

基于层自适应稀疏剪枝策略的高效SAR舰船目标检测方法研究

韦栋辉*#

长江大学电子信息与电气工程学院, 湖北 荆州

收稿日期: 2026年3月18日; 录用日期: 2026年4月21日; 发布日期: 2026年6月5日

摘要

合成孔径雷达(SAR)舰船检测在海洋监视中至关重要,但现有的高精度模型(如YOLO11)参数量大、计算复杂,难以部署在星载或边缘计算等资源受限平台。针对上述问题,文章提出了一种面向SAR任务的YOLO11结构化剪枝框架SASP (SAR-Aware Structured Pruning)。该框架对原始模型进行了深入的架构感知冗余性分析,量化了C3k2模块内部及跨层连接中的特征相关性,精准定位了冗余结构。通过对基准模型进行多梯度的通道剪枝与充分微调,实验结果表明适度的剪枝操作能有效去除网络冗余特征,产生显著的正则化效应。该方法不仅大幅降低了计算成本,更实现了检测精度的逆势增长:在2.5加速比下,达到了89.9% mAP50高精度的同时实现了近1000 FPS的实时推理速度。研究证明了结构化剪枝在特定策略下可作为一种高效的正则化手段,为资源受限场景下的实时高精度目标检测提供了新的解决方案。

关键词

合成孔径雷达, YOLO11, 结构化剪枝, 模型轻量化

Research on an Efficient SAR Ship Target Detection Method Based on Layer-Adaptive Sparse Pruning Strategy

Donghui Wei*#

School of Electronic Information and Electrical Engineering, Yangtze University, Jingzhou Hubei

Received: March 18, 2026; accepted: April 21, 2026; published: June 5, 2026

*第一作者。

#通讯作者。

文章引用: 韦栋辉. 基于层自适应稀疏剪枝策略的高效 SAR 舰船目标检测方法研究[J]. 仪器与设备, 2026, 14(2): 201-207. DOI: 10.12677/iae.2026.142023

Abstract

Synthetic Aperture Radar (SAR) ship detection is crucial in maritime surveillance, but existing high-precision models (such as YOLO11) have large parameter counts and computational complexity, making them difficult to deploy on resource-constrained platforms such as spaceborne or edge computing. To address these issues, this paper proposes a YOLO11 structured pruning framework, SASP (SAR-Aware Structured Pruning), for SAR tasks. This framework performs in-depth architecture-aware redundancy analysis on the original model, quantifies the feature correlations within the C3k2 module and across layer connections, and accurately locates redundant structures. Through multi-gradient channel pruning and sufficient fine-tuning of the benchmark model, experimental results show that appropriate pruning operations can effectively remove network redundant features and produce a significant regularization effect. This method not only significantly reduces computational costs but also achieves a counter-trend increase in detection accuracy: at a speedup ratio of 2.5, it achieves a high accuracy of 89.9% mAP50 while realizing a real-time inference speed of nearly 1000 FPS. This study demonstrates that structured pruning, under specific strategies, can serve as an efficient regularization method, providing a new solution for real-time, high-precision target detection in resource-constrained scenarios.

Keywords

Synthetic Aperture Radar, YOLO11, Structured Pruning, Lightweight Model

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

海洋监测在现代国防安全和民用经济中扮演着不可或缺的角色。从维护国家海洋权益、打击海盗与非法捕捞，到保障海上交通安全和执行紧急搜救任务，实时、准确的海洋态势感知能力至关重要。在众多对地观测手段中，合成孔径雷达凭借其独特的成像机制，展现出显著优势。与依赖光照条件且易受云雾雨雪干扰的光学和红外传感器不同，SAR 是一种主动式微波成像系统，具备全天候和全天时的工作能力。这使得 SAR 能够在恶劣气象条件或夜间环境下，依然稳定地获取海面舰船信息，填补了光学遥感的盲区[1]。

然而，将先进的深度学习检测算法实际应用于移动平台(如侦察无人机、小型巡逻艇或微纳卫星)仍面临严峻挑战。这些应用场景通常要求算法在嵌入式设备上运行，并满足严格的实时性指标。以单阶段目标检测模型 YOLO11 [2]架构为例，该模型是 Ultralytics 于 2024 年 9 月发布的最新一代 YOLO 模型。作为 YOLOv8 的继任者，YOLO11 在保持高效率的同时进一步提升了检测精度，引入了 C3k2 模块、C2PSA 注意力机制等创新设计，并在多任务支持、模型部署等方面进行了全面优化，其模型结构如图 1 所示。现有的高性能 SAR 检测网络往往为了追求极致的平均精度均值，在其骨干网络中堆叠了深层的 C3k2 模块及复杂的注意力机制。这种设计虽然显著提升了特征表达能力，却也导致模型推理速度较慢。且这种“计算密集型”特性与边缘设备有限的存储容量、计算能力形成了矛盾，使得原始的 YOLO11 模型难以直接部署于资源受限的星载或机载终端[3]-[5]。

在许多紧急任务中，一个虽然精度稍低但能快速给出结果的轻量级模型，往往比一个精度高但需

要数秒钟才能完成推理的庞大模型更具实际价值。鉴于此，本文提出了一种基于模型结构化剪枝的 YOLO11 轻量化重构方法。该方法旨在精准识别并剔除 YOLO11 架构中的冗余通道与滤波器，在最大限度保留其强大特征提取能力的前提下，显著降低计算复杂度。通过这种瘦身策略，致力于打破精度与效率之间的权衡，为资源受限环境下的 SAR 舰船实时检测提供一套兼具高精度与高效率的可行解决方案[6]-[9]。

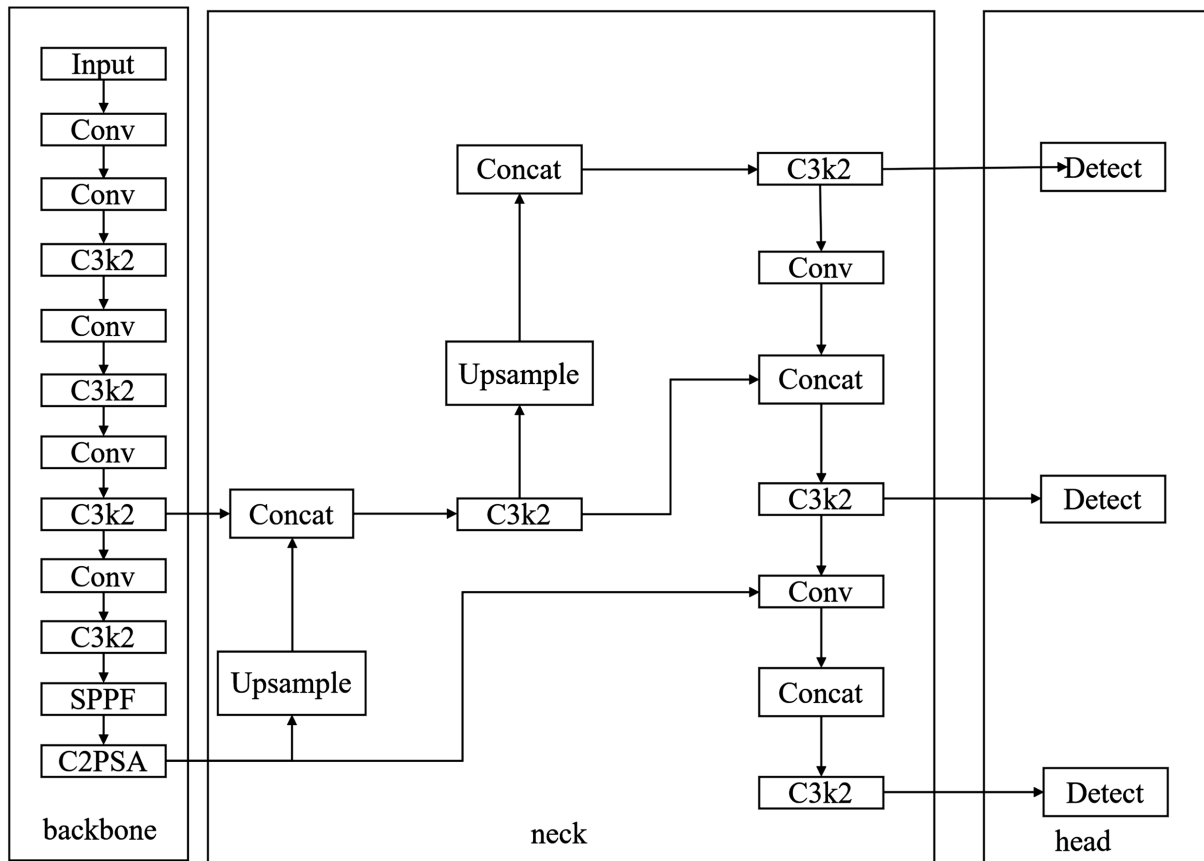


Figure 1. YOLO11 model structure diagram

图 1. YOLO11 模型结构图

2. 方法

2.1. 整体框架概述

本文提出了一种面向边缘端 SAR 舰船检测的自适应结构化剪枝框架 SASP，旨在对 YOLO11 架构进行轻量化重构。如图 2 所示，该方法并非简单地移除权重，而是一个包含“冗余性分析 - 重要性评估 - 结构化剪枝 - 微调”的闭环优化过程。

具体流程如下：首先，预训练一个高精度的 YOLO11 基线模型；其次，设计一种 SAR 感知的重要性评估准则，量化每个卷积通道对舰船特征提取的贡献度；随后，依据评估得分剔除冗余通道，生成结构紧凑的稀疏网络；最后，通过一次性微调策略恢复并提升模型性能。该框架的核心优势在于能够在大幅降低参数量和浮点运算量的同时，最大程度保留 YOLO11 对 SAR 图像中微弱舰船目标的特征表达能力。

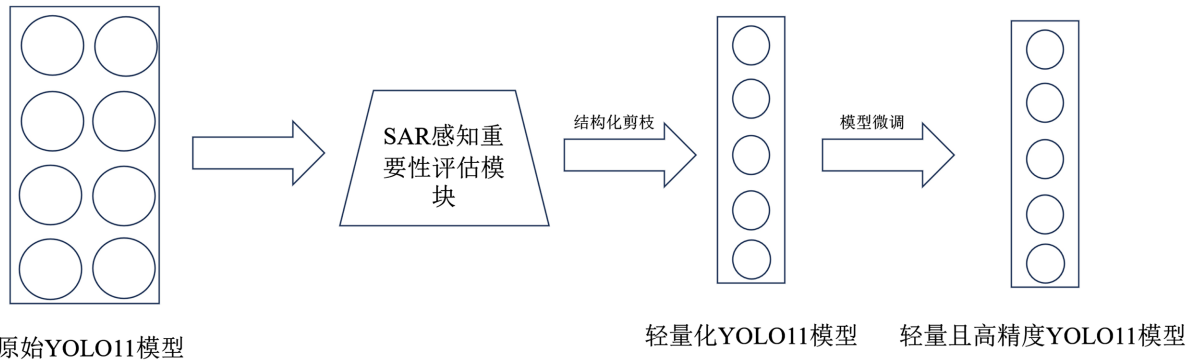


Figure 2. Overall framework diagram
图 2. 整体框架图

2.2. 基于幅值的层自适应稀疏性剪枝方法(LAMP)

YOLO11 模型广泛采用卷积操作，特别是在 C3k2 结构中级联了多层卷积核。这种深度的层级堆叠不可避免地导致了通道间的高度相关性与激活稀疏性，表明模型中存在大量非关键计算路径，使得通过剪枝移除冗余结构成为提升效率的可行策略。

LAMP [10]的重要性评估原则可以概括为“基于全局能量分布的相对重要性”，即它不再孤立地看待单个通道的权重范数，而是将其置于整个网络的权重能量分布背景中进行评估。一个通道权重重组的重要性，取决于移除它相对于保留所有比它更大的通道权重重组所造成的“相对失真”程度。因此，LAMP 定义的重要性分数 $S_{LAMP}(w_i)$ 本质上是该通道权重重组的能量占“所有大于等于该权重的能量总和”的比例。对于某一层的所有通道权重重组参数展平并排序，其中排序后的第 i 个参数 w_i 的重要性分数定义如下：

$$S_{LAMP}(w_i) = \frac{|w_i|^p}{\sum_{j:|w_j| \geq |w_i|} |w_j|^p}$$

其中， p 表示范数的阶数，默认取值为 2。

重要性分数越接近 1，表示该通道在网络中的重要性越大；反之，则表示该通道在网络中的重要性较低，可以优先剪枝。通过这个公式，LAMP 成功地将一个通道对应的权重重组转化为了一个相对的比率，即该通道在全局能量分布中的地位，从而实现了无需人工干预的层自适应剪枝。在完成所有通道的重要性评分后，将评分序列按升序排列(长度为 N)。根据预设的目标剪枝率 α ，计算需剪除的通道数量 $K = \lfloor N \times \alpha \rfloor$ 。选取排序后第 K 个位置的评分值作为全局截断阈值 T 。最终执行策略为：保留所有评分 $S > T$ 的通道，剪除所有评分 $S \leq T$ 的通道，从而实现精确的全局稀疏度控制。

剪枝操作虽然显著降低了模型复杂度，但也不可避免地破坏了原有的特征提取结构，导致初始精度下降。因此，微调(Fine-Tuning)是剪枝流程中不可或缺的闭环步骤。在此阶段，将基于剪枝后的稀疏拓扑结构，在原始训练数据集上采用较小的学习率进行重训练。这一过程旨在重新校准剩余权重的参数分布，使网络适应新的稀疏架构，从而最大限度地补偿精度损失，甚至在某些情况下通过正则化效应实现精度的反超。

3. 实验及结果分析

3.1. 数据集介绍

HRSID (High-Resolution SAR Images Dataset) [11]是一个专门用于高分辨率合成孔径雷达(SAR)图像中船舶检测、语义分割和实例分割任务的开源数据集，包含 5604 张高分辨率 SAR 切片图像，大多数图像大小为

800×800 像素, 含有超过 16,000 个船舶实例标注。其中训练集包含 3642 张图片, 测试集包含 1962 张图片。

3.2. 实验平台及参数设置

软硬件平台配置如表 1 所示。本研究以 YOLO11n 作为基准架构(Baseline)。首先, 在输入分辨率 640×640 的条件下对原始模型进行全量训练, 以确立性能基准。随后, 基于该基准模型开展结构化剪枝实验。加速比(Speed-Up)的设定对应于模型参数量或计算复杂度(GFLOPs)的预期降幅。本研究分别构建了模型参数量或计算量缩减至原模型 1/2.0、1/2.5、1/3.0、1/3.5 及 1/4.0 的五种剪枝变体, 以此形成不同压缩强度的实验梯度, 旨在全面评估模型在各类轻量化约束下的性能表现。所有实验均统一采用以下训练配置: 优化器选用 SGD (动量设为 0.937), 初始学习率为 0.01, 批次大小(Batch Size)为 16, 数据加载线程数(Workers)为 4, 总训练轮次(Epochs)设定为 400。

Table 1. Experimental platform configuration

表 1. 实验平台配置

软硬件平台	配置
操作系统	Win11
CPU	I5 12,600 kf
显卡	RTX 5060ti
Python	3.10
PyTorch	2.8.0
CUDA	12.9

3.3. 评价指标

本研究采用多维度的评价指标体系: 在检测精度方面, 选用平均精度均值(mAP@50 和 mAP@50:95)作为核心指标; 在模型复杂度方面, 通过参数量(Parameters)和浮点运算数(GFLOPs)评估模型规模; 在推理效率方面, 则利用 GPU 帧率(FPS)综合衡量模型的实时推理性能。

3.4. 实验结果

Table 2. Results of the pruning experiment

表 2. 剪枝实验结果

Model	Speed-Up	mAP50	mAP50:95	Parameters (M)	GFLOPS	FPS
YOLO11n		88.2	64.3	2.582	6.3	530.5
exp1	2.0	90.5	67	1.301	3.1	843.2
exp2	2.5	89.9	66.4	1.040	2.4	958.8
exp3	3.0	89.7	65.8	0.860	2.0	1066.1
exp4	3.5	89.6	65.4	0.776	1.8	1128.3
exp5	4.0	89.3	64.6	0.693	1.6	1337.4

表 2 展示了不同压缩比例下的实验结果。与传统认知中“剪枝导致精度下降”不同, 本研究发现适度的结构化剪枝能显著提升模型的检测性能。具体而言, exp1 (Speed-Up = 2.0)取得了最佳的综合性能, 其 mAP50 达到 90.5%, mAP50:95 达到 67%, 较基准模型 YOLO11n 分别提升了 2.3%、2.7%。这一现象可归因于剪枝操作充当了有效的正则化器, 通过移除网络中的冗余通道和噪声特征, 增强了模型在测试

集上的泛化能力，从而缓解了过拟合问题。在推理效率方面，剪枝带来了巨大的收益。随着压缩比的增加，模型的计算复杂度(GFLOPs)和参数量显著降低。例如，exp5 (Speed-Up = 4.0)将推理速度从 530.5 FPS 提升至 1337.4 FPS，同时保持了 89.3%的 mAP50，甚至优于原始模型。

综合考量精度与速度，exp2 (Speed-Up = 2.5)被视为本研究的最佳平衡点。它在仅牺牲 0.6% mAP50 (相对于 exp1)的情况下，实现了近 1000 FPS 的实时推理速度，且模型体积缩小了约 61.9%，非常适合部署在对延迟敏感的边缘计算设备上。exp2 模型的检测结果如图 3 所示。

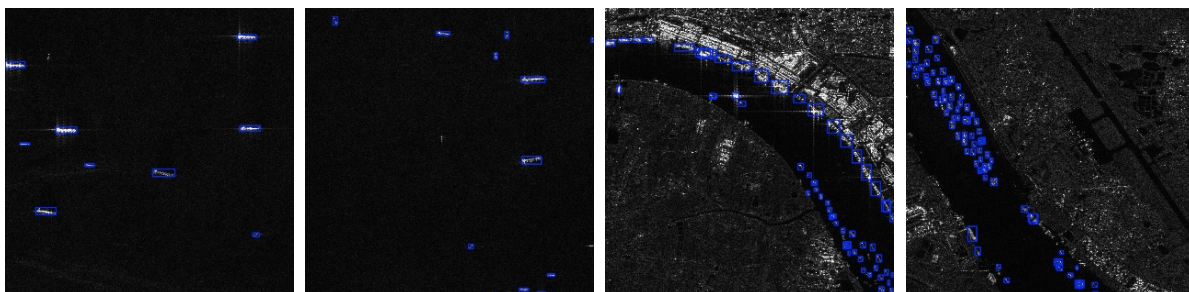


Figure 3. Visualization of detection results
图 3. 检测结果可视化

3.5. 特征可视化分析

如图 4 所示，原始稠密模型不仅在舰船目标上激活，还对海杂波背景有响应，表明可能存在对噪声的过拟合。相比之下，剪枝后的模型表现出更尖锐的激活图，高度集中在目标结构上。这表明 LAMP 剪枝有效地移除了对背景噪声敏感的冗余通道，起到了一种强大的正则化作用，迫使网络学习更鲁棒和更具判别力的特征。

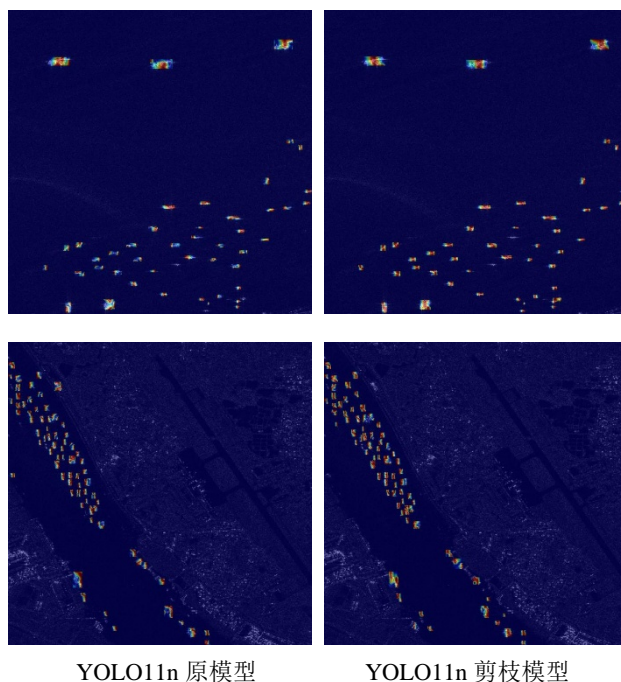


Figure 4. Visual comparison of the thermal effects of the model before and after pruning
图 4. 剪枝前后模型热效应可视化效果对比

4. 结论

本文提出了一种基于幅值的层自适应稀疏的 YOLO11n 结构化剪枝框架，旨在解决目标检测模型在资源受限设备上部署难的问题。通过对基准模型进行多梯度的通道剪枝与微调，本研究得出了以下主要结论：

首先，实验结果揭示了剪枝操作的正则化潜力。与传统观点认为剪枝必然导致精度损失不同，本研究发现适度的结构化剪枝能有效去除网络冗余噪声，显著提升模型的泛化能力。在 Speed-Up 为 2.0 的实验组中，剪枝模型的 mAP50 和 mAP50-95 分别达到 90.5% 和 67%，较原始 YOLO11n 模型分别提升了 2.3% 和 2.7%，实现了精度与效率的“双赢”。

其次，该框架在推理效率上取得了突破性进展。随着剪枝比例的增加，模型计算复杂度显著降低。在最高 4.0 倍加速比下，模型参数量缩减至 0.693 M (减少约 73%)，GFLOPs 降至 1.6 G，而推理速度从基准的 530.5 FPS 飙升至 1337.4 FPS，增幅达 152%。即便在此极端压缩下，检测精度仍优于未剪枝的基准模型，证明了该方法在保持高性能的同时具备极强的轻量化能力。

最后，本研究为边缘端实时检测提供了最优解决方案。综合考量精度保持率与加速收益，Speed-Up 为 2.5 的模型被确定为最佳平衡点，其在维持 89.9% 高精度的同时实现了近 1000 FPS 的实时推理速度。未来的工作将聚焦于将该剪枝策略迁移至更多样的骨干网络架构，并探索其在超低功耗嵌入式设备(如 MCU、FPGA)上的硬件协同优化部署。

参考文献

- [1] 车永辉. 基于 SAR 图像的海上目标识别算法研究[J]. 现代传输, 2026(1): 32-36.
- [2] Jocher, G., Qiu, J. and Chaurasia, A. (2024) Ultralytics YOLO (Version 8.3.9) [Computer Software]. <https://github.com/ultralytics/ultralytics>
- [3] 罗德艳, 王明刚, 徐杨. 基于时频特征融合的轻量化 SAR 图像目标检测[J/OL]. 无线电工程, 1-10. <https://link.cnki.net/urlid/13.1097.TN.20260130.1729.008>, 2026-03-16.
- [4] 闫晨宇, 耿亮, 杜伟伟, 等. 一种 FCDIS-YOLOv11s 轻量化 SAR 图像智能检测方法[J]. 指挥控制与仿真, 2026, 48(1): 45-54.
- [5] 赵梓翔, 王樱洁, 张华春. 基于改进 YOLOv11n 的轻量化 SAR 图像舰船目标检测方法[J]. 科学技术创新, 2025(20): 208-212.
- [6] 韩萍, 白继睿, 周杰龙, 等. 基于轻量化改进和模型剪枝的 SAR 图像飞机目标检测[J]. 仪器仪表学报, 2025, 46(9): 110-124.
- [7] 李卓阳, 郭荣佐, 车思韬, 等. 基于剪枝的 SAR 图像舰船目标检测模型压缩方案[J]. 计算机应用与软件, 2025, 42(5): 255-262+281.
- [8] 张帅. 基于深度学习的 SAR 图像舰船目标检测模型研究[D]: [硕士学位论文]. 大连: 大连理工大学, 2024.
- [9] 邱春, 路红, 秦彬鑫, 等. 基于模型剪枝和知识蒸馏的船舶目标检测方法[J]. 南京工程学院学报(自然科学版), 2022, 20(2): 13-17.
- [10] Lee, J., Park, S., Mo, S., *et al.* (2020) Layer-Adaptive Sparsity for the Magnitude-Based Pruning. arXiv: 2010. 07611.
- [11] Wei, S., Zeng, X., Qu, Q., Wang, M., Su, H. and Shi, J. (2020) HRSID: A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation. *IEEE Access*, 8, 120234-120254. <https://doi.org/10.1109/access.2020.3005861>