

光电工程

Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

融合空-频域的动态SAR图像目标检测

沈学利, 王嘉慧, 吴正伟

引用本文:

沈学利, 王嘉慧, 吴正伟. 融合空-频域的动态SAR图像目标检测[J]. *光电工程*, 2025, 52(1): 240245.

Shen X L, Wang J H, Wu Z W. Dynamic SAR image target detection by fusing space-frequency domain[J]. *Opto-Electron Eng*, 2025, 52(1): 240245.

<https://doi.org/10.12086/oe.2025.240245>

收稿日期: 2024-10-19; 修改日期: 2024-12-03; 录用日期: 2024-12-10

相关论文

合成孔径激光雷达成像技术研究进展

徐晨, 晋凯, 魏凯

光电工程 2024, 51(3): 240007 doi: 10.12086/oe.2024.240007

远场合成孔径计算光学成像技术: 文献综述与最新进展

李晟, 王博文, 管海涛, 梁坤瑶, 胡岩, 邹燕, 张许, 陈钱, 左超

光电工程 2023, 50(10): 230090 doi: 10.12086/oe.2023.230090

基于双分支多尺度融合网络的毫米波SAR图像多目标语义分割方法

丁俊华, 袁明辉

光电工程 2023, 50(12): 230242 doi: 10.12086/oe.2023.230242

更多相关论文见光电期刊集群网站 



<http://cn.oejournal.org/oe>



 OE_Journal



Website

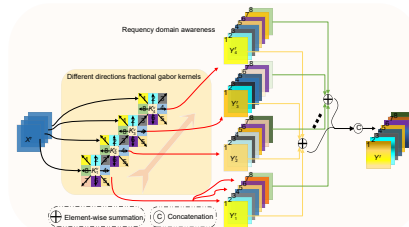
DOI: 10.12086/oe.2025.240245

CSTR: 32245.14.oe.2025.240245

融合空-频域的动态 SAR 图像目标检测

沈学利, 王嘉慧*, 吴正伟

辽宁工程技术大学软件学院, 辽宁 葫芦岛 125105



摘要: 针对合成孔径雷达 (synthetic aperture radar, SAR) 图像样本特征差异大、目标尺度不均衡、背景散斑噪声高所导致的检测精度低、推理速度慢问题, 提出一种融合空-频域的动态 SAR 图像目标检测算法。首先, 采用分流感知策略构造空-频域感知单元, 结合动态感受野及分数阶 Gabor 变换法, 增强算法对空间多样性特征和频率散射特征的捕获能力与感知力, 优化模型对全局上下文信息的保留能力, 加快推理速度, 降低特征映射模式相似性与背景噪声干扰, 有效改善漏检、误检情况。其次, 采用重参数学习方法设计自适应特征融合模块, 优化多尺度特征间的交互与整合, 丰富特征的多样性, 缓解特征采样引起的差异映射与信息丢失问题, 加强小目标信息与关键频率信息在融合过程中的显著性, 提高多尺度样本检测精度。最后, 引入 DY_IoU 动态回归损失函数, 利用自适应尺度惩罚因子与动态非单调注意力机制解决锚框膨胀和位置偏差问题, 进一步增强模型对多尺度目标的定位与检测能力, 加快模型收敛速度, 减少模型计算量。在公开数据集 SAR-Acraft-1.0 和 HRSID 上进行相关实验, 实验结果表明: 该方法 mAP@0.5 数值达到了 95.9% 和 98.8%, 较基线模型分别提升 5.2% 和 1.2%, 且优于其他对比算法。表明该算法显著提升了检测精度, 具备良好的鲁棒性与泛化性。

关键词: SAR 图像; 分流感知; 分数阶 Gabor 变换法; 特征融合; 多尺度样本; 小目标; DY_IoU

中图分类号: TP391.4

文献标志码: A

沈学利, 王嘉慧, 吴正伟. 融合空-频域的动态 SAR 图像目标检测 [J]. 光电工程, 2025, 52(1): 240245

Shen X L, Wang J H, Wu Z W. Dynamic SAR image target detection by fusing space-frequency domain[J]. *Opto-Electron Eng*, 2025, 52(1): 240245

Dynamic SAR image target detection by fusing space-frequency domain

Shen Xueli, Wang Jiahui*, Wu Zhengwei

School of Software, Liaoning Technical University, Huludao, Liaoning 125105, China

Abstract: A dynamic SAR image target detection algorithm integrating space-frequency domains is proposed to address challenges such as significant feature differences in SAR image samples, imbalanced target scales, and high speckle noise in the background, which result in low detection accuracy and slow inference speed. First, a dual-stream perception strategy constructs spatial-frequency perception units, leveraging dynamic receptive fields and fractional-order Gabor transforms to enhance the model's ability to capture spatial diversity and frequency scattering features. This way improves the retention of global contextual information, accelerates inference, reduces the similarity of feature mapping patterns, and mitigates background noise interference, effectively

收稿日期: 2024-10-19; 修回日期: 2024-12-03; 录用日期: 2024-12-10

基金项目: 国家自然科学基金面上项目 (62173171)

*通信作者: 王嘉慧, 2245414310@qq.com。

版权所有©2025 中国科学院光电技术研究所

reducing missed and false detections. Second, a re-parameterization-based adaptive feature fusion module is designed to optimize interactions across multi-scale features, enriching feature diversity, alleviating mapping discrepancies and information loss caused by feature sampling, and enhancing the salience of small target and key frequency information during fusion, thereby improving detection precision. Finally, the DY_IoU dynamic regression loss function is introduced, utilizing adaptive scale penalty factors and a dynamic non-monotonic attention mechanism to address anchor box expansion and positional deviation, further enhancing the localization and detection capabilities for multi-scale targets. This way also accelerates model convergence and reduces computational overhead. Experiments conducted on the public datasets SAR-Acraft-1.0 and HRSID demonstrate that the proposed method achieves mAP@0.5 values of 95.9% and 98.8%, respectively, representing 5.2% and 1.2% improvements over baseline models and outperforming other comparison algorithms. These results indicate that the proposed algorithm improves detection accuracy and exhibits strong robustness and generalization capabilities.

Keywords: SAR image; dual-stream perception; fractional-order Gabor transform method; feature fusion; multi-scale samples; small target; DY_IoU

1 引言

合成孔径雷达 (SAR) 是一种有源微波成像遥感设备, 具备全天候、全天时、不受天气影响的特点。SAR 技术通过发射微波信号并接收反射回波, 利用运动平台的相对位置产生合成孔径, 从而获得地表高分辨率图像。SAR 图像因其独特的成像机制和丰富的电磁散射特征, 被广泛应用于海洋监测、土地覆盖、灾害调查、军事等领域^[1-3]。然而, 在目标检测领域中, SAR 图像检测面临较大挑战, 因为 SAR 图像中的目标样本尺度不一, 形状多变, 易受散斑噪声和复杂背景的干扰。因此, 提高 SAR 图像检测精度, 加快检测算法推理速度是当前研究的热点和重点。

随着深度学习技术的不断发展, 基于深度学习的 SAR 图像目标检测算法得到了广泛研究。与传统的手工提取特征不同, 深度学习通过多层神经网络自适应地从数据中提取特征, 无需人工干预, 使得模型能够捕捉到图像中更丰富和复杂的映射信息, 因此具有更高的准确性和鲁棒性。基于深度学习的目标检测算法主要分为两类: 双阶段和单阶段。

以 Faster R-CNN^[4] 为代表的双阶段算法通过级联 RPN (region proposal network) 与 CNN (convolutional neural network) 实现对图像样本的精细化分类与检测。单阶段算法将 RPN 与 CNN 进行结合, 实现了端到端的检测方式, 简化了检测流程, 显著提升算法检测效率并降低模型资源消耗。代表性单阶段算法如 YOLO (you only look once) 系列^[5-9]、SSD (single shot multiBox detector) 等^[10]。

为进一步提高 SAR 图像目标检测模型的整体性

能, Zhang 等^[11] 提出了 YOLO-FA 算法, 通过引入频域注意力模块, 自适应处理 SAR 图像中的频域信息, 抑制背景噪声, 进而提高检测性能。Si 等^[12] 提出 YOLO-RSSD 算法, 在颈部特征融合部分嵌入改进的双向特征金字塔网络结构, 并在卷积单元中增加通道注意力机制, 增强目标特征信息捕获能力, 但在大尺度场景下的检测效果仍不佳。Qin 等^[13] 提出了一种基于自蒸馏的半软标签引导 SAR 舰船检测网络, 相比经典的硬标签或软标签, 该方法更能为目标提供引导知识, 但在复杂场景下检测效果不佳。Xu 等^[14] 针对 SAR 图像复杂场景中的正负样本不均衡及定位精度不佳问题, 提出了 A3-IOUSNet 网络模型, 通过自适应锚框分配正负样本, 并在检测头增加交并比监督机制改善目标定位质量, 增强检测性能, 但误检率略高。Xiao 等^[15] 在 YOLOv7 算法中引入 U-Net Denoising 去噪模块及上下文信息提取模块 (contextual transformer block, COT), 有效抑制相干斑噪声对图像的干扰, 改善漏检情况, 但增加了额外的计算量和参数量。Sun 等^[16] 提出了改进 YOLOv5 算法, 利用 GhostConv 优化网络主干, 降低模型参数量, 并引入 CBAM (convolutional block attention module) 注意力机制提高特征融合质量, 但对小目标检测效果不佳。Li 等^[17] 提出了高效的轻量级检测网络 SFS-CNet, 通过分流感知选择策略增加每个卷积层内的特征多样性, 显著增强模型对多尺度特征的提取能力, 但算法的检测准确性不足。Zhou 等^[18] 提出了一种 SAR 图像飞机目标检测网络 (DiffDet4SAR), 利用散射特征增强模块 (scattering feature enhancement module, SFEM) 有效减少了杂波

干扰并增强目标显著性, 但算法的泛化能力受限, 收敛能力较差。

为解决上述问题, 本文在 YOLOv10s 算法基础上, 提出了一种融合空-频域的动态 SAR 图像目标检测算法。首先, 构建空-频域感知 (spatial-frequency domain sensing, SFDS) 单元, 设计并利用全局空间感知 (global spatial awareness, GSA) 模块、频域感知 (frequency domain awareness, FDA) 模块, 增强算法对多尺度代表性特征与纹理变化频率特征的捕获能力及全局语义信息的保留能力, 改善散斑噪声引起的误检情况, 减少冗余映射资源消耗, 加快模型推理速度。其次, 设计自适应特征融合 (adaptive feature fusion, AFF) 单元, 利用重参数学习法, 自适应整合多尺度特征映射, 丰富特征多样性, 优化全局空间特征与频率散射特征间的协同与融合, 增强小尺度样本信息的保留能力, 进一步提高算法的检测精度。最后, 使用动态回归损失函数 (dynamic iou, DY_IoU), 增强算法对多尺度目标的定位与识别能力, 加快模型收敛速度, 降低模型参数量。

2 YOLOv10s 模型

YOLOv10s^[19] 算法由主干网络 (back-bone)、颈部

网络 (neck)、检测头部 (head) 三部分共同构成。YOLOv10s 算法结构图如图 1 所示。为满足主干网络的输入条件, 算法对输入图像进行预处理, 将原图像大小缩减至 $640 \text{ pixel} \times 640 \text{ pixel}$ 。

主干网络由 2 个标准卷积模块 CBS (conv+bN+siLU)、2 个空间通道解耦采样单元 (spatial-channel decoupling downsampling, SCDown)、3 个 C2f 模块和 1 个 C2fCIB 模块构成。CBS 模块采用核值为 3×3 , 步长为 2 的卷积核进行特征采样。同样作为采样单元的 SCDown, 利用点卷积 PWC (pointwise convolution)^[20] 与深度卷积 (depthwise convolution, DWC)^[21] 对特征提取过程进行了解耦, 在降低模型计算量的同时, 最大限度地增强了下采样期间的算法信息保留能力。C2f 模块利用标准卷积 CBS、残差结构^[22] 与 Bottleneck 块设计多分支整合单元, 进一步增强模型对多尺度目标的特征捕获能力, 加快反向传播时梯度的流动。C2fCIB 模块通过在 C2f 基础上引入紧凑倒置结构 (compact inverted block, CIB) 替换 Bottleneck 单元, 进一步提高了模型推理效率与检测精度, 解决了因下采样引起的特征差异问题。

颈部网络采用 PAFPN (path aggregation feature

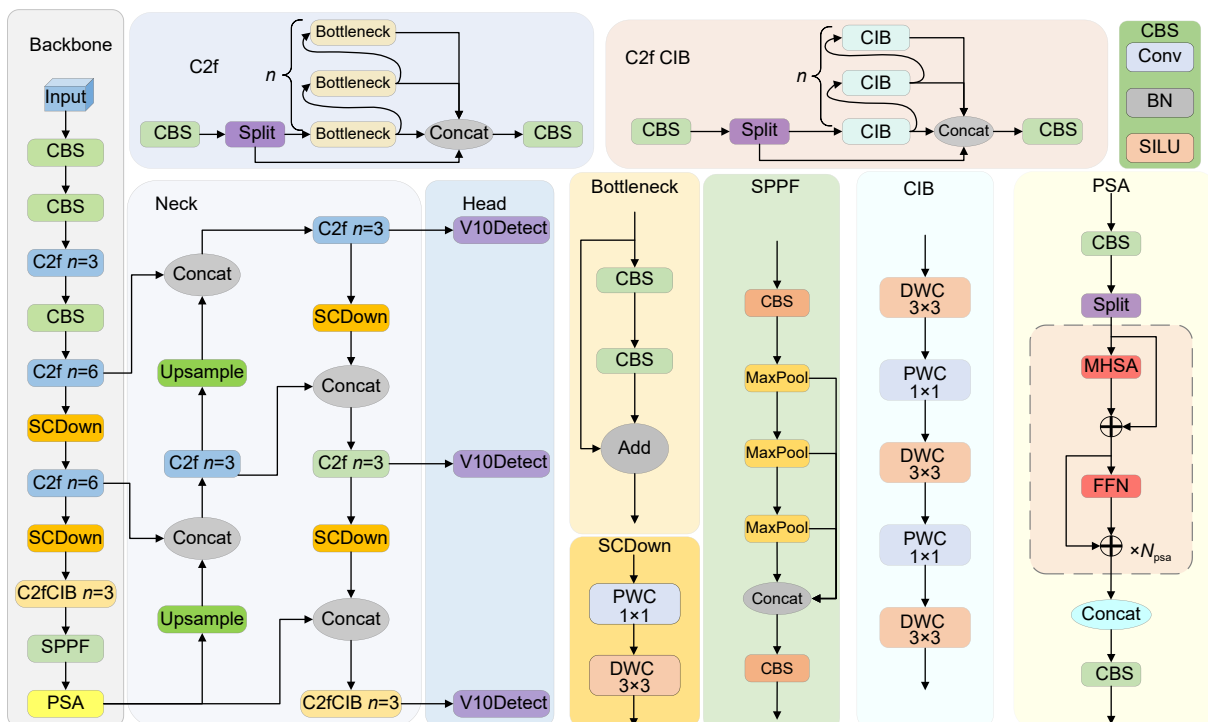


图 1 YOLOv10s 算法结构图

Fig. 1 YOLOv10s algorithm structure diagram

pyramid networks) 结构, 包含 1 个 CBS 模块、SPPF (spatial pyramid pooling-fast) 结构、3 个 C2f 模块、1 个 C2fCIB 模块、1 个 SCDown 下采样单元、4 个 Concat 拼接单元、2 个 Upsample 上采样单元与 1 个轻量化的部分自注意力机制 (partial self-attention, PSA)^[23] 模块。其中 PSA 采用两个 1×1 卷积对跨通道特征进行均匀分割与融合, 加强特征的交互与融合。并利用多头自注意力机制 (multihead self-attention, MSA)^[24] 与前馈网络 (feedforward neural network, FNN)^[25] 搭建 N_{psa} 单元, 使网络能够自关注更多的重要特征信息, 增强其对关键信息的判别能力, 提高模型的检测性能, 减少模型的计算开销。

YOLOv10s 采用 3 种尺度检测头对颈部网络输出的待检测图像进行分类和回归预测, 并给出各样本检测结果。为提高检测精度与效率, YOLOv10s 在分类头 (classification head) 上采用了轻量化架构, 使用 2 个深度可分离卷积 (depthwise separable convolution, DSConv)^[26] 替换原标准卷积。同时, 采用双标签分配机制与非极大值抑制 (non-maximum suppression, NMS)^[27] 的训练策略, 进一步提升了模型检测性能。

3 融合空-频域的动态 SAR 图像目标检测

为解决 SAR 图像目标检测算法精度低、推理速度慢等问题, 本文基于 YOLOv10s 提出融合空-频域的动态 SAR 图像目标检测算法, 结构如图 2 所示。

相较于原模型, 动态 SAR 图像目标检测算法主要做出了以下改进: 1) 构造空-频域感知 SFDS 单元替换部分 CBS 模块与 SCDown 下采样单元。利用全局空间感知模块 GSA 中的动态感受野与频域感知模块 FDA 中的分数阶 Gabor 变换法, 增强算法对关键性特征与变化频率特征的判别力与感知力, 改善对细粒度特征信息的识别与保留能力, 加快模型推理速度, 抑制背景冗余, 降低模型参数量与计算量; 2) 设计自适应特征融合单元 AFF, 替换 Concat 操作。运用重参数学习法提高多样化特征间的整合效率, 加强算法对重要映射信息的敏感度, 提高特征的表示能力与算法的检测精度; 3) 检测头处采用动态边界回归损失函数 DY_IoU 替换 CIou 损失, 结合额外惩罚项与非单调注意力机制解决位置偏差问题, 进一步增强模型对多尺度目标的定位与检测能力, 降低模型计算量。

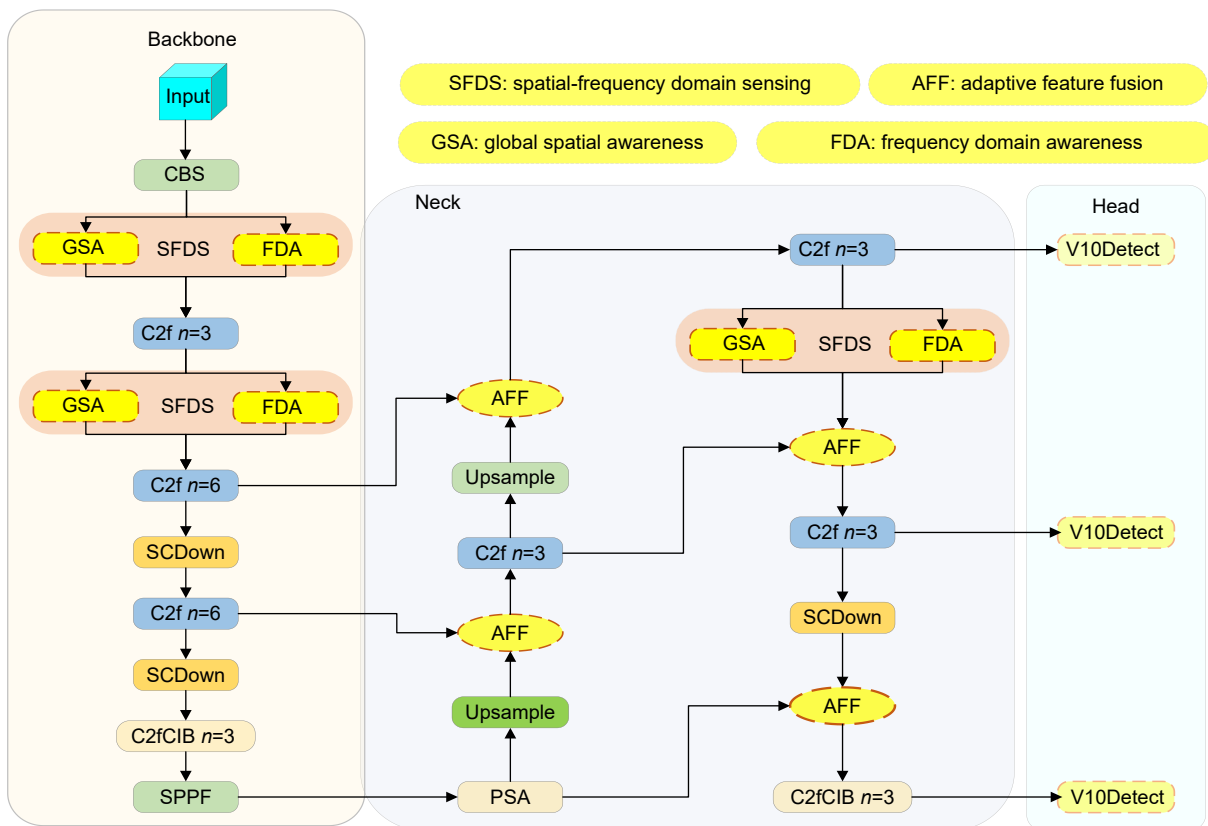


图 2 融合空-频域的动态 SAR 图像目标检测算法结构图

Fig. 2 Structure of the dynamic SAR image target detection algorithm fusing spatial-frequency domain

3.1 空-频域感知单元

SAR 图像的拍摄视角多变, 导致样本特征呈现多尺度和密集分布。同时, 受回波信号成像机制的影响, 图像中包含大量用于表示目标散射特征的频率分量。采用局部运算法则的标准卷积, 因固定感受野和缺乏对上下文建模能力, 仅能提取样本的基本纹理与边缘信息, 难以捕捉多尺度目标的细节特征和全局结构, 易受背景噪声干扰, 且无法有效捕获频率特征, 进而影响模型的检测性能与稳定性。为此, 本文采用分流感知策略, 设计空-频域感知 SFDS 单元, 结合动态感受野及分数阶 Gabor 变换法, 实现多尺度特征与多方向散射特征的全面提取和充分学习, 加快模型推理速度, 提高差异特征的融合效率。SFDS 单元结构如图 3 所示。

对于输入尺寸为 $W \times H$, 通道数为 C 的特征图 X , 首先采用分流操作, 沿通道维度进行分流压缩, 以减少冗余信息, 降低计算成本。分流过程如下:

$$X^S = X^{\delta C \times W \times H}, \quad (1)$$

$$X^F = X^{(1-\delta)C \times W \times H}, \quad (2)$$

式中: δ 表示分流比 ($0 < \delta < 1$), 本文 $\delta = 0.5$; $X^S \in \mathbb{R}^{\delta C \times W \times H}$ 表示具有多尺度、多变性和不同结构的空间特征; $X^F \in \mathbb{R}^{(1-\delta)C \times W \times H}$ 表示频率分量的多向偏振与散射特征。

随后, 利用点卷积 PWC (pointwise convolution) 对 X^S 和 X^F 中的代表特征进行提炼, 进一步增强模型对样本特征的代表能力, 减少特征冗余, 降低计算复杂度。再将经特征提炼后的 X^S 和 X^F 分别输入全局空间感知模块 GSA 和频域感知模块 FDA 中, 依据动态感受野技术及分数阶 Gabor 变换法获取多尺度全局空

间特征 $Y^S \in \mathbb{R}^{W \times H \times C}$ 和丰富频域特征 $Y^F \in \mathbb{R}^{W \times H \times C}$ 。GSA 能够捕捉全局空间中的长程依赖关系, 有效识别目标在复杂场景下的全局分布特征, 从而增强模型对多尺度目标形状和结构的辨别能力; FDA 模块专注于不同频率分量的特征学习与增强, 通过选择性地保留目标的关键频率成分, 抑制背景噪声和伪影干扰, 使模型能够更好地提取目标的散射特征与边缘细节。

最后借鉴空间通道重构卷积 (spatial and channel reconstruction convolution, SCConv)^[28] 的加权选择融合策略, 增加网络对多尺度空间特征 Y^S 和不同频域特征 Y^F 的感知力与判别力, 提高模型再复杂场景下的检测精度和鲁棒性, 缓解背景噪声干扰问题。具体而言, 首先使用全局平均池化 (global average pooling, GAP)^[29] 操作, 沿通道维度采集 Y^S 和 Y^F 全局范围内的空间和频率信息 $P^N \in \mathbb{R}^{C \times 1 \times 1}$, $N \in \{S, F\}$, 过程如下:

$$P^N = \text{GAP}(Y^N) = \frac{1}{H \times W} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} Y_{i,j}^N. \quad (3)$$

继而使用通道软注意力机制^[30] 对 P^S 、 P^F 叠加后的特征生成选择性权重 $\eta, v \in \mathbb{R}^C$ 。最后在选择性权重的指导下, 完成 Y^S 和 Y^F 的自适应融合, 得到包含全局上下文信息与丰富频率特征映射的特征图 $X' \in \mathbb{R}^{W \times H \times C}$ 。上述过程表示为

$$\begin{cases} \eta = \frac{e^{P^S}}{e^{P^S} + e^{P^F}}, \\ v = \frac{e^{P^F}}{e^{P^S} + e^{P^F}}, \\ \eta + v = 1, \end{cases} \quad (4)$$

$$X' = \eta \times Y^S + v \times Y^F. \quad (5)$$

3.1.1 全局空间感知 (GSA) 模块

为增强模型对多尺度特征的感知力和差异映射信

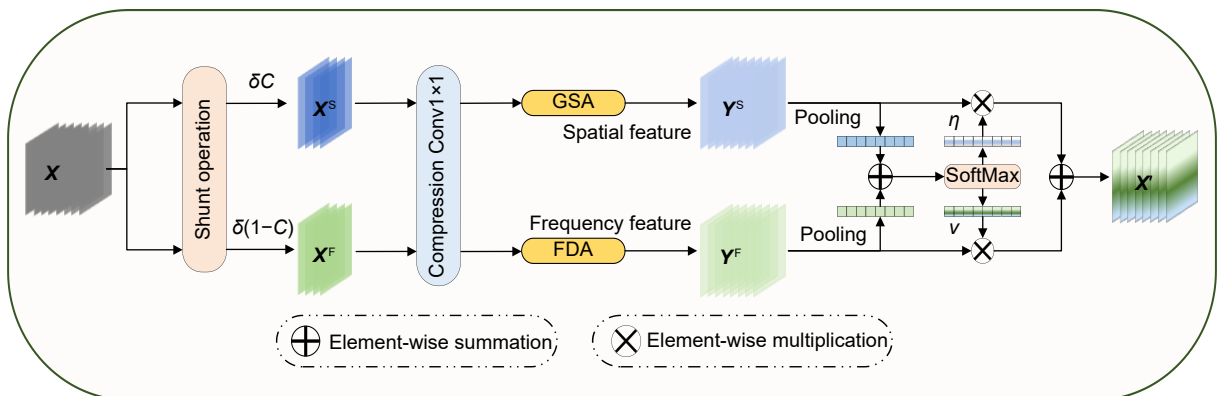


图 3 SFDS 结构

Fig. 3 Structure of SFDS

息的判别能力, 本课题组设计了全局空间感知模块 GSA, 通过动态感受野实现跨尺度动态建模上下文信息与全局空间特征的智能学习, 进而加强算法对关键目标的定位、识别能力, 并在减少背景冗余的同时降低漏检率。全局空间感知模块 GSA 结构如图 4 所示。

具体而言, 本文对提炼后的特征 X^S 通道进行等分操作, 得到 r 个特征映射组 X_t^S , 其中 $t \in \{1, 2, 3, \dots, r\}$ 。每个特征映射组 X_t^S 都与 X^S 具有相同的空间大小, 通道数为 X^S 的 $1/r$ 。为充分提取 X^S 中全局空间特征, 本文为每组 X_t^S 分配对应的独立卷积核 $K_t \in \mathbb{R}^{C \times k_t \times k_t \times C}$ 。其中, 核值大小 k_t 如下:

$$\begin{cases} k_{t+1} = k_t + 2 \\ k_1 = 3 \end{cases} \quad (6)$$

此外通过在卷积核之间构建高层次残差连接, 进一步扩展卷积核的感受野, 实现上下文信息的动态建模与感受野的自适应优化。过程如下:

$$RF_{t+1} = RF_t + (k_{t+1} - 1), \quad (7)$$

式中: RF_t 表示第 t 组卷积核对应的感受野。

由于每组卷积核 k_t 都能接受前序特征组 $\{X_i^S, i < t\}$ 的特征映射信息, 因此经特征学习后的输出特征组 Y_t^S 有着比前序组 $\{Y_i^S, i < t\}$ 更大的感受野。输出特征组 Y_t^S 表示如下:

$$Y_t^S = \begin{cases} X_t^S * K_t, t = 1 \\ (X_t^S + Y_{t-1}^S) * K_t, 1 < t \leq r \end{cases} \quad (8)$$

式中: $*$ 表示卷积运算操作。

最后将各组输出特征 Y_t^S 沿通道进行拼接, 并采用与输出通道数 C 相同个数的 1×1 卷积 $K_{i,C}$ 对其多尺度上下文信息进行融合, 得到全局空间细粒度特征 $Y^S \in \mathbb{R}^{W \times H \times C}$ 。过程如下:

$$Y^S = \sum_{i=0}^{C-1} (Y_1^S \odot Y_2^S \odot Y_3^S \odot \dots \odot Y_i^S)_i * K_{i,C}, \quad (9)$$

式中: \odot 表示拼接操作; i 表示每个输出通道的索引。

利用动态感受野处理多尺度特征, 能使模型自适应关注差异尺度的映射变化, 捕获丰富且有价值的视觉上下文信息。在减少资源开销的同时, 进一步增强算法对全局空间特征的定位与检测能力。

3.1.2 频域感知 (FDA) 模块

因 SAR 系统成像机制与目标运动之间的耦合效应引发的多普勒频移问题^[31], 给 SAR 图像目标检测带来了巨大挑战。分数阶傅里叶变换 (fractional Fourier transform, FrFT)^[32] 利用非线性旋转坐标转换, 将不同目标频率分量映射到不同时频平面, 以实现多普勒频移效应的精准描述与分离, 在加强算法特征表示力的同时, 有效抑制散射特征混叠与成像模糊问题。Gabor 变换 (gabor transform, GT)^[33] 是一种方向敏感的时频分析方法, 其使用不同方向和尺度的 Gabor 滤波器提取图像的局部高频特征, 解决了因目标旋转、尺度变化引起的方向特征丢失与失配问题。本文结

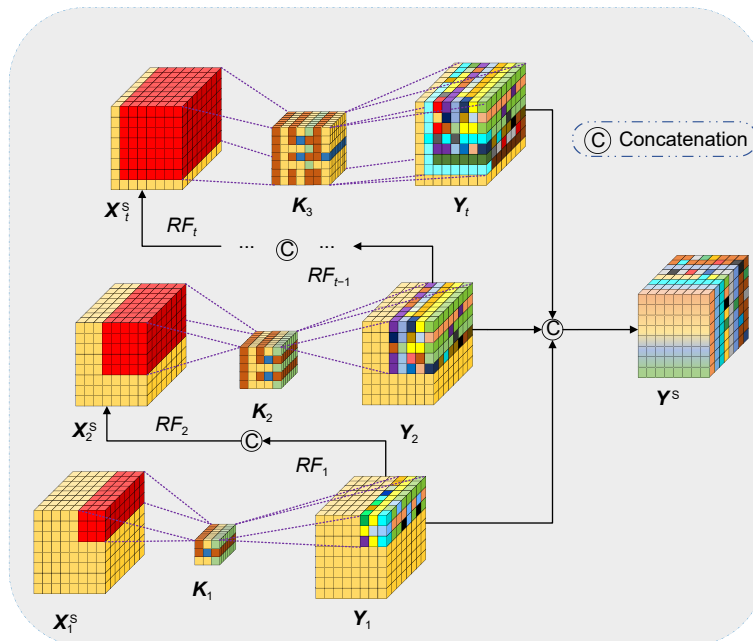


图 4 全局空间感知 (GSA) 模块结构

Fig. 4 Global spatial awareness (GSA) module structure

合 FrFT 与 GT, 提出分数阶 Gabor 变换法 (fractional gabor transform, FrGT), 并利用该方法设计频域感知模块 FDA。FDA 通过多向卷积核与尺度扩张, 提高算法对频率特征与纹理特征的捕获能力并进一步加强模型对多尺度特征的学习能力, 降低背景噪声干扰的同时有效改善了误检情况。频域感知模块 FDA 结构如图 5 所示。

分数阶 Gabor 变换法 (FrGT) 利用分数阶傅里叶变换 (FrFT) 核函数与 Gabor 窗口函数对输入特征 $X \in \mathbb{R}^{W \times H \times C}$ 进行多尺度、多方向的频域特征提取与空间频域间的高效转换。转换过程为两步计算: 首先对输入特征的行像素进行内层变换, 将其映射为 FrFT 频域特征, 并利用 Gabor 窗口函数调节每个位置的频率响应, 生成不同尺度及方向的局部频率特征。随后对所有行频率特征应用 FrFT 核函数进行外层加权求和, 再利用窗口函数进行二次方向权重调制, 以获取特征图在 FrGT 变换域中的全局频率特征表示。该过程表示为

$$G_f^\alpha(x, y, u, v) = \sum_{i=0}^{H-1} \sum_{m=0}^{V-1} B\left(i, \frac{m}{VT_1}, \alpha\right) \bar{g}(i-m) \left[\sum_{j=0}^{W-1} \sum_{n=0}^{U-1} f(i, j) B\left(j, \frac{n}{UT_2}, \alpha\right) \bar{g}(j-n) \right], \quad (10)$$

式中: $G_f^\alpha(x, y, u, v)$ 为 FrGT 核函数, 用于将时域或空间特征转换为 FrGT 频域特征; x, y 为输入特征空间坐标; u, v 为 FrFT 变换域中的频率坐标; $f(i, j)$ 为空间像素值; i, j 分别为水平、垂直坐标; W, H 分别为特征图高度与宽度; V, U 为 FrFT 核函数在水平和垂直方向上的样本数量, 通过调整 V, U 可以控制 FrFT 在两个方向上的分辨率; T_1, T_2 为采样间隔; $B(Y, Q, \alpha)$ 为 FrFT 核函数, 其利用旋转角度 α 实现空间特征到频域特征的转换, 其中 $\alpha = P\pi/2$, P 为变换阶数, Y, Q 表示空间和 FrFT 域坐标; $g(\cdot)$ 为 Gabor 窗口函数, 用于控制 FrFT 核函数在特定行或列上的局部响应; $-$ 表示共轭。 $B(Y, Q, \alpha)$ 核函数与 $g(\cdot)$ 窗口函数如下式所示:

$$B(Y, Q, \alpha) = \sqrt{1 - i \cot \alpha} \cdot \exp [i\pi(Y, Q) \cot \alpha - 2i\pi Y Q \csc \alpha], \quad (11)$$

$$g(x, y) = \exp\left(-\frac{x^2 + \gamma^2 y^2}{2\sigma^2}\right) \cdot \exp[2\pi i(v_0 x + u_0 y)], \quad (12)$$

式中: γ 为纵横比参数, 用于调整 Gabor 函数在空间坐标方向上的延展比例; v_0, u_0 为频率分量, 用于调制 Gabor 函数的频率响应。

为使同一卷积层内的标准卷积能够捕获不同尺度、方向和频率特征, 本文利用 FrGT 核函数调制标准卷积核, 调制过程如下:

$$K_{m,q}^{v'} = K_{m,o} * G(v', u'), \quad (13)$$

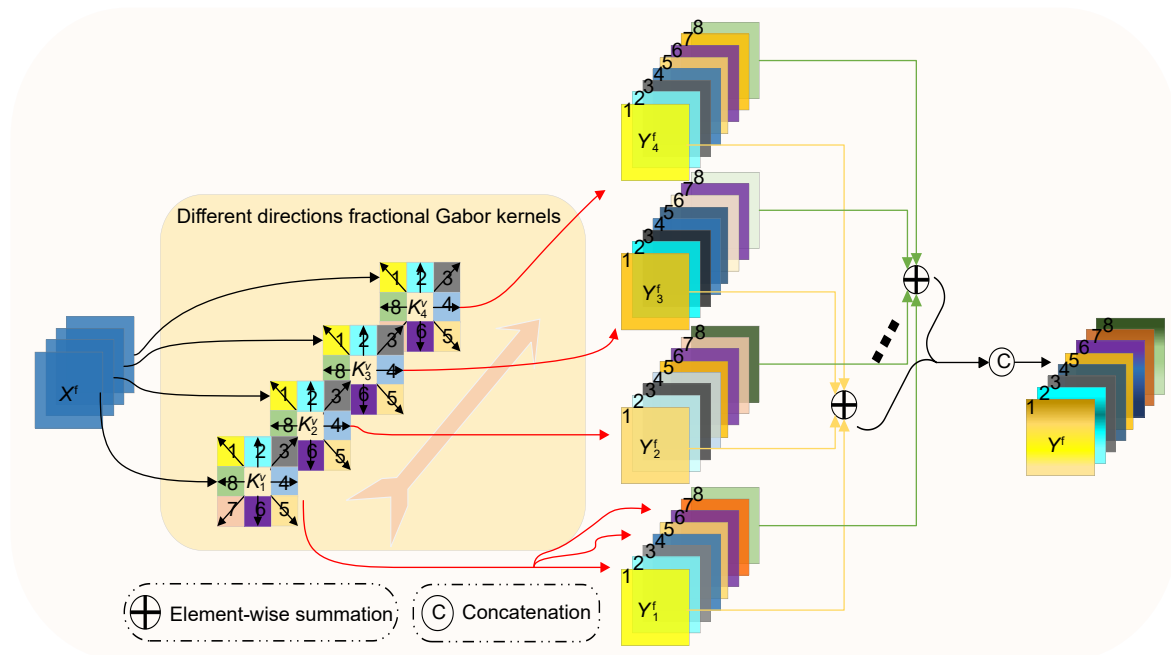


图 5 频域感知 (FDA) 模块结构
Fig. 5 Structure of frequency domain awareness (FDA) module

式中: $K_{m,q}^v$ 为调制后的卷积分数 Gabor 核 (convolutional fractional gabor kernels, CFrGK), 用于提取丰富纹理特征与多方向频率特征; m, q 为输入输出通道数; v', u' 分别为水平和垂直方向的频率分量, $u' \in \{0, U-1\}, v' \in \{0, V-1\}$; $K_{m,o}$ 表示核值大小为 $k \times k$, 输入输出通道数为 m, o 的标准卷积核; $G(v', u')$ 表示一组不同方向和尺度的 FrGT 核函数。

由于 FrGT 具有旋转等变性, 因此可在随机水平频率分量 v' 下共享 FrGK 参数。本文将 FrGK 通道数设置为 Gabor 方向数 U 的 N 倍。因此, 具有 m 个输入通道和 $U \cdot N$ 个输出通道的卷积核 $K_{m,U \cdot N}^v$ 通过拼接 N 个相同尺度但不同方向的 FrGK 卷积核得到, 表达式为

$$K_{m,U \cdot N}^v = K_{m,U_0}^v \odot K_{m,U_1}^v \odot K_{m,U_2}^v \odot \dots \odot K_{m,U_{N-1}}^v \quad (14)$$

对于通道数为 C 的特征 X^f , 本文将其分为 E 个组 $X_{E'}^f$, 其中 $E' \in \{0, E-1\}$ 。每个 $X_{E'}^f$ 需要 $N = C/E \times U$ 个卷积核, 以生成对应频率特征 $Y_{E'}^f$, 表示如下:

$$Y_{E'}^f = X_{E'}^f \cdot K_{m,C/E}^v \quad (15)$$

最终, 将各个 $Y_{E'}^f$ 进行拼接, 得到频率精细化特征 Y^f 。

$$Y^f = Y_0^f \odot Y_1^f \odot Y_2^f \odot \dots \odot Y_{E-1}^f \quad (16)$$

3.2 自适应特征融合单元

针对 SAR 图像目标尺度不均衡导致的检测精度低问题, 当前检测算法采用逐元相加法或通道拼接法, 优化不同尺度信息间的交流与融合, 提高特征的表达质量, 加强模型检测性能, 两种方法如图 6 所示。

但二者由于缺少特征间相关性与复杂性的自主学习, 导致在特征融合过程网络无法精准的捕捉到部分浅层纹理信息与深层语义信息, 从而出现漏检与误检问题。为此, 本文设计自适应特征融合 (adaptive feature fusion, AFF) 模块, 结构如图 7 所示。

AFF 利用重参数学习法, 实现不同尺度特征图间

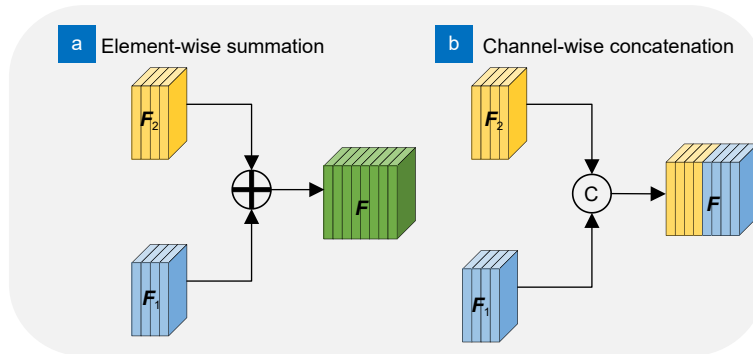


图 6 当前特征融合方法。(a) 逐元相加法; (b) 通道拼接法

Fig. 6 Current feature fusion methods. (a) Element-by-element summation method; (b) Channel splicing method

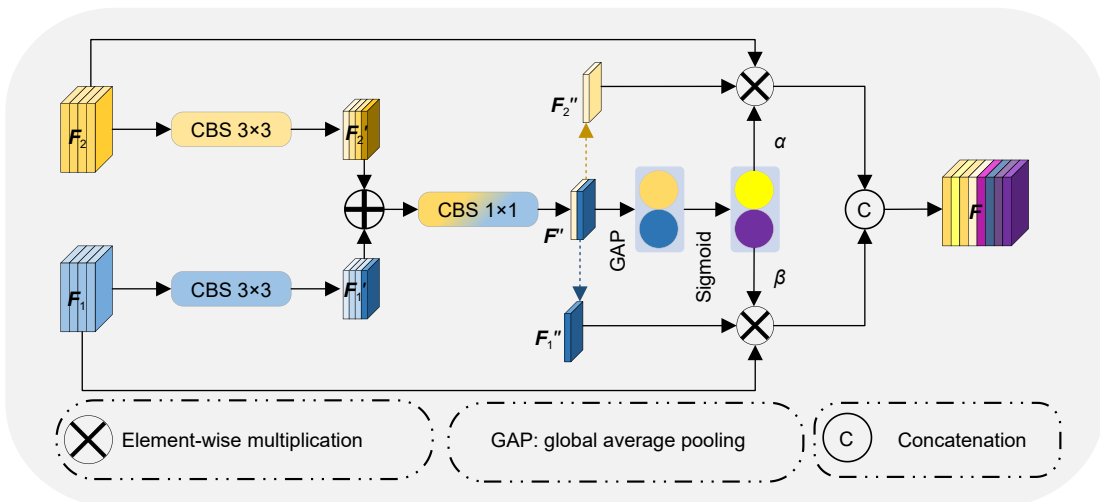


图 7 自适应特征融合 (AFF) 模块结构

Fig. 7 Adaptive feature fusion (AFF) module structure

的高效融合与动态调优, 提高算法精测精度。具体而言, 首先采用核值大小为 3×3 的 CBS 卷积单元, 对相邻层特征图 $F_1 \in \mathbb{R}^{H \times W \times C}$ 和 $F_2 \in \mathbb{R}^{H \times W \times C}$ (H 为特征图宽度, W 为特征图高度, C 为特征图通道数) 进行下采样, 使网络能够捕获特征图中的样本复杂模式与空间关联性, 提高算法对重要语义信息和高级映射信息的感知度, 增强模型对关键特征的识别与定位能力。随后, 对采样结果 $F'_1 \in \mathbb{R}^{H \times W \times C}$ 和 $F'_2 \in \mathbb{R}^{H \times W \times C}$ 进行拼接操作, 并利用 1×1 卷积对拼接后的特征映射进行降维, 以降低模型的复杂度与计算量, 生成代表性特征图 $F'' \in \mathbb{R}^{W \times H \times 2}$ 。为更好的反映不同位置特征的重要性, 加强算法对全局映射信息的保留能力, 本文对 F'' 上下层权值矩阵 F''_2 、 F''_1 执行全局平均池化操作 (global average pooling, GAP), 再经 sigmoid 归一化^[34] 后, 生成用于反映每个通道在全局范围内的重要性权重向量 α 与 β 。为突出每个通道在全局范围内的重要性, 使用权重向量 α 与 β 对权值矩阵 F''_2 、 F''_1 重新加权, 并根据加权后的权值矩阵重新加权原特征图 F_1 和 F_2 , 确保每个位置上的特征不仅受全局权重的影响, 还能保留原始特征图中的局部信息。

AFF 根据可学习权值矩阵与权值向量, 能够更好地捕捉输入特征图之间的复杂关系, 并通过多层处理和动态调参提高特征融合的灵活性。该方法不仅增强了特征图的全局和局部表示能力, 同时提高了小目标信息与关键频率信息在融合过程中的显著性, 有效缓解漏检及误检问题。

3.3 动态回归损失 DY_IoU

以 IoU 为基础的衍生回归损失函数根据锚框边界整体回归法^[35], 提高了算法对回归变量相关性的关注度与 SAR 图像尺度差异样本的敏感度, 显著优化了模型检测精度与收敛速率。其计算形式如下:

$$IoU = 1 - L_{IoU} = 1 - \frac{I}{U}, 0 \leq L_{IoU} \leq 1, \quad (17)$$

$$Loss = L_{IoU} + R[a'(B, B_{gt}), b'(B, B_{gt}), c'(B, B_{gt}), \dots], \quad (18)$$

式中: IoU 为预测框与目标框交集 I 与并集 U 的比值, 用于反应预测框与目标框的匹配程度; L_{IoU} 为 IoU 损失; I 为预测框与目标框交集; U 为预测框与目标框并集; $Loss$ 为以 IoU 为基础的衍生回归损失函数; $R(\cdot)$ 为自变量 a', b', c' 构成的惩罚项, a', b', c' 为惩罚因子, 用于量化锚框和目标框之间匹配程度的几何度量; B, B_{gt} 分别代表锚框与目标框。

但衍生损失中存在不合理的惩罚项且缺少锚框质量评估方法, 导致回归过程易出现锚框膨胀和定位偏差等问题。对此, 本文结合自适应尺度惩罚项与非单调注意力机制设计动态回归损失函数 DY_IoU, 进一步优化算法对多尺度目标, 尤其小目标的定位与检测能力。其中, 自适应尺度惩罚项由动态尺度惩罚因子与基于锚框质量的梯度调整函数构成, 旨在缓解因锚框膨胀引起的定位失准与收敛退化问题, 避免梯度爆炸或消失情况。非单调注意力机制利用单个超参数调整算法对中等质量锚框的兴趣度, 以加快模型的收敛速度, 保证回归过程稳定性。

3.3.1 自适应尺度惩罚项

基线模型使用 $CIoU$ ^[36] 作为回归损失, 通过直接最小化边界框中心点之间的归一化距离与添加长宽比惩罚项, 促进锚框向目标框移动, 加快收敛速度与定位性能。计算过程如下:

$$\begin{cases} L_{CIoU} = L_{IoU} + \frac{d^2}{c^2} + V' * \xi, 0 \leq L_{CIoU} \leq 2 \\ V' = \frac{4}{\pi^2} \left(\arctan \frac{w_{gt}}{h_{gt}} - \arctan \frac{w}{h} \right)^2 \\ \xi = \frac{V'}{1 - IoU + V'} \end{cases}, \quad (19)$$

式中: d 为锚框中心点与目标框中心点之间的欧氏距离; c 为二者最小外接矩阵的对角线长度; d^2/c^2 为惩罚因子; w_{gt} 与 h_{gt} 分别为目标框的高度与宽度; w 与 h 分别为锚框的高度和宽度, \arctan 函数用于限制高宽比的差异, 以保持数值稳定性; $4/\pi^2$ 为归一化项, 旨在保证损失的平衡, 维持损失的合理补充; V' 为平衡 IoU 与高宽比、衡量锚框与目标框高宽比差异的权重因子; ξ 为用于调整高宽比向 V' 的动态权重因子。 V' 利用平方、归一化操作及 \arctan 函数平衡二者高宽比的敏感性和数值稳定性, 优化锚框的形状匹配, 进而提升检测效果。 ξ 通过动态加权机制与归一化操作使模型在低 IoU 时优先优化位置偏差, 在高 IoU 时注重高宽比匹配, 进而提升目标检测的精度与训练的收敛稳定性。计算过程显示图如图 8(b) 所示。

$CIoU$ 受惩罚项的局限, 仅能反映中心点间距离与纵横比差异, 无法直接捕获锚框与目标框间的形状差异, 从而导致不希望的收敛行为。此外, 通过计算最小外接矩阵对角线 c 的 R_D 梯度可知, 当 c 增大时 R_D 减小, 反之亦然。因此, 当锚框与目标框不重叠时, L_{IoU} 保持不变, $CIoU$ 通过降低回归损失迫使锚框尺寸变大, 这可能导致锚框膨胀, 进而影响回归效率。

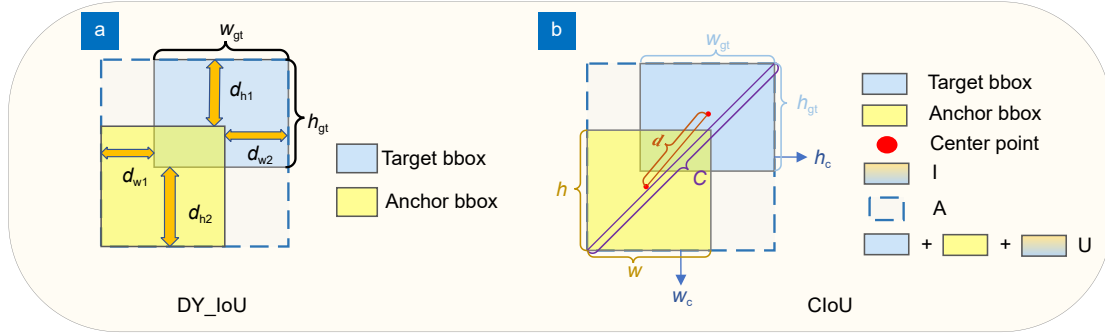


图 8 DY_IoU 与 CloU 回归计算可视化。(a) DY_IoU; (b) CloU

Fig. 8 Visualisation of DY_IoU and CloU regression calculations. (a) DY_IoU; (b) CloU

梯度计算如下:

$$\begin{cases} \frac{\partial R_D}{\partial d} = \frac{2d}{c^2} \\ R_D = \frac{d^2}{c^2} \end{cases} \quad (20)$$

为进一步提高回归效率, 增强模型定位能力, 本文利用动态尺度惩罚因子 M 与基于锚框质量的梯度调整函数 $f(x)$ 设计自适应尺度惩罚项, 解决不利惩罚因子引起的锚框膨胀问题, 使模型根据锚框尺度自适应选择最优回归路径。动态尺度惩罚因子 M 的表达式为

$$M = \left(\frac{d_{w1}}{w_{gt}} + \frac{d_{w2}}{w_{gt}} + \frac{d_{h1}}{h_{gt}} + \frac{d_{h2}}{h_{gt}} \right) \div 4, \quad (21)$$

式中: d_{w1} 、 d_{w2} 、 d_{h1} 、 d_{h2} 表示预测框与目标框对应边缘距离的绝对值, w_{gt} 、 h_{gt} 表示目标框宽高值。惩罚因子 M 对目标框的大小具有良好的自适应性, 因其分母仅依赖目标框的大小, 与预测框及最小外接矩阵无关, 这使得预测框在整个回归过程中均受目标框尺寸的约束, 有效抑制了锚框膨胀问题的产生。同时 M 数值不因锚框扩张而改变, 只有预测框与目标框完全重叠时 M 数值退化为 0。

为巩固回归过程稳定性, 实现算法对尺度样本的高效精准定位, 本文提出基于锚框质量的梯度调整函数 $f(x)$, 结构如图 9 所示。

该函数依据自适应梯度幅度调配, 加强对中等质量锚框的关注度, 缓解因低质量与高质量锚框产生的梯度爆炸与消失问题。 $f(x)$ 函数表达如下:

$$f(x) = 1 - e^{-x^2}. \quad (22)$$

经结合动态尺度惩罚因子 M 与基于锚框质量的梯度调整函数 $f(x)$ 后的 DY_IoU_1 计算过程表示为

$$DY_IoU_1 = IoU - f(M), -1 \leq DY_IoU_1 \leq 1, \quad (23)$$

$$\begin{aligned} L_{DY_IoU_1} &= 1 - DY_IoU_1 = L_{IoU} + f(M) \\ &= L_{IoU} + 1 - e^{-M^2}, 0 \leq L_{DY_IoU_1} \leq 2. \end{aligned} \quad (24)$$

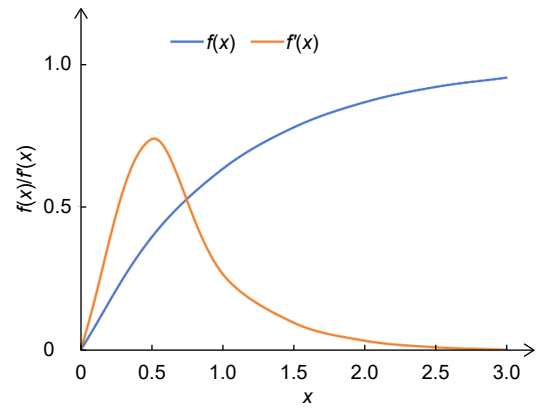


图 9 基于锚框质量的梯度调整函数结构图

Fig. 9 Structure of the gradient adjustment function based on the quality of the anchor frame

3.3.2 非单调注意力机制

为增强回归过程中网络对中等质量锚框的聚焦能力, 抑制低质量锚框产生的有害梯度, 提高 DY_IoU 回归损失的普适性, 本文借鉴 Focal EIou^[37] 与 WIou^[38] 损失函数, 提出了一种与 DY_IoU_1 结合的非单调注意力机制函数 $u(x)$, 该函数由单个超参数引导, 旨在简化梯度调整过程, 加快回归速率。非单调注意力机制函数 $u(x)$ 与 DY_IoU 回归函数的定义如下:

$$N = e^{-M}, N \in (0, 1], \quad (25)$$

$$u(x) = 3x \cdot e^{-x^2}, \quad (26)$$

$$L_{DY_IoU} = u(\Psi N) \times L_{DY_IoU_1} = 3 \times (\Psi N) \times e^{-(\Psi N)^2} \times L_{DY_IoU_1}, \quad (27)$$

式中: N 用于度量锚框质量, 其大小反映了锚框质量的优劣; Ψ 为用于控制注意力机制的超参数。当 N 趋于 0 时, 锚框质量变差; 当 $N=1$ 时, 预测框与目标

框完全重叠。DY_IoU 充分发挥了聚焦机制的潜力与优势, 在每个锚框回归过程中, 通过自适应地调整对不同质量锚框的关注程度, 提升了模型的整体回归效果。其计算过程显示图如图 8(a) 所示。

相比 CIoU, DY_IoU 损失利用自适应尺度惩罚项与非单调注意力机制, 解决了回归过程中锚框膨胀与极端质量锚框引起的回归偏差问题, 显著增强了模型对多尺度目标的定位与检测能力, 有效缓解因回归误差导致的漏检、误检问题, 节约资源开销, 二者回归过程比较图, 如图 10 所示。

4 实验结果与分析

4.1 数据集及实验环境

本文选用 SAR-AIRcraft-1.0 数据集^[39]和 HRSID 数据集^[40]进行实验验证。SAR-AIRcraft-1.0 数据集图像来自中国的高分三号 Gaofen-3 卫星, 该数据集包含图像分辨率为 0.5、1.0、3.0 m 的 4368 张图像, 其中 3058 张用于模型训练, 873 张用于验证, 437 张用于测试。图像尺寸分别为 800 pixel×800 pixel、1000 pixel×1000 pixel、1200 pixel×1200 pixel 和 1500 pixel×1500 pixel, 数据集中共有 16463 个飞机实例, 覆盖了 7 种不同类别飞机: A220、A320/321、A330、ARJ21、Boeing737、Boeing787 和 other。HRSID 数据集图像采集于 TerraSAR-X、Sentinel-1B 和 TanDEM-

X 卫星, 采用 horizontal bounding boxes (HBB) 格式进行标注, 共包含 5604 张大小为 800 pixel×800 pixel, 分辨率为 0.5、1.0、3.0 m 的 SAR 图像与 16951 个不同目标尺度的舰船实例。

实验环境为 ubuntu 20.04 操作系统, CPU 为 Intel (R) Xeon (R) Platinum 8474C, 基准频率为 3.60 GHz, GPU 为 NVIDIA GeForce RTX 4090 24 GB, 编程语言为 Python 3.8, 深度学习框架为 pytorch 1.11.0+ CUDA 11.3。为了公平对比算法的性能, 所有方法采用相同的超参数设置进行训练、验证, 批次大小 (batch_size) 设为 64, 初始学习率为 0.01, 训练轮数 (epochs) 为 300。

4.2 评价指标

本实验采用查准率 (Precision)、召回率 (Recall)、平均精度均值 (AP)、加权平均精度均值 mAP@0.5 (IoU 阈值取 0.5 时的 mAP 平均值)、加权调和平均 (F1)、模型参数量 (Params) 与计算量 (GFLOPs) 作为测评模型与目标类别的评价指标。对应计算公式为

$$Precision = \frac{TP}{TP+FP}, \quad (28)$$

$$Recall = \frac{TP}{TP+FN}, \quad (29)$$

$$AP = \int_0^1 Precision(Recall)d(Recall), \quad (30)$$

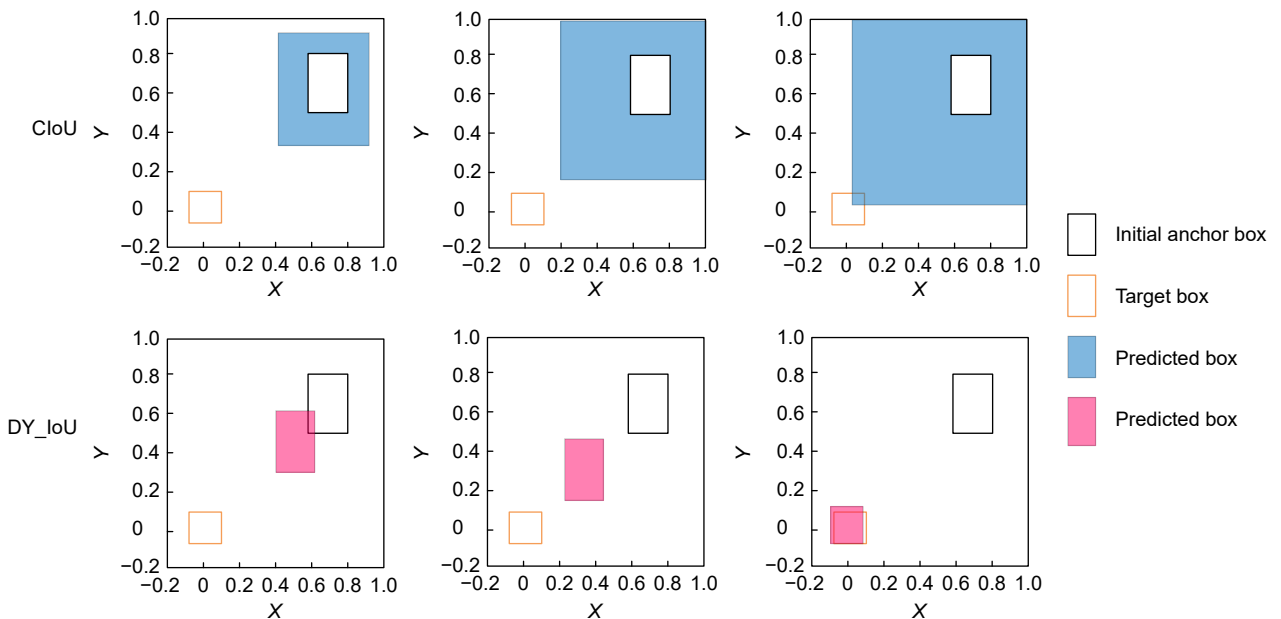


图 10 回归过程比较

Fig. 10 Comparison of regression processes

$$mAP = \frac{1}{n} \sum_{i=1}^n \int_0^1 Precision(Recall) d(Recall), \quad (31)$$

$$F1 = 2 \frac{Precision \times Recall}{Precision + Recall}, \quad (32)$$

式中: TP 表示模型正确预测实例为正样本的数量; FP 表示模型错误预测实例为正样本的数量; FN 表示模型错误预测实例为负样本的数量。

4.3 消融实验

为验证改进模型对 YOLOv10s 的每个改进策略的有效性, 使用 SAR-AIRcarft-1.0 数据集以 YOLOv10s 为基线进行消融实验, 实验所有数据参数和环境配置严格一致。其中“√”表示使用该改进策略, “—”表示未使用。

表 1 中的实验结果表明, 单独引入空-频域感知单元 SFDS, Precision 提高 0.7%, Recall 增加 2.7%, mAP0.5 增加 2.2%, Params 降低 0.62 M, GFLOPs 减少 1.8, 说明 SFDS 利用空间感知模块与频域感知模块能够增强算法对全局关键特征的学习能力与多尺度、多方向频域信息的捕获能力, 降低背景冗余, 提高模型检测性能, 加快推理速度。将 AFF 模块替换 Concat 单元后, Recall 增加 0.6%, mAP0.5 增加 0.1%, 添加 AFF 模块能够有效改善信息流失问题, 突出关键特征, 提高多尺度特征的融合质量与效率。DY_IoU 采用新型动态惩罚机制, 能够更准确地衡量预测框与真实框的重合程度, 进而优化边界框的位置和形状, 增强算法对多尺度目标的定位与识别能力, 并降低了模型复杂度, Recall 增加 1.5%, mAP0.5 增加 1.0%, Params 降低 0.03 M, GFLOPs 减少 0.6。

同时加入 SFDS 与 AFF 模块, 进一步提高了算法的特征提取与交互能力, 优化了模型整体复杂度, Precision 增加 1.9%, Recall 增加 4.6%, mAP0.5 增加 3.8%, Params 降低 0.02 M, GFLOPs 减少 0.3。同时加入 SFDS 与 DY_IoU, 显著增强了模型对代表性特征的判别能力与细粒度映射信息感知力, 进一步改善了模型的定位与检测能力, Precision 增加 4.0%, Recall 增加 6.3%, mAP0.5 增加 4.4%, Params 降低 0.17 M, GFLOPs 减少 0.7。同时加入 AFF 与 DY_IoU 相比原模型 Recall 增加 3.4%, mAP0.5 增加 2.4%。将 SFDS、AFF 与 DY_IoU 同时加入基线模型中, 即所提改进方法, 相比原模型, Precision 增加 10.2%, Recall 增加 1.7%, mAP0.5 增加 5.2%, Params 降低 0.05 M, GFLOPs 减少 0.4。根据消融实验结果可知, 改进模型完善了特征捕获与融合策略, 显著提高了综合检测性能, 证明了改进模型的有效性。

4.4 对比试验

本文将改进算法同当前流行检测算法与经典算法在 SAR-AIRcarft-1.0 数据集上进行对比, 以进一步验证算法有效性, 对比算法包括目标检测领域经典算法和近年流行算法在 VisDrone2019 数据集上进行对比, 包括 Faster-RCNN^[4]、RetinaNet^[41]、YOLOv5s、SKG-Net^[42]、YOLOv8s、YOLOv10s^[19]、SFS-CNet^[17]。对比实验结果如表 2 所示。

对比实验结果表明, 本文基于 YOLOv10s 单阶段目标检测算法提出的改进模型, 在查准率、召回率和推理速度方面全面优于多种经典算法。相较于 Faster R-CNN, 改进模型解决了两阶段框架的计算冗

表 1 所提算法在 SAR-AIRcarft-1.0 数据集的消融实验

Table 1 Ablation experiments of the proposed algorithm on the SAR-AIRcarft-1.0 dataset

YOLOV10	SFDS	AFF	DY_IoU	Precision/%	Recall/%	mAP0.5/%	Params/10 ⁶	GFLOPs
√	—	—	—	86.9	89.4	90.7	2.60	8.4
√	√	—	—	87.6	92.1	92.9	1.98	6.6
√	—	√	—	82.3	90.0	90.8	3.18	9.3
√	—	—	√	84.9	90.9	91.7	2.47	7.8
√	√	√	—	88.8	94.0	94.5	2.58	8.1
√	√	—	√	90.9	95.7	95.1	2.43	7.7
√	—	√	√	86.0	92.8	93.1	2.61	8.4
√	√	√	√	97.1	91.1	95.9	2.55	8.0

表 2 SAR-AIRcarft-1.0 数据集对比实验结果

Table 2 Results of comparison experiments on the SAR-AIRcarft-1.0 dataset

Model	Precision/%	Recall/%	mAP50/%	F1/%	GFLOPs
Faster R-CNN	79.0	68.5	75.9	73.4	137.5
Center-Net	62.8	71.9	70.8	67.6	51.6
YOLOv5s	90.5	81.1	86.9	85.5	16.5
SKG-Net	85.6	75.8	70.6	59.7	120
YOLOv8s	92.3	81.8	90.0	86.7	28.6
YOLOv10s	96.9	89.4	90.7	88.1	8.4
SFS-CNet	94.7	84.5	89.9	89.3	6.9
Ours	97.1	91.1	95.9	94.0	6.2

余问题, 推理速度显著加快, GFLOPs 降低至其 1/22。在复杂背景下, 查准率和召回率分别提高了 18.1% 和 22.6%, 有效提升了多尺度目标的检测性能。相较于 CenterNet, 改进模型在查准率、召回率和 F1 值上分别提升了 34.3%、19.2% 和 26.4%, 展现了其在小目标检测中的显著优势。此外, 特征融合模块增强了对 SAR 图像复杂纹理的鲁棒性。与 YOLOv5s 和 YOLOv8s 相比, 改进模型采用动态回归损失函数, 进一步优化了检测精度及对多尺度目标的敏感性。其中, 查准率分别提高 6.6% 和 4.8%, mAP@0.5 分别提升 9.0% 和 5.9%, GFLOPs 分别减少 10.3 和 22.4, 体现了其在计算效率与性能权衡上的显著优势。此外, 与 SKG-Net 相比, 本文模型减少了冗余计算, 使 GFLOPs 降低至 SKG-Net 的 1/19。在此基础上, 查准率、召回率和 F1 值分别提高了 11.5%、15.3% 和 34.3%, 进一步验证了其在推理效率和泛化性能方面的优势。与基线模型 YOLOv10s 相比, 本文模型在查准率、召回率和 mAP@0.5 关键指标上分别提升了 0.2%、1.7% 和 5.2%。此外, 改进的动态回归损失函数有效缓解了锚框膨胀问题, 增强了模型对中等质量锚框的敏感性, 并进一步提升了 SAR 图像中小目标和多尺度目标的检测性能。实验结果充分验证了本文改进方法的有效性和广泛适用性。

综上所述, 改进模型相比同领域的算法具有更高的检测精度、更快的检测速度和最佳的综合性能。在 SAR 图像目标检测任务中, 更具优势。

为了更加直观地对比上述 8 种方法, 本文从 SAR-AIRcarft-1.0 数据集中选取“多尺度目标”、“复杂背景”和“密集分布”三组代表性图像进行对比可视

化, 各算法可视化结果如图 11、图 12 所示。从图 11、图 12 可以看出, 改进算法在 3 组代表图像中的检测效果均优于其他对比算法, 有效抑制了背景干扰, 并解决了因采样不充分, 定位失准导致的检测精度低、漏检和误检等问题。进一步验证了改进算法的有效性。具体而言, 在“多尺度目标”场景中, 改进算法能够更精准地检测不同尺度的目标, 尤其是在小目标检测方面表现突出。这一优势归功于算法中引入的动态感受野与多尺度特征融合策略, 二者优化了其多尺度特征的捕获与整合能力, 增强了模型对小目标敏感性, 有效缓解了其他算法在小目标检测中的漏检问题。在“复杂背景”场景下, 传统算法常出现误检和漏检问题, 特别是在背景纹理复杂或目标边缘模糊的区域。改进算法通过引入空-频域感知单元, 更有效地提取全局上下文信息, 降低背景噪声和散斑干扰。在“密集分布”场景中, 改进算法在处理高密度目标分布时表现出优异的分辨能力和定位精度。这得益于本文设计的自适应特征融合模块, 该模块通过优化多尺度特征交互, 增强了模型对密集目标的分辨能力, 并结合 DY_IoU 动态回归损失函数, 显著提高定位精度, 降低误判率和漏判率。

为进一步验证改进模型的泛化能力和评估算法的鲁棒性, 本文在 HRSID 数据集上对各算法进行了对比实验。实验结果如表 3 所示。

结果表明, 改进模型在 Precision 评价指标上仅次于基线模型 YOLOv10s, 但在 Recall、mAP@0.5、F1 和 GFLOPs 指标上均优于其他算法。综上, 本文改进模型能够更好的兼顾算法检测精度与推理速度, 综合检测性能最佳。进一步说明, 改进模型具有较高的鲁棒性与普适性。

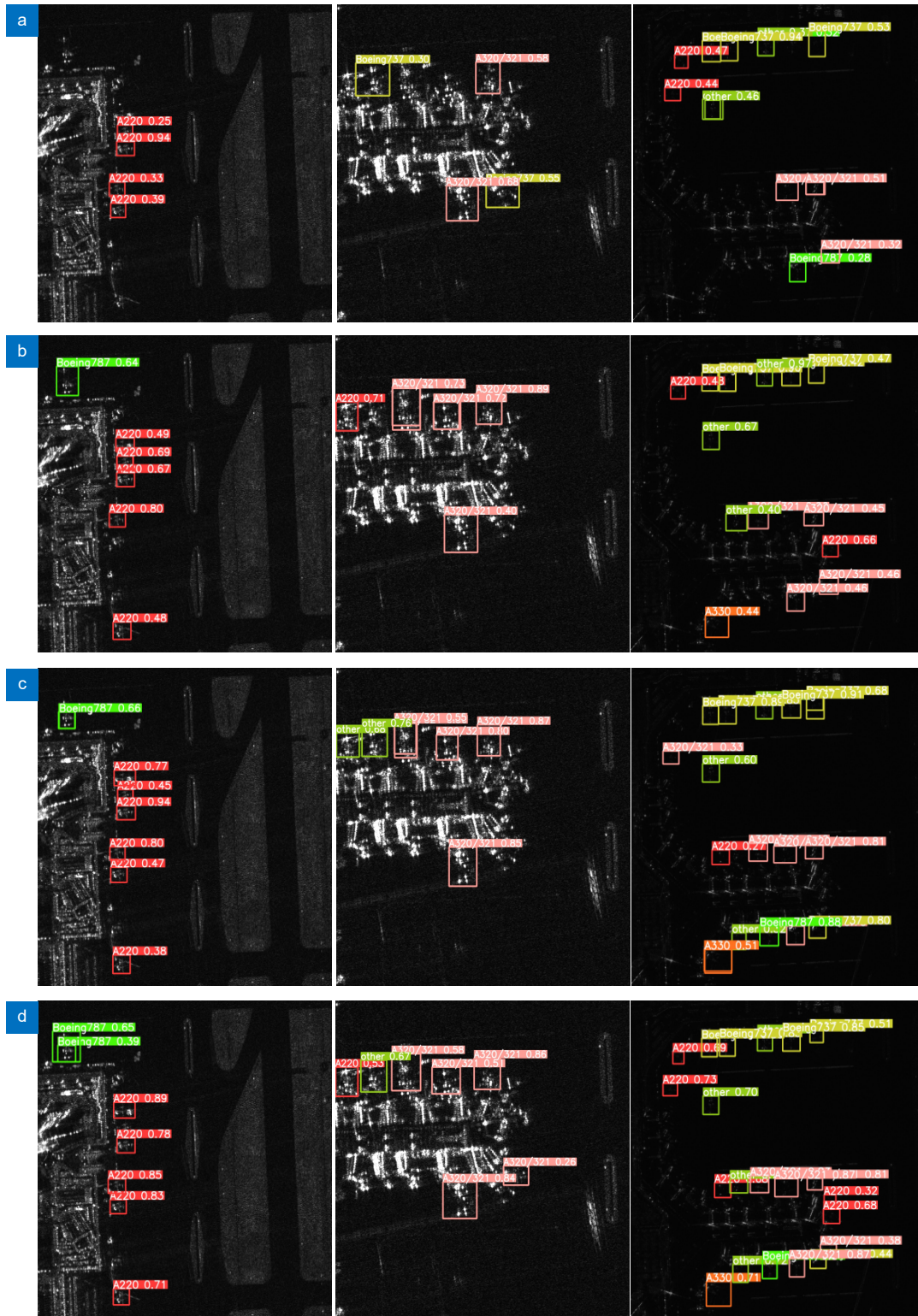


图 11 对比试验可视化结果 (一)。 (a) SKG-Net; (b) Center-Net; (c) Faster-RCNN; (d) YOLOv5s
Fig. 11 Comparison test visualisation results (I). (a) SKG-Net; (b) Center-Net; (c) Faster-RCNN; (d) YOLOv5s

5 结论

目标尺度不均衡、背景复杂度高所导致的检测精度低、推理速度慢等问题, 本文提出了一种融合空-频域的动态 SAR 图像目标检测算法。该算法通过引入空-频

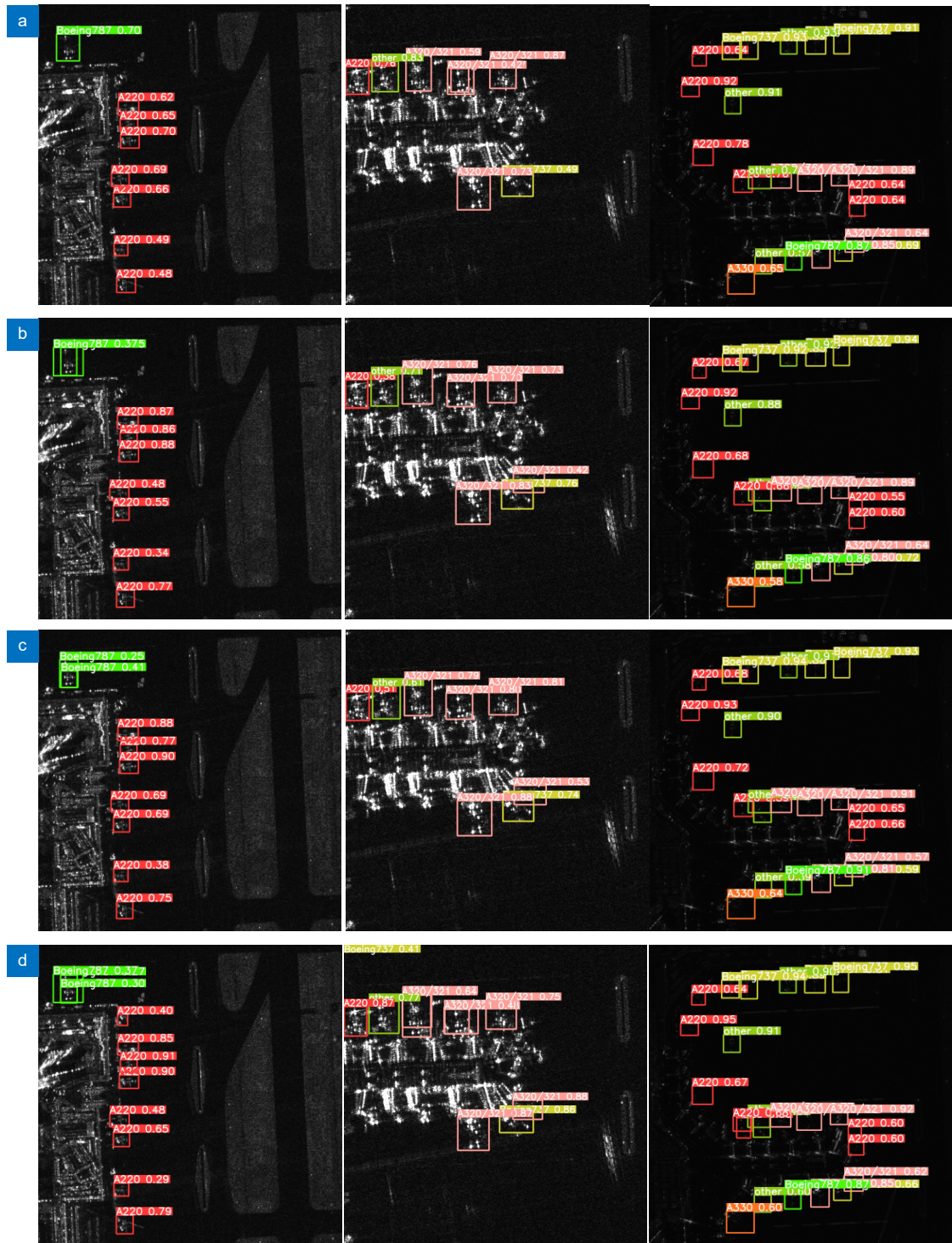


图 12 对比试验可视化结果 (二)。(a) SFS-CNet; (b) YOLOv8s; (c) YOLOv10s; (d) 所提算法
 Fig. 12 Comparison test visualisation results (II). (a) SFS-CNet; (b) YOLOv8s; (c) YOLOv10s; (d) Proposed algorithm

域感知单元 (SFDS), 增强对全局空间特征与频率散射特征的捕获能力与全局上下文信息的保留能力, 并提高采样精度与效率, 减少资源消耗。此外, 设计自适应特征融合模块 (AFF), 通过学习权重参数动态调整特征融合的程度, 加强了不同特征图之间的信息交

流和融合。最后采用动态损失函数 DY_Iou 优化检测头, 并结合自适应惩罚项和非单调注意力机制, 提高中等质量预测框的回归速率, 解决位置偏差、误检和漏检问题, 提升了算法在多目标检测中的精度和稳定性。在 SAR-AIRcraft-1.0 数据集与 HRSID 数据集上

表 3 HRSID 数据集对比实验结果

Table 3 Results of comparison experiments on the HRSID dataset

算法	Precision/%	Recall/%	mAP50/%	F1/%	GFLOPs
Faster R-CNN	86.3	81.6	86.3	83.9	137.5
Center-Net	90.5	74.1	83.3	81.5	51.6
YOLOv5s	94.3	89.4	88.1	91.7	16.5
SKG-Net	77.8	81.5	81.7	79.6	120
YOLOv8s	90.1	94.4	95.9	92.2	28.6
YOLOv10s	97.8	96.2	97.6	95.9	8.4
SFS-CNet	88.8	95.3	92.9	91.9	6.9
Ours	96.2	97.8	98.8	97.0	6.2

的实验表明, 改进后的模型在 mAP@0.5 指标上达到了 95.9% 和 98.8%, 检测精度优于对比算法, 尤其在复杂背景和密集小目标检测方面表现出色。然而, 该方法在计算资源的占用上仍存在一定局限性。未来研究可在保持检测精度的同时, 探索模型的轻量化, 并将深度学习与传统图像处理技术相结合, 以进一步提升模型的可解释性和实际应用价值。

参考文献

- Liang L M, Chen K Q, Wang C B, et al. Remote sensing image detection algorithm integrating visual center mechanism and parallel patch perception[J]. *Opto-Electron Eng*, 2024, **51** (7): 240099.
梁礼明, 陈康泉, 王成斌, 等. 融合视觉中心机制和并行补丁感知的遥感图像检测算法[J]. *光电工程*, 2024, **51** (7): 240099.
- Xiao Z J, Zhang J H, Lin B H. Feature coordination and fine-grained perception of small targets in remote sensing images[J]. *Opto-Electron Eng*, 2024, **51** (6): 240066.
肖振久, 张杰浩, 林渤翰. 特征协同与细粒度感知的遥感图像小目标检测[J]. *光电工程*, 2024, **51** (6): 240066.
- Ma L, Gou Y T, Lei T, et al. Small object detection based on multi-scale feature fusion using remote sensing images[J]. *Opto-Electron Eng*, 2022, **49** (4): 210363.
马梁, 苟于涛, 雷涛, 等. 基于多尺度特征融合的遥感图像小目标检测[J]. *光电工程*, 2022, **49** (4): 210363.
- Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, **39** (6): 1137–1149.
- Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 779–788. <https://doi.org/10.1109/CVPR.2016.91>.
- Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//*IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 6517–6525. <https://doi.org/10.1109/CVPR.2017.690>.
- Redmon J, Farhadi A. YOLOv3: an incremental improvement[Z]. arXiv: 1804.02767, 2018. <https://doi.org/10.48550/arXiv.1804.02767>.
- Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[Z]. arXiv: 2004.10934, 2020. <https://doi.org/10.48550/arXiv.2004.10934>.
- Ge Z, Liu S T, Wang F, et al. YOLOX: exceeding YOLO series in 2021[Z]. arXiv: 2107.08430, 2021. <https://doi.org/10.48550/arXiv.2107.08430>.
- Lyu Z W, Jin H F, Zhen T, et al. Small object recognition algorithm of grain pests based on SSD feature fusion[J]. *IEEE Access*, 2021, **9**: 43202–43213.
- Zhang L P, Liu Y, Zhao W D, et al. Frequency-adaptive learning for SAR ship detection in clutter scenes[J]. *IEEE Trans Geosci Remote Sens*, 2023, **61**: 5215514.
- Si J H, Song B B, Wu J X, et al. Maritime ship detection method for satellite images based on multiscale feature fusion[J]. *IEEE J Sel Top Appl Earth Obs Remote Sens*, 2023, **16**: 6642–6655.
- Qin C, Wang X Q, Li G, et al. A semi-soft label-guided network with self-distillation for SAR inshore ship detection[J]. *IEEE Trans Geosci Remote Sens*, 2023, **61**: 5211814.
- Xu X W, Zhang X L, Zhang T W, et al. SAR ship detection in complex scenes based on adaptive anchor assignment and IOU supervise[J]. *J Radars*, 2023, **12** (5): 1097–1111.
胥小我, 张晓玲, 张天文, 等. 基于自适应锚框分配与 IOU 监督的复杂场景 SAR 舰船检测[J]. *雷达学报*, 2023, **12** (5): 1097–1111.
- Xiao Z J, Lin B H, Qu H C. SAR ship detection with multi-mechanism fusion[J]. *J Image Graphics*, 2024, **29** (2): 545–558.
肖振久, 林渤翰, 曲海成. 融合多重机制的 SAR 舰船检测[J]. *中国图象图形学报*, 2024, **29** (2): 545–558.
- Sun P S, Wen X B. An improved algorithm for detecting ship target in SAR images based on YOLOv5 model[J]. *Electron-Opt Control*, 2024, **31** (8): 32–37,85.
孙培双, 温显斌. 基于改进 YOLOv5 模型的 SAR 图像舰船目标检测算法[J]. *电光与控制*, 2024, **31** (8): 32–37,85.
- Li K, Wang D, Hu Z Y, et al. Unleashing channel potential: space-frequency selection convolution for SAR object detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 17323–17332. <https://doi.org/10.1109/CVPR52733.2024.01640>.
- Zhou J, Xiao C, Peng B, et al. DiffDet4SAR: diffusion-based aircraft target detection network for SAR images[J]. *IEEE Geosci Remote Sens Lett*, 2024, **21**: 4007905.
- Wang A, Chen H, Liu L H, et al. YOLOv10: real-time end-to-end object detection[Z]. arXiv: 2405.14458, 2024. <https://doi.org/10.48550/arXiv.2405.14458>.
- Zhang P F, Lo E, Lu B T. High performance depthwise and pointwise convolutions on mobile devices[C]//*Proceedings of the 34th AAAI Conference on Artificial Intelligence*, 2020: 6795–6802. <https://doi.org/10.1609/aaai.v34i04.6159>.
- Guo Y H, Li Y D, Wang L Q, et al. Depthwise convolution is all you need for learning multiple visual domains[C]//*Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, 2019: 8368–8375. <https://doi.org/10.1609/aaai.v33i01.33018368>.

- [22] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- [23] Wu H B, Kuo H C, Zheng N J, et al. Partially fake audio detection by self-attention-based fake span discovery[C]//*ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022: 9236–9240. <https://doi.org/10.1109/ICASSP43922.2022.9746162>.
- [24] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017: 6000–6010.
- [25] Bebis G, Georgiopoulos M. Feed-forward neural networks[J]. *IEEE Potentials*, 1994, 13 (4): 27–31.
- [26] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 1800–1807. <https://doi.org/10.1109/CVPR.2017.195>.
- [27] Neubeck A, Van Gool L. Efficient non-maximum suppression[C]//*18th International Conference on Pattern Recognition (ICPR'06)*, 2006: 850–855. <https://doi.org/10.1109/ICPR.2006.479>.
- [28] Li J F, Wen Y, He L H. SCConv: spatial and channel reconstruction convolution for feature redundancy[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 6153–6162. <https://doi.org/10.1109/CVPR52729.2023.00596>.
- [29] Hsiao T Y, Chang Y C, Chou H H, et al. Filter-based deep-compression with global average pooling for convolutional networks[J]. *J Syst Archit*, 2019, 95: 9–18.
- [30] McClenny L, Braga-Neto U. Self-adaptive physics-informed neural networks using a soft attention mechanism[Z]. arXiv:2009.04544, 2020. <https://doi.org/10.48550/arXiv.2009.04544>.
- [31] Chen J W, An D X, Ge B B, et al. Detection, parameters estimation, and imaging of moving targets based on extended post-Doppler STAP in multichannel WasSAR-GMTI[J]. *IEEE Trans Geosci Remote Sens*, 2024, 62: 5223515.
- [32] Koç E, Alikashişoğlu T, Aras A C, et al. Trainable fractional Fourier transform[J]. *IEEE Signal Process Lett*, 2024, 31: 751–755.
- [33] Luan S Z, Chen C, Zhang B C, et al. Gabor convolutional networks[J]. *IEEE Trans Image Process*, 2018, 27 (9): 4357–4366.
- [34] Yin X Y, Goudriaan J, Lantinga E A, et al. A flexible sigmoid function of determinate growth[J]. *Ann Bot*, 2003, 91 (3): 361–371.
- [35] Gevorgyan Z. SloU loss: more powerful learning for bounding box regression[Z]. arXiv: 2205.12740, 2022. <https://doi.org/10.48550/arXiv.2205.12740>.
- [36] Zheng Z H, Wang P, Ren D W, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. *IEEE Trans Cybern*, 2022, 52 (8): 8574–8586.
- [37] Zhang Y F, Ren W Q, Zhang Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. *Neurocomputing*, 2022, 506: 146–157.
- [38] Tong Z J, Chen Y H, Xu Z W, et al. Wise-loU: bounding box regression loss with dynamic focusing mechanism[Z]. arXiv: 2301.10051, 2023. <https://doi.org/10.48550/arXiv.2301.10051>.
- [39] Wang Z R, Kang Y Z, Zeng X, et al. SAR-AIRcraft-1.0: High-resolution SAR aircraft detection and recognition dataset[J]. *J Radars*, 2023, 12 (4): 906–922. 王智睿, 康玉卓, 曾璇, 等. SAR-AIRcraft-1.0: 高分辨率 SAR 飞机检测识别数据集[J]. *雷达学报*, 2023, 12 (4): 906–922.
- [40] Wei S J, Zeng X F, Qu Q Z, et al. HRSID: a high-resolution SAR images dataset for ship detection and instance segmentation[J]. *IEEE Access*, 2020, 8: 120234–120254.
- [41] Wang Y Y, Wang C, Zhang H, et al. Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery[J]. *Remote Sens*, 2019, 11 (5): 531.
- [42] Pan D C, Gao X, Dai W, et al. SRT-net: scattering region topology network for oriented ship detection in large-scale SAR images[J]. *IEEE Trans Geosci Remote Sens*, 2024, 62: 5202318.

作者简介



沈学利(1969-), 男, 博士研究生, 教授, CCF 高级会员, 主要从事推荐系统、网络及信息安全、图像与视觉信息计算方面的研究等。
E-mail: Shenxueli@intu.edu



【通信作者】王嘉慧(2000-), 女, 硕士研究生, CCF 学生会员, 主要主要从事遥感图像目标检测方面的研究、推荐系统研究。
E-mail: 2245414310@qq.com



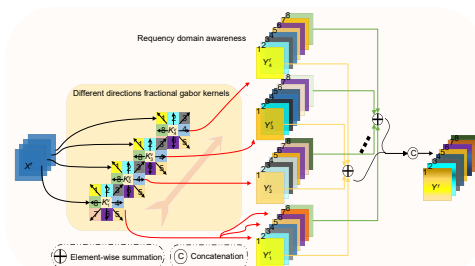
吴正伟(1998-), 男, 硕士研究生, 主要从事机器学习 and 图像与视觉信息计算方面的研究。
E-mail: 1525545769@qq.com



扫描二维码, 获取PDF全文

Dynamic SAR image target detection by fusing space-frequency domain

Shen Xueli, Wang Jiahui*, Wu Zhengwei



Structure of frequency domain awareness (FDA) module

Overview: A dynamic SAR image target detection algorithm integrating spatial-frequency domains is proposed to address several challenges inherent to SAR imagery, including significant feature variability, imbalanced target scales, and high speckle noise in background regions. These challenges contribute to decreased detection accuracy and slower inference speeds, posing difficulties for real-time applications. The proposed method is specifically designed to overcome these limitations through multiple innovative components that enhance both detection performance and computational efficiency. The algorithm first employs a dual-stream perception strategy to construct spatial-frequency perception units. This design integrates both dynamic receptive fields and fractional-order Gabor transforms, significantly improving the model's ability to capture spatial diversity and frequency scattering features. By expanding the receptive fields adaptively, the algorithm captures both local and global contexts, leading to more effective extraction of complex patterns in the input data. Using fractional-order Gabor transforms further enhances the model's sensitivity to fine-grained texture and frequency features, which helps retain important global contextual information. These improvements collectively speed up inference by minimizing redundant feature representations, reducing the interference from background noise, and decreasing the similarity of feature mapping patterns. Consequently, the algorithm effectively addresses common issues such as missed and false detections, are typical in cluttered SAR images. In the next stage, a re-parameterization-based adaptive feature fusion module is introduced to optimize the interaction between multi-scale features. This module facilitates the efficient integration of features across different scales, enriching feature diversity and mitigating the discrepancies introduced during the sampling process. Additionally, the fusion process highlights the salience of small targets and key frequency information, often challenging to detect in traditional SAR detection frameworks. This enhanced multi-scale feature integration improves the detection accuracy, particularly for small and subtle objects, which are crucial in applications like maritime surveillance and remote sensing. To further enhance the algorithm's effectiveness, a dynamic regression loss function, DY_IoU, is incorporated. This loss function employs adaptive scale penalty factors and a dynamic non-monotonic attention mechanism to address anchor box expansion and positional deviations. By dynamically adjusting the focus during training, the model achieves more precise localization of multi-scale targets. Moreover, the improved loss function facilitates faster convergence, reduces the computational burden, and ensures the algorithm remains lightweight and efficient for practical deployment. The proposed method was evaluated on two publicly available datasets, SAR-Aircraft-1.0 and HRSID. Experimental results show that the algorithm achieves mAP@0.5 values of 95.9% and 98.8%, respectively, representing 5.2% and 1.2% improvements over baseline models. Additionally, the proposed approach outperforms other comparison algorithms, demonstrating its superiority. These results confirm that the algorithm not only enhances detection accuracy but also exhibits strong robustness and generalization capabilities, making it suitable for a wide range of real-world applications.

Shen X L, Wang J H, Wu Z W. Dynamic SAR image target detection by fusing space-frequency domain[J]. *Opto-Electron Eng*, 2025, 52(1): 240245; DOI: [10.12086/oe.2025.240245](https://doi.org/10.12086/oe.2025.240245)