electronics

MDPI

Article

SAR Image Ship Target Detection Based on Receptive Field Enhancement Module and Cross-Layer Feature Fusion

Haokun Zheng, Xiaorong Xue *, Run Yue, Cong Liu and Zheyu Liu

School of Electronics and Information Engineering, Liaoning University of Technology, Jinzhou 121001, China
* Correspondence: xr_986@163.com

Abstract: The interference of natural factors on the sea surface often results in a blurred background in Synthetic Aperture Radar (SAR) ship images, and the detection difficulty is further increased when different types of ships are densely docked together in nearshore scenes. To tackle these hurdles, this paper proposes a target detection model based on YOLOv5s, named YOLO-CLF. Initially, we constructed a Receptive Field Enhancement Module (RFEM) to improve the model's performance in handling blurred background images. Subsequently, considering the situation of dense multi-size ship images, we designed a Cross-Layer Fusion Feature Pyramid Network (CLF-FPN) to aggregate multi-scale features, thereby enhancing detection accuracy. Finally, we introduce a Normalized Wasserstein Distance (NWD) metric to replace the commonly used Intersection over Union (IoU) metric, aiming to improve the detection capability of small targets. Experimental findings show that the enhanced algorithm attains an Average Precision (AP50) of 98.2% and 90.4% on the SSDD and HRSID datasets, respectively, which is an increase of 1.3% and 2.2% compared to the baseline model YOLOv5s. Simultaneously, it has also achieved a significant performance advantage in comparison to some other models.

Keywords: ship detection; SAR images; Feature Pyramid Network; NWD metric

Citation: Zheng, H.; Xue, X.; Yue, R.; Liu, C.; Liu, Z. SAR Image Ship Target Detection Based on Receptive Field Enhancement Module and Cross-Layer Feature Fusion. *Electronics* **2024**, *13*, 167. https://doi.org/10.3390/electronics13010167

Academic Editor: Byung-Gyu Kim

Received: 12 November 2023
Revised: 18 December 2023
Accepted: 26 December 2023
Published: 29 December 2023

1. Introduction

Synthetic Aperture Radar (SAR) is a remote sensing system that operates effectively under all weather conditions and at all times; its primary function is terrestrial monitoring, but its unique attributes have led to its extensive application in marine surveillance [1–4]. The marine economy's progression has led to an annual surge in the number of vessels navigating across diverse sea regions globally. Consequently, ship detection technology plays a crucial role in modern maritime surveillance and maritime safety. SAR images overcome the issues of optical remote sensing images, which can be affected by time and weather conditions. They possess unique characteristics that allow them to penetrate cloud cover and operate in darkness, offering better adaptability for sea surface scenarios [5]. Therefore, studying how to handle ship target detection under various sea surface scenarios has significant practical significance and application value [6–8].

In traditional SAR detection tasks, the Constant False Alarm Rate (CFAR) method is most widely used [9–11]. CFAR is a dynamic threshold technology that adaptively adjusts the detection threshold according to the statistical characteristics of background interference to achieve the required false alarm rate. The balance between the false alarm rate and the detection rate is a challenge in the design of a constant false alarm rate target detection system. A lower false alarm rate implies that the system can more accurately identify real targets without false alarms, but it may also lead to an increase in undetected real targets. Conversely, a higher false alarm rate may improve the detection rate but will also increase the number of false alarms. Therefore, the performance of CFAR highly depends on the threshold setting. Moreover, if the signal characteristics of the ship are highly similar to the background, or when the background interference changes significantly, it may

lead to instability in the false alarm rate. Extracting ship feature signals then becomes extremely difficult, inevitably leading to false alarms. To reduce the false alarm rate and improve detection accuracy, some researchers have improved the traditional CFAR method, achieving good results [12–14].

In contrast to conventional Constant False Alarm Rate (CFAR) techniques, deep learning methodologies exhibit substantial benefits in the realm of target detection tasks [15]. Current deep learning-oriented target detection methodologies can be broadly bifurcated into two categories: those that are based on region proposal, which are two-stage detection algorithms, and those that are founded on regression, which are single-stage detection algorithms. Two-stage target detection algorithms bifurcate the detection task into two stages. Initially, they engender a collection of prospective target zones. Subsequently, these candidate areas undergo classification to ascertain the presence of targets, followed by regression to pinpoint the precise location of the target. Representative algorithms include the R-CNN series [16–18]. Single-stage target detection algorithms directly apply classifiers to the entire image, performing target classification and location regression simultaneously. Representative algorithms include SSD [19], RetinaNet [20], YOLO [21], etc. Deep learning detection methods have significantly improved accuracy compared to traditional methods. Nevertheless, owing to certain intrinsic attributes of SAR ship images, the task of SAR ship detection still faces the following challenges.

(1) SAR ship images are easily affected by natural factors on the sea surface, such as waves and ripples, as well as terrestrial backgrounds, such as land and ports. This interference causes ships to be difficult to separate from the background, and the features of the ships themselves are not prominent. Figure 1a shows an undisturbed ship target with a clear ship surface contour and distinct foreground and background. Figure 1b shows a ship target against a blurred sea surface background, where the ship and the background merge into one in the image, making it difficult to distinguish.

(2) Due to the coexistence of ship targets of different sizes in the same image, the ship targets in the same dataset cover a wide range of sizes. Ship targets of varying sizes in SAR have distinct radar reflection characteristics, leading to numerous missed detections and false alarms in existing methods. In addition, unlike conventional detection targets, multi-size ship targets often dock densely at portcs. As shown in Figure 1c, ships of varying sizes dock densely together, further increasing the detection difficulty.

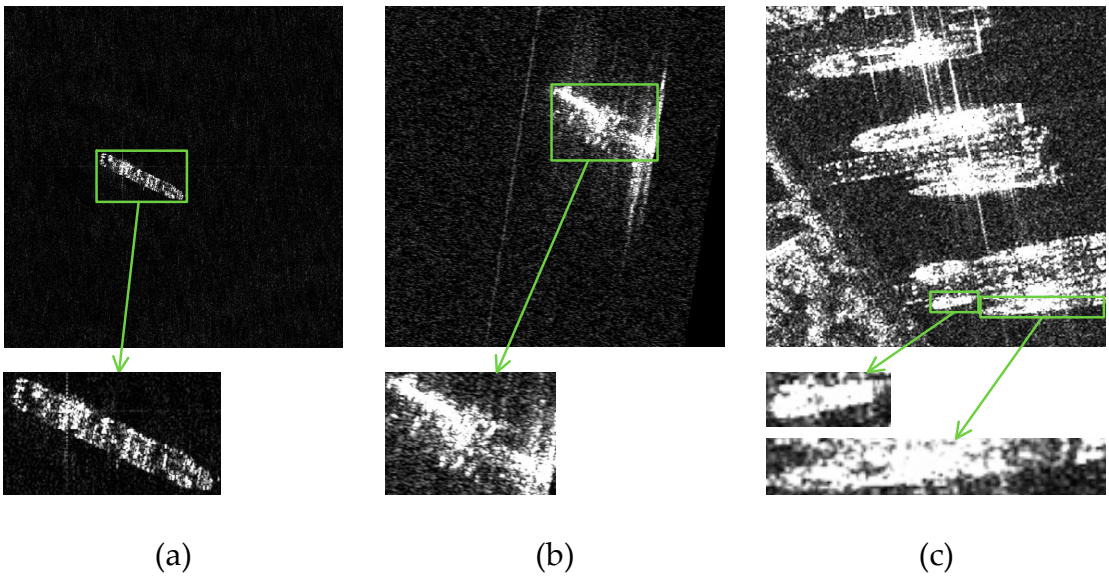

Figure 1. Ship targets in different scenarios: (**a**) conventional undisturbed targets; (**b**) fuzzy background targets; (**c**) near-shore dense multi-size ship targets.

In this paper, following an examination of the evolution of contemporary deep learning target detection algorithms and taking into account the SAR ship images, we propose the YOLO-CLF architecture based on YOLOv5 to address the above problems. The principal contributions of this paper can be summarized as follows:

(1) For the problem of blurred backgrounds in SAR ship images caused by interference from natural and terrestrial factors, we propose a Receptive Field Enhancement Module (RFEM). By utilizing the large receptive field characteristics of dilated convolution, we enhance the model's feature extraction ability for ship targets with blurred backgrounds.
(2) We propose a Cross-Layer Fusion Feature Pyramid Network (CLF-FPN). Through cross-channel concatenation of multiple feature layers and the use of Coordinate Attention (CA) modules after concatenation, we achieve an improvement in multi-scale feature fusion effects, enhancing the model's detection performance for multi-size targets.
(3) We introduce the NWD loss function to reduce the model's sensitivity to small target position deviations and improve the accuracy of bounding box regression.

This paper is divided into five sections, each playing a role in exploring the content related to the main theme of the article. The subsequent Section 2 is dedicated to introducing the latest advancements in ship target detection, providing a contemporary backdrop for the ensuing discussions. Section 3 extensively explores the proposed method, intricately detailing its design and functionality. In Section 4, experimental details and an analysis of the experimental results are included. The final summary is encapsulated in Section 5.

2. Related Work

In this section, a review of the current mainstream SAR ship detection methods is provided, followed by discussions on bounding box regression losses and multi-scale feature fusion.

2.1. CNN-Based SAR Ship Detector

In recent years, a multitude of researchers have put forth a variety of Synthetic Aperture Radar (SAR) ship detection algorithms, each tailored to distinct detection scenarios. Kang et al. [22] combined traditional CFAR algorithms with deep learning, using faster R-CNN as the protection window for CFAR. As a result, smaller and structurally ambiguous targets are often falsely reported. To address this issue, the authors proposed utilizing the CFAR algorithm for target detection. This method aims to reduce misclassification of smaller, structurally ambiguous targets by determining the detection threshold based on pixel intensities within the target and background regions. However, combining these two methods requires a considerable amount of experimentation and tuning to achieve optimal performance. Zhang et al. [23] proposed a detection network that includes a novel Spatial Cross-Scale Attention (SCSA) module. This innovative module utilizes features from each scale of the backbone's output to calculate the network's spatial focal points, aiming to eliminate interference from noise and complex land backgrounds. This method addresses challenges posed by complex backgrounds in near-shore scenes containing ship targets and densely arranged ships. However, the module occupies a large number of parameters, and its impact on computational resources is significant and cannot be ignored. Pan et al. [24] designed a ship anomaly detection method for feature learning using a SuperPixel (SP) processing unit. This method performs multi-feature extraction on SP units, enhancing information discrimination capabilities and improving the clutter feature learning (COFL) strategy for efficient classification of ships and clutter. However, the method still lacks detection performance for small-sized, weakly scattering ship targets. Hou et al. [25] developed a fully automated program dedicated to the automatic matching and annotation of samples, leading to the establishment of a high-resolution FUSAR-Ship dataset specifically tailored for ship detection and recognition. Additionally, Aao et al. [26] devised a multi-scale CFAR detector aimed at detecting candidate targets andefficiently

filtering potential targets. Subsequently, the utilization of feature ellipses in conjunction with the maximum likelihood method for identification effectively eliminated non-ship objects. This sophisticated ship detection algorithm demonstrates adaptability across diverse and complex environmental conditions. Currently, anchor-based methods still have some drawbacks, such as difficulty in assigning anchor boxes for each target when overlaps exist between targets. To address this issue, Zhu et al. [27] introduced an enhanced anchor-free detector based on the FCOS + ATSS network. This paper embeds an improved residual module (IRM) and deformable convolutions (Dconv) into the feature extraction network (FEN) to enhance accuracy. Additionally, it proposes a novel joint representation of classification scores and localization quality, coupled with a redesigned detection process to refine localization performance. Cui et al. [28] introduced a method of spatial random group enhanced attention in CenterNet to extract stronger semantic features while suppressing certain noise, aiming to reduce false positives caused by nearshore and inland interferences. The anchor-free detector does not require predefining a set of prior boxes, which makes the model more flexible and not limited by the quantity and shape of anchor boxes. However, this can also result in some difficulties in detecting occluded or overlapped regions.

2.2. The Loss of Bounding Box Regression

IoU is the most commonly used metric for measuring the similarity between bounding boxes. However, IoU fails to function if two bounding boxes do not intersect. CIOU [29] addresses the limitations of IoU by considering properties such as the distance between the centers of true and predicted bounding boxes, and the overlapping area. SIoU [30] aids in the rapid convergence of the network by considering the vector angle between true and predicted bounding boxes. Qin et al. [31] enhanced SAR near-shore ship target detection accuracy by introducing a factor related to angles, computed through weighted correlations of aspect ratios and center distances, describing the rotation of ship targets. Xu et al. [32] proposed TDIoU loss specifically designed for rotational bounding box regression in multi-angle SAR ship target detection. However, slight deviations in bounding boxes significantly alter the IoU for small targets. Our motivation lies in utilizing superior evaluation metrics to address convergence challenges in bounding box regression for small objects.

2.3. Multi-Scale Feature Fusion

Due to the lack of feature detail information in small-scale ship targets, there are often missed detections and false alarms. Therefore, designing detection algorithms that adapt to multi-scale targets has been a hot research direction in recent years. Since Lin et al. [33] proposed the Feature Pyramid Network (FPN), it has been the mainstream method for solving multi-scale problems, and many researchers have improved upon it. NAS-FPN [34] utilizes neural architecture search algorithms to automatically discover feature pyramid connections suitable for specific tasks. However, the neural architecture search process requires significant computational resources and time. Bi-FPN [35] introduces bidirectional network connections and feature fusion strategies, enabling information propagation within the network both bottom-up and top-down. This bidirectional structure enhances information flow and feature reuse within the network, but its performance might heavily rely on the characteristics of the training dataset and task. Due to longer paths between high-resolution and low-resolution features, semantics can be weakened during propagation, which is disadvantageous for the detection of SAR ships with specific scales and shapes. Sun et al. [36] introduced the proposed Bi-directional Feature Fusion Module (Bi-DFFM) into the YOLO framework. The introduced module adeptly consolidates features across various scales by employing a bidirectional mechanism that fosters information interaction from both top-down and bottom-up perspectives. This strategic approach significantly enhances the capability to discern ships of diverse scales. Zhou et al. [37] proposed a method that synergistically leverages the advantages of both Doppler features, capable of capturing motion-related information, and spatial features, which encompass crucial background and positional characteristics essential for accurate ship detection. This approach aims to

address challenges in ship detection, such as multi-scale transformations of ships within images and image blurring caused by motion. When constructing a feature pyramid, feature information may become blurred or lost, necessitating appropriate improvements tailored to the specific task at hand. In summary, the introduction of FPN brings more possibilities to the field of target detection. In our design, the improvement of FPN further enhances the detection performance of multi-scale targets.

3. Methods

Considering the characteristics of SAR ship images, this paper improves upon YOLOv5s and proposes a target detection method based on the Receptive Field Enhancement Module (RFEM) and Cross-layer Fusion Feature Pyramid Network (CLF-FPN)—YOLO-CLF. Figure 2 shows the algorithm structure of this paper, which is mainly composed of a backbone, neck, and head. We embed the Receptive Field Enhancement Module in the backbone. Next, we improve the structure of FPN and propose CLF-FPN for the head. Finally, we introduce the NWD loss function. This integration allows the model to have a larger receptive field and stronger multi-scale feature fusion capability. These contents will be detailed in the remaining part of this section.

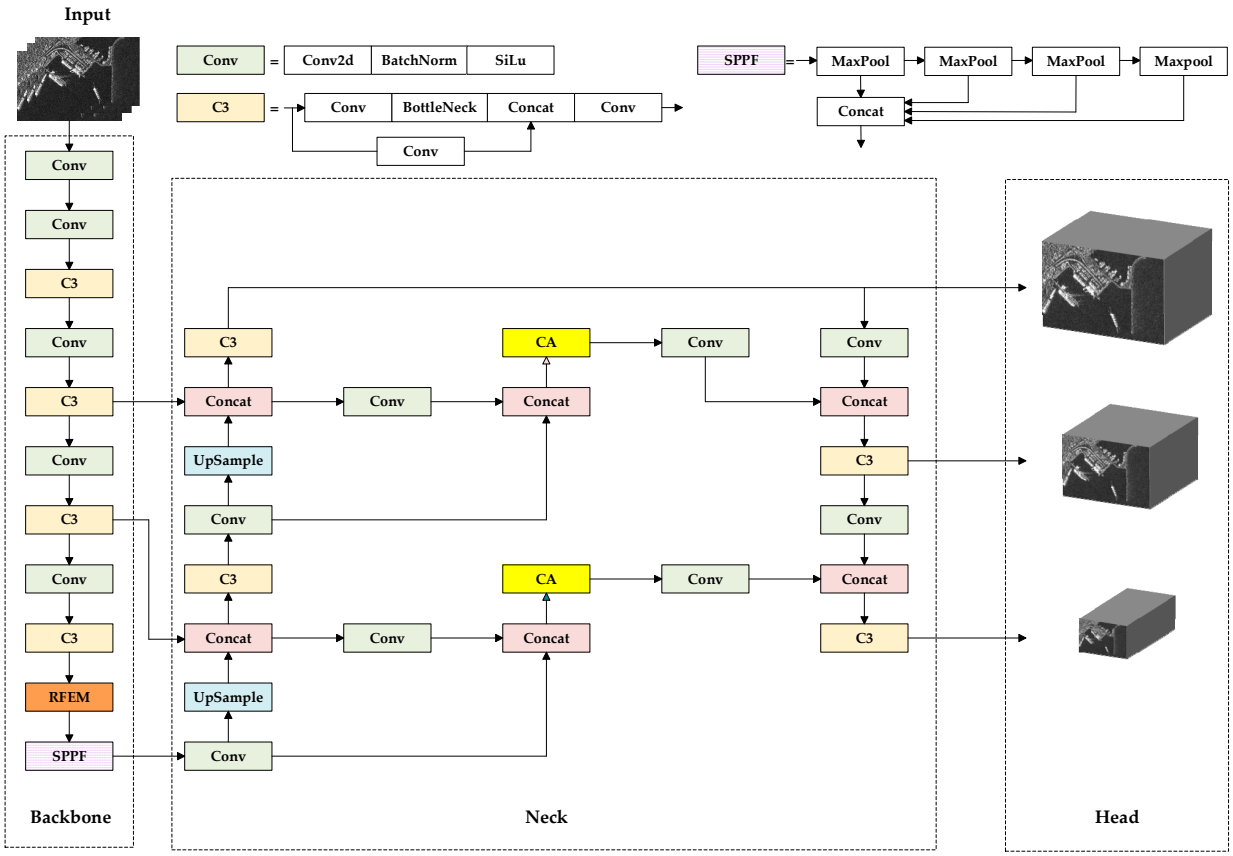

Figure 2. The general structure of the algorithm is described in this paper.

3.1. Receptive Field Enhancement Module

In the process of feature extraction within convolutional neural networks, individual layers are characterized by a predetermined receptive field size. This fixed receptive field size presents limitations for effectively capturing objects across various scales. The static nature of these fields hinders the network's capability to discern and accurately identify objects that exhibit diverse size characteristics within an image, especially for targets in SAR images that have blurred target edges, insufficient texture information, and large-scale changes. It is difficult for the model to detect such targets.

Inspired by the module proposed by Xiao et al. [38], this paper proposes the Receptive Field Enhancement Module (RFEM). RFEM amalgamates the traits of dilated convolution and employs branches of dilated convolution with varying dilation rates. This enhances the model's comprehension of contextual message information, thereby augmenting the model's detection efficacy. Dilated convolution has a wider receptive field and can perceive more contextual information. However, dilated convolution also has problems. A single dilation rate will cause some pixels to be omitted during the calculation.

An illustration of the RFEM module's network structure can be found in Figure 3. The RFEM module employs three sets of dilated convolutions with dilation rates of 3, 5, and 7 respectively to perform convolution operations on the input feature map. Varied dilation rates result in different degrees of receptive field enlargement, catering to diverse scales of feature learning and application. This approach prevents the oversight of individual pixels during computation by avoiding a singular dilation rate. Leveraging higher dilation rates in dilated convolutions aids neural networks in effectively capturing relationships between distant pixels, thereby enhancing the detection and identification of object boundaries, textures, and intricate details. The process unfolds in a sequence where each branch initially applies a 1×1 convolutional kernel to efficiently reduce the channel number within the feature map. Subsequently, dilated convolutions are performed independently within each branch before concatenating the resulting feature maps. Finally, a shortcut structure is incorporated to mitigate potential information loss during the convolutional process, ensuring the preservation of crucial data integrity.

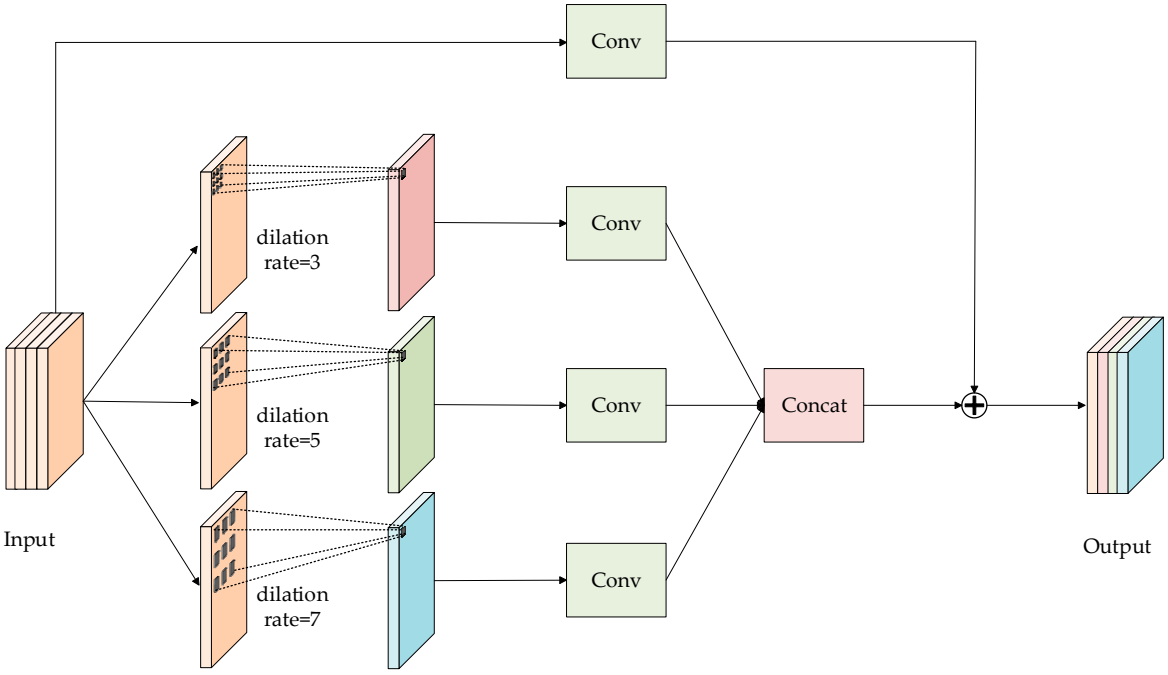

Figure 3. Structural diagram of the RFEM module.

For the fusion method of the three branch feature maps, attention fusion and adaptive fusion can also be used, as shown in Figure 4. Figure 4a is the previously mentioned concatenation fusion, which involves combining feature maps from different branches. Figure 4b shows attention fusion. This approach employs attention mechanisms to amplify beneficial feature channels while attenuating those that are deemed unproductive, retaining key feature information and playing a role in feature screening and enhancement. Figure 4c is adaptive fusion, which uses convolution and Softmax to map the dimension-reduced feature map to the weights corresponding to the three branches. In Section 4.4.1 of this paper, we compared the three fusion methods and chose the concatenation fusion method, which has better results with a simple structure.

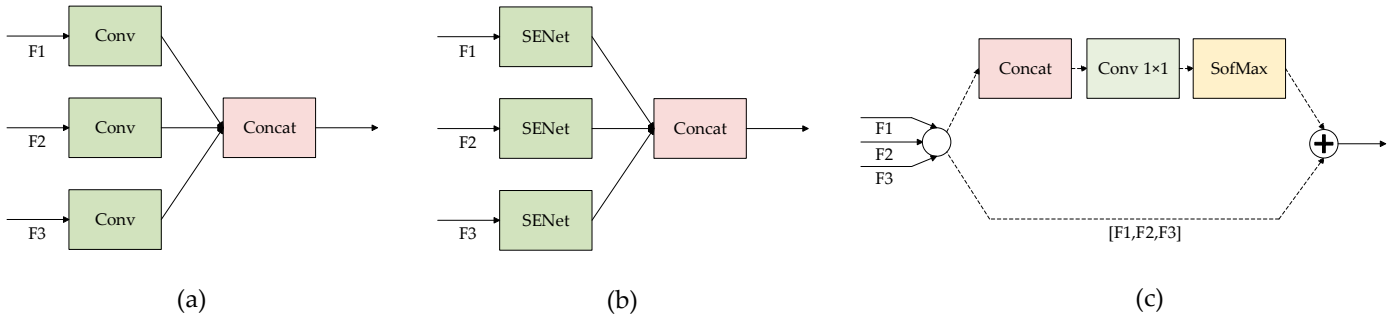

Figure 4. Different fusion methods: (**a**) Concatenation Fusion; (**b**) Attentional fusion; (**c**) Adaptive fusion.

3.2. Cross-Layer Fusion Feature Pyramid Network

In neural networks, shallow features have higher resolution and more detailed information, such as edges and textures. Deep features have lower resolution and rich semantic information, such as shape and model. By fusing shallow features and deep features, the model's feature expression ability can be enhanced, thereby improving the model's performance. FPN can fuse feature maps from different levels together, allowing the model to pay attention to different scale information in the image at the same time. PAN [39] first proposed the concept of feature bidirectional fusion, improving feature reuse. Since then, many researchers have gone further and tried more complex bidirectional fusion networks, such as AFPN [40] and ASFF [41]. In order to enhance the model's ability to integrate both shallow and deep information, we propose the Cross-layer Fusion Feature Pyramid Network (CLF-FPN), which is capable of incorporating multi-scale information. This network builds upon the original feature pyramid network of YOLOv5. Figure 5a illustrates the FPN of YOLOv5, while Figure 5b presents the overall network structure of CLF-FPN.

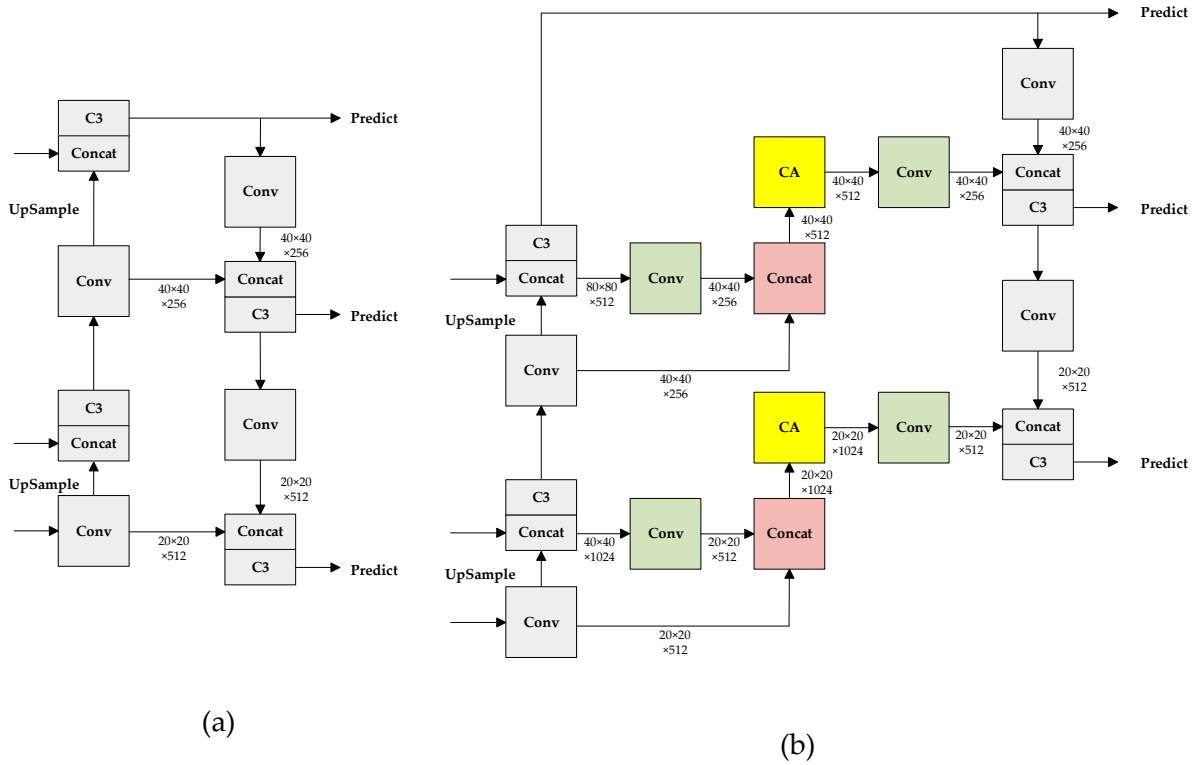

Figure 5. (**a**) The feature pyramid structure of YOLOv5; (**b**) The network architecture of CLF-FPN.

CLF-FPN establishes additional connections between feature layers, enabling low-resolution feature maps to receive feature information from higher-resolution feature maps, and vice versa. Through these cross-layer connections, low-resolution feature maps with stronger semantic information and high-resolution feature maps with more detailed information can complement each other, thereby providing a more comprehensive feature representation for the model.

CLF-FPN performs convolution on feature maps at different levels, adds additional feature fusion branches, and uses Coordinate Attention (CA) [42] after concatenation with the original branch. Coordinate attention incorporates locational data into channel attention, aggregating features along two spatial trajectories. One of these trajectories is designed to capture dependencies that span over long distances. In contrast, the other trajectory is engineered to retain precise locational information, simply and quickly enhancing the extraction ability of long-distance relationships. Its structure is shown in Figure 6.

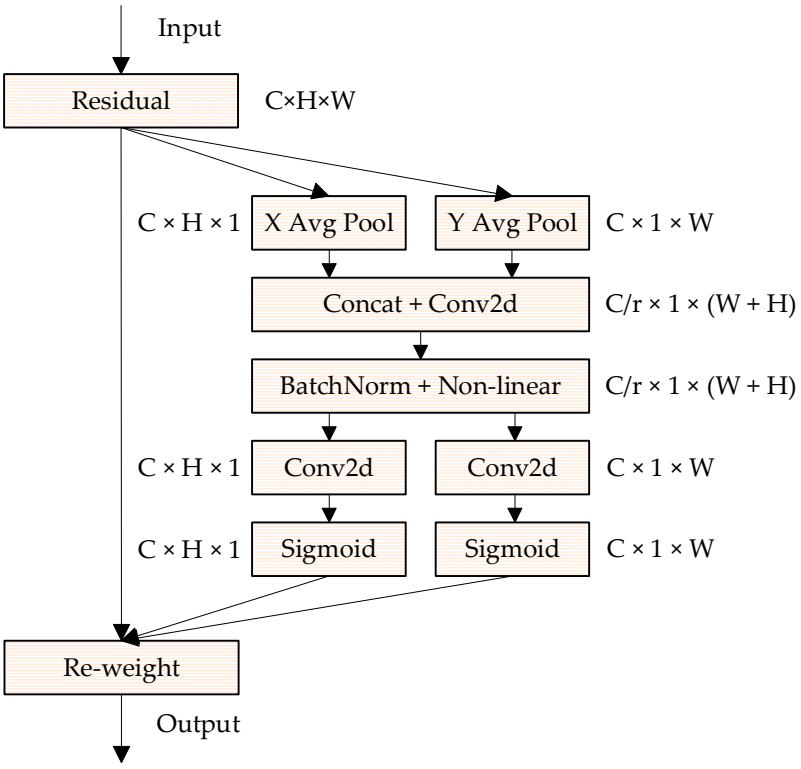

Figure 6. Coordinate Attention Module.

In Coordinate Attention, the input feature map is divided along its height and width directions to perform global average pooling. As indicated by Equations (1) and (2), where C represents the number of channels, H stands for height, and W denotes width, pooling operations are separately conducted along the height and width directions for the input feature map of dimensions C × H × W. This process generates feature maps of dimensions C × H × 1 and C × 1 × W, respectively. This segmentation pooling strategy aids in independently extracting fine-grained positional information along the vertical and horizontal axes, allowing the network to focus on and process specific spatial cues in each direction separately.

$$z_c^h(h) = \frac{1}{W} \sum_{0 \le i < W} x_c(h, i) \tag{1}$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \le j < H} x_c(j, w) \tag{2}$$

Afterward, the feature maps in the height and width directions are concatenated, and after batch normalization and non-linear activation functions, global information is obtained and separated. Finally, after convolution and Sigmoid, the output of Coordinate Attention y_c is represented as:

$$y_c(i,j) = x_c(i,j) \cdot g_c^h(i) \cdot g_c^w(j) \tag{3}$$

where x_c represents the input, and g_c^h and g_c^w represent the attention weights of the input feature map in the height and width directions.

By placing Coordinate Attention after the concatenation of the two branches, more channels can be extracted for attention. Cross-layer fusion can combine information from different-level feature maps, allowing the model to use local information and global information at the same time. These feature information can complement each other and play a key role in feature enhancement and context feature transmission.

3.3. Normalized Wasserstein Distance

The Intersection over Union (IoU) metric, a widely used quantitative measure, computes the ratio of the intersection to the union of the predicted and actual bounding boxes. This metric serves as an effective gauge of the extent of overlap between the predicted and actual bounding boxes, thereby providing a measure of the accuracy of the prediction. The formulas for IoU and IoU loss are as follows:

$$\text{IoU} = \frac{|\, B \cap B_{gt} \,|}{|\, B \cup B_{gt} \,|} \tag{4}$$

$$\text{IoU loss} = 1 - \text{IoU} \tag{5}$$

In this context, B and B_{gt} represent the coordinates of the center points and the width and height of the predicted and actual bounding boxes, respectively.

However, for small targets in SAR ship images, which have fewer pixels, slight positional deviations can cause significant changes in the IoU index, greatly reducing detection accuracy. As can be seen in Figure 7, for instance, an 18 × 6 pixel target A experiences an IoU value of 0.69 with a blue predicted bounding box's positional deviation, resulting in an IoU loss of 0.31. whereas a target B occupying only 5 × 2 pixels shows a reduced IoU value of 0.11 under the same predicted bounding box deviation, leading to an IoU loss of 0.89. This indicates that smaller targets, affected by minor positional deviations in their predicted bounding boxes, undergo a substantial increase in IoU loss, potentially causing instances of missed detection.

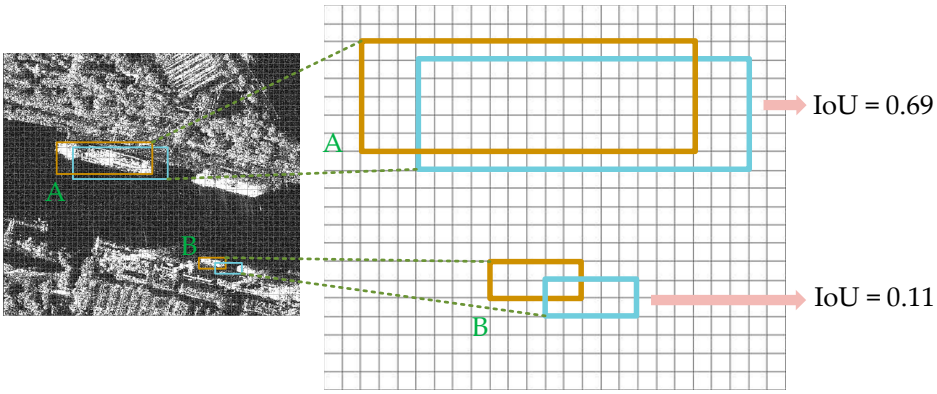

Figure 7. IoU comparison for normal-sized and small-sized targets under the same predicted bounding box deviation. Here, the orange bounding box represents the ground truth box, while the blue bounding box represents the predicted box. Object A and B respectively denote normal-sized and small-sized targets.

In order to enhance the detection performance of small targets, this paper introduces the Normalized Wasserstein Distance (NWD) [43] to measure the similarity between the predicted and actual bounding boxes. NWD abandons the commonly used IoU metric and initially models the bounding box $R = \left[c_x, c_y, \frac{w}{2}, \frac{h}{2}\right]^T$ as a two-dimensional Gaussian distribution:

$$\mu = \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \Sigma = \begin{bmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix} \tag{6}$$

Here, (c_x, c_y) represents the coordinates of the center point, while w and h denote the width and height, respectively. The bounding box is modeled as a two-dimensional Gaussian distribution $N(\mu, \Sigma)$. After modeling, the Normalized Wasserstein Distance is used to calculate the distribution distance between bounding box $A = \left[cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2}\right]^T$ and bounding box $B = \left[cx_b, cy_b, \frac{w_b}{2} \cdot \frac{h_b}{2}\right]^T$. The calculation formula is as follows:

$$W_2^2(N_a, N_b) = \| \left(\left[cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2}\right]^T, \left[cx_b, cy_b, \frac{w_b}{2} \cdot \frac{h_b}{2}\right]^T \right) \|_2^2 \tag{7}$$

However, this formula measures distance and cannot be directly used to measure similarity distance (i.e., a value between 0 and 1 for IoU). Therefore, it needs to be normalized using an exponential operation to obtain the NWD metric:

$$NWD(N_a, N_b) = \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{c}\right) \tag{8}$$

In this formula, C is a constant closely related to the dataset. Compared to IoU, NWD compensates for its sensitivity to small-scale targets, making it more suitable for detecting small objects.

4. Experiments

4.1. Dataset

The effectiveness of the proposed model are evaluated using two datasets: SSDD [44] and HRSID [45]. The SSDD dataset comprises 1160 Synthetic Aperture Radar (SAR) images from three different SAR satellites, containing 2587 ship targets. The resolution of these images ranges from 1 to 15 m. The dataset includes small, medium, and large ship targets in oceanic and nearshore areas, accounting for 60.2%, 36.8%, and 3% of the total targets, respectively.

The HRSID dataset includes 5604 SAR ship images with a resolution ranging from 0.5 to 3 m, containing a total of 16,951 ship targets. The proportions of small, medium, and large ships in this dataset are 54.5%, 43.5%, and 2%, respectively.

The SAR images in both datasets exhibit various characteristics, including different sea conditions, polarization modes, resolutions, complex scenes, and ship sizes. The datasets were arbitrarily partitioned into training, validation, and testing subsets, adhering to a proportion of 7:1:2, respectively. Table 1 presents comprehensive details of two datasets, where the polarization describes the orientation of radar waves during their propagation and reception concerning the Earth's surface. HH represents a dual horizontal polarization configuration, indicating the transmission and reception of horizontally polarized waves. HV signifies transmission with horizontal polarization received as vertical polarization, while VV denotes the mutual transmission and reception of vertically polarized waves. Conversely, VH refers to transmission with vertical polarization received as horizontal polarization. These polarization configurations offer distinct information about surface features.

Table 1. Detailed information on SSDD and HRSID.

Datasets	Image Number	Ship Number	Resolution (m)	Polarization	Image Size	
					Height	Width
SSDD	1160	2587	1–15	HH, HV, VV, VH	190~526	214~668
HRSID	5604	16,951	0.5, 1, 3	HH, HV, VV	800	800

4.2. Experimental Setup and Parameters

The experiments were conducted on a platform based on Pytorch 2.0.0 and CUDA 11.8, using an AMD EPYC 9654 processor with a clock speed of 2.4 G, an RTX4090 graphics card with 24 G of memory, and 60 G of system memory. The operating system used was Ubuntu 20.04.5 LTS.

For the experimental parameters, a consistent batch size of 16 was employed, and the model underwent training for an aggregate of 300 epochs. The optimization process was facilitated by the Stochastic Gradient Descent (SGD) optimizer, with an inaugural learning rate pegged at 0.01. The learning rate decay followed a cosine annealing schedule, tapering down to a final learning rate of 0.0001. Additionally, we set the weight decay factor to 0.0005.

4.3. Evaluation Metrics

The efficacy of the detection was assessed using several metrics, including precision (P), recall (R), average precision (AP), F1 score, and frames per second (FPS). The computation of precision and recall adheres to the following formulas:

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{9}$$

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{10}$$

where TP (true positives) is the number of correct detections, FP (false positives) is the number of false alarms where the negative class is predicted as the positive class, and FN (false negatives) is the number of missed detections where the positive class is predicted as the negative class.

The formula for calculating AP is as follows:

$$\mathrm{AP} = \int_0^1 \mathrm{P}(\mathrm{R})\mathrm{dR} \tag{11}$$

AP50 refers to the detection precision with an IoU threshold of 50%. AP50:95 refers to the average AP calculated with IoU values ranging from 50% to 95%, with a step size of 5%. In general, precision and recall are likely to be negatively correlated. The F1 score, as the harmonic mean of precision and recall, measures both metrics comprehensively. The formula is as follows:

$$\mathrm{F1} = \frac{2\mathrm{PR}}{\mathrm{P} + \mathrm{R}} \tag{12}$$

The FPS metric is indicative of the quantity of images that the model is capable of detecting within a span of one second. A higher FPS value is synonymous with a faster detection speed for the model.

4.4. Module Effectiveness Analysis

To verify the effectiveness of the modules proposed in this paper, an analysis was conducted on the SSDD dataset.

4.4.1. Verification of RFEM

We compared different feature fusion methods within the RFEM module to explore better fusion methods. We conducted five repeated experiments, and the mean and stan-

dard deviation results are shown in Table 2. It is evident that the concatenation fusion method performs the best in RFEM, with an increase of 0.5% in AP50 and 1.3% in AP50:75.

Table 2. Comparison of different fusion methods for sensory wild enhancement modules.

Fusion Method	AP50 (%)	AP50:95 (%)	F1 (%)
Baseline	96.9 ± 0.4	60.8 ± 0.6	94.7 ± 0.2
Attentional Fusion	96.9 ± 0.3	61.1 ± 0.5	94.5 ± 0.3
Adaptive Fusion	97.1 ± 0.2	62.1 ± 0.4	94.9 ± 0.3
Concatenation Fusion	**97.4 ± 0.3**	**62.1 ± 0.5**	**95.1 ± 0.2**

Bold font indicates the optimal values of each metric.

In an endeavor to further scrutinize the extraction efficacy of the network post- incorporation of the RFEM module, we elected to conduct a heatmap experiment utilizing images characterized by complex and ambiguous backgrounds. The outcomes of this experiment are depicted in Figure 8. In comparison to the baseline model, the network, upon the addition of the RFEM module, exhibits an enhanced ability to distinguish crowded multi-scale targets. This suggests a superior capability to handle intricate scenarios.

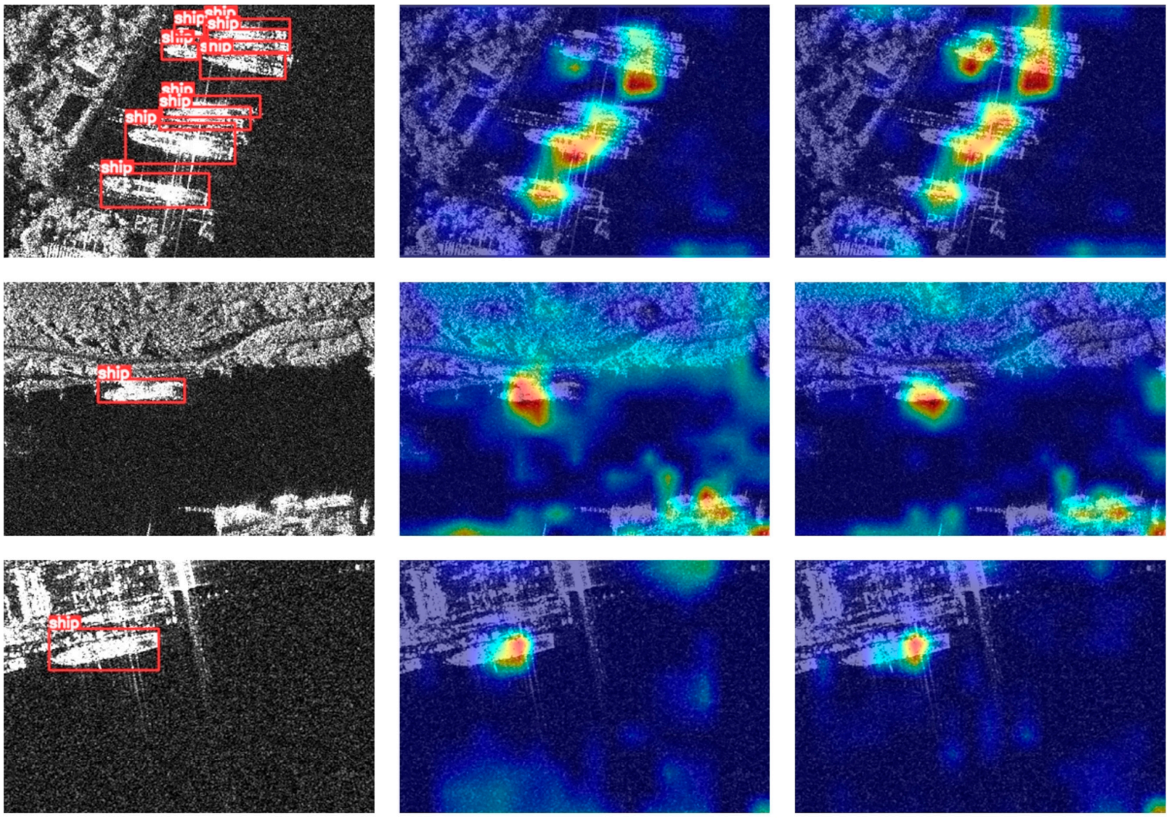

Figure 8. Results of the heat map visualization with the introduction of the RFEM module (**a**) Ground truth; (**b**) YOLOv5s; (**c**) YOLOv5s + RFEM.

4.4.2. Verification of NWD

In the experiments, we observed the convergence of four loss functions: CIOU [29], SIOU [30], WIOU [46], and NWD, as shown in Figure 9. Through these experiments, we found that NWD converges faster and achieves the lowest loss value. This validates the rationale behind choosing NWD for this study.

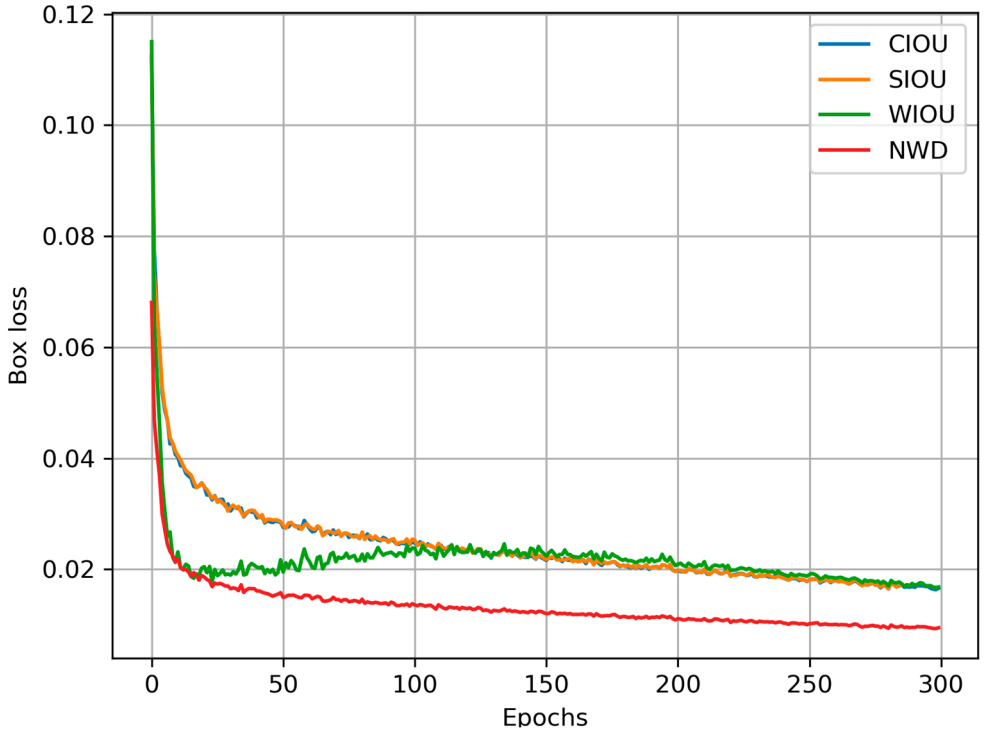

Figure 9. Comparison of different loss functions.

4.5. Ablation Experiment

To delve deeper into the efficacy of the diverse modules of the algorithm proposed in this paper, an ablation study was executed on the SSDD dataset. The outcomes of this study are presented in Table 3, where the presence of a check mark ("√") signifies the incorporation of the module in the baseline model.

Table 3. Ablation experiments of this paper's algorithm on the SSDD dataset.

ID	RFEM	CLF	NWD	P (%)	R (%)	AP50 (%)	AP50:95 (%)	F1 (%)
1				94.1	95.8	96.9	60.8	94.7
2	√			94.5	95.5	97.4	62.1	95.1
3		√		94.7	95.7	97.1	62.5	95.2
4			√	94.1	**97.0**	97.2	61.9	95.5
5	√	√		**95.4**	95.2	97.9	62.8	95.3
6	√	√	√	95.0	96.7	**98.2**	**63.1**	**95.6**

The '√' symbol indicates the use of this module in the model, while bold type represents the optimal value for each metric.

From the experimental data, it can be observed that compared to the baseline model, the second and third groups, which respectively added the RFEM and CLF-FPN modules, showed improvements in accuracy and AP. However, there was a decrease in recall to varying degrees. In the fifth group, where both modules were added simultaneously, the situation was more pronounced. The AP indicator improved significantly, but the F1 score, which comprehensively measures accuracy and recall, had limited improvement. This is because there is a certain trade-off between accuracy and recall, especially in the current situation where the accuracy and recall of the baseline model are both high. Therefore, we introduced the NWD loss function to balance this situation. As can be seen, the sixth group, which included all modules, had the highest AP and F1 scores. Compared to the baseline model, AP50 and AP50:95 increased by 1.3% and 2.3%, respectively, and F1 increased by 0.9%. This demonstrates that YOLO-CLF achieved a balance in all evaluation indicators.

We also conducted an ablation study on the HRSID dataset, which is more challenging than the SSDD dataset and has a larger room for improvement in detection results. The accuracy and recall have not yet reached the trade-off situation as before. The results are shown in Table 4.

Table 4. Ablation experiments of this paper's algorithm on the HRSID dataset.

ID	RFEM	CLF	NWD	P (%)	R (%)	AP50 (%)	AP50:95 (%)	F1 (%)
1				89.5	80.7	88.2	59.6	84.5
2	√			90.2	81.3	88.7	60.4	85.5
3		√		91.2	81.9	89.4	61.5	86.3
4			√	90.6	83.5	88.9	61.2	86.9
5	√	√		91.4	83.1	89.8	62.1	87.1
6	√	√	√	**91.8**	**83.7**	**90.4**	**62.7**	**87.8**

The '√' symbol indicates the use of this module in the model, while bold type represents the optimal value for each metric.

As discernible from Table 4, in the first group of baseline models, the original YOLOv5s yielded an AP50 and AP50:95 of 88.2% and 59.6%, respectively, with an F1 score of 84.5%. Upon the incorporation of various modules in the second, third, and fourth groups, all evaluation metrics experienced varying degrees of enhancement. Notably, the CLF-FPN module outperformed the others, with an increase of 1.2% and 1.9% in AP50 and AP50:95, respectively. In the fifth group, based on the third group, the simultaneous introduction of the CLF-FPN and RFEM modules into the model led to a further increase of 0.4% and 0.6% in AP50 and AP50:95, respectively. In the seventh group, YOLO-CLF, compared to the original algorithm, the model's AP50 and AP50:95 each increased by 2.2% and 3.1%, and the F1 score increased by 3.3%.

Furthermore, we conducted a visual comparison of the detection performance of the original YOLOv5s and the improved YOLO-CLF under different scenarios, the results of which are illustrated in Figure 10. In the figure, false detection targets are represented by blue ellipses, while missed detection targets are represented by yellow ellipses. The analysis reveals that under ordinary scenarios, the original model exhibits excellent detection performance. However, when ship targets are interfered with by other factors, making it difficult to distinguish between the foreground and background, or when ships of different scales are docked together, YOLO-CLF demonstrates a lower number of false and missed detections compared to the baseline model. Even in scenarios where small-scale ship targets are densely arranged, although there are a few instances of missed and false detections, there is a significant improvement compared to the baseline model.

4.6. Comparative Experiment

To objectively assess the performance of the improved algorithm, we compared YOLO-CLF with other prevalent algorithms under identical conditions. We conducted five repeated experiments to minimize experimental variability, and the mean and standard deviation results of the comparative experiments are presented in Table 5. The results derived from the analysis indicate that Libra R-CNN, as a two-stage detection algorithm, holds an edge in detection accuracy when compared to single-stage algorithms such as SSD and FCOS. However, its detection accuracy is marginally subpar when juxtaposed with some of the more superior single-stage algorithms. Centernet++, as an anchor-free detection algorithm, has good overall performance, but it is still weaker than our algorithm. Due to the increase in model parameters, while YOLO-CLF exhibits a marginally slower detection speed compared to the original model, it still significantly surpasses the benchmark for real-time detection. The F1 indicator on the HRSID dataset is slightly lower than YOLOv7x, but we have a great advantage in detection speed. Ultimately, the model we proposed achieved a balance between speed and accuracy compared to other popular detection algorithms.

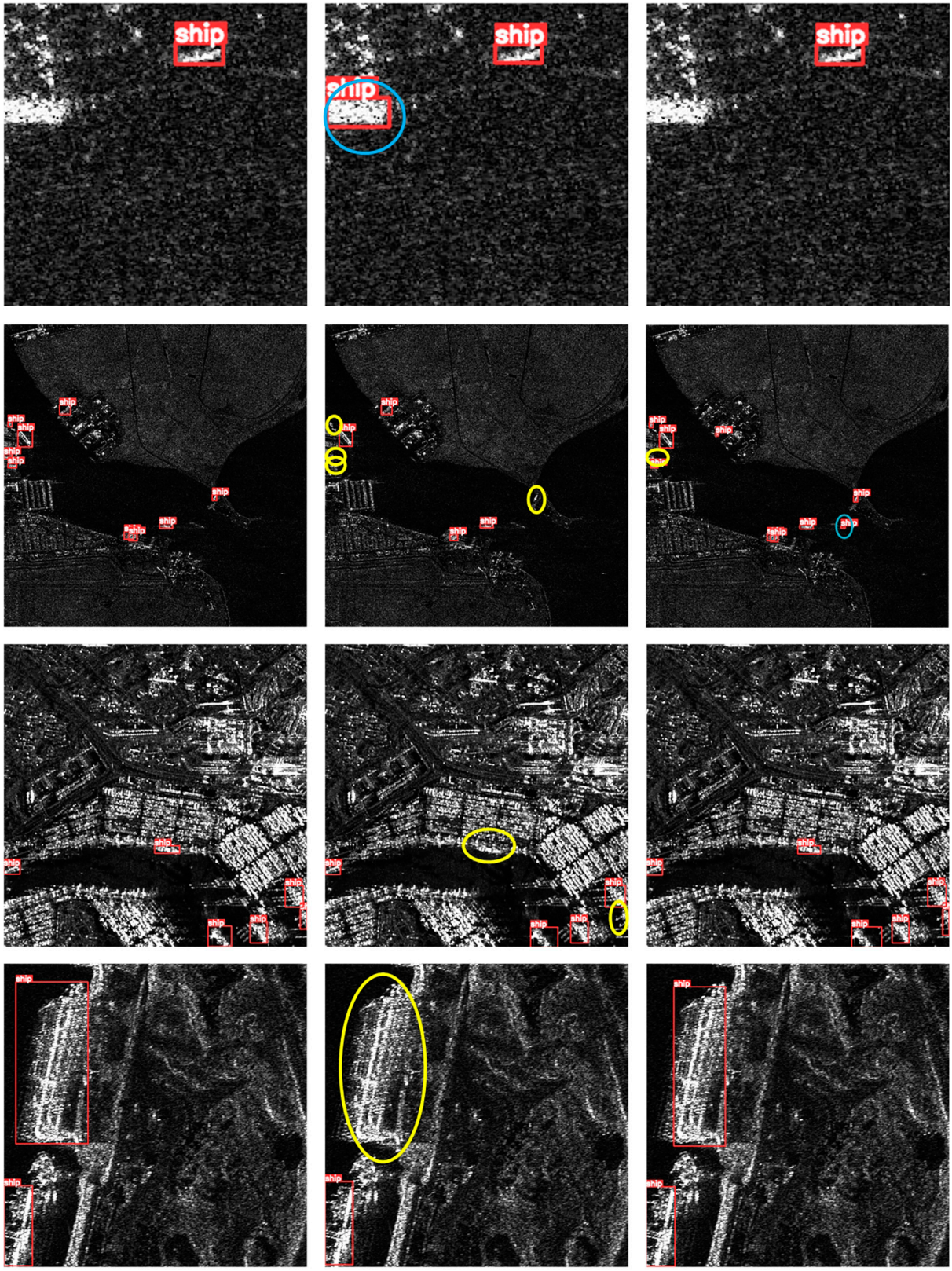

Figure 10. Test results of the baseline model and YOLO-CLF in different scenarios (**a**) Ground truth; (**b**) YOLOv5s; (**c**) YOLO-CLF; The blue oval indicates the misdetected target, and the yellow oval indicates the missed target.

Table 5. Algorithm comparison experiments.

Method	SSDD			HRSID		
	AP50 (%)	**F1 (%)**	**FPS**	**AP50 (%)**	**F1 (%)**	**FPS**
SSD512	85.7 ± 0.4	83.4 ± 0.4	31.3 ± 4.6	85.1 ± 0.5	82.2 ± 0.4	27.3 ± 3.3
FCOS [47]	87.0 ± 0.3	83.8 ± 0.2	22.5 ± 3.9	79.7 ± 0.4	75.5 ± 0.3	20.6 ± 1.8
Libra R-CNN [48]	89.6 ± 0.5	85.6 ± 0.4	8.4 ± 1.2	87.9 ± 0.4	83.1 ± 0.4	7.3 ± 0.9
Centernet++ [49]	92.7 ± 0.3	91.9 ± 0.4	42.1 ± 4.1	86.8 ± 0.3	83.5 ± 0.3	37.9 ± 4.3
YOLOv5s	96.9 ± 0.4	94.7 ± 0.2	**96.8 ± 10.3**	88.2 ± 0.4	84.5 ± 0.3	**79.2 ± 8.0**
YOLOv7x [50]	97.4 ± 0.1	95.2 ± 0.3	49.1 ± 5.6	90.3 ± 0.2	**88.1 ± 0.2**	42.9 ± 4.1
YOLO-CLF	**98.2 ± 0.3**	**95.6 ± 0.2**	81.4 ± 8.7	**90.4 ± 0.3**	87.8 ± 0.3	68.0 ± 7.3

Bold font indicates the optimal values of each metric.

5. Conclusions

This paper addresses the issues of complex backgrounds, dense distribution, and diverse scales of SAR ship targets by proposing the YOLO-CLF detection algorithm. We designed the RFEM module and the CLF-FPN structure to enhance detection performance. Empirical evidence demonstrates that, compared to other algorithms, YOLO-CLF achieves the best overall performance. However, there are still shortcomings in the algorithm we proposed. For instance, the complexity of the model has increased due to the addition of multiple structures. In future research, we plan to explore methods such as pruning and knowledge distillation to deploy the model on lightweight platforms.

Author Contributions: Conceptualization, H.Z. and X.X.; methodology, H.Z. and R.Y.; software, H.Z.; validation, H.Z.; investigation, H.Z. and C.L.; resources, X.X.; data curation, H.Z.; writing—original draft preparation, H.Z.; writing—review and editing, H.Z. and Z.L.; visualization, H.Z.; project administration, X.X.; funding acquisition, X.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Science and Technology Plan Project (2021JH2/10200023) of Liaoning Province, China, and the Key Project (LJKZ0618) of scientific research of the Education Department of Liaoning Province, China.

Data Availability Statement: Data are contained within the article.

Acknowledgments: We express our heartfelt gratitude to the reviewers and editors for their meticulous work, and we thank the authors of SSDD and HRSID.

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

References

1. 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]
2. Xiong, G.; Wang, F.; Yu, W.; Truong, T.K. Spatial Singularity-Exponent-Domain Multiresolution Imaging-Based SAR Ship Target Detection Method. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–12. [CrossRef]
3. 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]
4. Reigber, A.; Scheiber, R.; Jager, M.; Prats-Iraola, P.; Hajnsek, I.; Jagdhuber, T.; Papathanassiou, K.P.; Nannini, M.; Aguilera, E.; Baumgartner, S.; et al. Very-High-Resolution Airborne Synthetic Aperture Radar Imaging: Signal Processing and Applications. *Proc. IEEE* **2013**, *101*, 759–783. [CrossRef]
5. Zhang, T.; Zeng, T.; Zhang, X. Synthetic Aperture Radar (SAR) Meets Deep Learning. *Remote Sens.* **2023**, *15*, 303. [CrossRef]
6. Li, J.; Xu, C.; Su, H.; Gao, L.; Wang, T. Deep Learning for SAR Ship Detection: Past, Present and Future. *Remote Sens.* **2022**, *14*, 2712. [CrossRef]
7. 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]
8. Zhou, Y.; Liu, H.; Ma, F.; Pan, Z.; Zhang, F. A Sidelobe-Aware Small Ship Detection Network for Synthetic Aperture Radar Imagery. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–16. [CrossRef]
9. Joshi, S.K.; Baumgartner, S.V. Automatic CFAR Ship Detection in Single–Channel Range-Compressed Airborne Radar Data. In Proceedings of the 2019 20th International Radar Symposium (IRS), Ulm, Germany, 26–28 June 2019; pp. 1–8.

10. Liu, T.; Zhang, J.; Gao, G.; Yang, J.; Marino, A. CFAR Ship Detection in Polarimetric Synthetic Aperture Radar Images Based on Whitening Filter. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 58–81. [CrossRef]
11. Leng, X.; Ji, K.; Yang, K.; Zou, H. A Bilateral CFAR Algorithm for Ship Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1536–1540. [CrossRef]
12. Zhou, J.; Xie, J. An Improved Quantile Estimator With Its Application in CFAR Detection. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [CrossRef]
13. Bezerra, D.X.; Lorenzzetti, J.A.; Paes, R.L. Marine Environmental Impact on CFAR Ship Detection as Measured by Wave Age in SAR Images. *Remote Sens.* **2023**, *15*, 3441. [CrossRef]
14. Zhou, J.; Xie, J. Robust CFAR Detector Based on KLQ Estimator for Multiple-Target Scenario. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–16. [CrossRef]
15. Zaidi, S.S.A.; Ansari, M.S.; Aslam, A.; Kanwal, N.; Asghar, M.; Lee, B. A survey of modern deep learning based object detection models. *Digit. Signal Process.* **2022**, *126*, 103514. [CrossRef]
16. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
17. 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] [PubMed]
18. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
19. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the Computer Vision—ECCV 2016, Cham, The Netherlands, 11–14 October 2016; pp. 21–37.
20. Liu, S.; Huang, D. Receptive field block net for accurate and fast object detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 385–400.
21. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.
22. Kang, M.; Leng, X.; Lin, Z.; Ji, K. A modified faster R-CNN based on CFAR algorithm for SAR ship detection. In Proceedings of the 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP), Shanghai, China, 18–21 May 2017; pp. 1–4.
23. Zhang, L.; Liu, Y.; Qu, L.; Cai, J.; Fang, J. A Spatial Cross-Scale Attention Network and Global Average Accuracy Loss for SAR Ship Detection. *Remote Sens.* **2023**, *15*, 350. [CrossRef]
24. Pan, X.; Li, N.; Yang, L.; Huang, Z.; Chen, J.; Wu, Z.; Zheng, G. Anomaly-Based Ship Detection Using SP Feature-Space Learning with False-Alarm Control in Sea-Surface SAR Images. *Remote Sens.* **2023**, *15*, 3258. [CrossRef]
25. Hou, X.; Ao, W.; Song, Q.; Lai, J.; Wang, H.; Xu, F. FUSAR-Ship: Building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition. *Sci. China Inf. Sci.* **2020**, *63*, 140303. [CrossRef]
26. Ao, W.; Xu, F.; Li, Y.; Wang, H. Detection and Discrimination of Ship Targets in Complex Background From Spaceborne ALOS-2 SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 536–550. [CrossRef]
27. Zhu, M.; Hu, G.; Li, S.; Zhou, H.; Wang, S.; Feng, Z. A Novel Anchor-Free Method Based on FCOS + ATSS for Ship Detection in SAR Images. *Remote Sens.* **2022**, *14*, 2034. [CrossRef]
28. Cui, Z.; Wang, X.; Liu, N.; Cao, Z.; Yang, J. Ship Detection in Large-Scale SAR Images Via Spatial Shuffle-Group Enhance Attention. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 379–391. [CrossRef]
29. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 12993–13000.
30. Gevorgyan, Z. SIoU loss: More powerful learning for bounding box regression. *arXiv* **2022**, arXiv:2205.12740.
31. Qin, C.; Wang, X.; Li, G.; He, Y. A Semi-Soft Label-Guided Network With Self-Distillation for SAR Inshore Ship Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–14. [CrossRef]
32. Xu, Z.; Gao, R.; Huang, K.; Xu, Q. Triangle Distance IoU Loss, Attention-Weighted Feature Pyramid Network, and Rotated-SARShip Dataset for Arbitrary-Oriented SAR Ship Detection. *Remote Sens.* **2022**, *14*, 4676. [CrossRef]
33. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
34. Ghiasi, G.; Lin, T.-Y.; Le, Q.V. Nas-fpn: Learning scalable feature pyramid architecture for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7036–7045.
35. Li, Y.; Chen, Y.; Wang, N.; Zhang, Z. Scale-aware trident networks for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6054–6063.
36. 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]
37. Zhou, Y.; Fu, K.; Han, B.; Yang, J.; Pan, Z.; Hu, Y.; Yin, D. D-MFPN: A Doppler Feature Matrix Fused with a Multilayer Feature Pyramid Network for SAR Ship Detection. *Remote Sens.* **2023**, *15*, 626. [CrossRef]
38. Xiao, J.; Zhao, T.; Yao, Y.; Yu, Q.; Chen, Y. Context augmentation and feature refinement network for tiny object detection. In Proceedings of the Tenth International Conference on Learning Representations, Virtual Event, 3–7 May 2021.

39. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
40. Yang, G.; Lei, J.; Zhu, Z.; Cheng, S.; Feng, Z.; Liang, R.J. AFPN: Asymptotic Feature Pyramid Network for Object Detection. *arXiv* **2023**, arXiv:2306.15988.
41. Liu, S.; Huang, D.; Wang, Y.J.a.p.a. Learning spatial fusion for single-shot object detection. *arXiv* **2019**, arXiv:1911.09516.
42. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.
43. Wang, J.; Xu, C.; Yang, W.; Yu, L.J. A normalized Gaussian Wasserstein distance for tiny object detection. *arXiv* **2021**, arXiv:2110.13389.
44. Zhang, T.; Zhang, X.; Li, J.; Xu, X.; Wang, B.; Zhan, X.; Xu, Y.; Ke, X.; Zeng, T.; Su, H.; et al. SAR Ship Detection Dataset (SSDD): Official Release and Comprehensive Data Analysis. *Remote Sens.* **2021**, *13*, 3690. [CrossRef]
45. Wei, S.; Zeng, X.; Qu, Q.; Wang, M.; Su, H.; Shi, J. HRSID: A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation. *IEEE Access* **2020**, *8*, 120234–120254. [CrossRef]
46. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R.J. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. *arXiv* **2023**, arXiv:2301.10051.
47. Tian, Z.; Shen, C.; Chen, H.; He, T. Fcos: Fully convolutional one-stage object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9627–9636.
48. 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 821–830.
49. Guo, H.; Yang, X.; Wang, N.; Gao, X.J.P.R. A CenterNet++ model for ship detection in SAR images. *Pattern Recognit.* **2021**, *112*, 107787. [CrossRef]
50. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 7464–7475.