

Article

A Fast Facet-Based SAR Imaging Model and Target Detection Based on YOLOv5 with CBAM and Another Detection Head

Qingkuan Wang¹ , Jing Sheng², Chuangming Tong¹, Zhaolong Wang¹, Tao Song^{1,*}, Mengdi Wang³ and Tong Wang¹

¹ Air Defense and Antimissile School, Air Force Engineering University, Xi'an 710054, China; qingkuan_wang@163.com (Q.W.); cmtong156@163.com (C.T.); wangzhaolong2022@163.com (Z.W.); tong_wang001@163.com (T.W.)

² Science and Technology on Electromagnetic Scattering Laboratory, Beijing 100854, China; shengjing1122@126.com

³ Troops 94789, People's Liberation Army, Nanjing 210000, China; mendi_wang@163.com

* Correspondence: chinataoguo1990@163.com

Abstract: Synthetic Aperture Radar (SAR) image target detection is of great significance in civil surveillance and military reconnaissance. However, there are few publicly released SAR image datasets of typical non-cooperative targets. Aiming to solve this problem, a fast facet-based SAR imaging model is proposed to simulate the SAR images of non-cooperative aircraft targets under different conditions. Combining the iterative physical optics and the Kirchhoff approximation, the scattering coefficient of each facet on the target and rough surface can be obtained. Then, the radar echo signal of an aircraft target above a rough surface environment can be generated, and the SAR images can be simulated under different conditions. Finally, through the simulation experiments, a dataset of typical non-cooperative targets can be established. Combining the YOLOv5 network with the convolutional block attention module (CBAM) and another detection head, a SAR image target detection model based on the established dataset is realized. Compared with other YOLO series detectors, the simulation results show a significant improvement in precision. Moreover, the automatic target recognition system presented in this paper can provide a reference for the detection and recognition of non-cooperative aircraft targets and has great practical application in situational awareness of battlefield conditions.

Keywords: electromagnetic scattering calculation; SAR image; YOLOv5 network; convolutional block attention module



Citation: Wang, Q.; Sheng, J.; Tong, C.; Wang, Z.; Song, T.; Wang, M.; Wang, T. A Fast Facet-Based SAR Imaging Model and Target Detection Based on YOLOv5 with CBAM and Another Detection Head. *Electronics* **2023**, *12*, 4039. <https://doi.org/10.3390/electronics12194039>

Academic Editors: Massimiliano Pieraccini, Weike Feng, Xiaowei Hu and Xingyu He

Received: 24 May 2023

Revised: 31 August 2023

Accepted: 17 September 2023

Published: 26 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Due to its all-day, all-weather features, Synthetic Aperture Radar (SAR), which can serve as a persistent surveillance and reconnaissance platform [1,2], has been widely used in civilian and military fields [3,4], such as geological and ecological surveillance, automatic driving, target detection, and so forth [5–7]. Effective detection of targets in SAR images under different conditions is of great significance in the practical applications of SAR images, especially in civil surveillance and military reconnaissance. However, there are fewer publicly released SAR image datasets, such as the SAR Ship Detection Dataset (SSDD, 1.16 K images with one category) [8], the Moving and Stationary Target Acquisition and Recognition (MSTAR, 3671 target chips with 10 classes) [9], than optical datasets such as Microsoft COCO (328 K images with 80 categories) [10] by the research institutes. In particular, the image dataset used for SAR aircraft detection is barely open access, which makes it impossible to train a robust model for non-cooperative target detection [11].

To efficiently detect non-cooperative targets in SAR images, transfer learning has been widely used in convolutional neural networks (CNN) due to the lack of publicly released aircraft SAR image datasets for data-driven solutions. By utilizing the existing knowledge

to learn new knowledge in the target domain, it can achieve a state-of-the-art result. SAR ship data has been utilized for model pretraining [12]. However, it will be more accurate to train a target detection model based on SAR images of non-cooperative aircraft.

To obtain a large number of SAR images of non-cooperative aircraft, the computational electromagnetism method and SAR imaging algorithm are studied in this research. As for the computation of electromagnetic (EM) scattering and SAR imaging processes, several facet-based methods have been proposed to reduce the computational burden, especially for electrically large targets. To estimate the electromagnetic backscattering of electrically large sea surfaces more accurately, an improved facet-based two-scale model is proposed, and then SAR imaging of the sea surface is conducted [13]. A facet model for rough Creamer sea surfaces under high sea states is presented in [14], and based on the scattering results, SAR images of sea surfaces are simulated under different conditions. For the SAR image simulation of ship wake, a semi-deterministic facet scattering model is proposed in [15], which shows good agreement with the experiment results. Extensive pieces of literature have researched the facet-based scattering model. However, few studies have researched fast facet-based SAR imaging algorithms for target and environment composite scenes. Based on the iterative physical optics method and Kirchhoff approximation, which can obtain the scattering coefficient of each facet, a fast facet-based SAR imaging model is studied to simulate SAR images of non-cooperative targets under different conditions.

After obtaining the SAR image simulation results of non-cooperative targets, a SAR image dataset can be obtained. Then, target detection and recognition based on the established dataset becomes feasible. As for the research on SAR image target detection, according to a comprehensive review and analysis of current research, the traditional SAR image target detection methods have limitations in feature extraction and computational resources, such as the constant false alarm rate (CFAR) method based on the statistical distribution of background clutter and threshold extraction [16], support vector machine (SVM) [17], and multi-feature extraction and fusion algorithm [18]. Transfer learning cannot achieve satisfactory performance for SAR image target detection because of the apparent discrepancy between optical and SAR images [19]. As a comparison, deep learning, which is based on convolutional neural networks (CNN), has been widely utilized in target detection and recognition due to its remarkable performance and accuracy [3], mainly owing to the massive labeled datasets. In addition, target detection based on deep learning can be divided into one-stage detectors such as Single Shot MultiBox Detector (SSD) [20], You Only Look Once (YOLO), and two-stage detectors such as Fast-RCNN [21], Faster-RCNN [22], and Mask R-CNN. Compared with the two-stage detectors, YOLOv5, a one-stage network and the latest architecture in the series, can boost efficiency and reduce the complexity of SAR image detection. To realize multi-target real-time detection for vehicles, aircraft, and nearshore ships, several methods are proposed based on YOLOv5 [23–25]. Based on the aforementioned deep learning methods, research on target detection for SAR images has gradually developed [2].

In this paper, through analyzing the composite EM scattering of ultra-low altitude targets above the rough surface and processing the radar echo signal, a novel SAR image data set is constructed based on which the experiment is conducted. Moreover, target detection and recognition are realized by utilizing an improved YOLO network. To validate the proposed systematic framework, some experiments were conducted, and the results showed that the improved YOLO network had better performance in SAR target detection and recognition.

The rest of this paper is organized as follows: The EM scattering of ultra-low altitude targets above the rough surface is analyzed in Section 2. Furthermore, based on the backscattering coefficient of the face, a fast facet-based SAR image simulation method was proposed. Section 3 presents the original YOLOv5 network and introduces another detection head, an adaptive and robust edge detection detector, and the convolutional block attention module to improve the YOLOv5 model for SAR image identification. Section 4 shows the pretreatment of the dataset, including the construction of the dataset

and labeling of the images. Then, several experiment examples are presented in Section 5 to demonstrate the feasibility, accuracy, and efficiency of the proposed method. Finally, it draws a conclusion in Section 6.

2. Electromagnetic Scattering Computation and SAR Imaging Algorithm

2.1. Iterative Physical Optics

Based on the magnetic field integral equation (MFIE) derived from the Stratton-Chu equation, the surface-induced electric current that meets the error requirements can be obtained through the iteration and updating of the surface-induced current [26]. As a result, far-field scattering can be obtained. The principal value integral of the MFIE can be expressed as:

$$\mathbf{H}(\mathbf{r}) = 2\mathbf{H}^{inc}(\mathbf{r}) + 2\text{P.V.} \int_S [\mathbf{J}(\mathbf{r}') \times \nabla G(\mathbf{r}, \mathbf{r}')] dS', \mathbf{r} \in S \quad (1)$$

where $G(\mathbf{r}, \mathbf{r}')$ is the free space Green's function, \mathbf{H}^{inc} denotes the incident magnetic field, \mathbf{r} and \mathbf{r}' are the observation point and source point, respectively.

From (1), \mathbf{J}_i can be obtained as:

$$\mathbf{J}_i = 2\hat{\mathbf{n}} \times \mathbf{H}^{inc}(\mathbf{r}) + 2\hat{\mathbf{n}} \times \text{P.V.} \int_S [\mathbf{J}_i \times \nabla G(\mathbf{r}, \mathbf{r}')] dS', \mathbf{r} \in S \quad (2)$$

The zero-order equivalent surface electric current density is [27]:

$$\mathbf{J}_i^0 = \frac{2E_i}{Z} \hat{\mathbf{n}} \times (\hat{\mathbf{k}}_i \times \hat{\mathbf{e}}_i) \exp(-jk\hat{\mathbf{k}}_i \cdot \mathbf{r}') \quad (3)$$

where Z is the surface impedance and $\hat{\mathbf{k}}_i$ denotes the unit propagation vector of the incident wave. Classical Jacobi iteration can be utilized to solve (3), and the $(k+1)$ th iteration of the current can be given by [28]:

$$\mathbf{J}_i^{(k+1)} = \mathbf{J}_i^{inc} + 2 \sum_{j=1, i \neq j}^N \hat{\mathbf{n}}_i \times \text{P.V.} \int_S [\mathbf{J}_j^{(k)} \times \nabla G(\mathbf{r}, \mathbf{r}')] dS', \mathbf{r} \in S \quad (4)$$

As the iteration process continues, the surface-induced electric current will fully converge. The iterative procedure will be terminated after the set number of iterations is reached or the relative error tends to be zero in a finite time [29]. Error convergence can be defined as:

$$\varepsilon_n = \frac{\|\Delta \mathbf{J}_i^{(n)}\|_2}{\|\mathbf{J}_i^{(n)}\|_2} \times 100\% \quad (5)$$

where $\Delta \mathbf{J}_i^{(n)} = \mathbf{J}_i^{(n)} - \mathbf{J}_i^{(n-1)}$. In general, when the error convergence is less than 3%, the iteration process can be halted. The geometry of typical targets is shown in Figure 1.

Additionally, fast far-field approximation, which consists of aggregation, translation, and disaggregation [28], is introduced to reduce operational complexity and time consumption. The fast far-field approximation (FaFFA) consists of aggregation, translation and disaggregation, is shown in Figure 2. Then, far-field scattering can be obtained utilizing the physical optics method.

2.2. Kirchhoff Approximation

To analyze the composite electromagnetic scattering of typical ultra-low altitude targets above the random rough surface, the facet-based iterative Kirchhoff approximation method is introduced to calculate the scattering of the rough surface [30]. In addition, the Kirchhoff approximation method is suitable for large, rough surfaces with small incident angles [31]. Figure 3 shows typical ultra-low altitude targets above the rough surface.

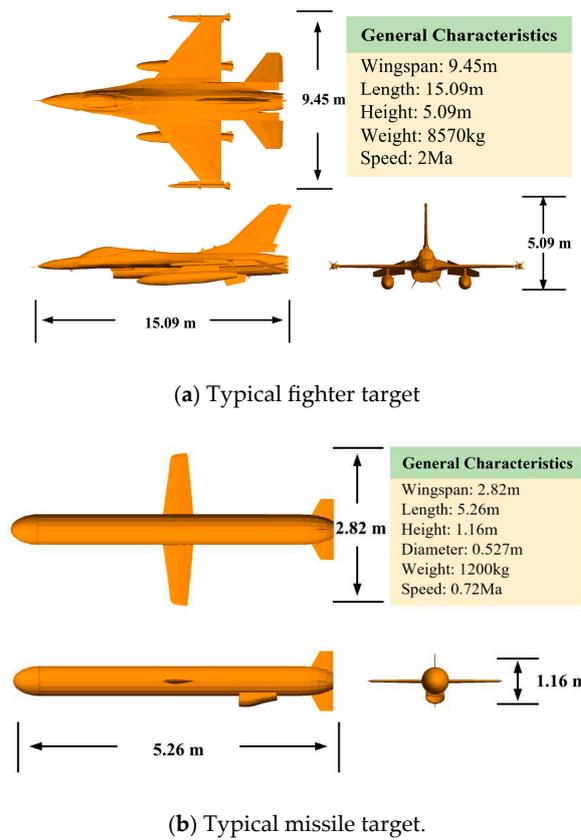


Figure 1. The geometry of typical targets with general characteristics.

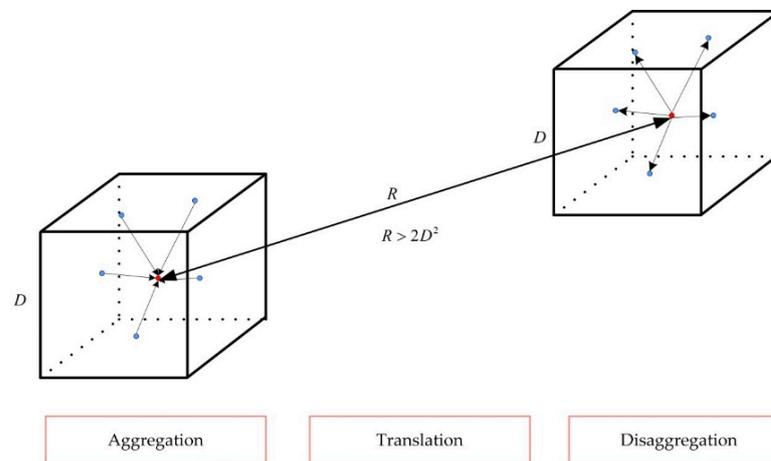


Figure 2. Far-field approximation with FaFFA acceleration, including aggregation translation and disaggregation.

Based on the Kirchhoff approximation, the induced electric currents J can be expressed as:

$$J_i = 2\hat{n}_i \times H^{inc} \tag{6}$$

where \hat{n}_i is the normal vector of the facet on the rough surface, H^{inc} denotes the incident magnetic field. According to the magnetic field integral equation, the induced currents J can be rewritten as:

$$J_i = 2\hat{n}_i \times H^i + 2\hat{n}_i \times [L(\mathbf{M}) + K(\mathbf{J})] \tag{7}$$

where

$$L(\mathbf{M}) = \iint_s -j\omega\epsilon_0 \left[\mathbf{M} \cdot \mathbf{G}(\mathbf{r}, \mathbf{r}') + (1/k_0^2) \nabla' \cdot \mathbf{M} \cdot \nabla G(\mathbf{r}, \mathbf{r}') \right] ds' \tag{8}$$

$$K(\mathbf{M}) = \iint_s \mathbf{J} \times \nabla G(\mathbf{r}, \mathbf{r}') ds' \tag{9}$$

where ϵ_0 is the relative dielectric constant and k_0 denotes the wave number of free space. Based on the impedance boundary condition, which can be given as:

$$\mathbf{M} = Z_S \mathbf{J} \times \hat{\mathbf{n}} \tag{10}$$

Then, the induced current \mathbf{J} can be expressed as:

$$\mathbf{J}_i = 2\hat{\mathbf{n}}_i \times \mathbf{H}^i + 2Z_S \hat{\mathbf{n}}_i \times L(\mathbf{J} \times \hat{\mathbf{n}}) + 2\hat{\mathbf{n}}_i \times K(\mathbf{J}) \tag{11}$$

The surface current can be given as:

$$\mathbf{J}_i^{n+1} = 2\hat{\mathbf{n}}_i \times \mathbf{H}^i + 2Z_S \sum_{j=1, j \neq i}^N \hat{\mathbf{n}}_j \times [L(\mathbf{J}_j^n \times \hat{\mathbf{n}}) + K(\mathbf{J}_j^n)] \tag{12}$$

where $n = 1, 2, \dots, N$ denote the facet on the rough surface. In general, the rough surface is meshed into triangular patches to calculate the electromagnetic scattering from the rough surface. Therefore, the scattering of the facet can be obtained by Kirchhoff approximation.

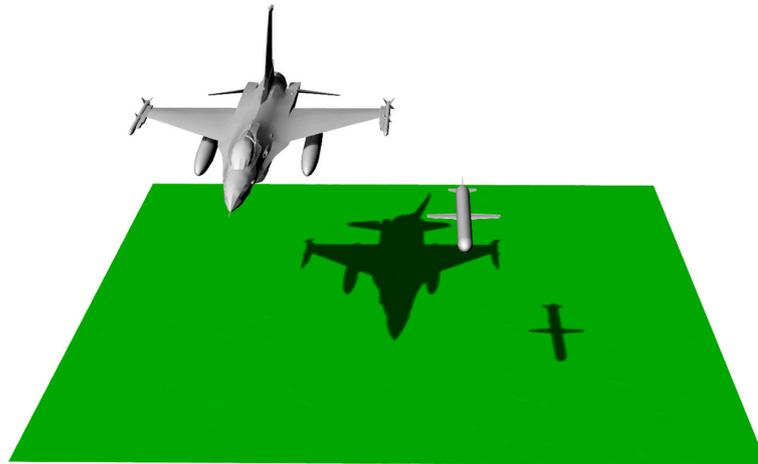


Figure 3. Typical ultra-low altitude targets above the rough surface.

2.3. SAR Imaging

Synthetic aperture radar, which can be installed on unmanned aerial vehicles (UAV) and airborne early warning aircraft, is a type of active observation system for object monitoring and information acquisition [32]. Especially, SAR imaging plays an important role in the implementation of target detection and identification. In order to obtain the SAR images, a facet-based SAR image simulation method is proposed [33]. The airborne SAR imaging system is illustrated in detail in Figure 4.

The SAR raw signal can be expressed as:

$$h_{strip}(x', y') = \iint \sigma(y, r) \exp \left[-j \frac{4\pi}{\lambda} \frac{\Delta f}{fc\tau} (r' - r - \Delta R)^2 \right] \cdot \omega^2 \left(\frac{y' - y}{X} \right) \exp \left(-j \frac{4\pi}{\lambda} \Delta R \right) \cdot \text{rect} \left[\frac{2(r' - r - \Delta R)}{c\tau} \right] dy dr \tag{13}$$

where $\sigma(y, r)$ denotes the backscattering coefficient of the facet, λ presents the wavelength in the free space, R denotes the slant range between the facet and the platform, Δf is the

bandwidth of the transmitting wave, and $\omega^2(\cdot)$ is the beam pattern amplitude modification. c represents the velocity of light in the free space and τ is the radar pulse duration.

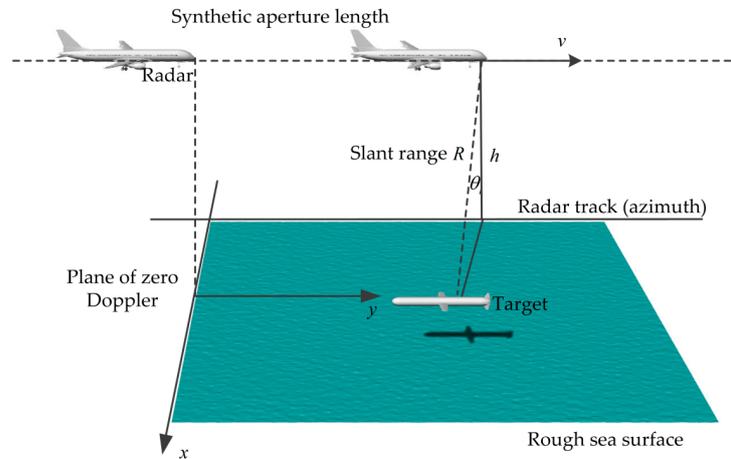


Figure 4. The airborne SAR imaging system.

Then, the SAR images of the target and rough surface can be reconstructed by a two-dimensional fast Fourier transform (FFT). The SAR simulation results for single and multiple targets can be seen in Figure 5. The range resolution and azimuth resolution are both 0.3 m.

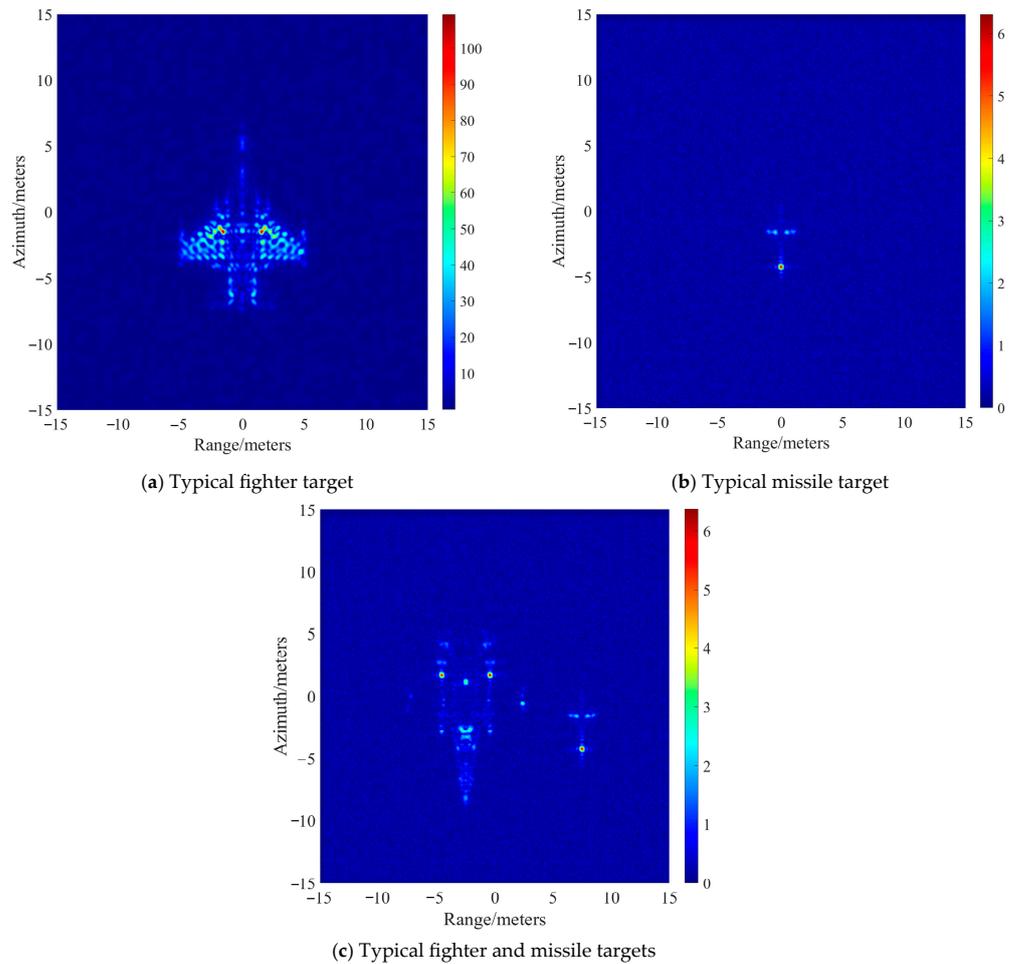


Figure 5. SAR images of typical targets.

The imaging results retain the structural information of the target, and compared with the SAR imaging method based on scattering center extraction, it achieves super-resolution imaging. The consumption of memory resources and computation time is less than that of other super-resolution imaging algorithms. Meanwhile, it is easy to conduct the simulation process, which is of some reference significance for real-time airborne SAR imaging.

3. Improved YOLOv5 Target Detection Model

In this section, the main idea of YOLOv5 is introduced in detail. YOLOv5 is a state-of-the-art object detection algorithm that solves object detection as a regression problem [34], and it meets the application requirements due to its fast recognition speed, high accuracy, and small model size [35].

3.1. Original YOLOv5

The main network of YOLOv5 consists of five parts: input, backbone, neck, prediction, and output. The overall network structure of YOLOv5 is shown in Figure 6.

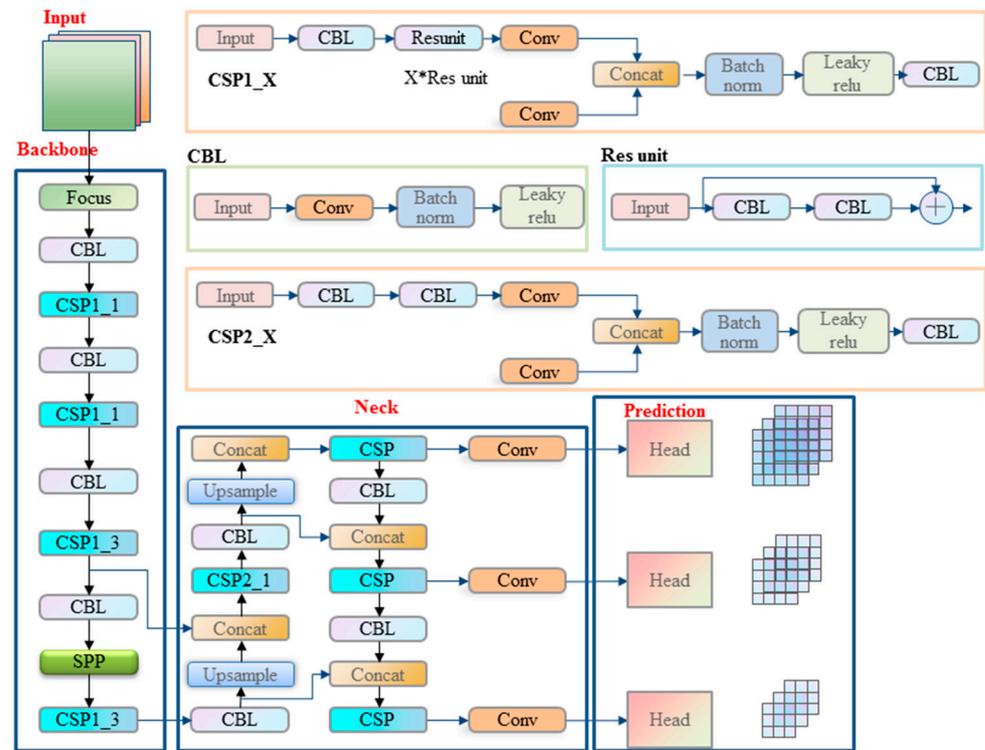


Figure 6. The network structure of YOLOv5.

3.1.1. Input

Data augmentation methods such as the Mosaic method and Cutmix are adopted to enhance the robustness of the model through randomly scaling, cropping, arranging, and splicing four or nine pictures [36,37]. As one of the most fundamental applications in the field of image analysis, the edge detection method is adopted to pre-process the images, utilizing the semantic and detail features to extract contiguous edge segments for high-level tasks [17]. Recently, many algorithms have been proposed in the field of edge detection.

3.1.2. Backbone

The raw backbone of YOLOv5 is CSPDarknet53, which contains a cross-stage partial (CSP) network that is designed for feature extraction. Meanwhile, the spatial pyramid pooling (SPP) structure was introduced to increase the perceptual field and separate contextual features [17].

3.1.3. Neck

In the neck part, the feature pyramid network (FPN) module and path aggregation network (PAN) module are introduced [38]. Through the FPN layer, high-level semantic features can be conveyed from top to bottom utilizing up-sampling [2,16]. On the contrary, the feature map is down-sampling again for localization feature fusion of different layers through PANet. Therefore, the two networks jointly enhance the capability of feature fusion, which will ultimately improve the performance of the overall model. The structure of the FPN and PAN modules is shown in Figure 7.

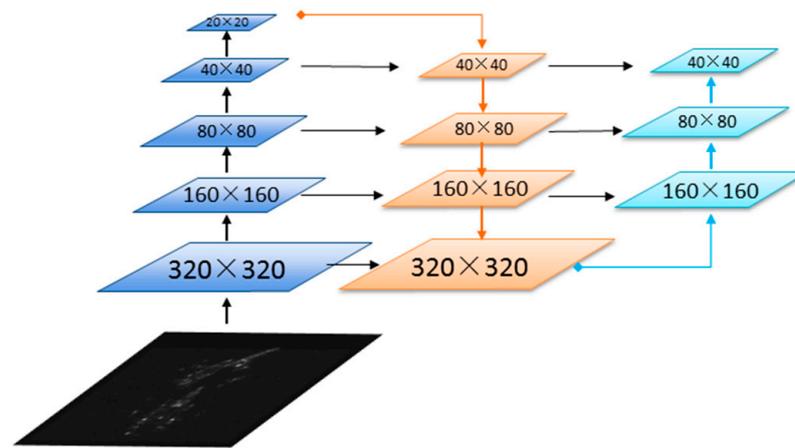


Figure 7. The structure of FPN and PAN.

3.1.4. Prediction

As shown in the network structure of YOLOv5, the prediction part consists of three prediction layers with different scales 80×80 , 40×40 , 20×20 , respectively, which are suitable for detecting different sizes of targets. Non-maximum suppression (NMS) is adopted to filter the preselector in the process of prediction.

3.2. Improved Method

3.2.1. Small Target Detection

As the UAV aerial images have few pixel features, the detection and recognition model is required to have a strong ability to detect small targets. At first, a higher resolution reconstruction method is introduced in this model to increase the detailed information for target detection and recognition. Through convolution operation, the image feature information is extracted and nonlinearly mapped into high-dimensional vectors. Then, the deconvolution operation is conducted for up-sampling to reconstruct a higher-resolution image. In addition, before the convolution operation, zero filling is performed to guarantee the input and output images are the same size.

Moreover, another prediction head (PH) is added to the model to improve the performance of small target detection. A low-level and high-resolution feature map is generated by a low-level fusion layer, through which the feature map is further up-sampled and concatenated with the feature map belonging to the low-level backbone networks. Although the memory and computation costs will increase, the target detection performance for small targets will improve.

In this paper, an adaptive and robust edge detection detector was adopted to extract the edge contours of the SAR images. The suspicious targets in the SAR image are extracted by edge detection using the Canny operator. The target centers extracted by edge detection are used as the initial cluster centers, and all targets extracted are clustered with the k-means method. The clustering results are recognized as aircraft candidates, which are discriminated by a YOLOv5 network afterward to obtain the final detection results. The input image and the edge detection result are shown in Figure 8. The experimental results

have demonstrated that this method can generate a semantically meaningful contour without significant discontinuities. However, some scattered noisy details were noticed in the edge detection result.

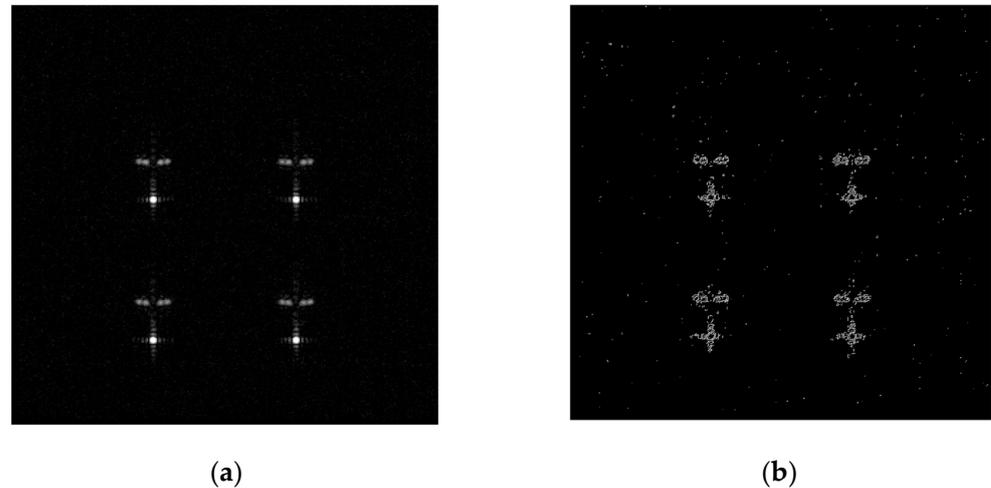


Figure 8. Original image and the edge detection result: (a) Original image; (b) Edge detection result.

3.2.2. Convolutional Block Attention Module

As is well known, the attention mechanisms of the human eyeball and brain are vitally important in human perception. Up until now, several researchers have made attempts to introduce the attention module into the large-scale target detection model to improve the performance of the deep learning model in target detection and recognition tasks. In this paper, a Convolutional Block Attention Module (CBAM) that contains both channel and spatial attention modules is introduced into the original YOLOv5 architecture.

On the one hand, channel attention focuses on what to pay attention to. When the spatial information is aggregated through average pooling and maximum pooling, two different spatial descriptors can be obtained. Through the shared multilayer perceptron, the output feature vectors can be merged, and the channel attention map \mathbf{M}_c can be produced by:

$$\mathbf{M}_c(\mathbf{F}) = \sigma(\text{MLP}(\text{AvgPool}(\mathbf{F})) + \text{MLP}(\text{MaxPool}(\mathbf{F}))) \quad (14)$$

where $\sigma(\cdot)$ denotes the sigmoid function, MLP presents the multilayer perceptron weights, and \mathbf{F} is the input feature.

On the other hand, different from the channel attention mentioned above, spatial attention focuses on where to pay attention. At first, average pooling and maximum pooling operations are applied along the channel axis to aggregate channel information features, and then a 2D spatial attention map \mathbf{M}_s can be obtained by concatenating them.

$$\mathbf{M}_s(\mathbf{F}) = \sigma\left(f^{7 \times 7}([\text{AvgPool}(\mathbf{F}); \text{MaxPool}(\mathbf{F})])\right) \quad (15)$$

where $\sigma(\cdot)$ denotes the sigmoid function and $f^{7 \times 7}$ present a 7×7 convolution operation.

By combining the channel and spatial attention modules, the composite module can focus on what and where to pay attention, respectively. According to the previous literature, the results of sequential arrangement are substantially better than parallel arrangement. As for the sequential process, experiment results with different attention arrangement orders show that channel-first attention performs at a higher level than spatial-first attention [39]. The diagrams of the channel and spatial attention modules are shown in Figure 9.

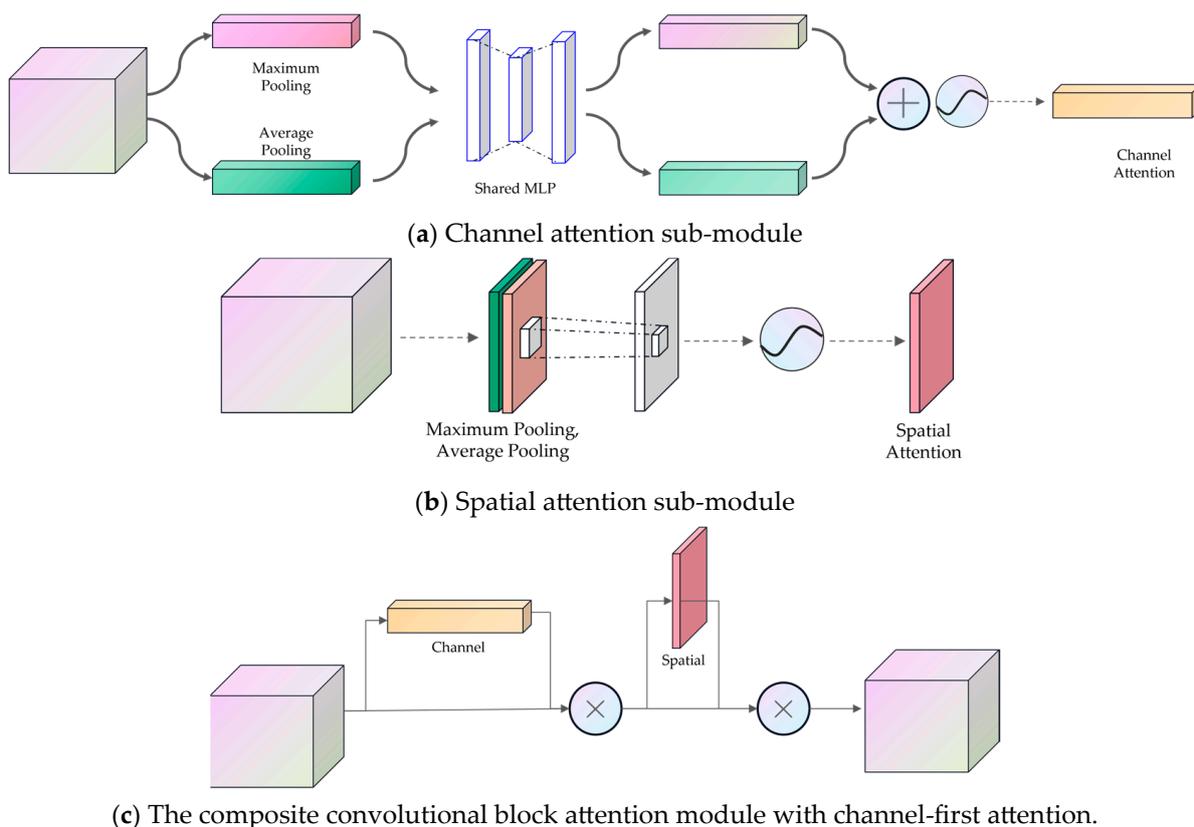


Figure 9. A diagram of the channel and spatial attention modules.

4. Pretreatment of the SAR Image

In this section, the SAR image dataset includes two kinds of military targets. All experiments are conducted on a WorkStation with an Intel (R) Xeon (R) Silver 4110 CPU @2.1 GHz (32 CPUs). The GPU version is NVIDIA Quadro P5000 3503 MB, the simulation platform is Pytorch 1.8.2 with CUDA 10.2 on Python 3.8.8, and the operating system is Windows 7.

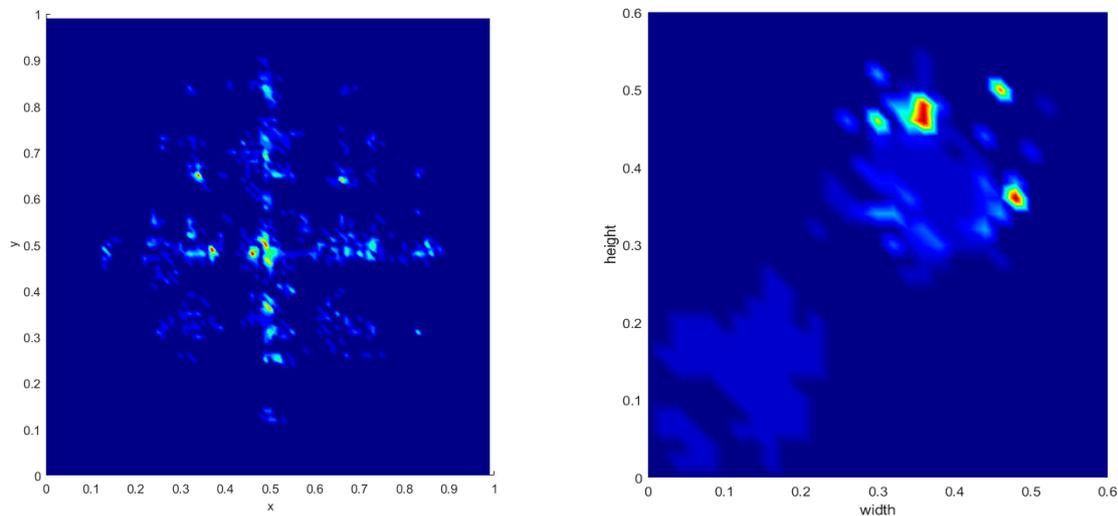
4.1. Dataset Construction

According to the facet-based SAR imaging model mentioned above, SAR images are simulated and obtained under different conditions. It contains different numbers of targets, such as single targets and multiple targets with different target orientations 30°, 45°, 60°, 75°, and 90°, and different positional relationships, such as parallel distribution and tandem distribution between multiple targets. As a result, the SAR images have diverse target types and relative position relationships in the dataset, which meets the requirements for detection and recognition. Moreover, the original dataset was adapted to increase the robustness of the model through brightness and contrast adjustment, bit depth conversion [36], and horizontal and vertical mirroring. Table 1 shows the detailed distribution of the raw dataset.

Table 1. SAR image information for the dataset.

Polarization	Target Type	
	Fighter	Missile
HH	550	450
HV	550	450
VH	550	450
VV	550	450

As shown in Figure 10, the targets have traits such as wide distribution, and dispersive position. The length and width distribution of the labels is shown in Figure 10b. It can be seen that the small and medium-sized targets account for a large proportion, which can strengthen the robustness of the proposed method [40].



(a) Heat map of the location distribution (b) Heat map of the length and width distribution.

Figure 10. Heat map of the target trait which maps the minimum value of the data to dark blue, the maximum value to red, and the middle value to yellow.

4.2. Labelling the SAR Images

Before the training process, the SAR image data needs to be preprocessed, manually labeled, and framed utilizing the image annotation tool `labelImg` [17] as shown in Figure 11. The dots indicate the coordinate points of the box in which the target is located, with the missile target in the red box and the fighter in the green box. Then make a dataset concerning the COCO dataset format, generating the TXT file [2]. Moreover, the images without targets were deleted. As a result, an experiment dataset with about 1000 targets was established. In addition, the dataset is divided into a training set and a test set in the ratio of 8:2.

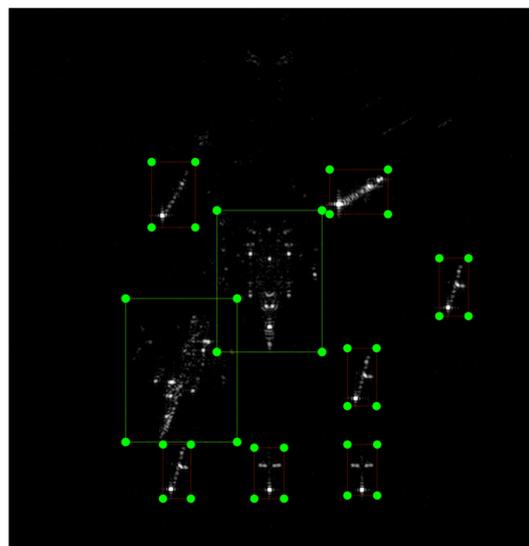


Figure 11. Manually labeled and framed utilizing the image annotation tool.

5. Results and Discussion

5.1. Model Training

Initial training was conducted utilizing pre-trained weights for YOLOv5s on the COCO dataset. The batch size is set to 300, and the learning rate is 0.001. Meanwhile, stochastic gradient descent is adopted as the optimizer.

5.2. Test and Evaluate

To comprehensively evaluate the performance of the model, precision rate, recall rate, and average precision (AP) are adopted as evaluation indexes. Therefore, the mAP value of different targets can be calculated. The calculation formulas can be expressed as [41]:

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (16)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (17)$$

$$\text{mAP} = \int_0^1 P(R)d(R) \quad (18)$$

where, TP , FP , and FN denote true positives, false positives, and false negatives, respectively.

To validate the detection and recognition performance of the proposed method, two published methods are chosen for comparison. Table 2 shows the performance comparisons between the proposed method and other detectors such as dense nested attention networks (DNANet) and attentional local contrast networks. It can be seen from the comparison results that the mAP value of the proposed method is 85.8%, 7.5% higher than that of the original YOLOv5 model, which proves the proposed method can realize a significant improvement in detection performance. In particular, it is more effective to detect small targets. To validate the simulated data generated by the SAR imaging model, we carried out the validation experiment on SAR-AIRCRAFT-1.0, a high-resolution SAR aircraft detection and recognition dataset published by the Aerospace Information Research Institute, Chinese Academy of Sciences. Based on a small amount of SAR image data, good recognition results can be obtained.

Figure 12 shows the result of the proposed method in the test set. It can be seen from Figure 12a that the loss value declines rapidly and shows rapid convergence of the iteration procedure. It also indicates that the mAP@0.5 is up to 98.4% and the precision reaches 98.36% after 100 iteration steps. It demonstrates that the module is more effective at target detection. Moreover, the mAP@0.5:0.05:0.95 reaches 0.78 through 150 epochs.

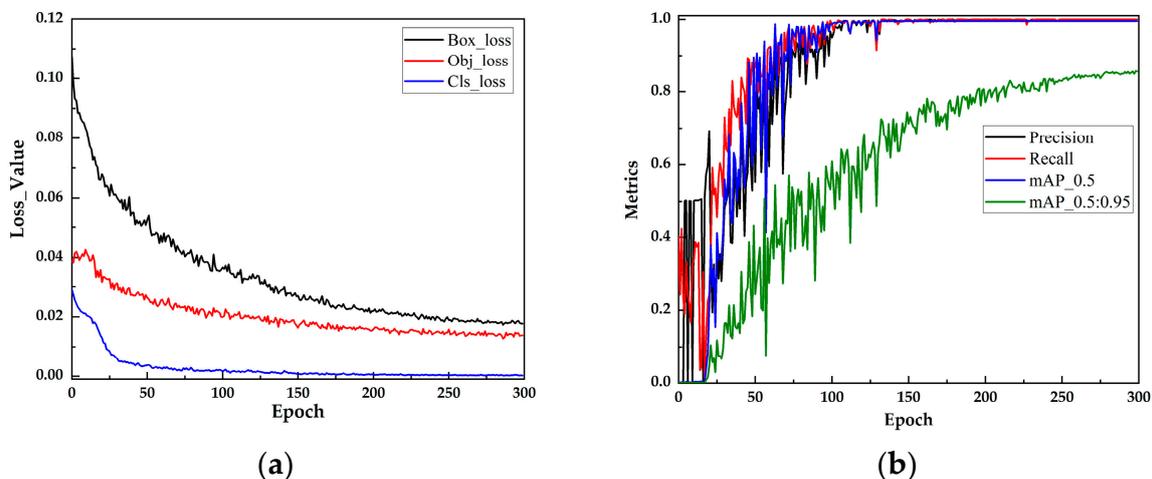


Figure 12. The loss value and metrics versus epoch curves: (a) The train loss value curves; (b) The metrics value curves.

Figure 13 shows the comparison results of the original YOLOv5 model and the proposed model, which introduced another prediction head (YOLOv5 + PH), another prediction head, and a convolution block attention module (YOLOv5 + PH + CBAM), respectively. It can be seen that the precision and average precision of the proposed algorithm are higher than those of the original YOLOv5 method.

Table 2. Comparison of dataset test results.

Detector	mAP	Precision
YOLOv4	56.8	77.6
ALCN	76.7	82.1
YOLOv5	78.3	89.2
DNANet	83.6	95.4
Proposed	85.8	98.4

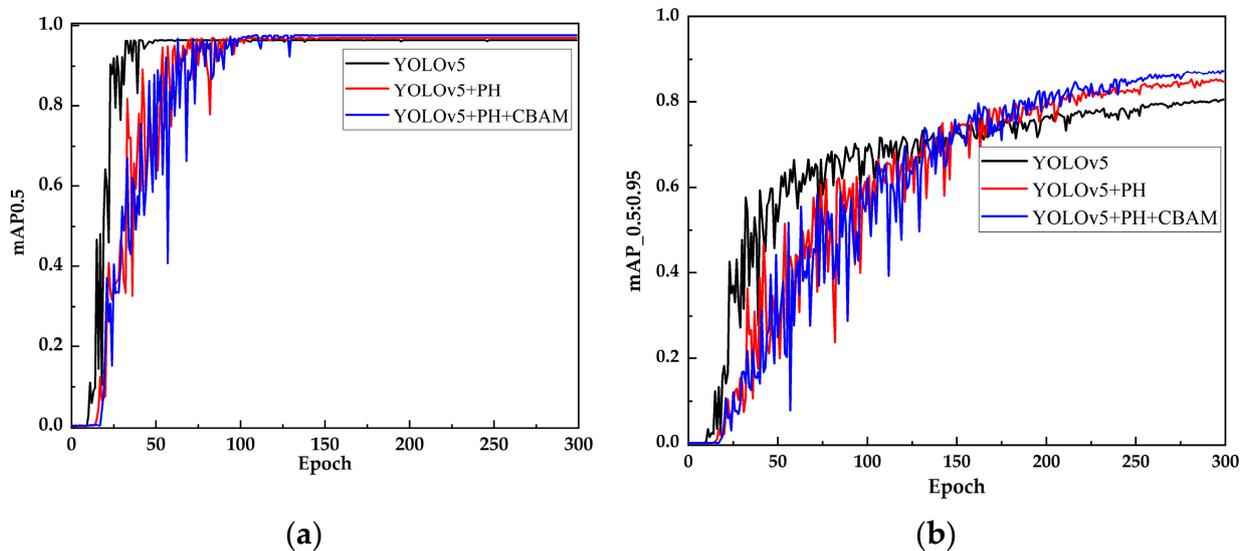
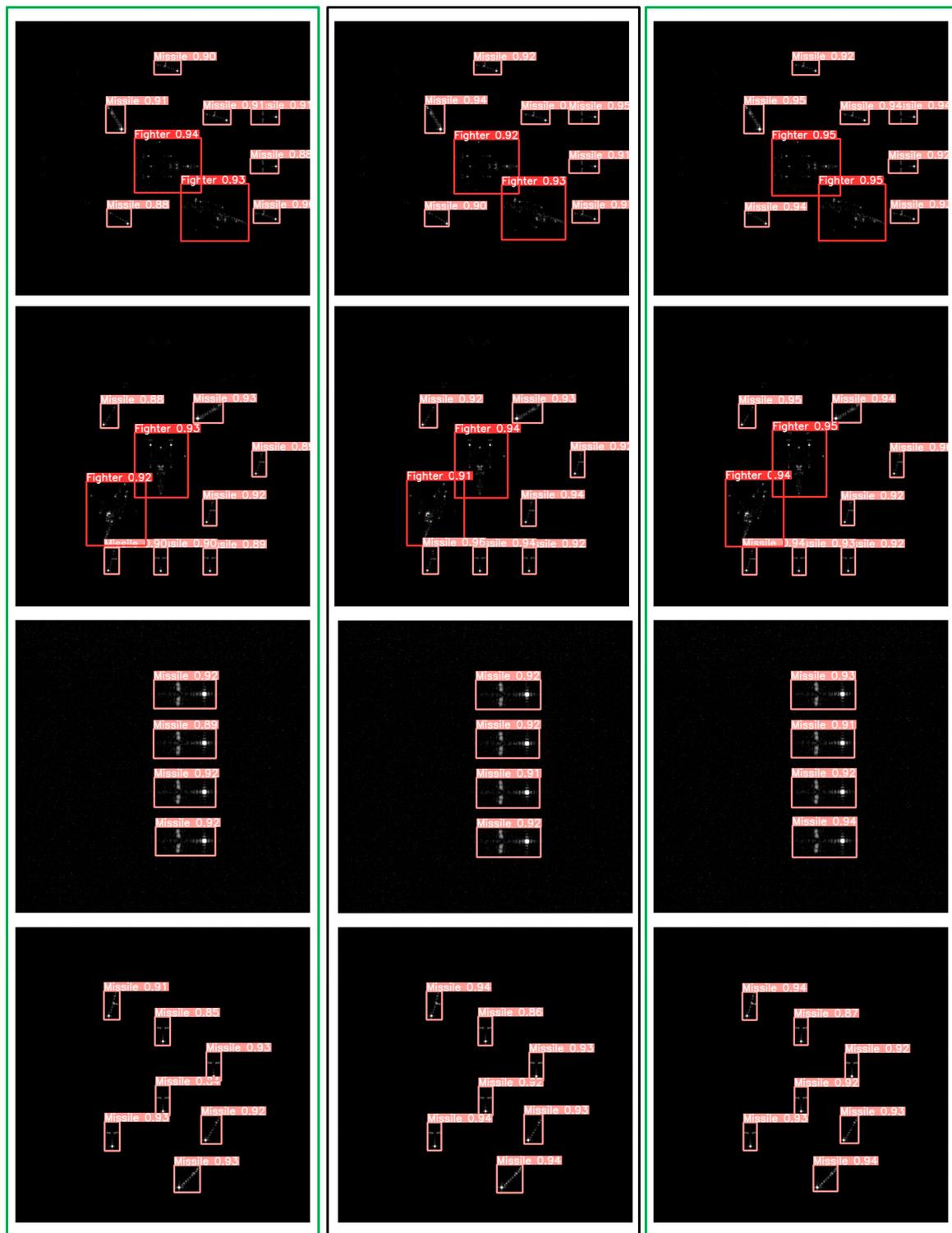


Figure 13. The mean average precision versus the epoch of the original YOLOv5 model and the proposed model: (a) The mAP@0.5; (b) The mAP@0.5:0.95.

However, due to the introduction of another prediction head and a convolutional block attention module, the improved model has great instability in the preliminary stage of deep learning training. Moreover, the overall network converges more slowly, so it takes more training epochs to reach stability. Through comprehensively analyzing the advantages and disadvantages of the proposed model, it can be concluded that the method proposed in this paper has significant improvements compared with the original model.

Figure 14 shows some of the results with different types and numbers of targets to demonstrate the detection performance. It can be seen that the proposed method shows good detection performance for images with multiple targets, especially for the detection of small targets, which demonstrates the better accuracy and robustness of the improved model in this paper. Meanwhile, the false detection rate has dropped significantly, which also verifies the improvement of the mAP indicators in Figure 13. In addition to using the self-established dataset, the improved model can be adapted to detect targets on the publicly available SAR image dataset.



(a) The original YOLOv5 model

(b) YOLOv5 model + Prediction Head

(c) YOLOv5 model + Prediction Head + CBAM module

Figure 14. The comparison results of the SAR images for target detection with missile target in orange box and fighter target in red box.

6. Conclusions

Effective target detection and recognition based on high-resolution SAR images is very important in civilian and military fields. However, there are few publicly released

SAR image datasets of typical non-cooperative targets. In this paper, a fast facet-based SAR imaging model was proposed to simulate the SAR images of typical non-cooperative targets under different conditions. As a result, a dataset can be constructed. Moreover, a target detection model based on an improved YOLOv5 model is introduced.

Combining the iterative physical optics which was accelerated by fast far-field approximation and the Kirchhoff approximation algorithm, the composite EM scattering of ultra-low altitude targets above rough surfaces was analyzed. Furthermore, based on the backscattering coefficient of the facet on the target surface, the radar echo can be obtained. Then, a fast facet-based SAR image simulation method was proposed. Based on the SAR images obtained by simulations under different conditions, a dataset was constructed. Furthermore, we introduced another detection head, an adaptive and robust edge detection detector, and the convolutional block attention module to improve the YOLOv5 model. Then, target detection and recognition are realized utilizing the improved YOLOv5 network. Compared with other YOLO series detectors, the simulation results show a significant improvement in the accuracy of detection and recognition. Although the method has achieved good performance in target detection, there are several shortcomings in the present study, especially that the overall network converges more slowly during the training process, so it takes more training epochs to reach stability.

In future research, the automatic target recognition system presented in this paper can provide a reference for the detection and recognition of non-cooperative aircraft targets and has great practical application in situational awareness of battlefield conditions. For example, the SAR imaging system and the automatic target recognition system can be equipped in unmanned reconnaissance aircraft, which can realize real-time detection and recognition of non-cooperative targets and provide data support for the ground-to-air missile weapon system in combating incoming targets. In addition, with the rapid development of unmanned aerial vehicle technology, real-time SAR imaging and target detection of typical unmanned aerial vehicles are of great importance for air defense. Moreover, detecting various typical targets in complex environments such as coastal cities, islands, and valley areas will be an important research direction.

Author Contributions: Conceptualization, C.T.; Formal analysis, J.S.; Investigation, M.W.; Methodology, Q.W., T.W., J.S., C.T. and T.S.; Software, Z.W. and T.S.; Supervision, C.T. and M.W.; Validation, Z.W.; Visualization, Q.W.; Writing—original draft, Q.W.; Writing—review and editing, T.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant numbers 62201611 and 62301597, and the Natural Science Foundation of Shaanxi Province, grant numbers 2021JQ-362 and 2023-JC-QN-0647, The Program of The Youth Innovation Team of Shaanxi University.

Data Availability Statement: Not applicable.

Acknowledgments: Thanks to my instructor, Tong, for his technical assistance throughout the project. Meanwhile, I would like to acknowledge all my friends for their hard work on this paper.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Wu, W.; Li, H.; Li, X.; Guo, H.; Zhang, L. PolSAR Image Semantic Segmentation Based on Deep Transfer Learning—Realizing Smooth Classification With Small Training Sets. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 977–981. [[CrossRef](#)]
2. Yang, W.; Zhang, Z. SAR Images Target Detection Based on YOLOv5. In Proceedings of the 2021 4th International Conference on Information Communication and Signal Processing (ICICSP), Shanghai, China, 24–26 September 2021; pp. 342–347.
3. Sun, Z.; Lei, Y.; Leng, X.; Xiong, B.; Ji, K. An Improved Oriented Ship Detection Method in High-resolution SAR Image Based on YOLOv5. In Proceedings of the 2022 Photonics & Electromagnetics Research Symposium (PIERS), Hangzhou, China, 25–29 April 2022; pp. 647–653.
4. Zhang, J.; Xing, M.; Xie, Y. FEC: A Feature Fusion Framework for SAR Target Recognition Based on Electromagnetic Scattering Features and Deep CNN Features. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 2174–2187. [[CrossRef](#)]

5. Deng, J.; Bi, H.; Yin, Y.; Lu, X.; Liang, W. Sparse SAR Image Based Automatic Target Recognition by YOLO Network. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 3597–3600.
6. Selvam, N.; Nagesa, Y.; Negesa, F. Deep Learning Approach with Optimization Algorithm for Reducing the Training and Testing Time in SAR Image Detection and Recognition. *Indian J. Sci. Technol.* **2022**, *15*, 371–385. [[CrossRef](#)]
7. Zhou, Z.; Cui, Z.; Zang, Z.; Meng, X.; Cao, Z.; Yang, J. UltraHi-PrNet: An Ultra-High Precision Deep Learning Network for Dense Multi-Scale Target Detection in SAR Images. *Remote. Sens.* **2022**, *14*, 5596. [[CrossRef](#)]
8. Zhang, T.; Zhang, X.; Li, J.; Xu, X.; Wang, B.; Zhan, X.; Xu, Y.; Ke, X.; Zeng, T.; Su, H.; et al. SAR Ship Detection Dataset (SSDD): Official Release and Comprehensive Data Analysis. *Remote. Sens.* **2021**, *13*, 3690. [[CrossRef](#)]
9. Furukawa, H. Deep Learning for Target Classification from SAR Imagery: Data Augmentation and Translation Invariance. *arXiv* **2017**, arXiv:1708.07920.
10. Lin, T.; Maire, M.; Belongie, S.J.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014.
11. Zhang, P.; Xu, H.; Tian, T.; Gao, P.; Li, L.; Zhao, T.; Zhang, N.; Tian, J. SEFEPNet: Scale Expansion and Feature Enhancement Pyramid Network for SAR Aircraft Detection With Small Sample Dataset. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 3365–3375. [[CrossRef](#)]
12. Guangshuai, L.L.; Juan, S.U.; Yihong, L.L.; Xiang, L.I. Aircraft detection in SAR images based on convolutional neural network and attention mechanism. *Syst. Eng. Electron.* **2021**, *43*, 3202–3210.
13. Li, D.; Zhao, Z.; Zhao, Y.; Huang, Y.; Liu, Q.-H. An Improved Facet-Based Two-scale model for Electromagnetic Scattering from Sea Surface and SAR Imaging. In Proceedings of the 2019 IEEE Radar Conference (RadarConf), Boston, MA, USA, 22–26 April 2019; pp. 1–4. [[CrossRef](#)]
14. Li, J.; Zhang, M.; Fan, W.; Nie, D. Facet-Based Investigation on Microwave Backscattering From Sea Surface With Breaking Waves: Sea Spikes and SAR Imaging. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2313–2325. [[CrossRef](#)]
15. Wang, J.-K.; Zhang, M.; Chen, J.-L.; Cai, Z. Application of Facet Scattering Model in SAR Imaging of Sea Surface Waves with Kelvin Wake. *Prog. Electromagn. Res. B* **2016**, *67*, 107–120. [[CrossRef](#)]
16. Fu, Q.; Chen, J.; Yang, W.; Zheng, S. Nearshore Ship Detection on SAR Image Based on Yolov5. In Proceedings of the 2021 2nd China International SAR Symposium (CISS), Shanghai, China, 3–5 November 2021; pp. 1–4.
17. Jiang, J.; Fu, X.; Qin, R.; Wang, X.; Ma, Z. High-Speed Lightweight Ship Detection Algorithm Based on YOLO-V4 for Three-Channels RGB SAR Image. *Remote. Sens.* **2021**, *13*, 1909. [[CrossRef](#)]
18. Guo, Q.; Wang, H.; Kang, L.; Li, Z.; Xu, F. Aircraft Target Detection from Spaceborne SAR Image. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July 2019–2 August 2019; pp. 1168–1171.
19. Huang, Z.; Pan, Z.; Lei, B. What, Where, and How to Transfer in SAR Target Recognition Based on Deep CNNs. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2324–2336. [[CrossRef](#)]
20. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.E.; Fu, C.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016.
21. Girshick, R.B. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 11–18 December 2015; pp. 1440–1448.
22. Zhang, Y.; Song, C.; Zhang, D. Small-scale aircraft detection in remote sensing images based on Faster-RCNN. *Multimed. Tools Appl.* **2022**, *81*, 18091–18103. [[CrossRef](#)]
23. Song, X.; Gu, W. Multi-objective real-time vehicle detection method based on yolov5. In Proceedings of the 2021 International Symposium on Artificial Intelligence and its Application on Media (ISAIAM), Xi'an, China, 21–23 May 2021; pp. 142–145.
24. Guo, Y.; Chen, S.; Zhan, R.; Wang, W.; Zhang, J. SAR Ship Detection Based on YOLOv5 Using CBAM and BiFPN. In Proceedings of the IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 2147–2150.
25. Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 11–17 October 2021; pp. 2778–2788.
26. Rashidi-Ranjbar, E.; Dehmollaian, M. Target Above Random Rough Surface Scattering Using a Parallelized IPO Accelerated by MLFMM. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1481–1485. [[CrossRef](#)]
27. Liao, S.; Ou, L. IPO: Iterative Physical Optics Image Approximation. *arXiv* **2020**, arXiv:2007.01861.
28. Li, C.; Tong, C.; Bai, Y.; Qi, L. Iterative physical optics model for electromagnetic scattering and Doppler analysis. *J. Syst. Eng. Electron.* **2016**, *27*, 581–589. [[CrossRef](#)]
29. Burkholder, R.J.; Lundin, T. Forward-backward iterative physical optics algorithm for computing the RCS of open-ended cavities. *IEEE Trans. Antennas Propag.* **2005**, *53*, 793–799. [[CrossRef](#)]
30. Han, C.; Li, Y.; Wu, Z. SAL imaging method based on Kirchhoff approximation. In Proceedings of the 2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE), Hangzhou, China, 3–6 December 2018; pp. 1–5.
31. Tian, J.; Tong, J.; Shi, J.; Gui, L. A new approximate fast method of computing the scattering from multilayer rough surfaces based on the Kirchhoff approximation. *Radio Sci.* **2017**, *52*, 186–195. [[CrossRef](#)]

32. Xiong, X.; Li, G.; Ma, Y.; Chu, L. New slant range model and azimuth perturbation resampling based high-squint maneuvering platform SAR imaging. *J. Syst. Eng. Electron.* **2021**, *32*, 545–558. [[CrossRef](#)]
33. Wang, T.; Tong, C.; Gao, Q.; Li, X. A Synthetic Aperture Radar Image Model of Ship on Sea. In Proceedings of the 2018 International Conference on Smart Materials, Intelligent Manufacturing and Automation (SMIMA 2018), MATEC Web of Conferences, Nanjing, China, 24–26 May 2018; Volume 173.
34. Sruthi, M.S.; Poovathingal, M.J.; Nandana, V.N.; Lakshmi, S.; Samshad, M.; Sudeesh, V.S. YOLOv5 based Open-Source UAV for Human Detection during Search And Rescue (SAR). In Proceedings of the 2021 International Conference on Advances in Computing and Communications (ICACC), Kochi, Kakkannad, India, 21–23 October 2021; pp. 1–6.
35. Sozzi, M.; Cantalamessa, S.; Cogato, A.; Kayad, A.; Marinello, F. Automatic Bunch Detection in White Grape Varieties Using YOLOv3, YOLOv4, and YOLOv5 Deep Learning Algorithms. *Agronomy* **2022**, *12*, 319. [[CrossRef](#)]
36. YKun; Man, H.; Yanling, L. Multi-target Detection in Airport Scene Based on Yolov5. In Proceedings of the 2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Changsha, China, 20–22 October 2021; pp. 1175–1177.
37. Zhou, J.; Jiang, P.; Zou, A.; Chen, X.; Hu, W. Ship Target Detection Algorithm Based on Improved YOLOv5. *J. Mar. Sci. Eng.* **2021**, *9*, 908. [[CrossRef](#)]
38. Ding, K.; Li, X.; Guo, W.; Wu, L. Improved object detection algorithm for drone-captured dataset based on yolov5. In Proceedings of the 2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 14–16 January 2022; pp. 895–899.
39. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the 15th European Conference on Computer Vision (ECCV), European Conference on Computer Vision (2018), Munich, Germany, 8–14 September 2018; Volume 11211, pp. 3–19. [[CrossRef](#)]
40. Xiao, B.; Guo, J.; He, Z. Real-Time Object Detection Algorithm of Autonomous Vehicles Based on Improved YOLOv5s. In Proceedings of the 2021 5th CAA International Conference on Vehicular Control and Intelligence (CVCI), Tianjin, China, 29–31 October 2021; pp. 1–6.
41. Liu, Y.; He, G.; Wang, Z.; Li, W.; Huang, H. NRT-YOLO: Improved YOLOv5 Based on Nested Residual Transformer for Tiny Remote Sensing Object Detection. *Sensors* **2022**, *22*, 4953. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.