remote sensing

MDPI

Article

FEMSFNet: Feature Enhancement and Multi-Scales Fusion Network for SAR Aircraft Detection

Wenbo Zhu [1,2], Liu Zhang [1,2], Chunqiang Lu [1,2], Guowei Fan [1,2], Ying Song [1,2], Jianbo Sun [3,4] and Xueying Lv [1,2,*]

[1] National Geophysical Exploration Equipment Engineering Research Center, Jilin University, Changchun 130026, China; zhuwb23@mails.jlu.edu.cn (W.Z.); zhangliu@jlu.edu.cn (L.Z.)
[2] College of Instrumentation & Electrical Engineering, Jilin University, Changchun 130061, China
[3] Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
[4] University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: lvxueying@jlu.edu.cn

Abstract: Aircraft targets, as high-value subjects, are a focal point in Synthetic Aperture Radar (SAR) image interpretation. To tackle challenges like limited SAR aircraft datasets and shortcomings in existing detection algorithms (complexity, poor performance, weak generalization), we present the Feature Enhancement and Multi-Scales Fusion Network (FEMSFNet) for SAR aircraft detection. FEMSFNet employs diverse image augmentation and integrates optimized Squeeze-and-Excitation Networks (SE) with residual network (ResNet) in a SdE-Resblock structure for a lightweight yet accurate model. It introduces ssppf-CSP module, an improved pyramid pooling model, to prevent receptive field deviation in deep network training. Tailored for SAR aircraft detection, FEMSFNet optimizes loss functions, emphasizing both speed and accuracy. Evaluation on the SAR Aircraft Detection Dataset (SADD) demonstrates significant improvements compared to the contrasted algorithms: precision rate (92%), recall rate (96%), and F1 score (94%), with a maximum increase of 12.2% in precision, 12.9% in recall, and 13.3% in F1 score.

Keywords: residual network; feature enhancement; multi-scales fusion; SAR; aircraft detection

Citation: Zhu, W.; Zhang, L.; Lu, C.; Fan, G.; Song, Y.; Sun, J.; Lv, X. FEMSFNet: Feature Enhancement and Multi-Scales Fusion Network for SAR Aircraft Detection. *Remote Sens.* **2024**, *16*, 1589. https://doi.org/10.3390/rs16091589

Academic Editor: Dusan Gleich

Received: 23 January 2024
Revised: 23 April 2024
Accepted: 27 April 2024
Published: 29 April 2024

1. Introduction

Synthetic Aperture Radar (SAR) stands out as an active microwave remote sensing technology, offering uninterrupted, all-day, and all-weather Earth surface observation [1]. It remains unaffected by conditions such as illumination, clouds, or weather, making it indispensable in remote sensing [2]. SAR has found extensive applications in both military and civilian domains, emerging as a pivotal tool for information acquisition [3]. In military contexts, the detection of aircraft holds a central position in air defense research, leveraging the distinctive advantages offered by SAR images [4]. Consequently, there is a global research emphasis on enhancing aircraft target detection in SAR imagery.

SAR imaging, distinct from optical methods, poses challenges in detecting and identifying aircraft targets due to its longer wavelength and complex mechanism [5]. The irregular distribution of land clutter, marked by bright backscattering points, introduces interference [6]. SAR images often showcase intricate terrain features, complicating aircraft target detection as these features may mimic the representation of targets. Targets in SAR images manifest as irregular bright spots, necessitating spot integration for effective recognition. The varied imaging characteristics of aircraft targets in SAR images, combined with fluctuating scattering conditions, reduce the relevance of traditional manually designed features [6–9]. Detecting aircraft targets in SAR images is, therefore, a significant but intricate research direction.

Traditional SAR image target detection relies heavily on model features, encompassing characteristics like the target's outline, size, texture, and scattering center [10–13].

A common traditional algorithm is the constant false alarm rate (CFAR) based on clutter statistics and threshold extraction [14]. Scholars have delved into statistical features and non-uniform backgrounds, proposing improved CFAR algorithms like CA-CFAR [15], SOCA-CFAR [16], GOCA-CFAR [17], OS-CFAR [18], and VI-CFAR [19]. Ai et al. [20] proposed an SAR detection algorithm using bilateral fine-tuning thresholds, enhancing performance in ocean backgrounds by fitting the target to clutter with higher contrast. Chen et al. [21] introduced an improved constant false alarm rate detection algorithm based on multiscale contrast and variable windows, elevating target detection accuracy in SAR images. Model-based recognition methods achieve heightened target recognition accuracy with an evolving template database. However, this approach demands multiple iterations for high-precision simulated images, taxing computation speed, and model accuracy. Furthermore, model-based methods suffer from high computational complexity and low efficiency. Consequently, researchers are increasingly exploring machine learning algorithms, such as support vector machines, neural networks, and adaptive enhancement, for automatic interpretation of SAR targets.

Recently, deep learning-based target detection has experienced rapid development across various fields [22–25]. Notable strides have been achieved in aircraft target recognition in Synthetic Aperture Radar (SAR) images through deep learning [26–28]. This progress is largely credited to the automatic learning and pattern recognition capabilities inherent in deep learning methods for handling complex features. Zhao et al. [29] introduced an SAR aircraft detection algorithm leveraging dilated convolution and attention mechanisms, creating a novel pyramid dilation network to optimize aircraft feature extraction in SAR images. Wang et al. [30], utilizing the SSD object detection framework, applied a strategy integrating transfer learning and data augmentation to enhance SSD's target detection performance in SAR images. In SAR aircraft target detection, scholars commonly modify existing algorithms to meet specific requirements for satisfactory results [31,32]. However, achieving a balance between model complexity and detection accuracy can be challenging, and existing algorithms may struggle in such scenarios. Moreover, aircraft in SAR images may experience deformation due to radar geometry effects, altering the target shape and complicating detection. Additionally, deep learning methods rely on ample samples for supervised training, and inadequate samples can result in overfitting, adversely affecting detection performance.

To address these challenges, we propose a Feature Enhancement and Multi-Scales Fusion Network (FEMSFNet). This network aims to improve detection accuracy while minimizing model complexity, achieving a balance between the two. Firstly, FEMSFNet employs a diverse image enhancement technique [33–35], applying methods like noise, mosaic, mixup, rotation, and cropping to enhance image features. This addresses the scarcity of SAR aircraft image data, enhancing sample diversity for improved generalization and robust network training. Secondly, drawing inspiration from Yolov4-tiny [36–38], FEMSFNet utilizes the CSPDarknet53-tiny network [39] as the backbone to create a lightweight model. It incorporates a residual module based on an improved attention mechanism, focusing more on critical image regions for enhanced recognition accuracy. Thirdly, the paper introduces superior CSP [40] structures based on an improved feature pyramid [41], preventing a reduction in the network's receptive field and the loss of target feature information in deep structures. Finally, tailored for SAR aircraft target detection, FEMSFNet optimizes loss functions [42,43] while implementing cosine annealing learning rate decay [44], and includes label smoothing [45] techniques to prevent overfitting, expedite convergence, and improve regression accuracy.

The main contributions of our work are as follows:

- FEMSFNet, as proposed, prioritizes both speed and accuracy in target detection. It leverages the lightweight CSPDarknet53-tiny as the backbone for efficient feature extraction. The model's performance has undergone evaluation using the SAR Aircraft Detection Dataset (SADD).

- To maintain a lightweight model without compromising detection accuracy, we integrate the optimized Squeeze-and-Excitation Networks (SE) attention module with the ResNet module in the backbone, forming the SdE-Resblock structure.
- To prevent the deep network from causing a deviation in the receptive field during training, leading to ineffective global feature fusion and loss of feature information, we propose a CSP structure based on an improved pyramid pooling model, called ssppf-CSP.
- Considering the unique characteristics of SAR aircraft target detection, FEMSFNet optimized the network's loss functions while implementing techniques such as learning rate cosine annealing decay and label smoothing to prevent overfitting, ultimately enhancing convergence speed and regression accuracy.

The rest of the paper is arranged as follows. Section 2 describes the proposed aircraft detection network in detail. Experimental results, as well as performance evaluation, are presented in Section 3, and a detailed discussion of the results is provided at the end of this section. Section 4 briefly summarizes this paper.

2. Materials and Methods

2.1. Overview of FEMSFNet

Yolov4-tiny, a lightweight variant of the YOLO (You Only Look Once) object detection series, is tailored for real-time object detection on devices with constrained computational resources. In contrast to Yolov4 [36], Yolov4-tiny boasts a smaller model size and reduced computational complexity. With outstanding performance in natural settings, we adopt Yolov4-tiny as the baseline for our work. The overall structure of FEMSFNet is illustrated in Figure 1.

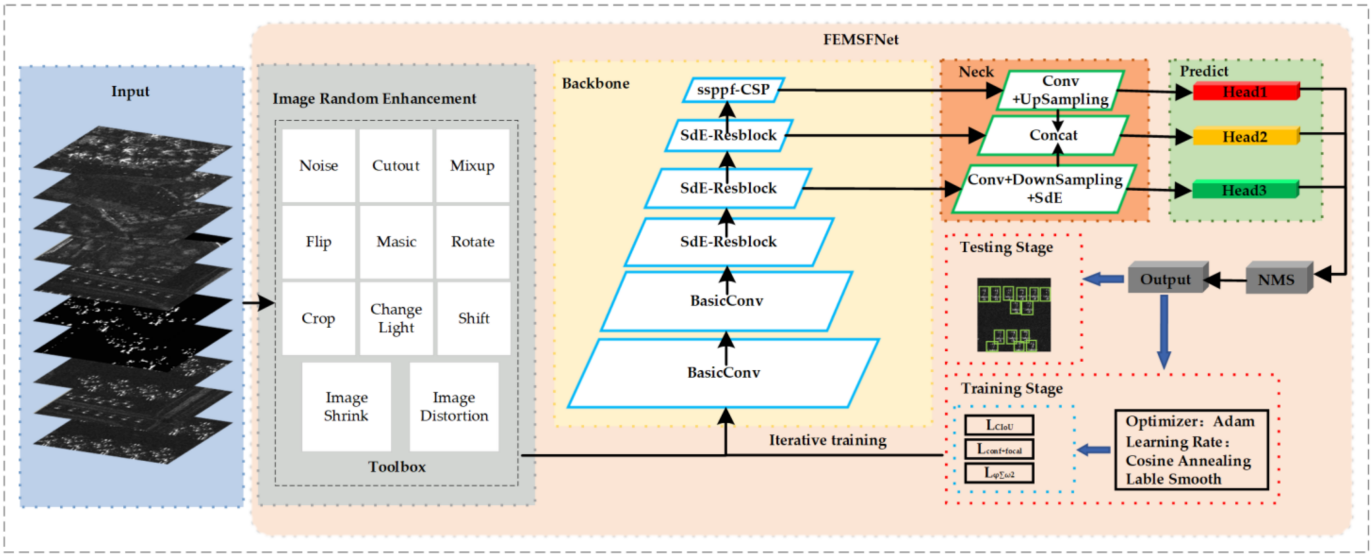
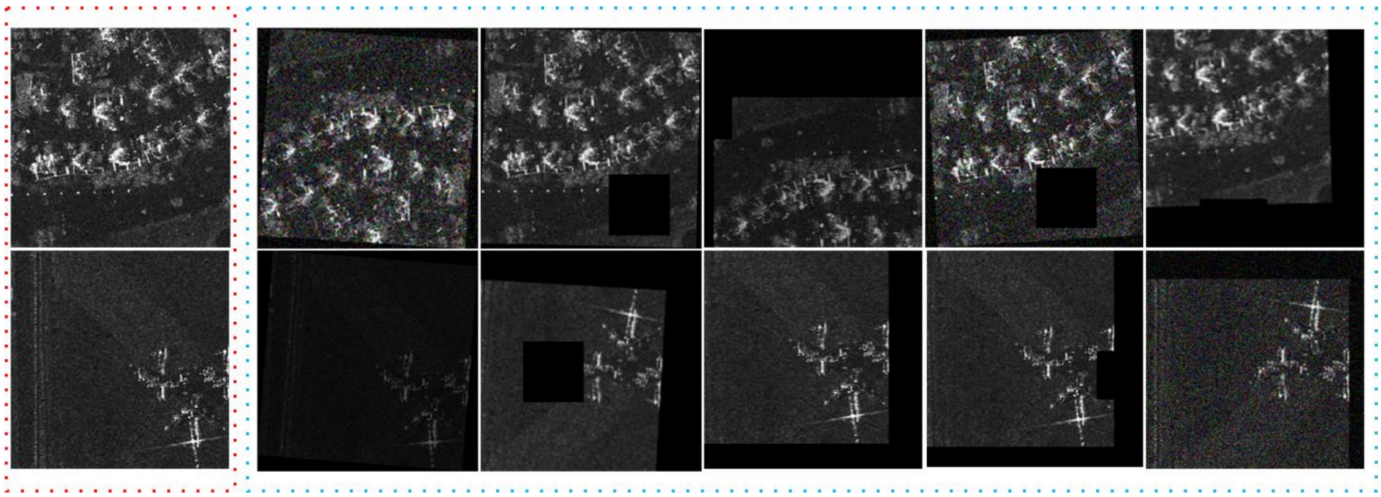


Figure 1. The overall structure of FEMSFNet.

FEMSFNet comprises five key modules: Input, Image Enhancement, Backbone, Neck, and Prediction. Recognizing the pivotal role of sample quality in network training effectiveness, the Image Enhancement module employs various methods, including mosaic, mixup, rotation, scaling, and cropping, with a certain probability. These diverse image enhancement techniques augment the dataset, addressing limitations in quantity and diversity in SAR aircraft image datasets. The results of image enhancement are illustrated in Figure 2. Drawing inspiration from Yolov4-tiny, we select the lightweight CSPDarknet53-tiny as the Backbone feature extraction network for FEMSFNet to strike a balance between detection accuracy and efficiency. The Backbone network integrates the Basic-Conv convolutional module, SdE-Resblock residual module, and ssppf-CSP deep feature fusion module. The

Basic-Conv convolutional module consists of a convolution block, batch normalization (BN) [46] block, and the SiLU [47] activation function block.



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

Figure 2. The results of images enhancement. (**a**) The saw images; (**b**) the enhanced images.

Following image enhancement and feature extraction by the Backbone network, FEMSFNet channels feature maps of different scales to the Neck layer. This layer incorporates a feature pyramid structure with an attention mechanism, fusing feature maps from the Backbone at three scales. During feature down-sampling, an attention mechanism prioritizes the target region for high-resolution, semantically rich features. The resulting feature maps are then input into the Prediction module for inference. Subsequently, predictions are compared with ground truth labels, and loss is computed. The error undergoes backpropagation through the network, and the Adam optimizer [48] with a cosine annealing learning schedule adjusts weights and parameters iteratively until the end of the training loop. To address the nuances of SAR image aircraft single-object detection and prevent issues like model overfitting, we utilize an optimized loss function in the calculation. The subsequent section provides a detailed explanation of our methods.

2.2. SdE-Resblock

Residual Network [49,50] (ResNet) is a deep learning model architecture specifically crafted to mitigate challenges like vanishing gradients and exploding gradients encountered in the training of deep neural networks. ResNet facilitates the construction of exceptionally deep networks without experiencing performance degradation. The fundamental concept behind ResNet is the incorporation of residual blocks, allowing the input signal to be directly forwarded to the block's output via skip connections. This mechanism, termed residual learning, enables the model to learn residuals rather than mapping directly. This approach simplifies the training of deep networks and enhances their effectiveness. The success of ResNet has served as inspiration for the design of numerous subsequent deep learning architectures, establishing it as a classic example in the construction of deep neural networks [51]. The residual structure in the Yolov4-tiny network is depicted in Figure 3. In the illustrated residual network, there are four DarkNet blocks and one MaxPooling block. The DarkNet module consists of fundamental convolutional layers, Batch Normalization (BN) layers, and ReLU activation layers. It serves as one of the essential building blocks of the entire FEMSFNet architecture. Here, the convolutional layers are tasked with extracting features from images. The BN layers address the issue of increasing training difficulty and slower convergence as the depth of the neural network grows. The ReLU activation layers map the input from neurons to the output, introducing sparsity within the network by rendering outputs of some neurons to zero. This sparsity reduces the interdependency

of parameters, thereby mitigating the risk of overfitting. The input data undergo the first convolutional block, producing Map1, which is then split into two parts. The second half passes through the second convolutional block, resulting in Map2. Map2 undergoes further computations in the third convolutional block, generating results. The obtained result is concatenated with Map2 and input into the fourth convolutional block, producing Map3. Finally, Map1 and Map3 are concatenated, and the result undergoes max-pooling to obtain the final output.

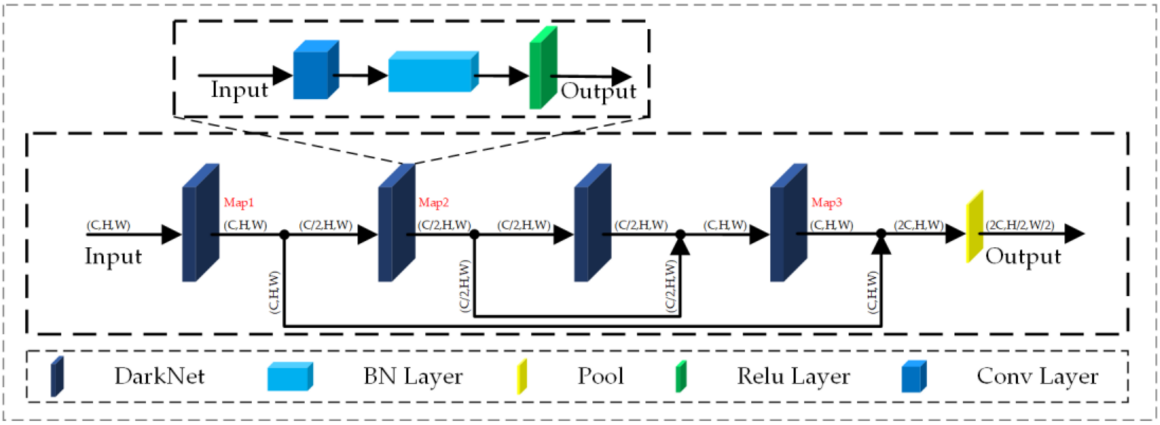


Figure 3. The residual structure in the Yolov4-tiny network.

Attention mechanisms, known for selectively focusing on important information, improve model performance on relevant tasks [52,53]. Integrating attention mechanisms into ResNet enhances the network's attention to crucial features [54]. Hence, we devised the SdE-Resblock structure by combining the optimized Squeeze-and-Excitation (SE) attention [55] module with the ResNet module in the Backbone. This structure assigns varied weights to different channels of the network's feature maps, prioritizing the target region for enhanced training efficiency. To counter overfitting during the Excitation phase, where nonlinear mapping and adjustment of squeezed features occur, we introduced Dropout [56] techniques for regularization into SE module. The detailed network architecture is depicted in Figure 4.

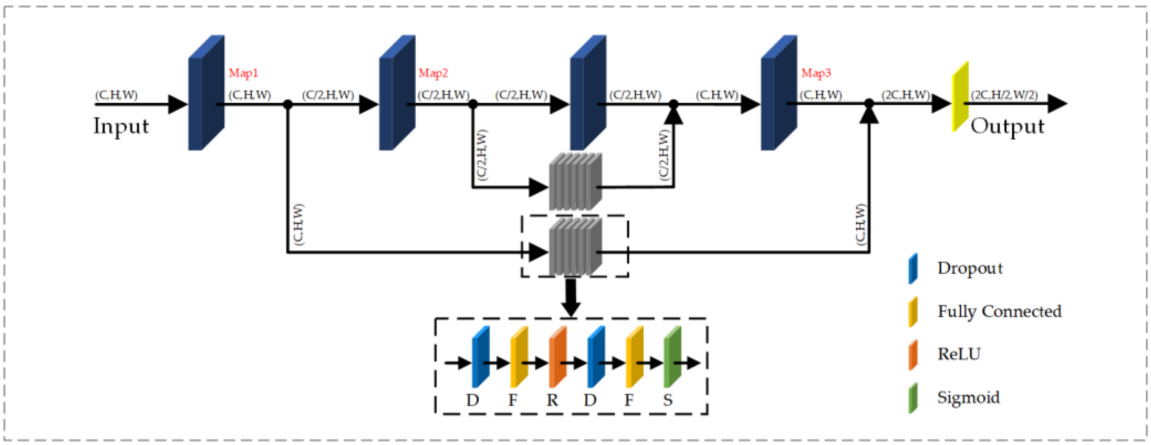


Figure 4. The detailed network architecture of SdE-Resblock.

Compared to the residual structure in Figure 3, the SdE-Resblock introduces an attention mechanism into the residual branch. Without altering the size and number of channels of the feature map, it employs Squeeze-and-Excitation along with regularization operations to assign varying weights to different channels in the residual branch. This enables the network to concentrate more attention on the target region, facilitating easier and more emphasized recognition of target features.

To gain deeper insights into the model's decision-making process, the distribution of the attention mechanism was visualized through heatmaps. As illustrated in Figure 5, the heatmap unveils the areas of focus while the model processes the input data. In this figure, warmer colors (such as red) denote higher degrees of attention by the model to those areas, whereas cooler colors (like blue) indicate lower levels of focus. This visualization method clearly demonstrates the model's tendency to concentrate on specific parts of the input data. Notably, the model significantly zeroes in on the airplane regions within the input image, aligning with expectations, as these areas typically contain crucial information necessary for object recognition. Moreover, the heatmap further reveals the model's ability to effectively ignore background noise or information irrelevant to the task at hand, underscoring the efficacy of the attention mechanism in enhancing the model's focus on vital information and improving overall performance. In summary, the attention mechanism's heatmap not only provides an intuitive view of the model's learning and decision processes but also affirms the model's capability to efficiently identify and utilize key information within the input data for accurate predictions.

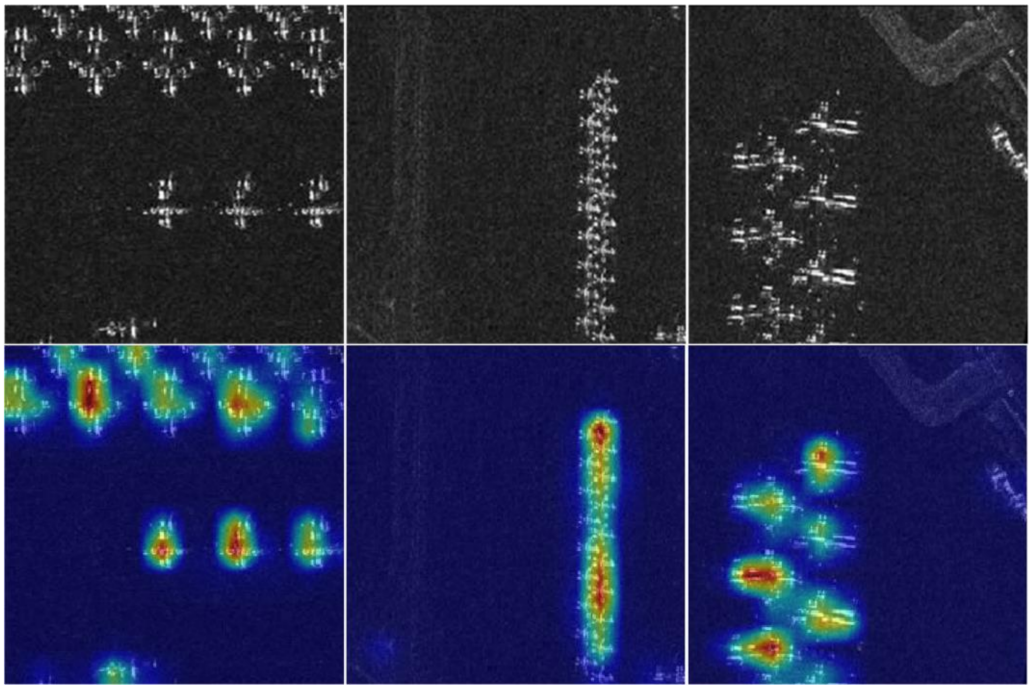


Figure 5. Attention mechanism heatmaps.

2.3. ssppf-CSP

In typical scenarios, deep neural networks often exhibit a bias in the receptive field during training, hampering effective integration of global features and causing loss of feature information. To address this challenge, we propose the ssppf-CSP structure (soft-spatial pyramid pooling-fast-CSPnet), leveraging an enhanced pyramid pooling model. This structure employs multi-level pooling operations to extract and merge local and global features, thereby improving the model's receptive field. The sppf (spatial pyramid pooling-fast) structure [57] achieves pooling effects of large-sized layers through stacking smaller layers, enhancing the network's expressive capability through multi-scale feature fusion. In contrast to traditional sppf modules using max-pooling, we employ soft-pooling in the ssppf module to minimize information loss during pooling, preserving detailed information for detection. Integrated with the final feature extraction module, CSPnet, in FEMSFNet, the ssppf module uses soft-pooling with four window sizes (6×6, 3×3, 2×2, 1×1) to map high-level feature information to low-level features. Skip connections concatenate high-level semantic information with shallow-level information after pooling, overcoming

performance loss in deeper network structures. The detailed architecture of ssppf-CSP is illustrated in Figure 6.

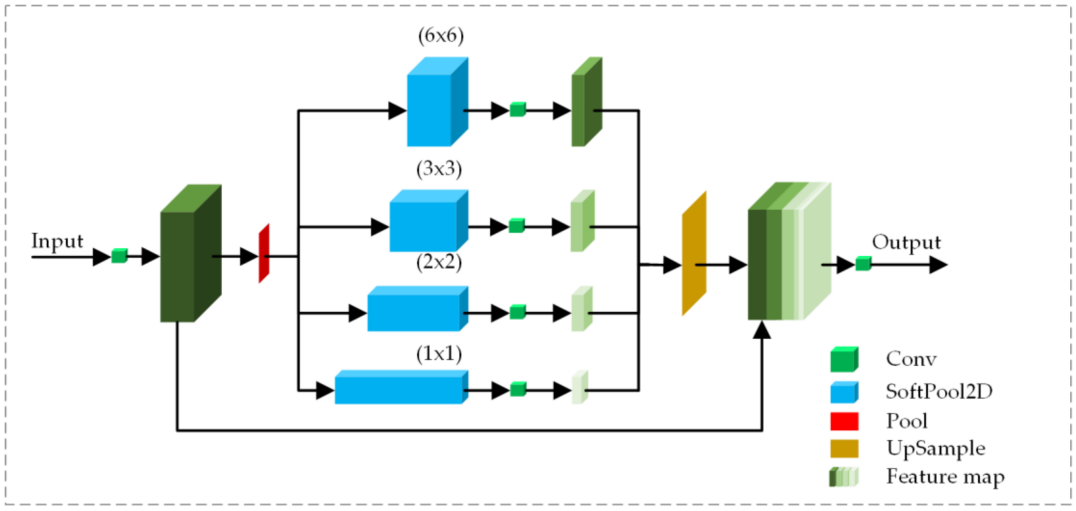


Figure 6. The detailed network architecture of ssppf-CSP.

In the ssppf-CSP structure, the input image undergoes convolutional operations, leading to the division of feature channels into two segments. One part directly forms a larger residual branch through skip connections, becoming an integral part of the final output feature map. The other segment undergoes pooling operations and contributes to a feature pyramid structure comprising four differently sized soft pooling layers. This process maps high-level feature information to low-level feature maps. Finally, it is cascaded and fused with high-level semantic information that has not undergone pooling, preserving global information to the maximum extent. This approach helps avoid performance loss caused by the limited feature representation capability of deep layers in the network.

2.4. Loss Function

The loss function plays a crucial role in deep learning models, quantifying the disparity between predicted values and true labels [58]. Yolov4-tiny's loss function includes bounding box (L_{CIoU}), confidence (L_{conf}), and class components (L_{class}). The detailed process is shown in Equations (1)–(7).

$$L_{\mathrm{loss}} = L_{\mathrm{CIoU}} + L_{\mathrm{conf}} + L_{\mathrm{class}} \tag{1}$$

$$L_{\mathrm{CIoU}} = \sum_{\mathrm{i}=0}^{\mathrm{S}\times\mathrm{S}} \sum_{\mathrm{j}=0}^{\mathrm{M}} I_{\mathrm{ij}}^{\mathrm{obj}} (2 - w_{\mathrm{i}} \times h_{\mathrm{i}})(1 - A_{\mathrm{CIoU}}) \tag{2}$$

$$A_{\mathrm{CIoU}} = 1 - A_{\mathrm{IoU}} + \frac{\rho^2\left(b, \hat{b}\right)}{c^2} + \alpha v \tag{3}$$

$$\alpha = \frac{v}{1 - A_{\mathrm{IoU}} + v} \tag{4}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{\hat{w}}{\hat{h}} - \arctan\frac{w}{h}\right)^2 \tag{5}$$

$$L_{\mathrm{conf}} = -\sum_{\mathrm{i}=0}^{\mathrm{S}\times\mathrm{S}} \sum_{\mathrm{j}=0}^{\mathrm{M}} I_{\mathrm{ij}}^{\mathrm{obj}} \left[\hat{C}_{\mathrm{i}} \log(C_{\mathrm{i}}) + (1 - \hat{C}_{\mathrm{i}}) \log(1 - C_{\mathrm{i}})\right] - \lambda_{\mathrm{noobj}} \sum_{\mathrm{i}=0}^{\mathrm{S}\times\mathrm{S}} \sum_{\mathrm{j}=0}^{\mathrm{M}} I_{\mathrm{ij}}^{\mathrm{noobj}} \left[\hat{C}_{\mathrm{i}} \log(C_{\mathrm{i}}) + (1 - \hat{C}_{\mathrm{i}}) \log(1 - C_{\mathrm{i}})\right] \tag{6}$$

$$L_{\mathrm{class}} = -\sum_{\mathrm{i}=0}^{\mathrm{S}\times\mathrm{S}} \sum_{\mathrm{j}=0}^{\mathrm{M}} I_{\mathrm{ij}}^{\mathrm{obj}} \sum_{\mathrm{c}\in\mathrm{classes}} \left[\hat{p}_{\mathrm{i}}(c) \log(p_{\mathrm{i}}(c)) + (1 - \hat{p}_{\mathrm{i}}(c)) \log(1 - p_{\mathrm{i}}(c))\right] \tag{7}$$

For FEMSFNet, focusing on SAR image aircraft single-object detection with a single category output, the class loss function is omitted for computational simplicity. In object detection, the imbalance between positive and negative samples, where most pixels are background, is addressed by introducing the focal loss function in the confidence component of the loss function. This helps the model handle class imbalance and concentrate on challenging examples. To expedite model convergence and prevent overfitting, L2 regularization [59] is incorporated into the loss function for weight decay of model parameters. Here, φ represents the regularization coefficient. The detailed process is shown in Equations (8)–(11).

$$Loss_{\text{total}} = L_{\text{CIoU}} + L_{\text{conf+focal}} + \varphi \sum_{i=0}^{n} \omega^2 \tag{8}$$

$$L_{\text{conf+focal}} = L_{\text{conf}} + \lambda L_{\text{focal}} 4 \tag{9}$$

$$L_{\text{conf}} = -\sum_{i=0}^{S\times S}\sum_{j=0}^{M} I_{\text{ij}}^{\text{obj}}\left[\hat{C}_{\text{i}}\log(C_{\text{i}}) + \left(1-\hat{C}_{\text{i}}\right)\log(1-C_{\text{i}})\right] - \lambda_{\text{noobj}}\sum_{i=0}^{S\times S}\sum_{j=0}^{M} I_{\text{ij}}^{\text{noobj}}\left[\hat{C}_{\text{i}} log(C_{\text{i}}) + \left(1-\hat{C}_{\text{i}}\right) log(1-C_{\text{i}})\right] \tag{10}$$

$$L_{\text{focal}} = -(1-\rho_{\text{t}})^{\gamma}\log(\rho_{\text{t}}) 4 \tag{11}$$

where $I_{\text{ij}}^{\text{obj}}$ is used to determine whether there is a target object in the network—if yes, it is set to 1; otherwise, it is set to 0; A_{IoU} represents the intersection over union ratio of the area between the true box and the predicted box; $\rho^2\left(b, \hat{b}\right)$ is the Euclidean distance between the center points of the predicted and true boxes; c is the diagonal distance of the minimum closed region between the true box and the predicted box; $\hat{w}$, $\hat{h}$ and w, h represent the width and height of the true and predicted boxes, respectively; $\hat{C}_{\text{i}}$ and C_{i} are the confidences of the true and predicted samples; $I_{\text{ij}}^{\text{noobj}}$ is the inverse of $I_{\text{ij}}^{\text{obj}}$—it is set to 0 if there is an object in the grid, and 1 if there is no object; λ is an adjustment parameter used to balance the importance of two losses; ρ_{t} is the model's probability of a sample being positive; γ is the parameter controlling attention.

3. Experiments and Results

3.1. Datasets

Currently, there is a scarcity of publicly available datasets for SAR image aircraft detection. Consequently, we exclusively utilize the SAR Aircraft Detection Dataset (SADD) [60] to validate our approach.

SADD is derived from the German Terra-SAR-X satellite, operating in the x-band and HH polarization mode. It provides image resolutions ranging from 0.5 to 3 m. Expert SAR Automatic Target Recognition (ATR) analysts manually annotate the ground truth of aircraft based on prior knowledge and corresponding optical images. After cropping large images, the SADD comprises 2966 non-overlapping 224 × 224 slices, containing 7835 annotated aircraft targets with clear structures, outlines, and main components. The dataset includes aircraft targets of varying sizes, with a significant number being small-scale targets. The SADD backdrop features a complex environment with diverse scenes such as airport runways, aprons, and civil aviation facilities. Negative samples are predominantly found in areas surrounding the airport, including open spaces and forests. Refer to Figure 7 for visual representations of sample images within the SADD.

In this article, to validate our method and enable comparison with other relevant papers using the same dataset, we randomly divide the SADD images into the training and test sets at a 5:1 ratio [60]. The training set comprises 799 positive samples and 1673 negative samples, totaling 6948 annotated aircraft boxes. The test set includes 85 positive samples, 409 negative samples, and a total of 887 annotated aircraft boxes, as summarized in Table 1.

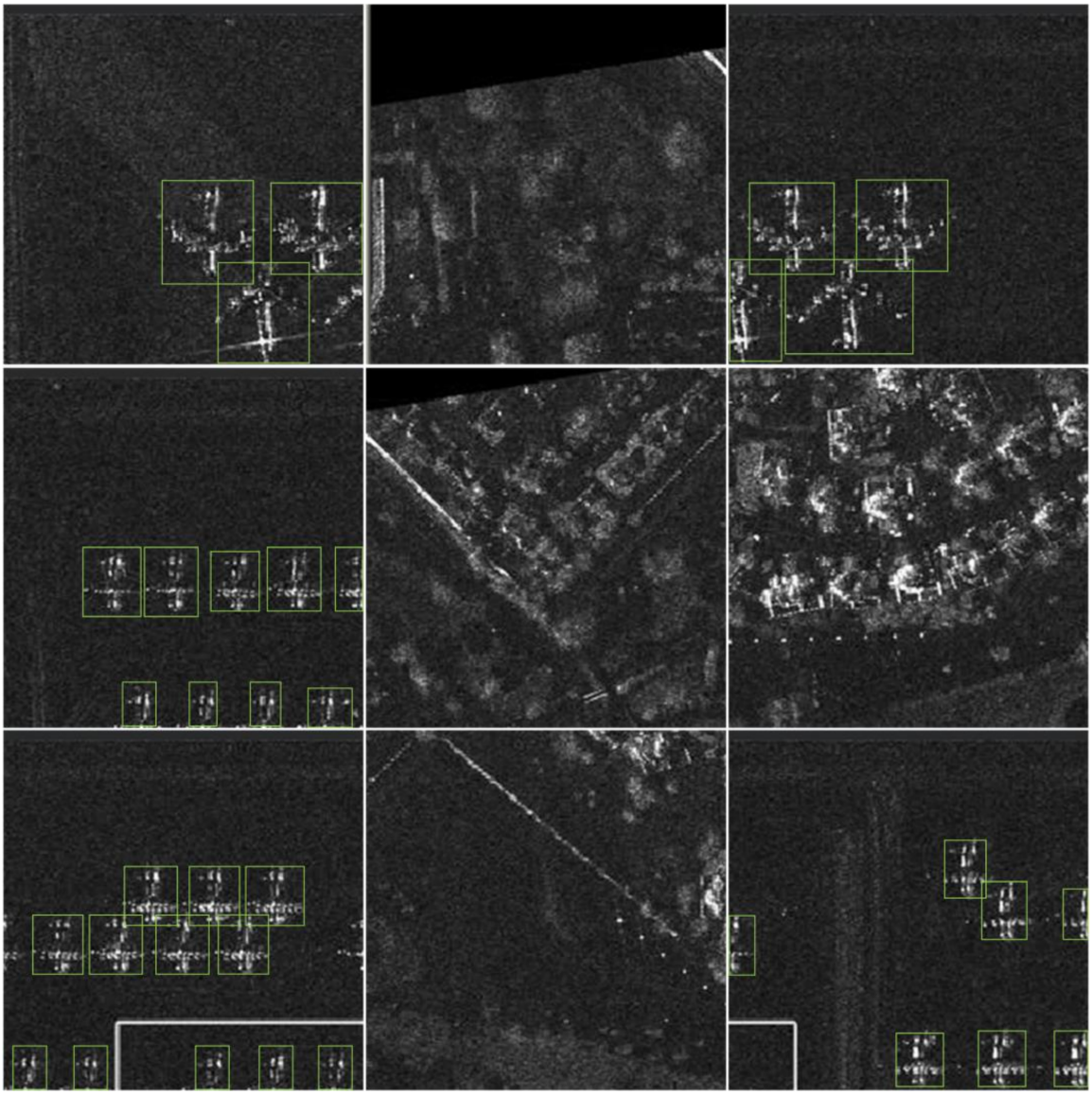
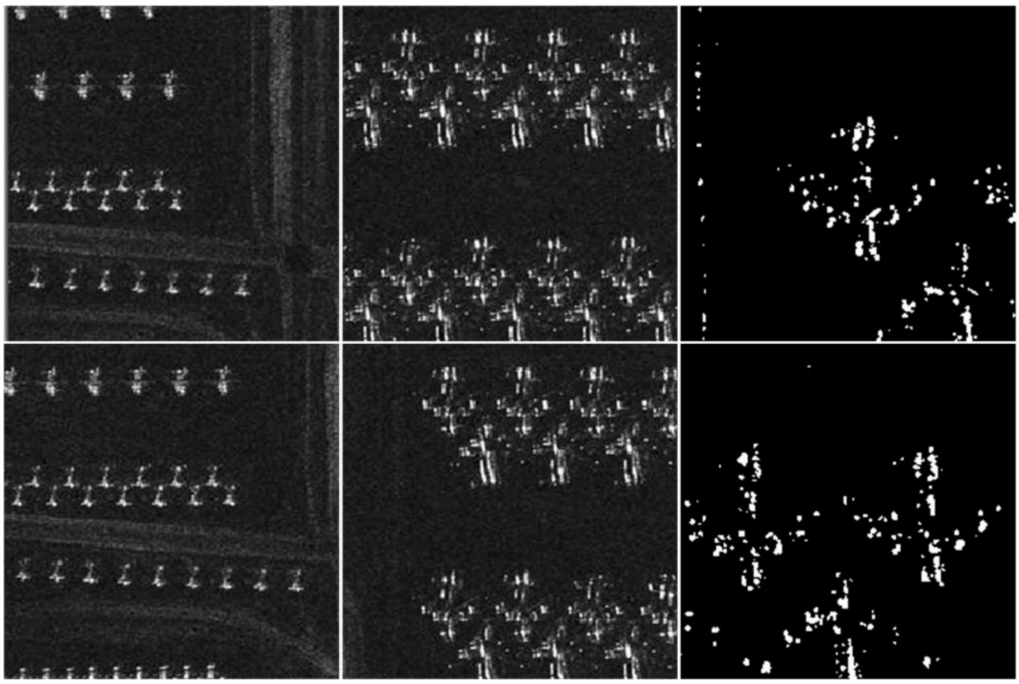


Figure 7. Visual representations of sample images within the SADD. The green boxes are detected aircrafts.

Table 1. Division of the dataset.

Dataset	Positive Samples	Negative Samples	Ground Truth
Train	799	1673	6974
Test	85	409	861

Figure 8 showcases sample images from the SADD. The images exhibit notable variations in the size of aircraft targets, and the backgrounds surrounding specific aircraft targets are particularly intricate, presenting challenges for accurate aircraft positioning. Furthermore, the intricate background clutter points may be misleadingly identified as aircraft components, further complicating the detection process.



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

Figure 8. The variational size of aircraft targets within SADD. (**a**) The size of aircraft targets is less than 20 pixels. (**b**) The size of aircraft targets is between 20 and 40 pixels. (**c**) The size of aircraft targets is more than 40 pixels.

3.2. Evaluation Metrics

In SAR aircraft detection, precision rate and recall rate serve as common evaluation criteria. Yet, there is often a trade-off between precision and recall, meaning enhancing one may reduce the other. To mitigate this, we introduce the F1 score as a supplementary metric, providing a holistic indicator that balances accuracy and recall. The calculation formulas are detailed in Equations (12)–(14):

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{14}$$

where the meaning of TP, FN, and FP are as shown in Table 2.

Table 2. The meaning of TP, FN, and FP.

Label \ Prediction	Positive	Negative
Positive	TP	FN
Negative	FP	TN

3.3. Experimental Setup

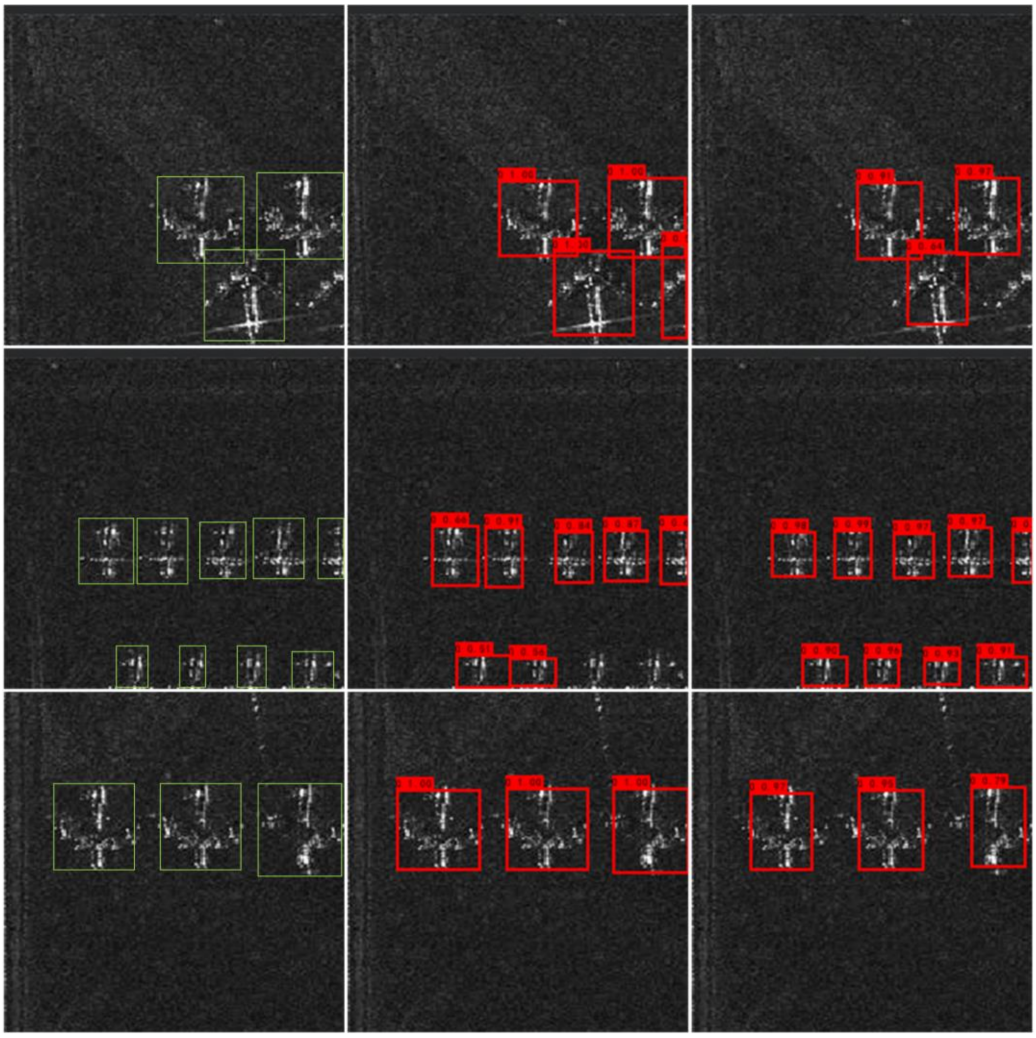
FEMSFNet utilizes a 224 × 224 image resolution for both training and testing phases. During the training stage, random rotation, random mixup, and mosaic data-enhancement methods are employed. All these methods are validated on an NVIDIA 4060Ti GPU. The configuration details of experimental parameters are outlined in Table 3.

Table 3. Hyperparameter settings during model training.

Hyperparameters	Value
Optimizer	Adam
Learning rate	0.001
Learning decay method	Cosine Annealing
Batch size	128
Momentum	0.937
Worker number	4

3.4. Comparison to Existing Algorithms

To showcase the efficacy of FEMSFNet, we visualized its detection confidence maps in comparison to the baseline Yolov4-tiny, depicted in Figure 9. FEMSFNet demonstrates superior accuracy in locating aircraft targets and effectively suppressing background clutter interference, resulting in agiler detection confidence compared to the baseline Yolov4-tiny.



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

Figure 9. Detection confidence visualization. (**a**) Ground truth of SADD; (**b**) Baseline detection confidence map; (**c**) FEMSFNet detection confidence map. The green box indicates the ground truth, while the red box shows the predicted result.

To rigorously validate our algorithm, we compare it against two two-stage methods (Faster R-CNN [61] and Cascade R-CNN [62]) and four one-stage methods (SSD [63], Yolov3 [64], SEFEPNet [60], and Yolov4-tiny [36]), as outlined in Table 4. To specify,

the model complexities are as follows: Faster R-CNN at 160MB, Cascade R-CNN at 319 MB, SSD at 96 MB, Yolov3 at 236 MB, Yolov4-tiny at 19 MB, and FEMSFNet at 23 MB. The results demonstrate that FEMSFNet excels across various metrics, particularly in Precision and Model Size, with a maximum improvement of 92.7% in model size (compared to cascade R-CNN), 12.2% in precision, 12.9% in recall, and 13.3% in F1 score compared to the contrasted algorithms. This success is attributed to enriching dataset diversity through image enhancement, incorporating attention modules in the residual structure, and integrating pyramid modules in the CSP network within the FEMSFNet architecture. These components enable FEMSFNet to capture deeper, broader, and more accurate target information. Figure 10 illustrates the detection results of FEMSFNet and other methods. In support of further research in Synthetic Aperture Radar (SAR) aircraft target detection, we plan to open-source FEMSFNet soon at https://github.com/WenboEth/Sar-Aircaft-Target-Detection, accessed on 10 May 2024.

Table 4. Comparisons to the state-of-the-art models.

Indicator / Model	P	R	F1	Model Size (MB)
faster R-CNN	0.86	0.89	0.87	160
cascade R-CNN	0.90	0.95	0.92	319
SSD	0.84	0.89	0.86	96
Yolov3	0.83	0.97	0.89	236
Yolov4-tiny	0.82	0.85	0.83	19
SEFEPNet [1]	0.89	0.98	0.93	×
FEMSFNet (ours)	0.92	0.96	0.94	23

[1] The data come from the original article [60], and there are no data about the SEFEPNet's model size.

The results below highlight that while FEMSFNet may make mistakes in complex target recognition scenarios, it consistently outperforms the compared algorithms, especially in terms of precision, recall, and model size. We anticipate that FEMSFNet has untapped potential, and with further optimization, broader expansion, and deeper evolution, it can achieve even better results in the future.

In the realm of deep learning-based object detection, assessing a network's generalization ability is crucial. A network model with robust generalization performance excels on unseen, complex, or even distorted data, serving as a key metric of the model's practical utility. The FEMSFNet network enhances its training data through methods including rotation, scaling, and cropping, aiding the model in learning a broader range of variations to boost generalization performance. Moreover, techniques like dropout are utilized within the SdE-Resblock module, randomly "dropping" a portion of neurons during training to decrease the model's dependency on specific data and enhance its generalization capability. To validate the generalization performance of the FEMSFNet network, a non-learned dataset (n-LD) was created. This dataset consists of targets the network has never "learned" from, not included in training, validation, or test sets. To challenge the network's generalization and robustness, the n-LD was complicated and distorted, increasing the difficulty of target recognition. Examples of the n-LD cases are shown in Figure 11. The n-LD was then fed into both the FEMSFNet and state-of-the-art models for comparative testing and validation of FEMSFNet's generalization ability, with results presented in Table 5. The charts reveal that FEMSFNet performs exceptionally well across various metrics, including accuracy, error rate, and omission rate. Such outcomes are attributable to the superior SdE-Resblock module and the ssppf-CSP structure of FEMSFNet.

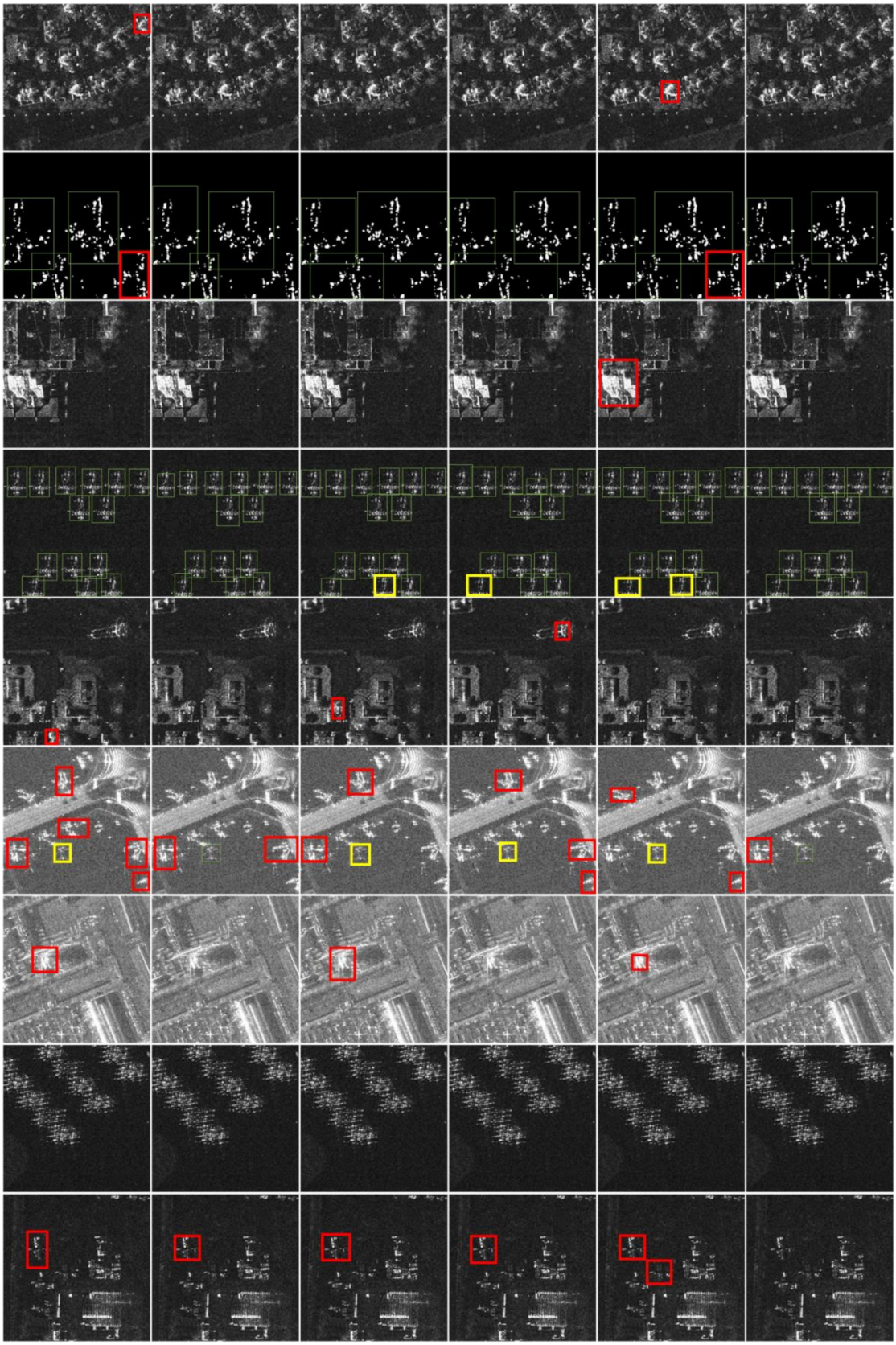


Figure 10. Visualization of SAR aircraft detection results of different algorithms. Yellow boxes indicate false alarms, red boxes represent missing targets, and green boxes denote correct detections.

Table 5. Generalization performance comparison.

Indicator \ Model	Faster R-CNN	Cascade R-CNN	SSD	Yolov3	Yolov4-Tiny	FEMSFNet
Accuracy Rate	0.91	0.93	0.89	0.88	0.85	0.94
Error Rate	0.04	0.04	0.06	0.04	0.06	0.02
Omission Rate	0.05	0.03	0.05	0.08	0.08	0.04

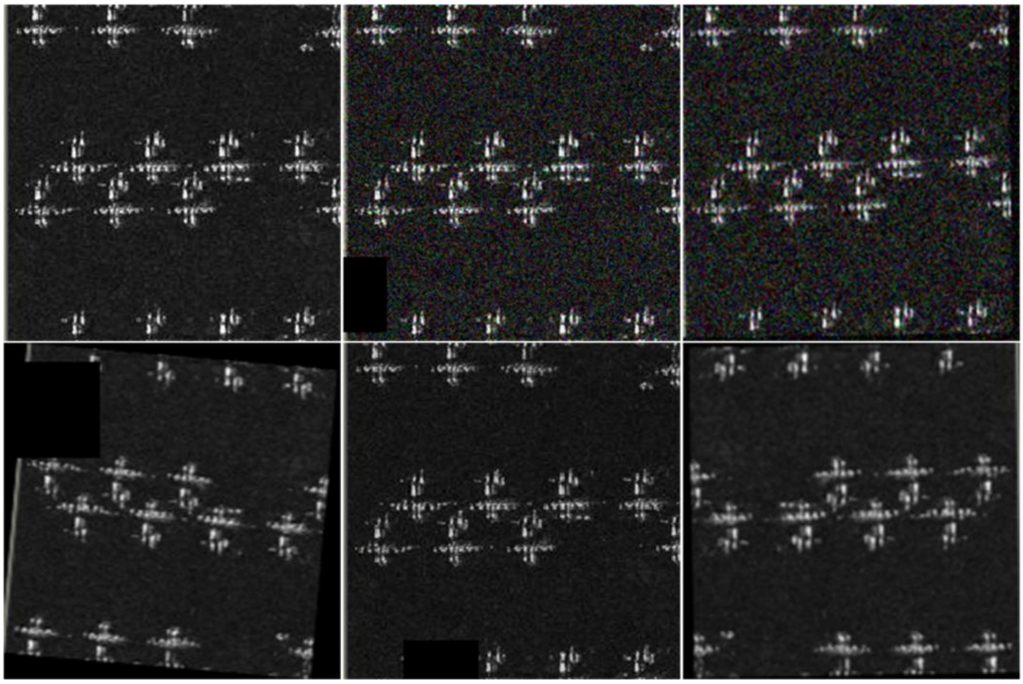


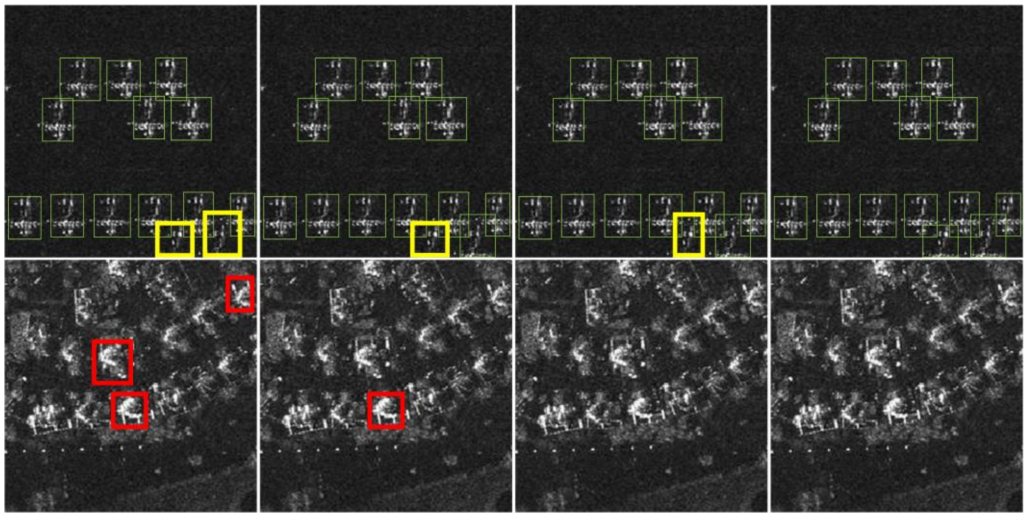
Figure 11. Examples of the n-LD cases.

3.5. Ablation Study

To enhance readers' understanding of each module's impact, we conducted an ablation study. Yolov4-tiny serves as the base framework for evaluating the effects of various feature modules on aircraft detection in SAR images, as outlined in Table 6. In the ablation comparison experiments, the "×" mark in the loss function column indicates the use of the pre-improvement Yolov4-tiny loss function, detailed in Equations (1)–(7). The "√" mark indicates the use of the improved loss function, as detailed in Equations (8)–(11). Simultaneously, Figure 12 presents the results of the ablation study conducted on the SADD dataset. The charts illustrate sequential application of different modules leading to improvements across various metrics. Precision sees the highest enhancement, with a 7.0% increase, while recall and F1 score exhibit improvements of 6.9% and 6.8%, respectively. It highlights that the ssppf-CSP module, integrating a soft feature pyramid into the CSP architecture, enhances network depth and incorporates features from multiple scales. This results in a substantial improvement in overall accuracy and recall metrics. The SdE-Resblock module, integrating an optimized attention mechanism into the residual structure, enhances the network's focus on target detection, leading to improvements in recognition rate and recall. While the improved loss function module may not show a significant increase in evaluation metrics, its main impact lies in accelerating convergence and reducing computational complexity. As illustrated in Figure 13, we have compared the iteration process of the loss function before and after improvements. It is evident from the figure that the improved loss function demonstrates an overall reduction in loss values compared to its prior state, along with a notably faster convergence rate. Furthermore, when comparing the same iteration round, there is a maximum enhancement of 26% in the loss value.

Table 6. Ablation study.

	ssppf-CSP	SdE-Resblock	Loss Function	P	R	F1
(a)	×	×	×	0.86	0.90	0.88
(b)	√	×	×	0.90	0.93	0.91
(c)	√	√	×	0.92	0.95	0.93
(d)	√	√	√	0.92	0.96	0.94



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

Figure 12. Ablation study results presentation. Yellow boxes indicate false alarms, red boxes represent missing targets, and green boxes denote correct detections. (**a**) None; (**b**):ssppf-CSP; (**c**) ssppf-CSP and SdE-Resblock; (**d**) ssppf-CSP, SdE-Resblock, and loss function.

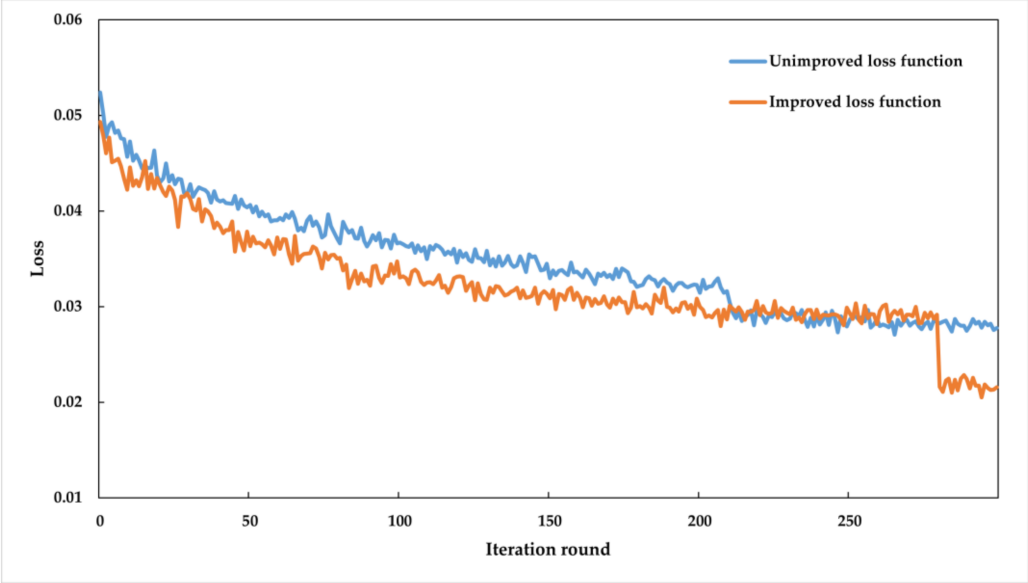


Figure 13. Comparison of accelerated iteration of loss function before and after optimization.

3.6. Result and Discussion

From the comparative validation analysis presented earlier, it is clear that in the domain of aircraft target detection in remote sensing SAR imagery, the FEMSFNet model excels across several metrics, including detection precision, recall rate, F1 score, and model size, achieving varying degrees of improvement. Specifically, it surpasses other algorithms, with a maximum increase of 12.2% in precision, 12.9% in recall, and 13.3% in

F1 score. Moreover, in terms of model size, FEMSFNet achieves an 88% reduction compared to the largest model size benchmarked. The primary reason for this performance is that, unlike general-purpose state-of-the-art neural networks, FEMSFNet is a unique network specifically tailored for SAR imagery of aircraft targets. It eliminates unnecessary components for this task, such as the Region Proposal Network and multiple cascading detection heads, refining the network architecture. Initially, to address the challenges of limited remote sensing SAR image data, complex backgrounds due to noise, and various obstructions making target detection difficult, FEMSFNet incorporates a data augmentation module. This module complexifies or even distorts the source image data, tackling the issue of data scarcity while training the network to enhance its generalization capability and robustness with complicated images. Furthermore, addressing the diversity in target scales and complex angles in SAR imagery, FEMSFNet introduces the ssppf-CSP and SdE-Resblock modules. Optimized residual modules are employed to extract more feature information, and multi-scale fusion modules ensure the detection and recognition of targets across various scales, thereby improving the network's accuracy. Lastly, in the specialized task of single-object detection for aircraft in SAR images, focusing on a singular category output, the class loss function is omitted to simplify computations. This approach addresses the imbalance between positive and negative samples—mostly background—by incorporating the focal loss function into the confidence measure of loss calculation. This modification significantly enhances the model's capability to manage class imbalances and prioritize complex examples. Moreover, to ensure faster convergence and prevent overfitting, L2 regularization is integrated into the loss function, facilitating the weight decay of model parameters.

4. Conclusions

This paper introduces FEMSFNet, an SAR aircraft detection model prioritizing both speed and accuracy. FEMSFNet utilizes the lightweight CSPDarknet53-tiny as its backbone for efficient feature extraction. To maintain a lightweight yet accurate model, we integrate the optimized SE attention module with the ResNet module, forming the SdE-Resblock structure. A novel CSP structure, ssppf-CSP, prevents deviations in the receptive field during training, enhancing global feature fusion. Addressing unique characteristics in SAR aircraft target detection, FEMSFNet optimizes loss functions and employs techniques like learning rate cosine annealing decay and label smoothing to prevent overfitting, improving convergence speed and regression accuracy. For increased sample diversity, FEMSFNet employs multi-faceted image augmentation with techniques like noise addition, mosaic, mixup, rotation, and cropping, enhancing training generalization and robustness. Experiments on the SADD dataset demonstrate FEMSFNet's effectiveness, surpassing state-of-the-art object-detection algorithms. FEMSFNet exhibits significant improvements compared to the contrasted algorithms in terms of precision rate (92%), recall rate (96%), and F1 score (94%). Notably, it surpasses contrasted algorithms with a maximum increase of 12.2% in precision, 12.9% in recall, and 13.3% in F1 score. Anticipating untapped potential, further optimization, broader expansion, and deeper evolution are expected to propel FEMSFNet to even better results in the future.

Author Contributions: Methodology, W.Z.; software, W.Z.; validation, L.Z., X.L., and W.Z.; formal analysis, G.F.; investigation, Y.S.; resources, J.S.; data curation, W.Z.; writing—original draft preparation, W.Z.; writing—review and editing, W.Z.; visualization, C.L. and W.Z.; supervision, W.Z.; project administration, L.Z.; funding acquisition, L.Z. 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 (General Program, No. 62073150, 62227812), the 173 Key Projects of Basic Research (2021-JCJO-ZD-025-11), and a pre-research project (6B2B5347).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: We sincerely thank the editor and reviewers for their constructive comments.

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

References

1. Chan, Y.K.; Koo, V.J. An introduction to synthetic aperture radar (SAR). *Prog. Electromagn. Res. B* **2008**, *2*, 27–60. [CrossRef]
2. Franceschetti, G.; Migliaccio, M.; Riccio, D. The SAR simulation: An overview. In Proceedings of the 1995 International Geoscience and Remote Sensing Symposium, IGARSS'95. Quantitative Remote Sensing for Science and Applications, Firenze, Italy, 10–14 July 1995; pp. 2283–2285.
3. Gens, R.; Van Genderen, J.L. Review Article SAR interferometry—Issues, techniques, applications. *Int. J. Remote Sens.* **1996**, *17*, 1803–1835. [CrossRef]
4. Qian, G.; Haipeng, W.; Feng, X.J. Research progress on aircraft detection and recognition in SAR imagery. *J. Radars* **2020**, *9*, 497–513.
5. Fuentes Reyes, M.; Auer, S.; Merkle, N.; Henry, C.; Schmitt, M.J. Sar-to-optical image translation based on conditional generative adversarial networks—Optimization, opportunities and limits. *Remote Sens.* **2019**, *11*, 2067. [CrossRef]
6. Singh, P.; Diwakar, M.; Shankar, A.; Shree, R.; Kumar, M. A Review on SAR Image and its Despeckling. *Arch. Comput. Methods Eng.* **2021**, *28*, 4633–4653. [CrossRef]
7. Gao, G. Statistical modeling of SAR images: A survey. *Sensors* **2010**, *10*, 775–795. [CrossRef] [PubMed]
8. Sansosti, E.; Berardino, P.; Manunta, M.; Serafino, F.; Fornaro, G. Geometrical SAR image registration. *Geosci. Remote Sens.* **2006**, *44*, 2861–2870. [CrossRef]
9. Lattari, F.; Gonzalez Leon, B.; Asaro, F.; Rucci, A.; Prati, C.; Matteucci, M. Deep learning for SAR image despeckling. *Remote Sens.* **2019**, *11*, 1532. [CrossRef]
10. Wang, X.; Chen, C. Ship detection for complex background SAR images based on a multiscale variance weighted image entropy method. *IEEE Geosci. Remote Sens. Lett.* **2016**, *14*, 184–187. [CrossRef]
11. Dong, G.; Wang, N.; Kuang, G. Sparse representation of monogenic signal: With application to target recognition in SAR images. *IEEE Signal Process. Lett.* **2014**, *21*, 952–956.
12. Gao, G.; Ouyang, K.; Luo, Y.; Liang, S.; Zhou, S. Scheme of parameter estimation for generalized gamma distribution and its application to ship detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *55*, 1812–1832. [CrossRef]
13. Leng, X.; Ji, K.; Zhou, S.; Xing, X. Ship detection based on complex signal kurtosis in single-channel SAR imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6447–6461. [CrossRef]
14. Steenson, B.O. Detection performance of a mean-level threshold. *IEEE Trans. Aerosp. Electron. Syst.* **1968**, *AES-4*, 529–534. [CrossRef]
15. Hm, F. Adaptive detection mode with threshold control as a function of spatially sampled clutter-level estimates. *RCA Rev.* **1968**, *29*, 414–465.
16. Hansen, V.G. Constant false alarm rate processing in search radars. In Proceedings of the IEEE Conference Publication no. 105 "Radar-Present and Future", London, UK, 23–25 October 1973; pp. 325–332.
17. Trunk, G.V. Range resolution of targets using automatic detectors. *IEEE Trans. Aerosp. Electron. Syst.* **1978**, *AES-14*, 750–755.
18. Kuttikkad, S.; Chellappa, R. Non-Gaussian CFAR techniques for target detection in high resolution SAR images. In Proceedings of the 1st International Conference on Image Processing, Austin, TX, USA, 13–16 November 1994; pp. 910–914.
19. Smith, M.E.; Varshney, P.K. VI-CFAR: A novel CFAR algorithm based on data variability. In Proceedings of the 1997 IEEE National Radar Conference, Syracuse, NY, USA, 13–15 May 1997; pp. 263–268.
20. Ai, J.; Mao, Y.; Luo, Q.; Xing, M.; Jiang, K.; Jia, L.; Yang, X. Robust CFAR ship detector based on bilateral-trimmed-statistics of complex ocean scenes in SAR imagery: A closed-form solution. *IEEE Trans. Aerosp. Electron. Syst.* **2021**, *57*, 1872–1890. [CrossRef]
21. Chen, S.; Li, X. A new CFAR algorithm based on variable window for ship target detection in SAR images. *Signal Image Video Process.* **2019**, *13*, 779–786. [CrossRef]
22. Liang, F.; Zhou, Y.; Chen, X.; Liu, F.; Zhang, C.; Wu, X. Review of target detection technology based on deep learning. In Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence, Sanya, China, 14–16 January 2021; pp. 132–135.
23. Khan, M.J.; Yousaf, A.; Javed, N.; Nadeem, S.; Khurshid, K. Automatic target detection in satellite images using deep learning. *J. Space Technol.* **2017**, *7*, 44–49.
24. Liu, Y.; Jiang, D.; Xu, C.; Sun, Y.; Jiang, G.; Tao, B.; Tong, X.; Xu, M.; Li, G.; Yun, J. Deep learning based 3D target detection for indoor scenes. *Appl. Intell.* **2023**, *53*, 10218–10231. [CrossRef]
25. Wang, J.; Liu, C.; Fu, T.; Zheng, L. Research on automatic target detection and recognition based on deep learning. *J. Vis. Commun. Image Represent.* **2019**, *60*, 44–50. [CrossRef]
26. Chen, S.; Wang, H. SAR target recognition based on deep learning. In Proceedings of the 2014 International Conference on Data Science and Advanced Analytics (DSAA), Shanghai, China, 30 October–1 November 2014; pp. 541–547.
27. El Housseini, A.; Toumi, A.; Khenchaf, A. Deep Learning for target recognition from SAR images. In Proceedings of the 2017 Seminar on Detection Systems Architectures and Technologies (DAT), Algiers, Algeria, 20–22 February 2017; pp. 1–5.

28. Soldin, R.J. SAR target recognition with deep learning. In Proceedings of the 2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 9–11 October 2018; pp. 1–8.
29. Zhao, Y.; Zhao, L.; Li, C.; Kuang, G. Pyramid attention dilated network for aircraft detection in SAR images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 662–666. [CrossRef]
30. Wang, Z.; Du, L.; Mao, J.; Liu, B.; Yang, D. SAR target detection based on SSD with data augmentation and transfer learning. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 150–154. [CrossRef]
31. Guo, Y.; Du, L.; Lyu, G. SAR target detection based on domain adaptive faster R-CNN with small training data size. *Remote Sens.* **2021**, *13*, 4202. [CrossRef]
32. An, Q.; Pan, Z.; Liu, L.; You, H. DRBox-v2: An improved detector with rotatable boxes for target detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8333–8349. [CrossRef]
33. Singh, G.; Mittal, A. Various image enhancement techniques-a critical review. *Int. J. Innov. Sci. Res.* **2014**, *10*, 267–274.
34. Aghagolzadeh, S.; Ersoy, O.K. Transform image enhancement. *Opt. Eng.* **1992**, *31*, 614–626. [CrossRef]
35. Qi, Y.; Yang, Z.; Sun, W.; Lou, M.; Lian, J.; Zhao, W.; Deng, X.; Ma, Y. A comprehensive overview of image enhancement techniques. *Arch. Comput. Methods Eng.* **2021**, *29*, 583–607. [CrossRef]
36. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
37. Jiang, Z.; Zhao, L.; Li, S.; Jia, Y. Real-time object detection method based on improved YOLOv4-tiny. *arXiv* **2020**, arXiv:2011.04244.
38. Wang, L.; Zhou, K.; Chu, A.; Wang, G.; Wang, L. An improved light-weight traffic sign recognition algorithm based on YOLOv4-tiny. *IEEE Access* **2021**, *9*, 124963–124971. [CrossRef]
39. Mahasin, M.; Dewi, I.A. Comparison of CSPDarkNet53, CSPResNeXt-50, and EfficientNet-B0 Backbones on YOLO V4 as Object Detector. *Int. J. Eng. Sci. Inf. Technol.* **2022**, *2*, 64–72. [CrossRef]
40. Wang, C.-Y.; Liao, H.-Y.M.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W.; Yeh, I.-H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 390–391.
41. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
42. Wang, Q.; Ma, Y.; Zhao, K.; Tian, Y. A comprehensive survey of loss functions in machine learning. *Ann. Data Sci.* **2020**, *9*, 187–212. [CrossRef]
43. Tian, Y.; Su, D.; Lauria, S.; Liu, X. Recent advances on loss functions in deep learning for computer vision. *Neurocomputing* **2022**, *497*, 129–158. [CrossRef]
44. Feng, Y.; Li, Y. An overview of deep learning optimization methods and learning rate attenuation methods. *Hans J. Data Min.* **2018**, *8*, 186–200. [CrossRef]
45. Schindler, K. An overview and comparison of smooth labeling methods for land-cover classification. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4534–4545. [CrossRef]
46. Jian-Wei, L.; Hui-Dan, Z.; Xiong-Lin, L.; Jun, X. Research progress on batch normalization of deep learning and its related algorithms. *Acta Autom. Sin.* **2020**, *46*, 1090–1120.
47. Dubey, S.R.; Singh, S.K.; Chaudhuri, B. Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing* **2022**, *503*, 92–108. [CrossRef]
48. Haji, S.H.; Abdulazeez, A.M. Comparison of optimization techniques based on gradient descent algorithm: A review. *PalArch's J. Archaeol. Egypt/Egyptol.* **2021**, *18*, 2715–2743.
49. Supani, A.; Andriani, Y.; Indarto, H. Enhancing Deeper Layers with Residual Network on CNN Architecture: A Review. In Proceedings of the 6th FIRST 2022 International Conference (FIRST 2022), Singapore, 14–16 October 2023; p. 449.
50. Gugulothu, V. Deep residual networks based image recognition-review. *J. Innov. Dev. Pharm. Tech. Sci.* **2022**, *5*, 14–17.
51. Zagoruyko, S.; Komodakis, N. Wide residual networks. *arXiv* **2016**, arXiv:1605.07146.
52. Guo, M.-H.; Xu, T.-X.; Liu, J.-J.; Liu, Z.-N.; Jiang, P.-T.; Mu, T.-J.; Zhang, S.-H.; Martin, R.R.; Cheng, M.-M.; Hu, S.-M. Attention mechanisms in computer vision: A survey. *Comput. Vis. Media* **2022**, *8*, 331–368. [CrossRef]
53. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]
54. Burt, P.J. Attention mechanisms for vision in a dynamic world. In Proceedings of the 9th International Conference on Pattern Recognition, Rome, Italy, 14 May–17 November 1988; pp. 977–987.
55. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
56. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
57. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef] [PubMed]
58. Christoffersen, P.; Jacobs, K. The importance of the loss function in option valuation. *J. Financ. Econ.* **2004**, *72*, 291–318. [CrossRef]
59. Van Laarhoven, T. L2 regularization versus batch and weight normalization. *arXiv* **2017**, arXiv:1706.05350.
60. 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]

61. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of the 29th Annual Conference on Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 7–12 December 2015.
62. Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving into High Quality Object Detection. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.
63. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.
64. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.