(Received 10 August 2025, Accepted 26 September 2025, Scheduled 2 October 2025)

SAR Aircraft Detection Network Based on Multi-Branch Collaborative Calibration and Feature Enhancement

Zengyuan Guo[1, 2, *], Wei Xu[1, 2], Pingping Huang[1, 2], Weixian Tan[1, 2], and Zhiqi Gao[1, 2]

[1] College of Information Engineering, Inner Mongolia University of Technology, Hohhot, China
[2] Inner Mongolia Key Laboratory of Radar Technology and Application, Hohhot, China

ABSTRACT: Aircraft target detection in synthetic aperture radar (SAR) images faces numerous challenges, primarily including weak contrast, diverse morphologies, and faint signals, which are even more pronounced in complex backgrounds. Meanwhile, practical deployment environments are constrained by limited computational resources and energy consumption, making it essential to balance detection accuracy with model lightweight design. To address this, this paper proposes a lightweight detection network that integrates multi-branch feature enhancement. First, a Parallel Aggregation and Calibration (PAC) module is designed to achieve collaborative modeling of local and global information through multi-scale dilated convolutions; second, a Moment Channel Attention (MCA) module based on higher-order statistical features is introduced to enhance the model's sensitivity to weak signals and target boundaries; finally, during the network fusion stage, the branch calibration connections in the PAC module are removed, and a frequency-domain-driven Efficient Discriminative Frequency domain-based FFN (EDFFN) module is incorporated to improve the detailed representation of low-contrast and blurred targets. Experimental results on the SAR-Aircraft-1.0 dataset demonstrate that the proposed method achieves 93.94% mAP, while reducing model parameters by 56% and computational complexity by 36% compared to YOLOv12s, effectively balancing performance and lightweight requirements.

1. INTRODUCTION

Synthetic Aperture Radar (SAR), as an active microwave remote sensing technology, offers the fundamental advantage of all-weather and all-day imaging capabilities [1]. By overcoming the limitations imposed by weather conditions and illumination in optical imaging [2], SAR has been widely applied in military reconnaissance, natural disaster monitoring, and maritime observation [3], serving as a key component in modern information acquisition systems. In civil aviation, SAR imagery provides essential technical support for airspace [4] surveillance and aviation safety management, particularly serving an irreplaceable role [5] in aircraft target detection.

With the advancement of techniques and the increasing complexity of application scenarios, deep learning has become a primary approach for improving the performance of target detection [6] in SAR imagery. Research in this direction has proceeded step by step. In earlier studies, in order to address challenges posed by speckle noise and target scattering characteristics in SAR images [7], many efforts focused on accuracy improvement by designing deep network architectures to enhance the extraction and representation of complex features [8]. For example, Xiao et al. [9] enhanced the representation of scattering features in SAR images for aircraft detection by incorporating spatial feature fusion and deformable convolution structures. Kang et al. [10] introduced Scattering Topology Module (STM) to enhance the spatial relationships and semantic information interaction of scattering points, addressing the challenges posed by imaging variability. Zhao et al. [11] employed Graph Convolutional Networks (GCN) to extract structural and semantic features and combined them with an improved VGGNet to extract image domain features, aiming to enhance aircraft recognition performance in SAR images. Zhou et al. [12] proposed a method based on a bounding box denoising diffusion process that removes anchor size selection and introduced a scattering feature enhancement module to reduce clutter interference, thereby improving target detection performance.

Although deeper networks have improved feature representation capabilities [13], their large parameter counts and high computational demands limit their applicability in resource-constrained environments [14], such as satellite platforms, where efficiency and power consumption are key considerations. These constraints have led to a shift in focus from accuracy-oriented strategies to solutions that balance accuracy and computational efficiency. Compared with post hoc compression and pruning methods, direct architectural optimization has become a more effective and widely adopted approach [15]. In this context, lightweight network designs have shown greater practical suitability for SAR-based aircraft target detection and are receiving increasing attention in related research [16]. For example, Luo et al. [17] proposed a lightweight detection network that integrates Involution-Enhanced Path Aggregation (IEPA) module and Efficient Residual Shuffle Attention (ERSA) module to address challenges in multi-scale feature extraction and background interference. Luo et al. [18] proposed a detection model incorporating a lightweight recurrent gated spatial interaction

* Corresponding author: Zengyuan Guo (20231100107@imut.edu.cn).

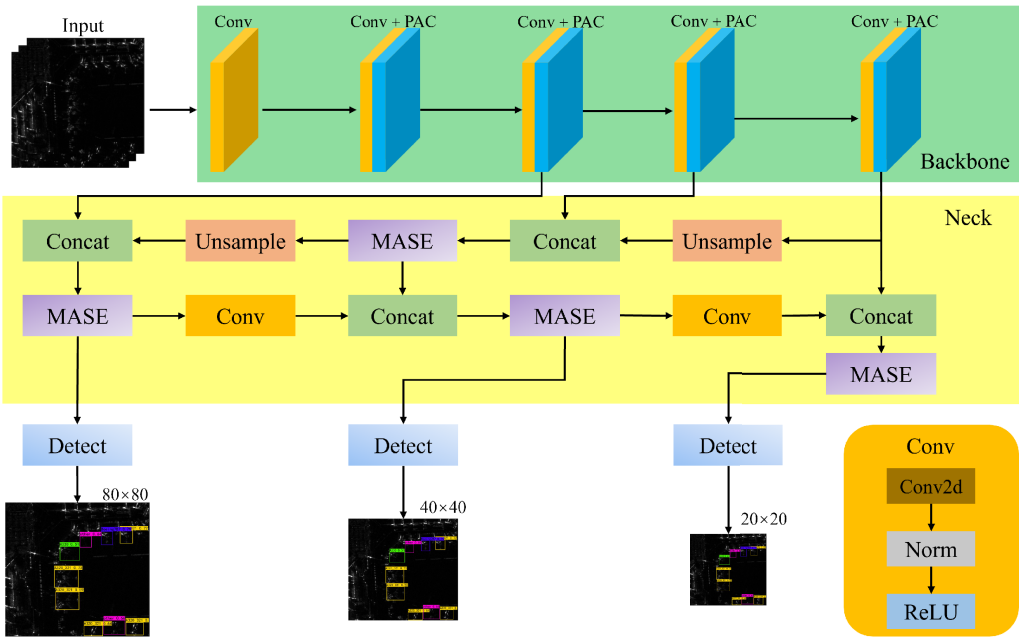

FIGURE 1. Overall structure diagram.

module to enhance feature modeling in panoramic SAR imagery. Zhu et al. [19] combined a lightweight SdE-Resblock structure with an improved ssppf-CSP module to enhance feature representation while reducing computational complexity. Yang et al. proposed a lightweight pseudo-Siamese feature extraction module and a random mask reconstruction mechanism to efficiently extract heterogeneous image features and enhance noise robustness [20].

Although existing studies have advanced lightweight design for SAR target detection from various perspectives such as feature modeling, attention mechanisms, and structural compression, these methods primarily focus on isolated optimizations of local modules, lacking systematic integration of hierarchical structural coordination and perceptual modeling capabilities. When facing typical SAR imaging challenges such as low contrast, small-scale targets, blurred boundaries, and complex background interference, current approaches still struggle to balance detection accuracy and computational efficiency, making the trade-off between lightweight design and high performance a critical challenge. To address this, this paper proposes an efficient detection framework with hierarchical structural optimization and enhanced perceptual capability. Specifically, a Parallel Aggregation and Calibration (PAC) module integrating multi-scale dilated convolutions is designed to improve collaborative modeling of local and global information; Moment Channel Attention (MCA) module [21] based on higher-order statistical features is introduced to increase the model's sensitivity to weak signals and boundary targets. An Efficient Discriminative Frequency domain-based FFN (EDFFN) [22] is incorporated into the neck and integrated with the PAC module to form a new component, Multi-scale Aggregation and Enhancement (MSAE), which improves multi-scale feature aggregation and enhances the model's perceptual capacity. In the neck stage, PAC branch connections are decoupled, reducing computational complexity while maintaining detection performance.

2. MATERIALS AND METHODS

2.1. Network Structure Overview

Figure 1 illustrates the overall architecture of the proposed network. The entire process consists of three stages: First, the input SAR image is fed into the backbone network to extract multi-level features. Next, these features are transmitted to the neck network for cross-scale information fusion and semantic enhancement. Finally, the fused multi-scale features are used for classification and localization in the target detection task. The backbone is constructed using Parallel Aggregation and Calibration (PAC) module as its fundamental building block. This module employs a multi-branch structure to perform differentiated modeling of input features, taking into account the cluttered background and varied target scales in SAR images. The outputs from each branch are concatenated and passed through a Moment Channel Attention (MCA) module, which performs cross-branch feature weighting and channel-wise calibration to improve the representation of discriminative features.

The neck incorporates Multi-scale Aggregation and Enhancement (MSAE) module to consolidate the multi-scale feature maps output from the backbone. MSAE retains the multi-branch design of PAC module to associate features across different semantic levels, while omitting multiplicative enhancement paths to reduce computational cost and structural complexity. It also embeds a frequency domain feed-forward submodule, named Efficient Discriminative Frequency domain-based FFN (EDFFN), which models local correlations in the frequency domain, complementing the representation of scale features from both spatial and frequency perspectives.

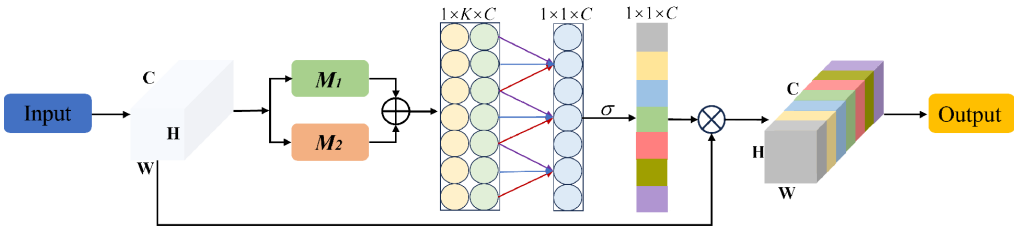

FIGURE 2. Moment Channel Attention (MCA) module.

The aggregated feature maps are then passed to the detection head to generate bounding boxes and class labels of the targets.

2.2. Moment Channel Attention (MCA) Module

To capture more discriminative statistical characteristics along the channel dimension, we introduce Moment Channel Attention (MCA) module into the network, as illustrated in Figure 2. This module incorporates both mean and skewness information of each channel in a lightweight manner, generating attention weights that reflect channel-wise response differences to enhance the input feature map [20].

Structurally, MCA first computes the first-order moment (mean) and third-order central moment (skewness) of the input feature map $X \in R^{B\times C\times H\times W}$ along the spatial dimensions, formulated as:

$$M_1(c) = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} X_c(i,j) \tag{1}$$

$$M_2(c) = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(\frac{X_c(i,j)-M_1(c)}{\sigma_c}\right)^3 \tag{2}$$

where σ_c is the standard deviation of channel C. These two statistics represent the shape of the channel feature distribution — the mean indicates basic intensity, while the skewness captures asymmetry in the distribution.

The concatenated M_1 and M_2 vector of size 2C is then passed through a 1D convolution followed by a Sigmoid activation to generate the channel attention weight F:

$$F = \sigma\left(Conv1D(Concat(M_1, M_2))\right) \tag{3}$$

This attention vector F is applied to the input feature X via channel-wise multiplication to produce the output feature map Y:

$$Y = X \cdot F \tag{4}$$

2.3. Parallel Aggregation and Calibration (PAC) Module

As illustrated in Figure 3, PAC module first applies pointwise convolution to expand input feature channels to twice of the original number. The expanded features are then evenly divided into two parts along the channel dimension with each part containing the same number of channels. One part is further partitioned into three sub-branches. 5×5 and 7×7 branches employ dilated convolutions, which enlarge the receptive field and increase the complexity of structural representations; each

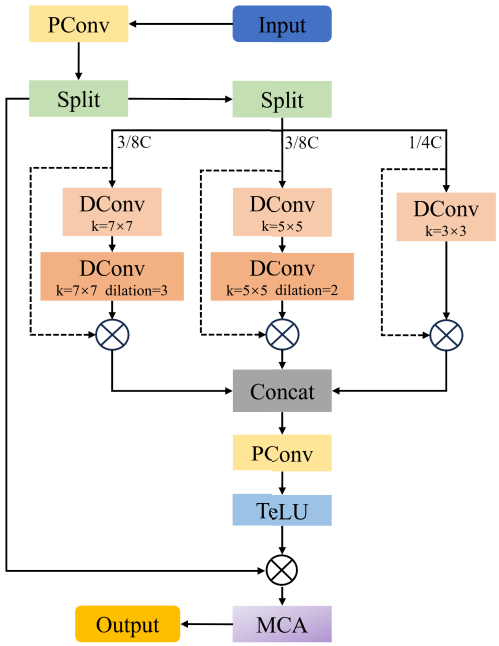

FIGURE 3. Parallel Aggregation and Calibration (PAC) module.

of these branches is therefore assigned 3/8C channels, providing additional capacity to represent broader contextual dependencies while preserving structural details. The 3×3 branch, in contrast, is specifically designed to capture fine-scale spatial details that require high-resolution local information but relatively lower representational capacity. Therefore, it is assigned 1/4C channels, which balances precise local feature extraction with computational efficiency.

During the fusion stage, two connection strategies are employed: in one, each sub-branch retains a shortcut copy, which is multiplied element-wise with the processed features to enhance intra-branch interaction, suitable for the backbone stage to strengthen low-level feature representation; in the other, only the processed features from each sub-branch are concatenated, without shortcut connections, which is more compatible with feature integration in the neck stage and subsequent frequency domain enhancement modules. These two structures differ only in their fusion strategies, allowing adaptation to different task requirements. The fused features obtained via channel-wise multiplication are then activated [23] and further multiplied by the other part of the features, which remain unprocessed, to calibrate the overall response intensity. Finally, Moment Channel Attention module is introduced to enhance inter-channel discriminability.

2.4. Efficient Discriminative Frequency Domain-Based FFN (EDFFN) Module

In complex environments, where high noise interference and multi-scale features coexist, improving feature representation and enhancing model adaptability are practical issues that require attention. EDFFN module [22] adopts a progressively enhanced feature extraction strategy, optimizing feature representation through multi-level processing, which effectively adapts to the needs of subsequent tasks.

First, the module employs pointwise convolutions (PConv) to expand the number of channels, thereby increasing feature dimensions and laying the foundation for enhanced feature expression. Next, deep convolution (DConv) is applied to feature extraction, successfully preserving key high-dimensional features while reducing computation, thus improving feature extraction capability without compromising efficiency. Specifically, the feature map is split into two parts: one is processed through the TeLU activation function, and the other (unprocessed part) is multiplied with it, followed by a channel-wise operation (as shown in Figure 4). Unlike the commonly used GeLU function TeLU [23] applies a controlled non-linearity that amplifies subtle variations in feature values, particularly for low-magnitude features that are often critical in complex scenarios. This selective emphasis allows the network to process weak or fine-grained features more distinctly, supporting the preservation and discrimination of small or low-contrast information during feature extraction. After spatial feature extraction, the module introduces a frequency-domain enhancement strategy, combining reinforcement from both the spatial and frequency domains. Using two-dimensional fast Fourier analysis (FFT), the feature map is moved from spatial domain to frequency domain, where frequency components are selectively processed. The frequency components of the feature map are multiplied by predefined parameters to amplify important frequency components and suppress irrelevant noise, allowing the model to focus on critical information. Finally, after inverse fast Fourier analysis (IFFT), the feature map is returned to the spatial domain, further refining spatial feature representation.

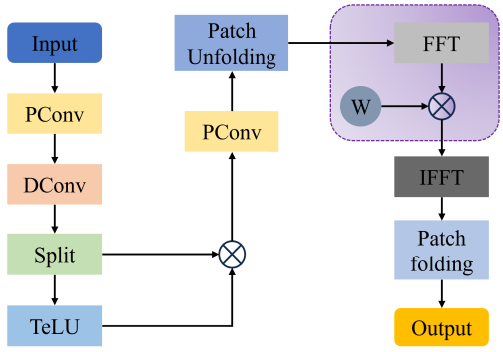

FIGURE 4. Efficient discriminative frequency domain-based FFN (EDFFN) module.

2.5. Multi-Scale Aggregation and Enhancement (MSAE) Module

The neck of the network is located between the backbone and detection head. Its primary role is to integrate features from different levels and transmit them to subsequent stages. In this process, both fine-grained spatial information and high-level semantic representations need to be considered. This aspect is particularly important for SAR imagery. For example, in airport scenes, aircraft targets are usually small in size and densely distributed, and easily confused with complex structures such as runways and buildings, together with strong background noise. These characteristics make the integration and selection of features in the neck more challenging. The output feature maps of the neck cover multiple resolutions, providing a foundation for handling such complexity across different scales and making both detail preservation and semantic enhancement necessary.

As shown in Figure 5, the design of MSAE module is developed around these requirements. Considering the role of the neck in feature integration and transmission, the module organizes different functional units into a continuous processing path. The PAC module captures spatial patterns through multi-scale convolutions, supplying structured information for subsequent stages. Building on this, MCA module adjusts channel-wise responses to emphasize potential key features. Finally, EDFFN refines and supplements feature representations by combining spatial and frequency domain processing, thereby improving adaptability to complex and variable environments. In addition, normalization and residual connections are introduced to maintain numerical stability during transmission and to alleviate gradient degradation caused by network depth.

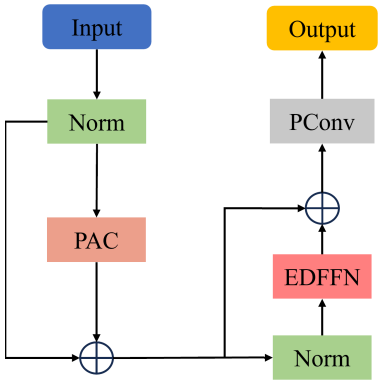

FIGURE 5. Multi-Scale Aggregation and Enhancement (MSAE) module.

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1. Dataset

All experiments in this paper are conducted based on SAR-AIRcraft-1.0 dataset [24]. This dataset was constructed by the Aerospace Information Research Institute, Chinese Academy of Sciences and is one of the representative publicly available benchmarks in current research on high-resolution SAR-based aircraft detection and recognition. The image data were collected by the Gaofen-3 satellite, with a spatial resolution of 1 meter, covering multi-temporal observations of three civil airports: Shanghai Hongqiao, Beijing Capital, and Taoyuan Airport in Taiwan.

The dataset contains a total of 4368 image patches, with spatial dimensions ranging from 800×800 to 1500×1500 pixels. A total of 16463 aircraft instances are annotated and categorized into seven fine-grained classes: A220 (3730), A320/321 (1771), A330 (309), ARJ21 (1187), Boeing737 (2557), Boeing787 (2645), and Other (4264), where the "Other" class includes aircraft types not belonging to the six aforementioned categories. In this study, the dataset is randomly divided into training, validation, and test sets at a ratio of $8:1:1$ to evaluate the effectiveness and generalization capability of the proposed model in multi-scale target representation and fine-grained recognition tasks.

3.2. Experimental Environment and Details

All experiments were conducted on a high-performance workstation equipped with an Intel Core i9-13900K processor and an NVIDIA GeForce RTX 4090 GPU with 24 GB of memory, capable of efficiently supporting large-scale deep learning model training and inference tasks. The operating system used was 64-bit Ubuntu 22.04 LTS, ensuring system compatibility and operational stability. Model development and training were based on the PyTorch 2.0.1 framework, programmed in Python 3.9.23, with CUDA 11.8 employed to accelerate GPU computations and improve training efficiency.

During training, an Adam optimizer was used for parameter updates. Its adaptive learning rate feature contributed to enhanced training stability and convergence speed. The initial learning rate was set to 0.001, and a cosine annealing learning rate scheduler was applied to gradually reduce the learning rate to promote model convergence. Training was performed for a total of 500 epochs with a batch size of 16, balancing training efficiency and GPU memory utilization.

All experiments were carried out on a single GPU, with GPU power consumption and temperature maintained within normal operating ranges throughout training, ensuring hardware stability during extended runtime.

3.3. Evaluation Criteria

To thoroughly evaluate the performance of the proposed model in object detection, four widely used and representative metrics are adopted: Recall, Precision, F1 Score, and mean Average Precision at an Intersection over Union (IoU) threshold of 0.5 (mAP@0.5). These metrics evaluate different aspects of the model's behavior, including its ability to identify targets, avoid false detections, and maintain overall balance.

Recall measures the model's ability to correctly detect actual targets, defined as:

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

where *TP* denotes the number of correctly detected targets, and *FN* is the number of missed ground-truth targets.

Precision evaluates the proportion of correct predictions among all predicted positive instances:

$$Precisin = \frac{TP}{TP + FP} \tag{6}$$

with *FP* representing the number of false alarms, i.e., non-targets incorrectly identified as targets.

F1 Score, defined as the harmonic mean of precision and recall, balances the trade-off between these two metrics:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{7}$$

It is particularly useful when both false positives and false negatives need to be simultaneously minimized.

Average Precision (AP) summarizes the precision-recall relationship for a single class. Without requiring curve plotting, it can be approximated as the area under the precision-recall curve and is defined as:

$$AP = \int_0^1 P(R)dR \tag{8}$$

where $P(R)$ is the precision as a function of recall R. In practice, this integral is estimated by numerical interpolation over discrete recall levels.

Mean Average Precision (*mAP*@0.5) is the mean of AP values across all target classes under an IoU threshold of 0.5:

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{9}$$

where N is the total number of classes, and AP_i is the AP for class i. This metric provides a comprehensive view of the model's detection performance in terms of both localization and classification accuracy.

These metrics collectively provide a comprehensive basis for measuring model effectiveness and enabling comparison across different settings.

3.4. Ablation Experiment

To examine the roles of different modules in the SAR aircraft detection task and to evaluate their behavior under varying combinations, we conducted an ablation study, with the results summarized in Table 1. The investigation focused on three modules: MCA module introduces channel-level attention guided by statistical features; PAC module organizes multi-scale feature aggregation; and EDFFN module refines feature representation through the joint use of spatial- and frequency-domain operations.

When considering individual contributions, applying MCA module alone (Number 1) improves responses to weak targets while keeping the structure lightweight, which suggests that its moment-driven attention offers useful cues for initial localization. Extending this setting by adding EDFFN module (Number 5) leads to a notable gain in detection accuracy. The additional frequency-domain enhancement helps delineate target boundaries more reliably and stabilizes feature quality under noisy conditions.

Interactions among modules provide further insights. Embedding MCA module into the multi-branch structure of PAC

TABLE 1. Ablation results of the proposed modules.

Number	MCA	PAC	EDFFN	mAP (%)	Params (M)	FLOPs (G)
1	√			89.63	2.3	9.8
2	√	√		92.12	2.9	11.2
3	√	√	√	93.94	4.1	13.7
4		√	√	93.40	3.8	13.5
5	√		√	92.56	3.6	12.4

module (Number 2) results in more coherent integration of features across scales, showing advantages when processing low-contrast targets and complex backgrounds. In contrast, removing MCA and retaining only PAC and EDFFN (Number 4) continues to improve accuracy, implying that the direct link between structural aggregation and frequency-domain refinement can strengthen semantic modeling with relatively few modules.

The complete configuration (Number 3) integrates all three modules: MCA refines channel selection; PAC reinforces scale-aware organization; and EDFFN enriches feature representations through spatial and frequency domain processing. In this setting, detection performance remains stable across different image resolutions and complex scenes, demonstrating the capacity of these modules to operate in combination.

Taken together, the results indicate that the three modules contribute from different perspectives. MCA emphasizes channel-level attention with efficiency; PAC enhances structural organization across scales; and EDFFN provides additional descriptive capacity through frequency-domain refinement. Their combinations produce complementary effects, leading to better detection performance while keeping the overall model complexity under control.

3.5. Proposed Model on SAR-AIRcraft-1.0

On the SAR-AIRcraft-1.0 dataset, the proposed model achieves stable detection performance across all aircraft categories, with AP values exceeding 91% in every case as shown in Table 2. These results indicate that the model maintains strong generalization ability under varying aircraft structures and imaging conditions, particularly in the presence of the complex textures and noise patterns typical of SAR imagery. For categories such as A220, ARJ21, and Boeing737, the model demonstrates high precision and recall, suggesting that it can provide reliable detection even against challenging background interference.

TABLE 2. Proposed model on SAR-AIRcraft-1.0.

Category	AP (%)	F1	Precision (%)	Recall (%)
A220	93.73	0.87	84.55	90.22
A320_321	92.61	0.87	80.60	94.19
A330	91.91	0.86	77.50	96.88
ARJ21	94.28	0.90	86.32	94.39
Boeing737	94.68	0.90	89.68	91.30
Boeing787	97.60	0.92	91.84	92.83
other	92.76	0.86	83.87	88.78

For more demanding categories such as A330 and "other", the model also delivers consistent results. A330, characterized by a larger airframe and frequent overlap with background clutter, presents higher detection difficulty, and yet the model still achieves 91.91% AP and nearly 97% recall, reflecting its ability to capture large-scale spatial patterns. The "other" category, which aggregates a wide range of aircraft instances outside the six common types, involves substantial intra-class variation that typically complicates classification and localization. Nonetheless, the model attains 92.76% AP and maintains a balanced trade-off between precision and recall, suggesting that the designed architecture adapts effectively to category diversity and structural complexity.

3.6. Comparison with Other Aircraft Detection Models

As shown in Table 3, we compare the proposed model with several mainstream detection methods on the SAR-AIRcraft-1.0 dataset. It can be observed that traditional approaches such as Faster Region-based Convolutional Neural Network (RCNN) and Single Shot MultiBox Detector (SSD) exhibit clear limitations in either detection accuracy or inference efficiency. The former has a parameter count of 139.1M and 370.2G Floating Point (FLOP) operations, making it unsuitable for efficient deployment; the latter, while more compact, achieves relatively lower accuracy (86.52%), limiting its applicability in complex target scenarios.

TABLE 3. Proposed model on SAR-AIRcraft-1.0.

Model	mAP (%)	Params (M)	FLOPs (G)
Faster RCNN	82.69	139.1	370.2
SSD	86.52	26.3	62.8
YOLOv5s	92.87	7.2	16.5
YOLOv8s	93.05	11.2	28.6
RT-DETR-R18	93.26	20.0	60.0
YOLOv11s	93.31	9.4	21.5
YOLOv12s	93.40	9.3	21.4
ours	93.94	4.1	13.7

RT-DETR-R18 achieves a slightly higher mAP of 93.26% than YOLOv8s, but its larger model size (20.0 M parameters and 60.0 G FLOPs) makes it less suitable for resource-constrained environments, where smaller architectures are preferred. YOLO-based models, including YOLOv5s, YOLOv8s, YOLOv11s, and YOLOv12s, demonstrate a gradual progres-

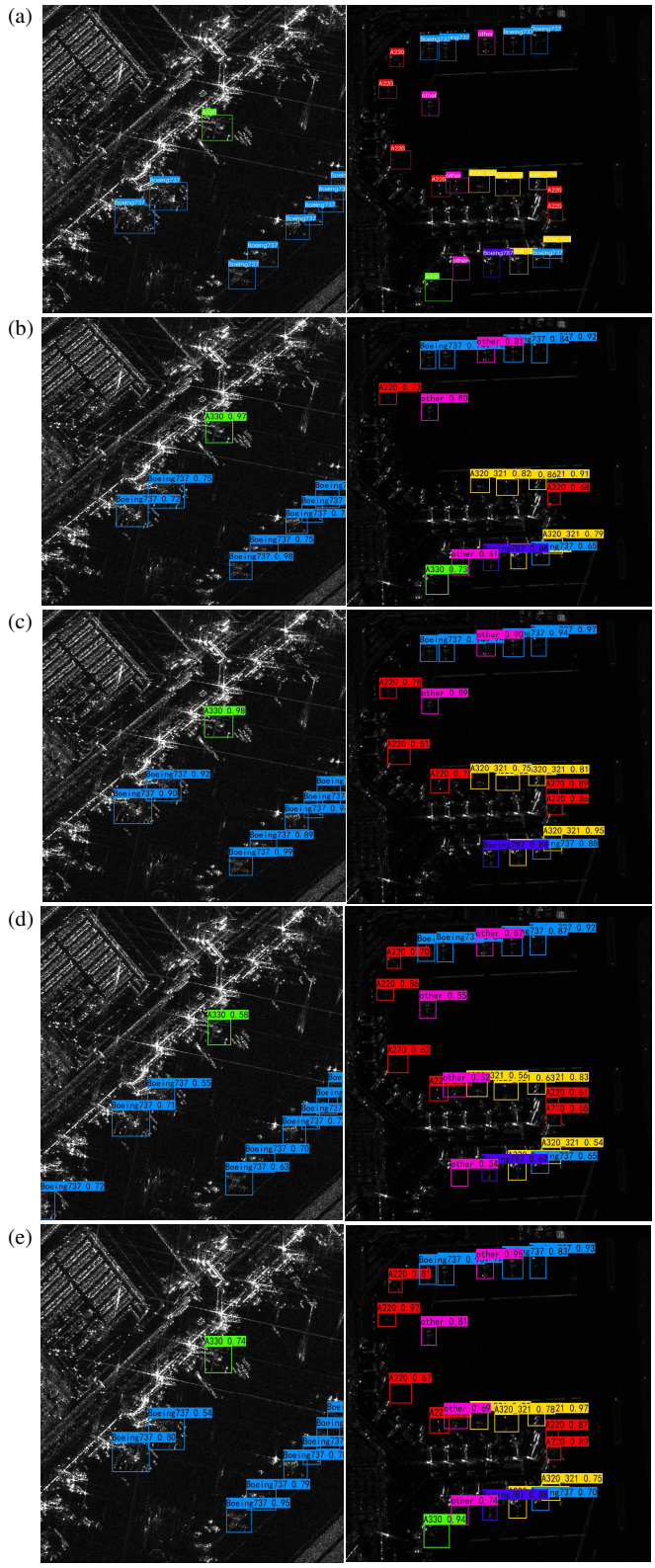

FIGURE 6. Comparison of detection results on the SARAircraft1.0 dataset. (a) Ground truth. (b) RT-DETR-R18. (c) YOLOv11s. (d) YOLOv12s. (e) Ours.

sion in detection performance, with mAP values ranging from 92.87% to 93.40%. While these results reflect steady advances, there remains room for further exploration in balancing detection accuracy and model efficiency, particularly in scenarios that demand both high precision and lightweight design.

In contrast, our method achieves an mAP of 93.94% using only 4.1 M parameters and 13.7 G FLOPs, outperforming all compared models. This result not only demonstrates the model's strong target recognition capability within an extremely compact structure but also emphasizes its suitability for real-world scenarios with strict computational and storage constraints, such as on-orbit satellites or unmanned aerial vehicle (UAV)-based systems. The design strikes a well-balanced trade-off between structural efficiency and detection performance, demonstrating its effectiveness and application potential in aircraft target detection on SAR imagery. As shown in Figure 6, compared to other advanced models, our model is the only one that successfully detects and accurately identifies all targets, further validating its superiority in detection completeness.

4. CONCLUSION

This paper proposes a lightweight SAR aircraft target detection network that combines multi-branch collaborative calibration and frequency-domain feature enhancement. Experimental results show that the network achieves an mAP of 93.94% on a SAR-AIRcraft-1.0 dataset, surpassing current leading detection models such as YOLOv12s and RT-DETR-R18, demonstrating the effectiveness and advanced nature of the designed architecture. The proposed model has only 4.1 M parameters and 13.7 G FLOPs, emphasizing its excellent lightweight characteristics while maintaining high accuracy. This makes the model highly suitable for the deployment on resource-constrained platforms (such as satellites and drones) and provides a practical technical solution for SAR image target detection tasks.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China under grant numbers U22A2010.

REFERENCES

[1] Zhang, Q., H. Fan, Y. Qin, and Y. Zhou, "Advances in interferometric synthetic aperture radar technology and systems and recent advances in Chinese SAR missions," *Sensors*, Vol. 25, No. 15, 4616, 2025.

[2] Yoon, S., M. Kim, M. Jang, Y. Choi, W. Choi, S. Kang, and W. Choi, "Deep optical imaging within complex scattering media," *Nature Reviews Physics*, Vol. 2, No. 3, 141–158, 2020.

[3] Liu, T., Z. Yang, Y. Jiang, and G. Gao, "Review of ship detection in polarimetric synthetic aperture imagery," *Journal of Radars*, Vol. 10, No. 1, 1–19, 2021.

[4] Hu, C., Z. Chen, Y. Li, X. Dong, and S. Hobbs, "Research progress on geosynchronous synthetic aperture radar," *Fundamental Research*, Vol. 1, No. 3, 346–363, 2021.

[5] Kıvanç, E., G. Tuzkaya, and O. Vayvay, "Safety management system and risk-based approach in aviation maintenance: A systematic literature review," *Safety Science*, Vol. 184, 106755, 2025.

[6] Ahmed, S. F., M. S. B. Alam, M. Hassan, M. R. Rozbu, T. Ishtiak, N. Rafa, M. Mofijur, A. B. M. S. Ali, and A. H. Gandomi,

"Deep learning modelling techniques: Current progress, applications, advantages, and challenges," *Artificial Intelligence Review*, Vol. 56, No. 11, 13 521–13 617, 2023.

[7] Singh, P. and R. Shree, "Analysis and effects of speckle noise in SAR images," in *2016 2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA) (Fall)*, 1–5, Bareilly, India, 2016.

[8] Khan, A., A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, Vol. 53, No. 8, 5455–5516, 2020.

[9] Xiao, X., H. Jia, P. Xiao, and H. Wang, "Aircraft detection in SAR images based on peak feature fusion and adaptive deformable network," *Remote Sensing*, Vol. 14, No. 23, 6077, 2022.

[10] Kang, Y., Z. Wang, H. Zuo, Y. Zhang, Z. Yang, X. Sun, and K. Fu, "ST-Net: Scattering topology network for aircraft classification in high-resolution SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 61, 1–17, 2023.

[11] Zhao, C., S. Zhang, R. Luo, S. Feng, and G. Kuang, "Scattering features spatial-structural association network for aircraft recognition in SAR images," *IEEE Geoscience and Remote Sensing Letters*, Vol. 20, 1–5, 2023.

[12] Zhou, J., C. Xiao, B. Peng, Z. Liu, L. Liu, Y. Liu, and X. Li, "DiffDet4SAR: Diffusion-based aircraft target detection network for SAR images," *IEEE Geoscience and Remote Sensing Letters*, Vol. 21, 1–5, 2024.

[13] Islam, M. T. and L. Xing, "Deciphering the feature representation of deep neural networks for high-performance AI," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 46, No. 8, 5273–5287, 2024.

[14] Rodriguez-Conde, I., C. Campos, and F. Fdez-Riverola, "Optimized convolutional neural network architectures for efficient on-device vision-based object detection," *Neural Computing and Applications*, Vol. 34, No. 13, 10 469–10 501, 2022.

[15] Zhang, S., Y. Luo, Z. Lyu, and X. Chen, "ShiftKD: Benchmarking knowledge distillation under distribution shift," *Neural Networks*, Vol. 192, 107838, 2025.

[16] Wu, H., H. Sang, Z. Zhang, and W. Guo, "LRMSNet: A new lightweight detection algorithm for multi-scale SAR objects," *Remote Sensing*, Vol. 16, No. 12, 2082, Jun. 2024.

[17] Luo, R., L. Chen, J. Xing, Z. Yuan, S. Tan, X. Cai, and J. Wang, "A fast aircraft detection method for SAR images based on efficient bidirectional path aggregated attention network," *Remote Sensing*, Vol. 13, No. 15, 2940, 2021.

[18] Luo, R., Q. He, L. Zhao, S. Zhang, G. Kuang, and K. Ji, "Geospatial contextual prior-enabled knowledge reasoning framework for fine-grained aircraft detection in panoramic SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 62, 2024.

[19] Zhu, W., L. Zhang, C. Lu, G. Fan, Y. Song, J. Sun, and X. Lv, "FEMSFNet: Feature enhancement and multi-scales fusion network for SAR aircraft detection," *Remote Sensing*, Vol. 16, No. 9, 1589, 2024.

[20] Yang, C., G. Q. Gong, C. Liu, J. W. Deng, and Y. X. Ye, "RMSO-ConvNeXt: A lightweight CNN network for robust SAR and optical image matching under strong noise interference," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 63, 2025.

[21] Jiang, Y., Z. Jiang, L. Han, Z. Huang, and N. Zheng, "MCA: Moment channel attention networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38, No. 3, 2579–2588, Vancouver, Canada, Feb. 2024.

[22] Kong, L., J. Dong, J. Tang, M.-H. Yang, and J. Pan, "Efficient visual state space model for image deblurring," in *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12 710–12 719, Nashville, TN, USA, 2025.

[23] Fernandez, A. and A. Mali, "TeLU activation function for fast and stable deep learning," *arXiv:2412.20269*, 2024.

[24] Wang, Z., Y. Kang, X. Zeng, Y. Wang, T. Zhang, and X. Sun, "SAR-AIRcraft-1.0: High-resolution SAR aircraft detection and recognition dataset," *Journal of Radars*, Vol. 12, No. 4, 906–922, 2023.