Spotlight on Small-scale Ship Detection: Empowering YOLO with Advanced Techniques and a Novel Dataset

Lingya Li[1][0009−0008−3361−8196], Zhixing Hou[2][0009−0006−3335−2768], Ming Ma[1], Jing Xiang[1], Chuangxin Yuan[1], and Guihua Xia[†1]

[1] Harbin Engineering University, Harbin, 150001 Heilongjiang, China
[2] Nanjing University of Science and Technology, Nanjing, 210094 Jiangsu, China
lyli18@hrbeu.edu.cn,[†]xiaguihua@hrbeu.edu.cn

Abstract. In recent years, significant advancements have been made in deep learning-based ship detection methods on the ocean surface. However, publicly available maritime datasets that include categories for small-scale ships and network frameworks optimized explicitly for small-scale ship detection on the ocean surface are still limited in availability. To address the data scarcity in small-scale ship detection, bridge the gap between small-scale ship detection and general object detection, and mitigate the impact of small objects on maritime safety, we collect a multi-scale dataset with a particular emphasis on detecting small objects on the ocean surface, named the iShip-1. Leveraging this dataset, we train the $\mathbf{S}^3Det$ for small-scale ship detection, which remarkably detects small-scale ships on the ocean surface. Specifically, the iShip-1 comprises 17,236 images encompassing six categories captured from multiple perspectives and under various weather conditions. Notably, the Other Ship category focuses explicitly on small-scale ships. The S^3Det is optimized for detecting small-scale ships through improved backbone and neck architecture. It employs the NWD Loss instead of the traditional IoU Loss and utilizes the Feedback Cut&Paste technique for effective data augmentation. We evaluate the performance of the S^3Det on both the Seaships and iShip-1. For small-scale ship detection, S^3Det achieved a recall rate of 68.9%, a mAP50 of 73.9%, and a mAP50:90 of 39.4%. These results indicate improvements of 5.9%, 2%, and 1.2% compared to the original YOLOv8 model, respectively. Our code and dataset are available here `https://github.com/li01233/S3Det`.

Keywords: small-scale ship detection · maritime dataset · maritime target recognition

1 Introduction

Recently, deep learning techniques have been introduced by many researchers into maritime surveillance to ensure navigation safety on the ocean. An essential part of maritime surveillance is object detection, however, the current techniques

still have low accuracy for several reasons: 1) more complicated light conditions on the ocean surface compared with traditional object detection; 2) the pictures captured from the sea surface are frequently affected by clouds, waves, and fog; 3) different ship types have more similar appearance features visually, while the front and sides of the same ship differ more; and 4) the current maritime public datasets do not adequately cover some ship types. All of these result in challenges with object detection on the ocean surface.

Researchers have tried several methods to deal with the various challenges. To enhance accuracy, some researchers attempt to modify the network, such as adding the attention module[20, 22], changing the neck[16, 20] and backbone[3] of the original network. While others make significant performance improvements by optimizing the NMS(Non-Maximum Suppression) algorithms and loss function[23] or by innovating in data augmentation[16].

For complex environmental conditions on the ocean surface, Shao Zhenfeng[10] limit the detection range by Sea-Sky Line to minimize interference. Nie Xin[8] take the initiative to produce low-quality images to increase the model's robustness. Appropriate research has additionally been conducted by Zeng Guangmiao[19] to re-identify a ship's front and side. In the meantime, an increasing amount of researchers have contributed to publicly available maritime datasets. One of the most popular maritime datasets is Seaships[10], which includes six common ship types. Singapore maritime dataset(SMD)[9] provides Video footage that can be used for object tracking and detection.

Fig. 1: Detection results of YOLOv8 model trained on Seaships.

However, in real-world applications, there are still a lot of difficulties, it is evident from Fig. 1 that the model trained on the current publicly available marine dataset struggles to identify small objects, and performs poorly in low-light situations (such as dusk or foggy), which would be fatal in the real world where missed detection could lead to ship collisions. The reason for this poor performance is twofold: firstly, a dataset encompassing the small object category is still unavailable; secondly, there are few network frameworks particularly optimized for small objects. To enhance small-scale ship detection on the sea and to guarantee the security of navigation, a dataset comprising small object categories and multiple perspectives in multi-weather environments should be created, and specifically algorithm optimization based on small-scale ships is required.

To address the aforementioned challenges, we develop our dataset, iShip-1, and propose a specialized model called $\mathbf{S}^3Det$ for efficient small-scale ship detec-

tion. Our work can be divided into two main steps. Firstly, we collect the iShip-1 dataset, which includes images of six different types of ships (Bulk Carrier, Cargo Ship, Passenger Ship, Fishing Vessel, Pleasure Craft), as well as small-scale ships (Other Ship) under various weather conditions, both on the shore and onboard from multiple perspectives. We carefully annotate the images and correct the original SeaShips labels to ensure accurate training data. Secondly, we utilize the iShip-1 dataset to train the S^3Det (as depicted in Fig. 2). The model is built upon the YOLOv8 network architecture, utilizing EfficientNetV2[12] as its backbone and incorporating the neck structure from Gold-YOLO[14]. To enhance the detection performance of small objects, we replace the traditional IoU Loss with NWD Loss[15] in the loss function. Additionally, we introduce a feedback signal for Feedback Small-Cut&Paste to address the issue of imbalanced data distribution and accelerate the recognition speed of small object detection.

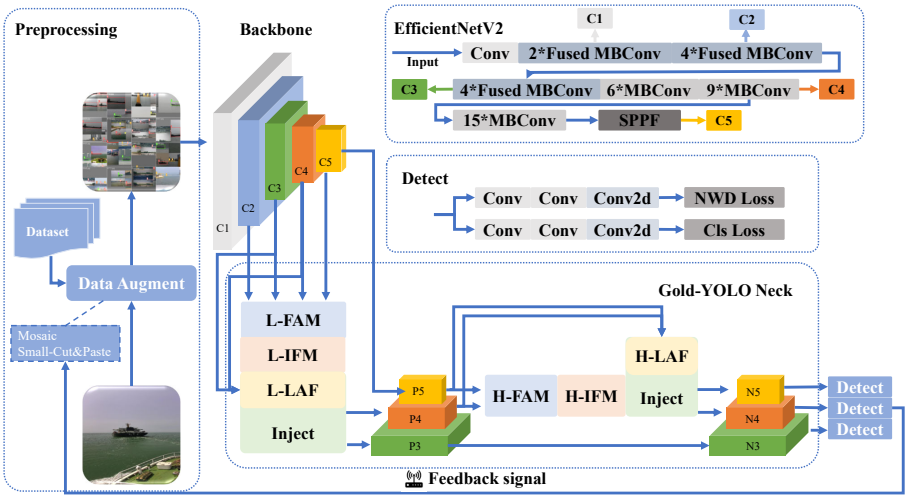


Fig. 2: General overview of S^3Det framework. In the preprocessing part, Mosaic and Feedback Small-Cut&Paste can enhance the data, and the feedback signal from the network output is used to adaptively align the threshold of Feedback Small-Cut&Paste. EfficientNetV2 is adopted as the backbone, and Fused MBConv and MBConv are in place of the original C2F module for feature extraction. Gold-YOLO's neck is chosen in the neck section instead of the original PAN, which integrates the information of C2, C3, C4, and C5 and finally outputs the feature maps of the three layers to the detect head. The detect head will output the loss for position and classification respectively, NWD Loss is used as a substitute for the original IoU Loss.

The contributions of our paper are as follows:

- We propose a ship benchmark dataset named iShip-1, which is available for small-scale ships and encompasses multiple perspectives under multi-weather conditions, to enhance the model's accuracy for detecting small-scale ships on the ocean surface and safeguard navigation safety in real-world scenarios.
- Our S^3Det improves based on the original YOLOv8 model, especially for small objects. The model is tested on Seaships and iShip-1, respectively, which exhibit notable performance in the recognition accuracy of small-scale ships. We have also already deployed the model on the "Haitun-1" scientific research experimental ship for experiments.
- We introduce a novel data augmentation technique called Feedback Small-Cut&Paste to address the dataset imbalance issue and enhance the network's ability to detect small objects. This technique significantly improves the network's sensitivity to small objects, thereby enhancing its overall performance in small object detection tasks.

2 Related Works

2.1 Current Maritime Public Datasets

Many existing maritime public datasets have been captured by researchers or enhanced from the original datasets. Apart from Seaships[10] and SMD[9] mentioned above, ABOShips[4], SeaSAw[5] and MCMOD[11] are also excellent maritime public datasets. Table 1 displays these datasets' specifics.

Table 1: Details of Maritime Public Datasets.

Name	Types	Images	Instances	Image Size
Seaships[10]	6	7,000	31,455	1920*1080
SMD[9]	2	17,450[1]	192,980	1920*1080
ABOShips[4]	9	9,880	41,967	1280*720
SeaSAw[5]	12	19,000,000	146,000,000	1920*1080[2]
MCMOD[11]	10	16,166	98,590	1920*1080

[1] These frames are extracted from the 36 videos in the dataset.
[2] SeaSAw dataset offers images in four resolutions: 7680*1408, 3840*2056, 3648*2052, 1920*1080.

2.2 Object Detection in Marine Field

Much of the work has proven to be effective in the field of marine object detection. The YOLO-ship model[23] replaces the convolutional layer in YOLOv5 with MinxConv and inserts the CA attention module, which results in an mAP up to 72.8%. On the other hand, Zhang Alun[20] also add the attention module CA to YOLOv5, and they choose BiFPN as the neck architecture, which raises

the recognition rate by 3.3% to 99.1% compared with the original one. Zhao Hangyue[22] base their YOLO-sea on YOLOv7, adding SimAM to the path, as a result of an AP of 59% and an improvement of 7%. Tan Xiangyu[13] revise the Faster R-CNN by introducing soft NMS and Focal Loss, which can improve the accuracy of inshore vessel recognition rate by 6% to 80.4% and mAP by 1.1% to 84.5%.

2.3 Small Object Detection

The performance of small object detection is still limited because this task often suffers from insufficient information and low resolution. The FPN[7], which combines high-level semantic and low-level spatial information to boost the features, is a representation of feature pyramid-based techniques. By producing high-resolution features or images, GAN-based techniques provide small objects with additional features. Stitcher[2] addresses the issue at the data level. It applies the ratio of small object loss to total loss as the feedback to decide whether to combine four images in the next iteration, enhancing the detection performance by increasing the number of small objects.

Small object detectors are now widely used in remote sensing and other fields. Such as the novel detector mentioned in [18], which can handle freely rotated objects of arbitrary sizes in remote sense images through its method based on CNN. Similarly, some researchers study the detection of small-scale ships. The enhanced-YOLOv7[6] is an effective small-ship detection algorithm because of its unique feature extraction module and multi-branch residual structures.

2.4 Data Augmentation in Object Detection

Data augmentation techniques can extend the dataset, avoid the disparity in the datasets, strengthen model robustness, and prevent overfitting. Object detection has witnessed steady performance improvements from MixUp[21], Mosaic[1], and CutMix[17] in terms of both accuracy and robustness.

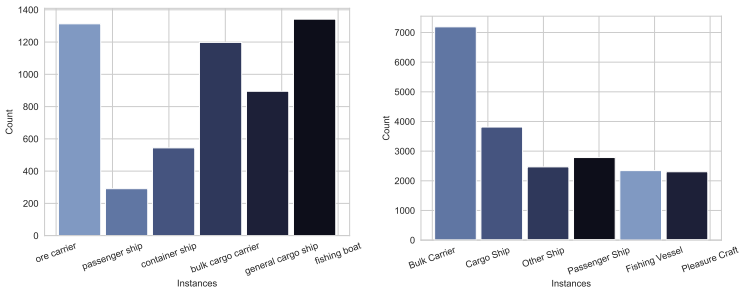
3 Dataset Details

The Seaships dataset is gathered onshore, containing six types of ship categories: ore carrier, bulk cargo carrier, container ship, general cargo ship, fishing boat, and passenger ship. This classification method raises category overlap and scarcity of certain common categories (such as passenger ship, sailboat). Collecting on shore contributes to a lack of frontal features of the ship.

We re-label the Seaships by adding labels for small-scale ships and correcting the wrong ones. Additionally, images of real ships captured ashore and on board under different weather conditions and a wide range of perspectives on the sea surface, are added to construct the iShip-1. iShip-1 has a total of 17,236 images and corresponding annotations, the resolution of the images ranges from 800x600 to 6000x4000. We classify the instances into six categories: Bulk Carrier, Cargo

Ship, Other Ship, Passenger Ship, Fishing Vessel, and Pleasure Craft, of which Other Ship is for small-scale ships notably. The mean value of image sizes (width, height) is (1969.60, 1211.65), and the standard deviation is (843.12, 615.68).

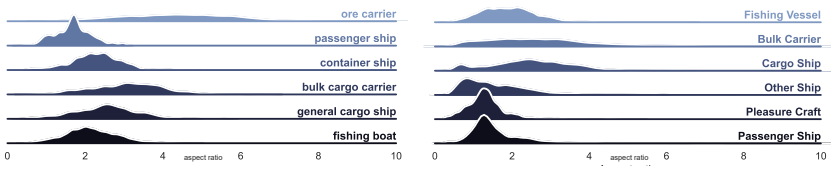
As for the definition of the small-scale ships, We usually use the object to image area ratio to define the small object in the computer vision field. However, small-scale ships usually refer to ships with weak radar characteristics, without AIS, and at long distances in the maritime field. To unify the definitions in these two domains, we take an object beyond the distance (400 m) as a small object and calculate the threshold to be 0.39% for a 25m ship based on the camera's intrinsic parameters. The labeling of our collected images is completed based on this threshold.



(a) Seaships (b) iShip-1

Fig. 3: Category statistics for training set of Seaships and iShip-1.

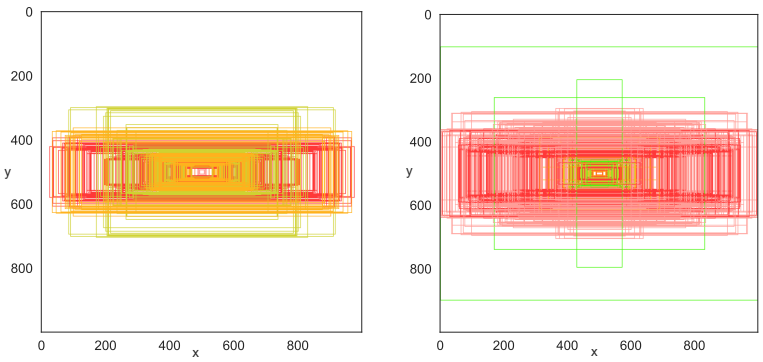
Fig. 3 demonstrates the classification statistics of these two training sets respectively, it can be found that the Seaships have fewer instances in general, and its categories are not well distributed. Our iShip-1 contains more instances like the number of objects under the category Bulk Carrier is more than 7,000; it is more balanced, with the ratio of the smallest category to the largest being no less than 1:3; and its classification method is more scientific, containing five major categories of ships and small-scale ships that are common in reality.



(a) Seaships (b) iShip-1

Fig. 4: The aspect ratio distributions of the annotation boxes for six categories in both the Seaships and iShip-1.

The annotation box aspect ratio distributions under each category for both two datasets are shown in Fig. 4. In the Seaships dataset, the aspect ratio of the ore carrier spans a wide range from 2:1 to 6:1, while the passenger ship and fishing boat category concentrates on 1:2, and 1:2.5 for the container ship. iShip-1 has a more abundant distribution of aspect ratios. The Passenger Ship and Pleasure Craft have similar distributions, and both center on around 1:1.5; The Bulk Carrier and Cargo Ship are also a pair of akin categories, spanning a wide spectrum; Fishing Vessel peaks between 1:1.5 and 1:2; as for the small object category Other Ship, it has a smaller aspect ratio compared with other categories, the spike of it is even less than 1:1. Abundant and highly differentiated aspect ratios can provide more training information while posing a challenge to the detector. At the same time, the iShip-1's rich aspect ratio is due to the fact that it contains images taken offshore of the boat from more perspectives.

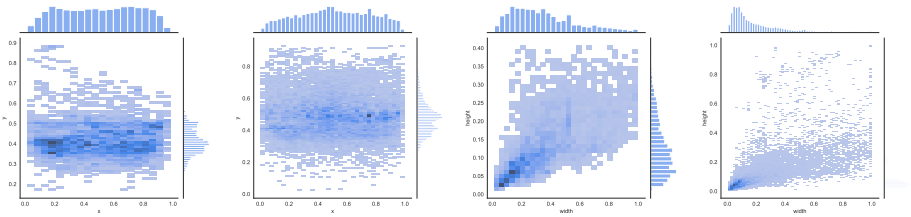


(a) Seaships (b) iShip-1

Fig. 5: Annotation boxes details for Seaships and iShip-1.

As it can also be noticed from their annotation box details in Fig. 5, iShip-1 provides a much wider variety of ships with more plentiful aspect ratios for training than Seaships. The mean value of relative center positions (x, y) is (0.51, 0.49), and the standard deviation is (0.25,0.12). The mean value of relative sizes (width, height) is (0.22, 0.11), and the standard deviation is (0.21, 0.11).

The relationships between the relative center positions (x, y) and relative sizes (width, height) of the annotation boxes in the two datasets are illustrated in Fig. 6. The annotation boxes of Seaships are centered on the central horizontal line of the frame, while iShip-1's annotation boxes are more widely distributed, which is more in line with the situation in real-world navigation. The height and width of the Seaships and iShip-1 are both clustered from 0 to 0.2, but the iShip-1 has a more diffuse distribution of height and width than Seaships, offering multi-scale data.



(a) Seaships (b) iShip-1 (c) Seaships (d) iShip-1

Fig. 6: Relations between the annotation boxes' (x, y), (width, height) for Seaships and iShip-1.

4 Method

As Fig. 2 shows, S^3Det is mainly refined based on YOLOv8, and its framework consists of 4 major parts: 1) the data augmentation module; 2) EfficientNetV2 as the backbone for extracting the features of the data; 3) the fusion of the four feature maps through Gold-YOLO's neck; 4) detect head which exports inference results and loss.

4.1 Data Augmentation: Feedback Small-Cut&Paste

Seaships and iShip-1 both comprise an imbalance of instance classes, which leads to poor performance, especially for small objects. Feedback Small-Cut&Paste can solve this problem well by automatically adjusting the number of small objects for training. It can select small objects from the whole database and paste them into the training image according to the feedback signal. In the meantime, the feedback signal ensures the training speed at the later stage.

Algorithm 1 Feedback Small-Cut&Paste

1: $\int_{-\infty}^{small\ threshold} f(x)\,dx = \frac{1}{3} \Rightarrow small\ threshold$ (Sthr)
2: **while** training **do**
3: $Loss_{all} = S^3Det(image, label)$ (Original Images)
4: $Loss_{small} = Loss_{all}[\frac{Area_{gt\_box}}{Area_{image}} \leqslant Sthr]$
5: $small\ ratio = \frac{Loss_{small}}{Loss_{all}}$ (Sr)
6: **if** $random.uniform > Sr$ **then**
7: $Sthr* = 0.999$
8: $SmallCut\&Paste(image, Objects[\frac{Area_{gt\_box}}{Area_{image}} \leqslant Sthr])$
9: **end if**
10: **end while**

Fig. 7 and pseudo-code 1 show the overall process of Feedback Small-Cut&Paste, the feedback signal means the small ratio given in the figure. In the forward

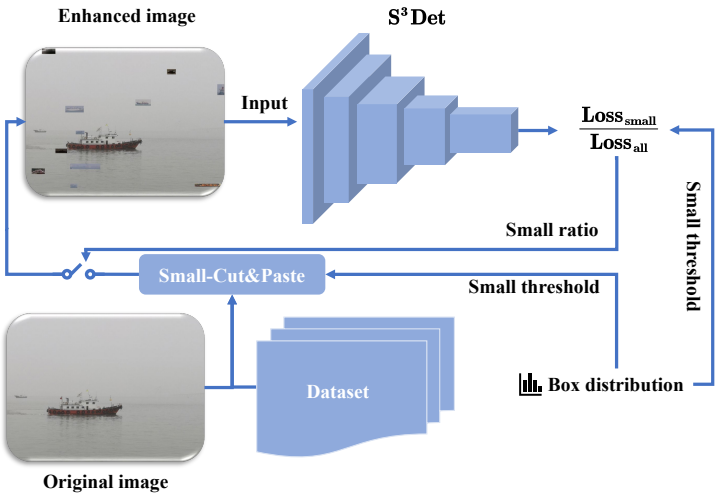


Fig. 7: The pipeline of Feedback Small-Cut&Paste.

process, the image is sent to S^3Det to get the forecast results and the corresponding loss of each prediction. Then in the backward process, the small threshold is calculated based on the distribution of the ground-truth boxes to define small objects, further to get the loss of small objects. The calculated small ratio is utilized as the feedback signal to manage whether or not to turn Feedback Small-Cut&Paste on. Eventually, the small objects are extracted from the entire dataset based on the aforementioned small threshold and pasted into the original image to generate the enhanced image, which is then sent for training.

Generally, it is directly considered objects satisfying the condition of $\frac{Area_{gt\_box}}{Area_{image}} \leqslant 0.5\%$ are small objects. However, for different datasets, the difficulty of small object detection varies, using a fixed threshold does not facilitate the network well. Therefore, it is more reasonable to define the small threshold depending on the distribution of the dataset. Fig. 8 presents the top three fitted distributions based on the aspect ratios of the ground-truth boxes in i-Ship, it reveals that the aspect ratios in the iShip-1 dataset are best represented by the lognormal distribution $\log_a X \sim N\left(\mu, \sigma^2\right)$, with the following probability density function:

$$f(x) = \begin{cases} \frac{1}{x \ln a \sqrt{2\pi}\sigma} \exp\left[-\frac{(\log_a x - \mu)^2}{2\sigma^2}\right], & x \in (0, +\infty) \\ 0, x \in (-\infty, 0] \end{cases} . \tag{1}$$

We expect small objects to account for $1/3$ of the entire dataset, so solving $\int_{-\infty}^{small\ threshold} f(x)\, dx = \frac{1}{3}$ yields the small threshold.

Fig. 9 displays the augmentation results through Feedback Small-Cut&Paste.

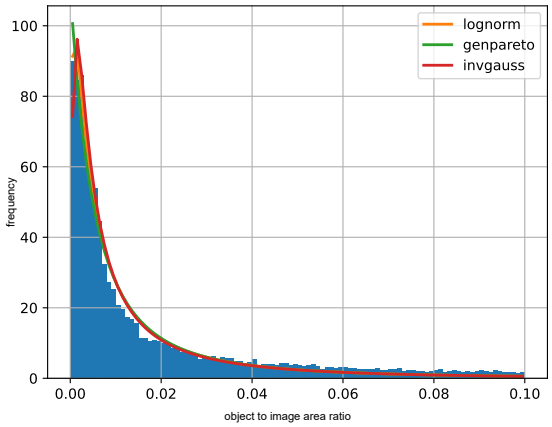


Fig. 8: Top three fitted distributions.

(a) original image (b) small threshold: .01 upper limit: 10 (c) small threshold: .1 upper limit: 5 (d) small threshold: .01 upper limit: 5

Fig. 9: Augmentation results of Feedback Small-Cut&Paste with different hyper-parameters: small threshold and upper limit of pasted instances.

4.2 Backbone: EfficientNetV2

As compared with the original YOLOv8's backbone CSPDarkNet's main component module C2F, EfficientNetV2's MBConv[12] upscales the inputs by the expansion ratio at the beginning, which enables the module to deliver much richer features, leading to be more sensitive for the small object detection. Meanwhile, as an evolution of EfficientNet, EfficientNetV2 changes the low-level layers from MBConv to Fuse MBConv to improve inference speed and be more friendly with the deployment on the real ship.

4.3 Neck: the neck structure from Gold-YOLO

It often encounters the case of detecting objects at multiple scales in the task of object detection on the ocean surface. YOLOv8 adopts PAN to enhance the expression of multi-scale features. However, we find that in the case of cross-scale objects existing in the same image, this fusion-by-path method can not integrate the cross-layer features directly. Therefore we utilize the neck structure from Gold-YOLO[14], which can fuse non-adjacent feature maps, which is essential

for the detection task at various scales. At the same time, it also introduces the information of C2, which will offer richer semantic information for small objects.

4.4 Loss Function: NWD Loss

The sensitivity of the IoU for objects of differing scales varies greatly. In terms of small objects, a small position deviation can significantly degrade the IoU, resulting in small objects that would be easily neglected. Hence we choose Normalized Wasserstein Distance (NWD) instead of IoU to describe the distances between the ground-truth boxes and the predicted boxes. The ground-truth boxes will have background pixels on the border, while the objects are concentrated more at the center[15], so 2D Gaussian distribution can model this kind of situation quite well. The ground-truth boxes (c_x, c_y, w, h) with center point x,y, width, and height can be modeled into 2D Gaussian distribution $N(\mu, \Sigma)$ with

$$\mu = \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \Sigma = \begin{bmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix}. \tag{2}$$

Thus, for the target box $B_{target} = \left(c_x^t, c_y^t, w^t, h^t\right)$ and the prediction box $B_{pred} = \left(c_x^p, c_y^p, w^p, h^p\right)$ their second order Wasserstein distance is:

$$\mathcal{W}_2^2\left(N_{target}, N_{pred}\right) = \left\| \left(\begin{bmatrix} c_x^t, c_y^t, \frac{w^t}{2}, \frac{h^t}{2} \end{bmatrix}^T, \begin{bmatrix} c_x^p, c_y^p, \frac{w^p}{2}, \frac{h^p}{2} \end{bmatrix}^T \right) \right\|_2^2. \tag{3}$$

Normalization is applied to Wasserstein distance so that Normalized Wasserstein Distance can be limited between 0 and 1, as a similarity metric to replace IoU:

$$NWD\left(N_{target}, N_{pred}\right) = \exp\left[-\frac{\sqrt{\mathcal{W}_2^2\left(N_t, N_p\right)}}{C} \right]. \tag{4}$$

The NWD is designed as a part of the loss function for the box regression task, box loss is calculated as follows:

$$Loss_{box} = (1 - r) * (1 - CIoU\left(Bt, B_p\right)) + r * \left(1 - NWD\left(N_t, N_p\right)\right), \tag{5}$$

where r ranging from 0 to 1, is the coefficient for balancing NWD and IoU. The larger r is, the more suitable the loss function is for datasets with many small objects.

5 Experiments

5.1 Performance Comparison

We test the network performance on the test sets of Seaships and iShip-1 respectively, choosing YOLOv8 as the baseline, comparing the S^3Det proposed in

Table 2: Performance comparison on Seaships.

Model	Precision Rate	Recall Rate	mAP50	mAP50:90
YOLOv5	96.8%	97.3%	98.6%	78.2%
YOLOx	78.9%	77.8%	93%	75.3%
YOLOv7	97%	**98.2%**	**99.2%**	80.8%
RT-DETR	95%	96.6%	98.1%	79.4%
YOLOv8(baseline)	97.6%	97.1%	98.7%	**84%**
S^3Det(ours)	**97.9%**	97.3%	99%	**84%**

Table 3: Ablation Study on Seaships.

Model	Precision Rate	Recall Rate	mAP50	mAP50:90
baseline	97.6%	97.1%	98.7%	84%
+NWD Loss	97.7%	96.9%	98.9%	84.1%
+EfficientNetV2	97.6%	97.4%	**99%**	**84.2%**
+Gold-YOLO's neck	96.8%	**97.7%**	98.7%	84.1%
+Small-Cut&Paste	97.8%	96.8%	98.9%	**84.2%**
S^3Det(ours)	**97.9%**	97.3%	**99%**	84%

this article with the results of the baseline and other models. Table 2 exhibits the experiment results on Seaships of various methods.

Based on the evaluation results, we can observe that the S^3Det outperforms the other mentioned models in terms of precision rate, recall rate, mAP50, and mAP50:90. Specifically, it achieves first place in precision rate and mAP50:90, and second place in recall rate and mAP50. Notably, compared to both YOLOv7 and the baseline model, the S^3Det exhibits a remarkable improvement of 3.2% in mAP50:90 and a 0.3% enhancement in precision rate.

To examine the contribution of each component towards the whole network, we conduct an ablation experiment on Seaships to analyze the Feedback Small-Cut&Paste, EfficientNetV2, Gold-YOLO's neck, and NWD Loss these four components' effects separately. Notably, S^3Det includes all four components. As shown in Table 3, NWD Loss and Feedback Small-Cut&Paste improve the network by 0.1%-0.2% with almost no consumption of inference resources; EfficientNetV2 contributes to improving the mAP50 by about 0.3% and the mAP50:90 by about 0.2%; the neck structure from Gold-YOLO increases the mAP50:90 by about 0.3%.

Table 4 exhibits the experiment findings of different methods on iShip-1. It can be observed that S^3Det has a larger improvement compared to other models, gaining improvements of 0.4%, 1.6%, 2.2%, 0.5%, 1.1% in precision rate, recall rate, mAP50, and mAP50:90 respectively over baseline in all categories. S^3Det has the highest recall rate and mAP50:90 at 89.6% and 65.8% respectively, and the second highest precision rate and mAP50. For small object detection, S^3Det shows superior performance with a 5.9% improvement in recall rate versus baseline, while mAP50 and mAP50:90 are even improved by 2% and 1.2% respectively.

Table 4: Performance comparison on iShip-1.

Method	ALL				Other Ship			
	P	R	mAP50	mAP50:90	P	R	mAP50	mAP50:90
YOLOv5	90.5%	86%	90%	58.8%	78.3%	61.3%	67%	33.1%
YOLOx	82.8%	69.4%	86.1%	61.3%	/	/	69.8%	/
YOLOv7	**91.9%**	89.2%	**92.4%**	63.6%	**83%**	64.8%	72.7%	36.3%
RT-DETR	85.7%	83.2%	86.4%	58.5%	72.8%	45.8%	56.7%	29.4%
YOLOv8(baseline)	90.6%	88%	91.8%	64.7%	80.5%	63%	71.9%	38.2%
S^3Det(ours)	91%	**89.6%**	92.3%	**65.8%**	79.2%	**68.9%**	**73.9%**	**39.4%**

Table 5: Ablation Study on iShip-1.

Method	ALL				Other Ship			
	P	R	mAP50	mAP50:90	P	R	mAP50	mAP50:90
baseline	90.6%	88%	91.8%	64.7%	80.5%	63%	71.9%	38.2%
+NWD Loss	90.9%	89.4%	**92.3%**	65.3%	78.7%	67.8%	73%	38.3%
+EfficientNetV2	**91.6%**	**89.8%**	**92.3%**	**65.9%**	81%	68.1%	72.9%	38.6%
+Gold-YOLO's neck	90.9%	89.1%	92%	65.6%	80.5%	67.3%	72.2%	38.3%
+Small-Cut&Paste	**91.6%**	88.8%	92.1%	65.3%	**83.5%**	65%	73.2%	38.7%
S^3Det(ours)	91%	89.6%	**92.3%**	65.8%	79.2%	**68.9%**	**73.9%**	**39.4%**

The result of ablation experiment on iShip-1 is displayed in Table 5, and it can be seen that for all categories, EfficientNetV2 contributes the most, with 1.6%, 1.8%, 0.5%, and 1.2% enhancements to baseline in precision rate, recall rate, mAP50, and mAP50:90 respectively; NWD Loss has achieved an increase of 0.5% on mAP50, besides Feedback Small-Cut&Paste have 1% improvement on precision rate over baseline. As far as small objects are concerned, all four components have significant improvements against the baseline. S^3Det possesses the highest recall rate, mAP50 and mAP50:90 for small object detection at 68.9%, 73.9%, 39.4% in Other Ship, and 91%, 89.6%, 92.3%, 65.8%. which are the first or second highest of all models.

5.2 Real Ship Deployment

S^3Det is deployed on the "Haitun-1" scientific research experimental ship under TensorRT8.6.1.6, CUDA11.8, and Pytorch2.1, using an Nvidia RTX3050 (8G) GPU, maintaining a stable inference speed of around 30fps, achieving real-time standards for operation.

We further evaluate the performance of the S^3Det in detecting ships under four extreme weather conditions: dusk, rainy, foggy, and cloudy, both in the harbor and the open sea. The details are shown in Table 6 below.

Table 6: Detect metrics under four extreme weather.

Weather	Precision	Recall	mAP50	mAP50:90
rainy	55%	57%	57.5%	40.6%
foggy	59.9%	59.9%	60.9%	46.3%
cloudy	55%	56%	56.7%	41.2%
dusk	58%	57.1%	58.2%	42.6%

6 Conclusion

In practical object detection on the sea surface, we usually have to consider the detection of small-scale ships to guarantee safety for navigation. In this paper, we build our own multi-scale and multi-perspective iShip-1 with small object category, train the S^3Det which is optimized by EfficientNetV2, Gold-YOLO's neck, NWD Loss, and Feedback Cut&Paste for detecting small objects in Seaships and iShip-1. Significantly, S^3Det boosts the performance of small-ship detection. Meanwhile, we also deploy the S^3Det on the "Haitun-1" scientific research experimental ship to prove its practicality.

Acknowledgments. This work was supported in part by the National Key R&D Program of China under Grant 2019YFE0105400, in part by the National Natural Science Foundation of China under Grant 52171302, in part by the Young Elite Scientists Sponsorship Program by CAST under Grant 2022QNRC001, in part by the Fundamental Research Funds for the Central Universities under Grant 3072023CFJ0402, in part by the New Era Longjiang Excellent Dissertations under Grant LJYXL2022-004.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Bibliography

[1] Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)

[2] Chen, Y., Zhang, P., Li, Z., Li, Y., Zhang, X., Qi, L., Sun, J., Jia, J.: Dynamic scale training for object detection. arXiv preprint arXiv:2004.12432 (2020)

[3] Gao, X., Sun, W.: Ship object detection in one-stage framework based on swin-transformer. In: Proceedings of the 2022 5th International Conference on Signal Processing and Machine Learning. pp. 189–196 (2022)

[4] Iancu, B., Soloviev, V., Zelioli, L., Lilius, J.: Aboships—an inshore and offshore maritime vessel detection dataset with precise annotations. Remote Sensing **13**(5), 988 (2021)

[5] Kaur, P., Aziz, A., Jain, D., Patel, H., Hirokawa, J., Townsend, L., Reimers, C., Hua, F.: Sea situational awareness (seasaw) dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2579–2587 (2022)

[6] Li, J., Ding, N., Gong, C., Jin, Z., Li, G.: Effective small ship detection with enhanced-yolov7. In: PRCV 2023. vol. 14434, pp. 249–260 (2024)

[7] Liu, S., Qi, L., Qin, H., Shi, J., Jia, J.: Path aggregation network for instance segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8759–8768 (2018)

[8] Nie, X., Yang, M., Liu, R.W.: Deep neural network-based robust ship detection under different weather conditions. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). pp. 47–52. IEEE (2019)

[9] Prasad, D.K., Rajan, D., Rachmawati, L., Rajabally, E., Quek, C.: Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. IEEE Transactions on Intelligent Transportation Systems **18**(8), 1993–2016 (2017)

[10] Shao, Z., Wu, W., Wang, Z., Du, W., Li, C.: Seaships: A large-scale precisely annotated dataset for ship detection. IEEE transactions on multimedia **20**(10), 2593–2604 (2018)

[11] Sun, Z., Hu, X., Qi, Y., Huang, Y., Li, S.: Mcmod: The multi-category large-scale dataset for maritime object detection. CMC-COMPUTERS MATERIALS & CONTINUA **75**(1), 1657–1669 (2023)

[12] Tan, M., Le, Q.: Efficientnetv2: Smaller models and faster training. In: International conference on machine learning. pp. 10096–10106. PMLR (2021)

[13] Tan, X., Tian, T., Li, H.: Inshore ship detection based on improved faster r-cnn. In: MIPPR 2019: Automatic Target Recognition and Navigation. vol. 11429, pp. 39–45. SPIE (2020)

[14] Wang, C., He, W., Nie, Y., Guo, J., Liu, C., Wang, Y., Han, K.: Gold-yolo: Efficient object detector via gather-and-distribute mechanism. Advances in Neural Information Processing Systems **36** (2024)

[15] Wang, J., Xu, C., Yang, W., Yu, L.: A normalized gaussian wasserstein distance for tiny object detection. arXiv preprint arXiv:2110.13389 (2021)
[16] Wu, W., Li, X., Hu, Z., Liu, X.: Ship detection and recognition based on improved yolov7. Comput. Mater. Contin **76**(1), 489–498 (2023)
[17] Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6023–6032 (2019)
[18] Zand, M., Etemad, A., Greenspan, M.: Oriented bounding boxes for small and freely rotated objects. IEEE Transactions on Geoscience and Remote Sensing **60** (2022)
[19] Zeng, G., Wang, R., Yu, W., Lin, A., Li, H., Shang, Y.: A transfer learning-based approach to maritime warships re-identification. Engineering Applications of Artificial Intelligence **125**, 106696 (2023)
[20] Zhang, A., Zhu, X.: Research on ship target detection based on improved yolov5 algorithm. In: 2023 5th International Conference on Communications, Information System and Computer Engineering (CISCE). pp. 459–463. IEEE (2023)
[21] Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
[22] Zhao, H., Zhang, H., Zhao, Y.: Yolov7-sea: Object detection of maritime uav images based on improved yolov7. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 233–238 (2023)
[23] Zhou, S., Yin, J.: Yolo-ship: a visible light ship detection method. In: 2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE). pp. 113–118. IEEE (2022)