Ship Detection in Synthetic Aperture Radar Images with Improved YOLOv8

Chushi Yu
School of Electronic Engineering
Soongsil University
Seoul 06978, Korea
Email: csyu@soongsil.ac.kr

Yoan Shin*
School of Electronic Engineering
Soongsil University
Seoul 06978, Korea
Email: yashin@ssu.ac.kr

***Abstract*—In recent years, ship detection using remote sensing images has emerged as a crucial task for coastal countries due to advancements in remote sensing technology. Synthetic aperture radar (SAR) is a prominent active imaging sensor, unaffected by clouds and capable of day-night operation. However, SAR images pose challenges such as unclear contours, complex backgrounds, and strong scattering, leading to the misdetection and missed detection of small ship targets. To address these issues, this paper proposes an improved ship detection model for SAR images based on the YOLOv8 framework. Our approach introduces a small target detection layer to the original YOLOv8 architecture and adapt the loss function using the wise-intersection over union. Experimental evaluations conducted on the HRSID and the SSDD datasets demonstrate the effectiveness of our method, improving detection accuracy, recall, and robustness in complex marine environments.**

Keywords—Synthetic aperture radar, deep learning, ship detection, YOLOv8, wise-intersection over union

I. INTRODUCTION

Synthetic aperture radar (SAR) is a microwave imaging sensor that can actively detect in all-day and all-weather conditions. It has very good applicability for monitoring oceans with changing climates [1]. In the task of detecting ships, SAR offers significant advantages as it remains unaffected by variable ocean weather, and enables real-time monitoring of ship targets from all directions [2].

Deep learning techniques have revolutionized various remote sensing applications, including object and oil spill detection, traffic monitoring, terrain mapping, coastline monitoring, and marine fisheries management, to name a few. Object detection, in particular, has garnered significant attention in remote sensing research. Deep learning-based target detection algorithms can be categorized into two-stage algorithms like faster region-based convolutional neural network (Faster R-CNN) [3] and single-stage algorithms like RetinaNet [4] and you only look once (YOLO) [5]. While YOLO exhibits fast speed and high accuracy compared to other deep learning algorithms, detecting small objects remains still challenging. Researchers are working to improve the detection accuracy of small objects while maintaining speed. The YOLOv8 [6] is the latest version of YOLO model for object detection, image classification, and instance segmentation tasks.

The objective function, particularly the bounding box regression loss, plays an essential role in deep learning-based detection models. A well-defined objective function can lead to significant improvements. Addressing the problem of balancing the bounding box regression for high- and low-quality examples is of utmost importance. In a previous study [7], Tong *et al.* proposed wise-intersection over union (WIoU), an IoU-based loss function incorporating a dynamic non-monotonic focusing mechanism (FM), which was validated on the YOLOv7 model [8]. The dynamic non-monotonic FM utilizes the outlier degree instead of IoU to evaluate the quality of anchor boxes, and provides a wise gradient gain allocation strategy. This strategy reduces the competitiveness of high-quality anchor boxes, while reducing the harmful gradient generated by low-quality anchor boxes and improve the detector's overall performance.

This paper focuses on improving SAR ship detection using an enhanced YOLOv8 framework. We introduce a detection layer for small targets on the basis of original YOLOv8 and modify the loss function by using the WIoU. The effectiveness and applicability of our approach are evaluated using the high resolution SAR images dataset (HRSID) [9] and the SAR ship detection dataset (SSDD) [10] in a comparative study.

II. METHODOLOGY

A. Overview of the Proposed Method

The YOLOv8 model is an enhanced version of the influential YOLOv5 model [11]. It introduces the C2f module which is a cross-stage partial bottleneck [12] comprising two convolutions that effectively combines high-level features with contextual information, leading to improved detection accuracy. Additionally, a modified version of spatial pyramid pooling [13], known as spatial pyramid pooling fast module, is incorporated at the end of the backbone. In the neck part, the feature fusion method employed is the path aggregation network-feature pyramid networks [14, 15], enhancing the fusion and utilization of feature layer information at different scales. The decoupled head is used in the last part of the neck, combining the classification and regression branches. Notably, YOLOv8 is an anchor-free model that identifies the object's center and estimates the distance between the center and the bounding box, as opposed to the anchor-based YOLOv5 and YOLOv7 models.

To address the issue of learning feature information for small targets in YOLOv8, an additional small target detection layer is proposed. This layer combines shallow and deep feature maps and increases the number of detection heads from 3 to 4. By incorporating this layer, the network focus more on detecting small targets and thereby enhance the overall detection performance. This modification aims to overcome the limitations imposed by large down sampling factor in the original YOLOv8 architecture.

The proposed improvement on YOLOv8-based architecture comprises three main components: backbone, neck, and head, as illustrated in Fig. 1.

*Corresponding author
This work was supported in part by the NRF grant funded by the Korea government (MSIT) (No. RS-2023-00251595), and by the MSIT, Korea, under the ITRC support program (IITP-2023-2018-0-01424) supervised by the IITP.

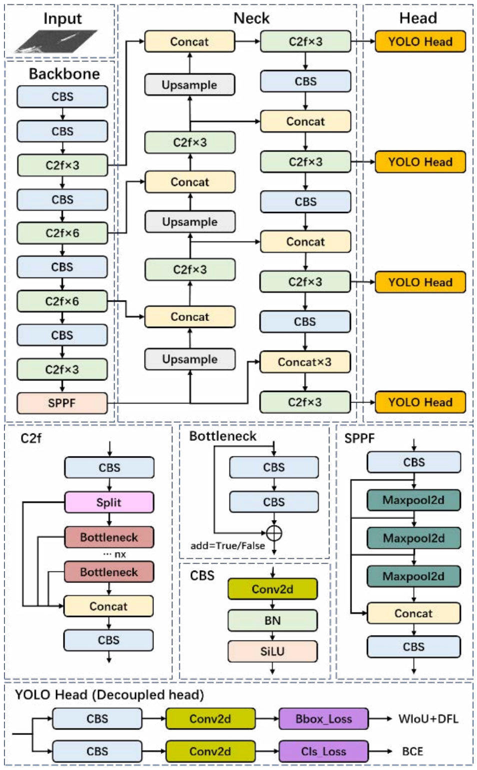

Fig. 1. Detailed illustration of the proposed model architececure.

B. Loss Function

In the original YOLOv8, the classification is based on the binary cross-entropy (BCE) and the bounding box loss is based on the complete IoU (CIoU) with the distribution focal loss (DFL) [16]. In our proposed approach, the WIoU loss function [7] is used as the bounding box regression loss. The modified loss function is then expressed as follows.

$$Loss = L_{cls} + L_{reg} = L_{BCE} + L_{DFL} + L_{WIoU}. \quad (1)$$

Here, the BCE loss is given as

$$L_{BCE} = -w[y_i \log(x_i) + (1-y_i)\log(1-x_i)], \quad (2)$$

where w is the weight, y_i is the labeled value, and x_i is the predicted value of the model.

DFL is an optimized version of the focal loss function that extends the discrete classification results to continuous ones, leading to improved performance.

$$L_{DFL(S_i,S_{i+1})} = -((y_{i+1}-y)\log(S_i) + (y-y_i)\log(S_{i+1})), \quad (3)$$

where y_i, y_{i+1} represent the values from the left and right sides near the consecutive labels y, satisfying $y_i < y < y_{i+1}$, $y = \sum_{i=0}^{n} P(y_i)y_i$. Here, P can be implemented through a softmax $S(\cdot)$ layer, with $P(y_i)$ being denoted as S_i for simplicity. $S_i = \frac{y_{i+1}-y}{y_{i+1}-y_i}$, $S_{i+1} = \frac{y-y_i}{y_{i+1}-y_i}$.

The WIoU loss adjusts based on the labeling quality of the training data. It uses a dynamic non-monotonic focusing mechanism to evaluate the anchor frame quality and avoids excessive penalties on geometric factors. When the predicted box closely matches the target box, the loss function promotes better generalization with less training intervention by reducing penalties on geometric factors. The expression for WIoU can be described as

$$L_{WIoU} = rL_{IoU}R_{WIoU} = r(1-IoU)R_{WIoU}$$
$$= \frac{\beta}{\delta\alpha^{\beta-\delta}}\left(1-\frac{W_iH_i}{S_u}\right)\exp\left(\frac{(x-x_{gt})^2+(y-y_{gt})^2}{(W_g^2+H_g^2)^*}\right), \quad (4)$$

where β indicates the degree of abnormality of the predicted box, and a smaller degree of abnormality means that the quality of the anchor box is higher. Therefore, using β to construct a non-monotonic focal number can assign small gradient gains to prediction boxes with large outliers, effectively reducing harmful gradients of low-quality training samples. Also, α and δ are hyper-parameters, and x and y represent the coordinate values of the prediction box, while x_{gt} and y_{gt} represent the coordinate values of the ground truth. W_g, H_g are the size of the smallest enclosing box, the superscript $*$ indicated the operation that detach W_g, H_g from the computational graph in order to prevent R_{WIoU} from producing the gradient that hinders convergence. The corresponding W and H values represent the width and height of the two boxes, respectively. We also have $S_u = wh + w_{gt}h_{gt} - W_iH_i$, where W_i and H_i are the width and height of the overlapping region of the anchor box and ground truth box.

III. Experimental Results

A. Dataset and Training Strategy

We use the HRSID [9] and the SSDD [10] to verify the performance of the proposed method. The HRSID is widely used for ship detection and instance segmentation. It contains 5,604 SAR image slices and 16,591 ship targets. We divide images into a training set, a validation set and a test set as 7:1:2. All the experiments are implemented on PyTorch. The batch size is set to 8 and the number of training epochs to 300. The stochastic gradient descent optimizer is utilized with an initial learning rate of 0.01 and a momentum of 0.9.

B. Results and Discussion

In this experiment, the precision (P), the recall (R) and the mean average precision (mAP) are used to evaluate the detection performance of the models.

Table 1 lists the experimental results on the HRSID with YOLOv8 and several improved models. In addition, the comparative experiments with state-of-the-art methods are conducted based on the YOLOv5 [11] and YOLOv7 [8, 17]. The evaluation metrics demonstrate that the improved method outperforms the baseline YOLOv5 and YOLOv7 methods in terms of accurate ship detection and real-time performance. The precision and recall are 0.5% and 0.4% higher than the

TABLE I. Results of Different Methods on the HRSID

Models	P	R	mAP @.5	mAP @.5:.95	Model size(MB)	Speed (sec)
YOLOv5	0.923	0.865	0.929	0.675	5.0	0.017
YOLOv7	0.844	0.700	0.786	0.481	74.7	0.279
YOLOv8	0.934	0.873	0.934	0.713	6.3	0.014
YOLOv8 +SmallHead	0.922	0.874	0.922	0.684	6.5	0.017
YOLOv8 +WIoU	0.938	0.857	0.923	0.686	6.3	0.014
Improved YOLOv8	0.939	0.877	0.936	0.709	6.5	0.016

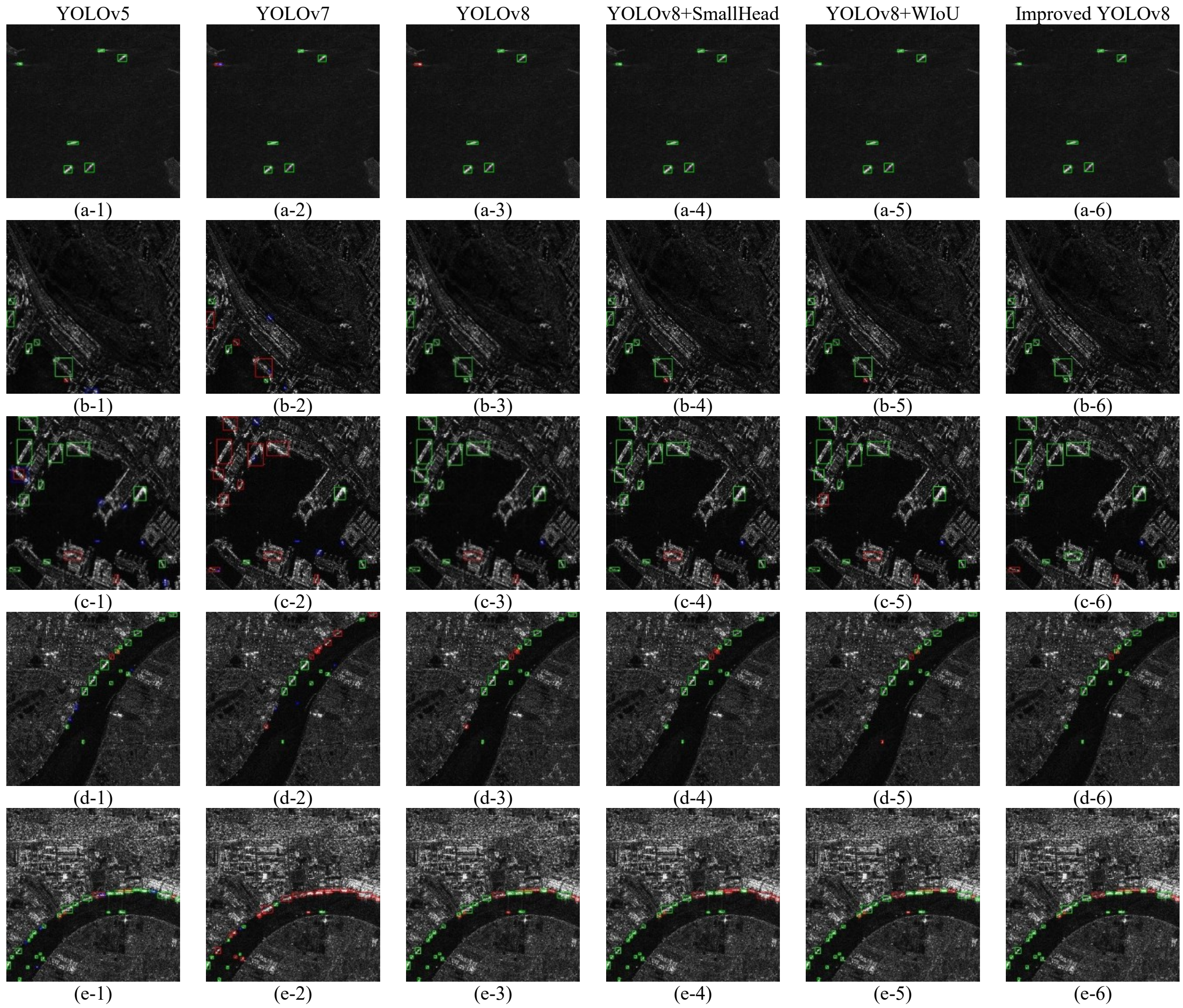

Fig. 2. Comparison of the qualitative ship detection results for the improved YOLOv8 and other methods in the HRSID. The green boxes represent correctly-detected ships, red boxes indicate missing ships, and blue boxes denote false alarms. (a), (b-c), (d-e) illustrate the detection results of offshore ships, inshore ships and densely distributed small-scale ships, respectively.

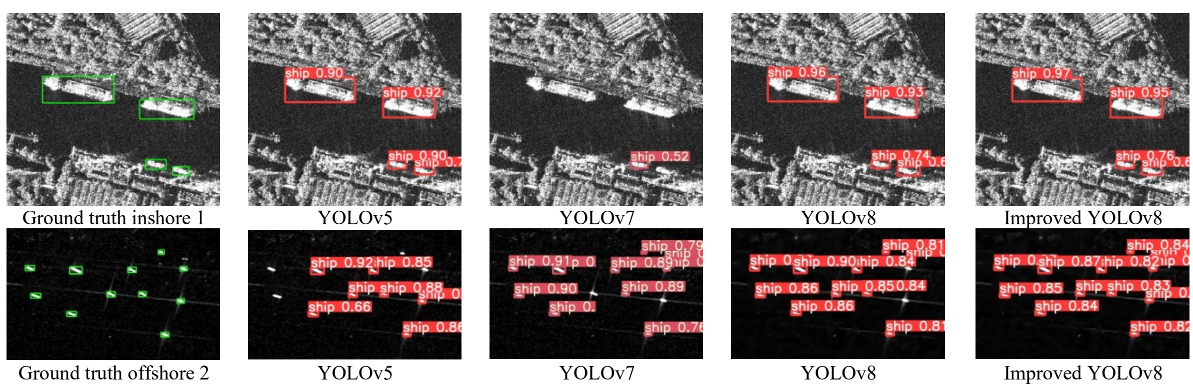

Fig. 3. Comparative ship detection results of the improved YOLOv8 and other state-of-the-art methods in the SSDD.

original YOLOv8, respectively, while the mAP are comparable to the original method according to different thresholds. Figure 2 showcases several qualitative results for visualizing the detection performance of the proposed scheme. The green, red, and blue boxes indicate the correctly-detected ships, the missing ships and false alarms. To assess the feasibility of the proposed method, the trained model and parameters is used to perform ship detection on the SSDD [10].

Figure 3 presents inshore and offshore ship detection sample results which are obtained by the improved YOLOv8 as well as other state-of-the-art methods like YOLOv5 and YOLOv7. Experimental results show that the proposed method is capable of recognizing most small-scale ships even without specific training, although there may be few missed detections and false alarms when ships are densely stacked.

IV. Conclusion

In this study, a YOLOv8-based SAR ship detection scheme is proposed after analyzing the existing state-of-the-art target detection algorithms. Our approach involved adding a small target detection layer to the original YOLOv8 and modifying the loss function using the WIoU, enabling fast and accurate ship target detection in complex marine environments. Different improvement of YOLOv8 models were trained and evaluated by comparing test results on the HRSID and the SSDD. Based on the analysis of experimental results, our proposed method demonstrates effective detection of small target ships, highlighting its potential applications in SAR ship detection. However, it has some limitations in accurately recognizing densely-stacked ships, which can be attributed to the characteristics of the dataset and certain factors inherent to the algorithm. In follow-up research, we consider the proposed method with the rotation head frame to achieve more efficient detection and localization of ships at multi-scales.

References

[1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. & Remote Sensing Mag.*, vol. 1, no. 1, pp. 6-43, Mar. 2013.

[2] J. Li, C. Xu, H. Su, L. Gao, and T. Wang, "Deep learning for SAR ship detection: Past, present and future," *Remote Sensing*, vol. 14, no. 11, pp. 2712, June 2022.

[3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137-1149, June 2017.

[4] T. -Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318-327, Feb. 2020.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proc. IEEE CVPR 2016*, pp. 779-788, Las Vegas, USA, Dec. 2016.

[6] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," https://github.com/ultralytics/ultralytics, Feb. 2023.

[7] Z. Tong, Y. Chen, Z. Xu, and R. Yu, "Wise-IoU: Bounding box regression loss with dynamic focusing mechanism," *arXiv preprint arXiv:2301.10051*, Jan. 2023.

[8] C. Wang, A. Bochkovskiy, and H. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv preprint arXiv:2207.02696*, July 2022.

[9] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234-120254, June 2020.

[10] T. Zhang, X. Zhang, J. Li, X. Xu, B. Wang, X. Zhan, Y. Xu, X. Ke, T. Zeng, H. Su, I. Ahmad, D. Pan, C. Liu, Y. Zhou, J. Shi, and S. Wei, "SAR ship detection dataset (SSDD): Official release and comprehensive data analysis," *Remote Sensing*, vol. 13, no. 18, pp. 3690, Sept. 2021.

[11] G. Jocher, "YOLOv5," available online at https://github.com/ultralytics/yolov5, 2020.

[12] C. Y. Wang, H. Y. M. Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," *Proc. IEEE CVPR 2020*, pp. 1571-1580, virtual conference, June 2020.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904-1916, Sept. 2015.

[14] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," *Proc. IEEE CVPR 2018*, pp. 8759–8768, Salt Lake City, USA, June 2018.

[15] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," *Proc. IEEE CVPR 2017*, pp. 2117–2125, Honolulu, USA, July 2017.

[16] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang, and J. Yang, "Generalized focal loss: Learning qualified and distributed bounding boxex for dense object detection," *arXiv preprint arXiv:2006.04388*, June 2020.

[17] C. Yu and Y. Shin, "A deep learning-based SAR ship detection," *Proc. ICAIIC 2023*, pp. 744-747, Bali, Indonesia, Feb. 2023.