv:2204.12511](https://arxiv.org/abs/2204.12511) (2022)
9. Li, Q., Mou, L., Liu, Q., Wang, Y., Zhu, X.X.: HSF-NET: multiscale deep feature embedding for ship detection in optical remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **56**(12), 7147–7161 (2018)
10. Liu, W., et al.: SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
11. Ma, N., Zhang, X., Liu, M., Sun, J.: Activate or not: Learning customized activation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8032–8042 (2021)
12. Prasad, D.K., Prasath, C.K., Rajan, D., Rachmawati, L., Rajabaly, E., Quek, C.: Challenges n video based object detection in maritime scenario using computer vision. arXiv preprint [arXiv:1608.01079](https://arxiv.org/abs/1608.01079) abs/1608.01079 (2016)

13. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788 (2016)
14. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6517–6525 (2017)
15. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement. arXiv preprint [arXiv:1804.02767](https://arxiv.org/abs/1804.02767) (2018)
16. Ren, L., Ran, X., Peng, J., Shi, C.: Saliency detection for small maritime target using singular value decomposition of amplitude spectrum. IETE Tech. Rev. **34**(6), 631–641 (2017)
17. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. Adv. Neural. Inf. Process. Syst. **28**, 91–99 (2015)
18. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520 (2018)
19. Shao, Z., Wang, L., Wang, Z., Du, W., Wu, W.: Saliency-aware convolution neural network for ship detection in surveillance video. IEEE Trans. Circuits Syst. Video Technol. **30**(3), 781–794 (2019)
20. Shao, Z., Wu, W., Wang, Z., Du, W., Li, C.: SeaShips: a large-scale precisely annotated dataset for ship detection. IEEE Trans. Multimedia **20**(10), 2593–2604 (2018)
21. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
22. Wang, B., Dong, L., Zhao, M., Xu, W.: Fast infrared maritime target detection: binarization via histogram curve transformation. Infrared Phys. Technol. **83**, 32–44 (2017)
23. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint [arXiv:2207.02696](https://arxiv.org/abs/2207.02696) (2022)
24. Ye, C., Lu, T., Xiao, Y., Lu, H., Qunhui, Y.: Maritime surveillance videos based ships detection algorithms: a survey. J. Image Graphics **27**, 2078–2093 (2022)
25. Zhang, Y., Li, Q.Z., Zang, F.: Ship detection for visual maritime surveillance from non-stationary platforms. Ocean Eng. **141**, 53–63 (2017)
26. Zhang, Y.F., Ren, W., Zhang, Z., Jia, Z., Wang, L., Tan, T.: Focal and efficient IOU loss for accurate bounding box regression. Neurocomputing **506**, 146–157 (2022)
27. Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D.: Distance-IOU loss: faster and better learning for bounding box regression. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 12993–13000 (2020)