applied sciences

MDPI

Article

ST-YOLOv8: Small-Target Ship Detection in SAR Images Targeting Specific Marine Environments

Fei Gao [1,2,3], Yang Tian [1,2,3,*], Yongliang Wu [1,2,3] and Yunxia Zhang [1,2,3]

[1] School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang 050043, China; gaof0105@stdu.edu.cn (F.G.); wuyongliang@stdu.edu.cn (Y.W.); zhangyx01@stdu.edu.cn (Y.Z.)
[2] Hebei Key Laboratory of Electromagnetic Environmental Effects and Information Processing, Shijiazhuang 050043, China
[3] Shijiazhuang Key Laboratory of Artificial Intelligence, Shijiazhuang 050043, China
* Correspondence: yangtian2022221@163.com

Abstract: Synthetic Aperture Radar (SAR) image ship detection faces challenges such as distinguishing ships from other terrains and structures, especially in specific marine complex environments. The motivation behind this work is to enhance detection accuracy while minimizing false positives, which is crucial for applications like defense vessel monitoring and civilian search and rescue operations. To achieve this goal, we propose several architectural improvements to You Only Look Once version 8 Nano (YOLOv8n) and present Small Target-YOLOv8(ST-YOLOv8)—a novel lightweight SAR ship detection model based on the enhance YOLOv8n framework. The C2f module in the backbone's transition sections is replaced by the Conv_Online Reparameterized Convolution (C_OREPA) module, reducing convolutional complexity and improving efficiency. The Atrous Spatial Pyramid Pooling (ASPP) module is added to the end of the backbone to extract finer features from smaller and more complex ship targets. In the neck network, the Shuffle Attention (SA) module is employed before each upsampling step to improve upsampling quality. Additionally, we replace the Complete Intersection over Union (C-IoU) loss function with the Wise Intersection over Union (W-IoU) loss function, which enhances bounding box precision. We conducted ablation experiments on two widely used multimodal SAR datasets. The proposed model significantly outperforms the YOLOv8n baseline, achieving 94.1% accuracy, 82% recall, and 87.6% F1 score on the SAR Ship Detection Dataset (SSDD), and 92.7% accuracy, 84.5% recall, and 88.1% F1 score on the SAR Ship Dataset_v0 dataset (SSDv0). Furthermore, the ST-YOLOv8 model outperforms several state-of-the-art multi-scale ship detection algorithms on both datasets. In summary, the ST-YOLOv8 model, by integrating advanced neural network architectures and optimization techniques, significantly improves detection accuracy and reduces false detection rates. This makes it highly suitable for complex backgrounds and multi-scale ship detection. Future work will focus on lightweight model optimization for deployment on mobile platforms to broaden its applicability across different scenarios.

Keywords: deep learning; SAR ship detection; ST-YOLOv8

Academic Editor: Atsushi Mase

Received: 20 May 2025
Revised: 12 June 2025
Accepted: 13 June 2025
Published: 13 June 2025

Citation: Gao, F.; Tian, Y.; Wu, Y.; Zhang, Y. ST-YOLOv8: Small-Target Ship Detection in SAR Images Targeting Specific Marine Environments. *Appl. Sci.* **2025**, *15*, 6666. https://doi.org/10.3390/app15126666

1. Introduction

Maritime vessel monitoring plays a vital role in ensuring maritime safety, protecting marine resources, facilitating crew operations, and supporting rescue missions at sea. Traditionally, monitoring relied on coast guard patrols or visible light technology. However, these approaches often fail to provide effective surveillance due to the unpredictability of natural conditions, such as weather and light interference [1–3]. As a result, developing

reliable vessel monitoring technologies has become an essential task not only for coastal nations but also for global research communities.

Currently, vessel monitoring primarily leverages advanced technologies such as infrared imagery, optical remote sensing, and SAR imagery [4–6]. SAR, an active microwave-based imaging system, has garnered significant attention due to its ability to operate under all weather conditions, detect metallic objects, and capture terrain features. SAR's potential in maritime traffic monitoring, illegal fishing detection [7], and maritime emergency operations has been well-established in the literature [8–10]. The technology's capacity for wide-area, high-resolution remote sensing, and its all-weather, all-day, multi-angle observation capabilities, have propelled advancements in spaceborne and airborne SAR systems.

However, despite the advantages, challenges persist in the effective real-time detection of vessels from SAR images. These challenges include issues such as detecting small and low-resolution targets, which may appear as mere bright spots, leading to missed detections or false positives [11,12]. Additionally, the varying scale of ship targets in SAR images further complicates detection efforts, emphasizing the need for more effective detection algorithms and end-to-end solutions [13,14].

The evolution of vessel detection algorithms can be divided into two main periods. Initially [15], traditional detection methods based on Constant False Alarm Rate (CFAR) [16–18] relied on expert knowledge and manually extracted features, which limited their robustness in dynamic environments [19]. Although these methods performed well on simple, singular SAR images, their effectiveness often fell short in complex maritime environments [20].

With advances in technology, Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in image and video analysis tasks [21], and are widely used in computer vision fields such as semantic segmentation [22], object detection [23], and image description [24]. CNNs can automatically learn stable features of objects from large datasets, reducing reliance on manual feature extraction and enhancing model robustness. Therefore, in the task of vessel detection in SAR images, CNN-based detection algorithms offer significant advantages over traditional methods.

Recent advances in deep learning have driven the development of two categories of CNN-based vessel detection: two-stage detectors and single-stage detectors [25–27]. Two-stage detectors, such as Faster R-CNN [28], Libra R-CNN [29], and Mask R-CNN [30], identify objects through two separate classification and regression processes. In contrast, single-stage detectors, such as the YOLO series [31–33], SSD [34], and FCOS [35], use fully convolutional networks to simultaneously perform classification and regression tasks in an end-to-end manner [36]. This gives them a speed advantage over two-stage detectors, although the latter may have an edge in accuracy.

Due to the vast resolution differences in SAR images, where the largest and smallest pixel areas in the same dataset can differ by nearly a thousandfold, detection models need to handle multi-scale targets effectively. Feature Pyramid Networks (FPNs) [37] are one effective solution to this issue. They extract and represent features hierarchically, efficiently handling images of varying sizes and resolutions. Other algorithms, such as Scale-Invariant Feature Transform (SIFT) [38] and Speeded-Up Robust Features (SURF) [39], also offer multi-scale feature extraction capabilities, but they may be less efficient than FPNs in terms of computational cost and handling image variations.

To further enhance the detection performance of small targets, some studies have modified network connections to enable independent predictions at different feature layers without increasing the computational burden of the original model. Additionally, to address the challenges of multi-scale vessel detection, researchers have developed various enhanced feature fusion networks [40]. For example, Cui et al. proposed the Dense Attention Pyramid

Network (DAPN) [41], which uses the convolutional block attention module (CBAM) to connect the top and bottom parts of the feature pyramid, extracting rich features that incorporate resolution and semantic information to solve multi-scale detection problems. Although DAPN demonstrates moderate performance in scene adaptability, it provides an effective solution for multi-scale vessel detection.

Zhou et al. [42] developed MSSD Net, a novel detector that incorporates the FC-FPN module, an improved FPN that enhances the fusion of feature maps by introducing learnable fusion coefficients. Sun et al. [43] proposed the Bi-DFFM module, which uses both top-down and bottom-up pathways to achieve higher-level feature integration, thereby improving the recognition capability for multi-scale vessels. H. Wu et al. [44] establishes the CTF net by proposing a detection algorithm that combines convolution and transformers to improve SAR ship detection by balancing global and local features. These innovative methods offer new possibilities for improving the efficiency and accuracy of detecting vessels of varying sizes in SAR images. This detector enhances the model's ability to detect vessels of different sizes through techniques like optimizing Feature Pyramid Networks (FPN). Despite these advancements, detecting vessels in complex maritime environments, especially in dense or cluttered areas, remains a significant hurdle. Firstly, many existing multi-scale vessel recognition models achieve high detection accuracy in simple scenarios, such as individual vessels in open waters (as shown in Figure 1a), where interference factors are minimal. However, vessel targets presented in SAR images are far more varied, often moving through complex maritime environments such as offshore operations, nearshore navigation, docking states, national defense missions and civilian rescue operations. The complexity of the background significantly increases when vessels are near the coastline, particularly when multiple vessels are densely arranged (as shown in Figure 1b). This makes it difficult for most models to accurately identify each vessel, leading to missed detections.

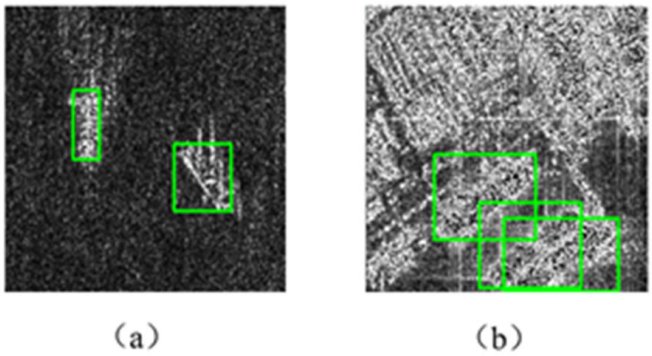

Figure 1. SAR images of different scenarios: (**a**) SAR images of no significant interference; (**b**) SAR images of coastal vessels with complex backgrounds.

Secondly, the detection of small vessel targets is particularly challenging. During the feature extraction process, repeated downsampling often leads to the loss of feature information for small-scale targets. These targets occupy fewer pixels in the image, making it difficult for the model to capture sufficient detail. Therefore, improving the recognition capability for small-scale targets is critical for enhancing the model's performance in multi-scale vessel detection.

To address the challenges in SAR ship detection, this paper introduces Small Target-YOLOv8 (ST-YOLOv8), a specialized architecture optimized for balancing detection accuracy and computational efficiency in marine environments. The selection of YOLOv8 Nano as the baseline model was validated through rigorous benchmarking against contemporary object detectors (Table 1). While Faster R-CNN demonstrates robust performance

for small object detection, its substantial computational cost and parameter count render it unsuitable for resource-constrained SAR applications. RetinaNet achieves comparable inference speeds but requires meticulous tuning of focal loss hyperparameters, exhibiting instability when processing SAR-inherent speckle noise. Although SSD maintains non-negligible throughput, its limited feature pyramid resolution compromises small target detection capabilities.

Table 1. Architectural comparison of object detection frameworks.

Model	FasterR-CNN	RetinaNet	SSD	YOLOv8n
Architecture	Two-Stage	Single-Stage	Single-Stage	Single-Stage
GFLOPs	112	83	31.2	14.5
Parameters (M)	41.3	38.6	26.3	9.3
AP_S (COCO)	34.7	33.2	24.3	32.1
Inference FPS	23	58	89	89
Memory Footprint (MB)	3212	2984	1640	784

In contrast, the adopted YOLOv8 [45] Nano exhibits superior computational efficiency with a lightweight parameter count, while maintaining competitive performance on the AP_S metric of the COCO dataset [46], making it the optimal choice for this study.

Moreover, YOLOv8 excels in multi-scale target detection. It effectively identifies and locates targets of varying sizes, whether they are small objects or large ones occupying most of the image, providing precise detection results. This capability is especially important in complex real-world scenarios, as it can handle targets of various sizes and shapes. Through comparative experiments and ablation studies on the SSDD and SSDv0, we demonstrate the superior performance of ST-YOLOv8 in complex maritime environments. The contributions of this study are as follows:

(1) Introduction of ST-YOLOv8, an advanced vessel detection model tailored for SAR imagery, with innovations in handling multi-scale targets and small object detection.
(2) Optimization of detection accuracy, leveraging techniques C_OREPA Model, ASPP Model, and Shuffle Attention Model to improve feature extraction and model robustness.
(3) Improve the YOLOv8 loss function by using W-IoU [47] to mitigate the impact of low-quality bounding boxes and enhance detection accuracy.

Comprehensive evaluation, through experiments on real-world SAR datasets, demonstrating the effectiveness of the proposed model in complex maritime scenarios. In summary, this study not only advances the field of vessel detection in SAR imagery but also provides a robust solution to the ongoing challenges of multi-scale and small object detection in dynamic maritime environments.

2. Method

2.1. Network Architecture

To more effectively detect multi-scale vessels in SAR image data, especially small vessels in complex backgrounds, we propose an optimized ST-YOLOv8 algorithm for multi-scale vessels and complex scenarios in SAR images. This algorithm maintains robust performance for multi-scale vessel detection. In the backbone network of the baseline model, the C_OREPA module and the ASPP module were integrated, and the SA module was added to the neck network. Firstly, the preprocessed SAR images are passed through a convolutional group and then into the feature extraction network reconstructed by OPERA for feature extraction. Subsequently, these features are processed through the ASPP module for feature fusion, resulting in enhanced spatial and semantic data (P1, P2, P3). Feature maps of various scales, augmented by SA at each stage, are then upsampled before entering

the detection network. Finally, the W-IoU loss function is used to continuously optimize the prediction results. The results of detection are obtained through Non-Maximum Suppression (NMS). Figure 2 shows us the model structure of ST-YOLOv8.

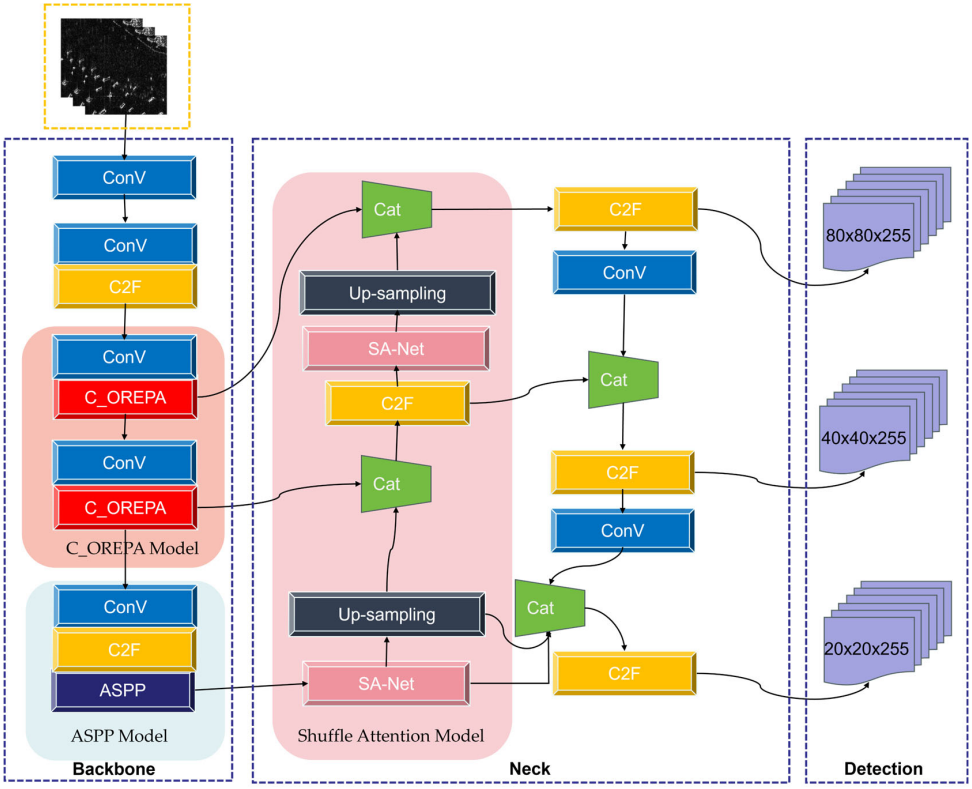

Figure 2. ST-YOLOv8 model diagram.

2.2. C_OREPA Model

Online Reparameterized Convolution (OREPA) [48] is a method designed to reduce the training cost and complexity of deep learning models through online convolutional reparameterization. It primarily involves two stages: firstly, optimizing the performance of online blocks using a specialized linear scaling layer; secondly, reducing training overhead by compressing complex training-time modules into a single convolution. This approach significantly reduces memory and computational costs during training, while also enhancing training speed. Figure 3 provides a detailed description of the composition of the C2f model.

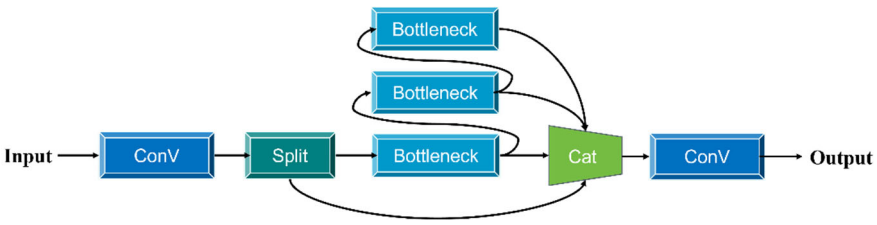

Figure 3. C2f module structure diagram.

The C2f module is an improved convolutional module used in YOLOv8. It enhances the model's performance and efficiency by introducing additional skip connections and Split operations. The C2f module first splits the input tensor into two parts. One part passes directly through multiple bottlenecks, while the other undergoes shortcut connections after each operational layer. The final output is produced through a convolution operation. Figure 4 illustrates how the C_OREPA module works. The core idea of OREPA is to dynam-

ically combine multiple convolution branches during training and reparameterize them into a single standard convolution layer during inference, achieving efficient computation.

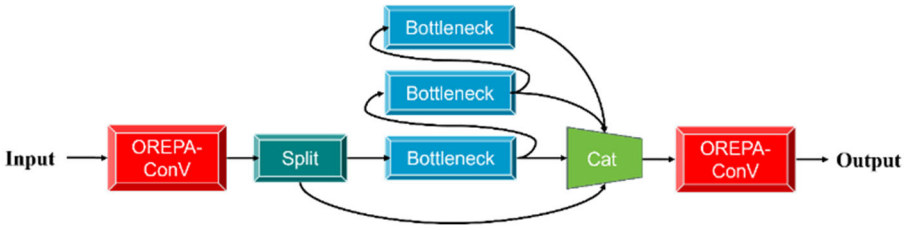

Figure 4. Structure diagram of C_OREPA module.

Here are the formulas:

Dynamic Combination during Training, the weights of multiple convolution branches are dynamically combined using learnable vectors:

$$\text{weight}_{\text{final}} = \Sigma \text{weight}_i \times \text{vector}_i \tag{1}$$

During inference, all branch weights are merged into a single standard convolution weight:

$$\text{weight}_{\text{fin}} = \text{weight}_{\text{ori}} + \text{weight}_{\text{avg}} + \text{weight}_{1\times1} + \cdots \tag{2}$$

This process simplifies complex convolution operations into a single standard convolution.

The proposed C_OREPA module enhances computational efficiency by replacing the original C2f block through decoupling spatial and channel operations via grouped convolutions. This design achieves a notable reduction in parameters while maintaining a high level of original feature expressiveness, as validated through ablation experiments. Furthermore, integrating OREPA's online re-parameterization within the C2f architecture enables replacement of complex convolutions with dynamically re-parameterized kernels. These kernels adaptively adjust weights during training to accommodate varying data distributions, thereby reducing computational complexity and improving end-to-end training/inference performance. This synergistic approach maintains model validity through kernel-level optimization while enhancing adaptability to diverse task requirements.

2.3. ASPP Model

To enhance small object detection performance in YOLOv8, we incorporate the Atrous Spatial Pyramid Pooling (ASPP) module at the backbone–neck junction, which significantly improves multi-scale feature representation capabilities. By replacing conventional pooling operations with atrous convolutions of varying dilation rates (6, 12, 18), ASPP constructs a hierarchical feature pyramid that captures contextual information across scales. This architecture addresses the inherent limitations of traditional pooling layers in small object detection [49].

ASPP (Atrous Spatial Pyramid Pooling), an advanced feature extraction module in deep convolutional neural networks, can be regarded as an enhanced version of traditional pooling layers. Its core design principle originates from an improvement upon Spatial Pyramid Pooling (SPP), where the introduction of atrous convolution (dilated convolution) technology enables multi-scale feature fusion while preserving the spatial resolution of feature maps.

From a functional perspective, ASPP shares the same fundamental objective as conventional pooling layers—to extract comprehensive and efficient deep feature representations of input data through hierarchical feature abstraction. However, unlike traditional pooling layers, which rely on a single fixed-scale feature extraction approach, ASPP constructs a

multi-scale feature extraction pyramid structure by employing parallel atrous convolutional layers with different dilation rates. This innovative design allows the module to simultaneously capture contextual information at varying scales: smaller dilation rates focus on local fine-grained features, while larger dilation rates capture broader semantic contexts.

In terms of implementation, the ASPP module typically consists of four parallel feature extraction branches: a standard 1×1 convolutional branch and three 3×3 atrous convolutional branches with different dilation rates (e.g., 6, 12, 18), supplemented by a global average pooling branch to incorporate image-level contextual information. This multi-scale feature fusion mechanism enhances the network's capability to process visual data with complex spatial structures. By integrating ASPP at the junction between the backbone and neck of YOLOv8, the receptive field is expanded, thereby improving multi-scale object detection performance—particularly in the recognition of small and densely clustered objects. The structure of the ASPP module is illustrated in the following Figure 5.

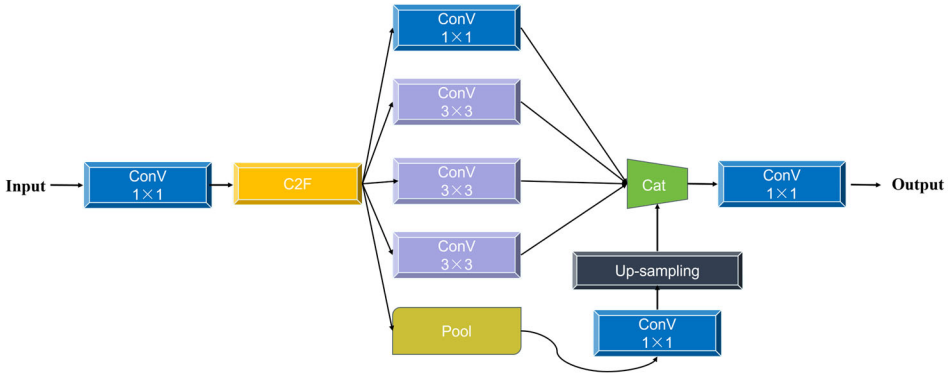

Figure 5. Structure diagram of ASPP module.

2.4. Shuffle Attention Model

Shuffle Attention (SA) introduces a novel mechanism to enhance feature representation learning by dynamically reallocating channel-wise attention weights. Unlike conventional attention modules that employ uniform channel scaling (e.g., SE-Blocks [50]) or sequential channel-spatial processing (e.g., CBAM), SA partitions feature channels into non-overlapping subgroups and computes attention weights in parallel across these subgroups. This group-wise attention computation significantly reduces computational complexity while preserving feature discriminability.

The proposed method comprises two sequential operations, as shown in Figure 6: intra-group attention calibration and inter-group feature shuffling. After parallel attention weighting within each subgroup, a stochastic channel shuffle operation is performed to redistribute features across subgroups. This procedure achieves two key advantages: it promotes feature diversity by enabling cross-group information exchange, and it mitigates overfitting risks through inherent stochasticity in feature redistribution. Experimental evaluations demonstrate that the shuffle operation effectively reduces model reliance on dominant features, thereby enhancing generalization capability [51].

When integrated into the YOLOv8 detection framework, SA yields substantial performance improvements, particularly in complex scenarios. During upsampling stages, the module bridges semantic gaps between multi-scale features through enhanced cross-channel interaction, leading to superior feature fusion quality. Quantitative results show notable detection accuracy gains for small objects and in cluttered environments, attributed to the module's dual benefits of computational efficiency and adaptive feature sampling.

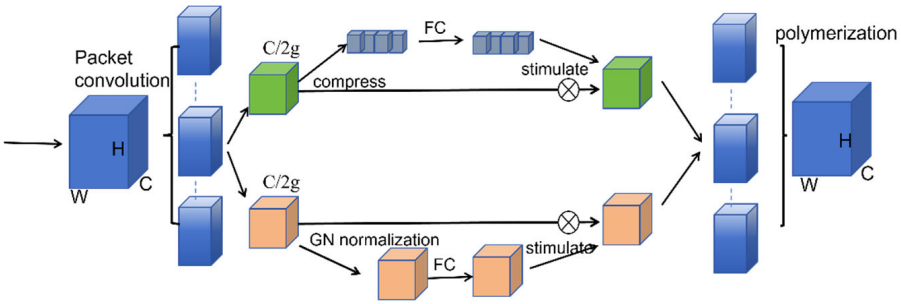

Figure 6. Structure diagram of Shuffle Attention.

2.5. Loss Functions

Object detection, as a core problem in computer vision, heavily relies on the design of its loss functions to achieve optimal performance. As a crucial component of object detection loss functions, a well-designed bounding box loss function can significantly enhance the performance of detection models. Recent studies often assume that the training data samples are of high quality and focus on enhancing the regression capabilities of bounding box loss functions. However, it is noteworthy that object detection training datasets often contain low-quality samples. Overly focusing on regressing bounding boxes for these low-quality samples can negatively impact model performance. Focal-EIoU v1 addresses this issue, but its static focusing mechanism does not fully exploit the potential of non-monotonic focusing mechanisms. Wise-IoU (W-IoU) is a bounding box regression loss function based on a dynamic focusing mechanism, significantly improving the performance of object detection models. Traditional IoU loss functions treat all samples equally during bounding box regression, ignoring the differences in difficulty among samples, while W-IoU dynamically adjusts sample weights, allowing the model to focus more on hard-to-learn samples (e.g., small or occluded targets), thereby enhancing detection accuracy. The core idea of W-IoU is to use a Dynamic Focusing Factor (DFF) to adjust the loss weight, formulated as

$$\mathcal{L}_{WIoU} = \mathcal{R}_{WIoU} \cdot \mathcal{L}_{IoU} \tag{3}$$

$$\mathcal{R}_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{\left(W_g^2 + H_g^2\right)^*}\right) \tag{4}$$

The meanings of each element are as follows:

x and y: The coordinates of the center point of the predicted bounding box.

x_{gt} and y_{gt}: The coordinates of the center point of the ground-truth bounding box.

W_g and H_g: The width and height of the minimum enclosing box, which is the smallest rectangle that contains both the predicted and ground-truth bounding boxes.

$\left(W_g^2 + W_g^2\right)*$: The sum of the squares of the width and height of the minimum enclosing box. The symbol $*$ may indicate a specific processing method, such as taking the average or maximum value, depending on the implementation details.

This formula calculates the squared distance between the center points of the predicted and ground-truth bounding boxes, normalized by the sum of the squares of the width and height of the minimum enclosing box, and then takes the exponential of this value. This term is used to adjust the Wiou loss function, such that the greater the distance between the center points of the predicted and ground-truth bounding boxes, the larger the loss. This encourages the model to predict bounding boxes that are closer to the center of the ground-truth boxes.

$$\mathcal{L}_{IoU} = 1 - \mathrm{IoU}\left(B_p, B_g\right) \tag{5}$$

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{6}$$

$$\text{IoU}(B_p, B_g) = \frac{\text{Area}(B_p \cap B_g)}{\text{Area}(B_p) + \text{Area}(B_g) - \text{Area}(B_p \cap B_g)} \tag{7}$$

Among them, B_p is the hypothetical prediction box, and B_g is the real box
Area ($B_p \cap B_g$) is the intersection area between the predicted box and the real box.
Area (B_p) and Area (B_g) are the areas of the predicted box and the real box, respectively.

3. Experiment and Results

3.1. Experiment Environment and Datasets

All experiments were conducted on a Windows 11 Pro system using PyTorch 1.7.1 with CUDA 11.0 acceleration, equipped with an Intel i5-12400F processor, 16 GB RAM, and an RTX 3060 GPU. Models were trained using SGD optimization with Nesterov momentum (0.937), L2 weight regularization (5×10^{-4}), batch size 32, and initial learning rate 0.01 across 100–300 epochs. Overfitting mitigation included data augmentation, learning rate decay, and early stopping with 15-epoch patience, with hyperparameters refined through iterative validation.

To validate the reliable performance of the model across various datasets, experiments were conducted using the SSDD and the SSDv0 to evaluate the proposed method. In addition to typical open-sea ship views, these two datasets include a variety of scenarios such as nearshore areas, ports, and islands.

To enhance the dataset's complexity and robustness, a series of advanced techniques were implemented. Specifically, data augmentation strategies including random scaling (0.8×–1.2×) and random cropping (85–100% area) were employed to expand the diversity of the SSDD, enriching the range of scenarios and features available for model training. Furthermore, cross-dataset validation was performed by evaluating model performance across both SSDD and SSDv0 datasets, providing a comprehensive assessment of generalization capabilities across different data distributions. Complementary subset analysis was also conducted to examine model behavior under varied operational conditions, offering granular insights into performance characteristics across specific data partitions.

To further examine the robustness of the models, subsets of the SSDD and SSDv datasets were created to simulate diverse environmental conditions. These included variations in weather conditions, sea states, and target densities, which allowed for a thorough assessment of the models' ability to perform reliably under different real-world scenarios. Through these meticulous approaches, a more comprehensive and reliable evaluation framework was established, ensuring that the models could effectively handle the inherent complexity and variability of the datasets. A detailed introduction to SSDD and SSDv0 follows below. Some of the detailed values are presented in Table 2. The annotation boxes in the dataset are all directly provided by the dataset.

The SSDD comprises 1160 SAR images of varying sizes, sourced from three satellite sensors. The image resolution spans from 1 million to 15 million pixels. The dataset encompasses diverse imaging scenarios, including complex environments such as docks and nearshore areas, as well as simpler open-sea scenes. Each SAR image contains a variable number of ship targets, with differing sizes and types. Specifically, small, medium, and large ships constitute 60.2%, 36.8%, and 3% of the total vessels, respectively.

Table 2. Detailed information of the dataset used in this article.

Dataset	SSDD	SSDv0
Sample quantity	1160	43,819
The number of samples for distant sea scenes	932	37,246
Scene type	Dense ship scene, Multi-scale ship scene	Multi-scale ship scene, Strong clutter background scene
The number of ships	2540	59,535
The number of small ships	1529	35,695
The number of medium-sized ships	935	23,660
The number of large ships	76	180
Image size	214 × 653 ~ 190 × 526	256 × 256
Resolution	1 m 10 m	3 m 25 m
Location	Yantai, China, Visakhapatnam, India	Yantai, Shanghai, China Busan, South Korea Tokyo, Japan, etc.
Sensor	Radar Sat-2, Terra SAR-X, Sentinel-1	Gaofen-3, Sentinel-1
Polarization mode	HH, VV, VH, HV	Single, Dual, Full

The SSDv0 dataset is primarily constructed using SAR data from China's Gaofen-3 and Sentinel-1 satellites. Specifically, it consists of 102 scenes from Gaofen-3 and 108 scenes from Sentinel-1 SAR images, forming a high-resolution SAR ship target deep learning sample library. The current library contains 43,819 ship patches. The imaging modes of Gaofen-3 include Strip-Map (UFS), Fine Strip-Map 1 (FSI), Full Polarization 1 (QPSI), Full Polarization 2 (QPSII), and Fine Strip-Map 2 (FSII), with resolutions of 3 m, 5 m, 8 m, 25 m, and 10 m, respectively. Sentinel-1 SAR data, on the other hand, are acquired in Strip Mode (S3 and S6) and Wide Swath Mode. These diverse imaging modes and resolutions provide a rich source of data for deep learning applications in ship detection.

3.2. Experimental Evaluation Indicators

To evaluate the effectiveness of ST-YOLOv8, mean average precision (mAP) [52], precision, recall [53], and F1-score [54] were used as evaluation metrics. The definitions of these metrics are as follows:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP(i) \tag{8}$$

$$AP = \int_0^1 P(R) * dR \tag{9}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$F_1 = 2 * \left(\frac{Precision * Recall}{Precision + Recall}\right) \tag{12}$$

In these formulas, mAP stands for mean average precision, which is calculated as the average of $AP(i)$ (the average precision of the i-th instance or class) over N instances or classes. AP (average precision) is defined as the integral of $P(R)$ (precision at a given recall R) with respect to dR (the differential of recall) from 0 to 1. *Precison* represents the ratio of true positives (TP the number of correctly predicted positive samples) to the sum of true positives and false positives (FP, the number of incorrectly predicted positive samples). *Recall* is the ratio of true positives (TP) to the sum of true positives and false negatives (FN, the number

of incorrectly predicted negative samples). F_1 is the harmonic mean of precision and recall, calculated as 2 times the product of precision and recall divided by the sum of precision and recall, combining the two metrics to evaluate the model's performance comprehensively.

3.3. Results and Discussion

The ablation experiment adopted the YOLOv8 Nano algorithm as the baseline model. (Referred to as YOLOv8n for short in the following text.) Known for its fast and accurate object detection capabilities, YOLOv8n served as the starting point for evaluating other improvement strategies. To ensure the practical applicability of our experimental results, we selected two different datasets for evaluation: SSDD and SSDv0.

The SSDD likely includes target detection samples from various complex environments, whereas the SSDv0 focuses specifically on ship detection tasks. This provides a diverse testing environment to assess the performance of ST-YOLOv8 across different scenarios.

During the experiments, certain components of YOLOv8n were systematically removed or replaced to observe how these changes affected the performance of ST-YOLOv8. Comprehensive tests were conducted on each improvement strategy, including but not limited to detection speed (time required for the model to process an image and produce predictions), accuracy (the proportion of correctly identified targets), recall (the model's ability to identify all targets), and IoU (Intersection over Union, measuring the overlap between predicted and ground truth bounding boxes). Multiple iterations were performed for each strategy to ensure the stability and reliability of the results. Experimental results were organized to show in Tables 3–8 form for easier comparison and analysis.

Table 3. Ablation experiment single-module impact assessment (SSDD).

Dataset	OPERA	SA	ASPP	W-IoU	P	R	F1
SSDD					0.891	0.767	0.823
SSDD	√				0.921	0.776	0.842
SSDD		√			0.913	0.832	0.870
SSDD			√		0.918	0.819	0.866
SSDD				√	0.926	0.812	0.865

Table 4. Ablation experiment progressive module integration analysis (SSDD).

Dataset	OPERA	SA	ASPP	W-IoU	P	R	F1
SSDD					0.891	0.767	0.823
SSDD	√				0.911	0.782	0.841
SSDD	√	√			0.923	0.787	0.849
SSDD	√	√	√		0.931	0.80	0.861
SSDD	√	√	√	√	0.941	0.821	0.876

Table 5. Ablation experiment single-module impact assessment (SSDV0).

Dataset	OPERA	SA	ASPP	W-IoU	P	R	F1
SSDv0					0.833	0.760	0.794
SSDv0	√				0.889	0.762	0.820
SSDv0		√			0.917	0.754	0.827
SSDv0			√		0.864	0.852	0.857
SSDv0				√	0.903	0.821	0.853

Table 6. Ablation experiment progressive module integration analysis (SSDV0).

Dataset	OPERA	SA	ASPP	W-IoU	P	R	F1
SSDv0					0.833	0.760	0.794
SSDv0	√				0.879	0.773	0.823
SSDv0	√	√			0.895	0.794	0.841
SSDv0	√	√	√		0.906	0.851	0.861
SSDv0	√	√	√	√	0.927	0.845	0.881

Table 7. Experimental results of different methods.

Model	Dataset	P	R	F1	mAP	Inference Time (FPS)	Model Size (MB)
Faster R-CNN	SSDD	0.871	0.730	0.794	0.889	35.6	107.0
SSD	SSDD	0.862	0.750	0.802	0.892	38.6	90.6
DAPN	SSDD	0.856	0.766	0.808	0.850	48.4	82.1
Quad-FPN	SSDD	0.923	0.833	0.875	0.935	57.1	56.7
RetinaNet	SSDD	0.912	0.821	0.864	0.934	69.1	528.5
CTF-Net	SSDD	0.909	0.796	0.857	0.911	198.4	25.6
YOLOv5n	SSDD	0.858	0.721	0.783	0.841	318.4	3.9
ST-YOLOv8(ours)	SSDD	0.941	0.820	0.871	0.951	300.5	8.6
Faster R-CNN	SSDv0	0.883	0.710	0.787	0.852	36.2	107.0
SSD	SSDv0	0.897	0.745	0.797	0.864	37.4	90.6
DAPN	SSDv0	0.863	0.773	0.815	0.843	44.6	82.1
Quad-FPN	SSDv0	0.912	0.815	0.861	0.928	59.8	56.7
RetinaNet	SSDv0	0.901	0.806	0.851	0.916	64.8	528.5
CTF-Net	SSDv0	0.912	0.856	0.870	0.927	208.5	25.6
YOLOv5n	SSDv0	0.867	0.786	0.824	0.853	320.8	3.9
ST-YOLOv8(ours)	SSDv0	0.921	0.927	0.881	0.943	299.7	8.6

Table 8. Performance comparison of network models on SSDv0.

Model	Model Size (MB)	Parameter (M)	FLOPs (G)	Speed (fps)
Faster R-CNN	107.0	27.6	459.0	36.2
SSD	90.6	23.6	87.4	37.4
DAPN	82.1	16.8	56.6	44.6
Quad-FPN	56.7	21.5	68.8	59.8
RetinaNet	528.5	145.2	190.6	64.8
CTF-Net	25.6	15.7	60.6	208.5
YOLOv5n	3.9	1.9	4.5	320.8
ST-YOLOv8(ours)	8.6	6.5	10.6	299.7

3.3.1. Ablation Experiment

To systematically evaluate the contributions of individual components and their synergistic interactions within the proposed framework, comprehensive ablation experiments were conducted on two benchmark datasets: SSDD and SSDv0. The experimental protocol was divided into two complementary stages: Single-Module Impact Assessment and Progressive Module Integration Analysis. This two-tiered ablation strategy provides rigorous empirical evidence for both the standalone efficacy and collaborative interactions of the proposed architectural innovations.

Firstly, in this paper, the SPPF module in YOLOv8n is replaced with the ASPP model. In the Ablation Experiment, Table 3 shows that ASPP significantly improves the performance of SAR ship target detection. This improvement is mainly because traditional

downsampling increases the receptive field but reduces spatial resolution. In contrast, dilated convolution expands the receptive field while maintaining resolution. Different dilation rates provide the network with various receptive fields, enabling the model to capture multi-scale contextual information. Table 4 shows that adding the ASSP module improves the recall rate and precision to 93.1% and 80%, respectively, indicating that the SA module reduces missed detections caused by unknown feature extraction in the YOLOv8n algorithm.

Tables 3 and 4 present a performance comparison of different improvement strategies on the SSDD, while Tables 5 and 6 focus on results from the SSDv0. These data include both quantitative metrics, such as accuracy and recall, and qualitative analyses, such as the model's performance in detecting specific object types. Through these detailed experiments and analyses, a scientific basis for further optimization of ST-YOLOv8 was provided, and directions for future research were identified. Specifically, Tables 3 and 5 illustrate the performance changes when individual new modules are incorporated into the base model. This approach allows us to clearly observe the distinct contributions of each module to the overall model performance. Tables 4 and 6, on the other hand, demonstrate the process of incrementally integrating these modules into the complete model. Through comparative analysis, these tables highlight the cumulative enhancement of model performance achieved by the combined integration of these modules.

Finally, the application of the W-IoU loss function introduces a dynamic non-monotonic focusing mechanism and a focusing coefficient. This enables the model to handle small targets more precisely and accurately, thereby improving performance in object detection tasks.

The loss function is considered a fundamental component in object detection tasks, as it quantifies the discrepancy between the predicted and true values, thereby directly influencing the overall performance of the model. Among the various loss components, including box loss, classification loss, and distribution focal loss, superior performance is demonstrated by ST-YOLOv8, primarily due to the incorporation of W-IoU. This enhancement results in more accurate bounding box predictions, which improve both localization and classification precision. When a comparison is made with YOLOv8n, significant improvements are achieved by ST-YOLOv8 across three key evaluation metrics: precision, recall, and mAP. Specifically, detection accuracy is enhanced, and recall is improved, ensuring that fewer objects are missed during detection. Additionally, the increase in mean average precision reflects a more robust model that excels in both the quality and quantity of detections. The marked improvements in these metrics are clearly illustrated in Figure 7, demonstrating the effectiveness of the proposed enhancements and highlighting the significant performance gains of ST-YOLOv8 over its predecessor, YOLOv8n. The importance of the W-IoU modification in driving improved detection capabilities across diverse object categories and detection scenarios is underscored by this comparative advantage.

The proposed methodology demonstrates enhanced comprehensiveness and integration, rendering it particularly advantageous for the training process. As illustrated in the comparative analyses Figure 8, ST-YOLOv8 exhibits superior performance relative to YOLOv8n across all evaluated metrics. Figure 8 presents the detection performance of both models when applied to diverse SAR ship imagery datasets. Notably, under complex environmental conditions, ST-YOLOv8 demonstrates significantly improved capability in detecting multi-scale ship targets compared to the baseline YOLOv8n architecture.

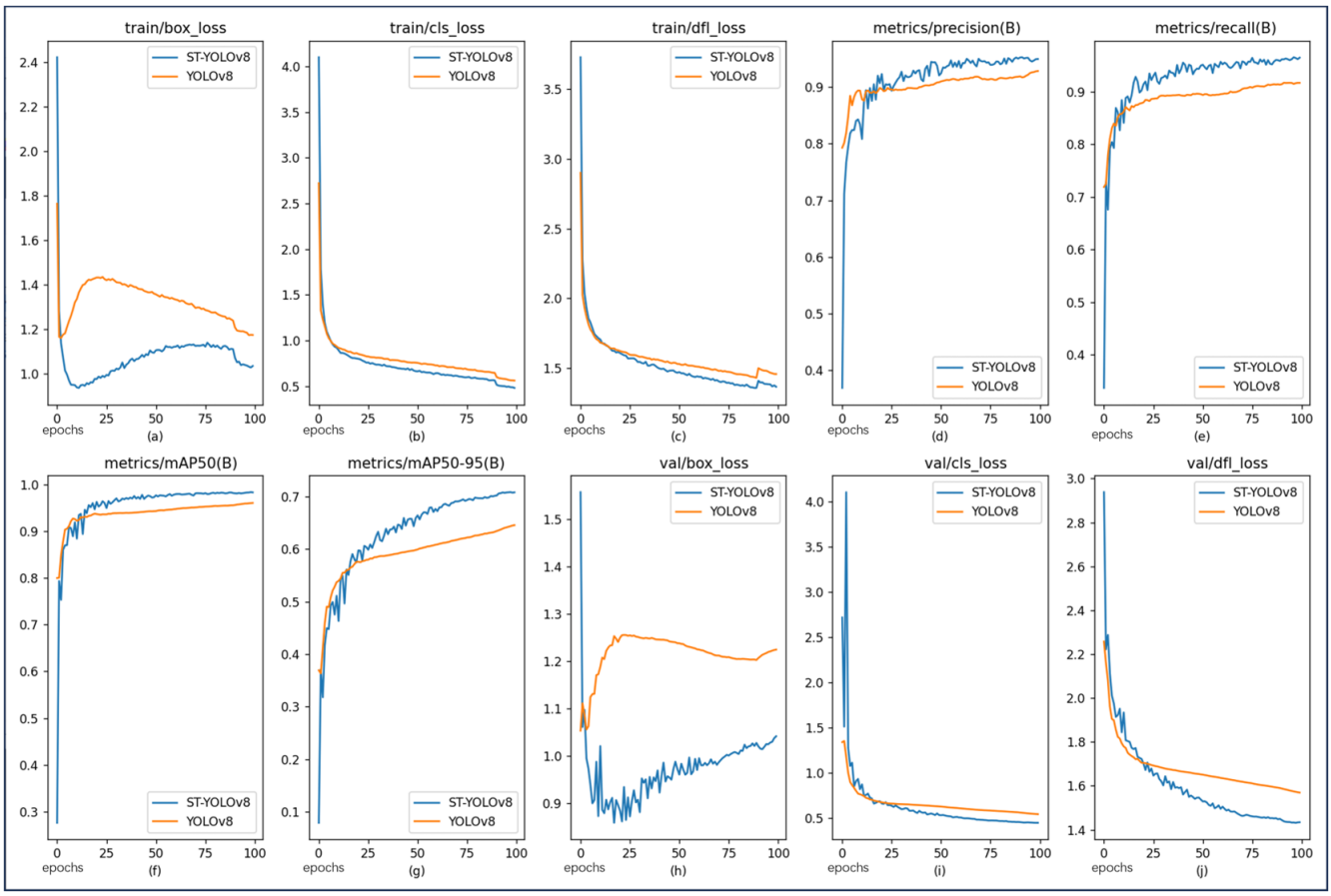

Figure 7. Comparison of loss function indicators between ST-YOLOv8 and YOLOv8n: (**a**) box loss during training; (**b**) classification loss during training; (**c**) distribution focal loss during training; (**d**) precision; (**e**) recall; (**f**) when the IoU threshold is 0.5, the average accuracy of the model; (**g**) the average accuracy of the model within the IoU threshold range of 0.5 to 0.95; (**h**) box loss during validation; (**i**) classification loss during validation; (**j**) distribution focal loss during validation.

This study further validates the effectiveness of the proposed algorithm for SSDv0 detection, with results illustrated in Figure 8. Among them, the first row is for small targets in large sea areas, and the second row is for small targets with complex backgrounds. The comparison demonstrates that ST-YOLOv8 achieves superior recognition accuracy. Figure 8a shows the true distribution of ships in the image. As shown in Figure 8b, YOLOv8n fails to detect small and medium-sized ships (highlighted by the blue circle), primarily because their limited pixel coverage leads to false negatives. In contrast, Figure 8c presents the results of our proposed ST-YOLOv8 model, which accurately identifies small ship targets with higher confidence.

For multi-scale SAR images of docked ships with complex backgrounds (e.g., medium-sized ships near the shore), Figure 8e reveals that YOLOv8n not only misses ships at the image boundary but also generates false positives by misclassifying coastal structures as vessels (marked in yellow). This error stems from the model's susceptibility to interference from cluttered dock backgrounds.

Compared to the YOLOv8n model, the proposed model incorporates SA, enabling it to focus more on ship features and ignore irrelevant background information. Therefore, this model has higher recognition accuracy when facing the pier background. In the SAR images of densely arranged multi-scale ships, the experimental results indicate that the YOLOv8n model exhibits more severe misdetection, represented by yellow circles, when it encounters ships closer to the shore.

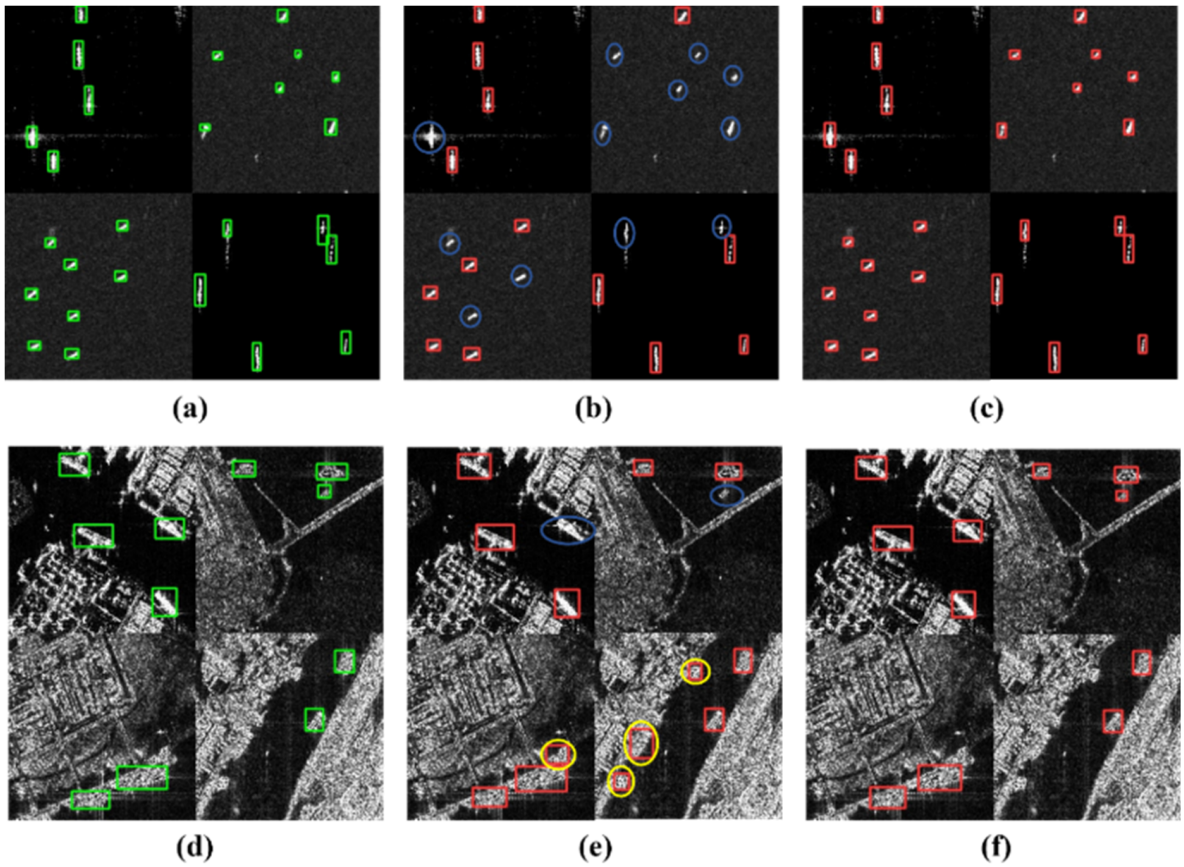

Figure 8. Comparison of detection results: (**a**) real and accurate box for small and medium-sized targets in the sea area (green box); (**b**) performance of YOLOv8n baseline model in the state of small targets in large sea areas; (**c**) performance of ST-YOLOv8 model in the state of small targets in large sea areas; (**d**) the true position of a ship in complex background situations such as docking and port (green box); (**e**) performance of YOLOv8n baseline model in complex backgrounds; (**f**) performance of ST-YOLOv8 model in complex backgrounds.

The ST-YOLOv8 model demonstrates better performance when dealing with densely parked multi-scale ships, as the model combines depthwise separable convolutions with different dilation rates and atrous convolutions, which collectively improve its multi-scale feature extraction capability. Even when faced with overlapping small targets, the model maintains good recognition performance.

In the validation experiments on the SSDv0 dataset, as illustrated in Figure 9, a visual analysis was conducted on 16 representative image tiles (average size: 256 × 256 pixels) containing densely distributed small ship targets (ranging from 8 × 8 to 32 × 32 pixels). The ST-YOLO model successfully detected all small ship targets across these tiles, identifying a total of 26 valid objects (with all targets in each image being correctly recognized), significantly reducing the missed detection rate compared to the baseline model. The Intersection over Union (IoU) values predominantly exceeded 0.6. Notably, in the complex sea clutter background of the second tile from the left in the fourth row, ST-YOLO effectively distinguished a blurred ship target measuring 16 × 20 pixels (confidence score: 0.4), demonstrating the efficacy of the proposed multi-scale feature enhancement module in extracting small target features. Quantitative analysis revealed that this set of tiles achieved the highest average precision, further highlighting the superior performance of ST-YOLO in small ship target detection tasks.

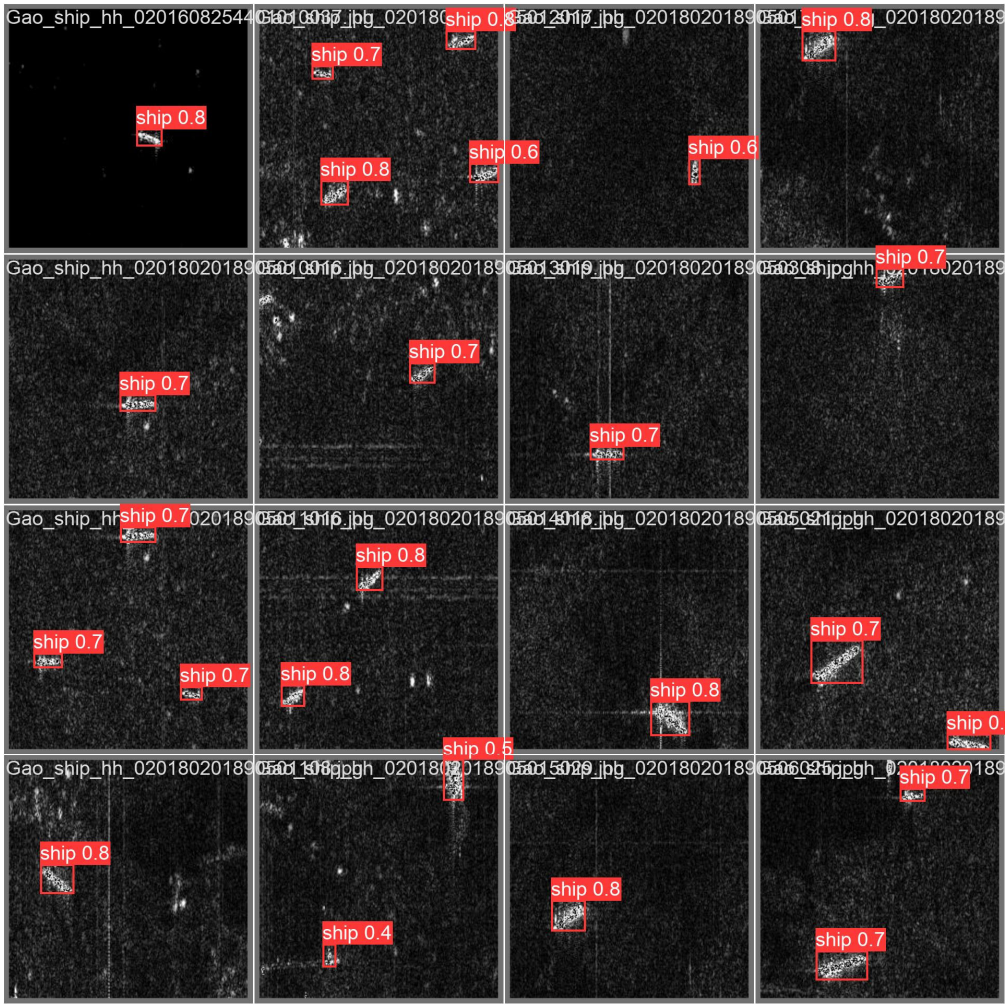

Figure 9. Verification results of small target ship image tiles in the SSDv0 dataset.

3.3.2. Comparison with Other Methods

This section presents comparative experiments using five popular ship detection models: Faster R-CNN, SSD, DAPN, Quad-FPN, Retina Net, CTF-Net, and YOLOv5 on the SSDD and SSDv0, alongside the proposed ST-YOLOv8 model. Faster R-CNN improves the R-CNN series algorithms by introducing the Region Proposal Network (RPN) to reduce the computational cost of candidate regions and improve detection speed. SSD is a single-shot detection algorithm that performs target detection on feature maps of different scales, capable of detecting objects of varying sizes. DAPN enhances the model's ability to perceive important features by introducing channel attention and spatial attention mechanisms, improving detection accuracy, and is used for image recognition and target detection. Quad-FPN is a variant of the Feature Pyramid Network (FPN) used for object detection tasks. It enhances the model's ability to detect multi-scale targets by constructing a four-layer feature pyramid to fuse features at different scales. It boosts the model's ability to detect multi-scale targets by constructing a four-layer feature pyramid to fuse features at different scales. Focal loss helps the model focus more on hard-to-classify samples during the training process. These algorithms have their own advantages and application scenarios in the field of object detection, using different mechanisms to improve detection accuracy and speed. The experimental results show that ST-YOLOv8 achieved the best mAP (95.60%) on SSDD and 94.3% on SSDv0.

This is attributed to the reallocation of the model's relevant parameters, elimination of irrelevant information, reduction in false alarm rates for targets such as islands, improve-

ment in recall, and enhanced sensitivity to medium and small ships. These adjustments enable the model to effectively extract discriminative features for detecting multi-scale ships in complex backgrounds. Table 7 presents the actual performance of different models on the SSDD and SSDv0. Table 8 provides a detailed comparison of our proposed model with other state-of-the-art models in terms of model size, number of parameters, computational speed, and floating-point operations. As shown in the table, our model achieves the smallest model size and the fewest number of parameters among all the compared models except YOLOv5n. However, the detection capability of YOLOv5n in Table 7 is not satisfactory. Our model can better demonstrate that it still performs relatively well under the condition of a compact structure, thereby highlighting its superior efficiency. This significant balance between model complexity and computational performance highlights the outstanding capabilities of the model we proposed, positioning it as a highly effective solution compared to other models.

4. Discussion

In the proposed ST-YOLOv8 model, the introduction of key design choices such as the ASPP module and Wise-IoU loss function has led to significant improvements in detection performance, particularly evident in the enhancement of precision, recall, and F1 score for small target ship detection in SAR images. The ASPP module, by capturing multi-scale contextual information through atrous convolutions with varying dilation rates, enables the model to more accurately identify ship targets, thereby reducing false positives and contributing to a noticeable increase in precision. Additionally, by expanding the receptive field, the ASPP module aids the model in capturing smaller targets, leading to an improvement in recall. The combined effect of these enhancements results in a higher F1 score, indicating a better balance between precision and recall.

Furthermore, the adoption of the Wise-IoU loss function further refines the model's detection capabilities. By dynamically adjusting sample weights, the Wise-IoU loss function allows the model to focus more on challenging samples during training, such as small targets or occluded objects. This dynamic focusing mechanism leads to a further increase in precision by reducing the impact of low-quality bounding boxes. Although the improvement in recall is relatively modest, the overall effect is still positive, contributing to a slight increase in recall and a subsequent elevation in the F1 score, demonstrating an optimized balance between precision and recall. Additionally, the model's generalization capabilities are expected to improve further in addressing complex sea conditions and weather variations in SAR images, adapting to diverse application scenarios, particularly in defense monitoring, maritime security, and ship search and rescue. By incorporating domain-specific prior knowledge or adaptive multimodal data fusion techniques, ST-YOLOv8's detection performance can be further enhanced, enabling it to play a key role in more specialized tasks in the future.

Synthetic Aperture Radar (SAR) ship detection has garnered significant attention due to its military applications, such as maritime surveillance and strategic target identification. However, the potential misuse of SAR technology in military contexts poses considerable risks. For instance, unauthorized detection and tracking of civilian vessels could lead to unwarranted harassment or even escalation of conflicts. Additionally, the high-resolution imaging capabilities of SAR systems may inadvertently capture sensitive information, raising privacy concerns.

To mitigate these risks, it is imperative to implement robust data encryption and anonymization techniques to safeguard sensitive information. Additionally, clear regulatory frameworks and ethical guidelines should be established to govern the use of

SAR technology and vessel monitoring systems, ensuring that privacy is respected while maintaining the benefits of advanced maritime surveillance.

5. Conclusions

Maritime SAR images contain ships of various sizes, but existing detection algorithms face challenges in identifying ships of different sizes, particularly smaller vessels in complex scenes. To address this, the ST-YOLOv8 model, which is based on YOLOv8n, is proposed and integrates the C_OREPA module. At the junction of the backbone and neck networks, the SA mechanism and ASPP are incorporated into the model. Spatial and channel attention are combined through the inclusion of the SA mechanism, enhancing accuracy while maintaining low computational cost, and detection performance is significantly improved. The receptive field is expanded by ASPP through global pooling and large dilation rate convolutional kernels. Feature maps from different levels are effectively fused by the model, further improving the detection rate of small vessels in complex scenes. Finally, the W-IoU loss function has been introduced to enhance the recognition of multi-scale ships by focusing on regions containing ship information. Detection accuracy is improved, and background interference is reduced. To improve the detection accuracy of small targets and reduce dependence on environmental variables, the highest detection accuracy and strong generalization capability are demonstrated by ST-YOLOv8 in comparisons on the SSDv0 and SSDD. The low recognition accuracy in complex environments is primarily caused by detection errors and missed detections. Compared to YOLOv8n's accuracy on the SSDD, ST-YOLO improves it from 89% to 94.1%, the recall rate from 76.7% to 82%, and F1 score from 82.3% to 87.6%. On the SSDv0, accuracy increased from 83.3% to 92.7%, recall rate from 76% to 84.5%, and F1 score from 79.4% to 88.1%. The best results are achieved by the ST-YOLOv8 model compared with other popular ship detection algorithms such as Faster R-CNN, SSD, DAPN, Quad-FPN, and Retina Net, with an accuracy of 94.1% and an mAP of 95.6% on the SSDD, and an accuracy of 92.7% and an mAP of 94.3% on the SSDv0. Multi-scale ships in SAR images are effectively detected by ST-YOLOv8, making it crucial for defense and civilian search and rescue missions. In search and rescue missions, the enhancement of model accuracy and the reduction in false alarm rates play a crucial role in improving response times, monitoring illegal activities, and optimizing port management. Specifically, higher model accuracy enables more precise target identification, thereby significantly reducing the time wasted on ineffective searches and shortening the overall response time. Moreover, a lower false alarm rate allows surveillance systems to more accurately detect illegal activities, minimizing unnecessary alerts and resource allocation, and thereby enhancing the efficiency of monitoring. Additionally, in port management, high-precision models can effectively track ship dynamics in real-time, optimize the allocation of port resources, and reduce congestion, ultimately improving overall operational efficiency. These improvements not only enhance the efficiency of mission execution but also provide robust support for safety management in relevant fields.

Author Contributions: Data curation, Y.T., Y.W. and Y.Z.; Investigation, Y.W. and Y.Z.; Methodology, F.G. and Y.T.; Validation, Y.T., Y.W. and Y.Z.; Writing—original draft, F.G. and Y.T.; Writing—review and editing, F.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of Hebei Province (No. F2024210005, No. F2022210023).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the authors, upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhang, Y.; Hao, Y. A Survey of SAR Image Target Detection Based on Convolutional Neural Networks. *Remote Sens.* **2022**, *14*, 6240. [CrossRef]
2. Eldhuset, K. An automatic ship and ship wake detection system for spaceborne SAR images in coastal regions. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 1010–1019. [CrossRef]
3. Ulaby, F.T.; Kouyate, F.; Brisco, B.; Williams, T.H.L. Textural Information in SAR Images. *IEEE Trans. Geosci. Remote Sens.* **1986**, *GE-24*, 235–245. [CrossRef]
4. Tebaldini, S.; Manzoni, M.; Tagliaferri, D.; Rizzi, M.; Monti-Guarnieri, A.V.; Prati, C.M.; Spagnolini, U.; Nicoli, M.; Russo, I.; Mazzucco, C. Sensing the Urban Environment by Automotive SAR Imaging Potentials and Challenges. *Remote Sens.* **2022**, *14*, 3602. [CrossRef]
5. Hovland, H.A.; Johannessen, J.A.; Digranes, G. Slick detection in SAR images. In Proceedings of the IGARSS'94—1994 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 8–12 August 1994; Volume 4, pp. 2038–2040. [CrossRef]
6. Dellepiane, S.G.; Angiati, E. Quality Assessment of Despeckled SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 691–707. [CrossRef]
7. Lattari, F.; Gonzalez Leon, B.; Asaro, F.; Rucci, A.; Prati, C.; Matteucci, M. Deep Learning for SAR Image Despeckling. *Remote Sens.* **2019**, *11*, 1532. [CrossRef]
8. Liu, L.; Lei, B. Can SAR Images and Optical Images Transfer with Each Other. In Proceedings of the IGARSS, Valencia, Spain, 22–27 July 2018; pp. 7019–7022. [CrossRef]
9. Wang, P.; Zhang, H.; Patel, V.M. SAR Image Despeckling Using a Convolutional Neural Network. *IEEE Signal Process. Lett.* **2017**, *24*, 1763–1767. [CrossRef]
10. Ma, J.; Gong, M.; Zhou, Z. Wavelet Fusion on Ratio Images for Change Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 1122–1126. [CrossRef]
11. Chierchia, G.; Cozzolino, D.; Poggi, G.; Verdoliva, L. SAR image despeckling through convolutional neural networks. In Proceedings of the IGARSS, Fort Worth, TX, USA, 23–28 July 2017; pp. 5438–5441. [CrossRef]
12. Reigber, A.; Ferro-Famil, L. Interference suppression in synthesized SAR images. *IEEE Geosci. Remote Sens. Lett.* **2005**, *2*, 45–49. [CrossRef]
13. Chen, S.; Wang, H.; Xu, F.; Jin, Y.-Q. Target Classification Using the Deep Convolutional Networks for SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4806–4817. [CrossRef]
14. Moreira, A.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.; Younis, M.; Lopez-Dekker, P.; Huber, S.; Villano, M.; Pardini, M.; Eineder, M.; et al. Tandem-L: A Highly Innovative Bistatic SAR Mission for Global Observation of Dynamic Processes on the Earth's Surface. *IEEE Geosci. Remote Sens. Mag.* **2015**, *3*, 8–23. [CrossRef]
15. Feng, S.; Fan, Y.; Tang, Y.; Cheng, H.; Zhao, C.; Zhu, Y.; Cheng, C. A Change Detection Method Based on Multi-Scale Adaptive Convolution Kernel Network and Multimodal Conditional Random Field for Multi-Temporal Multispectral Images. *Remote Sens.* **2022**, *14*, 5368. [CrossRef]
16. Zhou, Z.; Chen, J.; Huang, Z.; Lv, J.; Song, J.; Luo, H.; Wu, B.; Li, Y.; Diniz, P.S.R. HRLE-SARDet A Lightweight SAR Target Detection Algorithm Based on Hybrid Representation Learning Enhancement. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5203922. [CrossRef]
17. Yoshida, T.; Ouchi, K. Detection of Ships Cruising in the Azimuth Direction Using Spotlight SAR Images with a Deep Learning Method. *Remote Sens.* **2022**, *14*, 4691. [CrossRef]
18. Brusch, S.; Lehner, S.; Fritz, T.; Soccorsi, M.; Soloviev, A.; van Schie, B. Ship Surveillance with TerraSAR-X. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1092–1103. [CrossRef]
19. Steenson, B.O. Detection Performance of a Mean-Level Threshold. *IEEE Trans. Aerosp. Electron. Syst.* **1968**, *4*, 529–534. [CrossRef]
20. Touzi, R.; Lopes, A.; Bousquet, P. A statistical and geometrical edge detector for SAR images. *IEEE Trans. Geosci. Remote Sens.* **1988**, *26*, 764–773. [CrossRef]
21. Wang, S.; Gao, S.; Zhou, L.; Liu, R.; Zhang, H.; Liu, J.; Jia, Y.; Qian, J. YOLO-SD Small Ship Detection in SAR Images by Multi-Scale Convolution and Feature Transformer Module. *Remote Sens.* **2022**, *14*, 5268. [CrossRef]
22. Yu, X.; Salimpour, S.; Queralta, J.P.; Westerlund, T. General-Purpose Deep Learning Detection and Segmentation Models for Images from a Lidar-Based Camera Sensor. *Sensors* **2023**, *23*, 2936. [CrossRef]

23. Wei, D.; Du, Y.; Du, L.; Li, L. Target Detection Network for SAR Images Based on Semi-Supervised Learning and Attention Mechanism. *Remote Sens.* **2021**, *13*, 2686. [CrossRef]
24. Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; Kuksa, P. Natural language processing (almost) from scratch. *J. Mach. Learn. Res.* **2011**, *12*, 2493–2537.
25. Kang, M.; Ji, K.; Leng, X.; Lin, Z. Contextual Region-Based Convolutional Neural Network with Multilayer Fusion for SAR Ship Detection. *Remote Sens.* **2017**, *9*, 860. [CrossRef]
26. Wang, J.; Cui, Z.; Jiang, T.; Cao, C.; Cao, Z. Lightweight Deep Neural Networks for Ship Target Detection in SAR Imagery. *IEEE Trans. Image Process.* **2023**, *32*, 565–579. [CrossRef] [PubMed]
27. Jiang, Y.; Li, W.; Liu, L. R-CenterNet+ Anchor-Free Detector for Ship Detection in SAR Images. *Sensors* **2021**, *21*, 5693. [CrossRef]
28. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
29. Pang, J.; Chen, K.; Shi, J.; Feng, H.; Ouyang, W.; Lin, D. Libra R-CNN Towards Balanced Learning for Object Detection. In Proceedings of the CVPR, Long Beach, CA, USA, 15–20 June 2019; pp. 821–830. [CrossRef]
30. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397. [CrossRef]
31. Wang, C.Y.; Liao, H.Y.M. YOLOv1 to YOLOv10 The Fastest and Most Accurate Real-time Object Detection Systems. *APSIPA Trans. Signal Inf. Process.* **2024**, *13*, e29. [CrossRef]
32. Farhadi, A.; Redmon, J. Yolov3 An incremental improvement. *arXiv* **2018**, arXiv:1804.02767. [CrossRef]
33. Bochkovskiy, A.; Wang, C.; Liao, H. YOLOv4 Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. [CrossRef]
34. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD Single Shot MultiBox Detector. *ECCV* **2016**, *9905*, 21–37. [CrossRef]
35. Tian, Z.; Chu, X.; Wang, X.; Wei, X.; Shen, C. Fully Convolutional One-Stage 3D Object Detection on LiDAR Range Images. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 34899–34911.
36. Zhang, T.; Zhang, X.; Ke, X.; Zhan, X.; Shi, J.; Wei, S.; Pan, D.; Li, J.; Su, H.; Zhou, Y.; et al. LS-SSDD-v1.0 A Deep Learning Dataset Dedicated to Small Ship Detection from Large-Scale Sentinel-1 SAR Images. *Remote Sens.* **2020**, *12*, 2997. [CrossRef]
37. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 936–944. [CrossRef]
38. Zeng, L.; Zhou, D.; Pan, Q.; Lu, C.; Zhou, Y. Sar Target Detection Based on Psift Feature Clustering. In Proceedings of the IGARSS 2019, Yokohama, Japan, 28 July–2 August 2019; pp. 17–20. [CrossRef]
39. Wang, W.; Cao, T.; Liu, S.; Tu, E. Remote Sensing Image Automatic Registration on Multi-scale Harris-Laplacian. *J. Indian Soc. Remote Sens.* **2015**, *43*, 501–511. [CrossRef]
40. Zhu, H.; Xie, Y.; Huang, H.; Jing, C.; Rong, Y.; Wang, C. DB-YOLO A Duplicate Bilateral YOLO Network for Multi-Scale Ship Detection in SAR Images. *Sensors* **2021**, *21*, 8146. [CrossRef]
41. Cui, Z.; Li, Q.; Cao, Z.; Liu, N. Dense Attention Pyramid Networks for Multi-Scale Ship Detection in SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8983–8997. [CrossRef]
42. Zhou, K.; Zhang, M.; Wang, H.; Tan, J. Ship Detection in SAR Images Based on Multi-Scale Feature Extraction and Adaptive Feature Fusion. *Remote Sens.* **2022**, *14*, 755. [CrossRef]
43. Sun, Z.; Leng, X.; Lei, Y.; Xiong, B.; Ji, K.; Kuang, G. BiFA-YOLO A Novel YOLO-Based Method for Arbitrary-Oriented Ship Detection in High-Resolution SAR Images. *Remote Sens.* **2021**, *13*, 4209. [CrossRef]
44. Wu, H.; Yu, L.; Li, X.; Zhou, L.; Zhang, W.; Bai, G. CTF-Net A Convolutional and Transformer Fusion Network for SAR Ship Detection. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 4010005. [CrossRef]
45. Terven, J.; Córdova-Esparza, D.-M.; Romero-González, J.-A. A Comprehensive Review of YOLO Architectures in Computer Vision from YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [CrossRef]
46. Jain, S.; Dash, S.; Deorari, R. Object detection using coco dataset. In Proceedings of the 2022 International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 6–7 October 2022. [CrossRef]
47. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU bounding box regression loss with dynamic focusing mechanism. *arXiv* **2023**, arXiv:2301.10051.
48. Hu, M.; Feng, J.; Hua, J.; Lai, B.; Huang, J.; Gong, X.; Hua, X.S. Online convolutional re-parameterization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 568–577. [CrossRef]
49. Qiu, Y.; Liu, Y.; Chen, Y.; Zhang, J.; Zhu, J.; Xu, J. A2SPPNet: Attentive atrous spatial pyramid pooling network for salient object detection. *IEEE Trans. Multimed.* **2022**, *25*, 1991–2006. [CrossRef]
50. Gholamalinejad, H.; Khosravi, H. Vehicle classification using a real-time convolutional structure based on DWT pooling layer and SE blocks. *Expert Syst. Appl.* **2021**, *183*, 115420. [CrossRef]

51. Zhang, Q.-L.; Yang, Y.-B. Sa-net: Shuffle attention for deep convolutional neural networks. In Proceedings of the ICASSP, Toronto, ON, Canada, 6–11 June 2021. [CrossRef]
52. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
53. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the Computer Vision–ECCV, Zurich, Switzerland, 6–12 September 2014.
54. van Rijsbergen, C. *Information Retrieval*, 2nd ed.; Buttersworth: London, UK, 1979.