

DOI:10.16356/j.2097-6771.2026.01.016

## 基于动态频域卷积的 SAR 目标检测

纪晓平<sup>1</sup>, 陶 普<sup>2</sup>

(1. 中国航空工业集团公司雷华电子技术研究所, 无锡 214082; 2. 南京航空航天大学航天学院, 南京 211106)

**摘要:** 由于合成孔径雷达 (Synthetic aperture radar, SAR) 具有特殊的成像机制与电磁散射特性, 导致其采集的图像常伴随强散斑噪声和复杂背景干扰, 这些特性严重制约了针对该类图像的目标检测的精度与鲁棒性。为进一步降低噪声干扰并解决现有方法存在的多尺度特征建模不足和频域信息利用有限的问题, 本文提出一种融合频域动态卷积与空-频特征增强的 SAR 目标检测网络。首先引入了动态频域卷积模块, 通过可学习的傅里叶谱系数和分组重构卷积核, 结合特征修复机制实现对高低频成分的自适应调制, 从而提升了卷积核的频带响应多样性与干扰条件下的特征表达能力。随后, 通过联合频率自注意力与空间自注意力机制以及空频融合策略构建了空-频特征增强模块, 实现了目标特征的增强。实验结果表明, 所提方法在数据集 MSAR 与 SARDet-100K 上相较于基于注意力的可变形多子空间特征去噪的 SAR 图像目标检测 (Attention as deformable multisubspace feature denoising for target detection in SAR images, 记作 DenoDet) 网络、全卷积单阶段目标检测 (Fully convolutional one-stage object detection, FCOS) 网络、金字塔视觉 Transformer 轻量版 (Pyramid vision Transformer-tiny, PVT-T)、Faster 基于区域的卷积神经网络 (Region-based convolutional neural networks, R-CNN) 和仅需聚焦单层级特征 (You only look one-level feature, YOLOF) 网络等代表性方法, 在多项评价指标上均取得了显著提升, 展现出更高的检测精度、鲁棒性与良好的泛化能力, 为 SAR 图像目标检测提供了一种新的解决思路。

**关键词:** 合成孔径雷达; 目标检测; 动态频域卷积; 空-频特征增强; 自注意力

**中图分类号:** TP753 **文献标志码:** A **文章编号:** 1005-2615(2026)01-0173-10

## SAR Target Detection Based on Dynamic Frequency Convolution

Ji Xiaoping<sup>1</sup>, Tao Pu<sup>2</sup>

(1. Leihua Electronic Technology Research Institute, Aviation Industry Corporation of China, Wuxi 214082, China;  
2. College of Astronautics, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China)

**Abstract:** Due to the unique imaging mechanism and electromagnetic scattering characteristics of synthetic aperture radar (SAR), the acquired images often contain strong speckle noise and complex background interference, which severely limit object detection accuracy and robustness. To mitigate noise and address the limitations of insufficient multi-scale feature modeling and limited frequency information usage in existing methods, this paper proposes a SAR object detection network combining frequency dynamic convolution with spatial-frequency feature enhancement. A dynamic frequency-domain convolution module is first introduced. It adaptively modulates high and low frequency components via learnable Fourier coefficients and grouped reconstructed convolution kernels, combined with a feature restoration module, enriching the diversity of the kernel's frequency responses and enhancing the capability of feature representation. A spatial-frequency feature enhancement module is then constructed using joint frequency and spatial self-attention with a spatial-frequency fusion strategy, enhancing the features of targets. Experiments on MSAR and

**基金项目:** 国家自然科学基金(62471224)。

**收稿日期:** 2025-09-17; **修订日期:** 2025-10-31

**通信作者:** 纪晓平, 男, 高级工程师, E-mail: 52625271@qq.com。

**引用格式:** 纪晓平, 陶普. 基于动态频域卷积的 SAR 目标检测[J]. 南京航空航天大学学报(自然科学版), 2026, 58(1): 173-182. Ji Xiaoping, Tao Pu. SAR target detection based on dynamic frequency convolution[J]. Journal of Nanjing University of Aeronautics & Astronautics (Natural Science Edition), 2026, 58(1): 173-182.

SARDet-100K datasets demonstrate that, compared with attention as deformable multisubspace feature denoising for target detection in SAR images (Denoted as DenoDet), fully convolutional one-stage object detection (FCOS), pyramid vision Transformer-tiny (PVT-T), faster region-based convolutional neural networks (R-CNN), and you only look one-level feature (YOLOF), the proposed method achieves significant improvements across multiple evaluation metrics and shows higher detection accuracy, robustness, and generalization, offering a new approach for SAR image object detection.

**Key words:** synthetic aperture radar (SAR); target detection; dynamic frequency convolution; spatial-frequency feature enhancement; self-attention

合成孔径雷达(Synthetic aperture radar, SAR)是一种具有全天候成像能力的有源微波遥感设备,凭借其独特的成像机理和电磁散射特性,被广泛应用于军事侦察、海洋监测和灾害调查等领域<sup>[1-3]</sup>。近年来SAR目标检测领域中出现了具有代表性的包含MSAR<sup>[4]</sup>、SARDet-100K<sup>[5]</sup>在内的数据集,这使得SAR目标检测得到了更为广泛的研究。

虽然建立在空间与通道间的联合特征提取与融合上的深度学习技术在语义分割<sup>[6]</sup>、密集预测<sup>[7]</sup>等下游任务中取得了良好的效果。但面对SAR成像过程中固有的强散斑噪声和复杂背景,将该技术运用在目标检测<sup>[8]</sup>时仍面临严峻挑战。尽管SAR目标检测已经持续在公开数据集上取得了进展,但以卷积作为核心组件的检测网络采用的固定权重的方式在应对具有复杂特征的目标时缺乏自适应的建模能力,因此动态卷积<sup>[9]</sup>在其基础上应运而生,提供了自适应且高效的方法,其实现方法是结合注意力模块与多个并行权重,使得权重可以跟随输入自适应调整。然而,已知的动态卷积大多在空间域进行,并行权重对频域信息缺乏利用,这导致各个权重具有相似的频率响应,这一特性产生的频响多样性缺失降低了模型对频率信息的捕获能力,对于需要区分前景和背景的机器视觉任务产生了难以忽略的影响。同时,多数目标检测算法采用经过预训练的骨干网络提取图像的多尺度特征,其中涉及多次下采样以扩展感受野并降低维度,根据Nyquist-Shannon定理,如果存在欠采样会导致信号在频域中的分量存在混叠现象。文献[10]表明,混叠程度与下采样步长正相关,即越大的卷积步长会导致越严重的混叠效应。文献[11]采用低通滤波器去除高频成分克服这一问题,但人工设置低通核或采用过低的频率阈值导致有效信息损失,无法有效剔除影响下游任务精度的高频噪声。随后,一些算法提出在卷积层上的抗混叠滤波器<sup>[12]</sup>和模糊滤波器<sup>[13]</sup>作为对策,一些算法还通过将深度自适应模糊滤波器和去混叠激活函数<sup>[14]</sup>或小波变换中的低频分量相结合来解决该问题<sup>[15]</sup>。

虽然卷积神经网络在局部和多尺度信息建模

上取得了巨大成功,但缺乏对全局信息建模能力,受到了以自注意力机制为基础的Vision Transformer<sup>[16]</sup>和基于窗口的自注意力机制设计的Swin Transformer<sup>[17]</sup>等可以捕获图像中存在的长距离依赖的算法的挑战。近年来Transformer密集出现在图像修复<sup>[18]</sup>、去噪<sup>[19]</sup>等图像质量增强的任务中并达到最佳性能,但这些方法均在空间域进行而忽略频域特征的利用。文献[20]表明傅里叶变换能够将图像分解为不同频率的正弦和余弦函数并提供图像全局频率信息,因此通过在深度网络中结合频域特征可以促进网络对非局部特征的学习,针对噪声严重影响图像质量的情况,将图像变换到频域进行高低频分离以及频率选择,这可以保留并强化目标特征,同时完成背景噪声的去除,这对目标检测任务有积极意义。本文认为现有基于深度学习的SAR目标检测存在以下不足:(1)采用逐级下采样进行多尺度特征提取与融合通常只在空间域进行,不仅缺乏频域信息的利用还导致训练结果缺乏频响多样性;(2)多次下采样忽略了下采样步长过大导致信息丢失的问题,这使得特征域中存在噪声将目标特征淹没的情况。这些挑战限制了目标检测在SAR图像上的发展。因此,设计面向SAR图像特性的高效目标检测算法具有重要的研究意义,本文针对性地提出了一种基于动态频域卷积的SAR目标检测算法,其中包含有效提取频域特征的动态频域卷积模块,该模块是构成特征金字塔网络(Feature pyramid network, FPN)的主要组件;以及能够克服下采样导致的频率混叠现象的增强目标特征的空间-频域特征增强模块(Spatial-frequency feature enhancement module, SFFEM)。结果表明所设计的模块搭建的目标检测网络在现有公开数据集上取得了最佳的性能,证明了所提模块的有效性。

## 1 基于特征修复的动态频域建模与空间-频域特征增强机制

本文提出了一种基于特征修复的动态频域卷积模块(Dynamic frequency convolution module, DFCM),充分提取目标的潜在特征并减弱噪声干

扰增强目标特征,参考 Transformer 设计了 SF-FEM。基于以上两个模块,本文采用经典的金字塔结构设计了一个全新的 SAR 目标检测网络,整体网络架构如图 1 所示,其中包含 4 个主要模块,即 (1) 经过预训练的 ResNet-50 构成的主干网络 (Backbone),其主要作用是提取输入图像的多尺度特征实现对不同尺度目标语义信息的获取;(2) 采用 DFCM 模块设计的 FPN,其作用是充分融合主

干网络提取的多尺度特征,保持高语义表达的同时增强对不同尺寸目标的检测能力;(3) 由 SFFEM 模块构成的特征增强网络,通过有效整合空间和频域信息增强目标特征并抑制背景信息;(4) 全卷积单阶段目标检测 (Fully convolutional one-stage object detection, FCOS) 网络检测头,对输出的多尺度特征图进行密集预测,对不同尺寸的目标进行识别和定位。

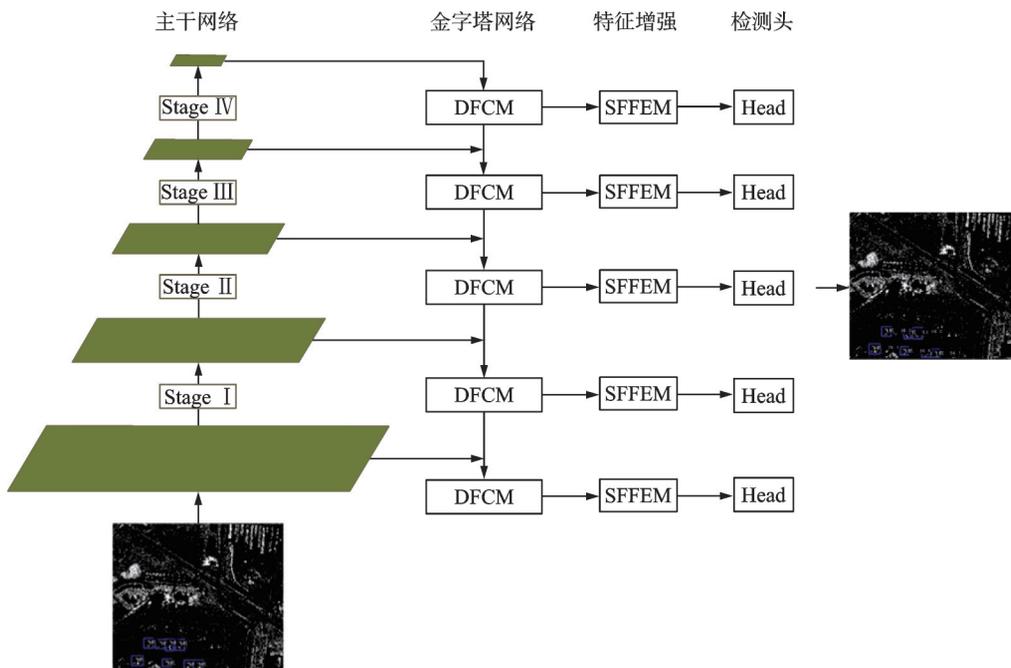


图 1 所提算法总体架构

Fig.1 Overall architecture of proposed method

### 1.1 DFCM

动态卷积通过结合多个并行权重与注意力机制实现自适应权重选择获得优异性能。但实践过程中发现并行权重的频响通常具有比较高的相似性,这一特性限制了模型对不同频带特征的捕获能力,虽然网络参数量大但对目标场景的适应能力依旧不足。针对以上问题,本节提出频域动态卷积模块,内容如下:

(1) 设计初始权重参数为  $W \in \mathbf{R}^{k \times k \times C_m \times C_m}$ , 将这些参数作为傅里叶域谱系数并重塑为  $F \in \mathbf{R}^{kC_m \times kC_m}$ , 并为每个参数设置作为傅里叶域中坐标的傅里叶索引  $(u, v)$ 。按照索引的  $L_2$  范数从低到高将参数重排并均匀分为  $n$  组  $\{F_0, F_1, \dots, F_{n-1}\}$ , 分别表示低频到高频成分。通过逆变换将各组参数转换到空间域,表达式为

$$S_i^{p,q} = \sum_{u=0}^{kC_m-1} \sum_{v=0}^{kC_m-1} F_i^{u,v} e^{i2\pi \left( \frac{p}{kC_m} u + \frac{q}{kC_m} v \right)} \quad (1)$$

式中:  $F_i^{u,v}$  表示第  $i$  组参数中傅里叶索引为  $(u, v)$  的参数,若该索引属于该组参数则保留原始值即  $F_i^{u,v} = F^{u,v}$ , 否则  $F_i^{u,v} = 0$ 。经过逆变换到空间域后

的  $S_i^{p,q}$  表示坐标在  $(p, q)$  处的元素,之后通过将  $S_i$  裁切重排为  $W_i \in \mathbf{R}^{k \times k \times C_m \times C_m}$ 。考虑到参数分组是根据频率进行的,因此  $S_i$  只会包含特定频段的频率组分进而多组权重  $W_i$  表现出不同的频响。参考传统动态卷积,本文将输入特征经过全局平均池化和全连接层生成动态权重  $\{\pi_1, \pi_2, \dots, \pi_n\}$  调整卷积核的贡献度,这使得网络可以根据输入进行自适应调整。

(2) 为了提升网络对每个  $k \times k$  卷积核的调整能力。本文设计了用于生成调制矩阵  $\kappa \in \mathbf{R}^{k \times k \times C_m \times C_m}$  的特征修复模块,其结构如图 2 所示。该模块对高于和低于奈奎斯特频率的频率分量进行自适应加权,表达式为

$$f'(c, h, w) = A^\dagger(c, h, w) * f^\dagger(c, h, w) + A^\ddagger(c, h, w) * f^\ddagger(c, h, w) \quad (2)$$

式中:  $f$  和  $f'$  分别表示特征修复模块的输入和输出特征图,  $c, h$  和  $w$  分别表示特征通道、高度和宽度,  $f^\dagger$  和  $f^\ddagger$  分别表示采用的高通、低通滤波器在傅里叶域中获取的高于和低于奈奎斯特频率的频率分量,  $A^\dagger$  和  $A^\ddagger$  分别是混合不同频率特征的权重。然而,

直接对  $A^{\uparrow}, A^{\downarrow} \in \mathbb{R}^{C \times H \times W}$  中的参数进行预测会引入较为沉重的计算负担,因此本文将权重预测分解到通道维度和空间维度分别进行,表达式为

$$f^{\uparrow}(c, h, w) = A_1^{\uparrow}(c) * A_2^{\uparrow}(h, w) * f^{\uparrow}(c, h, w) +$$

$$A_1^{\downarrow}(c) * A_2^{\downarrow}(h, w) * f^{\downarrow}(c, h, w) \quad (3)$$

式中:通道维度的权重值预测通过直接对特征图做全局平均池化后接一个全连接层实现,空间维度的权重值预测则采用卷积层实现。

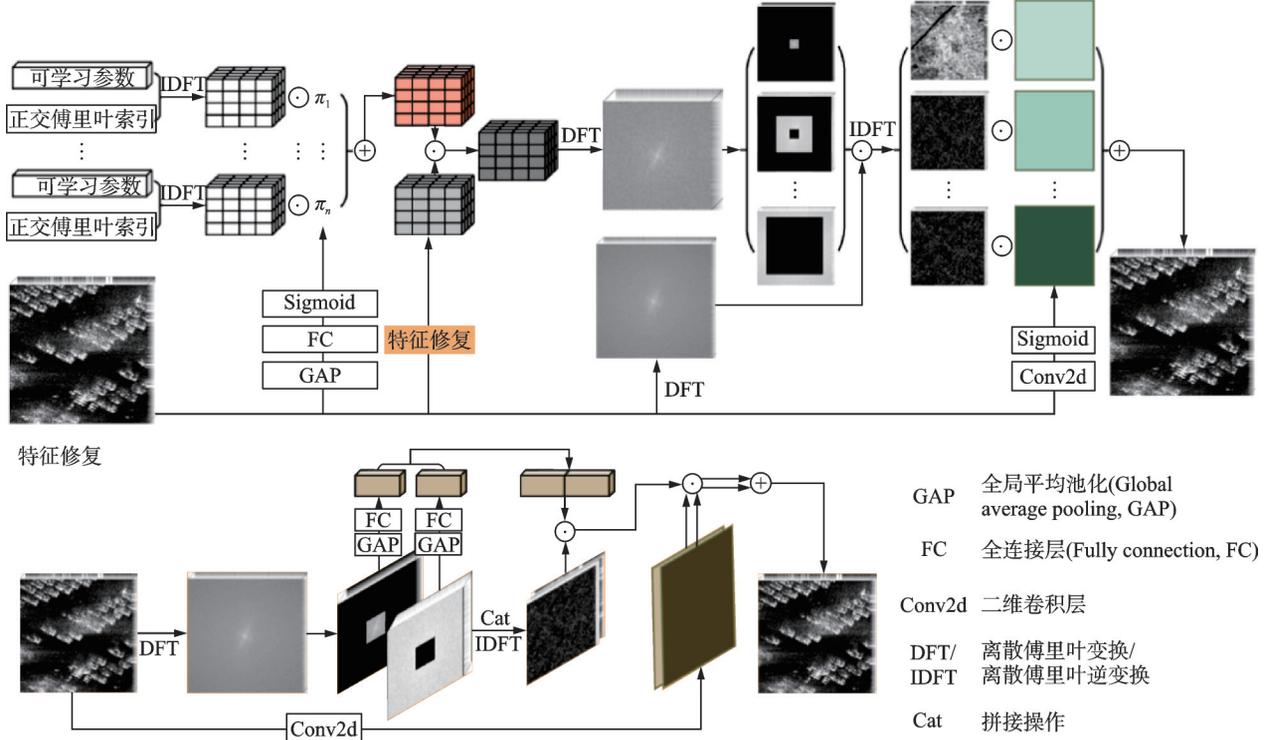


图2 频域动态卷积模块结构

Fig.2 Architecture of frequency dynamic convolution module

(3)由于经典动态卷积的权重在整个特征图上是共享的,这种空间不变性限制了卷积层通过动态调整频响捕获复杂空间特征的能力。本文针对这一问题将变换到频域的卷积核分解为多个子带,并将各个子带的卷积核填充至输入特征图尺寸后,实现在频域的卷积操作,最后对输入特征进行卷积和非线性变换生成调制图以调控不同频带对不同空间位置的影响。表达式为

$$Y = \sum_{b=0}^{B-1} ((F^{-1}(W_b \odot F(X))) \odot A_b) \quad (4)$$

式中: $Y, X$ 分别表示输出和输出特征, $W_b, A_b$ 分别表示第 **$b$** 个频带的权重和空间调制图, $F(\cdot)$ 表示傅里叶变换。

## 1.2 SFFEM

为了对空间和频域中的长程依赖进行建模以增强目标潜在特征,本文提出了空间-频域特征增强模块,其架构如图3所示,该模块由频率自注意力、空间自注意力和空频信息融合模块组成,使用该模块可以进一步整合纹理、颜色和光谱等信息,实现了对特征的多域联合增强。

**频率自注意力:**为了更好地处理SAR图像的频域信息,本节参考自注意力机制设计了频率自注

意力模块,其具体结构如图3所示。通过该模块本文实现了对不同频率信息间相关性的建模,为目标所处的频带分配更大的注意权重并抑制背景。该模块将输入 $P$ 通过离散傅里叶变换获取频域中的查询 $Q_f$ 、键 $K_f$ 和值 $V_f$ 即 $Q_f = \text{DFT}^Q(P)$ 、 $K_f = \text{DFT}^K(P)$ 、 $V_f = \text{DFT}^V(P)$ ,其中 $\text{DFT}(\cdot)$ 表示离散傅里叶变换。随后将 $Q_f \in \mathbb{R}^{H \times W \times C}$ 和 $K_f \in \mathbb{R}^{H \times W \times C}$ 重塑为 $\tilde{Q}_f \in \mathbb{R}^{C \times HW}$ 和 $\tilde{K}_f \in \mathbb{R}^{HW \times C}$ ,并对重塑后的查询和键进行点积得到频域注意力图 $\mathbf{A}_f = \tilde{Q}_f \odot \tilde{K}_f$ 。由于频域注意力图是复数,无法直接使用Softmax处理,本文提取实部 $\mathbf{A}_f^{\text{real}}$ (即 $\mathbf{A}_f^{\text{real}} = (\mathbf{A}_f + \text{Conj}(\mathbf{A}_f))/2$ )和虚部 $\mathbf{A}_f^{\text{img}}$ (即 $\mathbf{A}_f^{\text{img}} = (\mathbf{A}_f - \text{Conj}(\mathbf{A}_f))/2i$ ),其中 $\text{Conj}(\cdot)$ 代表共轭复数。随后,分别对实部和虚部进行激活并将两者进行拼接得到激活注意力图 $a\mathbf{A}_f$ ,其数学表达式为

$$a\mathbf{A}_f = \Theta(\text{Softmax}(\mathbf{A}_f^{\text{real}}), \text{Softmax}(\mathbf{A}_f^{\text{img}})) \quad (5)$$

式中: $\Theta(\cdot, \cdot)$ 表示将实部和虚部重组为复数形式; $\text{Softmax}(\cdot)$ 代表Softmax激活函数。通过激活注意力图优化频域特征 $V_f$ 的权重并采用离散傅里叶逆变换 $\text{IDFT}(\cdot)$ 将其转换到空间域。最后为增强频域信息引入频率残差连接 $P_f$ ,通过 $1 \times 1$ 卷积特征

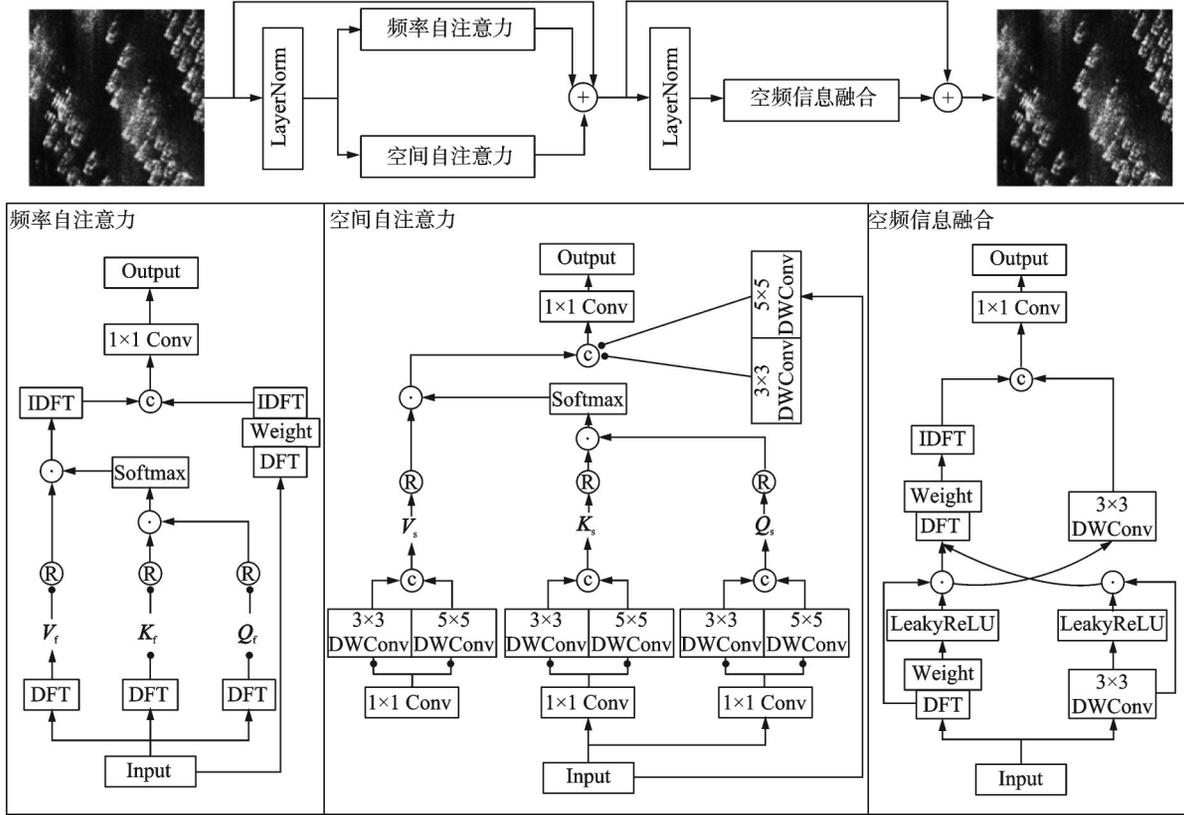


图 3 空间-频域特征增强模块具体架构

Fig.3 Architecture of spatial-frequency feature enhancement module

融合得到频域特征  $F_f^1$ , 公式表达为

$$F_f^1 = \text{Conv}_{1 \times 1} \text{Cat}(\text{IDFT}(a\mathbf{A}_f \odot \tilde{V}_f), P_f) \quad (6)$$

式中:  $\text{Cat}(\cdot, \cdot)$  表示拼接操作, “ $\odot$ ” 表示矩阵乘法,  $\tilde{V}_f$  为将  $V_f$  重排列后的特征。

空间自注意力: 由于 SAR 图像中目标类型多样且目标尺寸各异, 为了捕获多尺度特征, 本节采用空间自注意力, 具体结构如图 3 所示。对于输入特征  $P$ , 通过  $1 \times 1$  卷积对目标特征进行初步提取, 随后采用两个尺寸分别为  $3 \times 3$  和  $5 \times 5$  的深度可分离卷积 (Depth-wise separable convolution, DWConv) 得到查询  $Q_s$ 、键  $K_s$  和值  $V_s$ 。随后将  $Q_s$  和  $K_s$  重塑为  $\tilde{Q}_s$  和  $\tilde{K}_s$  并做矩阵乘法和 Softmax 激活得到注意力权重  $a\mathbf{A}_s = \text{Softmax}(\tilde{Q}_s \odot \tilde{K}_s)$ , 将  $a\mathbf{A}_s$  与值  $V_s$  做矩阵乘法得到空间注意力图。同理, 为了增强空间局部信息引入残差连接得到空间特征  $F_s^1$ , 表达式为

$$F_s^1 = \text{Conv}_{1 \times 1} \text{Cat}(a\mathbf{A}_s \odot \tilde{V}_s, P_s) \quad (7)$$

式中:  $P_s = \text{Cat}(\text{DWConv}_{3 \times 3}(P), \text{DWConv}_{5 \times 5}(P))$ ,  $\tilde{V}_s$  为  $V_s$  重排列后的特征。

空频信息融合: 频域特征和空间特征包含不同的信息, 其中频域特征包含全局能量分布及变化, 空间信息包含像素相关的局部结构信息, 对包含不同信息的两种特征进行有效融合有助于提升网络的表达能力。针对以上特点本文设计了空频信息融合模块, 结构如图 3 所示。对经过加和操作及残差

连接进行初步融合的特征  $\hat{P} = F_f^1 + F_s^1 + P$ , 将其作为输入进一步融合。输入经过  $3 \times 3$  DWConv 和离散傅里叶变换分别生成空间和频域组分, 经过 LeakyReLU 激活函数和门控机制分别得到了频域特征  $F_f^2$  和空间特征  $F_s^2$ , 数学表示为

$$\begin{cases} F_f^2 = \text{LeakyReLU}(\sigma(\text{DFT}(\hat{P})) * \text{DFT}(\hat{P})) * \\ \quad \sigma(\text{DFT}(\hat{P})) * \text{DFT}(\hat{P}) \\ F_s^2 = \text{LeakyReLU}(\text{DWConv}_{3 \times 3}(\hat{P})) * \\ \quad \text{DWConv}_{3 \times 3}(\hat{P}) \end{cases} \quad (8)$$

式中:  $\text{LeakyReLU}(\cdot)$  表示激活操作。最后, 将得到的两种特征进行交叉学习和拼接处理得到最终的输出特征  $P_{\text{out}}$ , 表达式为

$$P_{\text{out}} = \text{Conv}_{1 \times 1} \text{Cat}(F_f^3, F_s^3) \quad (9)$$

其中

$$F_f^3 = \text{IDFT}(\sigma(\text{DFT}(\text{Cat}(F_f^2, F_s^2))) * \text{DFT}(\text{Cat}(F_f^2, F_s^2))) \quad (10)$$

$$F_s^3 = \text{DWConv}_{3 \times 3} \text{Cat}(F_f^2, F_s^2) \quad (11)$$

经过以上多域特征学习、信息交互以及全局频率和局部空间特征的整合, 本文算法具备高效的 SAR 图像特征表达能力, 具备优异的目标检测性能。

## 2 实验结果及分析

### 2.1 数据集及实验环境

为验证本文所提出的 SAR 目标检测算法的性

能与泛化能力,选取两个最具有代表性的 SAR 目标检测数据集,其中包含 MSAR<sup>[4]</sup>和 SAR-Det-100K 数据集<sup>[5]</sup>,这两个数据集不仅涵盖多样化的场景和目标类别,还具备多分辨率、背景复杂度各异等特点,为全面测试算法的鲁棒性与适应性提供了有力支撑。其中,MSAR 数据集的图像均是由中国商业 SAR 卫星海丝一号(HISEA-1)所采集,共包含了 28 449 张图像,涵盖了 4 个目标类别,总计 60 396 个目标实例。采用该数据集主要面临的挑战有:(1)目标外观差异显著,形状受噪声干扰严重;(2)背景环境多变,包括港口、城市等多种强干扰场景;(3)桥梁、飞机等类别占比较低,类别数量不平衡。SARDet-100K 数据集整合自 10 个公开的高质量 SAR 图像数据集,来源包含中国科研机构、欧洲航天部门以及美国军事部门。该数据集覆盖了多种成像条件和地理区域,其中包含 6 种目标类别,总计 245 653 个目标实例。使用该数据集时面临的挑战主要包括:(1)图像中背景杂波干扰严重导致目标与背景对比度低;(2)目标尺度差异极大,从十像素的船舶目标到极大跨度的桥梁均有覆盖;(3)部分目标类别在不同成像模式下表现出显著的外观变化。

所有实验均在 Windows 11 操作系统下进行,硬件配置包括 Intel(R) Core(TM) i7-12700K CPU 以及 NVIDIA GeForce RTX 4090(显存 24 GB) GPU。深度学习框架采用 PyTorch 2.0.1 搭配 CUDA 11.8。为确保评估过程的一致性与公平性,所有算法在训练前均对输入图像进行统一的标准化数据预处理。对于 MSAR 数据集,所有算法均采用 DAdaptAdam 优化器进行训练,训练轮次设为 50,批次大小设为 64,初始学习率设为 1.0,权重衰减参数设为 0.01。对于 SARDet-100K 数据集,训练轮次设为 30,批次大小设为 16,其余训练参数一致。

## 2.2 评价指标

在目标检测任务中,常用的性能评价指标包括精度(Precision)和召回率(Recall)。这两个指标都是基于交并比(Intersection over union, IoU)进行计算,IoU 通过计算预测框与真实框之间的重叠度来反映预测结果与人工标注在空间位置上的一致性,其定义为预测框与真实框之间的交并面积比值,表达式为

$$\text{IoU} = \frac{\text{Area}(B_p \cap B_{gt})}{\text{Area}(B_p \cup B_{gt})} \quad (12)$$

式中: $B_p$ 表示预测框, $B_{gt}$ 表示真实框,“ $\cap$ ”和“ $\cup$ ”分别表示集合的交集和并集操作。通常,当 IoU 超过

预设阈值(如 0.5 或 0.75)时,预测结果即被视为一次正确检测,同时将该判定结果用于计算精度和召回率等指标。当计算得到的 IoU 大于或等于设定的置信度阈值时,预测框被判定为真阳性(True positive, TP),即该预测结果与真实标注匹配。当 IoU 小于设定的置信度阈值时,预测框被判定为假阳性(False positive, FP),表示该预测结果与真实标注不匹配或匹配程度不足。此外,若在真实数据集中不存在标注,但在预测结果中检测到了目标,则该预测被视为真阴性(True negative, TN),即原本无目标但错误检测到目标。相反,如果在真实数据集中存在标注的目标,但预测结果中未能检测到,则该情况被判定为假阴性(False negative, FN),表明该目标被漏检。基于上述 4 类判定结果,可以进一步计算目标检测的精度,其数学表达式为

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (13)$$

该公式反映了所有被预测为正样本的检测结果中真正为正样本的比例,反映了对目标的查准率。而召回率的数学表达式为

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (14)$$

召回率衡量的是在所有真实为正样本的目标中,被正确检测出来的比例,即反映检测算法对目标的查全率。以上两个指标中,高精度表示预测结果更为准确,而高召回率则表示低漏检率。在实际评估中,Precision 与 Recall 往往需要结合使用,以全面衡量目标检测算法的性能。综上,本文采用平均精度(Average precision, AP)作为核心评价指标。为保证评估结果的客观性与可靠性,实验中采用了多种评价方式,包括 AP@50、AP@75、基于插值的  $\text{AP}'_{0.7}$ 、 $\text{AP}'_{1.2}$  等指标。其中基于插值的指标计算方法概述如下:在目标检测任务中,通常采用设定不同的置信度阈值的方法来获得一系列预测结果,并将综合获取的结果进行更为可靠的准确度评价。以 Recall 为横坐标、Precision 为纵坐标,可以绘制出 Precision-Recall(PR)曲线。在此基础上,早期的 PASCAL VOC 2007 评估方法  $\text{AP}'_{0.7}$  在  $[0.0, 0.1, 0.2, \dots, 0.9, 1.0]$  共 11 个等间距的 Recall 点上取对应的 Precision 值,并对这些 Precision 值取平均作为最终精度估计,其数学定义为

$$\text{AP}'_{0.7} = \frac{1}{11} \sum_{r \in \{0.0, 0.1, \dots, 1.0\}} \text{AP}_r \quad (15)$$

然而, $\text{AP}'_{0.7}$  的计算方式存在采样点过少的问题,可能导致对 PR 曲线整体形态的刻画不够准确。为此,研究人员在 VOC 2012 中提出了一种基于积分的改进方法,即  $\text{AP}'_{1.2}$ 。该方法通过对 PR 曲

线进行插值处理,并计算插值曲线与横轴所围成的积分值来衡量模型的综合性能。数学表示为

$$AP'_{12} = \sum_{i=1}^{n-1} (r_{i+1} - r_i) p_{\text{interp}}(r_{i+1}) \quad (16)$$

式中: $r_i$ 表示第*i*个 Recall 取值, $p_{\text{interp}}(r_i)$ 表示在该 Recall 下的插值 Precision 值。当采样点数趋近于无穷大时,该公式即退化为 PR 曲线的积分,即

$$AP'_{12} = \int_0^1 \text{Precision}(\text{Recall})d(\text{Recall}) \quad (17)$$

在获得每个目标类别的平均精度 AP 值后,对所有类别的 AP 取算术平均即可得到平均精度 (mean Average Precision, mAP),其定义为

$$\text{mAP} = \frac{1}{n} \sum_{i=1}^n AP_i \quad (18)$$

式中: $n$ 表示目标类别总数, $AP_i$ 表示第*i*个类别的 AP 值。此外,根据设定的 IoU 阈值通常有两种评价指标,即 AP@50 和 AP@75,AP@50 反映了模型在较为宽松的匹配条件下的检测性能,AP@75 体现了模型在高精度定位方面的能力。因此,综合使用多种阈值下的指标,有助于更全面地评估模型的检测表现。

### 2.3 对比实验及可视化分析

为了验证本文算法的性能,本文在 MSAR 数

据集上进行了对比实验,并选取了基于注意力的可变形多子空间特征去噪的 SAR 图像目标检测 (Attention as deformable multisubspace feature denoising for target detection in SAR images, 记作 DenoDet) 网络<sup>[21]</sup>、全卷积单阶段目标检测 (Fully convolutional one-stage object detection, FCOS) 网络<sup>[22]</sup>、金字塔视觉 Transformer 轻量版 (Pyramid vision Transformer-tiny, PVT-T)<sup>[23]</sup>、Faster 基于区域的卷积神经网络 (Region-based convolutional neural networks, R-CNN)<sup>[24]</sup> 和 仅需聚焦单层级特征 (You only look one-level feature, YOLOF) 网络<sup>[25]</sup> 等现有具有代表性的算法作为对比算法。实验结果如表 1 所示,其中加粗表示性能最优结果,单下划线表示性能次优结果,双下划线表示第 3 优结果,本方法在  $AP'_{07}$  指标下 mAP 达到 70.51%,相较于对比算法分别提升了 1.60%、3.40%、36.64%、14.35% 和 22.66%;在  $AP'_{12}$  指标下 mAP 达到 71.82%,相较于对比算法分别提升了 0.74%、2.27%、36.96%、15.81% 和 24.67%。即使在各个目标类别下,本方法均取得了最优或接近最优的精度表现,进一步验证了其在处理多尺度目标和复杂背景时的鲁棒性与泛化能力。

表 1 MSAR 数据集上的对比实验结果

Table 1 Comparative experiments on MSAR dataset

算法	$AP'_{07}$					$AP'_{12}$				
	mAP	船只	油罐	桥梁	飞机	mAP	船只	油罐	桥梁	飞机
Proposed	<b>70.51</b>	<b>89.54</b>	<u>63.44</u>	79.88	49.18	<b>71.82</b>	<b>91.03</b>	<b>68.48</b>	<b>81.21</b>	<u>46.55</u>
DenoDet	<u>68.91</u>	<u>89.30</u>	<u>63.49</u>	<u>76.30</u>	<u>46.54</u>	<u>71.08</u>	<u>90.87</u>	<u>68.26</u>	<u>78.16</u>	<b>47.02</b>
FCOS	<u>67.11</u>	<u>89.09</u>	63.32	<u>74.17</u>	<u>41.86</u>	<u>69.55</u>	<u>90.15</u>	<u>68.14</u>	<u>77.48</u>	<u>42.43</u>
PVT-T	33.87	60.22	48.24	22.27	4.74	34.86	62.51	47.97	27.61	1.35
Faster RCNN	56.16	81.01	<b>63.52</b>	71.00	9.09	56.01	84.80	66.67	71.17	1.40
YOLOF	47.85	69.90	49.78	62.63	9.09	47.15	72.76	51.10	64.35	0.39

为直观验证所提算法的有效性,在图 4 中展示了从 MSAR 测试集中选取的“船只”“油罐”“飞机”“桥梁”4 类目标的可视化检测结果,其中,紫色框表示该目标被模型判定为船只,绿色框表示该目标被模型判定为油罐,蓝色框表示该目标被模型判定为飞机,红色框表示该目标被模型判定为桥梁。结果表明,DenoDet 表现出了较高的检测准确度,但在面对“飞机”这一强背景干扰目标类型时检测准确度不足。而 FCOS 则面临漏检船只和飞机目标导致的检测精度不足的问题。PVT-T 则在所有目标类别的检测上展示了错检或漏检现象。Faster RCNN 对“船只”“油罐”和“桥梁”表现出不错的检测精度,但面对“飞机”时其检测精度低,整体性能不足。YOLOF 针对“油罐”和“桥梁”目标表现

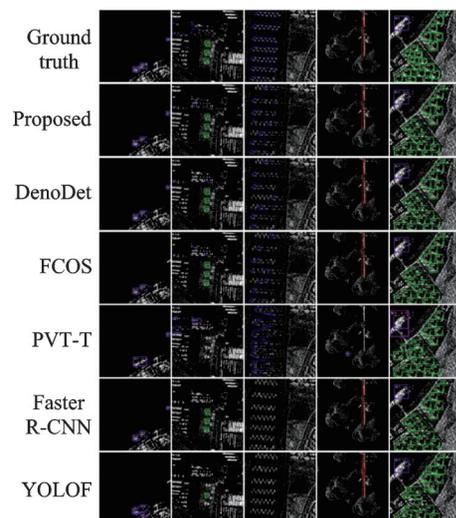


图 4 MSAR 数据集上的可视化检测结果

Fig.4 Visualize detection results on MSAR dataset

较好,但面对“船只”和“飞机”时出现了较为严重的错检和漏检现象,这使得其针对 SAR 目标检测任务时整体表现较差。相比之下,本文提出的算法虽然在“飞机”目标的检测中依然存在一定程度的漏检,但整体上在目标的识别与定位精度方面均优于上述方法,尤其是在复杂干扰环境下,能够更为准确地完成目标的判别与定位。从上述对比结果可以看出,本文所提出的算法在强干扰、弱目标等典型复杂场景下仍然保持了较高的检测与定位精度。为检验模型在大规模场景下的泛化能力,在 SARDet-100K 数据集上开展了对比实验,结果如

表 2 所示,其中加粗表示性能最优结果,单下划线表示性能次优结果,双下划线表示第 3 优结果。结果表明,本方法整体上仍优于对比算法。在以 AP@50 为评价指标时,平均精度达到了 87.07%,并分别提升了 0.86%、0.64%、6.64%、6.64% 和 8.22%;在指标 AP@75 下,平均精度达到 61.94%,并分别提升了 1.58%、5.12%、10.03%、13.20% 和 11.98%;在指标 mAP 下,本方法取得了 57.50%,分别提高了 1.22%、3.84%、9.08%、11.27% 和 10.04%。各目标类别下取得的最佳表现,证明了其稳定性。

表 2 SARDet-100K 数据集上的对比实验结果

Table 2 Comparative experimental results on SARDet-100K dataset

算法	mAP	AP@50	AP@75	船只	飞机	汽车	油罐	桥梁	港口
Proposed	57.50	87.07	61.94	67.06	58.15	65.20	47.53	38.10	68.96
DenoDet	56.28	<u>86.21</u>	60.36	<u>65.35</u>	57.78	<u>64.02</u>	46.20	36.76	<u>67.57</u>
FCOS	<u>53.66</u>	86.43	<u>56.82</u>	<u>60.86</u>	56.53	61.88	<u>42.84</u>	<u>35.27</u>	<u>64.56</u>
PVT-T	48.42	80.43	51.91	55.69	54.31	61.43	32.67	24.91	61.53
Faster R-CNN	46.23	80.43	48.74	57.41	<u>56.86</u>	<u>63.88</u>	31.93	25.70	41.61
YOLOF	47.46	78.85	49.96	58.24	56.11	57.54	27.31	28.44	57.12

此外,在 SARDet-100K 测试集中随机选取“油罐”“飞机”“桥梁”“船只”“港口”5 类目标进行可视化对比,并在图 5 中给出了可视化对比结果,其中,深蓝色框表示该目标被模型判定为飞机,浅蓝色框表示该目标被模型判定为港口,黄色框表示该目标被模型判定为桥梁,红色框表示该目标被模型判定为船只,绿色框表示该目标被模型判定为油罐,紫色框表示该目标被模型判定为汽车。结果显示: DenoDet 在“港口”和“汽车”出现错检;FCOS 在“飞机”错检及多类别漏检;PVT-T 在“船只”上表现较

好,但在“港口”“油罐”“桥梁”均严重错漏检;Faster R-CNN 在“港口”“桥梁”“汽车”错检明显;YOLOF 在“桥梁”错检,在“船只”“油罐”漏检。而本文方法仅在复杂背景下的“油罐”和“船只”目标中出现少量漏检,其余类别均优于对比方法。

由以上实验结果及对比可以看出,本文提出的算法在复杂背景干扰的 SAR 图像中表现出优异的检测精度和鲁棒性,显著优于现有高性能方法。但在杂波干扰严重的场景中仍存在少量误检与漏检,但这也点明了未来的研究工作中亟待改进的方向,即如何克服复杂噪声干扰。

2.4 消融实验

为了证明本文提出的 DFCM 和 SFFEM 在网络架构中确实发挥了无可替代的作用,这一节将所提出的模块逐步替换为卷积模块进行消融实验。实验采用的消融策略如表 3 所示,其中“√”表示这一模块保留在现有网络框架中,而“×”表示将其移除。Case 1 表示本文所提出的原始网络,Case 2 表示将 DFCM 替换为卷积模块并保留 SFFEM 的情况,Case 3 表示将本文所提出的所有模块替换为卷

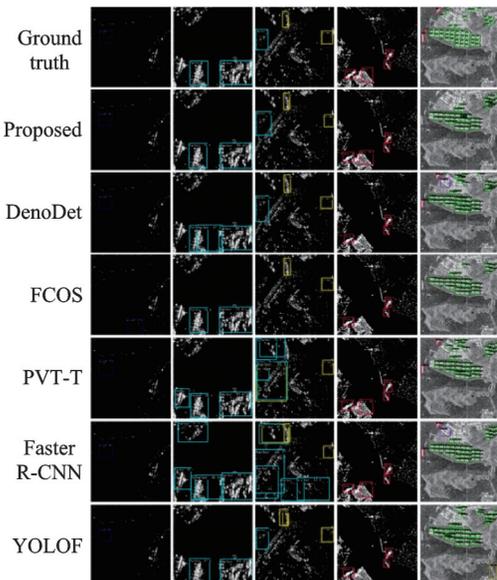


图 5 SARDet-100K 数据集上的可视化检测结果

Fig.5 Visualize detection results on SARDet-100K dataset

表 3 消融实验策略

Table 3 Strategy of ablation experiments

Case	DFCM	SFFEM
1	√	√
2	×	√
3	×	×

积模块的情况。消融实验采用 MSAR 作为实验数据集,评价指标采用  $AP'_{07}$  和  $AP'_{12}$ ,实验结果见表 4。其中,单下划线标注的结果为最佳结果,双下划线标注的结果为次优结果。可以看出,保留 DFCM 和 SFFEM 的网络相较于替换 DFCM 的网络和替换所有模块的网络在以  $AP'_{07}$  作为评价指标时 mAP 分别提升 1.18% 和 4.63%,4 种目标类型的检测精度分别提高 0.21%、0.17%、2.68%、1.67% 和 0.5%、0.09%、8.66%、9.29%。在以

$AP'_{12}$  作为评价指标时 mAP 分别提升了 1.04% 和 4.64%,4 种目标类型的检测精度分别提高 0.76%、0.11%、2.5%、0.78% 和 1.41%、0.2%、5.19%、11.74%。表 5 展示不同情况下本文所提出算法以浮点运算数衡量的计算复杂度和参数量对比,验证了本文所提出模块在未引入过大计算负担的同时获取了优异检测性能。以上结论表明所提出的 DFCM 和 SFFEM 对于提升 SAR 目标检测精度的有效性。

表 4 在 MSAR 数据集上的消融实验结果

Table 4 Ablation experimental results on MSAR dataset

Case	$AP'_{07}$					$AP'_{12}$				
	mAP	船只	油罐	桥梁	飞机	mAP	船只	油罐	桥梁	飞机
1	<u>70.51</u>	<u>89.54</u>	<u>63.44</u>	<u>79.88</u>	<u>49.18</u>	<u>71.82</u>	<u>91.03</u>	<u>68.48</u>	<u>81.21</u>	<u>46.55</u>
2	<u>69.33</u>	<u>89.33</u>	63.27	<u>77.20</u>	<u>47.51</u>	<u>70.78</u>	<u>90.27</u>	<u>68.37</u>	<u>78.71</u>	<u>45.77</u>
3	65.88	89.04	<u>63.35</u>	71.22	39.89	67.18	89.62	68.28	76.02	34.81

表 5 运算复杂度及参数量分析

Table 5 Computational complexity and parameter number analysis

Case	浮点运算次数/ $10^9$	参数量/ $10^6$
1	19.90	50.98
2	16.88	41.46
3	16.11	40.64

### 3 结 论

针对 SAR 目标检测中存在多尺度特征建模不足与频域信息利用有限的问题,本文提出了一种融合频域动态卷积与空间-频域特征增强的 SAR 目标检测网络,即通过引入可学习的傅里叶谱系数与特征修复机制提升模型频率响应的多样性,并结合频率与空间自注意力实现多域信息的高效提取及融合。在两个具有代表性的大规模数据集 MSAR 与 SARDet-100K 上进行的实验表明,本文所提方法相对于多种现有高性能目标检测算法实现了较大的性能提升,即使在复杂背景干扰和弱目标条件下依旧展现出了不错的检测精度,验证了其良好的适应性与应用潜力。然而,在目标尺度极小、背景噪声高度复杂的场景下,所提算法同样存在一定程度的误检与漏检。未来的研究可进一步探索动态特征建模、跨尺度上下文交互、动态频域分解等方向,以推动 SAR 目标检测在极端复杂场景中的稳定性和准确性。

#### 参考文献:

- [1] CUI Zongyong, LI Qi, CAO Zongjie, et al. Dense attention pyramid networks for multi-scale ship detection in sar images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(11): 8983-8997.
- [2] JIA Peng, DONG Tiancheng, WANG Taoyang, et al. Dual-domain joint dense multiple small ship target detection algorithm for spaceborne SAR images[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2024, 41(6): 725-738.
- [3] XU Xiaowo, ZHANG Xiaoling, ZENG Tianjiao, et al. Group-wise feature fusion R-CNN for dual-polarization SAR ship detection[C]//Proceedings of IEEE Radar Conference. [S.l.]: IEEE, 2023: 1-5.
- [4] 陈杰, 黄志祥, 夏润繁, 等. 大规模多类 SAR 目标检测数据集-1.0[J/OL]. 雷达学报, 2022, 14: 1488. CHEN Jie, HUANG Zhixiang, XIA Runfan, et al. Large-scale multi-class SAR image target detection dataset-1.0[J/OL]. Journal of Radars, 2022, 14: 1488.
- [5] LI Yuxuan, LI Xiang, LI Weijie, et al. SARDet-100K: Towards open-source benchmark and toolkit for large-scale sar object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 128430-128461.
- [6] CHEN Linwei, FU Ying, GU Lin, et al. Spatial frequency modulation for semantic segmentation[J]. arXiv preprint, 2507.11893, 2025.
- [7] CHEN Linwei, GU Lin, FU Ying. Frequency-dynamic attention modulation for dense prediction[J]. arXiv preprint, 2507.12006, 2025.
- [8] ZHOU Zheng, CHEN Jie, HUANG Zhixiang, et al. HRLE-SARDet: A lightweight SAR target detection algorithm based on hybrid representation learning enhancement[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-22.
- [9] VERELST T, TUYTELAARS T. Dynamic convolutions: Exploiting spatial sparsity for faster inference

- [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 2320-2329.
- [10] GRABINSKI J, JUNG S, KEUPER J, et al. Frequency lowcut pooling-plugin and play against catastrophic overfitting[C]//Proceedings of European Conference on Computer Vision. Cham, Switzerland: Springer, 2022: 36-57.
- [11] JIANG Zhenyu, LI Xiaodong, DU Chen, et al. YOLO-v8 with multidimensional attention and upsampling fusion for small air target detection in radar images[J]. Transactions of Nanjing University of Aeronautics & Astronautics, 2024, 41(6): 710-724.
- [12] ZHANG R. Making convolutional networks shift-invariant again[C]//Proceedings of International Conference on Machine Learning. [S.l.]: PMLR, 2019: 7324-7334.
- [13] ZOU Xueyan, XIAO Fanyi, YU Zhiding, et al. Delving deeper into anti-aliasing in convnets[J]. International Journal of Computer Vision, 2023, 131(1): 67-81.
- [14] HOSSAIN M T, TENG S W, LU G, et al. Anti-aliasing deep image classifiers using novel depth adaptive blurring and activation function[J]. Neurocomputing, 2023, 536: 164-174.
- [15] LI Qiufu, SHEN Linlin, GUO Sheng, et al. Wavecnet: Wavelet integrated cnns to suppress aliasing effect for noise-robust image classification[J]. IEEE Transaction on Image Process, 2021, 30: 7074-7089.
- [16] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth  $16 \times 16$  words: Transformers for image recognition at scale[J]. arXiv preprint, 2010.11929, 2020.
- [17] LIU Ze, LIN Yutong, CAO Yue, et al. Swin transformer: Hierarchical vision Transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision.[S.l.]: IEEE, 2021: 10012-10022.
- [18] LIANG Jingyun, CAO Jiezhong, SUN Guolei, et al. Swinir: Image restoration using swin Transformer [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2021: 1833-1844.
- [19] LI Miaoyu, LIU Ji, FU Ying, et al. Spectral enhanced rectangle Transformer for hyperspectral image denoising[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2023: 5805-5814.
- [20] LI Z, KOVACHKI N, AZIZZADENESHELI K, et al. Fourier neural operator for parametric partial differential equations[J]. arXiv preprint, 2010.08895, 2020.
- [21] DAI Yimian, ZOU Minrui, LI Yuxuan, et al. Deno-det: Attention as deformable multi-subspace feature denoising for target detection in sar images[J]. IEEE Transactions on Aerospace and Electronic Systems, 2024, 61(2): 4729-4743.
- [22] TIAN Zhi, SHEN Chunhua, CHEN Hao, et al. Fcos: Fully convolutional one-stage object detection [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2019: 9627-9636.
- [23] WANG Wenhai, XIE Enze, LI Xiang, et al. Pyramid vision Transformer: A versatile backbone for dense prediction without convolutions[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2021: 568-578.
- [24] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2015, 28: 91-99.
- [25] CHEN Qiang, WANG Yingming, YANG Tong, et al. You only look one-level feature[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 13039-13048.

(编辑:陈珺)