

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-16

论文引用格式: Wang Hanbin, Hu Haiyang, Ma Zhejie, Zhou Qiaoping. Integrating visual Mamba and frequency-domain priors for lightweight SAR ship detection[J/OL]. Journal of Image and Graphics, XXXX: 1-16. DOI: 10.11834/jig.260395. (王涵彬, 胡海洋, 马哲杰, 周桥平. 融合视觉 Mamba 与频域先验的轻量化 SAR 舰船目标检测[J/OL]. 中国图象图形学报, XXXX: 1-16. DOI: 10.11834/jig.260395.) [DOI: 10.11834/jig.260395]

融合视觉 Mamba 与频域先验的轻量化 SAR 舰船目标检测

王涵彬^{1,2}, 胡海洋^{1,2}, 马哲杰^{1,2}, 周桥平^{1,2}

1. 杭州电子科技大学计算机学院, 杭州 310018; 2. 杭州电子科技大学上虞科学与工程学院有限公司, 杭州 310018

摘要: 目的 合成孔径雷达(synthetic aperture radar, SAR)图像目标检测面临相干斑噪声干扰与离散散射中心长程关联困难双重挑战。现有卷积神经网络受限于局部感受野,而视觉 Transformer 存在算力瓶颈。为此,本文提出一种兼顾全局长程建模、抗噪与轻量化的新架构。方法 提出频空协同状态空间网络(frequency-spatial Mamba network, FSM-Net)。首先,构建可学习频域门控激活机制自适应滤除高频散斑噪声,提供纯净状态先验;其次,在主干引入选择性状态空间模型,融合深度可分离卷积与二维多向连续扫描,在线性复杂度下实现离散散射中心全图级拓扑关联;最后,设计自适应跨流融合机制,利用滑窗无参注意力与重参数化通道投影,以低推理开销补偿一维序列展开导致的高频空间细节丢失。结果 在 SAR 舰船检测数据集(SAR ship detection dataset, SSDD)和高分辨率 SAR 图像数据集(high-resolution SAR images dataset, HRSID)上进行对比实验。FSM-Net 在参数量仅约 8.72 M 的条件下,精度显著优于主流基线及视觉 Mamba(vision Mamba, VMamba)等模型。在 SSDD 数据集上,均值平均精度(mean average precision, mAP)与 mAP50-95 分别提升 1.9% 和 1.5%;在 HRSID 数据集上分别提升 1.5% 和 1.1%。结论 本文提出的 FSM-Net 有效打通了频域抗噪先验与空域长程演化的逻辑闭环,在保持模型轻量化计算开销的前提下,有效弥补了状态空间模型的局部感知短板,为复杂场景下的高精度轻量化 SAR 目标检测提供了极具潜力的全新范式。

关键词: SAR 舰船目标检测;选择性状态空间模型;频空协同;深度可分离卷积;无参注意力

Integrating visual Mamba and frequency-domain priors for lightweight SAR ship detection

Wang Hanbin^{1,2}, Hu Haiyang^{1,2}, Ma Zhejie^{1,2}, Zhou Qiaoping^{1,2}

1. Department of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China; 2. Shangyu Institute of Science and Engineering, Hangzhou Dianzi University, Hangzhou 310000, China

Abstract: Objective Synthetic aperture radar (SAR) is crucial for all-weather, all-day high-resolution Earth observation. However, accurate ship detection in SAR images faces two inherent physical and computational dilemmas. First, the coherent imaging mechanism generates severe multiplicative speckle noise, blurring target boundaries and inducing high false alarm rates in complex ocean clutter. Traditional methods and conventional convolutional neural networks (CNNs) struggle to filter this noise without losing structural information. Second, large artificial targets like ships often appear as discrete

收稿日期: 2026-07-01; 修回日期: 2026-10-03

基金项目: 国家自然科学基金项目(62202133, 61572162, 61802095); 浙江省重点研发计划“领雁”项目(2023C01145); 浙江省自然科学基金项目(LQ17F020003)

Supported by: National Natural Science Foundation of China (62202133, 61572162, 61802095); the Zhejiang Provincial Key Science and Technology "LingYan" Project Foundation (No. 2023C01145); the Zhejiang Provincial National Science Foundation of China (No. LQ17F020003)

©中国图象图形学报版权所有

scattering centers, requiring models with global receptive fields to extract long-range structural dependencies. While Vision Transformers (ViTs) introduce global modeling, their quadratic computational complexity restricts deployment on resource-constrained edge computing platforms. To overcome these limitations, this study proposes the Frequency-Spatial Mamba Network (FSM-Net), a novel lightweight architecture harmonizing global long-range modeling with robust speckle noise resistance. **Methods** FSM-Net deeply couples frequency-domain priors with selective state space models (SSMs) through three core components. First, a learnable frequency-guided activation (LFGA) mechanism is introduced at the frontend to prevent non-linear speckle noise amplification. LFGA utilizes a two-dimensional fast Fourier transform (2D-FFT) to map features into the frequency domain. Using a dynamic gating mask, it adaptively filters out high-frequency speckle disturbances while retaining low-frequency structural components, providing a purified initial hidden state prior via an inverse FFT. Second, a Depthwise Separable Convolution Mamba (DSCMamba) block serves as the core detection backbone. It replaces standard convolutions with depthwise separable convolutions (DSConv) to reduce parameters and employs a two-dimensional multi-directional continuous scanning strategy. This flattens feature maps across four directions, allowing state transition equations to traverse physical topologies with linear time complexity to associate discrete scattering centers. Third, an adaptive cross-stream fusion (ACF) mechanism is devised to compensate for local spatial details lost during sequence flattening. ACF integrates a local sliding window parameter-free attention (LSWA) branch, evaluating an analytical energy function based on local neuron variance and mean to precisely activate high-frequency edges. High-dimensional features are then compressed via a structural re-parameterization technique, enabling zero-latency inference. Finally, a spatial-frequency adaptive feature fusion module merges the global Mamba topological features with the ACF local high-frequency edge features. **Results** Extensive comparative experiments on the SAR Ship Detection Dataset (SSDD) and High-Resolution SAR Images Dataset (HRSID) demonstrate that FSM-Net establishes a new state-of-the-art baseline. On SSDD, FSM-Net achieves 98.2% mean average precision (mAP) and 73.0% mAP50-95, surpassing the Vision Mamba (VMamba) baseline by 1.9% and 1.5%. On the complex HRSID dataset containing dense inshore clusters, FSM-Net yields 93.1% mAP and 68.4% mAP50-95, improving upon VMamba by 1.5% and 1.1%. Crucially, the model maintains excellent computational efficiency, requiring only 8.72 million (M) parameters and 21.35 giga floating-point operations (GFLOPs), while reaching an inference speed of 151 frames per second (FPS). Statistical evaluation reveals a remarkably low log-average miss rate (LAMR) of 11.8%, significantly outperforming Faster R-CNN (29.5%) and YOLOv11s (13.5%). Feature activation heatmaps qualitatively confirm that FSM-Net effectively suppresses coastal noise and flawlessly connects discrete scattering points, mitigating target fragmentation and missed detections in dense scenarios. **Conclusion** FSM-Net effectively bridges the gap between frequency-domain speckle suppression priors and spatial-domain long-range evolution. By innovatively integrating the LFGA mechanism, a multi-directional scanning DSCMamba backbone, and the zero-latency ACF module, it completely resolves the dilemma between coherent speckle interference and the long-range association of discrete scattering centers. Adhering to a highly lightweight computational budget, FSM-Net provides a theoretically sound and empirically validated new paradigm for high-precision, real-time SAR target detection in complex maritime scenarios, exhibiting immense potential for future deployment on spaceborne and air-borne edge computing platforms.

Key words: SAR ship detection; selective state space model; frequency-spatial synergy; depthwise separable convolution; parameter-free attention

论文引用格式:[DOI:10.11834/jig.260395]

0 引言

合成孔径雷达(synthetic aperture radar, SAR)作为一种主动式微波遥感成像系统,凭借其独特的物理机制,展现出全天时、全天候的高分辨率观测能

力,能够穿透云层、雨雾等光学传感器难以逾越的障碍,实现对地表目标的持续性监测。近年来,随着高分三号、Sentinel-1等高性能星载SAR系统以及各类机载平台的广泛部署,为海洋资源监测以及灾害紧急预警等领域提供了极其丰富的信息源。在这些复杂多变的应用场景中,如何从海洋杂波剧烈干扰的图像中实现对舰船等高价值目标的精准定位与识

别,始终是遥感信息处理领域的重点研究对象。

然而,SAR图像是相干成像系统对地物散射回波进行矢量叠加后的结果,这种物理机制不可避免地引入了严重的乘性相干斑噪声,这构成了SAR图像目标检测所面临的首要困境。在频域表现上,相干斑通常呈现弥散的高频特性,不仅极大地降低图像的信噪比并且模糊目标的边缘纹理,还容易掩盖弱小目标的有效散射中心,导致检测模型产生较高的虚警率与漏检率(王梓祺等,2024)。

基于统计建模与信号处理的传统检测方法是研究者们最先尝试对抗这一噪声困境的方向。其中,最具代表性的是恒虚警率(constant false alarm rate, CFAR)系列算法(Li等,2022)。这类方法通过对背景杂波的统计分布如高斯分布、Gamma分布等进行假设,在滑动窗口内计算局部阈值以提取异常高亮像素。后续研究还衍生出了诸如单元平均CFAR(cell-averaging CFAR, CA-CFAR)(Medeiros等,2024)以及有序统计CFAR(order statistic CFAR, OS-CFAR)(Li等,2024)等变体,试图提升在非均匀背景下的鲁棒性。此外,早期的预处理方向还包括了基于小波变换的多尺度去噪(Rani等,2024)以及基于非局部均值的纹理保持去噪(Zhu等,2022)。然而,这些非参数化方法存在着致命的缺陷,它们往往高度依赖于人工对杂波模型的先验假设,在面临近岸港口、密集停泊等复杂场景时,参数估计极易失效,且计算开销随图像分辨率的提升而显著增长,泛化能力十分脆弱。

进入深度学习时代后,以卷积神经网络(convolutional neural networks, CNN)为基础的检测框架凭借其强大的非线性特征表征能力,逐渐取代了传统算法的统治地位。研究者通过构建双阶段模型如Faster R-CNN(Cai等,2023)、Cascade R-CNN(Chai等,2024),或是构建单阶段网络如YOLO系列(Guo等,2023)、SSD(Zhou等,2022)、RetinaNet(Miao等,2022)的检测范式。这些新范式的提出,极大提升了对复杂背景的适应性。

而为了针对性地缓解SAR图像独特的相干斑噪声的干扰,一系列改进方向应运而生:一些研究者聚焦基于注意力机制的特征提纯方向,通过引入卷积块注意力模块(convolutional block attention module, CBAM)(Wang等,2023)或跨尺度注意力机制,使网络在空间和通道维度上自适应地抑制背景噪声

权重。另一些研究者则是关注基于特征融合的增强方向,利用特征金字塔网络(feature pyramid network, FPN)(Zhang等,2023)还有路径聚合网络(path aggregation network, PANet)(Hu等,2026)或是引入双树复小波变换(dual-tree complex wavelet transform, DTCWT)(Tian等,2025)将浅层的高频细节与深层的语义信息进行多尺度耦合,防止弱小目标特征在下采样过程中流失。

尽管CNN在一定程度上缓解了噪声问题,但其固有的范式却面临着第二个核心挑战即目标散射中心的离散性与长程结构依赖。SAR图像中的人造目标如大型金属商船通常由多个强散射中心构成(Xiao Zhenjiu等,2024)。在高分辨率图像中,一个狭长目标往往占据数百个像素,其散射中心在物理空间上是离散的,但在语义逻辑上却是高度相关的。CNN受限于局部卷积核的物理尺寸,其感受野的扩大只能依靠网络层数的不断堆叠。这种局部信息聚合方式在处理极端长宽比目标时,可能会显得力不从心,极易将一艘完整的长船误判为多个断裂的局部目标。为了应对这一挑战,研究者探索了诸如空洞卷积以非对称扩大感受野(Pires de Lima等,2023)、可变形卷积(deformable convolutional network, DCN)(Chen等,2023)以自适应调整采样偏移,或是采用图神经网络(graph neural network, GNN)(Hou等,2024)以建立非欧几里得空间的特征关联等方向。

为了从根本上突破CNN局部感受野的局限,视觉Transformer(vision transformer, ViT)被引入遥感领域(Deng等,2023),开启了全局建模的新范式。Transformer的核心在于自注意力(Self-Attention)机制,能够直接计算全图所有像素块之间的相似度,从而在单层网络中建立起跨越长距离的散射中心关联,显著改善了大型目标的完整性。为了平衡计算开销,后续衍生出了诸如Swin Transformer的层次化滑动窗口机制(Pan等,2024)、基于特征采样的稀疏注意力机制(Bao等,2022)等方法,试图在局部细节保持与全局依赖建模之间寻找平衡。然而,Transformer架构始终面临着难以回避的算力瓶颈,标准自注意力的计算复杂度与输入序列长度呈二次方增长,使得模型难以部署在星载、机载等资源受限的边缘计算平台上,限制了其在实时观测场景下的应用价值。

在这一背景下,起源于控制论的现代状态空间模型(state space model, SSM),特别是以 Mamba 为代表的选择性架构(Gu 等, 2023),为解决全局建模与计算效率之间的博弈提供了全新的契机。Mamba 通过引入硬件感知的并行算法与选择性扫描机制,成功在保持全局有效感受野的同时,将计算复杂度压缩至与序列长度呈线性关系的级别,展现出了超越 Transformer 的推理效率与长序列处理能力。在视觉任务中, VMamba (Liu 等, 2024)、VisionMamba (Hatamizadeh 等, 2025)、MBSSNet (Li 等, 2025) 等工作的初步探索已经证明了其在图像分类与通用检测中的巨大潜力。

然而,要将 Mamba 理论真正应用于 SAR 图像目标检测,仍存在两个亟待解决的矛盾,这构成了本文研究的切入点。首先,标准 Mamba 的隐状态演化过程是一个高度非线性的累积过程,对输入信号的频率分布极其敏感。SAR 图像中广泛存在的高频相干斑噪声若不经处理进入状态空间,其误差会顺着序列演化被不断放大,造成有效隐状态的严重污染。而目前针对 SSM 去噪的研究方向尚处于起步阶段,传统的空间域预处理往往会破坏 Mamba 所需的连续信号拓扑。其次, Mamba 为了维持线性复杂度,其核心机制依赖于将二维图像展平为一维序列进行连续扫描。这一拓扑变换不可避免地导致了局部相邻像素间空间相关性的割裂,加之其对高频信号的内生平滑效应,导致模型在刻画舰船边缘、尖锐角反射器等局部细节时,模型往往会表现出明显的感知钝化,极易引发边界定位不准或密集目标的虚警。

因此,针对 SAR 图像的物理特性,对状态空间模型进行适配性改进,设计合理的去噪前端与特征补偿后端,是发挥线性长程建模优势的重要工作。基于上述分析,本文提出了一种轻量化的频空协同状态空间网络。本文的主要贡献可归纳为以下: 1) 提出了一种兼顾全局长程建模与轻量化的检测新架构。将选择性状态空间模型(Mamba)引入 SAR 目标检测。突破传统方案的感受野局限和算力瓶颈。2) 设计了可学习频域门控激活机制。通过频空域可逆变换,在全局频谱空间自适应滤除高频散斑干扰,在不破坏空间拓扑的前提下,为状态演化提供纯净的初始特征先验。3) 构建了自适应跨流融合机制与增强型 Mamba 主干。结合二维多向连续扫描与深度可分离卷积,利用滑窗无参局部感知分支,以极低

的推理开销弥补了一维序列展平导致的局部边界模糊,实现了全局拓扑与局部细节的频空协同闭环。

算法 1:FSM-Net 整体流程

Algorithm 1: The Overall Pipeline of FSM-Net

FSM-Net 算法流程图

```

1://阶段1:频域门控激活(LFGA)
2: $\tilde{X} \leftarrow \text{Stem}(I)$  //浅层特征提取

3:for  $c = 1$  to  $C$  do
4:  $A_c, P_c \leftarrow \text{Amplitude/Phase}(2D - \text{FFT}(X_c))$  //初始特征分解为幅度、相位
5:  $M_g \leftarrow \text{Sigmoid}(\text{Conv}_{1 \times 1}(W_g))$ 
6:  $\tilde{X}_c \leftarrow 2D - \text{IFFT}((A_c \odot M_g) \cdot \exp(jP_c))$ 
7:end for

8:  $h_0 \leftarrow \text{LN}(\text{Lin}(\tilde{X}))$  //生成初始隐状态
9://阶段2:状态选择性长程建模

10:for 每一个 DSCMamba 模块 do
11:  $x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)} \leftarrow \text{CrossScan}(\text{DSConv}(\tilde{X}))$ 
12: for  $k \in \{1, 2, 3, 4\}$  do
13:  $h_i^{(k)} \leftarrow \bar{A}_i^{(k)} h_{i-1}^{(k)} + \bar{B}_i^{(k)} x_i^{(k)}$  //隐状态转移
14:  $y_i^{(k)} \leftarrow C_i^{(k)} h_i^{(k)}$ 
15: end for
16:  $F_g \leftarrow \text{Lin}(\sum_k Y^{(k)}) + \tilde{X}$  //重塑并融合残差
17:end for

18://阶段3:自适应跨流融合(ACF)与输出
19:  $E^* \leftarrow \text{EnergyFunction}(\tilde{X})$  //滑窗无参能量评估
20:  $X_l \leftarrow \text{Sigmoid}(1/E^*) \odot \tilde{X}$  //激活局部高频边缘
21:  $Y_l \leftarrow \text{ReparamConv}(X_l)$  //重参数化降维
23:  $s \leftarrow \text{AvgPool}([F_g, Y_l])$ 
23:  $\alpha \leftarrow \text{Sigmoid}(W_2 \delta(W_1 s))$ 
24:  $F_o \leftarrow \alpha \odot F_g + (1 - \alpha) \odot Y_l$  //频空自适应融合
25: return  $\text{DynamicHead}(\text{FPN}(F_o))$ 

```

注:算法中 $\leftarrow$ 表示变量赋值操作; $\odot$ 表示 Hadamard 积矩阵逐元素相乘; δ 表示 ReLU 非线性激活函数。 $\text{LN}(\cdot)$ 为层归一化操作, $\text{Lin}(\cdot)$ 为线性投影。

1 本文方法

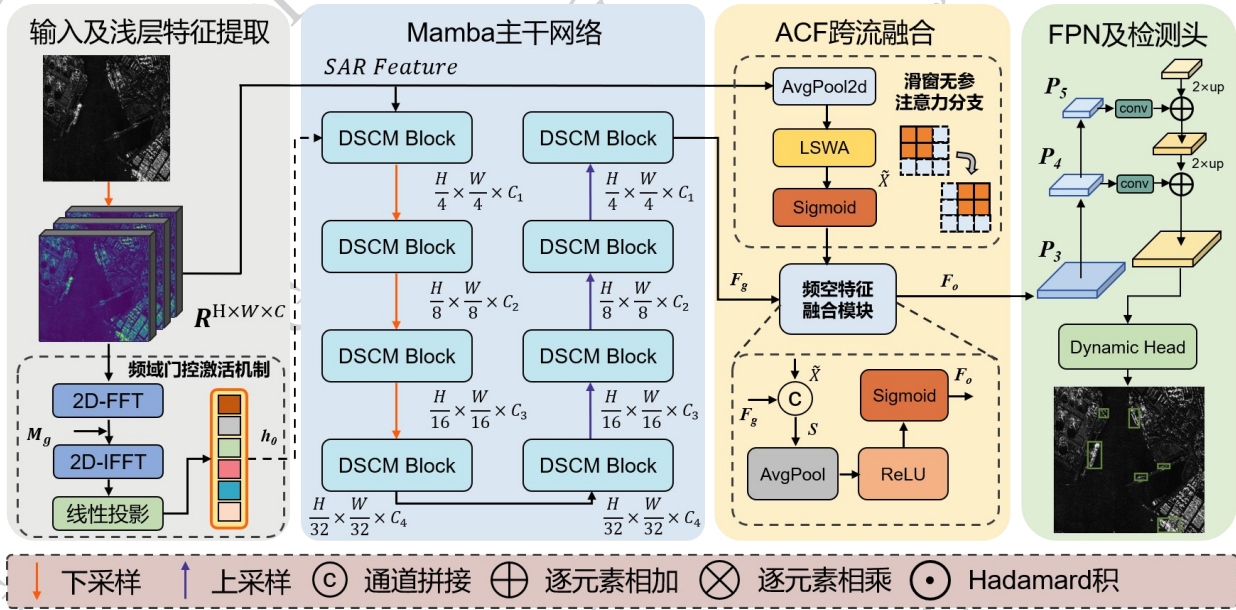
1.1 整体框架

针对复杂背景下 SAR 图像目标检测面临的相干斑噪声干扰、长程结构特征提取困难以及局部细节易丢失等问题,本文提出了一种频空协同状态空间网络(frequency-spatial Mamba network, FSM-Net)。该网络在保持轻量化架构的前提下,通过深度耦合频域先验与空间演化机制,实现了对 SAR 图像中高价值目标的精准定位与识别。

FSM-Net 的总体网络架构以高效的全卷积检测框架为基线,核心创新在于重构了主干网络中的深

层特征表达范式。整体架构如图 1 所示, 主要由浅层特征提取模块、核心频空协同状态空间模块、多尺

度特征融合颈部以及检测头四个部分协同构成。



FSM-Net 的完整执行流程如算法 1 (Algorithm 1) 所示。具体而言, 该算法包含以下三个关键阶段:

1) 频域门控激活 (LFGA): 将特征映射至频域并自适应滤除高频散斑噪声, 为后续模块提供纯净的初始隐状态先验; 2) 长程特征演化 (Mamba 主干层): 利用融合深度可分离卷积的二维多向连续扫描策略, 在线性时间复杂度下实现离散散射中心的全图级关联; 3) 局部细节补偿 (ACF 优化后端): 通过平行无参局部感知分支提取高频边缘特征, 并将其与 Mamba 输出的全局低频特征进行自适应加权融合。最终融合的特征送入 FPN 与解耦检测头, 完成目标类别与边界框的预测。上述各核心模块的具体实现机理将在后续小节详细阐述。

1.2 频域门控激活机制 (LFGA)

在 Mamba 的序列演化中, 初始隐状态若夹杂非结构化噪声, 误差将被累积放大。针对 SAR 图像固有的高频相干斑干扰, 本文提出可学习频域门控激活机制 (LFGA)。由于输入 LFGA 的特征图已通过浅层网络映射, 相干斑虽转化为复杂的高频扰动, 但与其呈低中频分布的连贯舰船主体在空间频域上依然存在解耦特性。基于该先验, LFGA 打破空域卷积限制, 将特征整体映射至频域, 构建动态门控掩码自适应抑制上述高频能量, 从而为后续演化提供纯净的状态底座。

图 1 网络框架图

Fig. 1 Network framework

LFGA 首先执行从二维空间域到频域的特征映射。浅层网络输出的基础特征图为 $X \in \mathbf{R}^{C \times H \times W}$, 其中 C, H, W 分别代表通道数、高度和宽度。为了捕获不同感受野下的局部到全局频域特性, LFGA 引入了多尺度的空间分块策略。对于该特征图数组的第 c 个通道矩阵 X_c , 如算法 1 第 4 行所示利用二维快速傅里叶变换将其转化为复数频谱矩阵 F_c 。而为了便于提取 SAR 目标的散射强度信息, 我们将复数张量分解为幅值谱 A_c 与相位谱 P_c 。通过上述映射, 网络成功将空域中杂乱无章的噪声干扰转化为幅值谱中边缘区域的高频响应, 为后续的参数化过滤奠定了数学基础。具体的模块细节如图 2 所示。

为解决传统滤波方法中固定频率截断方法导致目标高频边缘的丢失的问题, LFGA 提出了一种基于可学习权重的动态频域门控 (Frequency Gating) 机制。具体而言, LFGA 构建了一个与幅值谱具有相同维度的可学习参数矩阵 $W_g \in \mathbf{R}^{C \times H \times W}$ 。该权重矩阵经过激活函数后, 生成范围在 (0, 1) 之间的动态门控掩码 M_g :

$$M_g = \text{Sigmoid}(\text{Conv}_{1 \times 1}(W_g)) \# (1)$$

随后, LFGA 将该门控掩码与原始幅值谱进行逐元素相乘 (Hadamard 积, 记为 $\odot$), 实现对不同频

段能量的软性重新分配:

$$\tilde{A}_c(u,v) = A_c(u,v) \odot M_g(u,v) \# (2)$$

式中, A_c 为幅值谱; (u,v) 为空间坐标; $\odot$ 为 Hadamard 积。在此过程中, 网络会根据检测任务的最终损失自动优化权重矩阵 W_g 。对于代表目标连贯结构的中低频区域, M_g 的权重趋近于 1, 予以无损保留; 对于代表相干斑噪声的极端高频区域, M_g 的权重则趋近于 0 以压制噪声; 而对于目标边缘与噪声交叠的混合频带, 网络通过学习获得一个平滑的过渡权重。

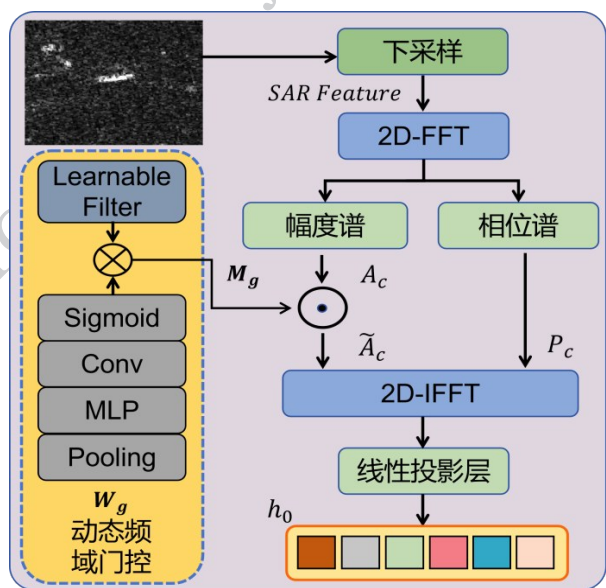


图2 LFGA模块细节图

Fig. 2 LFGA module details

在动态门控过滤后, LFGA 将提纯后的幅值谱 $\tilde{A}_c$ 与未经破坏的原始相位谱 P_c 进行复数重构, 并通过二维逆快速傅里叶变换将特征重新映射回空间域。至此, 输出的特征图 $\tilde{X} \in \mathbf{R}^{C \times H \times W}$ 已经在保留物理空间结构的同时, 完成了深度的频域去噪与特征提纯。该特征 $\tilde{X}$ 通过线性投影层, 专门生成用于 Mamba 长程扫描的初始隐状态条件 h_0 :

$$h_0 = \text{LN}(\text{Lin}(\text{Fl}(\tilde{X}))) \# (3)$$

式中, $\text{LN}(\cdot)$ 为层归一化操作; $\text{Lin}(\cdot)$ 为线性投影操作; $\text{Fl}(\cdot)$ 为展平操作。其在此基础上, Mamba 能够摒弃相干斑噪声的干扰, 专注于利用其线性复杂度的优势, 沿着二维扫描路径将零散的高价值目标散射中心串联起来, 从而实现频域去噪机制与空间状态演化机制的深度物理耦合。

1.3 状态选择性空间长程建模 (Mamba)

为在全局有效感受野与线性计算复杂度之间取得最优折中, 本文在 LFGA 提纯的特征基础上引入选择性状态空间模型 (Mamba)。通过融合深度可分离卷积与多向连续扫描策略, 实现对 SAR 目标离散散射特性的轻量化长程建模。

1.3.1 深度可分离增强的状态空间机制

Mamba 的核心思想源于现代控制论中的状态空间模型 (SSM)。在引入步长参数 Δ 并采用零阶保持等方法将连续系统离散化后, 其演化过程可直接描述为针对输入序列 x_i 和隐状态变量 h_i 的迭代更新方程:

$$h_i = Ah_{i-1} + Bx_i \# (4)$$

$$y_i = Ch_i \# (5)$$

式中, A 为状态转移矩阵, B 和 C 分别为输入与输出投影矩阵。传统的线性时不变系统无法区分输入信息的密度, 容易导致隐状态被杂波淹没。为此, Mamba 引入了动态选择性机制, 使得步长及投影矩阵成为输入 x_i 的动态函数。在处理 SAR 图像时, 当扫描路径经过大面积海面杂波, 模型会自适应输出极小的参数值, 选择忽略杂波并保持历史记忆。而当模型遇到强烈的舰船散射体时, 则会迅速将当前高频结构写入隐状态 h_i 中。由此, 选择性机制在维持线性复杂度的同时, 实现了类似注意力机制的动态特征聚合。

在标准的视觉 Mamba 块中, 其使用的标准卷积在处理深层高维通道时会引入巨大的参数量与计算冗余。针对这一问题, 本文将 Mamba 块中的原生卷积替换为深度可分离卷积 (depthwise separable convolution, DSCConv) (Shen 等, 2025), 整体架构如图 3 所示。

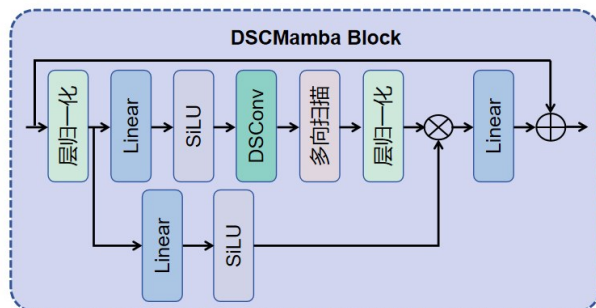


图3 融合深度可分离卷积的状态视觉空间块

Fig. 3 State visual space block with depthwise separable convolution

DSConv 将传统卷积操作巧妙解耦为逐通道的空间卷积和逐像素的通道投影。具体而言,特征序列首先进入逐通道空间卷积阶段,网络为每一个高维输入通道严格分配一个独立的二维空间滤波器进行滑动扫描。随后,再通过逐像素通道投影,即利用轻量的 1×1 卷积核,对上述相互独立的空间特征进行跨通道的线性映射与非线性组合。相比原生卷积,DSConv 在维持相同局部感受野的前提下,有效在堆叠的 Mamba 模块中削减了模型参数量与浮点运算次数,完美契合了本文轻量化 SAR 舰船检测的设计初衷。

1.3.2 面向 SAR 特征的多向连续扫描策略

标准 Mamba 的单向一维拓扑如逐行光栅扫描会撕裂二维特征图中垂直方向的像素相关性,不利于任意方向停泊的 SAR 舰船检测。为了在进行一维状态演化的同时最大限度地保留 SAR 图像的二维全向物理结构,本文参考 VMamba 中将二维信息转换为一维序列的扫描方法(Tu 等, 2025),引入了二维多向连续扫描策略(multi-directional continuous scanning)。考虑到 SAR 图像中目标的方位向与距离向特征分布不均,如图 4 所示本文构建了包含四个基础方向(从左上至右下、从右下至左上、从右上至左下、从左下至右上)的交叉扫描路径。

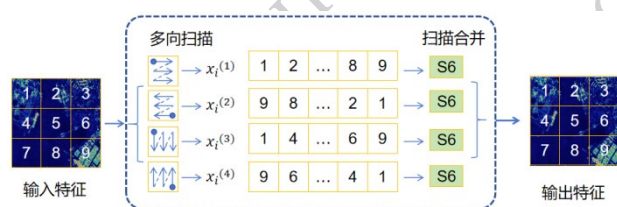


图 4 多向连续扫描

Fig. 4 Multiple continuous scanning

对于输入特征 $\tilde{X}$, 网络并行地将其沿四种路径展平为四个独立的一维序列 $x_i^{(1)}, x_i^{(2)}, x_i^{(3)}, x_i^{(4)}$, 长度均为 $L = H \times W$ 。这四个序列被分别送入共享权重的选择性状态空间, 进行独立的状态演化与特征映射。该机制确保了存在至少一条扫描路径能够沿着目标的主轴方向进行连续遍历, 建立坚实的长程结构依赖。完成四向状态演化后, 输出的一维特征序列 $y_i^{(k)}$ 被按原路径反向重塑回二维空间维度 $Y^{(k)} \in \mathbb{R}^{C \times H \times W}$ 。最后, 网络通过沿通道维度对四个方向的二维特征进行求和与线性投影, 完成多向长

程特征的深度融合, 并将融合结果与原始输入特征 $\tilde{X}$ 进行逐元素相加构成残差连接, 以确保网络在极深层结构中的梯度回传稳定性。经过上述改进, Mamba 模块在保持全局有效感受野的同时, 将其状态演化计算复杂度控制在 $O(L)$ 级别, 打破了高分辨率特征处理的算力瓶颈。这意味着网络能够在维持大分辨率特征输入, 不丢失弱小目标的前提下, 且在更具轻量化特性的前提下实现整图级别的感受野覆盖。

1.4 自适应跨流融合机制 (ACF)

Mamba 在实现线性时间复杂度时, 其核心的连续扫描机制需要将二维特征空间展平为一维序列。这一拓扑变换不可避免地破坏了局部像素间的强相关性, 导致输出的深层特征在刻画目标边缘、尖锐角反射器等局部高频细节时发生退化。

对于 SAR 图像而言, 局部高频细节是区分密集并排舰船、精准回归检测框边界的核心判别依据。为了在不破坏 Mamba 轻量化特性的前提下弥补这一缺陷, 解决全局感受野与局部空间保真度之间的矛盾, 本文提出了一种轻量级的自适应跨流融合机制 (adaptive cross-stream fusion mechanism, ACF)。该机制通过构建 Mamba 感受分支与无参局部感知分支的双流水线并行架构, 并引入重参数化技巧与自适应特征融合, 实现了极低推理开销下的全局拓扑加边缘分支的完美闭环。

1.4.1 局部显著性无参校准

针对序列展平丢失的局部空间细节, 本文构建独立于 Mamba 的局部感知分支。本文引入了受简单无参注意力模块 (simple parameter-free attention module, SimAM) (Yang 等, 2021) 启发的基于能量函数的滑窗无参注意力机制 (local sliding window parameter-free attention, LSWA)。如图 5 所示, 该无参注意力模块摒弃了常规引入大量参数的多层感知机拟合方式, 而是直接构建了一个解析能量函数来尝试捕捉图像中的关键信息。

对于给定的输入特征图 $X \in \mathbb{R}^{C \times H \times W}$, 每一个空间位置点 (h, w) 在特定通道 c 下的响应值均可定义为一个空间神经元。在本文的无参注意力机制中, 每一个神经元 t 被视为一个潜在的显著性信息源。通过评估该神经元与其周围邻域神经元之间的线性可分性, 可量化该点在杂波背景下的重要程度。

我们将特征图 X 中的每一个位置视为一个待
© 中国图象图形学报版权所有

评估的神经元 t 。定义一个 t 为中心、空间尺寸为 $K \times K$ 的局部滑动窗口 Ω_t ，该窗口构成了神经元 t 的局部拓扑领域。窗口内除 t 以外的其他 $M - 1$ 个神经元标记为 x_i (其中 $M = K \times K$)。为了衡量神经元 t 与其邻域 x_i 的重要性，我们构建如下解析能量函数：

$$e_i(w_i, b_i, x_i, t) = \frac{1}{M-1} \sum_{x_i \in \Omega_t, x_i \neq t} (-1 - (w_i x_i + b_i))^2 + (1 - (w_i t + b_i))^2 + \lambda w_i^2 \#(6)$$

该式通过寻找最优的权重 w_i 和偏置 b_i ，试图将目标神经元 t 与其邻域神经元 x_i 进行线性分类。若能量 e_i 的极小值越小，说明神经元 t 与背景 x_i 的差异越显著，其在局部特征提取中的地位就越高。其中， w_i 和 b_i 分别为线性变换的权重和

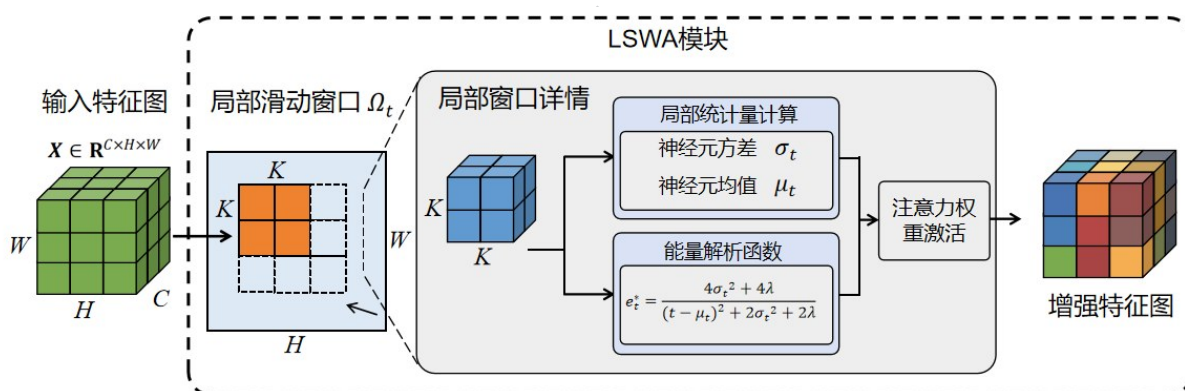


图5 局部滑动窗口无参注意力LSWA模块

Fig. 5 Local sliding window parameter-free attention

偏置， λ 为正则化系数。该能量函数的物理意义在于：当目标神经元 t 与背景 x_i 差异越大时，该方程的极小值就越小。

由于上述能量函数是一个典型的二次无约束优化问题，我们可以引入均值 μ_t 和方差 σ_t^2 ，并由均值 μ_t 和方差 σ_t^2 可以直接求得 w_t 和 b_t 的闭式解 (closed-form solution)。再将闭式解代入原方程，即可得到最小能量 e_t^* 的解析表达式：

$$e_t^* = \frac{4\sigma_t^2 + 4\lambda}{(t - \mu_t)^2 + 2\sigma_t^2 + 2\lambda} \#(7)$$

由此可见，最小能量 e_t^* 与神经元 t 和其邻域均值 μ_t 的差异成反比。基于此，我们通过取能量的倒数来构建三维注意力激活权重，再对原始特征进行校准：

$$\tilde{X} = \text{Sigmoid}\left(\frac{1}{E^*}\right) \odot X \#(8)$$

式中， E^* 为包含了所有通道和空间位置最小能量的张量矩阵， $\odot$ 表示逐元素相乘。该过程可以实现几乎零参数的条件下，精准激活 SAR 目标的局部高频边缘，完成了空间细节的初步找回。

1.4.2 极简通道重参数投影

Mamba 模块为了保证隐状态能够容纳充足的全局上下文信息，其特征输出通道数通常远大于原始输入通道数。若直接将经过显著性校准后的高维特征送入后续融合模块，将带来严重计算冗余。因此，必须在流水线中段引入通道投影机制进行降维。

为了在尽可能压缩通道维度的同时不损失特征的非线性表征能力，本文设计了基于结构重参数化 (structural reparameterization) 的高效投影模块。该模块将训练时的强表征与推理时的零开销进行了有效解耦。

如图6所示，在网络的训练阶段，为了捕获不同感受野下的通道耦合关系，模块采用了多分支并行结构。多分支结构赋予了网络在参数空间中更平滑的损失景观，加速了通道冗余特征的剥离。

而在模型的推理部署阶段，多分支结构会带来额外的内存访问成本和延迟。此时，我们利用卷积与线性变换的可加性定理，将上述结构等价折叠。具体而言，由于批归一化操作实际上是对特征的线性缩放与平移，它可以被很好地吸收到前置的卷积核权重中。基于此，通过对 1×1 卷积核进行四周补

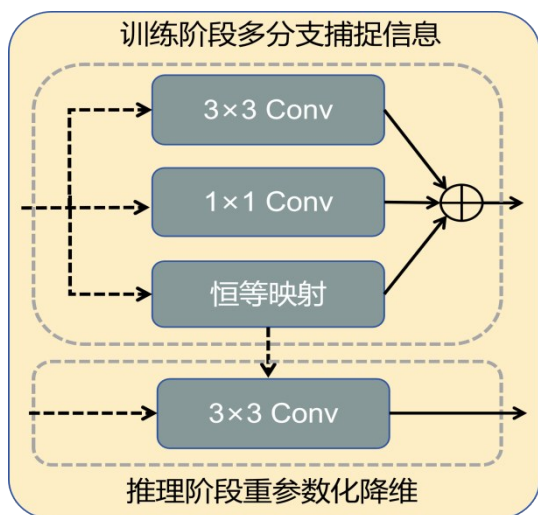


图 6 极简通道重参数投影

Fig. 6 Minimalist channel reparameterization projection

零转化为 3×3 的形状,并将恒等映射分支视作一个中心为 1、其余为 0 的特定 3×3 卷积核,训练期的三个分支最终被重参数化为一个等效的单路卷积算子。在实际推理时,特征只需经过一次精简的单分支卷积运算,即在零额外推理延迟的代价下,完成了高质量的通道降维。

1.4.3 频空特征自适应融合

经过空间无参校准与重参数通道投影,流水线输出的局部特征已具备了锐利的高频轮廓。双流架构的末端需要解决异构特征即 Mamba 的低频长程特征与 LSWA 的局部高频特征之间的平滑过渡与整合。为了形成最终的检测特征金字塔,需要将局部特征与源自 Mamba 主干的全局低频特征进行整合。

针对不同尺度目标对全局结构与局部细节依赖程度的动态变化,本文设计了自适应特征融合模块。首先,将两路特征进行通道维度拼接,并通过全局平均池化聚合全局上下文描述符 s 。随后,将描述符 s 馈入由两层全连接层构成的轻量级激发网络生成动态融合权重向量 $\alpha \in \mathbb{R}^c$:

$$\alpha = \text{Sigmoid}(W_2 \delta(W_1 s)) \# (9)$$

其中, δ 为 ReLU 激活函数, W_1 实现了降维以控制参数量, W_2 用于将维度恢复。最终,利用该权重向量对全局特征 F_g 与局部特征 F_l 进行自适应加权融合,输出 ACF 模块的最终特征 F_o :

$$F_o = \alpha \odot F_g + (1 - \alpha) \odot F_l \# (10)$$

这种动态分配权重的机制,使得网络能够在深层特征空间中自动权衡频域抗噪先验与空域细节响

应的比例,极大提升了模型面对异构尺度 SAR 目标的泛化鲁棒性。

2 实验与结果分析

为了验证本文提出的频空协同状态空间网络 (FSM-Net) 在复杂 SAR 图像目标检测任务中的有效性、鲁棒性及轻量化优势,本节在公开权威的 SAR 目标检测数据集上进行了广泛的定性与定量实验。首先介绍实验所用的数据集与评估指标;其次阐述网络训练的实施细节;随后通过消融实验深入剖析 LFGA、Mamba 主干与 ACF 单元的作用机理;最后,将 FSM-Net 与当前主流的检测算法进行全面的性能对比与可视化分析。

2.1 数据集与评估指标

本文实验选取了两个在 SAR 目标检测领域极具代表性且具有高度挑战性的公开数据集:SSDD (SAR Ship Detection Dataset, SSDD) 与 HRSID (High-Resolution SAR Images Dataset, HRSID)。

1) SSDD 数据集 (Zhang 等, 2021): 包含 1160 张多源 (RadarSat-2、TerraSAR-X 等)、多场景的 SAR 图像,共 2456 个目标。2) HRSID 数据集 (Wei 等, 2020): 包含 5604 张高分辨率图像,共 16951 个目标实例。相比 SSDD,该数据集中的近岸密集并排小目标及防波堤等强杂波干扰比例显著增加,用于验证模型在复杂高分辨场景下的泛化能力。本文按照通用划分标准,将数据集以 8:2 的比例划分为训练集和测试集。

本文采用了如下考察指标,主要精度相关指标包括精确率 (Precision)、召回率 (Recall) 以及均值平均精度 (mean Average Precision, mAP)。同时选取参数量 (Params) 和浮点运算次数 (FLOPs) 作为模型复杂度和计算开销的衡量标准。

2.2 实验环境与实施细节

实验的软硬件环境配置为:操作系统采用 Ubuntu,编程语言为 Python 3.8,基于 CUDA 11.8 并在单张 NVIDIA RTX4090U (24 GB 显存) GPU 上进行训练与测试。网络基于单阶段全卷积架构,输入图像统一缩放或裁剪至 640×640 像素。网络优化采用随机梯度下降 (stochastic gradient descent, SGD) 优化器,初始学习率设定为 0.01,动量因子为 0.937,权重衰减设定为 0.0005。模型共训练 300 个 Epoch,

批处理大小(batch size)为16。训练过程采用余弦退火策略动态调整学习率,并在前3个Epoch引入Warm-up预热策略,以稳定状态空间模型的初始状态演化。

2.3 消融实验与机制验证

为了深度解析FSM-Net中各个创新模块的实质贡献,本文在HRSID数据集上进行了严密的消融实验与模块替换实验。我们以不包含任何本文改进的纯净单阶段网络为基线模型(Baseline),逐步嵌入可学习频域门控激活机制(LFGA)、选择性状态空间主干(DSCMamba)以及自适应跨流融合机制(ACF)。具体的消融实验量化结果如表1所示。

表1 在HRSID数据集上的消融实验
Table 1 Ablation results on the HRSID dataset

LFGA	DSCMamba	ACF	P (%)	R (%)	mAP (%)
-	-	-	88.2	84.9	88.9
√	-	-	90.6	85.2	91.6
-	√	-	91.4	86.2	92.2
√	√	-	91.6	87.0	92.4
-	√	√	92.7	87.4	92.6
√	√	√	93.5	87.8	93.1

注:√表示引入该模块,-表示未引入该模块;加粗字体表示各列最优结果。

如表1所示,引入LFGA后,mAP提升至91.6%,特别是召回率的提升印证了该机制成功提纯了弱小目标的低频先验。进一步将主干替换为DSCMamba后mAP增至92.4%,大长宽比与近岸密集目标的漏检显著改善,但一维序列展平带来的局部细节平滑导致精确率(Precision)增幅受限。最终嵌入ACF模块构建完整的双流架构后,网络成功弥补了局部高频特征的损失,使得模型mAP达到最优的93.1%。该结果直观验证了本文“频域去噪-全局演化-局部保真”协同机制的系统级增益。

为了深入验证第1.4.2节所提出的极简通道重参数投影模块在特征表征与推理加速方面的有效性,本文在HRSID数据集上设计了针对通道投影策略的独立对比实验。实验中网络其他核心组件不变,仅对比不同投影结构对检测精度、参数量、计算量及推理速度的影响,实验结果表格见表2。

表2 不同通道投影策略在HRSID数据集上的性能开销对比

Table 2 Performance and computational cost comparison of different channel projection strategies on HRSID dataset

模块	Params/	FLOPs	Inference	mAP (%)
conv1x1	6.24M	20.86G	5.71ms(175fps)	92.2
conv3x3	8.04M	21.25G	6.24ms(160fps)	92.7
MultiBranch	8.72M	23.93G	7.04ms(142fps)	93.1
重参数投影	8.72M	21.25G	6.62ms(151fps)	93.1

注:加粗字体表示各列最优结果。

结果表明,采用标准1×1卷积计算量虽低(20.86 GFLOPs)但精度受限(mAP 92.2%);保留多分支并行结构虽能提升精度(93.1%),但会导致频繁的内存访问,推理延迟增至7.04 ms。本文采用的重参数投影策略在推理阶段等效折叠为单路3×3卷积,在保持mAP 93.1%高精度的同时,将计算量控制在21.25 GFLOPs,推理速度达151 FPS,有效兼顾了强表征力与部署期低延迟的轻量化需求。

2.4 与主流算法的对比实验

为了验证FSM-Net在SAR目标检测领域的领先地位,本文将其与当前先进目标检测算法进行了全面对比。对比对象涵盖了Faster R-CNN、RetinaNet、YOLOv8系列网络、YOLOv11系列网络。同时,还与基于Transformer架构的各种检测模型进行了对比,包括SwinTransformer、Deformable DETR (Zhu等, 2020)、RT-DETR模型,以及DETR架构下针对SAR图像检测进行改进的DCEA (DETR with concentrated deformable attention)模型(Lin等, 2024)和针对SAR舰船多尺度敏感问题设计的混合感受野与多位置编码M2S-DETR模型(Guo等, 2025)。此外,对比算法中也引入了最新的SAR目标基准测试方法UCR(Zhang等, 2025)以进行全面评估。

对比实验分别在SSDD和HRSID数据集上进行,结果汇总于表3与表4。同时在关注检测精度等指标的同时,我们同样关注模型的各种轻量化指标,并将其结果汇总于表5。

在SSDD数据集中,Faster R-CNN等大型网络虽然具备一定的精度基础,但其极大的参数量和高昂运算开销限制了其实际应用。相比之下,主流的轻量级单阶段网络如YOLO系列在计算效率上优势明

显,但在面对SAR特有的相干斑干扰和高长宽比目标时,精度遭遇了瓶颈。本文提出的FSM-Net在SSDD数据集上取得了最高的98.2%的mAP,相比之下,同样引入了全局建模能力的VMamba虽然取得了96.3%的mAP,但FSM-Net在此基础上依然实现了1.9%的显著提升。在更具挑战性的HRSID数据集中,高分辨率带来了更庞大的离散散射点阵列和更密集的近岸干扰。在此场景下,单纯依赖CNN局部感受野的模型如Faster R-CNN出现了严重的性能衰减。而FSM-Net凭借Mamba主干赋予的全图级别感受野,能够跨越防波堤与停泊舰船之间的背景杂波,建立起准确的目标整体认知,取得93.1%的mAP,较VMamba提升1.5%,保持了绝对的精度领先优势。

表3 不同方法在SSDD数据集上的性能指标对比结果
Table 3 Comparison of performance of different methods on SSDD dataset

方法 (Method)	P (%)	R (%)	mAP (%)	mAP50-95
Faster R-CNN(Li 等,2017)	89.6	88.3	91.2	68.5
RetinaNet-SAR(Yang 等,2021)	91.0	89.2	92.5	67.2
FCOS-SAR(Yang 等,2022)	85.5	83.4	88.1	64.3
Cascade R-CNN(Zhang 等,2020)	88.1	85.2	89.3	66.8
SwinTransformer(Liu 等,2021)	92.5	91.4	92.7	68.1
YOLOv8n(Yan 等,2023)	91.5	89.9	91.8	68.7
YOLOv8m(Yan 等,2023)	95.3	93.6	94.6	70.6
YOLOv11n(Bakirc 等,2024)	92.4	91.1	93.4	69.4
YOLOv11s(Bakirc 等,2024)	94.8	92.9	94.1	70.3
Deformable DETR(Zhu 等,2020)	93.7	94.2	94.4	70.1
RT-DETR(Zhao 等,2024)	95.4	95.1	96.0	71.3
DCEA(Lin 等,2024)	96.3	95.8	96.5	71.8
UCR(Zhang 等,2025)	93.2	92.1	93.1	69.2
M ² S-DETR(Guo 等,2025)	96.6	96.2	97.0	72.1
VMamba(Liu 等,2024)	95.1	95.7	96.3	71.5
FSM-Net (本文)	97.4	96.8	98.2	73.0

注:加粗字体表示各列最优结果。

在模型的轻量化指标上,如表5所示,本文提出的FSM-Net在HRSID数据集上以参数量仅为8.72 M、计算量为21.35 G的情况下,推理速度达到了151 FPS,检测精度mAP达到93.1%。

表4 不同方法在HRSID数据集上的性能指标对比结果
Table 4 Comparison of performance of different methods on HRSID dataset

方法 (Method)	P (%)	R (%)	mAP (%)	mAP50-95
Faster R-CNN (Li 等,2017)	86.4	82.7	87.3	65.1
RetinaNet-SAR (Yang 等,2021)	84.1	79.9	81.2	63.6
FCOS-SAR (Yang 等,2022)	88.6	82.3	89.2	65.7
Cascade R-CNN (Zhang 等,2020)	86.7	76.9	85.6	62.7
SwinTransformer (Liu 等,2021)	89.3	85.2	88.6	65.1
YOLOv8n(Yan 等,2023)	86.7	76.9	85.6	62.7
YOLOv8m(Yan 等,2023)	87.1	76.4	86.3	63.8
YOLOv11n(Bakirc 等,2024)	87.6	83.4	87.3	63.4
YOLOv11s(Bakirc 等,2024)	88.5	85.6	88.1	64.9
Deformable DETR(Zhu 等,2020)	89.2	86.1	89.4	66.2
RT-DETR(Zhao 等,2024)	89.9	86.2	90.8	67.0
DCEA(Lin 等,2024)	93.0	86.4	92.2	67.5
UCR(Zhang 等,2025)	90.1	85.7	90.4	65.9
M ² S-DETR(Guo 等,2025)	92.4	86.6	91.9	67.4
VMamba(Liu 等,2024)	90.9	86.8	91.6	67.3
FSM-Net (本文)	93.5	87.8	93.1	68.4

注:加粗字体表示各列最优结果。

表5 不同方法在HRSID数据集下的轻量化性能指标对比
Table 5 Comparison of lightweight metrics across various methods on the HRSID dataset

算法	Params (M)	FLOPs (G)	FPS	mAP (%)
Faster R-CNN	47.53	107.29	34.3	87.3
RetinaNet-SAR	36.33	52.09	131.3	81.2
YOLOv8n	3.70	5.67	246	85.6
YOLOv11s	9.42	15.93	205	88.1
SwinTransformer	38.85	98.24	96.6	88.6
FSM-Net (本文)	8.72	21.35	151	93.1

注:加粗字体表示最优结果。

而基于Transformer架构的SwinTransformer虽然取得了相较RetinaNet-SAR、YOLO等模型更好的检测结果,但是带来了仅次于Faster R-CNN的参数量和FLOPs以及相对落后的推理速度,相较于之下FSM-Net在计算效率和推理帧率上具有压倒性优势。这意味着FSM-Net用不到Faster R-CNN五分之

一的参数量,实现了全面超越的检测精度。而在面对轻量化模型时,FSM-Net在保持了良好的实时推理速率的同时,检测精度得到了大幅提升,实现了精度、参数量与推理时效性的良好平衡。

2.5 可视化与定性分析

为了更直观地展示 FSM-Net 的检测效果并与其他主流检测模型进行对比,本文首先展示了数据集下几种经典的挑战场景,包括复杂小目标、近岸靠泊以及受强烈海杂波干扰区域的检测结果,可视化对比如图 7 所示。

在受强烈海杂波或近岸陆地边缘干扰的区域,基线模型极易将相干斑噪声或陆地杂乱的相似人造结构误判为舰船,产生较多的误检框。相比之下,FSM-Net 显著降低了复杂背景下的虚警率,展现出

对纯净目标先验的良好提取能力与抗噪鲁棒性。

此外,对于大型狭长舰船,常规 CNN 模型受限于有限的局部感受野,难以跨越海面背景建立目标首尾的结构依赖,经常导致长目标检测不全,或将一艘完整的长船错误预测为多个分离的局部边界框。FSM-Net 引入了 DSCMamba 模块,有效改善了狭长目标的断裂与局部漏检问题。在近岸多船并排停泊的极端密集场景下,舰船间的物理间距极小且伴随强烈的海岸线散射干扰,基线模型往往无法有效区分紧靠的独立船体,易引发目标粘连或严重漏检。FSM-Net 在自适应跨流融合机制的辅助下,通过局部无参校准找回了锐利的高频边缘,对复杂密集小目标依然展现出了卓越的定位能力。

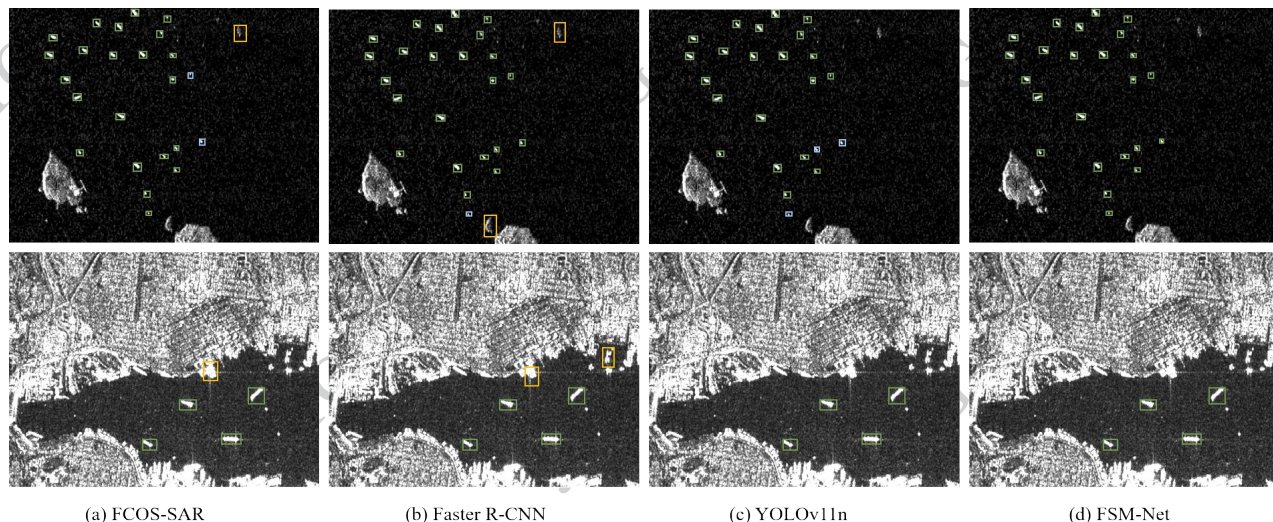


图 7 SSDD 数据集可视化对比

注:绿色框为正确检测结果,蓝色框为漏检结果,橙色框为误检结果。

Fig. 7 Visual Comparison on SSDD Dataset

Note: Green boxes indicate correct detections, blue boxes indicate missed detections, and orange boxes indicate false detections

((a) FCOS-SAR; (b) Faster R-CNN; (c) YOLOv11n; (d)FSM-Net)

如图 8 第一行所示,近岸环境存在严重陆地建筑强散射干扰。基线模型的特征激活图在非目标区域出现了明显的误激活。而 FSM-Net 能够精准聚焦于目标区域,有效抑制了背景杂波,在其热力图中背景呈现出低响应的纯净深蓝色,直观证明了其强大

的抗噪能力。

图 8 第二行展示了更加复杂的近岸堆叠停泊场景。大型舰船在此处呈现出典型的离散散射中心特性,目标首尾跨度大且特征断裂。FCOS-SAR 仅能局部激活目标的强散射点,导致特征严重碎片化,将画面中心的船只与背景混淆从而导致漏检;Faster R-CNN 则发生了严重的特征混叠,大量非目标背景被错误激活,且对周围小目标的特征捕捉极差。相比之下,FSM-Net 成功建立了离散散射中心之间的长程上下文依赖,其高响应区域几乎完整地覆盖了大型舰船的几何轮廓,将孤立的散射点关联为一个完整的语义实体。这一可视化结果与本文的设计初衷高度吻合,充分体现了 FSM-Net 的网络架构价值。

上述定性分析表明,FSM-Net 在抑制背景误检

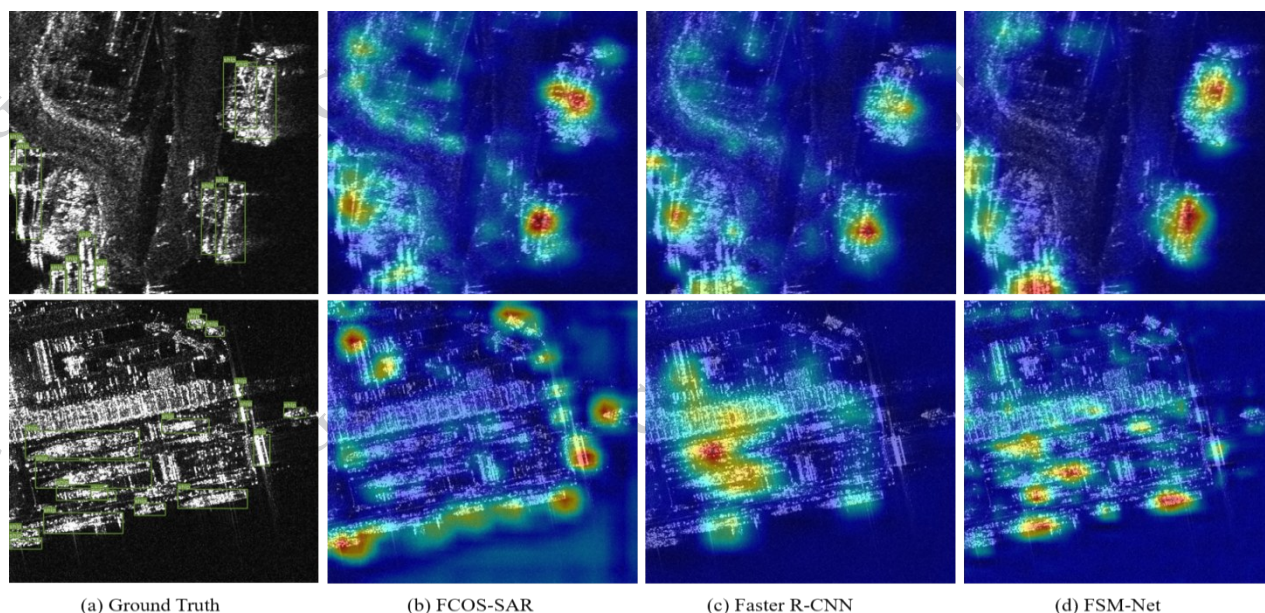


图8展示了利用特征激活热力图(heatmaps)对不同模型在复杂近岸场景下关注区域进行的可视化对比。本文通过该对比,进一步深层剖析并直观验证了FSM-Net在应对强相干斑噪声干扰与离散散射中心长程关联困难两大核心挑战时的工作机制。
(a)Ground truth;(b)FCOS-SAR;(c)Faster R-CNN;(d)FSM-Net)

图8 HRSID数据集特征热力图可视化对比
Fig. 8 Visual comparison of feature heatmaps on the HRSID dataset

和改善目标漏检方面具有显著优势。为了定量地评估模型在复杂SAR背景下控制虚警与漏检的全局综合能力,本文在完整测试集上计算了平均每图虚警数(false positives per image, FPPI)与漏检率(Miss rate)的统计分布,并绘制了FPPI-MR对比曲线,如图9所示。

定量分析表明,传统两阶段模型Faster R-CNN对SAR图像的复杂背景杂波极为敏感,其对数平均漏检率(log-average miss rate, LAMR)高达29.5%。

经典基线模型YOLO11s与FCOS-SAR的表现相近,LAMR分别为13.5%和13.3%。相比之下,本文提

出的FSM-Net展现出了极其优异的鲁棒性,取得了11.8%的最低LAMR。这一全量测试集的定量统计结果与前文的视觉分析高度一致,充分论证了FSM-Net卓越的综合检测性能。

3 结 论

针对复杂背景下SAR图像目标检测面临的相干斑噪声强干扰、目标散射中心高度离散以及长程建模与计算开销难以兼顾的困境,本文跳出了传统空域卷积的局部感知范式,提出了一种具有严密物理机理支撑的频空协同状态空间网络FSM-Net。本研究不仅突破了常规轻量化网络在感受野上的物理瓶颈,更系统性地论证了线性选择性状态空间模型Mamba在雷达视觉解译任务中的巨大潜力。

研究表明,在SAR特征空间中实现高效长程演化的核心,在于初始状态的纯净度控制与演化后期的局部细节闭环。本文创新的可学习频域门控激活机制LFGA,成功利用了SAR图像目标与噪声的频

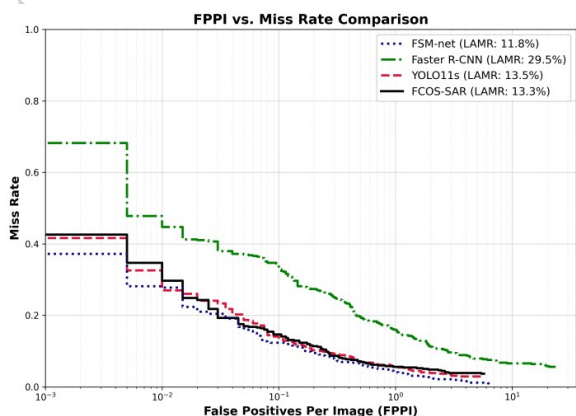


图9 不同检测模型的FPPI-MR曲线对比
Fig. 9 Comparison of FPPI-MR curves among different detection models

谱分布差异,在无损空间拓扑的前提下,为网络提供了免受高频散斑污染的纯净初始状态。在此先验底座上,Mamba主干通过二维多向连续扫描策略,以极低的时间复杂度跨越背景杂波,精准聚合了大型目标离散的散射中心。针对连续扫描不可避免引发的局部高频边界模糊问题,本文巧妙构建了自适应跨流融合机制ACF。该单元将滑窗无参注意力机制、极简重参数通道投影与频空自适应融合无缝嵌合,在几乎零推理算力添加的前提之下,有效地补偿了近岸密集舰船的局部轮廓特征。

综合实验结果充分验证了上述理论推演的科学性。在SSDD与HRSID两大数据集上,FSM-Net以仅约8.72M的极小参数量,在检测精度与抗虚警鲁棒性上全面超越了当前主流的单阶段与双阶段网络模型,实现了精度与效率的平衡。

尽管本文在轻量化SAR特征表达上取得了显著进展,但状态空间模型在处理更复杂的微波散射特性时仍有广阔的探索空间。未来的研究工作将聚焦于以下两个方向:一是探索SAR与光学数据的跨模态融合机制,利用本文提出的自适应跨流融合思想,实现异构传感器信息的动态互补;二是后续工作将致力于针对状态空间模型的模型量化与算子剪枝技术,并结合专用硬件加速器如FPGA或边缘端NPU进行深度调优,继续优化其轻量化特性,推动其在星载、机载边缘计算平台上的在轨实时部署。

参考文献(References)

- Bakirci M and Bayraktar I. 2024. Assessment of YOLO11 for ship detection in SAR imagery under open ocean and coastal challenges//Proceedings of the 2024 21st International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE): 1-6 [DOI: 10.1109/CCE62852.2024.10770926]
- Bao W, Hu J, Huang M, Xu Y, Ji N and Xiang X. 2022. Detecting fine-grained airplanes in SAR images with sparse attention-guided pyramid and class-balanced data augmentation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15: 8586-8599 [DOI: 10.1109/JSTARS.2022.3208928]
- Cai J, Zhang L, Dong J, Guo J, Wang Y and Liao M. 2023. Automatic identification of active landslides over wide areas from time-series InSAR measurements using Faster R-CNN. International Journal of Applied Earth Observation and Geoinformation, 124: 103516 [DOI: 10.1016/j.jag.2023.103516]
- Chai B, Nie X, Zhou Q and Zhou X. 2024. Enhanced cascade R-CNN for multiscale object detection in dense scenes from SAR images. IEEE Sensors Journal, 24 (12): 20143-20153 [DOI: 10.1109/JSEN.2024.3393750]
- Chen Y, Cong Y and Zhang L. 2023. Deformable scattering feature correlation network for aircraft detection in SAR images. IEEE Geoscience and Remote Sensing Letters, 20: 1-5 [DOI: 10.1109/LGRS.2023.3292243]
- de Medeiros D D S, García F D A, Alves D I, da Costa R F, Machado R and Santos Filho J C S. 2024. CA-CFAR detection for SAR systems over correlated gamma-distributed clutter. IEEE Geoscience and Remote Sensing Letters, 21: 1-5 [DOI: 10.1109/LGRS.2024.3450293]
- Deng J, Zhu Y, Zhang S and Chen S. 2023. SAR image recognition using ViT network and contrastive learning framework with unlabeled samples. IEEE Geoscience and Remote Sensing Letters, 21: 1-5 [DOI: 10.1109/LGRS.2023.3332364]
- Gu A and Dao T. 2023. Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint, arXiv: 2312.00752 [DOI: 10.48550/arXiv.2312.00752]
- Guo P, Li H C, Liu N, Lei S, Geng X and Pan J. 2025. M2S-DETR: A mixed receptive field and multiposition encoding for scale-sensitive SAR ship detection with transformer. IEEE Sensors Journal, 25 (7): 12184-12197 [DOI: 10.1109/JSEN.2025.3540901]
- Guo Y, Chen S, Zhan R, Wang W and Zhang J. 2022. LMSD-YOLO: A lightweight YOLO algorithm for multi-scale SAR ship detection. Remote Sensing, 14(19): 4801 [DOI: 10.3390/rs14194801]
- Hatamizadeh A and Kautz J. 2025. MambaVision: A hybrid Mamba-Transformer vision backbone//Proceedings of the Computer Vision and Pattern Recognition Conference: 25261-25270 [DOI: 10.1109/cvpr52734.2025.02352]
- Hou J, Bian Z, Yao G, Lin H, Zhang Y, He S, et al. 2024. Attribute scattering center-assisted SAR ATR based on GNN-FiLM. IEEE Geoscience and Remote Sensing Letters, 21: 1-5 [DOI: 10.1109/LGRS.2024.3391323]
- Hu J, Chen C, Xie J, Li B, Hong A, Yi X, et al. 2026. AWANet: Improved YOLOv11's adaptive weighted attention network for haystack detection in airborne SAR imagery. IEEE Access [DOI: 10.1109/ACCESS.2026.3672644]
- Li J, Liu Z, Liu S and Wang H. 2025. MBSSNet: A Mamba-based joint semantic segmentation network for optical and SAR images. IEEE Geoscience and Remote Sensing Letters, 22: 1-5 [DOI: 10.1109/LGRS.2025.3541895]
- Li J, Qu C and Shao J. 2017. Ship detection in SAR images based on an improved Faster R-CNN//Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSAR DATA): 1-6
- Li N, Pan X, Yang L, Huang Z, Wu Z and Zheng G. 2022. Adaptive CFAR method for SAR ship detection using intensity and texture feature fusion attention contrast mechanism. Sensors, 22 (21): 8116 [DOI: 10.3390/s22218116]

- Li Y, Wang Z, Chen H and Li Y. 2024. A density clustering-based CFAR algorithm for ship detection in SAR images. *IEEE Geoscience and Remote Sensing Letters*, 21: 1-5 [DOI: 10.1109/LGRS.2024.3397883]
- Lin H, Liu J, Li X, Wei L, Liu Y, Han B, et al. 2024. DCEA: DETR with concentrated deformable attention for end-to-end ship detection in SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17: 17292-17307 [DOI: 10.1109/JSTARS.2024.3461723]
- Liu Y, Tian Y, Zhao Y, Yu H, Xie L, Wang Y, et al. 2024. VMamba: Visual state space model//*Proceedings of the Advances in Neural Information Processing Systems*: 103031-103063 [DOI: 10.52202/079017-3273]
- Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. 2021. Swin Transformer: Hierarchical vision transformer using shifted windows//*Proceedings of the IEEE/CVF International Conference on Computer Vision*: 10012-10022
- Miao T, Zeng H, Yang W, Chu B, Zou F, Ren W, et al. 2022. An improved lightweight RetinaNet for ship detection in SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 4667-4679 [DOI: 10.1109/JSTARS.2022.3180159]
- Pan Y, Zhong L, Chen J, Li H, Zhang X and Pan B. 2024. SAR image despeckling based on denoising diffusion probabilistic model and Swin Transformer. *Remote Sensing*, 16(17): 3222 [DOI: 10.3390/rs16173222]
- Pires de Lima R, Vahedi B, Hughes N, Barrett A P, Meier W and Karimzadeh M. 2023. Enhancing sea ice segmentation in Sentinel-1 images with atrous convolutions. *International Journal of Remote Sensing*, 44 (17) : 5344-5374 [DOI: 10.1080/01431161.2023.2248560]
- Rani R, Singh R and Patra T K. 2024. Land ocean mask generation using HAAR transform on VH intensity band of SAR Sentinel-1 C-Band polarimetry image//*Proceedings of the 2024 Intelligent Systems and Machine Learning Conference (ISML)*: 163-169 [DOI: 10.1109/ISML60050.2024.11007457]
- Shen J, Cheng X, Yang X, Zhang L, Cheng W and Lin Y. 2025. Efficient CNN accelerator based on low-end FPGA with optimized depthwise separable convolutions and squeeze-and-excite modules. *AI*, 6(10): 244 [DOI: 10.3390/ai6100244]
- Tian Z, Wang W, Wu F, Zhou K, Liu S and Zhang H. 2025. SAR target recognition based on CNN with 2-D dual-tree complex wavelet transform decomposition. *Pattern Recognition*: 112585 [DOI: 10.1016/j.patcog.2025.112585]
- Tu H, Wang W, Guo Y and Chen S. 2025. Mamba-UDA: Mamba unsupervised domain adaptation for SAR ship detection. *IEEE Geoscience and Remote Sensing Letters* [DOI: 10.1109/LGRS.2025.3595843]
- Wang Q, Sheng J, Tong C, Wang Z, Song T, Wang M, et al. 2023. A fast facet-based SAR imaging model and target detection based on YOLOv5 with CBAM and another detection head. *Electronics*, 12 (19): 4039 [DOI: 10.3390/electronics12194039]
- Wang Z Q, Li Y, Zhang R, Wang J B, Li Y C and Chen Y. 2024. Few-shot SAR image classification: A survey. *Journal of Image and Graphics*, 29(7): 1902-1920 (王梓祺, 李阳, 张睿, 王家宝, 李允臣, 陈瑶. 2024. 小样本 SAR 图像分类方法综述. *中国图象图形学报*, 29(7): 1902-1920) [DOI: 10.11834/jig.230359]
- 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 [DOI: 10.1109/ACCESS.2020.3005861]
- Xiao Z J, Lin B H and Qu H C. 2024. SAR ship detection with multi-mechanism fusion. *Journal of Image and Graphics*, 29(2): 545-558 (肖振久, 林博涵, 曲海成. 2024. 基于多机制融合的 SAR 图像舰船目标检测. *中国图象图形学报*, 29(2): 545-558) [DOI: 10.11834/jig.230166]
- Yan S, Fu Y, Zhang W, Yang W, Yu R and Zhang F. 2023. Multi-target instance segmentation and tracking using YOLOv8 and BoT-SORT for video SAR//*Proceedings of the 2023 5th International Conference on Electronic Engineering and Informatics (EEI)*: 506-510 [DOI: 10.1109/EEI59236.2023.10212903]
- Yang L, Zhang R Y, Li L and Xie X. 2021. SimAM: A simple, parameter-free attention module for convolutional neural networks//*Proceedings of the International Conference on Machine Learning*: 11863-11874
- Yang R, Pan Z, Jia X, Zhang L and Deng Y. 2021. A novel CNN-based detector for ship detection based on rotatable bounding box in SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 1938-1958 [DOI: 10.1109/JSTARS.2021.3049851]
- Yang S, An W, Li S, Wei G and Zou B. 2022. An improved FCOS method for ship detection in SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 8910-8927 [DOI: 10.1109/JSTARS.2022.3213583]
- Zhang J, Xie Z, Sun J, Zou X and Wang J. 2020. A cascaded R-GNN with multiscale attention and imbalanced samples for traffic sign detection. *IEEE Access*, 8: 29742-29754 [DOI: 10.1109/ACCESS.2020.2972338]
- Zhang T, Zhang X, Li J, Xu X, Wang B, Zhan X, et al. 2021. SAR ship detection dataset (SSDD): Official release and comprehensive data analysis. *Remote Sensing*, 13(18): 3690 [DOI: 10.3390/rs13183690]
- Zhang X, Yang X, Li Y, Yang J, Cheng M M and Li X. 2025. RSAR: Restricted state angle resolver and rotated SAR benchmark. *arXiv preprint, arXiv:2501.04440* [DOI: 10.48550/arXiv.2501.04440]
- Zhang Z T, Zhang X and Shao Z. 2023. Deform-FPN: A novel FPN with deformable convolution for multi-scale SAR ship detection//*Proceedings of the IGARSS 2023-2023 IEEE International Geoscience*

- and Remote Sensing Symposium: 5273-5276 [DOI: 10.1109/IGARSS52108.2023.10282798]
- Zhao Y, Lv W, Xu S, Wei J, Wang G, Dang Q, et al. 2024. DETRs beat YOLO on real-time object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition: 16965-16974 [DOI: 10.1109/CVPR52733.2024.01605]
- Zhou F, He F, Gui C, Dong Z and Xing M. 2022. SAR target detection based on improved SSD with saliency map and residual network. Remote Sensing, 14(1): 180 [DOI: 10.3390/rs14010180]
- Zhu J, Wang F and You H. 2022. SAR image segmentation by efficient fuzzy C-means framework with adaptive generalized likelihood ratio nonlocal spatial information embedded. Remote Sensing, 14(7): 1621 [DOI: 10.3390/rs14071621]
- Zhu X, Su W, Lu L, Li B, Wang X and Dai J. 2020. Deformable

DETR: Deformable transformers for end-to-end object detection. arXiv preprint, arXiv: 2010.04159 [DOI: 10.48550/arXiv.2010.04159]

作者简介

王涵彬,男,硕士研究生,主要研究方向为图像处理与计算机视觉。E-mail:1940400366@qq.com

胡海洋,男,博士,教授、博士生导师,主要研究方向:计算机视觉、工作流技术。E-mail:huhaiyang@hdu.edu.cn

马哲杰,男,博士研究生,主要研究方向为计算机视觉,遥感小目标检测。E-mail:mzjcs@hdu.edu.cn

周桥平,男,硕士研究生,主要研究方向深度学习,小目标检测。E-mail:2401926257@qq.com