remote sensing

MDPI

Article

An Anchor-Free Method Based on Adaptive Feature Encoding and Gaussian-Guided Sampling Optimization for Ship Detection in SAR Imagery

Bokun He, Qingyi Zhang, Ming Tong and Chu He *

Electronic Information School, Wuhan University, Wuhan 430072, China; bokun.he@whu.edu.cn (B.H.); zhqy@whu.edu.cn (Q.Z.); tongming@whu.edu.cn (M.T.)
* Correspondence: chuhe@whu.edu.cn

Abstract: Recently, deep-learning methods have yielded rapid progress for object detection in synthetic aperture radar (SAR) imagery. It is still a great challenge to detect ships in SAR imagery due to ships' small size and confusable detail feature. This article proposes a novel anchor-free detection method composed of two modules to deal with these problems. First, for the lack of detailed information on small ships, we suggest an adaptive feature-encoding module (AFE), which gradually fuses deep semantic features into shallow layers and realizes the adaptive learning of the spatial fusion weights. Thus, it can effectively enhance the external semantics and improve the representation ability of small targets. Next, for the foreground–background imbalance, the Gaussian-guided detection head (GDH) is introduced according to the idea of soft sampling and exploits Gaussian prior to assigning different weights to the detected bounding boxes at different locations in the training optimization. Moreover, the proposed Gauss-ness can down-weight the predicted scores of bounding boxes far from the object center. Finally, the effect of the detector composed of the two modules is verified on the two SAR ship datasets. The results demonstrate that our method can effectively improve the detection performance of small ships in datasets.

Keywords: ship object detection; deep learning; remote sensing imagery; feature extraction; object sampling

Citation: He, B.; Zhang, Q.; Tong, M.; He, C. An Anchor-Free Method Based on Adaptive Feature Encoding and Gaussian-Guided Sampling Optimization for Ship Detection in SAR Imagery. *Remote Sens.* **2022**, *14*, 1738. https://doi.org/10.3390/rs14071738

Academic Editor: Stefano Perna

Received: 11 February 2022
Accepted: 31 March 2022
Published: 4 April 2022

1. Introduction

Synthetic aperture radar (SAR) is an active imaging radar that has the advantages of all-weather operation and a robust anti-jamming ability. It can effectively identify camouflage and penetrate masking objects, and is an essential means of ground monitoring [1]. Due to its imaging mechanism and characteristics, SAR has been widely used in marine monitoring, especially in detecting nearshore and ocean ships. SAR ship detection has high application value in both military and civil fields. For example, cruise counting and ocean rescue in the civil field [2] and battlefield detective and intelligence acquisition in the military field are inseparable from this technology. The traditional SAR ship detection usually adopts the constant false alarm rate (CFAR) [3] algorithm, which directly calculates the detection threshold adaptively according to the local clutter statistical characteristics, and judges whether it is a target according to the threshold. The advantage of the CFAR algorithm lies in its simple structure, a small amount of computation, and fast detection speed, but it is sensitive to the selection of clutter distribution. The breakthrough of deep learning makes convolutional neural network (CNN) shine in the field of computer vision. CNN has achieved outstanding results in the field of image classification. The object-detection algorithm based on deep learning usually takes a classification network as the backbone network to extract the features of the image and then sends it to the detection network for classification and regression after feature fusion. At present, they are often divided into two-stage networks represented by the RCNN series [4] and one-stage networks represented by Yolo [5], SSD [6] and retinanet [7].

The above RCNN series, Yolo, SSD and RetinaNet are anchor-based detectors. Since the emergence of CornerNet [8] in 2018, a large number of detection algorithms without anchor design have emerged in the field of object detection, such as FASF [9], CenterNet [10], and FCOS [11]. These algorithms further promote the transition from object detection to the end-to-end system.

At the same time, researchers in the field of SAR also try to migrate CNN in natural images to the field of SAR ship detection. However, due to the significant field differences between natural images and SAR images, the following characteristics deserve attention in algorithm design.

SAR image is a top view taken by airborne or space-borne satellites. The direction of objects is often arbitrary. Therefore, the rotation invariance of feature extraction should be considered when designing the network. In addition, the image used for SAR ship object detection is often a single-channel gray image preprocessed by geometric correction. In contrast, the natural image is often a three-channel color image. Limited by the demand of SAR long-range imaging, the object being detected occupies a small image area. It can be regarded as a small target with low resolution and insufficient detailed information.

As shown in Table 1, small objects are generally defined as objects that are less than 32×32 pixels in area or less than 1% of the image size. However, the current object detection algorithms are not ideal for detecting small targets. On general visual data sets such as Ms coco [12], the detection accuracy of small targets is about half that of large targets, while in SAR images with more small targets and sparse targets, the problem is more serious.

Table 1. Definition of large, medium, and small targets in MS COCO.

Pixel	Minimum Area	Maximum Area
Small target	0×0	32×32
Medium target	32×32	96×96
Large target	96×96	$\infty \times \infty$

One of the problems of small ship detector is that small targets occupy fewer pixels and lack detailed information, resulting in insufficient available feature information. To extract the deep-seated features of the target, the object-detection algorithm based on deep learning often needs to improve the semantics through continuous convolution and downsampling. If only the last layer feature map of the backbone network is used for detection, for small targets with an area of fewer than 32×32 pixels, there may be less than 1 pixel mapped to the feature map. This will undoubtedly lead to a smaller problem of information.

Unlike the classification task, in the training process of the detection task, the predicted samples need to be allocated to the actual target by manually designing rules to divide the positive and negative samples. This division process is called sampling. The quality of the sampling strategy directly affects the effect of training. If the target cannot be effectively sampled in the training process, the learning of that category cannot be well, resulting in the difficulty of detection. Due to the sparsity of small ship targets in SAR images, the imbalance between positive and negative samples often occurs, resulting in poor training effects.

Therefore, aiming at the two difficulties of small-target detection in SAR image, this paper first proposes an adaptive feature encoding method, which can directly learn the spatial weighting coefficient of deep feature fusion as to filter the deep useless information and enhance the useful information. Secondly, a target existence measure gauss-ness based on Gaussian distribution modeling is proposed. The weighted sampling method based on gauss-ness is used to recalibrate the weights of positive samples at different locations in the boundary box.

1.1. Problem Description and Motivations

Although object detection has achieved good results on public data sets such as MS coco [13] and pascal-VOC [14], the detection accuracy of small targets is still very low.

One problem of small target detection is that the detailed information is not rich, and the available feature information is too little [15]. Object-detection algorithms based on deep learning usually use a classification network as the backbone network for feature extraction. Compared with detection tasks, classification requires higher semantics. Therefore, in the classification network, multiple downsampling will be carried out in convolution to improve semantics, which is unfavorable for detecting small targets. Although introducing a feature pyramid network (FPN) greatly alleviates this problem, for small targets, although the receptive field of shallow features is small and the detailed features are retained, its semantic information representation ability is weak. For SAR images, this problem is more serious. For example, on public SAR datasets such as HRSID [16] and SSDD [17], the proportion of small targets exceeds 50%. In addition, due to the sparsity of targets in SAR image, the proportion of pixel area of small targets in SAR image is also smaller than that in the optical image. These problems have brought great challenges to detecting small targets in SAR images.

Another problem is that positive and negative samples are uneven in SAR ship detection, and the sampling quantity and quality of positive samples in a common sampling strategy are poor [18]. For the anchor-based detection algorithm, the anchor is difficult to design, and the sampling strategy based on IOU matching is easy to lead the absence of positive samples, and some targets with small area are ignored. For the anchor-free design method, the division of positive and negative samples is based on whether the pixels are in the actual boundary box. There will be pixel confusion at the edge of the boundary box, resulting in the reduction of the quality of positive samples.

1.2. Contributions and Structure

This paper has carried out a series of research on the SAR ship's small target problems. The main contents and innovations are as follows:

(1) Aiming at the problem of insufficient detail information and difficult feature extraction of small targets, an adaptive feature encoding method (AFE) is proposed. This method effectively integrates the deep high semantic features into the shallow layer to enhance the feature representation of small targets, so as to improve the detection performance of small targets. AFE first calculates the spatial weight of each deep feature map in the multi-scale feature pyramid by introducing a spatial attention mechanism, then weights the deep features in the way of pixel-by-pixel multiplication and fuses them into the shallow feature map. After normalization, the fused feature map with detailed information and high-level semantics is obtained. Experiments on HRSID and SSDD data sets show that the AFE method has a significant improvement in the problem of feature reuse conflict compared with other variants of FPN.

(2) Aiming at the sampling problem of small targets, this paper first determines the sampling method without anchor design and analyzes the quality problem of edge positive samples. Then a Gaussian-guided detection head (GDH) is introduced. It proposes a target existence measure Gauss-ness which is more suitable for SAR ships, and a Gauss-ness weighted sampling strategy. Experiments show that the sampling optimization method can achieve good improvement in small target detection.

(3) A detector suitable for small-target detection is constructed by combining the adaptive feature method and the sampling optimization method proposed above. The proposed AFE is embedded in the basic feature extraction module, while GDH plays a role in the object location. The effect of the detector has also been verified experimentally.

This article is organized as follows: Section 3 briefly introduces some existing methods that have inspired our work. Next, Section 3.2 describes the principle and significance of

the two proposed modules. The description of datasets and experimental results are shown in Section 4, and a final conclusion is stated in Section 5.

2. Related Work

2.1. General Object Detection

In 2014, Grishick et al. proposed regional convolutional neural network (RCNN) [4], which became the pioneering work of deep learning in the field of target detection. The subsequently proposed Faster-RCNN [19] is the first end-to-end detection network, which utilizes Region Proposal Network (RPN) [19] to integrate the extraction of candidate regions with target detection head. Different from RCNN series, which divides detection into two tasks of classification and positioning, You Only Look Once (YOLO) [5,20,21] removes the generation process of candidate regions and unify classification and positioning into regression problems, which greatly improves detection speed and is the pioneering work of single-stage network.

Both Faster-RCNN and YOLO are based on preset anchor frame for detection. The design of anchor frame parameters directly affects the effect of target detection and limits the generalization of the target-detection algorithm. In 2018, Hei Law et al. proposed CornerNet based on key-point detection, which transforms the detection box into the description of key points; that is, two points at the upper-left corner and the lower-right corner are used to determine a detection box, breaking the limitation of the preset anchor box of the target-detection algorithm. Since then, target detection has entered the anchor-free era. The CornerNet only provides information about the edges of the objects. For the objects, the most recognizable information should be the area inside them. CenterNet has added center-point detection to help screen candidate boxes. ExtremeNet [22] points out that corner points may not be on the object. It uses the top, bottom, left, and right poles of the object to describe the boundary of the object, and introduces the center point to judge the category of the object. The CornerNet, CenterNet, and ExtremeNet are all based on key-point detection to break through the limitations of anchors, while another important class of anchor-free detectors are based on segmentation methods. FCOS directly outputs the probability distribution of each position class and the distance of the four boundaries of the target on the feature map. ATSS [23] pointed out that FCOS is superior to RetinaNet because of its excellent positive and negative sample division strategy. Foveabox [24] directly divided the samples of the edge of the boundary box into irrelevant samples, and the definition of the boundary box was determined by two different stretching coefficients. FSAF explored the positive sample allocation problem under multi-scale prediction in FPN (feature pyramid network) [25]; that is, which layer of feature map should be selected for training for each positive sample. FSAF proposes a feature selective anchor-free module to enable the network to adaptively learn how to allocate samples to different layers.

2.2. Ship Detection

SAR ship detection is a popular research direction in the field of remote sensing target detection. The traditional SAR image detection usually adopts Constant False Alarm Rate (CFAR) [26] and its improved algorithm. The essence of CFAR lies in the statistical modeling of clutter in the sea surface background. According to the local clutter statistical characteristics, the false alarm threshold is adaptively calculated, and then each pixel value is compared one-by-one regarding whether it exceeds the false alarm threshold to achieve ship target detection. Traditional ship target detection algorithms are mostly targeted at specific scenes [27,28], and are highly dependent on predefined distribution or artificially designed features in the detection process, resulting in low robustness and poor generalization of the algorithm. Convolutional neural networks based on deep learning have the ability to learn parameters and extract features automatically, and can get rid of the dependence on hand-crafted features, making it the mainstream algorithm for current target detection.

Compared with traditional target detection methods, the current CNN-based methods have a high improvement in detection accuracy and robustness. Therefore, researchers

have widely applied deep learning technology to SAR image ship detection. The direction of improvement mainly focuses on the two key issues of feature extraction and anchor design. The summary and comparison of SAR ship detection algorithms are shown in Table 2.

Table 2. Summary of CNN-based SAR ship detection algorithms.

Category	Method
Target feature extraction	Characteristic pyramid Super dense connection Visual attention Context fusion
Target anchor design	Direction design Scale design

In the aspect of target feature extraction, since the ship shape in SAR image is variable, the feature fusion formula is often used to fuse the information of different feature layers. Yang et al. [29] carried out two cross-layer feature fusion in the network, which could effectively solve the detection problem of multi-scale ship targets at sea and port, but the detection efficiency was reduced. The method in [30] integrates the information of adjacent characteristic layers to fully utilize the semantic meaning and space information, and improves the detection performance of small targets, step by step. However, the vdetection of weak targets or low-intensity targets is easy to cause missed detection and false alarm. Miao et al. [31] generate regional proposals by fusing three layers with different resolutions, so as to improve the spatial resolution of RPN to the same level as that of the middle layer and improve the network's response to small- and medium-sized targets. Chen et al. [32] deployed forward-connected blocks from shallow feature to deep feature and reverse connected blocks to generate the enhanced intermediate feature. Gao et al. [33] use Split Convolution Block (SCB) to divide the input image into smaller pieces to improve the attention of dense objects and strengthen the target area. However, there was an increase in the test time. Inside-outside net [34] employed recurrent neural networks and transmitted spatial information in both horizontal and vertical directions through images. The densely connected multi-scale neural network (DCMSNN) [35] introduces the dense connections to deal with the detection of the multi-scale ships in the multi-scene SAR images.

On the other hand, the design of anchor mainly includes direction design and scale design. MSR2N [36] set six kinds of rotation angles to cover with full directions, but the calculation efficiency decreased. Zhang et al. [37] added rotation angle into the output regression parameters to directly predict the ship direction, but the accuracy was greatly reduced. The above two methods are still limited to the anchor-generating mechanism, resulting in a higher cost of calculation. Referring to CenterNet, Zhang et al. [38] designed a network to predict the target center point, and then performed regression on the scale and direction of the target at the center point. The algorithm in [39] used SSDKmeans clustering algorithm to generate anchor, improving the detection effect of small targets in complex background.

3. Methods

3.1. Explanation of Center-Ness in FCOS

This section summarizes the centrality measurement based on the center priori in FCOS, analyzes its shortcomings in SAR ship object detection.

In the field of object detection, the boundary box is usually used to label the target, and the shape of most targets is often irregular, so there is a certain gap between the representation of the target and the true value. Especially in ship detection in SAR imaging, due to the multi-directionality of the ship target, the marked boundary box of target contains a large number of background pixels, and these background pixels will be indiscriminately

marked as positive samples when sampling in anchor-free methods. These low-quality samples will affect the training optimization of the model.

Approaches such as FCOS, Noisy Anchor [40] and Auto Assign [41] that attempt to solve this problem are all based on an assumption: the central prior, that is, the sampling point located near the center of the boundary box, is usually the most effective, and the validity of the sampling point gradually decreases from the center to the surrounding. In order to quantitatively analyze the effectiveness of different positions in the bounding box, FCOS uses the ratio between the sample position and the distance from the boundary as the centrality of the position and takes it as the measurement of the existence of the target, adds a center-ness branch in the detection head to learn the centrality of the target position, and multiplies the predicted centrality with the classification score when testing. Center-ness suppresses the low-quality samples at the edge of the bounding box, and the suppression process of Gauss-ness is described in Figure 1.

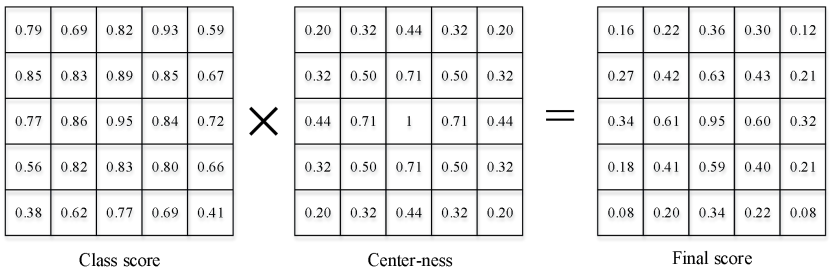


Figure 1. An example of the center-ness suppression process.

P refers to the position in the bounding box, l, r, t and b are the distance from P to the left boundary, right boundary, upper boundary, and lower boundary. The centrality of P in [11] can be described by:

$$center - ness = \sqrt{\frac{\min(l,r)}{\max(l,r)} \times \frac{\min(t,b)}{\max(t,b)}} \tag{1}$$

Figure 2 shows the distribution of centrality in the bounding box represented by Equation (2). Red represents the area with centrality of 1, blue represents the area with centrality of 0, and other colors represent the change process of centrality from 1 to 0.

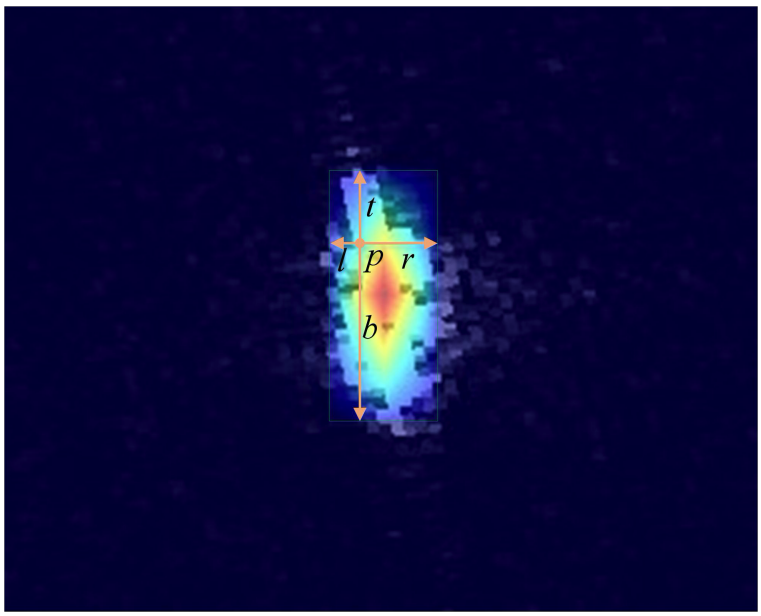


Figure 2. Visualization of center-ness.

We show more center-ness visualization results in the following Section 3.4.1. It is observed that although the centrality calculation method can meet the aforementioned central prior, it can not well fit the pixel distribution of ships due to the multi-directional problem of SAR ships.

3.2. The Overall Framework of the Proposed Method

How to design more efficient networks has been a hot topic in the field of deep learning research. Our proposed ship object-detection method is considered from feature extraction and sampling scheme aspects, and then two modules are designed to be combined with the basic network to improve the overall performance of SAR image ship detection, and Figure 3 is the overall process description of the proposed method.

First of all, our method eliminates the effect of anchors by adopting a general anchor-free strategy that directly learns the encoded bounding boxes. Multiple convolutional layers constitute a backbone to extract image basic features, and then the obtained features are fed into FPN to form multi-scale features. Subsequently, the AFE proposed in this paper adaptively fuses these features and obtains several new feature maps with refined detailed information and semantic information. When it comes to box prediction and positioning, Gauss-ness designed in GDH can be used to calculate the classification score of detected targets in the test process. Meanwhile, in the training process, loss function can be optimized by taking advantage of the characteristics that predicted bounding boxes at different locations have different Gauss-ness values.

Our method is an end-to-end convolutional neural network, which is the same as general detection networks, including a backbone for extracting features and a detection head for target localization. The proposed AFE and GDH have been improved in these two parts respectively.

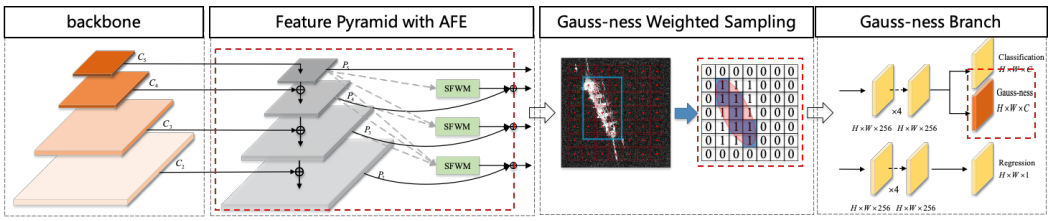


Figure 3. The overall framework of the proposed algorithm for ship target in SAR imagery.

3.3. Refined Feature Pyramid with Adaptive Feature Encoding

FPN is a mainstream solution to small target detection. Previously, SSDS detected targets on multiple scale feature maps. It is based on a rule: a deep feature map has a larger receptive field and stronger semantics that are suitable for large-target detection; while a shallow feature map has higher resolution and keeps enough detail, so it is suitable for small-target detection. Although SSD adopts a shallow feature map to improve the small-target detection ability to a certain extent, the method of directly extracting the lower layer of the backbone network as shallow feature map will lead to insufficient shallow semantic information. FPN combines a deep semantic feature map with a shallow feature map by top-down method, aiming to retain high-resolution detail/information while integrating deep semantic information. The appearance of FPN greatly improves the detection effect of small targets, but it still has some problems. FPN uses heuristic feature selection, which associates objects at different scales to feature maps at different levels through hand-designed scale partitioning strategies. After the target is assigned to a certain level of the feature map, the corresponding areas of the other levels of the feature map are treated as background and suppressed. When the network trains targets of different scales, the direct fusion of FPN will cause feature reuse conflicts among different levels and reduce the effectiveness of feature pyramid. In order to solve this problem, the Guided Anchoring technique [42] attempted to adopt predicted anchor-guided features, but only a single deformable convolution could not solve the problem perfectly. TridentNet [43]

builds multiple specific scale branches with different receptors to avoid conflicts among feature pyramids. However, it does not make good use of shallow feature maps with high-resolution, so its detection of small targets is limited. Therefore, this paper proposes an adaptive feature coding method to further solve this problem.

3.3.1. Initial Feature Pyramid Generation

The input SAR image x is first sent to the backbone network (RESNET, Darknet, etc.), and a series of feature maps $C = \{C2, C3, C4, C5\}$ with different scales can be obtained and after a series of convolution layers. The shallow feature map has high resolution and the detailed features are intact, making it suitable for detecting small targets. The deep feature map has low resolution, but has a larger receptive field and higher semantics, meaning it is suitable for large-object detection. These feature maps are combined as a top-down connection to form a feature pyramid for multi-scale detection. The specific connection process is shown in Figure 4. Firstly, the low-resolution feature map was up-sampled, and then it was added pixel-by-pixel to the shallow feature map with lateral connection. The initial pyramid feature map are obtained subsequently through a 3×3 convolution, and the 1×1 convolution of lateral connection was used to change the number of channels.

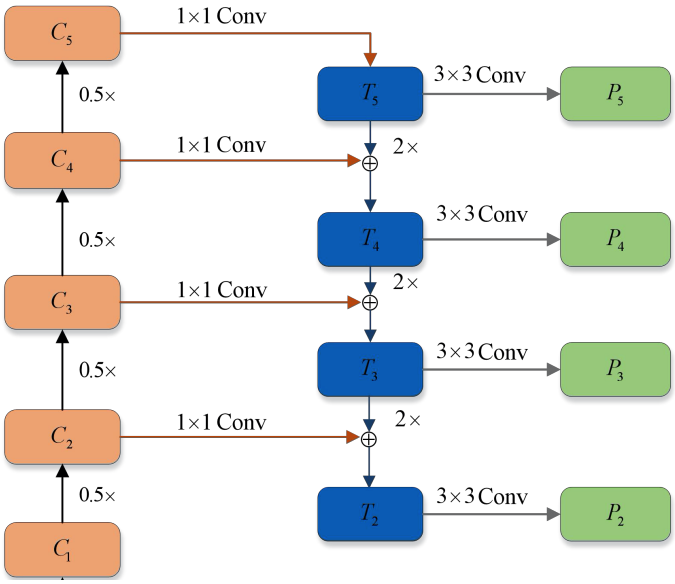


Figure 4. Architecture of the initial FPN.

With the increase of i, the resolution of the feature map gradually decreases, which can be expressed by:

$$\begin{cases} W_i = W/2^i \\ H_i = H/2^i \end{cases} \tag{2}$$

where W and H are the width and height of the image. In order to reduce the amount of calculation, C_1 is usually not used when constructing the feature pyramid. The initial feature pyramid constructed in this paper is $P = \{P2, P3, P4, P5\}$.

3.3.2. Adaptive Feature Encoding

Adaptive feature encoding is mainly divided into two steps: the first step is to up-sampling the deep feature map to realize the resolution matching between deep and shallow feature maps; the second step is to calculate the spatial weight of the matched deep feature map and encode the spatial features.

For a layer P_l in the initial feature pyramid, $P_5, P_4, \ldots P_{l+1}$ scale to P_l size in turn. Specifically, as shown in Figure 5, for P_{l+1}, bilinear interpolation is used for up-sampling, and

3×3 convolution is used for feature modification. For P_{l+2}, the operation of convolution after upsampling is adopted twice, and so on for other feature maps.

The $P_{n \to l} \in \mathbb{R}^{C \times H \times W}$ means that P_n is scaled to the feature map with the same resolution as P_l, and the number of channels is C, $M_{n \to l} \in \mathbb{R}^{C \times H \times W}$ represents the weighted feature map. The process of spatial feature weighting can be described as:

$$M_{n \to l} = \Gamma(P_{n \to l}) \otimes P_{n \to l} \tag{3}$$

where $\otimes$ represents the pixel-by-pixel multiplication with broadcasting attached, and $\Gamma(P_{n \to l}) \in \mathbb{R}^{H \times W}$ represents the two-dimensional weight tensor with the same size as $P_{n \to l}$ generated by the gating module Γ.

The structure of spatial attention mechanism is used for Γ. As shown in Figure 5, the maximum pooling and average pooling of $P_{n \to l}$ are performed in the channel dimension firstly. After the two are spliced together, they are input into the sigmoid function through a 3×3 convolution to obtain the normalized weight tensor. Γ is to obtain the spatial weight, and then conduct pixel-by-pixel multiplication with $P_{n \to l}$ to obtain the weighted $M_{n \to l}$.

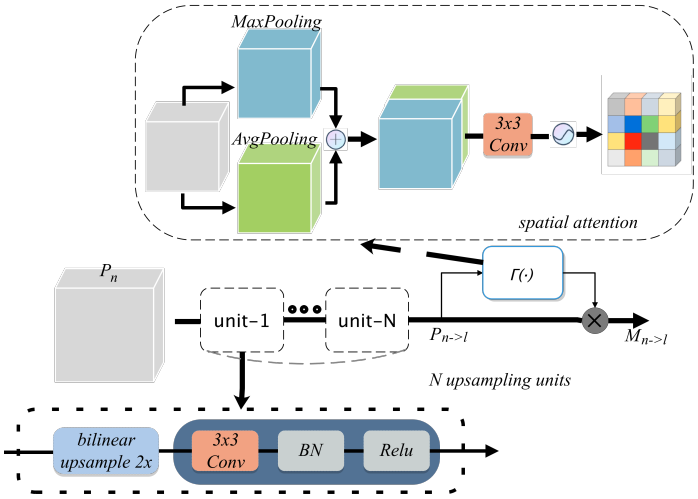


Figure 5. Description of the AFE.

In the specific process, we also normalize the spatial weight of each layer by Softmax. The reason for using Softmax is to keep the accumulation of multi-layer characteristic values in a reasonable range and not cause excessive influence on optimization. Compared with the concatenate operation, the method of per-pixel accumulation is more in line with the idea of weighting and has less computation. The final generation process of feature pyramid is shown in Figure 6.

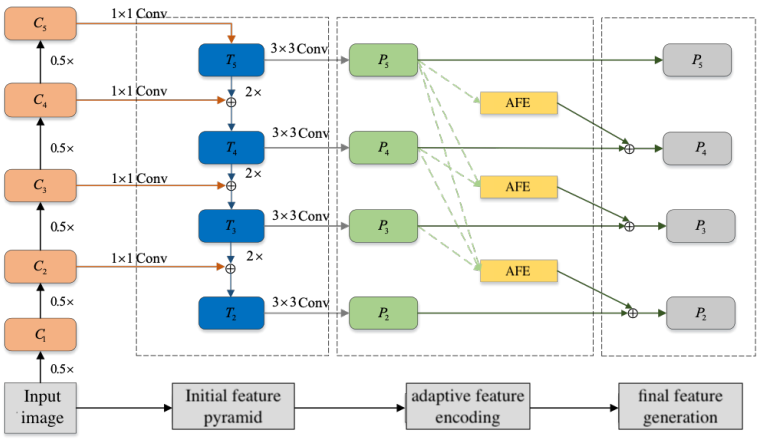


Figure 6. Description of the final feature pyramid generation.

3.4. Gaussian-Guided Detection Head

Different from the classification task, in the training process of the detection task, the predicted samples need to be allocated to divide into positive and negative samples by manually designing rules; this division process is called sampling. The quality of the sampling strategy directly affects the effect of training. The sampling strategies in current detection tasks can be roughly divided into IoU matching sampling, based on predefined anchors, and densely points sampling, without anchors. The IoU matching sampling strategy needs to calculate the IoU between the anchor and the real boundary box of the target to divide positive and negative samples. The problem of this strategy is that it is sensitive to the super-parameter setting of the anchor. The detector without anchor design can avoid this limitation. It usually adopts the method of densely points sampling that directly maps the real boundary box of objects to the final feature map. The points inside the boundary box region are selected as the center point of the positive samples, and the remaining positions are divided as negative samples. Although this method is not affected by hyper-parameters, it cannot effectively judge the quality of samples. The positions on the edge of the boundary box are usually the background center rather than the target, while the samples of different positions are not distinguished in network training. Although FCOS introduces the center-ness branch to suppress the low-quality samples at the edge, there are still some problems in applying it to SAR ship object detection: (1) center-ness cannot fit ships with changeable direction well; (2) FCOS only uses center-ness to restrict the classification score of targets in the test stage, and does not distinguish the contribution of samples from different positions in the bounding box in the training stage.

In view of the above problems in small target detection, we use Gaussian distribution to model the pixel distribution of targets in the horizontal boundary box, and obtain a new target existence measure Gauss-ness. A weighted sampling method based on Gauss-ness is also proposed from the perspective of soft sampling. Compared with the truncated sampling based on Gauss-ness, our method does not directly discard the prediction results of the edge of the bounding box, but reduces its weight when calculating the total classification loss.

3.4.1. Gauss-Ness Branch for Inference Process

It is observed that although the center-ness method can meet the priori of object, it can not well fit the pixel distribution of ship targets due to the multi-direction of SAR ships.

In order to judge the target existence of the location more accurately, a new measurement method needs to meet the following properties:

(1) The center prior is satisfied. The weight of the position closer to the center is larger, and the value closer to the edge is smaller. There is a gradual decay process from the center to the edge. (2) Center symmetry is satisfied. Since the direction of ships in SAR images are uncertain, this metric should not be simply symmetrical in horizontal and vertical directions. (3) To meet the normalization, the centrality of the center of the boundary box is 1, and the edge position should be attenuated to 0 as far as possible.

Through analysis, the bivariate Gaussian distribution is suitable for modeling the horizontal bounding box:

$$f(\boldsymbol{p} \mid \mu, \Sigma) = \frac{1}{2\pi|\sum|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{p}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{p}-\boldsymbol{\mu})\right) \tag{4}$$

As shown in Equation (5), the Gauss-ness is obtained by normalizing it, where $\boldsymbol{p}$ is the vector of position coordinates (x, y), $\boldsymbol{\mu}$ is the mean vector of Gaussian distribution, $\boldsymbol{\Sigma}$ is the covariance matrix of Gaussian distribution, and $|\sum|$ is the determinant of covariance matrix.

$$gauss - ness = \exp\left(-\frac{1}{2}(\boldsymbol{p}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{p}-\boldsymbol{\mu})\right) \tag{5}$$

Due to the obvious contrast between the gray value of the SAR ship and the surrounding sea area, the gray value of the pixels in the boundary box can be used as the sample weight for maximum likelihood estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

The second line in Figure 7 shows the fitting effect of Gauss-ness on the target in the bounding box. It can be observed that Gauss-ness can better fit the shape of ships, especially the ship direction, which has great advantages over the center-ness in the first line.

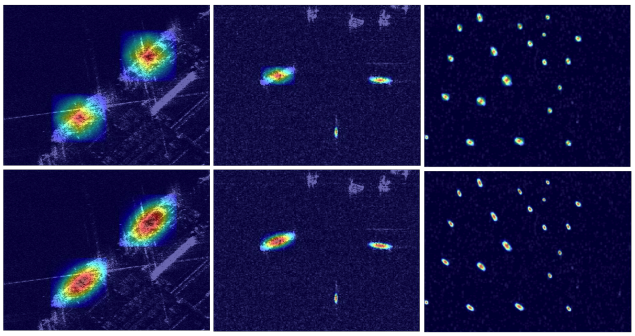


Figure 7. Visualization results of Center-ness and Gauss-ness.

3.4.2. Gauss-Ness Weighted Sampling in Training Process

The anchor-based detector screens positive samples through IoU value. Figure 8a shows the distribution of positive samples in anchor-based detectors, and the number of positive samples depends on the setting of IoU threshold. The threshold of RetinaNet in MS COCO detection task is 0.5 and approximately 25% of the locations in the boundary box will be classified as positive samples. Figure 8b is the positive sample distribution diagram of FCOS, which divides all samples with center points in the boundary box into positive samples without considering the influence of low-quality samples at the edge.

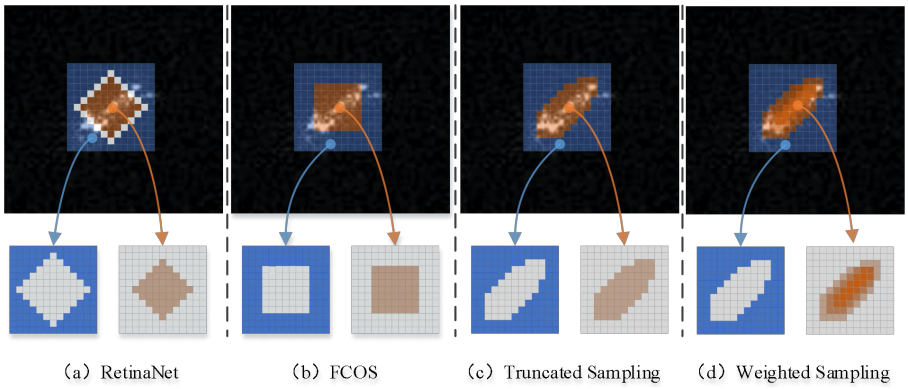


Figure 8. Positive sample distribution for different sampling strategies. (**a**) RetinaNet; (**b**) FCOS; (**c**) Truncated Sampling; (**d**) Weighted Sampling.

Since FCOS does not consider the low-quality samples, Gauss-ness is proposed. Specifically, we first calculate the sampling threshold by dividing 25% positions into positive samples, which is the quartile of the Gaussian distribution. The position where Gaussian value is greater than the threshold in the boundary box is divided into positive samples. Figure 8c shows the distribution of the Gaussian truncated sampling.

The low-quality samples generated at the edge of the bounding box are often hard samples in training. The existence of these samples brings noise to the training of the target detector. To reduce the influence of low-quality samples on the training process, we consider weighting the samples based on Gauss-ness, and introduce the adjustment factor k to regulate the degree of variance:

$$\text{Gauss-ness } = \exp\left(-\frac{1}{2k}(\boldsymbol{p}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{p}-\boldsymbol{\mu})\right) \quad (6)$$

k can adjust the variance of Gauss-ness between the center of the bounding box and the edge, and control the speed of edge attenuation. Figure 9 shows the change curve of Gauss-ness along the horizontal direction when k takes $\{1, 2, 3\}$ respectively. The abscissa adopts the distance after normalization. With the increase of k value, the attenuation velocity will gradually decrease. The experiment in Section 4.3.2 explores the impact of different k values on the detection performance. In the benchmark experiment, we take the k value as 2.

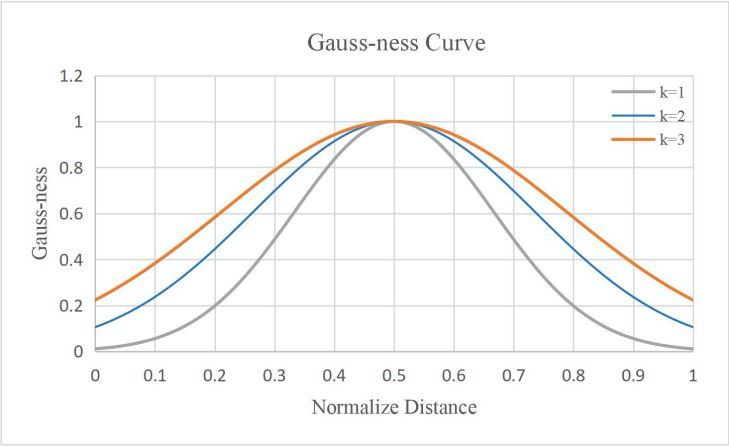


Figure 9. Gauss-ness curves at different K values.

In addition, considering that weighted sampling is a redistribution of the original samples and should not change the overall weight, normalization factor α is introduced to normalize the Gauss-ness in the boundary box of a single target so that its mean value remains to 1. The final sample weight can be described by the formula:

$$w = \begin{cases} \alpha \exp\left(-\frac{1}{2k}(\mathrm{x}-\mu)^T \sum^{-1}(\mathrm{x}-\mu)\right) & s \in S_{pos} \\ 0 & \text{others} \end{cases} \tag{7}$$

w is the sample weight, α represents the normalization factor, and is the average weight of positive samples in the bounding box. s represents the current sample and S_{POS} represents the set of positive samples. Figure 8d describes the distribution of positive samples with weighted sampling strategy. Orange represents positive samples and light to deep represents different weights. The final weighted loss function is described as follows:

$$L(p_{x,y}, t_{x,y}) = \frac{1}{N_{pos}} \sum_{x,y} w_{x,y} L_{cls}\left(p_{x,y}, c^*_{x,y}\right) + \frac{\lambda}{N_{pos}} \sum_{x,y} \delta_{c^*_{x,y}>0} w_{x,y} L_{reg}\left(t_{x,y}, t^*_{x,y}\right) \tag{8}$$

where L_{cls} is the focal loss [7] for classification, L_{reg} is the IoU-loss [44] for regression. Focal loss is proposed to solve the problem of severe imbalance in the proportion of positive and negative samples in end-to-end object detection. For regression loss, the commonly used $L2$ *loss* needs to optimize four independent variables of the bounding box at the same time, and leads to the model optimization focuses more on larger objects and ignores smaller ones. In order to solve this problem, we use *IoU loss* function to optimizes the bounding box as a whole. $w_{x,y}$, $p_{x,y}$ and $c^*_{x,y}$ separately represent the sampling weight, category score and real category at (x, y). $t_{x,y}$ stands for the predicted coordinates of (x, y), while $t^*_{x,y}$ is the ground truth value at (x, y). N_{pos} represents the number of positive samples, λ is set to 1 as the balance weight of L_{reg}. $\delta_{c^*_{x,y}>0}$ is an indication function, when $c^*_{x,y} > 0$, its value is 1, otherwise it is 0. Re-weighting the positive samples in the above way not only weakens the influence of noise samples that cannot be well modeled, but also helps to re-examine the central samples that are considered to be correctly classified, thus promoting the classification performance of detection.

4. Experiments and Analysis

4.1. Introduction to SAR Ship Dataset

To verify the effectiveness of the model in this paper, experiments on large-scale datasets are required. Due to the limitation of SAR imaging conditions, its related datasets are not as common as optical datasets. This paper conducts multiple experiments on two public SAR ship datasets. The statistics of these datasets are listed in Table 3. Detailed information, including the resolution, image size, number of images in the dataset, labeling method, is given for comparison. Each dataset is described in detail below.

Table 3. Statistics of two SAR ship detection datasets.

Datasets	Resolution (m)	Image Size (Pixel)	Images (n)	Annotations	Categories
SSDD [17]	1–15	190–668	1160	Bounding box	1
HRSID [16]	0.5, 1, 3	800 × 800	5604	Bounding box/Polygon	1

HRSID : HRSID is a -detection dataset of SAR images released by the University of Electronic Science and technology of China [16]. It adopts MS coco annotation format. The image sources are mainly sentinel-1b, TerraSAR-X and TanDEM-X, and the resolution is less than 3 m. The imaging area of the whole dataset includes the ocean surface with broad field of vision and simple background, and the offshore scene with more ships berthing and complex background. The offshore scenes are mainly ports, in which the ships are greatly affected by artificial facilities or buildings. In addition, the dense arrangement of ships is also a challenge. The ocean surface background is relatively simple, but its resolution is usually low and is greatly affected by wave clutter. The dataset cuts the target area to 800 × 800 size slices with resolutions ranging from 0.5 m to 3 m. Figure 10 shows some typical scenarios in the HRSID dataset.

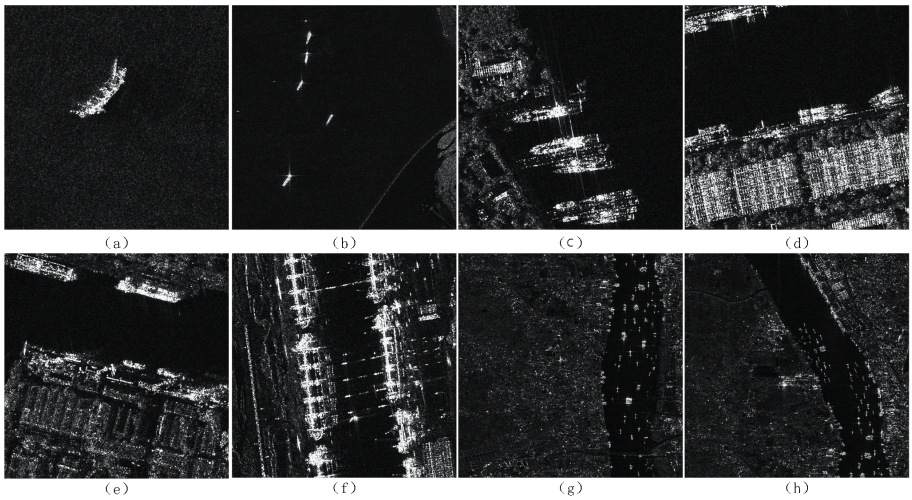


Figure 10. Typical legends in HRSID. (**a**) and (**b**) represent single-ship and multi-ship scenarios in the far sea, respectively; (**c**) and (**d**) for ships arranged in near shore; (**e**) and (**f**) are ships moored in the port; and (**g**) and (h) show small ships clustered in river channels.

HRSID contains 5604 high-resolution SAR images, providing detection and instance segmentation task annotation respectively, 65% of the images are divided into training sets and the remaining 35% are used as test sets. In Figure 11, the distribution of the area and width–height ratio of the bounding box in the training set and the test set is summarized, respectively. According to statistics, the percentages of small targets, medium targets and large targets in HRSID are 54.5%, 43.5%, and 2%, respectively, and small targets account

for more than half. Therefore, HRSID has the characteristics of small targets but large detection scene.

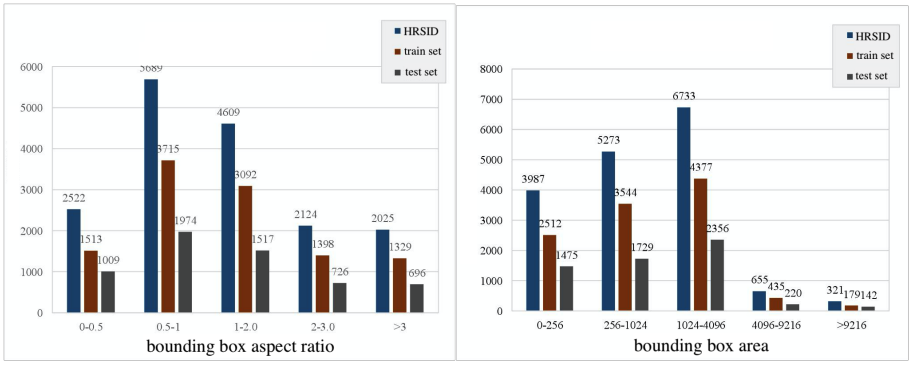
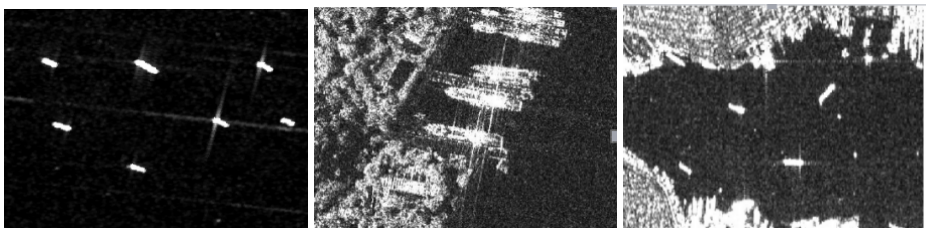


Figure 11. Statistical graph of target aspect ratio and area distribution in HRSID.

SSDD : SSDD is the first public SAR ship detection dataset constructed based on publicly downloaded SAR images [17]. It adopts PASCAL VOC annotation format. Images in the SSDD dataset were collected by radarSat-2, TerrasAR-X and Sentinel-1 sensors, including HH, HV, VV, and VH polarization modes, with resolutions ranging from 1 to 15 m. The image range includes the open sea area with relatively simple background and the nearshore port with complex background. Figure 12 shows typical scenarios in the SSDD dataset: (a) multiple ships in a simple background; (b) a multi-objective scenario against a complex near-shore background; (c) the scene of a dense array of ships close to shore. SSDD cuts the target area into 500 × 500 pieces, with a total of 1160 images, including 2456 ship targets. In this paper, the dataset are divided into training sets and test sets according to the ratio of 9:1, of which 930 are training sets and 230 are test sets. Compared with the HRSID, the size distribution of ship targets in SSDD is relatively concentrated and the average size is slightly larger.



(**a**) (**b**) (**c**)

Figure 12. Some samples from SSDD. (**a**) Simple background (**b**) complex background (**c**) nearshore targets.

4.2. Experimental Parameter Setting

In the adaptive feature encoding experiment, this paper added C_2 feature map to construct a new feature pyramid, and divided it into pyramid $\{M_2, M_3, M_4, M_5, M_6, M_7\}$. The target size range on feature maps are set as $(-1, 32)$, (32, 64), (64, 128), (128, 256), (256, 512), (512, INF). *GN* is selected as the normalization method, bilinear interpolation is selected as the up-sampling method in the adaptive feature encoding module.

In the experiment of optimized sampling method, the threshold of truncated sampling is set as the quartile of Gaussian distribution of corresponding boundary box, and the value of regulating factor k is 2.

For the setting of training parameters, the SGD optimizer was used, the initial learning rate was set to 0.005, and the attenuation rate of 0.1 was performed after the 16th and 22nd epochs, with a total of 24 epochs trained. To improve the stability of the mini-batch training, a linear Warmup of 500 iterations was used and the pre-training weight of ImageNet was used to initiate the backbone network. *GN* was used for regularization

processing, and the batch size was set to 4, training on NVIDIA Titan Xp GPU. During the test, the IoU threshold of the NMS is set to 0.5.

4.3. Experimental Results and Analysis

4.3.1. Experiment Evaluation of AFE

To evaluate the effect of the proposed adaptive feature encoding module, this section will analyze the experimental results from both qualitative and quantitative perspectives.

First of all, in order to verify the contribution of different configurations in AFE, this section uses FCOS as the baseline model to conduct ablation experiments on the HRSID. The experiments mainly include whether to construct the initial feature pyramid with high resolution feature map C_2 and whether to use GN in AFE. The test results under different conditions are listed in Table 4. FCOS obtained AP of 62.0% and AP_S of 63.6% on HRSID, and AP and AP_S increased to 64.7% and 66.2%, respectively, after the initial feature pyramid was reconstructed by C_2. After weight calculation by embedding the AFE module, AP and AP_S increased to 65.8% and 67.2%, respectively. After GN was added, the AP and AP_S of FCOS-AFE were further improved to 66.2% and 67.5%, respectively. Among them, the use of a high-resolution feature map showed the most obvious improvement in detection effect, and its AP and AP_S increased by 2.7% and 2.6%, respectively, while our AFE could continue to generate 1.5% and 1.3% improvement in AP and AP_S, respectively. This indicates that AFE can effectively enhance the semantic meaning of the introduced C_2 high-resolution feature map and enhance the detection ability of small objects.

Table 4. Ablation experiments for AFE on HRSID.

Detection Algorithm	C_2	Spatial Attention	GN	AP	AP_S
FCOS				62.0	63.6
FCOS	✓			64.7	66.2
FCOS	✓	✓		65.8	67.2
FCOS	✓	✓	✓	**66.2**	**67.5**

Then, the AP of AFE on HRSID and SSDD are evaluated. Table 5 shows the evaluation results on the HRSID dataset. AFE achieved the best performance at all IoU thresholds, with a 4.2% increase in AP relative to FCOS and a 3.9% increase in AP_S for small-target detection results. To compare the fairness, we try to use the parameters in MS COCO data set, without tuning the anchor parameters.

The AP of two-stage Faster RCNN is lower than that of single-stage RetinaNet, which is the worst because the advantage of two-stage Faster RCNN is the accuracy of location detection, while the more dense anchor setting is adopted in RetinaNet that are more suitable for HRSID with smaller target size on average, so the RetinaNet is 5.2% higher in AP_{50} than in Faster RCNN and 0.6% lower in AP_{75}, which is more accurate.

Table 5. Experimental results of the AFE Module on HRSID.

Detection Algorithm	Backbone Network	AP	AP_{50}	AP_{75}	AP_S
Faster RCNN	ResNet-50	60.4	80.6	69.1	61.1
RetinaNet	ResNet-50	61.6	85.8	68.5	62.9
FCOS	ResNet-50	62.0	87.3	69.7	63.6
FCOS + AFE(ours)	ResNet-50	**66.2**	**90.9**	**75.3**	**67.5**

Evaluation results on SSDD are shown in Table 6. Our AFE still achieved the best detection effect, and compared with FCOS, AP and AP_S increased by 1.7%. In contrast to the HRSID dataset, AP detected by Faster RCNN was only 0.1% lower than AFE, almost the same as AFE, and 1.6% higher than FCOS. According to the aforementioned analysis of target sizes in SSDD and HRSID, the average size of small targets in SSDD data sets is relatively large, which is more suitable for precise detection of Faster RCNN.

Table 6. Experimental results of the AFE Module on SSDD.

Detection Algorithm	Backbone Network	AP	AP_{50}	AP_{75}	AP_S
Faster RCNN	ResNet-50	53.7	91.3	53.4	46.8
RetinaNet	ResNet-50	52.0	91.8	51.6	46.5
FCOS	ResNet-50	52.1	93.8	53.2	46.3
FCOS + AFE(ours)	ResNet-50	**53.8**	**94.3**	**57.2**	**48.0**

Figure 13 shows the partial visualization results of our proposed AFE on the HRSID dataset. The first line of the figure shows the real annotation of the image, and the second and third lines show the detection results of FCOS and FCOS-AFE respectively. In the first column, FCOS has missed detection of small targets with extreme scales. However, our improved algorithm significantly improves the recall rate of small targets because it adopts a higher-resolution feature graph to construct a feature pyramid. The detection results of the second column show that the method based on adaptive feature encoding has high robustness for small-target detection in complex scenes, and has a good ability to distinguish objects easily confused with ships, such as land-based buildings and docks. The third section is the detection scenario of dense small and weak targets, which further proves the ability of our improved algorithm to detect small targets. The comparison of FCOS-AFE and FCOS detection results proves the effectiveness of our proposed adaptive feature encoding method for small-target detection.

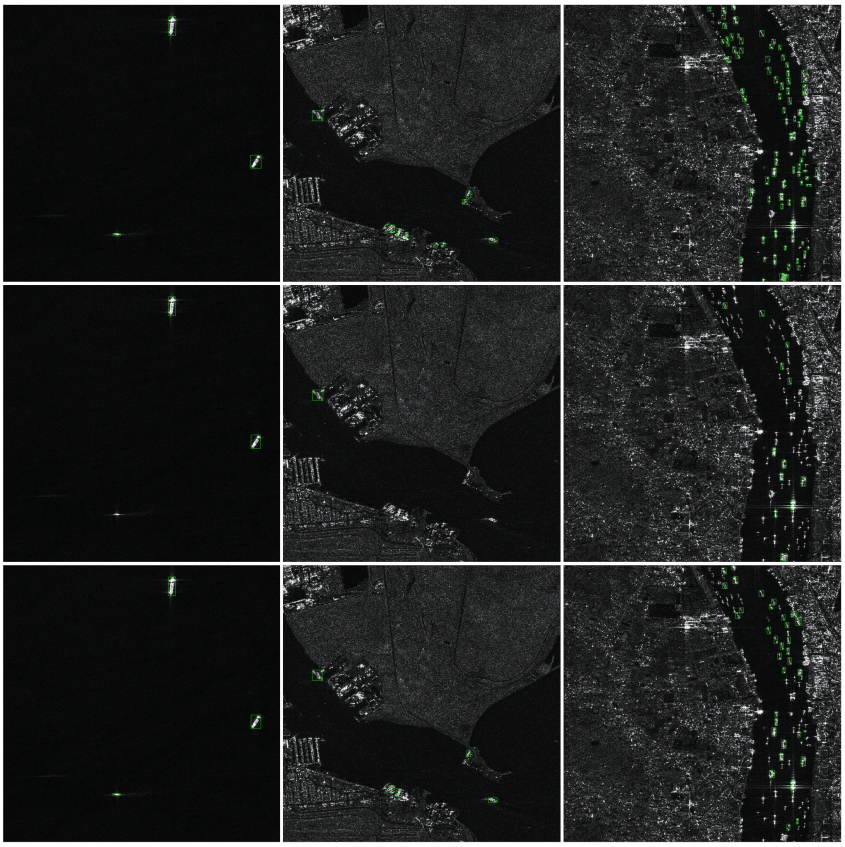


Figure 13. Comparison diagram of FCOS and AFE detection results.

In addition, we also selected P_2 from the feature pyramid and corresponding P2 feature map after feature encoding for visual analysis. P_2 and M_2 are both at the bottom of the feature pyramid and are responsible for detecting small targets with target area less than 32×32 pixels. The left side of Figure 14 is the visualization result of P_2. On the right is the visualization result of M_2. It can be seen that M_2 with adaptive feature encoding has a more obvious activation effect on small targets, confirming the effectiveness of AFE.

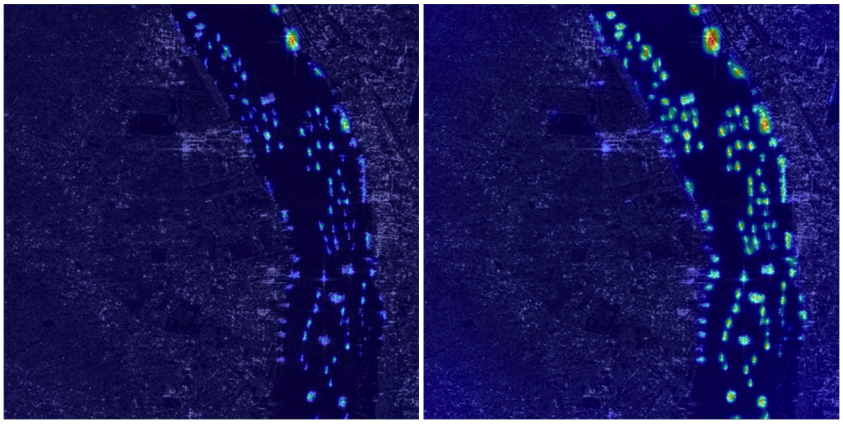


Figure 14. Feature activation heatmap w/o AFM.

4.3.2. Experiment Evaluation of Gaussian-Guided Detection Head

This section mainly carries on the quantitative experimental analysis to the sampling optimization method proposed in Section 3. In addition, the value of the regulator k of weighted sampling is also investigated experimentally.

To achieve a better detection effect, we conducted an experimental study on the value of the regulating factor k. The detection results of Gauss-FCOS with weighted sampling in Table 7 are obtained when k is set to 1, 2, 3 respectively. It can be seen that, as the value of k increases, the detection effect AP_{50} when IoU threshold is 0.5 gradually improves. However, AP_{75} with a more strict IoU threshold of 0.75 gradually decreases, indicating that with the gradual relaxation of the weight (that is, assuming that the pixel distribution of the target is more dispersed), the recall rate of the algorithm increases, but the location accuracy decreases. As AP and AP_S of all IoU thresholds are mainly used as evaluation indexes, k is 2 in the preceding experiment.

Table 7. Detection results on SSDD when k takes different values.

k	AP	AP_{50}	AP_{75}	AP_S
1	52.8	92.1	**55.6**	46.0
2	**53.2**	93.5	53.4	46.7
3	52.8	**93.9**	53.6	**46.8**

The evaluation results on the HRSID are shown in Table 8. The network replacing center-ness in FCOS with Gauss-ness is named Gauss-FCOS, which improved by 0.9% and 1.1% in AP and AP_S, respectively. Although the improvement is not much, its advantage is that it will not cause any burden to the network computation and detection speed. After soft sampling with Gauss-ness weighted sampling strategy (GWS is short for Gauss-ness weighted sampling), the AP_{50} and AP_{75} achieved improvements of 1.6% and 1.1%, respectively, and the detected AP and AP_S reached their best results: 63.4% and 66.1%, respectively.

Table 8. Experimental results of the Gaussian function-based detection head on HRSID and SSDD.

Detection Algorithm	Backbone Network	HRSID				SSDD			
		AP	AP_{50}	AP_{75}	AP_S	AP	AP_{50}	AP_{75}	AP_S
FCOS	ResNet-50	62.0	87.3	69.7	63.6	52.1	93.8	53.2	46.3
Gauss-FCOS	ResNet-50	62.9	88.2	68.3	64.7	53.5	94.3	52.9	47.4
Gauss-FCOS + GWS	ResNet-50	**63.4**	**88.9**	**70.8**	**66.1**	**54.0**	**95.6**	**53.6**	**48.8**

As in Table 8, after introducing the Gauss-ness branch, SSDD results of Gauss-FCOS on AP and AP_S increased by 1.4% and 1.1%, from 52.1% and 46.3% to 53.5% and 47.4%,

respectively. The AP in SSDD is improved greatly. Based on the weighted sampling strategy, AP and AP_S increased to 54.0% and 48.8%, respectively, and increased by 1.9% and 2.5% compared with FCOS. Experiments on SSDD show that, compared with truncation sampling (TS), weighted sampling can distinguish samples of different qualities more effectively, and improve the detection ability of the algorithm.

4.3.3. Ablation Experiment of the Overall Framework

We combined the proposed adaptive feature coding (AFE) with Gaussian-guided detection head (GDH), and constructed a detector suitable for small target detection based on the anchor-free idea. Experiments were conducted on HRSID and SSDD data sets, respectively.

Table 9 shows the ablation analysis of detector on HRSID. After AFE is embedded, AP and AP_S are increased from 62.0% and 63.6% to 66.2% and 67.5%, respectively. Finally, AP and AP_S are increased by 5.4% and 5.3%, respectively, after adding the proposed Gaussian-guided detection head, which verifies the effectiveness of the detector in detection accuracy. By analyzing its detection speed, it can be seen that the decrease of detection speed of our method is mainly caused by AFE, while GWS has almost no influence on its detection speed. Analysis of the reasons for the speed decline caused by AFE is as follows: first, by introducing C_2, the computation of feature map and positive sampling will be improved; second, the process of adaptive feature encoding increases the process of network.

Table 9. Experimental results of the two module on HRSID.

Baseline	AFE	GDH	AP	AP_S	FPS
FCOS			62.0	63.6	**18.9**
FCOS	✓		66.2	67.5	15.4
FCOS		✓	63.4	65.7	18.5
FCOS	✓	✓	**67.4**	**68.9**	15.2

Table 10 shows the ablation analysis of detector on SSDD. After AFE is embedded, the AP and AP_S are improved from 52.1% and 46.3% to 53.8% and 48.0%, respectively. And with the GDH method, AP increased by 1.9% and AP_S increased by 2.1%, reaching 54.0% and 48.4%, respectively. Combining the two methods results in up to a 4.1% AP boost and a 4.5% AP_S boost. The detector still achieves excellent detection effect on SSDD. Similarly, the decrease of its FPS was mainly caused by AFE, and GDH had almost no effect on it.

Table 10. Experimental results of the two module on SSDD.

Baseline	AFE	GDH	AP	AP_S	FPS
FCOS			52.1	46.3	**35.8**
FCOS	✓		53.8	48.0	28.9
FCOS		✓	54.0	48.4	35.5
FCOS	✓	✓	**56.2**	**50.8**	28.5

Table 9 shows the overall evaluation of the detector on the two dateset. Compared with other detector, AP_50 of our method on HRSID improved from 87.3% to 92.0% and AP_S improved from 63.6% to 68.9%. By comparing the prediction speed of these algorithms, we can see that FCOS has the highest frame number, reaching 18.9, while the frame number of our proposed detector drops obviously to only 15.2, which is also the biggest defect of our proposed algorithm.

As for SSDD dataset, our method achieves the best performance in all indicators, in which AP_50 improves by 2.7%, from 93.8% to 96.5%, and AP_S improves by 4.5%, from 46.3% to 50.8%. However, the frame number of the detector also produces a drop on SSDD, which is only 28.5.

4.3.4. Comparison of Performance Between the Proposed Overall Framework and the State-of-the-Art

In this section, the proposed detection framework is compared to several other detectors. Our method outperforms all other comparison methods on these data sets.

First, we conduct experiments to compare the performance with the commonly used two-parameter CFAR detector. The quantitative detection results are shown in Table 11. Traditional CFAR typically does not use AP to measure accuracy, so precision and recall are used to evaluate performance. Furthermore, CFAR typically runs on the CPU, while modern CNN-based approaches always run on the GPU. To ensure a reasonable comparison, we chose CPU time for speed comparison (t_{CPU}). It can be seen from Table 11 that our method is much better than CFAR in detection precision, which is nearly 30% higher than CFAR. The difference in recall is not so obvious, and the recall of our method is 13.26% higher than CFAR. It can be explained that the traditional CFAR algorithm cannot adapt to complex scenes and small targets, and many false alarms are detected, resulting in a significant decrease in precision. In terms of detection speed, our method is also much better than CFAR. In other words, the running time of our method on CPU is 0.356 s, which is far less than that of CFAR.

Table 11. Quantitative comparison with CFAR on SSDD.

Method	*Precision* (%)	*Recall* (%)	t_{CPU} (s)
CFAR	61.23	81.49	1.56
Ours	96.21	94.75	0.356

The comparison CNN-based methods in the experiments can be divided into two categories, namely anchor-based methods, such as YOLOv3-tiny, YOLOV3, RetinaNet, Faster R-CNN, Cascade R-CNN, Mask-RCNN, and anchor-free methods such as CornerNet and FCOS. Moreover, two SAR ship detection methods, DCMSNN and FBR-Net, are compared with our proposed detector. Note that some of these methods employ different backbones than our method, and the experimental settings on training of them are taken from the original articles. As shown in Table 12, the anchor-free method FCOS achieves a leading performance among all the compared methods. Our method used to improve FCOS achieves the best performance, demonstrating the effectiveness of the method for ship detection in remote sensing images. Furthermore, thanks to the complementary effect of the proposed components, the AP_s of our method is 5.3% higher than that of the FCOS method. It can be seen from the experiments on SSDD that the method proposed in this paper achieves an AP value of 96.5% and shows the best performance. In particular, our method significantly outperforms the baseline methods. Furthermore, the experimental results on these two datasets further demonstrate the robustness of our method in adapting to different datasets. Although the method proposed in this paper has outstanding performance on small targets, it also has the ability to detect medium and large targets (scale > 64 $\times$ 64 & scale > 128 $\times$ 128). Table 12 shows that AP_{50} is also improved while AP_S increases, which proves that the network does not lose the ability to detect larger-scale targets.

To demonstrate the advantages of our method over previous methods, we show some visual results. Figures 15 and 16 are the detection results of the overall detection framework on the two datasets, respectively. The numbers above the detected bounding boxes in the figure represent the confidence of it, the confidence of the detected boxes is high as can be seen in the figure. For the densely distributed ships in HRSID and the near-shore ships in SSDD, our method can achieve good results, indicating that our network has strong robustness to different scenes, and the quality of detected boxes is higher.

It is worth mentioning that our method still has deficiencies in some aspects. Specifically, AFE will lead to a certain decrease in detection speed. The comparison results of speed and accuracy are shown in Tables 9 and 10. Note that NMS post-processing is

included at runtime for each image. The results in the table show that after adding the AFM module, the test speed will drop by about 20%. This drop is mainly due to the increase in operations caused by feature fusion. This problem is a focus of follow-up work.

Table 12. Experimental results of the overall detection framework on HRSID and SSDD.

Algorithm		*HRSID*		*SSDD*	
		AP_{50}(%)	AP_S(%)	AP_{50}(%)	AP_S(%)
anchor-based	YOLOv3-tiny	70.4%	51.8%	64.0%	36.2%
	YOLOV3	74.0%	56.3%	67.0%	41.4%
	RetinaNet	74.0%	56.3%	67.0%	42.6%
	Faster R-CNN	79.2%	57.3%	85.9%	42.3%
	Cascade R-CNN	79.1%	59.9%	87.1%	44.9%
	Mask-RCNN	81.1%	57.2%	87.4%	43.1%
	DCMSNN [35]	83.4%	61.3%	89.4%	46.5%
	FBR-Net [45]	89.7%	65.8%	94.1%	48.6%
anchor-free	CornerNet	73.6%	50.2%	74.3%	36.7%
	FCOS	87.3%	63.6%	93.8%	46.3%
	FCOS(AFE)	90.9%	67.5%	94.3%	48.0%
	FCOS(AFE-GDH)	**92.0%**	**68.9%**	**96.5%**	**50.8%**

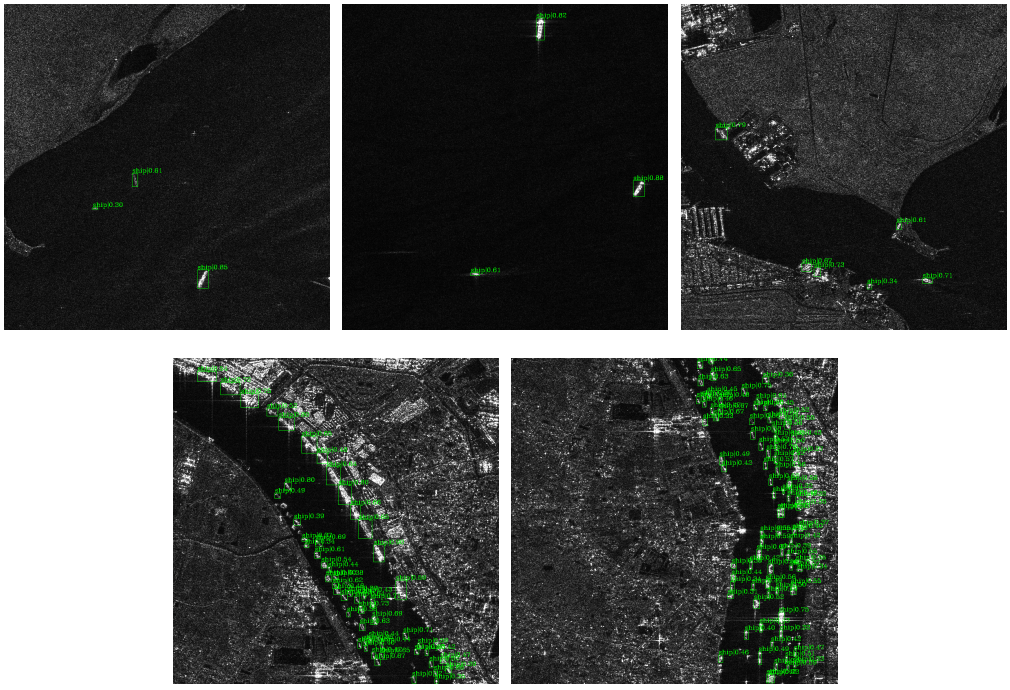


Figure 15. Some of the detection results obtained by the proposed overall framework on HRSID.

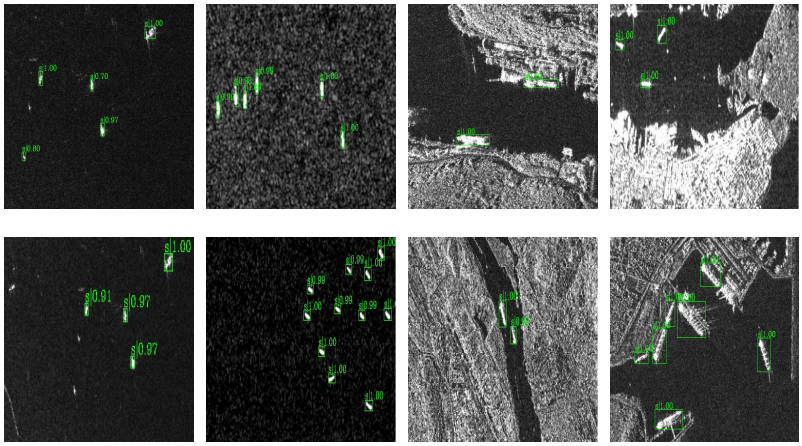


Figure 16. Some of the detection results obtained by the proposed overall framework on SSDD.

5. Conclusions

In this paper, the feature representation and sampling strategy of SAR small-target detection are studied respectively, and adaptive feature encoding (AFE) and Gauss-guided detection head (GDH) are proposed. AFE is used to effectively integrate the semantic information in the deep feature map into the shallow layer, so as to enhance the feature representation of small targets. Moreover, aiming at the problem of low-quality samples at the edge of the boundary box in the sampling of the anchor-free detectors, this paper proposes a series of sampling optimization methods by using the Gaussian prior distribution of the target and construct the GDH. Specifically, the proposed Gauss-ness uses two-dimensional Gaussian function to fit the target distribution in the boundary box during training, and is more consistent with the target shape. Then, truncation sampling and weighted sampling based on gauss-ness are proposed to optimize the network training process. Finally, the AFE proposed in Section 3.2 is combined with the GDH in Section 3.3. In the two-module fusion experiment, the effects of the two improved methods are confirmed, and the highest AP_s improvement is 5.3% on HRSID, and the highest AP_S improvement is 4.5% on SSDD.

Author Contributions: Conceptualization, B.H.; methodology, B.H.; software, M.T.; validation, B.H., Q.Z.; writing—original draft preparation, B.H.; writing—review and editing, Q.Z.; C.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (No. 2016YFC0803000) and the National Natural Science Foundation of China (No. 41371342).

Data Availability Statement: Not applicable.

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

References

1. Yang, X.; Sun, H.; Fu, K.; Yang, J.; Sun, X.; Yan, M.; Guo, Z. Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks. *Remote Sens.* **2018**, *10*, 132. [CrossRef]
2. Tian, L.; Cao, Y.; He, B.; Zhang, Y.; He, C.; Li, D. Image enhancement driven by object characteristics and dense feature reuse network for ship target detection in remote sensing imagery. *Remote Sens.* **2021**, *13*, 1327. [CrossRef]
3. Farina, A.; Studer, F.A. A review of CFAR detection techniques in radar systems. *Microw. J.* **1986**, *29*, 115.
4. 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.
5. 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, 27–30 June 2016; pp. 779–788.
6. 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 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
7. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.

8. Law, H.; Deng, J. Cornernet: Detecting objects as paired keypoints. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 734–750.
9. Zhu, C.; He, Y.; Savvides, M. Feature selective anchor-free module for single-shot object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 840–849.
10. Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. Centernet: Keypoint triplets for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 6569–6578.
11. 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, Korea, 27 October–2 November 2019; pp. 9627–9636.
12. Chen, X.; Fang, H.; Lin, T.Y.; Vedantam, R.; Gupta, S.; Dollár, P.; Zitnick, C.L. Microsoft coco captions: Data collection and evaluation server. *arXiv* **2015**, arXiv:1504.00325.
13. 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 2014, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
14. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
15. Xu, D.; Wu, Y. FE-YOLO: A feature enhancement network for remote sensing target detection. *Remote Sens.* **2021**, *13*, 1311. [CrossRef]
16. 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]
17. Li, J.; Qu, C.; Shao, J. Ship detection in SAR images based on an improved faster R-CNN. In Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), Beijing, China, 13–14 November 2017; pp. 1–6.
18. Hou, J.B.; Zhu, X.; Yin, X.C. Self-adaptive aspect ratio anchor for oriented object detection in remote sensing images. *Remote Sens.* **2021**, *13*, 1318. [CrossRef]
19. 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.* **2016**, *39*, 1137–1149. [CrossRef]
20. Redmon, J.; Farhadi, A. YOLO9000: better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
21. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
22. Zhou, X.; Zhuo, J.; Krahenbuhl, P. Bottom-up object detection by grouping extreme and center points. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 850–859.
23. Zhang, S.; Chi, C.; Yao, Y.; Lei, Z.; Li, S.Z. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 9759–9768.
24. Kong, T.; Sun, F.; Liu, H.; Jiang, Y.; Li, L.; Shi, J. Foveabox: Beyound anchor-based object detection. *IEEE Trans. Image Process.* **2020**, *29*, 7389–7398. [CrossRef]
25. 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.
26. Gao, G.; Liu, L.; Zhao, L.; Shi, G.; Kuang, G. An Adaptive and Fast CFAR Algorithm Based on Automatic Censoring for Target Detection in High-Resolution SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 1685–1697. [CrossRef]
27. Huo, W.; Huang, Y.; Pei, J.; Zhang, Q.; Gu, Q.; Yang, J. Ship Detection from Ocean SAR Image Based on Local Contrast Variance Weighted Information Entropy. *Sensors* **2018**, *18*, 1196. [CrossRef]
28. Gao, G.; Gao, S.; He, J.; Li, G. Ship Detection Using Compact Polarimetric SAR Based on the Notch Filter. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5380–5393. [CrossRef]
29. Yue, B.; Zhao, W.; Han, S. SAR Ship detection method based on convolutional neural network and multi-layer feature fusion. In *Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery*; Springer: Cham, Switzerland, 2019; pp. 41–53. [CrossRef]
30. Dai, W.; Mao, Y.; Yuan, R.; Liu, Y.; Pu, X.; Li, C. A Novel Detector Based on Convolution Neural Networks for Multiscale SAR Ship Detection in Complex Background. *Sensors* **2020**, *20*, 2547. [CrossRef]
31. 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]
32. Shiqi, C.; Ronghui, Z.; Jun, Z. Regional attention-based single shot detector for SAR ship detection. *J. Eng.* **2019**, *2019*, 7381–7384. [CrossRef]
33. Gao, F.; Shi, W.; Wang, J.; Yang, E.; Zhou, H. Enhanced Feature Extraction for Ship Detection from Multi-Resolution and Multi-Scene Synthetic Aperture Radar (SAR) Images. *Remote Sens.* **2019**, *11*, 2694. [CrossRef]
34. Bell, S.; Zitnick, C.L.; Bala, K.; Girshick, R. Inside-Outside Net: Detecting objects in context with skip pooling and recurrent neural networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016. [CrossRef]
35. Jiao, J.; Zhang, Y.; Sun, H.; Yang, X.; Gao, X.; Hong, W.; Fu, K.; Sun, X. A Densely Connected End-to-End Neural Network for Multiscale and Multiscene SAR Ship Detection. *IEEE Access* **2018**, *6*, 20881–20892. [CrossRef]
36. Pan, Z.; Yang, R.; Zhang, Z. MSR2N: Multi-Stage Rotational Region Based Network for Arbitrary-Oriented Ship Detection in SAR Images. *Sensors* **2020**, *20*, 2340. [CrossRef]

37. Tian, T.; Pan, Z.; Tan, X.; Chu, Z. Arbitrary-Oriented Inshore Ship Detection based on Multi-Scale Feature Fusion and Contextual Pooling on Rotation Region Proposals. *Remote Sens.* **2020**, *12*, 339. [CrossRef]
38. Zhang, F.; Wang, X.; Zhou, S.; Wang, Y.; Hou, Y. Arbitrary-Oriented Ship Detection Through Center-Head Point Extraction. *IEEE Trans. Geosci. Remote. Sens.* **2022**, *60*, 1–14. [CrossRef]
39. Chen, P.; Li, Y.; Zhou, H.; Liu, B.; Liu, P. Detection of Small Ship Objects Using Anchor Boxes Cluster and Feature Pyramid Network Model for SAR Imagery. *J. Mar. Sci. Eng.* **2020**, *8*, 112. [CrossRef]
40. Li, H.; Wu, Z.; Zhu, C.; Xiong, C.; Socher, R.; Davis, L.S. Learning from noisy anchors for one-stage object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10588–10597.
41. Zhu, B.; Wang, J.; Jiang, Z.; Zong, F.; Liu, S.; Li, Z.; Sun, J. Autoassign: Differentiable label assignment for dense object detection. *arXiv* **2020**, arXiv:2007.03496.
42. Wang, J.; Chen, K.; Yang, S.; Loy, C.C.; Lin, D. Region proposal by guided anchoring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2965–2974.
43. 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, Korea, 27 October–2 November 2019; pp. 6054–6063.
44. Yu, J.; Jiang, Y.; Wang, Z.; Cao, Z.; Huang, T. UnitBox. In Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016. [CrossRef]
45. Fu, J.; Sun, X.; Wang, Z.; Fu, K. An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1331–1344. [CrossRef]