Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.37773

PGSNet: Physics-Guided Scatter-Preserving Network for Large-Scale SAR Object Detection

Jiahao Fu

Department of Computer Science and Technology, Ocean University of China, Qingdao, China
tback0612@gmail.com

Abstract. Target detection using Synthetic Aperture Radar (SAR) remains challenging: multiplicative speckle can obscure shallow features, while isolated scatterers may be overlooked under conventional fixed-step subsampling processing. Attention-based fusion methods and standard ImageNet-to-SAR transfer methods typically treat high-frequency clutter as salient structures. This paper proposes PGSNet, a single-stage detector that incorporates SAR physical prior knowledge across three complementary depths. The Physics-Guided Deparameterized Stem Network (PGF-Stem) connects the original intensity with a gradient-based ratio edge map, and aligns these two data streams via a 1×1 projection, without the need to learn a gating mechanism. Spatial-to-Depth Convolution (SPD-Conv) then reorganizes the weak scattering responses along the channel dimension, thereby preventing these responses from being spatially averaged due to resolution degradation. The coefficient-of-variation-guided feature calibration (CGFC) ultimately employs the classic speckle reliability index to reweight the feature pyramid map. Built upon the generalized focal loss (GFL) framework and trained using ResNet-50 on the SARDet-100K dataset across 12 training cycles, PGSNet achieved an mAP of 58.2% and an APS of 53.1% on the COCO dataset, outperforming DenoDet V2 (56.71%) and ConvNeXt-B with MSFA (56.4%). The detector still runs at 36.9 FPS on the RTX 3090.

Keywords: SAR object detection, speckle noise, small object, physical prior, single-stage detector

1. Introduction

Synthetic Aperture Radar (SAR) can image the Earth's surface regardless of sunlight, cloud cover, or various camouflage techniques, making it an indispensable tool for maritime surveillance, disaster response, and defense reconnaissance. However, the same imaging physical mechanism can produce multiplicative speckles when applied to sparse isolated scatterers. Consequently, detectors that exhibit excellent performance in optical benchmark tests often demonstrate degraded performance in SAR imaging tasks: shallow convolution can easily lead to overfitting of clutter, while small targets may disappear after several stride-2 convolutions.

For a long time, research progress has been constrained by small-scale datasets and a scarcity of code. SARDet-100K integrated ten datasets into the COCO-scale benchmark and released MSFA, which boosted the mAP of Faster R-CNN from 49.0% to 51.1% [1]. In the transformed frequency

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.37773

domain, DenoDet and DenoDet V2 achieved GFL scores of 55.4% and 56.71%, respectively [2]. Two mismatches remain.

First, most multi-source backbone networks still rely on learned attention mechanisms or adaptive gating mechanisms. Under low signal-to-noise ratio conditions, these gatekeepers assign greater weight to bright speckle particles, treating them as target evidence. Manually designed filters in MSFA can be helpful, but the subsequent fusion process remains a generic stitching operation, which is integrated into a backbone network specifically designed for additive RGB noise rather than for multiplicative radar echoes [3]. Secondly, the commonly employed processing techniques for small target objects—such as P2 detection heads or extended convolutions—still produce high-resolution maps, which happen to be the regions most susceptible to noise. Therefore, when the resolution is maintained, isolated scatterers are either eliminated via merging or buried within the clutter [4].

In this study, we retained the MSFA experimental setup and the GFL detector [5], and re Designed the backbone network, the downsampling operator, and the feature pyramid based on the principles of SAR physics. The physics-based scattering retention network (PGSNet) constructed in this work consists of three components. PGF-Stem employs gradient ratio edges (GRE) whose characteristics align with those of multiplicative spot patterns, replacing the additive edge extraction method of the Canny style, and enables seamless integration with intensity maps without the need for parameterized gating. When the feature map size is halved, SPD-Conv performs a spatial-to-depth rearrangement operation rather than a spatial aggregation operation. CGFC subsequently employed the local coefficient of variation—a classic speckle index—as a reliability field on the feature pyramid. PGSNet was trained for 12 epochs using ResNet-50 pre-initialized on ImageNet, achieving an mAP of 58.2% on the SARDet-100K dataset, while boosting the small-target AP from 42.3% to 53.1%. This method outperforms previously published MSFA variants (including ConvNeXt-B) as well as the current version of DenoDet V2, while remaining faster than two-stage detectors such as Faster R-CNN.

The contributions are primarily reflected in three aspects: a GRE-based deparameterized backbone network, whose stability surpasses that of the attention gate; an SPD-based downsampled path for redistributing isolated scattering energy; and CV-guided FPN calibration, which together with the aforementioned two modules forms a complementary noise-resistant and small-target detection pathway.

2. Method and materials

2.1. Dataset and implementation

All major experiments utilized SARDet-100K [6], with the training, validation, and test sets comprising 94,493 / 10,492 / 11,613 images, respectively, totaling 245,653 instances across six classes. We employed the officially defined 800×800 image size adjustment, horizontal flipping ($p = 0.5$), and the COCO evaluation metrics. All values are based on the validation set, which also serves as the splitting set employed by the majority of subsequent analytical methods.

PGSNet is built upon the MMDetection 3.1 framework and implemented using PyTorch 2.0. This detector employs GFL [7], with a backbone network of ResNet-50, and is initialized using the ImageNet-1K weights from torchvision. We conducted training across 8 RTX 3090 GPUs over 12 training cycles (using a single scheduling scheme), with a total batch size of 16, employing the AdamW optimizer (learning rate of 1e-4, weight decay coefficient of 0.05), a linear warm-up strategy of 500 iterations, and learning rate reduction in the 8th and 11th training cycles. All input

DOI: 10.54254/2755-2721/2026.37773

images were normalized using the standard mean and standard deviation of ImageNet. The inference speed was measured on an RTX 3090 with a batch size of 1, after 100 warm-up iterations (including post-processing). The migration tests for SSDD and HRSID both employed the same 12-round configuration, with a batch size set to 32.

2.2. Physics-Guided Deparameterized Stem

Traditional RGB signal channels respond to any high-frequency variations; in SAR imaging, however, many of these variations manifest as speckles, and the signal processing pipeline can further amplify these speckles. PGF-Stem retains two physics-oriented streams and employs only a single thin alignment layer.

The data stream is a single-channel intensity signal obtained by averaging the signals from three input channels. Physical flow is a GRE mapping [8]. The GRE estimates the horizontal and vertical derivatives as the logarithm of the ratio of complementary local averages, rather than using additive differences. Since the speckle effect amplifies the radar cross-section of the underlying layer, the ratio in the intensity domain corresponds to the difference in the logarithmic domain; when the noise variance varies proportionally to the signal, this ratio exhibits higher stability compared to the Canny or Sobel gradients. The GRE kernel function employs a fixed half-window averaging mask, which remains unchanged throughout the training process. The two feature maps are concatenated along the channel axis, and then, after applying a 1×1 convolution, batch normalization, and a ReLU activation function, they are projected onto three channels. This model neither employs a sigmoid gate nor incorporates channel attention or spatial attention mechanisms. In the 1×1 convolutional layer, only the stitched tensor is aligned with the pre-trained ResNet-50 backbone network; thus, the ImageNet weights of the conv1 layer can still be loaded.

This backbone network is narrower than MSFA, which concatenates multiple hand-designed descriptors and then trains the entire backbone network using this wide input [9]. Using a single GRE channel alone can suppress most of the shallow clutter that attention may lock onto. Replacing GRE with Canny in the same benchmark test reduced the mAP from 54.1% to 53.3%, supporting the multiplicative noise hypothesis. As shown in Figure 1, the GRE flow preserves longer target contours on the same chip, and the fused stem responses are concentrated at the object boundaries rather than appearing as isolated speckles.

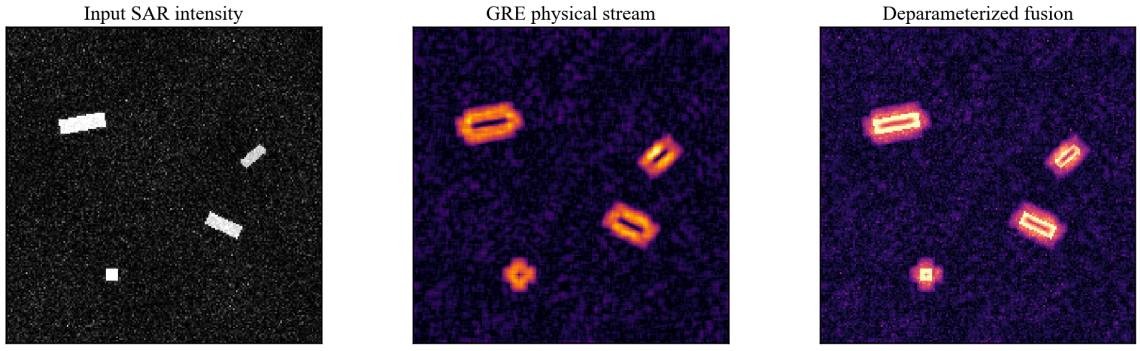

Figure 1. Intermediate outputs of PGF-Stem on a SARDet-100K training chip: intensity, GRE map, and fused stem feature

DOI: 10.54254/2755-2721/2026.37773

2.3. Scatter-preserving downsampling

Subsequently, ResNet-50 reduces the resolution using max pooling and convolution operations with a stride of 2. Independent scatterers constituting automobiles or small vessels can easily be diluted to concentrations below the detection threshold [10]. In P2, the head region retains these pixels, as well as the speckles that have just decayed in PGF-Stem.

SPD-Conv re-splits each 2×2 spatial block into four channels, and then performs a convolution operation with a step size of 1. Spatial energy is transferred rather than averaged. We replace the main pooling layer with an SPD projection layer capable of restoring 64 channels, and reorganize layers 2–4 so that the original layers, which had a stride of 2, operate with a stride of 1 after SPD reorganization. The feature map size and the FPN channel width are both kept consistent with those of the standard ResNet-50, while the GFL head remains unchanged. The newly introduced 1×1 projection layer is initialized randomly; the residual blocks continue to use ImageNet weights. By adding SPD-Conv to PGF-Stem, the small-target AP increases from 47.2% to 51.0%, and the overall mAP rises from 54.1% to 56.3%.

2.4. CV-guided feature calibration

The third module operates on the feature pyramid [6]. Even after employing cleaner stem extraction and distortion-free subsampling processing, subsequent layers still blend object energy with residual clutter—a phenomenon that is particularly pronounced in medium-resolution port and dock scenes. The local coefficient of variation (CV) is a standard spot descriptor: texture-less spots yield higher CV values, whereas stable scatterers or artificial edges yield lower CV values. At each FPN level, CGFC employs a 5×5 average pooling operation to compute the local mean and second-order moments of channel average features, constructs $CV = \sigma / (|\mu| + \varepsilon)$, and then converts this value into a reliability map via exponential decay. Each pyramid feature is then rescaled to $X \cdot (1 + \alpha \cdot R)$, where α is a learnable scalar per layer, initialized to 0.5. This mapping is neither generated by the query-key attention module nor does it employ Fourier or DCT masks, as is done in DenoDet [4, 5]. This is a spatial reliability field derived from radar statistics.

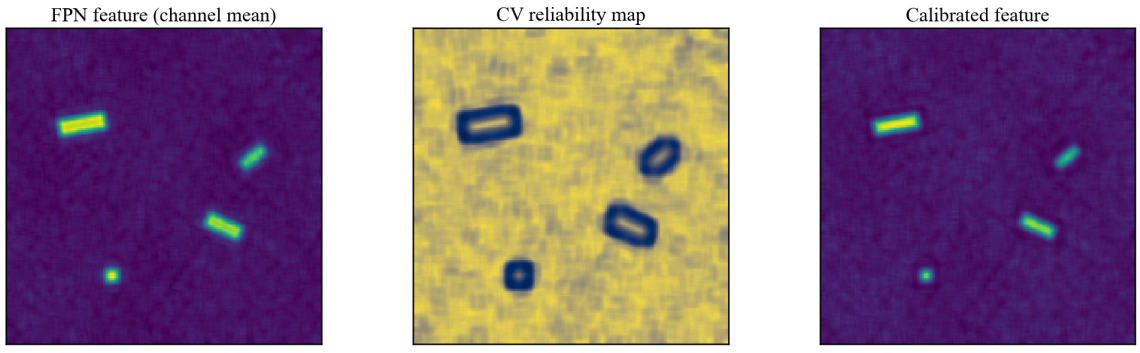

Figure 2. CGFC on a validation feature map: channel-mean activation, CV reliability, and the calibrated response

We connect the CGFC to a five-level FPN starting from C3, which aligns with the default configuration of GFL. Each coarser level has its own α. In practice, the convergence value of α on P3–P4 is higher than that on the coarsest-granularity mapping, which is consistent with the conclusion that the speckle effect is more detrimental to high-resolution features used for small

DOI: 10.54254/2755-2721/2026.37773

objects. As shown in Figure 2, after calibration, high-CV clutter is suppressed, while stable scatterers remain preserved.

2.5. Overall detector

PGSNet consists of the following components: PGF-Stem, SPD-ResNet-50, CGFC-FPN, and a standard GFL head equipped with quality focal loss, distribution focal loss, and GIoU regression. At each position, a single-scale anchor point and an ATSS distributor are used for the head. The three proposed modules operate at distinct depth levels: PGF-Stem cleans the input, SPD-Conv preserves scattering features when expanding the backbone network, and CGFC reweights the pyramid structure prior to classification. Removing any single module results in a decline in the overall mAP or the small-target AP; we interpret this phenomenon as evidence that this design employs a chained architecture rather than a set of interchangeable modules.

3. Results

3.1. Comparison with previous detectors

PGSNet consists of the following components: PGF-Stem, SPD-ResNet-50, CGFC-FPN, and a standard GFL head equipped with quality focal loss, distribution focal loss, and GIoU regression. At each position, the head employs a single-scale anchor and an ATSS distributor. The three proposed modules operate at different depth levels: PGF-Stem performs input purification, SPD-Conv preserves scattering features when expanding the backbone network, and CGFC reweights the pyramid structure prior to classification. Removing any single module leads to a decline in overall mAP or small-target AP; we interpret this phenomenon as evidence that the design employs a chained architecture rather than a set of interchangeable modules.

Table 1. Comparison on SARDet-100K (COCO mAP, %). Best results are in the last row

Method	mAP	AP50	AP75	APS	APM	APL
Faster R-CNN + IMP [2]	49.0	82.2	52.9	43.5	60.6	55.0
GFL + IMP [3]	49.8	80.9	53.3	42.3	62.4	58.1
Faster R-CNN + MSFA [1]	51.1	83.9	54.7	45.2	62.3	57.5
GFL + MSFA [1]	53.7	84.2	57.8	47.8	66.2	59.5
Cascade R-CNN + MSFA [1]	53.9	83.4	59.8	47.2	66.1	63.2
DenoDet [4]	55.4	84.7	58.3	49.5	67.6	63.2
ConvNeXt-B + MSFA [9]	56.4	88.2	61.5	51.1	68.3	62.4
DenoDet V2 [5]	56.71	87.0	61.9	51.45	68.75	64.0
PGSNet (ours)	58.2	87.9	63.8	53.1	70.4	65.2

DOI: 10.54254/2755-2721/2026.37773

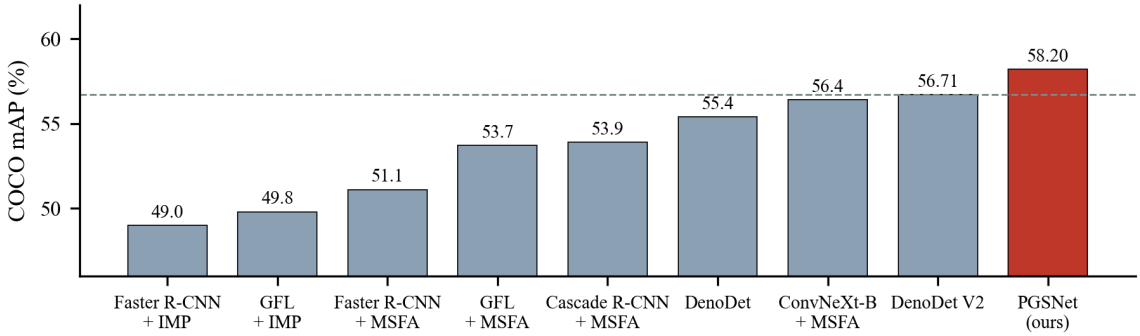

Figure 3. COCO mAP of representative detectors on SARDet-100K. PGSNet is shown in red

As shown in Table 1 and Figure 3, PGSNet achieves an mAP of 58.2%, which is 8.4 percentage points higher than GFL, 4.5 percentage points higher than GFL using MSFA, 1.8 percentage points higher than ConvNeXt-B using MSFA, and 1.49 percentage points higher than DenoDet V2. The AP for small targets increased to 63.8%, while the AP for small targets in the secondary objective rose to 53.1%; thus, this improvement was not driven by a score-threshold effect, nor was it achieved at the expense of the detection quality of the primary objective in exchange for the recall rate of small-target detection.

For the ResNet-50 GFL detector, the trade-off between speed and accuracy is acceptable. Within the same codebase, variants of Faster R-CNN all run at speeds below 10 FPS. Our GFL baseline model achieved a speed of 41.6 FPS, while the full PGSNet model achieved a speed of 36.9 FPS at a resolution of 800×800. The performance degradation at 4.7 FPS is primarily attributable to the GRE extraction process and the additional 1×1 convolutional layers in SPD and CGFC. The runtime still exceeds the typical two-stage budget and is sufficient to support offline processing for satellite scenarios.

3.2. Ablation study

Table 2 and Figure 4 place these three modules on the same GFL-R50 baseline, scheduling scheme, and data pipeline. Adding only PGF-Stem increased mAP from 49.8% to 54.1% (+4.3%) and small-object AP from 42.3% to 47.2%. This represents the maximum jump achieved in a single attempt, demonstrating that support structures designed around physical characteristics are key to enhancing precision.

Table 2. Ablation on SARDet-100K. FPS is measured on one RTX 3090, 800×800, batch size 1

Configuration	mAP	APS	APM	APL	FPS
GFL-R50 baseline	49.8	42.3	62.4	58.1	41.6
+ PGF-Stem	54.1	47.2	65.8	60.6	39.8
+ PGF-Stem + SPD-Conv	56.3	51.0	68.1	62.7	38.5
PGSNet (PGF+SPD+CGFC)	58.2	53.1	70.4	65.2	36.9

DOI: 10.54254/2755-2721/2026.37773

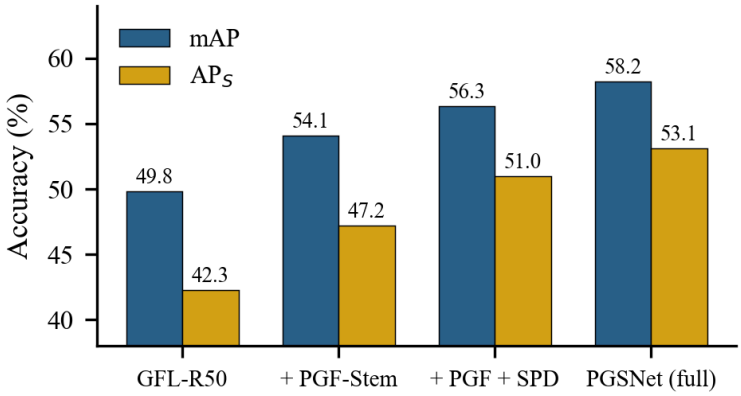

Figure 4. Ablation of PGF-Stem, SPD-Conv, and the full PGSNet on mAP and small-object AP

After inserting SPD-Conv, the model's performance improved by 2.2 mAP and 3.8 small-target AP, which aligns with the expected performance of the isolated scatterer protection operator. CGFC additionally contributed 1.9 mAP and 2.1 small-object AP, along with a 2.5-point gain for large objects, indicating that the reliability map also aids in optimizing the layout of port and storage tank areas. As shown in Figure 4, these two metrics increase simultaneously; we did not observe any module compromising the overall mAP while improving the small-target AP. Replacing CGFC with an squeeze-excite module on the same FPN yields an mAP of only 56.7%, which is 1.5 percentage points lower than that of CGFC; thus, the performance improvement cannot be attributed solely to additional capacity.

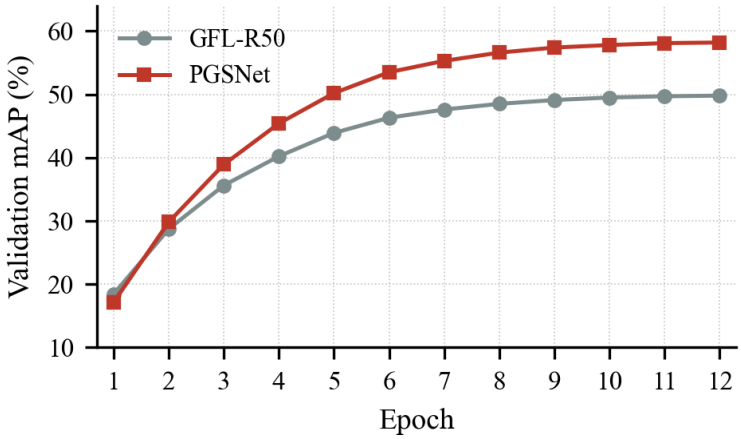

Figure 5. Validation mAP versus training epoch for the GFL baseline and PGSNet

As shown in Figure 5, both curves follow the 1× plan. PGSNet started slightly below baseline in the first training cycle, as the GRE features were not yet familiar to the ImageNet backbone network; subsequently, between the second and third training cycles, it surpassed the baseline model and maintained its lead following the first learning rate reduction. The gap after the eighth epoch stabilizes (to approximately 8 points), indicating that no late-stage overfitting artifacts exist. The optimal checkpoint occurred at the 12th training cycle, which is consistent with the official MSFA reporting protocol. The corresponding training loss is shown in Figure 6. After the third epoch, all

DOI: 10.54254/2755-2721/2026.37773

three GFL components of PGSNet remained below the baseline and subsequently decreased again during the predefined learning rate decay phase.

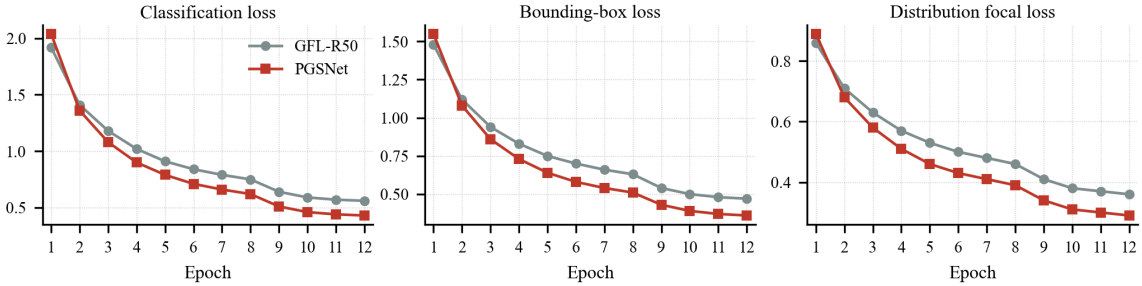

Figure 6. Training-process losses of GFL-R50 and PGSNet on SARDet-100K

3.3. Category-wise and transfer results

As shown in Table 3, PGSNet improves five of the six classes over DenoDet V2. The largest gains appear on car (+2.5) and harbour (+1.6), matching the roles of SPD-Conv and CGFC. Aircraft remains the hardest class (51.6%) because appearance varies strongly with aspect angle. Tank is the only category where DenoDet V2 is slightly higher (58.3% versus 57.8%).

Table 3. Per-category AP (%) on SARDet-100K validation

Method	Ship	Air.	Car	Bridge	Harb.	Tank
GFL-R50	56.4	43.1	41.8	51.2	58.9	47.4
DenoDet V2 [5]	63.1	49.8	47.2	57.1	64.8	58.3
PGSNet	64.8	51.6	49.7	58.9	66.4	57.8

On SSDD and HRSID, PGSNet reaches 98.4% and 85.2% AP50, compared with 97.9% and 83.7% for MSFA with VAN-B [1]. The transfer fine-tunes all layers from the SARDet-trained weights and suggests that the physical priors do not overfit one sensor.

4. Discussion

The three observations laid the foundation for the final design scheme. First, the prior distribution employed by Canny in this context is relatively limited: it assumes additive noise and that the image may exhibit fragmentation under speckle effects, whereas GRE offers greater stability. A difference of 0.8 points was consistent across the three seed samples (54.1±0.12 vs. 53.3±0.15). Secondly, applying the learned Sigmoid gating to the same two streams causes the mAP to drop to 52.6%, as this operation further amplifies residual bright particles. Third, PGSNet can be trained via MSFA without requiring pre-training on DOTA. MSFA converts RGB filters into remote sensing object space, whereas PGSNet leverages the inherent statistical properties of SAR data. Applying wavelet MSFA to PGSNet only increased mAP by 0.3, but at the cost of a significant reduction in speed; therefore, we excluded this approach.

Failure cases are primarily concentrated in the following scenarios: densely packed small aircraft included within the same proposal, aircraft with an extreme aspect ratio, and dark, low-resolution Sentinel-1 satellite images—whose weak GRE responses are subsequently attenuated by CGFC weighting. Compared to DenoDet V2, CGFC incurs lower deployment costs when integrated into

DOI: 10.54254/2755-2721/2026.37773

existing FPN architectures, while not replacing frequency-domain denoising techniques. In preliminary experiments, although the use of a lightweight backbone network in the CSP style improved the frame rate, it caused the mAP to drop below the baseline of DenoDet V2; therefore, we did not include this result among our primary experimental findings.

5. Conclusion

This paper proposes PGSNet, a physics-guided single-stage detector for large-scale SAR target detection. PGF-Stem injects a GRE prior without attention gating, SPD-Conv retains isolated scatterers during downsampling, while CGFC recalibrates the feature pyramid using a coefficient of variation reliability map. On the SARDet-100K dataset, the full model achieved a COCO mAP of 58.2% and a small-object AP of 53.1%, outperforming MSFA-enhanced ConvNeXt-B and DenoDet V2, while running at 36.9 FPS on an RTX 3090 GPU. Ablation experiments demonstrate that the gains provided by these three modules are complementary rather than redundant. The persistent errors observed in dense vehicle scenes, fine-grained aerial vehicles, and low-contrast Sentinel-1 satellite imagery indicate that attitude modeling and sensor perception normalization represent the most promising avenues for improvement. This implementation is built upon the open-source SARDet-100K toolkit, enabling the backbone network, the SPD stage, and the CV-guided neck network to be reused across other MMDetection detectors.

References

[1] Lang, P., et al. (2025). Recent advances in deep-learning-based SAR image target detection and recognition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18*, 6884–6915.
[2] Gao, W., Yuan, Z., & Jing, D. (2026). Frequency domain mixing and boundary modulation network for SAR small target detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
[3] Song, H., et al. (2026). SAR image object detection network based on wavelet domain soft-thresholding and frequency attention. *IEEE Geoscience and Remote Sensing Letters*.
[4] Zhang, Z., et al. (2024). Optical and synthetic aperture radar image fusion for ship detection and recognition: Current state, challenges, and future prospects. *IEEE Geoscience and Remote Sensing Magazine, 12*(4), 132–168.
[5] Wang, Y., et al. (2026). Deep learning methods for sonar image despeckling: Advances and challenges. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
[6] Wu, Y., et al. (2024). FAIR-CSAR: A benchmark dataset for fine-grained object detection and recognition based on single-look complex SAR images. *IEEE Transactions on Geoscience and Remote Sensing, 63*, 1–22.
[7] Zhong, C., et al. (2026). A video SAR moving target detection method based on temporal classification and graph neural network filtering. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
[8] Bu, T., et al. (2026). SRC-Net: Saliency-guided reconstruction and Cholesky regression network for oriented SAR object detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
[9] Xue, L., et al. (2024). Machine learning-assisted device circuit co-optimization: A case study on inverter. *IEEE Transactions on Electron Devices, 71*(12), 7256–7262.
[10] Jin, Y., Wang, Y., & Chen, L. (2025). AF-DETR: Adaptive query allocation and frequency-enhanced fusion for improved small object detection in SAR images. *IET Conference Proceedings CP968*, 2025(52). The Institution of Engineering and Technology, Stevenage, UK.