electronics

MDPI

Article

CCDS-YOLO: Multi-Category Synthetic Aperture Radar Image Object Detection Model Based on YOLOv5s

Min Huang [1], Zexu Liu [1], Tianen Liu [1] and Jingyang Wang [1,2,*]

[1] Hebei University of Science and Technology, Shijiazhuang 050018, China
[2] Hebei Technology Innovation Center of Intelligent IoT, Shijiazhuang 050018, China
* Correspondence: jingyangw@hebust.edu.cn

Abstract: Synthetic Aperture Radar (SAR) is an active microwave sensor that has attracted widespread attention due to its ability to observe the ground around the clock. Research on multi-scale and multi-category target detection methods holds great significance in the fields of maritime resource management and wartime reconnaissance. However, complex scenes often influence SAR object detection, and the diversity of target scales also brings challenges to research. This paper proposes a multi-category SAR image object detection model, CCDS-YOLO, based on YOLOv5s, to address these issues. Embedding the Convolutional Block Attention Module (CBAM) in the feature extraction part of the backbone network enables the model's ability to extract and fuse spatial information and channel information. The 1×1 convolution in the feature pyramid network and the first layer convolution of the detection head are replaced with the expanded convolution, Coordinate Conventional (CoordConv), forming a CRD-FPN module. This module more accurately perceives the spatial details of the feature map, enhancing the model's ability to handle regression tasks compared to traditional convolution. In the detector segment, a decoupled head is utilized for feature extraction, offering optimal and effective feature information for the classification and regression branches separately. The traditional Non-Maximum Suppression (NMS) is substituted with the Soft Non-Maximum Suppression (Soft-NMS), successfully reducing the model's duplicate detection rate for compact objects. Based on the experimental findings, the approach presented in this paper demonstrates excellent results in multi-category target recognition for SAR images. Empirical comparisons are conducted on the filtered MSAR dataset. Compared with YOLOv5s, the performance of CCDS-YOLO has been significantly improved. The mAP@0.5 value increases by 3.3% to 92.3%, the precision increases by 3.4%, and the mAP@0.5:0.95 increases by 6.7%. Furthermore, in comparison with other mainstream detection models, CCDS-YOLO stands out in overall performance and anti-interference ability.

Keywords: target recognition; SAR; YOLOv5; multi-category; deep learning

Citation: Huang, M.; Liu, Z.; Liu, T.; Wang, J. CCDS-YOLO: Multi-Category Synthetic Aperture Radar Image Object Detection Model Based on YOLOv5s. *Electronics* **2023**, *12*, 3497. https://doi.org/10.3390/electronics12163497

Academic Editor: Jyh-Cheng Chen

Received: 16 July 2023
Revised: 12 August 2023
Accepted: 16 August 2023
Published: 18 August 2023

1. Introduction

Known for its ability to observe with high resolution in any weather conditions and continuously, SAR is extensively applied in various domains, including earth observation, object detection, and classification [1–3]. Because of these characteristics, SAR is suitable for the military, disaster monitoring, and marine resource exploration, among other fields. With the rapid update of tools, information, and techniques, many SAR images have been acquired. SAR image object detection aims to realize automatic positioning and the recognition of specific targets. It provides a wide range of practical application prospects for various fields. Precisely obtaining the geographical coordinates of designated military objectives is significant in optimizing coastal defense early warning capabilities and facilitating strategic deployment in military scenarios. On the civilian side, detecting smuggling and illegal fishing vessels contributes to monitoring and managing maritime transportation [4,5]. In some special cases, detection goes beyond common ships or planes. It can involve the simultaneous detection of various strategic targets like ships, planes,

bridges, and oilcans in an area; site searching; real-time monitoring; and early warning. The SAR application scenarios are shown in Figure 1.

Figure 1. SAR application scenarios.

With the increase in SAR image data, SAR image object detection is still facing many difficulties. First of all, the gray value of neighboring pixel points will have some random changes by correlation. This stochastic change operates around a particular average, producing spots of image noise. The generation of speckle noise [6] is due to defects caused by the coherence principle of SAR imaging, so it needs manual resolution. Secondly, there are often significant differences in the sizes of targets in SAR images, especially in some scenes, posing challenges to current object detection techniques.

The Constant False Alarm Rate (CFAR) [7] is a technique based on measuring contrast while minimizing the occurrence of false alarms. A key benefit of the CFAR algorithm is its ability to maintain a constant false alarm rate under varying background conditions. The CFAR algorithm is robust and can achieve reliable object detection in tough marine environments. There are many ways to implement the CFAR algorithm, such as Cell Average CFAR (CA-CFAR), Sequence Average CFAR (SA-CFAR), and Training Cell Average CFAR (TCA-CFAR). The appropriate implementation strategy can be chosen according to specific application requirements. However, when dealing with complex backgrounds with multi-scale and multi-directional structures, the CFAR algorithm is susceptible to the influence of noise speckles. It might struggle to adapt to these changes effectively, which could result in a decrease in detection performance and may not meet the needs of current SAR target recognition. Traditional detection methods, which rely on prior information, can be easily affected by complex environments or background noise and have low calculation speeds.

The complex texture features of SAR images and the similarity between targets bring great challenges to object detection and recognition [8]. Recently, Convolutional Neural Networks (CNN) have been applied to computer vision, such as image detection [9], semantic classification, and other tasks [10]. However, a simple CNN cannot meet the detection tasks of SAR images. The size and scale of objects in SAR images vary widely, which may make it difficult for a simple CNN model to adjust to different scales and lead to the degradation of object detection performance. And SAR images are usually affected by noise and other disturbances, which may interfere with the visualization and analysis of targets. As CNNs can be sensitive to noise and interference, these factors can affect object detection accuracy. Therefore, they have great potential to apply deep learning methods to SAR image object detection and recognition. Currently, convolutional-neural-network-based algorithms mainly used for SAR image detection include a two-stage detection modus delegated by Region-CNN (R-CNN), Fast R-CNN, Faster R-CNN [11], and DETR [12]. The basic process of these two-stage algorithms includes candidate box generation, classification, and the refinement of candidate boxes. R-CNN first brings many candidate regions and then inputs these regions into the CNN for feature extraction. The extracted features are utilized by a Support Vector Machine (SVM) to perform classification while bounding box regression adjusts the predicted boxes. In addition, Fast R-CNN unifies the classification and bounding box regression tasks into one network to achieve end-to-end training. However, the two-stage algorithm also brings a lot of computational overhead while obtaining high precision. In comparison, the single-stage algorithm maintains a faster speed. SSD [13], FCOS [14], and the YOLO [15] series of single-stage algorithms directly predict bounding box coordinates and class labels rather than using a two-step process or employing an additional stage, where object location and class information need to be regressed. YOLO is the mainstream single-stage object detection algorithm at present. YOLO is simple and efficient and can perform object detection faster, with real-time performance [16]. Each generation of the YOLO algorithm has its outstanding advantages. Overall, the YOLO series algorithm is more suitable for processing SAR image detection tasks due to YOLO's high real-time performance, simple algorithm design, and lightweight model [17].

Based on these studies, this paper further improves the YOLOv5 model to meet the needs of multi-category SAR image target recognition tasks. As the fields of deep learning and object detection rapidly progress, the technology to tackle object detection and recognition problems in common scenarios is nearing perfection. However, since SAR images are very different from general natural scene pictures, several issues persist, such as the model being sensitive to the speckle noise specific to SAR images and insufficient adaptability to target deformation and occlusion. For multi-category SAR image datasets, the oilcans are usually closely clustered together, which is likely to cause a large problem of missed detection. Additionally, the large-scale gap between oilcans and bridges within the target category underlines the importance of multi-scale target recognition capabilities. To advance the detection technology of SAR images and make the model achieve better results in detecting multi-category objects, we must solve the above problems.

This paper addresses the task of multi-category SAR image object detection by refining the lightweight YOLOv5s algorithm and studying feature extraction methods to further enhance the detection performance. Compared with other related techniques, experimental results show that the proposed model excels in multi-category SAR image object recognition tasks.

To sum up, the main contributions of this paper are the following four points:

(1) This paper proposes a new multi-category SAR image detection model named CCDS-YOLO. The model exhibits robust detection capabilities for multi-scale and multi-category objects in complex backgrounds. Aiming at the compact aggregation of some targets in SAR images and the large gap between the scales of different types of targets, this paper embeds the CBAM in the backbone network. The attention

mechanism layer can better integrate each layer's feature information and improve the model's detection effect on targets of different scales.

(2) To tackle the challenges of significant target scale variance in SAR images and environmental disturbances, this paper proposes a CRD-FPN module. The feature extraction part and part of the ordinary convolution in the detection network are replaced by CoordConv. On the basis of ordinary convolution, CoordConv adds two coordinate channels to obtain the spatial information of the graph of features, which improves the model's processing ability for regression tasks.

(3) To improve the comprehensive ability of model detection and classification, the original detection head part is replaced by the decoupled head. This provides two distinct computation channels for classification and regression tasks, enabling the model to acquire appropriate feature information. Consequently, the model achieves superior results in classification and regression tasks.

(4) We replaced NMS in the detection network with the Soft-NMS algorithm. This mitigates traditional NMS methods' struggles with overlapping bounding boxes effectively and reduces the false detection rate and missed detection rate due to compact goals.

2. Related Work

The traditional SAR image object detection methods are as follows. The feature-based process mainly extracts some features of the target in the SAR image, such as shape, size, and texture. These features are manually designed and may be affected by factors like target scale, rotation, and noise. Tan et al. proposed a new SAR image adaptive aircraft target detection algorithm. The algorithm first detects the airport candidate area, then obtains the gradient texture saliency map, and finally uses CFAR to segment the saliency map to obtain the target. Template matching methods usually require high computational complexity and are affected by object rotation, scale change, and noise. Regarding SAR object detection, Hong et al. [18] advocated a method based on anchor box matching. Statistical-learning-based methods achieve object detection by learning the statistical properties between objects and backgrounds. These methods usually need a lot of training data and may be affected by the target class imbalance. Marti et al. [19] proposed a statistical-learning-based SAR object detection approach.

Deep learning has been widely employed in SAR image object detection in recent years. Among the two-stage object detection algorithms in deep learning are R-CNN, Fast R-CNN, and Faster R-CNN. Wang et al. [20] introduced a ship detection method suitable for SAR images. Traditional approaches are limited by factors such as weather, and their robustness needs improvement. In this study, a deep learning model, RetinaNet, is employed. It utilizes a Feature Pyramid Network (FPN) to extract multi-scale features and addresses class imbalance using focal loss, thereby enhancing the training weight of hard examples. Liao et al. [21] applied R-CNN to object detection in SAR images. Fast R-CNN is an enhanced version of R-CNN that incorporates improvements. By introducing the pooling layer, the features of multiple candidate regions are mapped to a feature graph with a fixed size, thus reducing the computational complexity. Li et al. [22] proposed an SAR image object detection method based on Faster R-CNN to balance detection accuracy and speed. Its innovative lightweight backbone network integrates feature amplification using relay and multi-scale feature skip connections, facilitating the recognition of objects at various scales and thereby improving accuracy. The use of RoIAlign instead of traditional Region of Interest pooling reduces quantization errors in localization.

Some scholars have attempted to apply the YOLO series models to SAR image object detection. For example, Chen et al. [23] developed a ship detection method for SAR images using YOLOv3. Acknowledging the need for both accuracy and speed, they leverage the efficient YOLOv3 architecture for its rapid detection. To improve small ship detection, a lightweight Dilated Attention Module (DAM) is introduced, aiding in discriminative feature extraction. DAM suppresses irrelevant regions and highlights ship-

related features. However, this approach's focus on detecting small-sized ships might impact its suitability for multi-scale object detection scenarios. Lin et al. [24] discussed the significance of ship target detection for preserving marine interests and introduced an SAR image ship detection algorithm based on an enhanced YOLOv4 model. This study employed K-means [25] clustering to adjust the anchor boxes of YOLOv4 in order to address the reduction in detection accuracy caused by mismatches between anchor boxes and ship target sizes. The algorithm's effectiveness is validated on SSDD datasets, exhibiting a 2.87% detection accuracy improvement over the original YOLOv4 while maintaining detection efficiency. A limitation of this algorithm lies in its dependency on k-means clustering, which might not optimally handle complex ship shapes in some cases. Luo et al. [26] introduced an Efficient Bidirectional Path Aggregation Attention Network (EBPA2N) for detecting aircraft in SAR images. To address challenges related to shattered features, size heterogeneity, and complex backgrounds, YOLOv5s serves as the base network, complemented by the integration of the Involution Enhanced Path Aggregation (IEPA) and Effective Residual Shuffle Attention (ERSA) modules. IEPA extracts semantic and spatial information from multi-scale scattering features, while ERSA enhances features to mitigate background interference and false alarms. Luo et al. [27] proposed an explainable AI (XAI) framework to enhance the interpretability of Deep Neural Networks (DNNs) for aircraft detection in SAR images. The framework includes Hybrid Global Attribution Mapping (HGAM) for network selection, Path Aggregation Network (PANet) for feature fusion, and Class-specific Confidence Scores Mapping (CCSM) for visualization. It improves DNN comprehension but may have implementation complexity. Overall, it is a valuable contribution to enhancing transparency in SAR image analytics. However, a potential drawback of this framework could be its complexity in implementation as it combines multiple techniques. Additionally, the scalability of this framework to other target detection tasks beyond aircraft detection might require further investigation.

3. Models

3.1. YOLOv5s

Considering the specific characteristics of SAR object detection, we chose YOLOv5 to construct the feature extraction network to ensure appropriate parameters and meet real-time detection requirements. The YOLO series is currently the most popular single-stage object detection model. YOLOv5 [28] includes four versions: YOLOv5l, YOLOv5m, YOLOv5x, and YOLOv5s. YOLOv5s not only has high precision and relatively fast speed but also has fewer parameters. YOLOv5s can be applied to real-time detection tasks for space-borne satellites with limited hardware resources. It is suitable for target detection in SAR images.

YOLOv5s usually comprises four components: the input, backbone network, network layer, and output prediction. On the input side, YOLOv5s uses a data augmentation operation to process the input image; randomly selects four training images for random cropping, zooming and other functions; and then generates a more diverse set of training samples. Before YOLOv5s training, the k-means clustering method is used to calculate the prior bounding box best suited for the current dataset. k-means makes YOLOv5 better adaptable to object detection tasks of different scales and aspect ratios.

The backbone network of YOLOv5 includes focus processing [29], rearranging the channels of the input feature map into four parts to obtain a new feature map. Focus processing aims to extract feature information with different scales and receptive fields to enhance object detection performance. Following this, Convolution, Batch Normalization, and Leaky ReLU (CBL) [30] processing is conducted. First, use convolution to extract features from the feature map in the previous step, and use Batch Normalization (BN) to normalize the input data of each batch to improve the model's adaptability to changes in the input distribution. Finally, the Leaky ReLU activation function processes the resultant data.

YOLOv5s adopts Cross Stage Partial (CSP) [31], which has fewer parameters and calculations while maintaining high performance. The Spatial Pyramid Pooling Feature

(SPPF) [32] in YOLOv5s pools the input feature of different sizes to generate fixed-length feature vectors. This approach can handle objects of different sizes and extract contextual information on distinct regions of the input image, thereby improving the function and robustness of the model.

The neck network structure mainly uses the Feature Pyramid Network (FPN) [33] and the Pyramid Attention Network (PAN) [34]. FPN adopts the top-down paths and lateral connections and fuses the underlying high-resolution features with the top-level semantic information. By aggregating features from different levels of the FPN, the PAN enables features at various levels to influence and interact with each other. FPN and PAN structures strengthen the understanding of context information and improve the network feature fusion ability.

The prediction model is equipped with three feature maps of different scales and filters the prediction frame using NMS [35]. It also employs Generalized Intersection Over Union (GIOU) [36] as the loss function, considering the location and size variations of the target box. This approach more accurately measures the overlap degree of the target box. GIOU helps the model improve the object localization and boundary regression capabilities of object boxes. The original model of YOLOv5s is shown in Figure 2. Since this paper does not make enhancements to the SPPF pooling layer, the SPPF module is not prominently depicted in the figure.

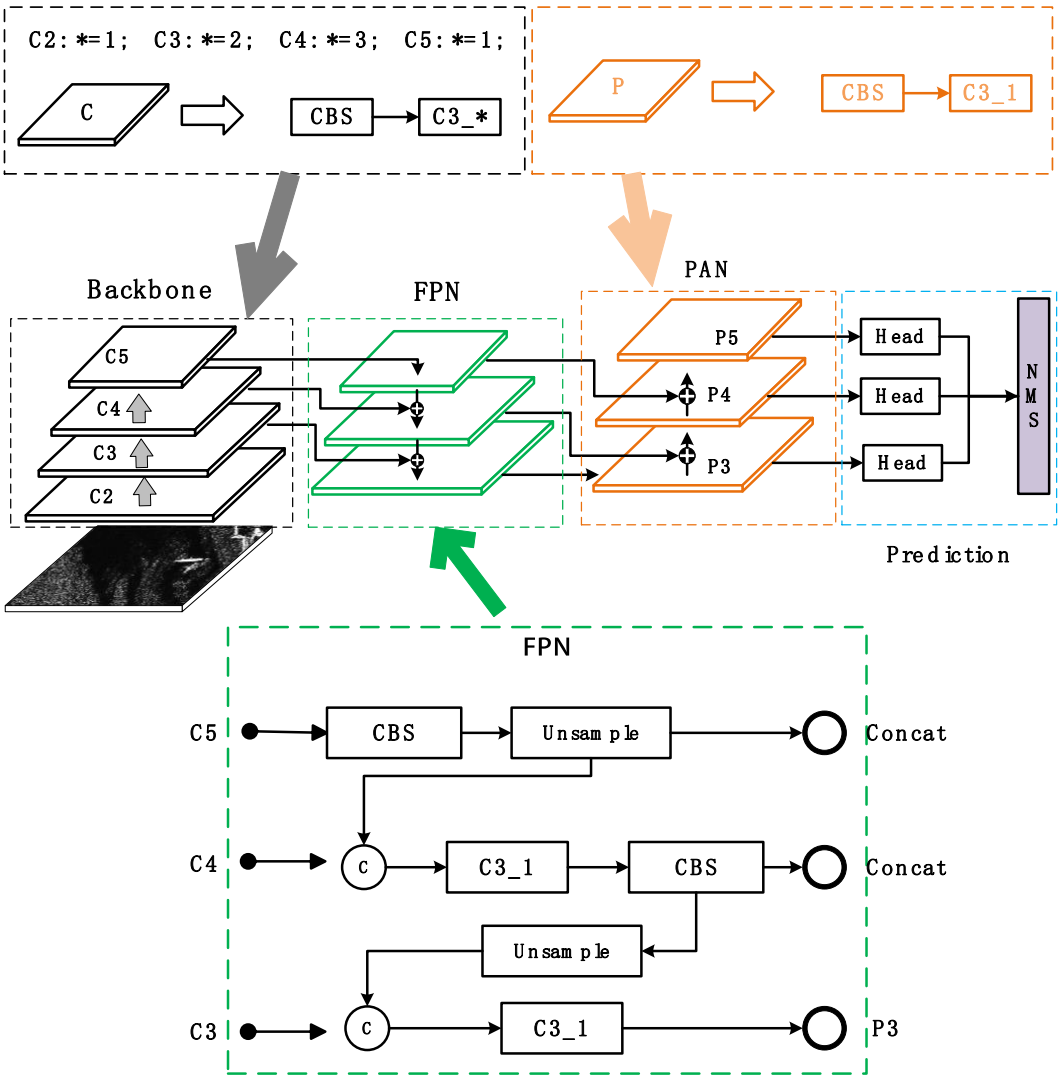


Figure 2. YOLOv5s structure.

3.2. CCDS-YOLO

This paper proposes a detection model, CCDS-YOLO, for multi-category targets in complex sets. The model not only achieves efficient detection results but also maintains a lightweight structure. Compared to the single-object SAR image object detection studies mentioned in the referenced literature, CCDS-YOLO is designed to detect four target categories. This approach presents a more challenging problem and carries higher significance for practical applications. Firstly, this paper chooses the appropriate anchor using k-means++ [37] clustering calculation on the dataset. SAR images are preprocessed and fed into the backbone network of the model. Due to the embedded CBAM, the backbone network can extract more useful messages, and this feature information will enter the CRD-FPN module for feature fusion, yielding better spatial and semantic information. The CRD-FPN module and the decoupled head can provide effective information for the model. Finally, the results are optimized by using Soft-NMS. Figure 3 illustrates the CCDS-YOLO model structure.

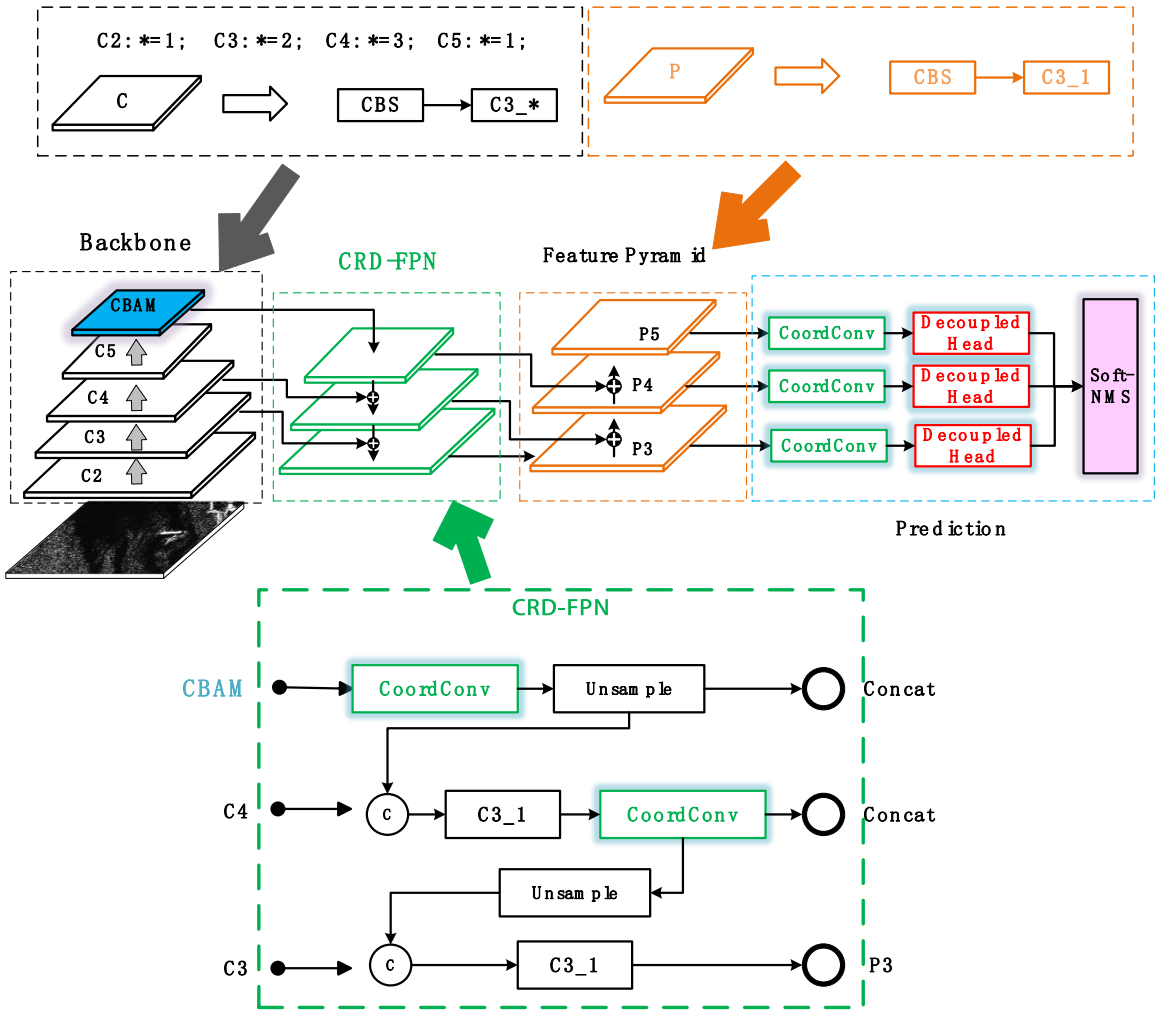


Figure 3. CCDS-YOLO structure.

3.2.1. CBAM Block

In the down-sampling stage, this paper integrates the CBAM, which significantly enhances the connection between the channel and the space of the model and increases the feature extraction ability. The CBAM block [38] is a simple and effective attention block for feed-forward convolutional neural networks. The CBAM can enhance the network's perceptual and expressive capabilities, thereby improving its performance and generalization ability. Additionally, the CBAM can seamlessly integrate into existing CNN architectures,

introducing minimal additional computational burden and thus achieving enhanced network performance without sacrificing computational efficiency. This module performs inference on the intermediate feature map to generate two distinct attention maps—one for channel and another for space. The input feature map is enhanced by multiplying the input feature map with the generated attention map.

The CBAM is different from Squeeze-and-Excitation Network (SE-Net) or Efficient Channel Attention Network (ECA-Net). It integrates the spatial attention module following the channel attention module and realizes the dual mechanism of channel and spatial attention. The choice of SE-Net or ECA-Net is typically contingent on whether the channel attention's connection utilizes Multi-Layer Perceptron (MLP) or one-dimensional convolution. Figure 4 describes the structure of the CBAM.

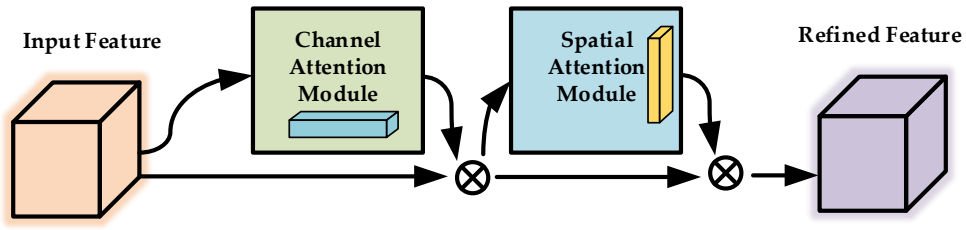


Figure 4. CBAM structure.

The Channel Attention Module (CAM) structure is shown in Figure 5. It performs global maximum pooling and average pooling on the input feature layer, adds the results obtained by a MLP, uses the sigmoid activation function to process, and then obtains the weight (ranging from 0 to 1) for each channel within the input feature layer. Finally, the consequence is the weighted channel-by-channel multiplication of the input feature layer.

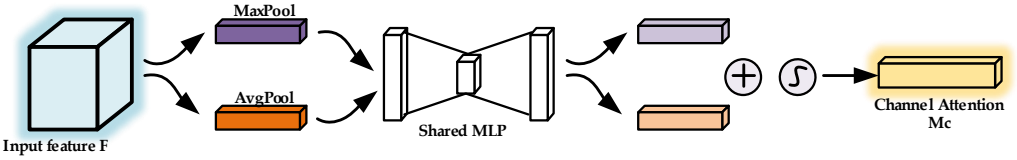


Figure 5. CAM structure.

To minimize parameter overhead, the CBAM sets the dimension of the closet activation to $r \times C/r \times 1 \times 1$, where r represents the reduction ratio. After each descriptor, we apply the shared network and output a feature vector by element-wise summation. In simple terms, channel attention is calculated using the following formula:

$$\begin{aligned} \mathbf{M_c}(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma\left(\mathbf{W}_1\left(\mathbf{W}_0\left(\mathbf{F}^{\mathbf{c}}_{\mathrm{avg}}\right)\right) + \mathbf{W}_1(\mathbf{W}_0(\mathbf{F}^{\mathbf{c}}_{max}))\right) \end{aligned} \quad (1)$$

where σ represents the sigmoid function, W_0 is a matrix with dimension $C/r \times C$, and W_1 is a matrix with size $C \times C/r$. It should be noted that the weights W_0 and W_1 of MLP are shared, and the ReLU function acts after W_0. Figure 6 illustrates the Spatial Attention Module (SAM) structure.

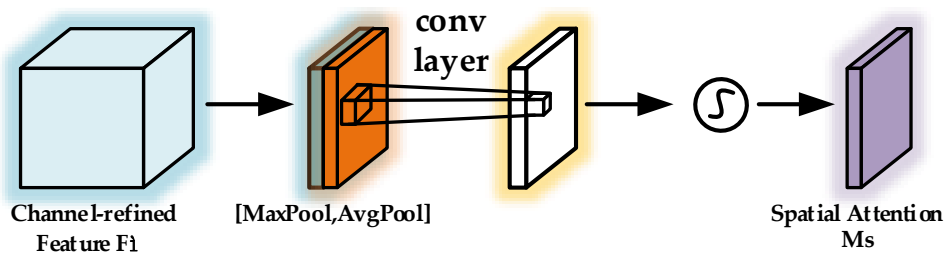


Figure 6. SAM structure.

As depicted, the channel submodule encompasses the outputs of max-pooling and average-pooling from the shared network. The spatial submodule leverages these two comparable outcomes, effectively diminishing redundancy by pooling them into the convolutional layers. The spatial attention mechanism formula is as follows:

$$\begin{aligned} \mathbf{M_s}(F) &= \sigma\left(f^{7\times7}([\text{AvgPool}(\mathbf{F}); \text{MaxPool}(\mathbf{F})])\right) \\ &= \sigma\left(f^{7\times7}\left(\left[\mathbf{F^s_{avg}}; \mathbf{F^s}_{max}\right]\right)\right) \end{aligned} \tag{2}$$

where σ represents the sigmoid function, and $f^{7\times7}$ represents the convolution operation with a filter size of 7×7.

Unlike the channel attention module, the spatial attention module focuses on the pivotal information parts of the input image, which is a supplement to the channel attention module.

3.2.2. CRD-FPN

To better solve the classification task in the detection task, this paper combines CoordConv [39] into the model network to construct the CRD-FPN module. CoordConv is a convolutional neural network technique for processing image data, which has some advantages in processing SAR image classification recognition tasks. Compared with optical images, given the low resolution of SAR images and their susceptibility to noise, traditional convolutional networks can encounter challenges. These challenges include geometric distortion, intense speckle noise, alterations in scattering centers, etc.

CoordConv not only maintains a reduced parameter count and exceptional computational efficiency but also provides the flexibility for the network model to either retain or discard the transformation invariance, which is meaningful for our research. Coordconv can be run as an extension of simple convolutional conv and improves the performance of convolutional neural networks by adding position coordinate information to feature maps. CoordConv adds the coordinate information of the feature map to the input as an extra channel, so the network can use this position information to understand better and process the image. This paper proposes to replace the 1×1 convolution in the feature pyramid network and the first layer convolution of the detection head part with the expanded convolution of CoordConv to construct the CRD-FPN module. The CRD-FPN module introduces positional information within the convolutional layers, enabling the network to more accurately comprehend changes in object position and size. CRD-FPN assists the network in effectively localizing and recognizing objects of different scales, which proves particularly beneficial for addressing scenarios with various scale objects, such as small and large targets. The structure of CoordConv is shown in Figure 7. It describes the operation of two coordinates i and j. Fundamentally speaking, the i-coordinate channel is an $h \times w$ matrix based on a rank of -1. The value of its first row is 0, the second row has a value of 1, and the third row has a value of 2. During the calculation, the values of i and j are linearly scaled so that their results fall in $[-1, 1]$. For two-dimensional convolution, the two (i, j) coordinates can specify an input pixel. Under further conditions, an extra channel can be integrated into the model to achieve specific results. The r coordinate of CoordConv uses the third channel, and the specific calculation formula is as follows:

$$r = \sqrt{(i - h/2)^2 + (j - w/2)^2} \tag{3}$$

In the SAR image classification and recognition task, CoordConv has the following advantages:

Dealing with geometric distortion. Due to factors like platform movement or terrain variations, SAR images might exhibit object shape distortions. By introducing location information, CoordConv enables the network to perceive better and capture the geometric shape of objects, thereby improving the accuracy of classification and recognition.

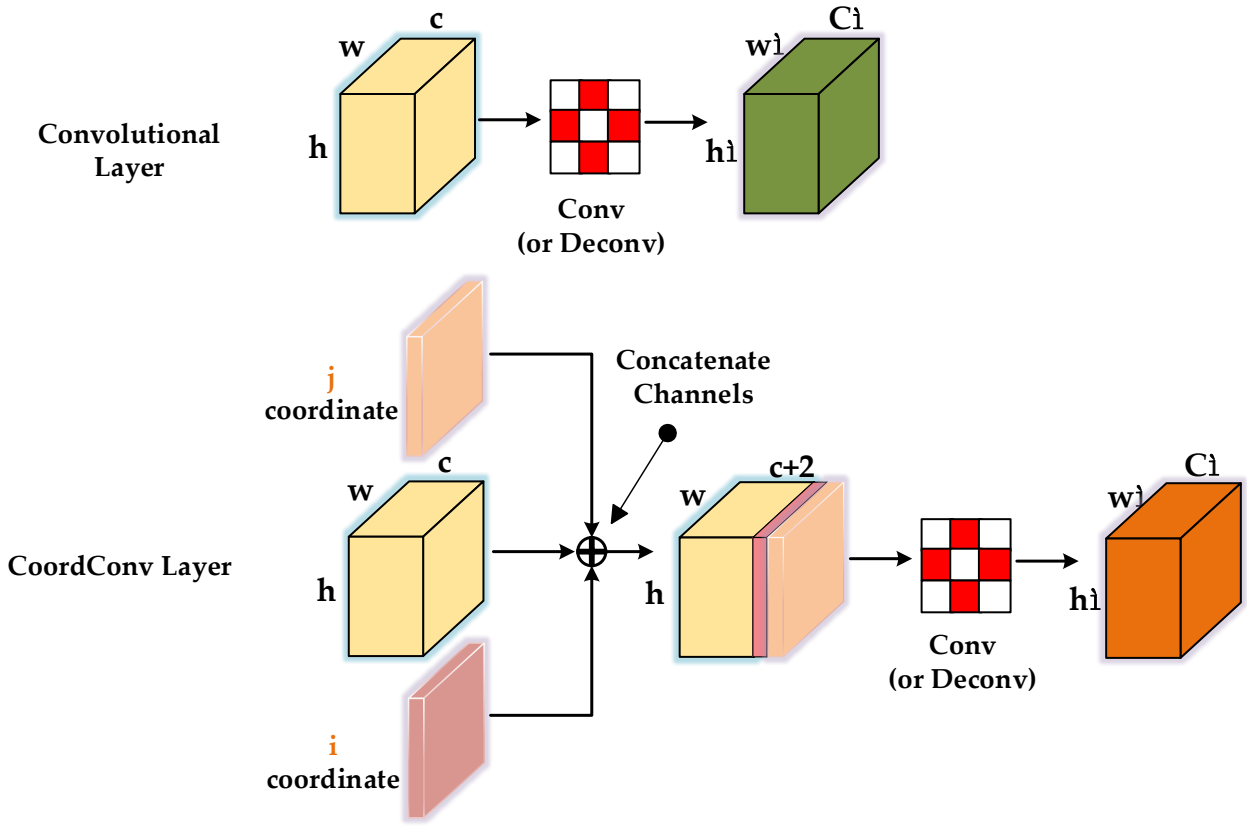


Figure 7. CoordConv structure.

Counter speckle noise. Speckle noise often exists in SAR images due to the superposition of multiple radar echoes. The location information of CoordConv can provide more contextual information, which enables the network to suppress noise better during the learning process, thus improving the quality of classification results.

3.2.3. Decoupled Head

The research in this paper includes both regression tasks and classification tasks. At present, the feature information for the network's regression and classification is derived from a singular head, leading to identical feature details. After simple calculation, it is applied to handling objectness, classification, and regression tasks, but it turns out that information suitable for classification tasks is not necessarily ideal for regression tasks. The concept of a decoupled detection head involves separating the prediction of object categories and object positions, carrying out these predictions on distinct channels. This approach enhances the model's prediction accuracy in multi-category scenarios and avoids interference between different categories.

The features of classification tasks and regression tasks satisfy different output forms. This paper proposes using the decoupled head [40] for feature extraction to obtain the optimal solution for classification and regression. For each layer of FPN features, an initial 1×1 conv layer reduces the feature channels to 256. Subsequently, two parallel branches are introduced, and each branch has two 3×3 conv layers for classification and regression tasks. Figure 8 illustrates the structure of a decoupled head.

Distinct network structures are designed for the classification task branch and the regression task branch for calculation so that the classification branch focuses on more local information and performs fine-grained recognition. The regression branch focuses on global

communication and strengthens spatial connections, thus enhancing the performance of both tasks.

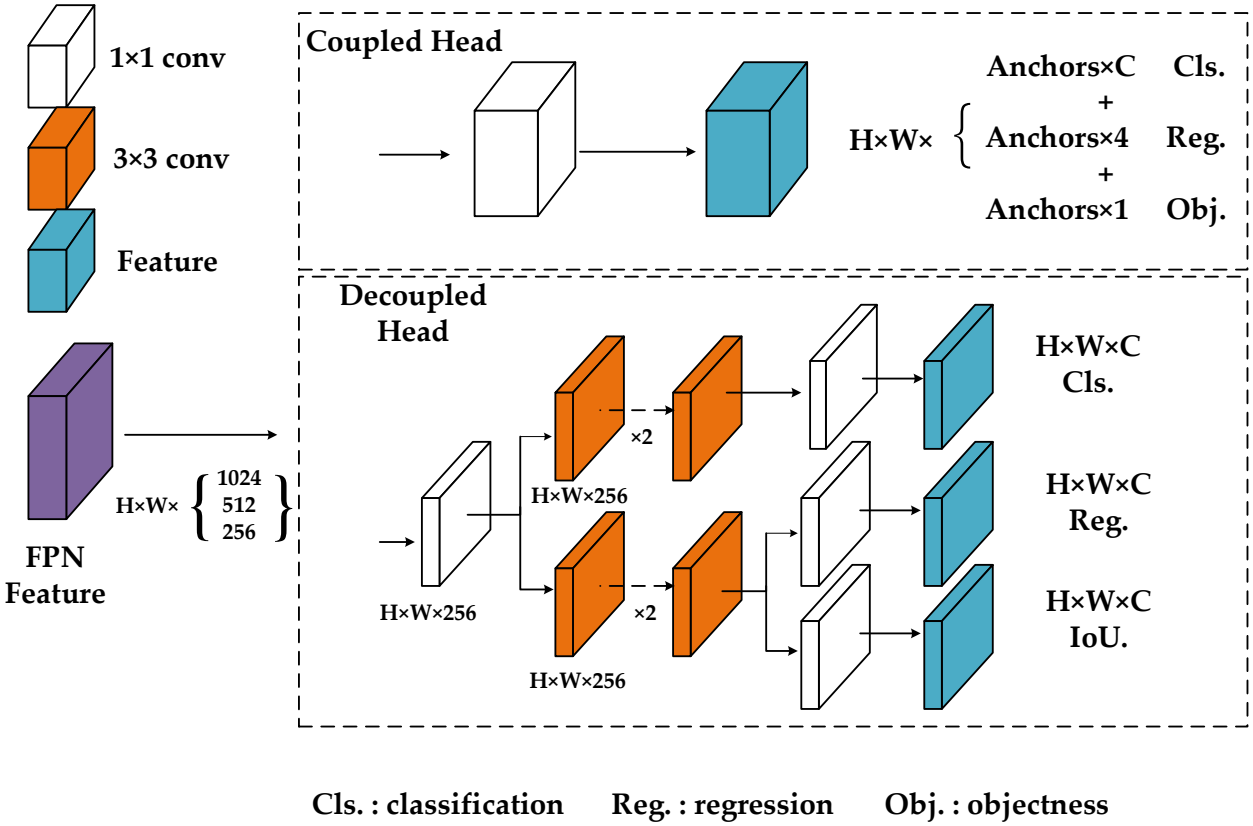


Figure 8. Decoupled head structure.

3.2.4. Soft-NMS

Soft-NMS [41] is an algorithm for NMS in object detection, which is designed to improve the traditional hard NMS approach to handle large overlapping target boxes better.

In object detection, the model usually uses a bounding box to represent the detected object location. When an object detector detects multiple overlapping bounding boxes in an image, it usually needs to use the NMS algorithm to select the most accurate object box to avoid repeatedly catching the same object. Traditional hard NMS methods select the object box with the high confidence score according to a predefined threshold and suppress other bounding boxes that highly overlap with this bounding box. However, this approach may cause some problems, especially when dealing with object boxes with large overlaps, leading to large errors. The working processes of NMS and Soft-NMS are shown in Figure 9. This paper replaces the traditional NMS algorithm with Soft-NMS. Soft-NMS retains the target box that overlaps the most with the maximum scoring box and then calculates and judges whether to keep the overlapping box using Formula (5).

The pruning algorithm in the NMS algorithm step is shown in Formula (4):

$$S_i = \begin{cases} S_i, \mathrm{iou}(M,b_i) < N_t \\ 0, \mathrm{iou}(M,b_i) \geq N_t \end{cases} \quad (4)$$

M represents the detection frame with the highest confidence level at present, N_t is the manually set threshold, and b_i refers to the detection frame to be processed. S_i is the confidence score during processing.

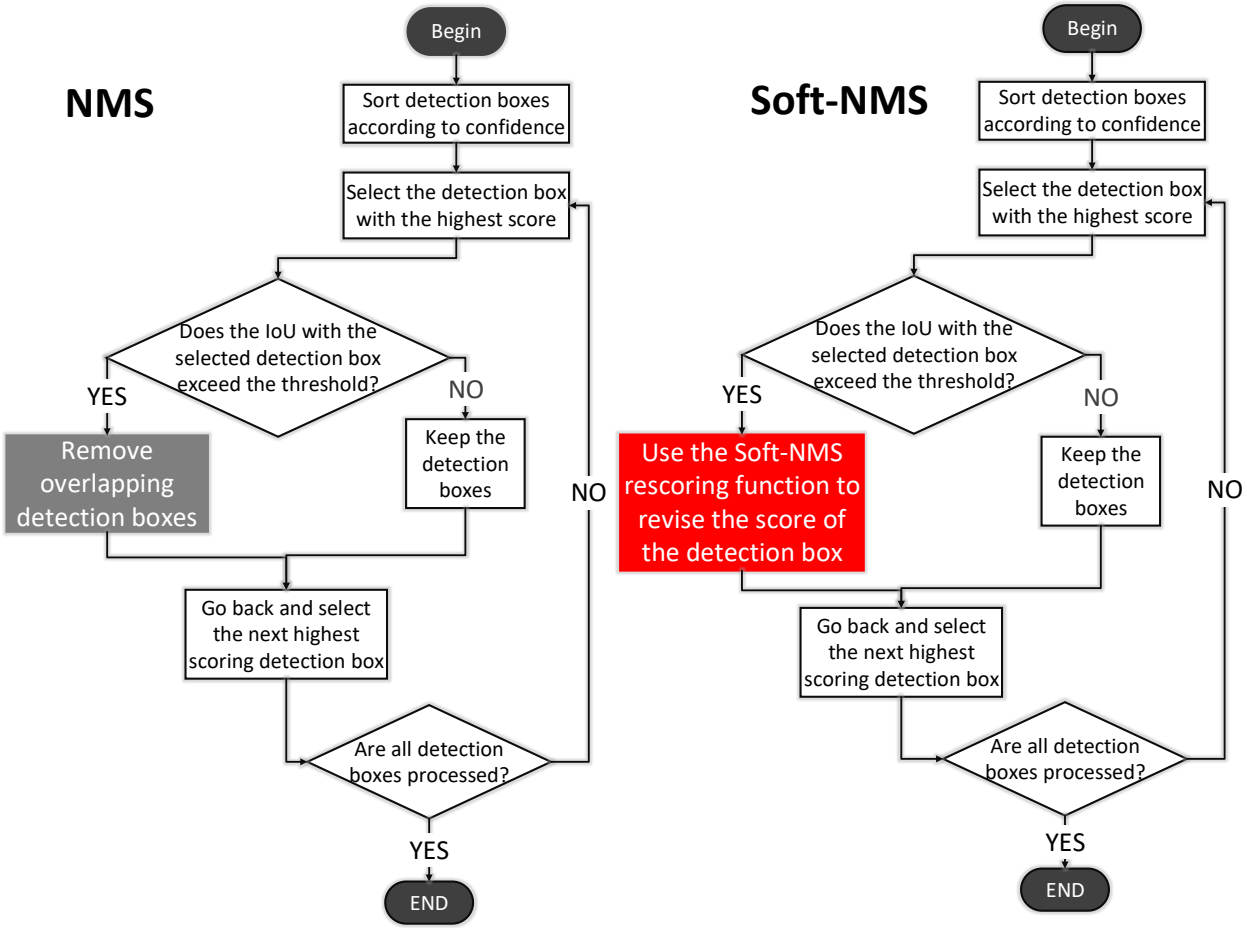


Figure 9. Working processes of NMS and Soft-NMS.

Soft-NMS improves the traditional hard NMS method by introducing a soft suppression mechanism. The basic idea is that, instead of directly removing the overlapping frame when calculating the overlapping area, the method decays its confidence. This approach preserves some overlapping object boxes, adjusting their importance based on their confidence levels. The Soft-NMS rescoring function is shown in Formula (5):

$$S_i = \begin{cases} S_i, iou(M,b_i) < N_t \\ S_i\,(1 - iou(M,b_i)), iou(M,b_i) \geq N_t \end{cases} \tag{5}$$

The above function uses the score of the detection box above the threshold value N_t to attenuate the linear function of M overlap. Considering that the penalty function should be continuous, update the pruning step using the Gaussian function as the following formula:

$$S_i = S_i e^{-\frac{iou(M,b_i)^2}{\sigma}}, \forall b_i \notin D \tag{6}$$

With Soft-NMS, the target frame with higher confidence can still retain a certain weight when calculating the overlap. This reduces the competition between overlapping target frames, thereby improving object detection accuracy. Soft-NMS has shown good performance and robustness in some object detection tasks.

Complex-shaped objects might exhibit significant overlap along their boundaries, rendering traditional NMS less effective. Soft-NMS adjusts confidence scores by considering distances, making it more suitable for handling these complex shapes.

4. Experiments

4.1. Experiment Settings

To ensure the rigor and credibility of the research, all the experiments are carried out under the same experimental environment. Table 1 shows the detailed configuration of the environment in this paper.

Table 1. Experimental environment.

Environment	Argument
Operating system	Ubuntu 18.04
CPU	AMD EPYC 7601
RAM	50 G
GPU	RTX 3070-8 G
PyTorch	1.11.0
Cuda	11.3

This study trains and tests all models on the same dataset, ensuring that each model is trained to convergence and obtains the most valuable evaluation metrics. All models utilize the same training dataset and undergo standard data augmentation. We set the learning rate to 0.001 and utilized the Stochastic Gradient Descent (SGD) optimizer along with the cosine annealing scheduler. Table 2 shows the parameter configuration used in this paper.

Table 2. Parameter configuration.

Parameter Configuration	Value
epoch	1000
lr	0.001
image	256 × 256
batch size	128
optimizer	SGD

4.2. Experiment Dataset

In this paper, the performance of the models is verified by the filtered MSAR dataset. The polarization modes of the MSAR [42] dataset include horizontal sending and horizontal receiving (HH), horizontal sending and vertical receiving (HV), vertical sending and horizontal receiving (VH), and vertical sending and vertical receiving (VV). The dataset scenarios include airports, ports, near the shore, islets, open seas, urban areas, etc., and the objects include four types of targets: planes, oilcans, bridges, and ships. First of all, the image resolution in the original MSAR dataset is inconsistent, which will affect the experiment. Given that the images in this dataset mainly possess a resolution of 256 pixels, we filter out images with varying resolutions. This approach standardizes the image format across the dataset. Secondly, the dataset comprises 39,858 samples of ship targets and only 1851 samples of bridge targets, leading to a significant imbalance in the sample quantities across different categories of targets, which could potentially affect the detection outcomes. This study aims to minimize the disparity in sample quantities among different categories of objects by reducing the number of samples for certain targets. This paper conducts a filtering procedure on the dataset, mitigating the impact of disparate sample quantities and non-uniform image resolutions on the model evaluation. The filtered dataset is better suited for research in the field of object detection. While increasing the number of images in the training set may enhance the model's performance, quantity alone is not the sole critical factor. Equally significant are the dataset's quality and diversity. Furthermore, we maintain consistent experimental conditions for all tests, including the use of the same dataset. Finally, the dataset includes a total of 4818 pictures of 256 × 256, in which the gap in the number of planes, bridges, oilcans, and ships is reduced, and the distribution of the number of different types of targets tends to be balanced. Table 3 is the detailed information before and after the dataset modification.

The dataset images in this research paper are divided into three subsets: the training set, validation set, and test set, with a distribution ratio of 6:2:2. The model uses mosaic data augmentation for data preprocessing during training.

Table 3. Comparison of MSAR original dataset and filtered dataset parameters.

Datasets	MSAR	MSAR Change
Total number	28,449	4818
Plane number	6368	5699
Bridge number	1851	1190
Oilcan number	12,319	8108
Ship number	39,858	3773

This paper uses Precision (P), Recall (R), Average Precision (AP), mean Average Precision (mAP), parameters, GFLOPS (Giga Floating-point Operations Per Second), and training time as comprehensive evaluation indicators to represent the effect of SAR image object detection. These indicators are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

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

$$\text{mAP} = \frac{1}{\text{N}} \sum_{\text{i}=1}^{\text{N}} \text{AP(i)} \tag{10}$$

4.3. Experiment Results

4.3.1. Ablation Experiments

This paper selected YOLOv5 as the benchmark and used the filtered MSAR dataset. The ablation experiment was carried out under the same environment and parameter settings to assess the impact of various improvements on the CCDS-YOLO model. The results of the ablation experiment are shown in Table 4.

Table 4. Ablation experiments.

NO.	Model	P	R	mAP@.5	mAP@0.5:0.9
1	YOLOV5s	87.1	86.0	89.0	61.3
2	YOLOV5s + CBAM	89.0	87.2	89.5	61.9
3	YOLOV5s + Decoupled head	90.2	87.6	90.4	65.8
4	YOLOV5s + CRD-FPN	90.4	87.4	90.6	62.3
5	YOLOV5s + Soft-NMS	87.1	86.0	90.8	61.3
6	YOLOV5s + CRD-FPN + Soft-NMS	91.0	86.8	91.7	66.1
7	YOLOV5s + CRD-FPN + Decoupled head + Soft-NMS	89.8	87.2	91.9	66.6
8	YOLOV5s + CBAM+ CRD-FPN + Decoupled head + Soft-NMS	90.5	88.0	92.3	68.0

Experiment 1 is the experimental result of the basic model of YOLOV5s, which is used as a comparison benchmark for the improved model later. YOLOv5s's detection mAP@0.5 value is 89.0%, the mAP@0.5:0.95 value is 61.3%, the P is 87.1%, and the R is 86.0%.

This paper replaces the detection head part's NMS with the Soft-NMS. Although other indicators have not improved, mAP@0.5 increases by 1.8%. Soft-NMS avoids the phenomenon that most coincidence boxes are deleted directly and reduces the misdetection

of compact and dense targets. The mAP@0.5 value is greatly improved. From experiment 6, when using the Soft-NMS and CRD-FPN modules at the same time, the experimental effect is the most significant. Various indicators have been greatly improved because CoordConv can better extract spatial location information. Under the premise of ensuring normal convolution, CoordConv adds two additional channels to transmit the i and j coordinates, which reduces the influence of noise and greatly improves the model's performance.

In Experiment 7, based on the previous experiment, the original detection head is replaced with the decoupled head, which provides the most suitable feature information for the two different detection tasks of classification and regression. Finally, this paper adds the CBAM to the feature extraction part of the backbone. The experiment proves that adding the CBAM enhances the feature extraction ability and reduces the missing detection caused by unknown feature information extraction. Compared to YOLOv5s, the model detection mAP@0.5 increases by 3.3%, mAP@0.5:0.95 increases by 6.7%, the P increases by 3.4%, and the R increases by 2%. These results indicate that the CCDS-YOLO model is generally superior to YOLOv5 in the ability to identify SAR multi-category targets.

4.3.2. Comparison Experiments

To thoroughly evaluate the CCDS-YOLO's detection capabilities, this section conducts a comprehensive performance test of the model. This paper employs nine object detection models that are currently popular, namely Faster R-CNN, RetinaNet, YOLOv5s, YOLOv5n, YOLOv5m, YOLOv5l, YOLOX-s, YOLOv7-tiny, and YOLOv7-x. It conducts comparative experiments with the proposed CCDS-YOLO model on the filtered MSAR dataset and ensures that the dataset division and experimental environment of each model are the same. Table 5 shows the experimental results.

Table 5. Experimental results of different models.

Model	Backbone	Precision	Recall	mAP@0.5	Parameters	GFLOPS	Time
RetinaNet	ResNet	50.4	83.3	67.6	37 M	198	4.2 h
Faster R-CNN	ResNet50	45.6	78.9	76.1	41.2 M	201	9.3 h
YOLOv7-tiny	MX + CBL + SPPCSP	80.9	78.7	83.1	6.1 M	13	8.5 h
YOLOX-s	CSPDarknet	83.4	80.7	86.3	9 M	26.8	12 h
YOLOv5n	CBS + CSP + FPN	87.7	81.8	86.4	1.8 M	4.1	3 h
YOLOv5s	CBS + CSP + FPN	87.1	86.0	89.0	7.0 M	15.8	4.3 h
YOLOv5m	CBS + CSP + FPN	88.3	85.4	91.0	21 M	47.9	5.8 h
YOLOv5l	CBS + CSP + FPN	90.8	86.0	91.4	46 M	107	6.8 h
YOLOv7-x	CBS + ELAN + MP + SPPCSPC	89.4	88.2	91.6	71 M	188	31 h
CCDS-YOLO	CBS + CSP+ CRD-FPN	90.5	88.0	92.3	16 M	34	4 h

Considering that this paper needs to deal with SAR images and limited satellite carrying resources, and that it cannot occupy too much resource space while ensuring real-time detection capabilities, this paper selects some models for model comparison experiments. Among them, more single-stage algorithms are chosen for comparison. The experimental results are shown in Figure 10. The experimental results prove that CCDS-YOLO achieves the best detection effect, with a mAP value of 92.3%. Compared to numerous mainstream object detection models, CCDS-YOLO achieves the highest mAP@0.5, with fewer parameters, making it more suitable for applications in the multi-class SAR target detection domain. This is because the improved model has a more effective and accurate feature information extraction ability, enhancing the detection ability of compact, dense, and multi-scale objects. Compared to the baseline model, YOLOv5s, the mAP@0.5 value of CCDS-YOLO increases from 89.0% to 92.3%, and the detection effect of CCDS-YOLO is also better than other comparison models.

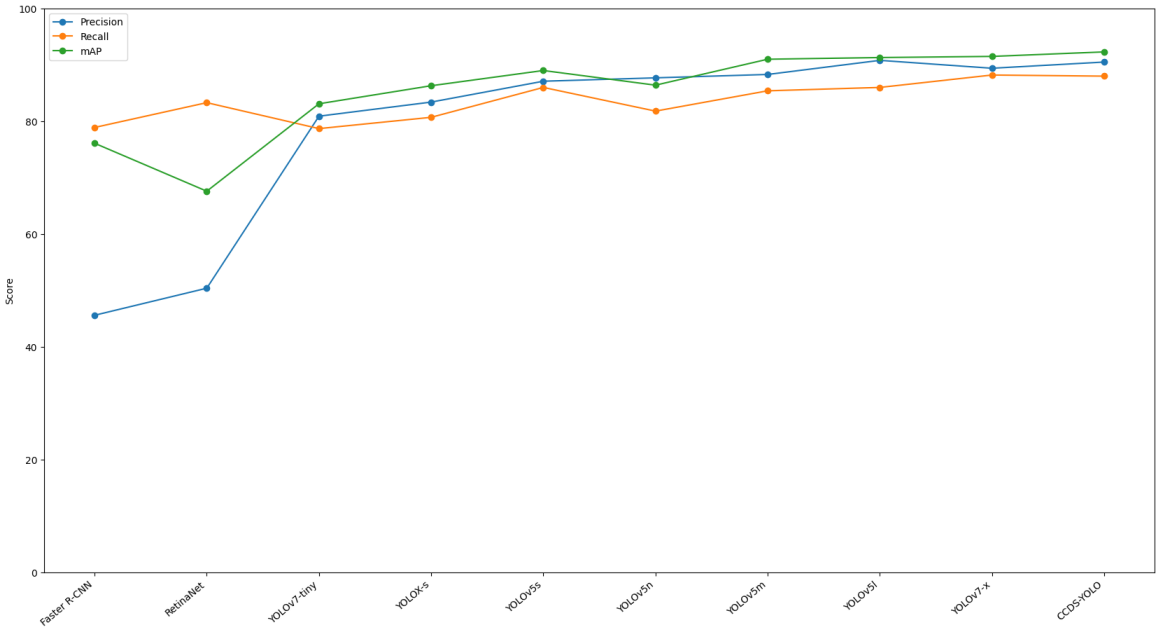


Figure 10. Results of different models.

4.3.3. Anti-Interference Experiments

In order to prove the stability of the CCDS-YOLO model, this paper conducts anti-interference experiments. This paper adds Gaussian noise to the dataset, where sigma is set to 8, 16, 22, 28, and 34.

Because of its characteristics, SAR images will produce noise effects under the influence of the environment and climate, and the additional Gaussian noise will bring great difficulties to the SAR image detection, which will lead to missed detection and false detection. This paper uses CCDS-YOLO, YOLOv5s, and the single-stage representative algorithm Faster R-CNN to make effective comparison experiments. The experimental results are shown in Figure 11. It can be seen that in the case of intense noise interference, Group (a) represents the ground truth of data without added noise, depicting the actual situation. Groups (b), (c), and (d), respectively, represent the detection results of CCDS-YOLO, YOLOv5s, and Faster R-CNN under the condition of added noise. The yellow circle in the figure represents missed detection, and the red circle represents false detection. In group (b), it can be observed that the overall detection performance of CCDS-YOLO is nearly consistent with the ground truth values. In groups (c) and (d), YOLOv5s and Faster R-CNN both exhibit numerous instances of missed detection and false detection. For example, in the first row of the result images, both YOLOv5s and Faster R-CNN missed some ship targets. In the third row of Figure 11, the red circle denotes an erroneous detection by Faster R-CNN, in which coastal rocks have been incorrectly identified as a ship. The results of the anti-interference experiments indicate that, in comparison to other models, CCDS-YOLO exhibits stronger resistance to interference.

This paper used six comparison models to verify the noise data. The results of the anti-interference experiment are shown in Figure 12, the CCDS-YOLO model is relatively less affected by noise, and the downward trend is more gradual. The CBAM is embedded in the model, which can extract more feature information; the CRD-FPN module can enhance the integration of spatial features to improve the detection model's capability; and the decoupled head and Soft-NMS enable the model to output more effective feature maps and effectively avoid deleting the correct detection frame, improving the stability of the model. It turns out that CCDS-YOLO is more suitable for processing SAR image detection tasks.

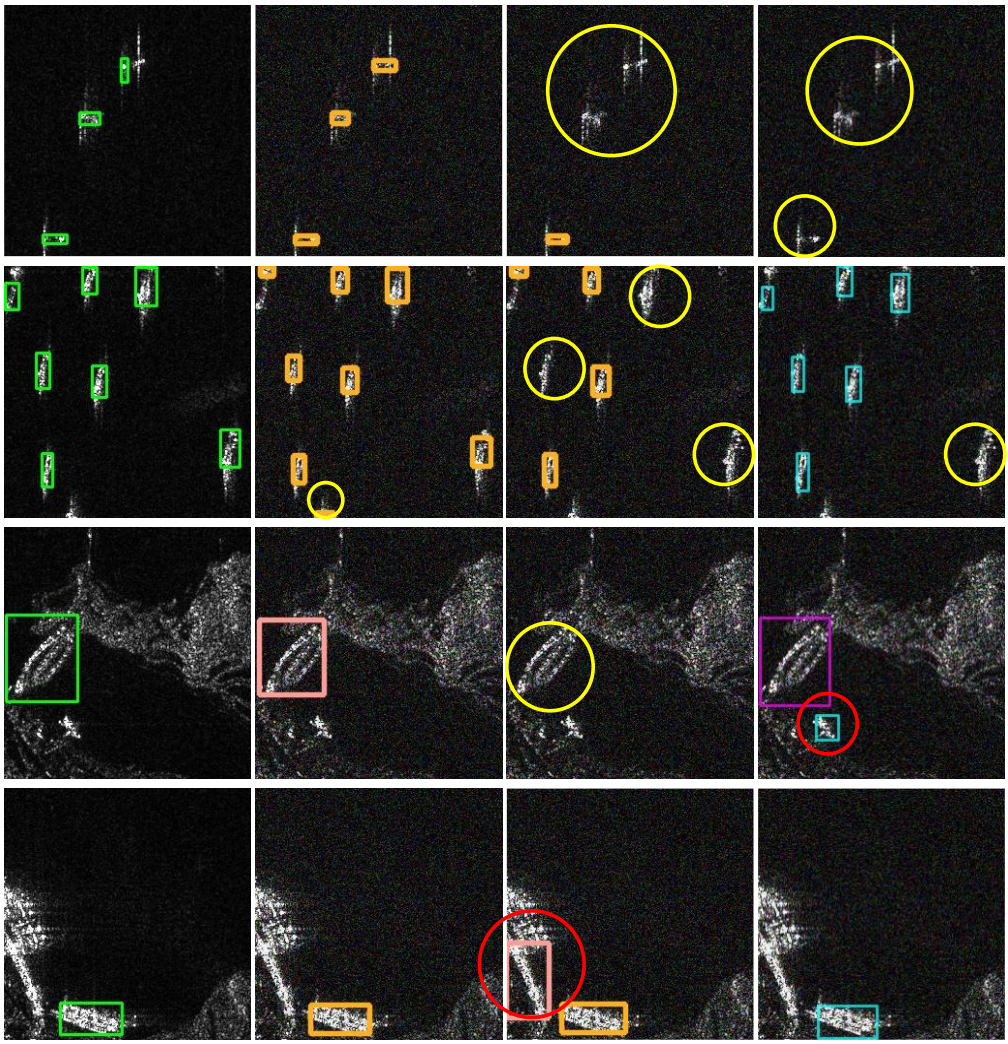


Figure 11. Comparison chart of model anti-noise effect. (**a**) Ground truth; (**b**) Sigma 34 for CCDS-YOLO; (**c**) Sigma 34 for YOLOv5s; (**d**) Sigma 34 for Faster R-CNN. The yellow circle in the figure represents missed detection, and the red circle represents false detection.

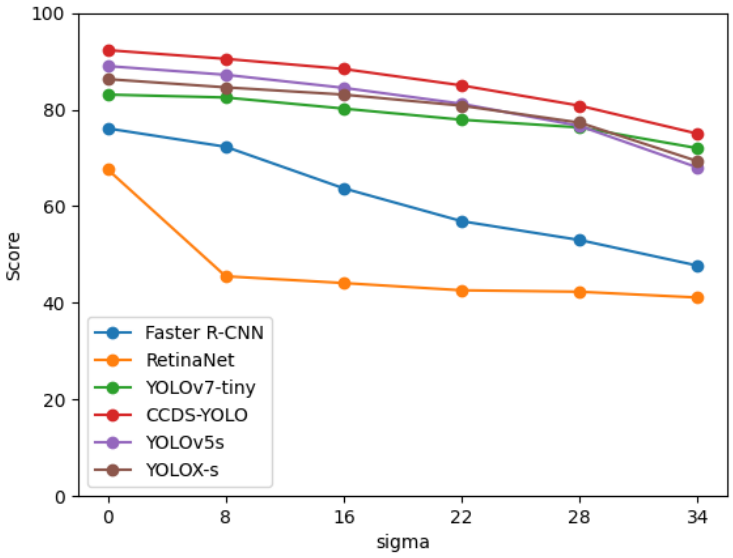


Figure 12. Model anti-interference effect comparison.

4.4. Visualization

According to the experimental results, this paper compares the results of the basic YOLOv5s and the improved model, CCDS-YOLO, in Figure 13. The figure indicates a 3.3% improvement in the overall mAP@0.5 of CCDS-YOLO, along with a significant 13.5% increase in the detection accuracy, specifically for planes. The plane detection scenes in the dataset are usually densely distributed, and other targets on land will interfere with the detection. This paper introduces the Soft-NMS algorithm to avoid the phenomenon that overlapping frames are easily missed effectively and make it easier for the model to detect dense, small objects.

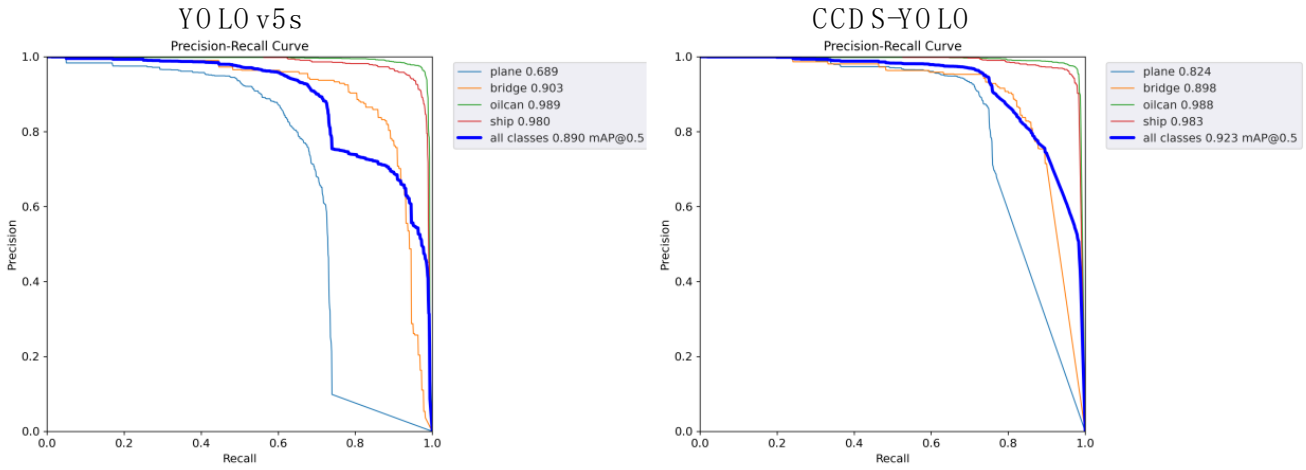
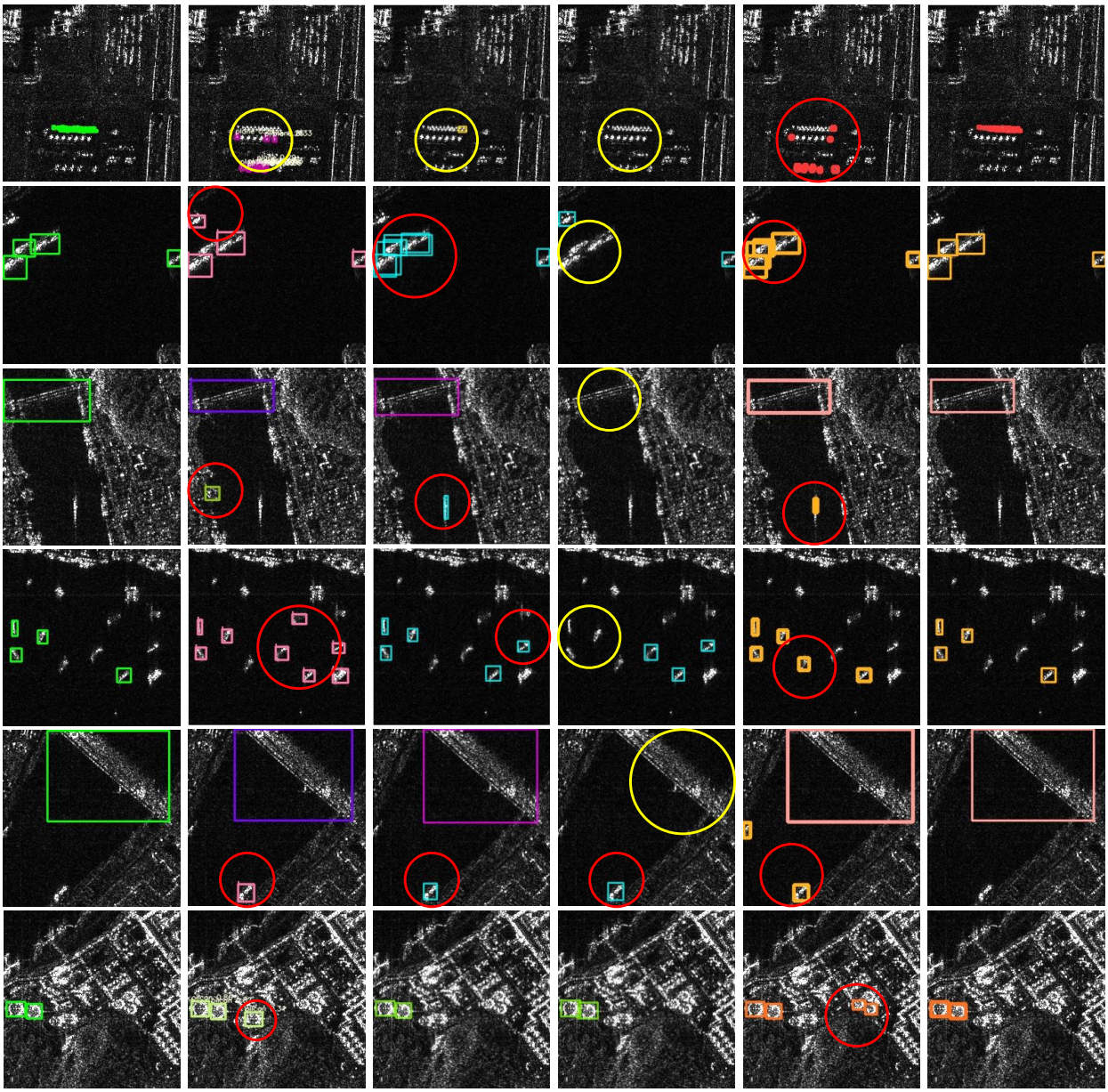


Figure 13. PR comparison between YOLOv5s and CCDS-YOLO.

In order to observe the advantages of the improved model more intuitively, this paper uses the real label frame picture as the standard and compares the test results of YOLOv5s, Faster R-CNN, RetinaNet, YOLOv7-tiny, and the enhanced CCDS-YOLO. Figure 14 presents the SAR image detection effect of different models. In Figure 14, the yellow circle represents missed detection, and the red circle represents false detection. Group (a) represents the ground truth of the data. From group (b) to group (e), this paper shows the detection effects of YOLOv7-tiny, Faster R-CNN, RetinaNet, and YOLOv5s. Group (f) shows the detection effect of the CCDS-YOLO model. Observing the images, it becomes evident that CCDS-YOLO accurately identifies various types of targets, showcasing excellent detection performance. For instance, in the comparison of result images in the first row, it is evident that the detection results of other models in groups (b) to (e) are not satisfactory. These models are unable to detect the actual plane targets and misidentify the image background as a plane. However, in group (f), CCDS-YOLO accurately detects the correct plane targets without any occurrences of false positives. In conclusion, compared with other mainstream models, CCDS-YOLO has better performance in multi-category target detection.

The comparison figure reveals that YOLOv5s, along with other models, exhibits a poor ability to collect feature information. This often leads to failure in identifying small targets, such as planes, ships, and oilcans, or results in misidentifying them as other types. Most notably, complex scenes such as coasts and mountains can affect the model's detection capability, causing it to detect background objects as targets that need to be identified. The CCDS-YOLO model effectively avoids the above problems. The CBAM improves the accuracy of the model to extract target feature information, enhances the anti-interference ability of the model, and reduces the false detection rate of the model. The CRD-FPN module and the improved detection network can enable the model to obtain more effective spatial position information, help the model to locate various small targets, and thus more accurately detect targets of multiple scales.

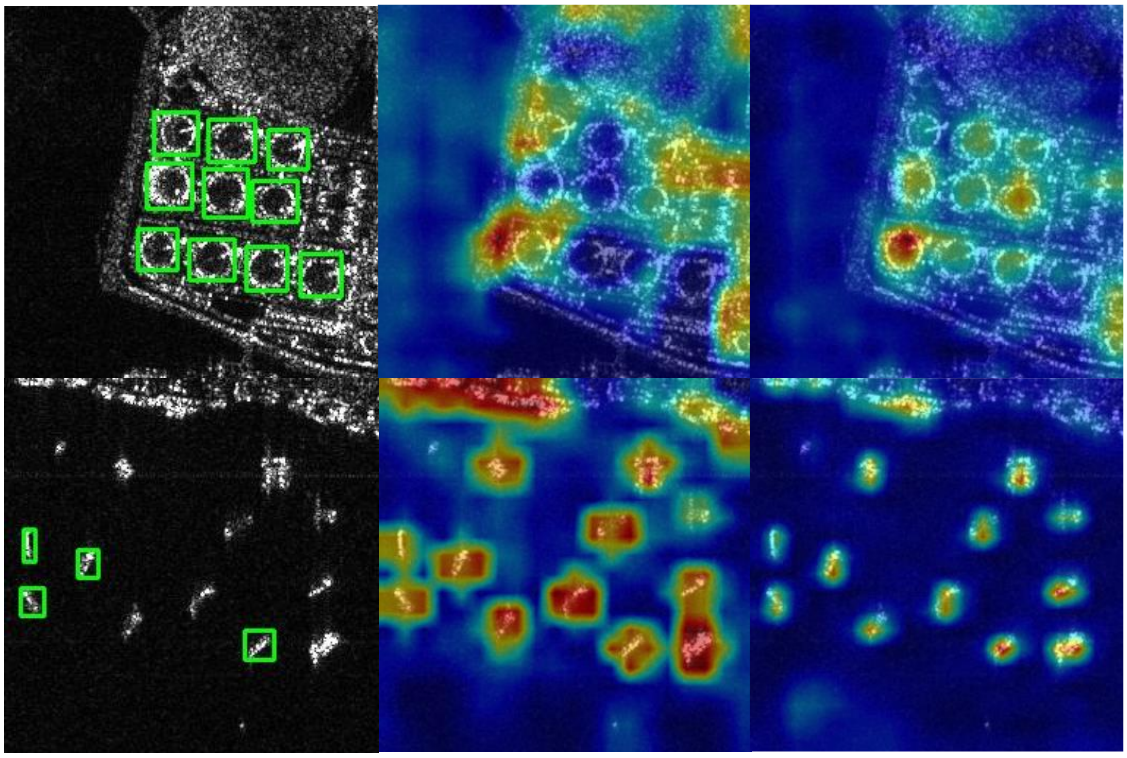


(**a**) (**b**) (**c**) (**d**) (**e**) (**f**)

Figure 14. Comparison of the effect of CCDS-YOLO and other models. (**a**) Ground truth; (**b**) YOLOv7-tiny; (**c**) Faster R-CNN; (**d**) RetinaNet; (**e**) YOLOv5s; (**f**) CCDS-YOLO. The yellow circle in the figure represents missed detection, and the red circle represents false detection.

This paper chooses the GradCam method to visualize the output of the model. The heat map, shown in Figure 15, helps to compare the visualization results of CCDS-YOLO with those of the basic model, YOLOv5s, providing a more intuitive view of the differences in feature processing between CCDS-YOLO and the basic model. In the result images, the blue area represents regions of low attention, the yellow area indicates regions of moderate attention, and the red area signifies the highest attention. Group (a) represents the ground truth of the data. Group (b) shows the feature extraction heat map of YOLOv5s. YOLOv5s focuses more attention on background objects while ignoring the target to be detected. In group (c), the CCDS-YOLO model can focus more attention on the target object and less on the background. For instance, in the result images of the first row, YOLOv5s allocates greater attention to the coastal background, thereby neglecting to emphasize the oilcans, which is the intended target. This indicates that YOLOv5s may incorrectly identify the

background, such as the coastline or rocks, as detection targets, leading to occurrences of false positives. The CCDS-YOLO model proposed in this paper can focus more on the oilcans while minimizing emphasis on the background. In summary, the CCDS-YOLO model effectively extracts features from the recognition target and mitigates the influence of complex backgrounds on detection.



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

Figure 15. GradCam visualization heat map. (**a**) Ground truth; (**b**) YOLOv5s heat map; (**c**) CCDS-YOLO heat map. The green squares in the figures show the real targets.

5. Conclusions

The multi-category SAR image dataset in complex scenes contains multiple categories of targets, creating challenges due to the large-scale span of object detection and the compact arrangement of targets in the pictures. This paper proposes a CCDS-YOLO model to solve the above problems. The model is embedded with CBAM, which enhances feature capture and improves multi-scale object detection. CCDS-YOLO introduces a CRD-FPN module based on CoordConv to obtain more spatial location information while maintaining a lightweight design. CCDS-YOLO also uses the decoupled head to replace the original detection head, which provides more effective feature information for both classification tasks and regression tasks, improving the overall performance of the model. After the model generates feature information, the Soft-NMS method is employed to address existing problems in traditional NMS. The CCDS-YOLO model has achieved a notable level of detection accuracy while also retaining its lightweight advantages. In comparison to the baseline model, YOLOv5s, the mAP@0.5 value of CCDS-YOLO on the filtered MSAR dataset increases by 3.3%, and the detection mAP@0.5:0.95 value increases by 6.7%, which ensures the lightweight advantages and improves the detection capability of the model.

The experiments prove that, in complex scenes, the CCDS-YOLO model proposed in this paper can effectively detect strategically important targets, such as planes, ships, oilcans, and bridges, in SAR images. The lightweight CCDS-YOLO model also makes it easier to deploy on satellites and has better applicability.

Author Contributions: Conceptualization, M.H. and Z.L.; methodology, J.W.; software, M.H. and T.L.; validation, M.H., Z.L. and J.W.; formal analysis, M.H.; investigation, Z.L.; resources, J.W.; data curation, M.H.; writing—original draft preparation, Z.L. and T.L.; writing—review and editing, Z.L.; visualization, M.H. and T.L.; supervision, Z.L.; project administration, M.H.; funding acquisition, M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Foundation of Hebei Technology Innovation Center of Intelligent IoT (grant number KFZD2201) and by the Defense Industrial Technology Development Program (grant number JCKYS2022DC10).

Data Availability Statement: This paper uses the MSAR dataset, and based on the research requirements, the original dataset has been filtered to make the distribution of the sample size tend to be balanced. Data sources: https://radars.ac.cn/web/data/getData?dataType=MSAR (accessed on 15 July 2023).

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

References

1. Qiao, K.; Sun, Z.; Yang, L.; Wei, W.; Wozniak, M.; Scherer, R. High-Resolution SAR Image Despeckling Based on Nonlocal Means Filter and Modified AA Model. *Appl.-Aware Multimed. Secur. Tech.* **2020**, *3*, 8889317. [CrossRef]
2. Cui, Z.; Wang, N.; 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]
3. Wang, X.; Cui, Z.; Cao, Z.; Tian, Y. Ship Detection in Large Scale Sar Images Based on Bias Classification. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020. [CrossRef]
4. Wang, X.; Chen, C.; Pan, Z.; Pan, Z. Fast and Automatic Ship Detection for SAR Imagery Based on Multiscale Contrast Measure. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1834–1838. [CrossRef]
5. Meng, Y.; Guo, C. Ship Detection in SAR Images Based on Lognormal ρ. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1372–1376. [CrossRef]
6. Lopez-Martinez, C.; Fabregas, X. Polarimetric SAR speckle noise model. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 2232–2242. [CrossRef]
7. Cui, Z.; Quan, H.; Cao, Z.; Xu, S.; Ding, C.; Wu, J. SAR Target CFAR Detection via GPU Parallel Operation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4884–4894. [CrossRef]
8. Araujo, L. Genetic programming for natural language processing. *Genet. Program. Evolvable Mach.* **2020**, *21*, 11–32. [CrossRef]
9. Yu, M.; Quan, S.; Kuang, G.; Ni, S. SAR Target Recognition via Joint Sparse and Dense Representation of Monogenic Signal. *Remote Sens.* **2019**, *11*, 2676. [CrossRef]
10. Dai, J.; Li, Y.; Sun, J. R-FCN: Object Detection via Region-based Fully Convolutional Networks. *arXiv* **2017**, arXiv:1605.06409.
11. Zhou, Q.; Li, X.; He, L.; Cheng, G.; Tong, Y. TransVOD: End-to-End Video Object Detection with Spatial-Temporal Transformers. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 7853–7869. [CrossRef]
12. Cui, Z.; Dang, S.; Cao, Z.; Wang, S.; Liu, N. SAR Target Recognition in Large Scene Images via Region-Based Convolutional Neural Networks. *Remote Sens.* **2018**, *10*, 776. [CrossRef]
13. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detection. *arXiv* **2019**, arXiv:1904.01355. [CrossRef]
14. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767. [CrossRef]
15. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2016**, arXiv:1506.02640. [CrossRef]
16. Qu, Z.; Zhu, F.; Qi, C. Remote Sensing Image Target Detection: Improvement of the YOLOv3 Model with Auxiliary Networks. *Remote Sens.* **2021**, *13*, 3908. [CrossRef]
17. Tan, Y.; Li, Q.; Li, Y.; Tian, J. Aircraft Detection in High-Resolution SAR Images Based on a Gradient Textural Saliency Map. *Sensors* **2015**, *15*, 23071–23094. [CrossRef]
18. Hong, Z.; Yang, T.; Tong, X.; Zhang, Y. Multi-Scale Ship Detection from SAR and Optical Imagery via a More Accurate YOLOv3. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6083–6101. [CrossRef]
19. Marti, M.; Maki, A. A multitask deep learning model for real-time deployment in embedded systems. In Proceedings of the Poster Presentation at Swedish Symposium on Deep Learning SSDL2017, Stockholm, Sweden, 20–21 June 2017.
20. Wang, Y.; Wang, C.; Zhang, H.; Dong, Y.; Wei, S. Automatic Ship Detection Based on RetinaNet Using Multi-Resolution Gaofen-3 Imagery. *Remote Sens.* **2019**, *11*, 531. [CrossRef]
21. Liao, L.; Du, L.; Guo, Y. Semi-Supervised SAR Target Detection Based on an Improved Faster R-CNN. *Remote Sens.* **2022**, *14*, 143. [CrossRef]
22. Li, Y.; Zhang, S.; Wang, W. A Lightweight Faster R-CNN for Ship Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4006105. [CrossRef]

23. Chen, L.; Shi, W.; Deng, D. Improved YOLOv3 Based on Attention Mechanism for Fast and Accurate Ship Detection in Optical Remote Sensing Images. *Remote Sens.* **2021**, *13*, 660. [CrossRef]
24. Lin, Q.; Wang, B.; Wang, Y. SAR image ship detection based on improved YOLOv4. In Proceedings of the 2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 28–30 June 2021. [CrossRef]
25. Liu, L.; Jia, Z.; Yang, J.; Kasabov, N.K. SAR Image Change Detection Based on Mathematical Morphology and the K-Means Clustering Algorithm. *IEEE Access* **2019**, *7*, 43970–43978. [CrossRef]
26. Luo, R.; Chen, L.; Xing, J.; Yuan, Z.; Tan, S.; Cai, X.; Wang, J. A Fast Aircraft Detection Method for SAR Images Based on Efficient Bidirectional Path Aggregated Attention Network. *Remote Sens.* **2021**, *13*, 2940. [CrossRef]
27. Luo, R.; Xing, J.; Chen, L.; Pan, Z.; Cai, X.; Li, Z.; Wang, J.; Ford, A. Glassboxing Deep Learning to Enhance Aircraft Detection from SAR Imagery. *Remote Sens.* **2021**, *13*, 3650. [CrossRef]
28. Ultralytics. yolov5. Available online: https://github.com/ultralytics/yolov5 (accessed on 18 May 2020).
29. Qiu, X.; Li, M.; Zhang, L.; Yuan, X. Guided filter-based multi-focus image fusion through focus region detection. *Signal Process. Image Commun.* **2019**, *72*, 35–46. [CrossRef]
30. Chen, H.; Jin, H.; Lv, S. YOLO-DSD: A YOLO-Based Detector Optimized for Better Balance between Accuracy, Deploy Ability and Inference Time in Optical Remote Sensing Object Detection. *Appl. Sci.* **2022**, *12*, 7622. [CrossRef]
31. Lin, H.; Yang, J. Ensemble cross-stage partial attention network for image classification. *IET Image Process.* **2022**, *16*, 102–112. [CrossRef]
32. Tang, H.; Liang, S.; Yao, D.; Qiao, Y. A visual defect detection for optics lens based on the YOLOv5-C3CA-SPPF network model. *Opt. Express* **2023**, *31*, 2628–2643. [CrossRef]
33. Zhang, T.; Zhang, X.; Ke, X. Quad-FPN: A Novel Quad Feature Pyramid Network for SAR Ship Detection. *Remote Sens.* **2021**, *13*, 2771. [CrossRef]
34. Wang, X.; He, N.; Hong, C.; Wang, Q.; Chen, M. Improved YOLOX-X based UAV aerial photography object detection algorithm. *Image Vis. Comput.* **2023**, *135*, 104697. [CrossRef]
35. Ma, W.; Zhou, T.; Qin, J.; Zhou, Q.; Cai, Z. Joint-attention feature fusion network and dual-adaptive NMS for object detection. *Knowl.-Based Syst.* **2022**, *241*, 108213. [CrossRef]
36. Cui, M.; Duan, Y.; Pan, C.; Wang, J.; Liu, H. Optimization for Anchor-Free Object Detection via Scale-Independent GIoU Loss. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 6002205. [CrossRef]
37. Atasever, U.H.; Gunen, M.K. Change Detection Approach for SAR Imagery Based on Arc-Tangential Difference Image and k-Means++. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 3509605. [CrossRef]
38. Woo, S.; Park, J.; Lee, J. CBAM: Convolutional Block Attention Module. *arXiv* **2018**, arXiv:1807.06521. [CrossRef]
39. Liu, R.; Lhman, J.; Molino, P.; Such, F.; Frank, E.; Sergeev, A.; Yosinski, J. An Intriguing Failing of Convolutional Neural Networks and the CoordConv Solution. *arXiv* **2018**, arXiv:1807.03247. [CrossRef]
40. Ge, Z.; Liu, F.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430. [CrossRef]
41. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L. Soft-NMS—Improving Object Detection with One Line of Code. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017. [CrossRef]
42. Xia, R.; Chen, J.; Huang, Z.; Wan, H.; Wu, B.; Sun, L.; Yao, B.; Xiang, H.; Xing, M. CRTransSar: A Visual Transformer Based on Contextual Joint Representation Learning for SAR Ship Detection. *Remote Sens.* **2022**, *14*, 1488. [CrossRef]