DOI:10.19651/j.cnki.emt.2519080

融合多尺度特征的 SAR 图像目标检测方法

赵 喆 李 勃 徐文校 李 尧
(南京航空航天大学电子信息工程学院 南京 210016)

摘 要:针对合成孔径雷达图像中相干斑噪声干扰、低信噪比及目标多尺度散射特性导致的目标检测精度衰减与小目标漏检问题,提出一种兼顾特征表征能力与实时性的轻量化检测模型 XMNet。XMNet 在主干网络部分引入改进型单头视觉 Transformer,通过全局注意力机制强化上下文语义关联;设计跨层级多路径聚合网络作为颈部结构,融合动态上采样与并行多尺度卷积模块,优化多尺度特征表征;新增高分辨率检测层,利用浅层高分辨率特征增强小目标细节捕捉能力。在 MSAR-1.0 数据集上的实验表明:全类别平均检测精度达 90.4%,较基准模型提升 8.7%;飞机类小目标检测精度显著提高 20.1%,参数量仅增加 2 M,推理检测速度达到 185 FPS;与 FCOS、CenterNet 等 9 个先进方法对比,XMNet 在检测精度与计算效率综合指标上排名首位。XMNet 通过跨层级注意力机制与多尺度特征融合的协同设计,有效解决了 SAR 图像中多尺度目标特征丢失与实时性难以兼顾的难题。其轻量化特性与高检测精度为各类 SAR 平台的实时遥感监测提供了可行的工程化解决方案,尤其在小目标密集的复杂场景中展现出显著优势。

关键词:合成孔径雷达;密集小目标检测;单头注意力机制;跨层多尺度特征融合;动态上采样
中图分类号:TN957.52;TN911.73 **文献标识码**:A **国家标准学科分类代码**:510.4050

Multiscale feature fusion for object detection in SAR images

Zhao Zhe Li Bo Xu Wenxiao Li Yao
(College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: To address the issues of target detection accuracy degradation and small target miss detection caused by speckle noise interference, low signal-to-noise ratio, and multi-scale scattering characteristics of targets in Synthetic Aperture Radar images, this paper proposes a lightweight detection model named XMNet, which balances feature representation capability and real-time performance. XMNet incorporates an improved single-Head vision Transformer into the backbone network to strengthen contextual semantic correlations through global attention mechanisms. A cross-layer multi-path aggregation network is designed as the neck structure, integrating dynamic upsampling and a parallel multi-scale convolution module to optimize multi-scale feature representation. An additional high-resolution detection layer is introduced to leverage shallow high-resolution features, enhancing detail capture capability for small targets. Experiments on the MSAR-1.0 dataset demonstrate that XMNet achieves a mean average precision of 90.4% across all categories, representing an increase of 8.7% over the baseline model. Detection accuracy for small aircraft targets significantly improves by 20.1%, with only a 2-million parameter increase while achieving an inference speed of 185 FPS. When compared against nine advanced methods including FCOS and CenterNet, XMNet ranks first in comprehensive metrics balancing detection accuracy and computational efficiency. Through the design of cross-layer attention mechanisms and multi-scale feature fusion, XMNet effectively resolves the challenge of balancing feature preservation for multi-scale targets and real-time processing in SAR imagery. Its lightweight and high detection accuracy provide a viable engineering-ready solution for real-time remote sensing monitoring across various SAR platforms, demonstrating significant advantages particularly in complex scenes with dense small targets.

Keywords: synthetic aperture radar; dense small object detection; single-head attention; cross-layer multi-scale feature fusion; dynamic upsampling

0 引 言

合成孔径雷达(synthetic aperture radar, SAR)是一种主动式对地观测传感器,具有全天时、全天候的监测能力,其微波穿透特性可突破云层与地表物体的限制,以稳定、连续的方式获取地球表面的观测数据。SAR 的应用可大致

收稿日期:2025-06-11

分为 3 大类:制图与土地分类、参数反演和目标检测[1]。其中,目标检测问题,尤其是针对舰船、飞机等小目标的检测,是一个典型的研究方向,可为国防安全、海洋监视以及机场管理提供重要信息,在军民领域具有重要的应用价值。

传统的 SAR 图像目标检测方法主要依赖于传统图像处理技术实现特征提取,主要方法有基于恒虚警率(constant false alarm rate,CFAR)的方法[2]、基于模板匹配的方法[3]等。然而,随着 SAR 图像数据集的增大以及分辨率的提升,对于目标检测方法的精确性和实时性有着更高的要求。

近年来,以深度学习为基础的目标检测算法被提出,以 YOLO[4-6](you only look once)、SSD[7](single shot-multibox detector)和 Faster R-CNN[8]为代表的检测框架有着高准确率、高鲁棒性的性能表现,成功应用于 SAR 目标检测领域。Sun 等[9]提出了一种双向特征融合检测模型 BiFA-YOLO,通过利用自顶向下和自底向上的信息交互,增强舰船检测能力,但缺点是对复杂背景虚警抑制不足。Su 等[10]提出的 SII-Net 通过引入通道位置注意力模块和高级特征增强模块进行舰船检测,沿不同的空间方向提取特征,增强了主干网络的检测能力,缓解目标位置信息的丢失,但对小目标召回率提升有限。Zhao 等[11]将 Swin-Transformer[12]和坐标注意力融合作为主干网络,提出了 ST-YOLOA,以应对 SAR 图像中局部干扰和语义信息丢失问题,提升了模型的检测能力。李波等[13]以 SSD 为基础模型,通过结合注意力机制与特征融合,实现了高效精确的舰船目标检测,却显著增加参数量,制约边缘部署。Liu 等[14]提出的 MDD-YOLOv8 利用动态卷积与可变形大核注意力解决 SAR 图像中目标尺寸微小、背景干扰导致的漏检问题,但参数量增加 20%且计算开销上升 53%,制约了边缘部署。胡欣等[15]以 YOLOv5 为基础,提出了注意力模块 MBAM,在不同的 3 个维度对提取特征进行信息融合,但在极端复杂背景下,对小尺寸目标的漏检率仍较高。

尽管针对 SAR 图像的目标检测技术已获得较大的提升,但是仍存在技术瓶颈需要突破,如小目标检测召回率低,多类别目标分辨能力差,实时处理效率低下。特别是当目标呈现密集小尺寸分布时,多数模型为了提高检测精度而增加模型深度,导致模型参数达到千万级别,严重制约了模型在边缘设备上的部署应用。

因此,为解决上述难题,基于 YOLOv8 架构提出一种实现了跨层多尺度特征融合的多类别 SAR 图像目标检测方法 XMNet。该方法在主干网络中引入改进型单头视觉 Transformer[16](single head vision Transformer, SHViT),通过全局注意力机制实现上下文的语义联系;将颈部网络替换为跨层级多路径聚合网络 XMPAN,并在特征融合阶段引入动态上采样(DySample)[17]和并行多尺度卷积(parallel multi-scale convolution, PMC)模块,以提高模型对于多尺度目标的检测能力;最后,针对 SAR 图像中小目标数量占比高的特点,引入小尺度高分辨率检测层,进一步提升模型对小目标的识别精确率。在大规模多类 SAR 目标检测数据集 MSAR-1.0 上与其他先进检测算法进行横向对比实验,验证了模型的检测性能;进行了消融实验以定量分析各优化的贡献度。此外还在 SADD(SAR aircraft detection dataset)[18]和 SAR-Ship-Dataset[19]两个数据集上验证了模型的泛化能力。

1　YOLOv8

YOLO 系列算法是单阶段目标检测算法的典型代表算法,其基于卷积神经网络构建的端到端检测框架,实现仅通过单次前向传播预测出所有的目标的边界框和类别概率,在实时检测领域保持优势,自 Redmon 等[4]首次提出架构以来一直被广泛应用。YOLOv8 是 Ultralytics 团队在 YOLOv5 架构基础上进行深度改进,其结构图如图 1 所示,网络结构主要包括主干网络(backbone)、颈部网络(neck)和检测头部(head)3 个部分。

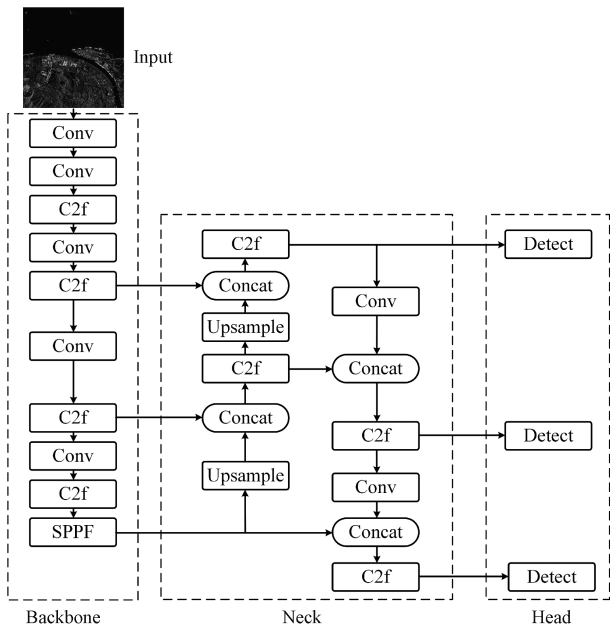


图 1　YOLOv8 结构图

Fig. 1　The architecture of YOLOv8

具体而言,主干网络部分使用 CSPDarkNet 对网络输入的图片进行特征提取,输出不同尺度的具有丰富语义信息的特征图;颈部网络部分的设计参考了 PANet[20],实现自顶向下与自底向上的双向信息交互;检测头部采用任务解耦机制,将目标定位与分类置信度预测进行独立分支处理,有效缓解了两项任务的相互干扰问题。

2　跨层多尺度特征检测方法 XMNet

针对 SAR 图像特有的复杂散射背景、目标几何形变多样性以及相干斑噪声干扰等,本文以 YOLOv8 框架为基

础,在特征提取及多尺度特征融合阶段进行改进,提出了一种专为 SAR 目标检测设计的深度神经网络架构 XMNet,其总体结构如图 2 所示。首先,在模型的特征提取阶段采用了改进型单头视觉 Transformer 作为主干网络,以增强模型的检测能力,SHViT 以内存高效的方式解决了设计层面的计算冗余问题,并且引入了全局注意力机制,实现了推理速度与准确性之间的最佳平衡,提高了模型在资源受限设备上的运行效率。其次,设计跨层级多路径聚合网络 XMPAN 作为特征融合架构,通过将不同层级的特征图聚合,实现浅层纹理信息与深层语义特征的动态交互,显著提升模型对多尺度目标的适应性。最后,针对 SAR 图像中小目标占比高、纹理特征微弱的特点,在检测头部分引入 4×4 超小尺度检测层,该层通过融合主干网络输出的高分辨率特征图实现细颗粒度的特征捕捉,极大提升飞机等微小目标的识别精确率。本章将深入解析各子模块的设计原理及其在 SAR 图像理解中的重要作用。

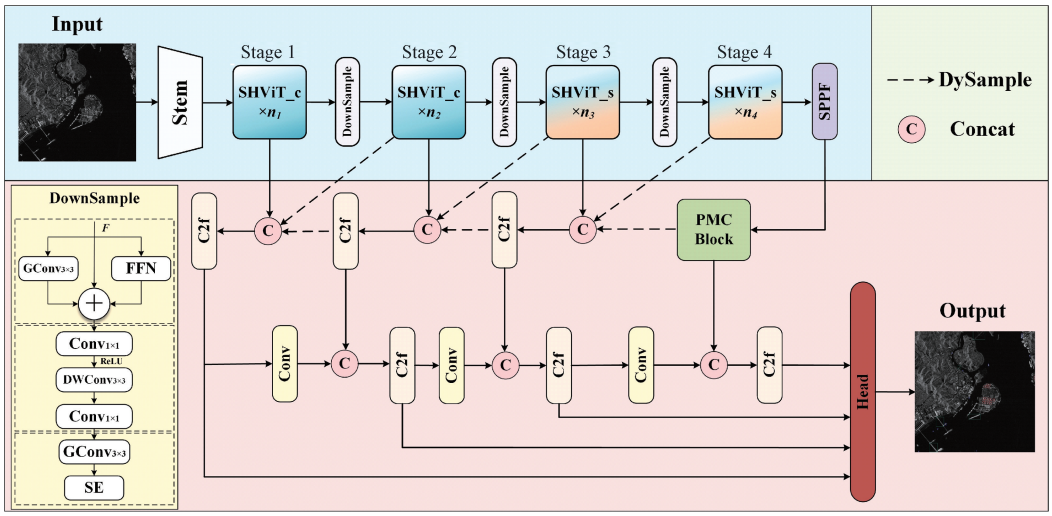


图 2　XMNet 总体结构

Fig. 2　The overall architecture of XMNet

2.1　改进型 SHViT

与传统卷积方法不同,Transformer 模型基于自注意力机制(self attention, SA),通过在输入的序列内建立全局依赖关系,更充分地利用上下文信息。该架构最初在机器翻译等自然语言处理领域取得突破性进展,后被 Vision Transformer(ViT)[21]成功迁移至视觉领域,通过将图片分割为序列化图像块(patch)进行处理,开创了以无卷积架构进行图像分类的先河。然而,ViT 存在两大局限性:首先,ViT 中缺乏 CNN 网络中的归纳偏置(inductive bias),需要依赖大量的训练数据进行参数学习;其次,全局注意力机制的计算复杂度与输入图像尺寸的平方呈正相关,当数据集为高分辨率图像时需要耗费大量的计算资源,严重限制了其部署能力。

近期的研究系统分析了多头注意力(multi-head self attention, MHSA)的计算特性,发现部分通道的计算存在冗余。此外,在模型的早期阶段,将注意力机制替换为卷积层能够有效降低计算复杂度,同时保持模型性能不受显著影响。基于此,Yun 等[16]进一步提出了 SHViT,其优势在于:

1) 采用单头注意力机制精简 MHSA 结构,相比 MobileViT v2 快了 3.3 倍;

2) 采用不同层次的特征处理策略。其浅层模块由深度可分离卷积、批归一化和前馈网络构成,侧重局部特征提取;深层模块保留自注意力机制,专注全局关系建模。

尽管 SHViT 凭借其低计算复杂性在光学图像分类、场景分割等任务中展现其部署优势,但其在 SAR 图像多尺度目标检测任务中的性能仍有局限性。具体而言,SAR 图像中舰船、飞机等小目标因成像机理导致纹理特征稀疏、相干斑噪声明显,而 SHViT 对于小尺度目标特征提取能力有限,经连续下采样后小目标位置偏移误差较大。

为此,选择在 SHViT 基础上进行优化,提出改进型 SHViT(improved SHViT),其结构图如图 3 所示。

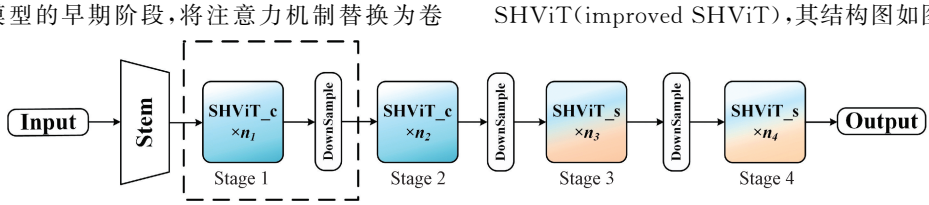


图 3　改进 SHViT 结构图(虚线框内为改进新增阶段,n_1-n_4 为各阶段深度)

Fig. 3　Improved SHViT structure diagram (the dashed box indicates the newly added stages, and n_1-n_4 indicates the depth of each stage)

首先，针对池化层与跨步卷积等传统下采样操作引起的信息丢失问题，设计了三分支高效下采样模块 DownSample，结构如图 2 左下角所示。在预处理分支中，通过分组卷积与前馈网络 FFN 构建的残差连接，强化了特征表达能力，公式表达为：

$$\boldsymbol{F}^{pre} = \boldsymbol{F} + GConv(\boldsymbol{F}) + FFN(\boldsymbol{F}) \tag{1}$$

式中：$GConv(\ )$ 为分组卷积，$FFN(\boldsymbol{x}) = GeLU(\boldsymbol{W}_1 x + \boldsymbol{b}_1)\boldsymbol{W}_2 + \boldsymbol{b}_2$，$\boldsymbol{W}_1$，$\boldsymbol{W}_2$，$\boldsymbol{b}_1$ 与 $\boldsymbol{b}_2$ 为全连接层的参数。

在压缩路径分支采用三级卷积层联级架构，通过深度可分离卷积（depthwise separable convolution），降低了计算复杂度的同时仍能保持特征图的完整性。分支计算表达为：

$$\boldsymbol{F}^{down} = Conv(DW(ReLU(Conv(\boldsymbol{F}^{pre})))) \tag{2}$$

式中：$DW(\ )$ 为深度可分离卷积。

在后处理分支中，进一步引入了 SE 注意力机制[22]实现通道自适应校准：

$$\boldsymbol{F}^{post} = SE(GConv(\boldsymbol{F}^{down})) \tag{3}$$

在 SE 注意力机制中，设定自适应比例参数 $\rho = 1/\log_2 \mathrm{dim}$，其中 dim 表示该层特征图的维度，使网络能够动态调整压缩比例。

其次，针对 SAR 图像目标的多尺度特性，将原三阶段结构扩展为四阶段层次结构（Stage1-Stage4），在 Stage1-Stage2 过程中，采用 SHViT_c 模块，在 Stage3-Stage4 过程中，采用 SHViT_s 模块。SHViT_c 模块采用纯卷积的结构增强局部特征，通过深度可分离卷积与残差连接构建轻量化特征提取，确保浅层细节高效传递；SHViT_s 模块融合卷积与单头注意力机制，在深层语义信息建立全局依赖关系。图 4 为 SHViT_c 模块与 SHViT_s 模块结构示意图。

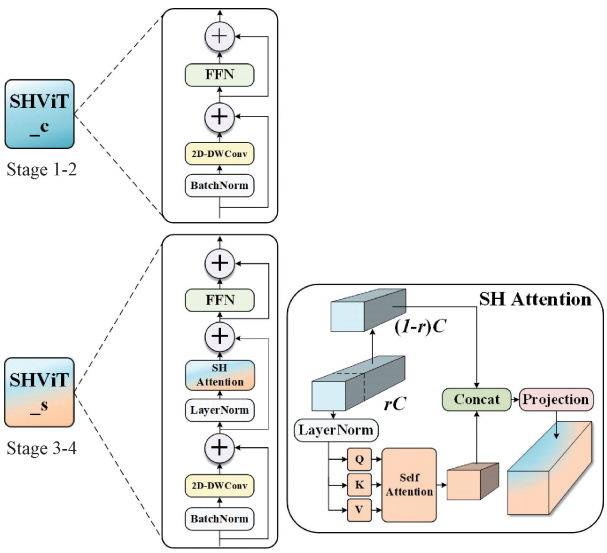


图 4　SHViT Block 的结构图

Fig. 4　The architecture of SHViT Block

SHViT Block 的构建方式与 Transformer Block 类似，但在模块设计与归一化策略上有所改动。该模块由 3 个阶段构成：在结构前期使用二维 DW Conv 进行空间特征提取，中期阶段配置单头注意力（SHA）机制完成全局信息交互，最后阶段通过 FFN 实现非线性变换。归一化策略上，在 SHA 层前应用层归一化（layer norm，LN）进行特征分布校准，而在 DW Conv 操作前则采用批归一化（batch norm，BN）提升训练稳定性。特别地，SHA 层计算流程可表示为：

$$\begin{cases} \boldsymbol{F}_{att}, \boldsymbol{F}_{res} = S_r(\boldsymbol{F}) \\ \tilde{\boldsymbol{F}}_{att} = Attention(\boldsymbol{F}_{att}\boldsymbol{W}^{Q}, \boldsymbol{F}_{att}\boldsymbol{W}^{K}, \boldsymbol{F}_{att}\boldsymbol{W}^{V}) \\ SHA(\boldsymbol{F}) = Concat(\tilde{\boldsymbol{F}}_{att}, \boldsymbol{F}_{res})\boldsymbol{W}^{o} \end{cases} \tag{4}$$

式中：$\boldsymbol{F}_{att}$ 和 $\boldsymbol{F}_{res}$ 分别表示将输入特征 $\boldsymbol{F}$ 按比例 r 和 $(1-r)$ 进行通道分割后的两部分；$S_r(\ )$ 表示通道分割操作，r 为分割通道比例；$\tilde{\boldsymbol{F}}_{att}$ 表示经注意力机制处理后的特征；$Attention(\ )$ 是注意力计算操作，计算公式如式（5）所示；$\boldsymbol{W}^{Q}$，$\boldsymbol{W}^{K}$，$\boldsymbol{W}^{V}$ 分别为查询（query）、键（key）和值（value）的权重矩阵，将 F_{att} 线性变换得到查询向量、键向量和值向量；$SHA(\boldsymbol{F})$ 表示输出特征，$Concat(\ )$ 表示拼接操作，$\boldsymbol{W}^{o}$ 是用于融合拼接后特征的权重矩阵。

$$Attention(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = Softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_{qk}}}\right)\boldsymbol{V} \tag{5}$$

式中：$\boldsymbol{Q}$，$\boldsymbol{K}$，$\boldsymbol{V}$ 分别表示查询、键和值向量，$\sqrt{d_{qk}}$ 为缩放因子，d_{qk} 表示查询、键向量的维度，$Softmax(\ )$ 为归一化激活函数。

2.2　*XMPAN*

在 SAR 图像中包含着多尺度目标，为了将浅层特征图中的位置回归信息与深层特征图中的语义信息相结合，现有的检测方法通常采用特征金字塔的策略融合多尺度的特征信息，如 FPN（feature pyramid network）[23]，PAN（path aggregation network）以及它们的变体结构。尽管这些结构的应用提升了多尺度目标的检测精度，但是这类结构仍然存在着一定的局限性。浅层特征与深层特征的融合需要经过多个中间层，在传播路径上要经历多次卷积以及低效的上、下采样的处理，在这个过程中，特征信息容易丢失。因此，为了解决这些问题，提出一种新的特征融合网络 XMPAN 作为颈部网络。该网络构建了跨层特征融合路径，通过引入动态上采样技术，能够根据输入特征图的内容自适应调整上采样策略，从而减少信息丢失，并使得不同尺度的特征能够进行交互和互补，有效弥补了传统路径信息流通受限的问题。如图 5 所示，为 XMPAN 的简要示意图。

1）跨层多尺度特征融合

针对 SAR 图像中特有的相干斑噪声强烈、目标尺度分布广泛以及地物背景干扰显著等瓶颈，本文采用的层级化颈部网络设计（图 5）具有以下优势：

（1）多尺度特征机制以应对目标尺度变异现象，颈部网络底层接收骨干网络提取的 4 个尺度的特征图（P_2 ～

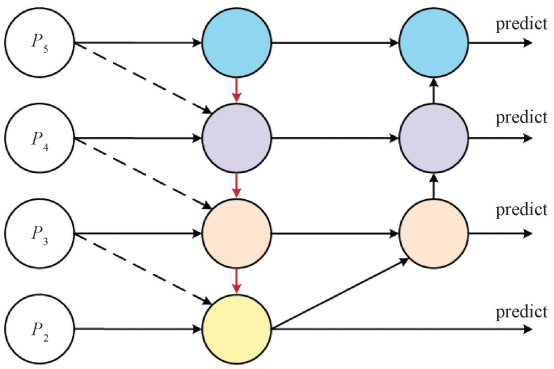


图 5　XMPAN 的简要示意图

Fig. 5　Schematic diagram of XMPAN

P_5)覆盖高分辨率细节与低分辨率语义信息。通过融合节点实现高层语义信息向低层传导,增强小目标特征响应,以及低层纹理细节向高层补充(如图 5 虚线箭头),提升大目标边缘定位精度。该设计有效缓解了 SAR 图像中因成像角度不同导致的尺度变化问题,极大改善了小目标的漏检率。

(2) 跨层特征融合机制以抑制相干斑噪声,由图 5 虚线箭头表示的非对称跨层连接构建了噪声抑制路径,将特征图跳跃跨层拼接,不仅可以保留原始散射点空间分布信息,避免多次下采样导致的弱目标淹没,而且引入中层语义约束,区分噪声与真实目标。

(3)多层次特征抽象增强目标表征能力,这些不同层级抽象出来的特征具有递进性的语义,如浅层提取局部散射单元特征,深层整合全局结构特征。这种分层机制可以适应 SAR 图像的方位角敏感性,利用多层特征互补可确保不同观测角度下的稳定检测。

具体而言,在 XMPAN 中,不同层级的特征可以表示为:

$$\boldsymbol{F}_i^{(l)} = Concat\{\boldsymbol{P}_i, Conv^{(l)}(\boldsymbol{F}_{i+1}^{(l)}), u^{(l)}(\boldsymbol{P}_{i+1})\} \tag{6}$$

式中:l 表示网络层级,$\boldsymbol{P}_i$ 表示骨干网络第 i 级输入特征,$Conv^{(l)}(\ )$ 表示第 l 层可学习卷积,$u^{(l)}(\ )$ 表示第 l 层动态上采样算子。

2) 动态上采样

在目标检测模型中,卷积和池化操作通过逐层缩减特征图的空间分辨率实现深层特征提取,而上采样则是恢复高分辨率表征的关键步骤。传统的上采样方法如最近邻插值法(nearest neighbor interpolation)和双线性插值法(bilinear interpolation),采用固定几何规则生成采样点,缺乏对图像内容的自适应能力。为此,在 XMPAN 中,引入了轻量级的动态上采样 DySample。其核心思想如图 6 所示。

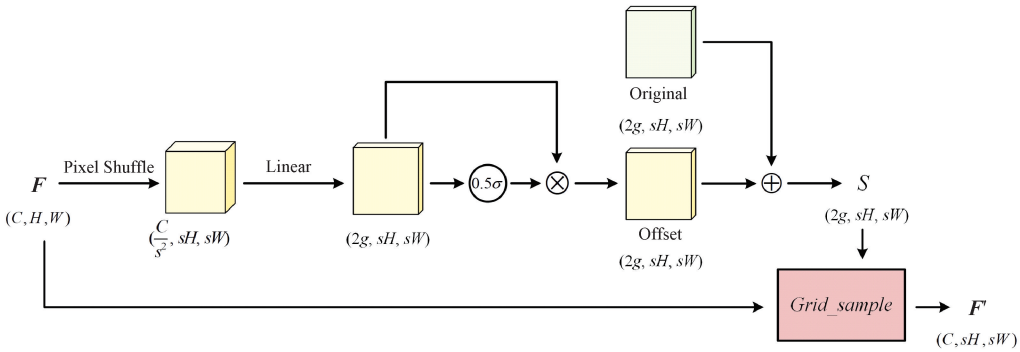


图 6　动态上采样过程图

Fig. 6　Schematic diagram of DySample process

首先,原始特征图 $\boldsymbol{F} \in \boldsymbol{R}^{C\times H_{in}\times W_{in}}$ 经过 Pixel Shuffle 进行像素重排,进行通道-空间维度置换;随后经过线性层进行通道数变换,将通道数变换为 $2g$,其中 2 表示偏移量的两个维度坐标 (x, y),g 为分组数;再后经过一个类残差结构,通过 $sigmoid$ 进行激活后乘以范围参数 0.5 调整目标尺度,引入偏移动态范围得到偏移量 $\boldsymbol{O}$;最后在初始位置网格 $\boldsymbol{G}$ 上加入偏移量 $\boldsymbol{O}$,得到动态采样点集合 $\boldsymbol{S} \in \boldsymbol{R}^{2g\times sH\times sW}$。计算过程表示为:

$$\begin{cases} \boldsymbol{O} = \dfrac{1}{2}\sigma(\boldsymbol{W}_1\boldsymbol{F}) \otimes \boldsymbol{W}_2\boldsymbol{F} \\ \boldsymbol{S} = \boldsymbol{G} \oplus \boldsymbol{O} \end{cases} \tag{7}$$

式中:$\boldsymbol{O}$ 为采样偏移量,σ 为激活函数,$\boldsymbol{W}_1$,$\boldsymbol{W}_2$ 为不同的线性变换权重矩阵,$\boldsymbol{F}$ 为原始特征输入,$\boldsymbol{G}$ 为初始位置网格,$\boldsymbol{S}$ 为最终动态采样点集合。

对于动态采样点集合需进行如下预处理:首先,对采样点集合进行维度重排,将张量调整为 $(sH, sW, 2g)$ 的维度排列;之后沿通道维度拆分为 g 个独立采样组 $\{\boldsymbol{S}^{(k)}\}_{k=1}^{g} \in \boldsymbol{R}^{sH\times sW\times 2}$,通过二维双线性插值采样函数实现特征上采样。具体来说,对于每个独立采样组进行如下操作:

$$\boldsymbol{F}'(k) = grid\_sample(\boldsymbol{F}, \boldsymbol{S}^{(k)}), k \in \{1, \cdots, g\} \tag{8}$$

最终将各分组输出按照通道数进行拼接,得到上采样特征图 $\boldsymbol{F}' \in \boldsymbol{R}^{C\times H_{out}\times W_{out}}$。其中,$grid\_sample(\ )$ 函数先通过采样点集合的尺寸确认输出特征图的尺寸,然后对于输出特征图的每个像素位置,根据采样点集合得到原输入

特征图上的对应的坐标，最终根据坐标，使用双线性插值方法计算出输出图像的像素值。

相对于传统上采样的直接复制像素值的操作，动态上采样 DySample 能够根据特征图内容自适应调整采样偏移量，使采样点向关键边缘聚集，更好地保留原始特征的细节信息。

3）并行多尺度卷积模块

SAR 图像因成像机制与观测视角差异，呈现出显著的多尺度跨度特性。针对这一难题，设计并行多尺度卷积 PMC 模块嵌入 SAR 目标检测模型，核心思想是通过并行多尺度深度卷积密集提取局部特征，其无膨胀的卷积设计有效避免了 SAR 图像中相干斑噪声敏感问题；进一步结合上下文锚点注意力机制（context anchor attention，CAA）[24]捕获长距离轴向空间信息依赖关系，并强化中心区域特征，该机制对于桥梁等细长形状目标识别效果出色。并行多尺度卷积模块结构如图 7 所示。

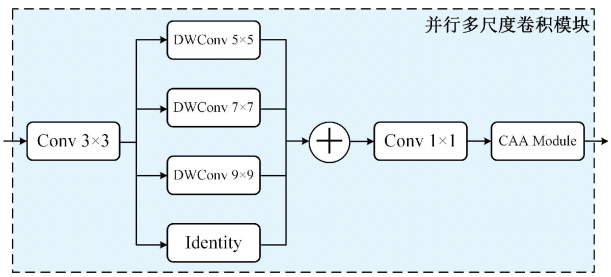


图 7　并行多尺度卷积模块结构图

Fig. 7　The architecture of parallel multi-scale convolution module

首先，输入特征图经过一个小核卷积来获取局部信息，然后通过一组并行多尺度的深度卷积来捕获多尺度细节特征，随后通过 1×1 卷积整合各分支特征，实现跨尺度的信息交互，最后经过上下文锚点注意力机制强化特征表达。并行多尺度卷积模块计算过程可被表示为：

$$\begin{cases} \boldsymbol{F}^{(i)} = DW_{k(i)}(Conv(\boldsymbol{F})), i = 1, \cdots, 3 \\ \boldsymbol{F}^{out} = CAA(Conv(\boldsymbol{F}^{pre} + \sum_{i=1}^{3} \boldsymbol{F}^{(i)})) \end{cases} \tag{9}$$

式中：$\boldsymbol{F}$ 表示输入特征图，$Conv(\ )$ 表示标准卷积操作，$DWConv_{k(i)}(\ )$ 表示卷积核尺寸为 $k(i)$ 的深度卷积，$\boldsymbol{F}^{(i)}$ 表示不同尺寸深度卷积得到的特征图，$\boldsymbol{F}^{out}$ 表示并行多尺度卷积模块输出特征图。上下文锚点注意力机制结构如图 8 所示。

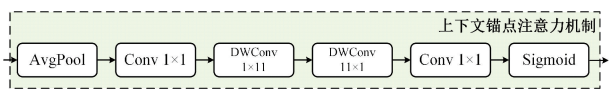


图 8　上下文锚点注意力机制结构图

Fig. 8　The architecture of context anchor attention module

首先采取平均池化和一个 1×1 卷积，获得部分区域特征，然后通过两个不同方向的条带深度卷积，使用这种卷积的优势在于，与同尺寸的大核深度卷积相比，仅用 $k/2$ 的参数量，就可以实现相近的卷积性能[25]，其中，k 为选定的卷积核尺寸；最后，通过 1×1 卷积与 $sigmoid$ 函数生成一个注意力权重，该权重被用于增强 PMC 模块的输出特征图。CAA 模块计算过程可被表示为：

$$\begin{cases} \boldsymbol{F}^{pool} = Conv(P_{avg}(\boldsymbol{F})) \\ \boldsymbol{F}^{w} = DW_{1\times k}(\boldsymbol{F}^{pool}) \\ \boldsymbol{F}^{h} = DW_{k\times 1}(\boldsymbol{F}^{w}) \\ \boldsymbol{F}^{att} = \sigma(Conv(\boldsymbol{F}^{h})) \end{cases} \tag{10}$$

式中：$\boldsymbol{F}$ 表示输入特征图，$P_{avg}(\ )$ 表示平均池化操作，$DW_{1\times k}(\ )$ 与 $DW_{k\times 1}(\ )$ 分别表示卷积核尺寸为 $1\times k$ 和 $k\times 1$ 的深度卷积，$\boldsymbol{F}^{pool}$ 表示池化后的特征，$\boldsymbol{F}^{w}$ 和 $\boldsymbol{F}^{h}$ 分别表示经水平方向和经垂直方向深度卷积后的特征图，$\boldsymbol{F}^{att}$ 表示输出的注意力权重特征图。

4）小尺寸目标检测头

在 SAR 图像中，小尺寸目标检测面临着多重挑战。首先是在目标尺度层面，如图 9(a)～(b)所示，在典型的机场区域中，飞机目标尺寸集中在 15 pixel×15 pixel，且同类目标密集排布导致目标间距小于 0.5 倍特征感受野，导致特征混叠效应发生；其次是在背景干扰层面，如图 9(c)～(d)海岸线区域中，相干斑噪声与地形地貌形成多尺度的背景干扰，其纹理特征与小目标高频分量存在重叠。传统的检测器在多次下采样后，如岸边船只等小尺寸目标保留特征衰减严重，最终导致小尺寸目标检测的漏检概率上升。

针对小目标特性，本文在检测头部分进行了优化。在主干网络部分的 4×下采样层增设高分辨率检测分支，通过 XMPAN 的跨层特征聚合，直接利用浅层特征图中未被稀释的高频分量，如边缘响应与角点分布。相较于传统方法将小目标检测依赖于深层语义的做法，XMPAN 实现的优化避免了高频细节被多次下采样过程所稀释，增强目标与背景的区分度，可有效缓解小目标与背景及其他目标的混淆问题，在保持检测效率的同时提升模型对小目标的敏感度。

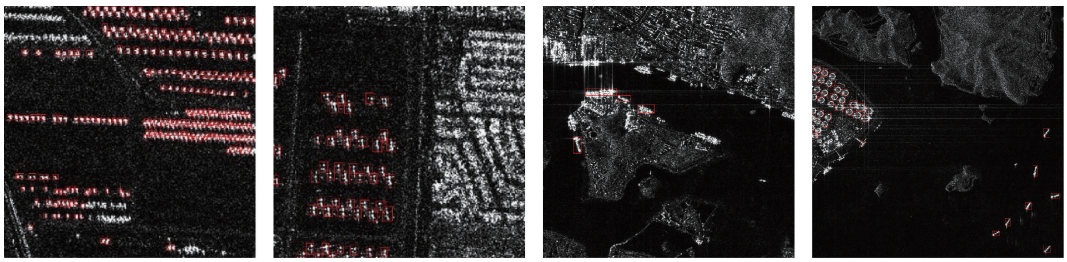
3　实验与分析

本章通过横向的对比实验和纵向的消融实验相结合对模型进行评估，系统性地验证了提出的 XMNet 网络对 SAR 图像目标检测的有效性。此外，实验设计特别引入可解释性分析工具，结合特征响应热力图可视化技术，揭示了模型在复杂散射特性、目标尺度多变等 SAR 图像检测特有挑战下的性能，为算法优化提供理论依据。

3.1　实验环境与参数设置

本章全部实验，均使用图像处理器（GPU）NVIDIA GeForce RTX4060 加速计算，实验平台使用编程语言 Python 3.11，基于开发框架 PyTorch 2.2.2 进行设计，使用 CUDA 12.1 进行加速训练。

综合考虑训练速度以及数据集图像尺寸，在训练过程中，设置批处理大小（batch size）为 16，训练轮次（epochs）为



(a) 机场停机坪场景1 (a) Airport apron(scene 1)　(b) 机场停机坪场景2 (b) Airport apron(scene 2)　(c) 海岸线场景1 (c) Coastline (scene 1)　(d) 海岸线场景2 (d) Coastline (scene 2)

图 9　MSAR-1.0 数据集中的典型小尺寸目标

Fig. 9　Typical small-sized targets in the MSAR-1.0 dataset

300,优化器使用带动量的随机梯度下降(stochastic gradient descent, SGD)进行优化,动量设置为 0.937,初始学习率(learning rate)为 0.01 且随轮次提升而降低,并采取在前 3 个轮次中设置学习率预热的策略以提升训练稳定性。

3.2　数据集

MSAR-1.0 数据集共包括 28 449 张检测切片,采用了海斯一号卫星和高分三号卫星数据。该数据集包含了大规模多类 SAR 目标,场景包括机场、港口、近岸、岛屿、远海以及城区,目标类型包括飞机、油罐、船只和桥梁 4 类。其中,船只实例 39 858 条,油罐实例 12 319 个,飞机实例 6 368 架,桥梁实例 1 851 架。MSAR-1.0 数据集切片尺寸为 256 pixel×256 pixel,部分桥梁切片为 2 048 pixel×2 048 pixel。在训练过程中,按照 6∶2∶2 的比例将数据集随机划分为训练集、验证集和测试集。

3.3　评价指标

实验选用了目标检测领域常用的性能指标来衡量模型的性能,其中包括平均精度(average precision, AP),平均精度均值(mean average precision, mAP)以及模型参数量(parameters)。特别地,平均精度和平均精度均值指标均在 IoU 阈值为 0.5 情况下进行统计,记作 AP@50 和 mAP@50,分别简记为 AP50 和 mAP50。具体定义如下:

AP 的计算遵循 MS COCO 标准,采用 101 点插值法:

$$AP = \sum_{i=0}^{101} \max_{\widetilde{r} \geqslant r} P(\widetilde{r}) \Delta r \tag{11}$$

式中:$r \in \{0, 0.01, \cdots, 1.00\}$ 为预设召回率阈值,$\max_{\widetilde{r} \geqslant r} P(\widetilde{r})$ 表示大于当前阈值的实际召回率对应的最大精度。

mAP 计算通过对 C 个类别的 AP 取均值:

$$mAP = \frac{1}{C} \sum_{C=1}^{C} AP_C \tag{12}$$

AP 的数学含义为某类别的精度-召回率曲线(P-R Curve)下的面积,精度与召回率的计算公式如下:

$$\begin{cases} Precision = \dfrac{TP}{TP + FP} \\ Recall = \dfrac{TP}{TP + FN} \end{cases} \tag{13}$$

式中:TP (true positive)表示正确预测的正样本数,FP (false positive)表示错误预测的正样本数,FN (false negative)表示错误预测的负样本数。

模型的复杂度通过参数量评估,单位为百万(M),反映了模型对于计算资源和内存资源的需求,通常来说,参数量越低表示模型越轻量化。

模型的推理速度通过 FPS(frames per second)进行评估,该数值反映了模型每秒钟可以处理的图像数量,在同一个数据集上的 FPS 可以用以评估模型推理速度以及实时性。

3.4　消融实验

为了探究算法中不同的模块设计对 XMNet 检测性能的影响,本节以 YOLOv8 作为基准方法,进行了多项消融实验,重点验证主干网络优化、颈部特征融合结构以及跨尺度特征融合对检测性能的贡献。消融实验结果如表 1 所示,实验通过逐步引入各创新结构,分别测试了不同情况下的各类检测精度。

首先将基线模型的主干网络升级为改进型 SHViT 架构后,全类别检测平均精度均值提升 2.4%。这得益于该网络引入的全局注意力机制,通过跨通道交互与长程依赖建模显著增强了目标特征的判别性表征能力,此阶段推理速度降至 192.7 FPS,主要因注意力计算增加了计算开销。在基线模型中新增优化检测头后,为模型全类别检测平均精度均值带来了 4.8%的提升,推理速度降至 202.3 FPS,特别地,模型在飞机等小目标检测任务中取得 13.6%的精度突破,这是因为新增的检测头增强了浅层特征的利用,进一步验证了针对 SAR 图像小目标特性设计的特征强化机制的有效性。

进一步采用 XMPAN 结构重构颈部网络,通过跨层特征融合策略实现多尺度特征互补,模型的检测平均精度均值提升了 1.9%。该策略通过将不同层级的特征图进行聚合,极大地提升了模型的多尺度检测能力,实验结果表明,应用了 XMPAN 结构后,对油罐类中型目标检测精度提升达 5%,证明了该策略能够有效解决 SAR 图像中目标尺度差异大的难题,同时推理速度回升至 193.5 FPS,优于未应用 XMPAN 结构的模型,这是因为 XMPAN 结构设计中特

征复用减少了重采样操作的计算量。值得关注的是，颈部网络 XMPAN 中引入动态上采样和并行多尺度卷积模块后，使模型的平均检测精度均值分别提升至 90.1% 和 90.4%。前者通过可学习的自适应采样核实现特征空间重校准，后者借助异构卷积并行计算捕获跨尺度上下文信息，二者协同作用使桥梁目标的召回率提升 2.2%，在维持定位精度的同时显著缓解了细长目标漏检问题，这些实验数据验证了各模块设计的协同优化作用。

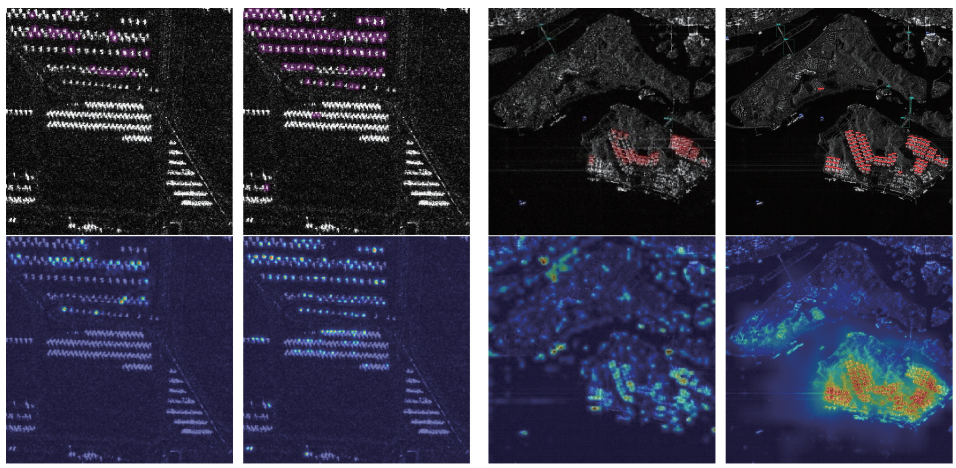
在图 10（a）～（d）中对比展示了基准模型（左）与 XMNet（右）在不同场景下的 SAR 图像检测性能及其热力

表 1　XMNet 的消融实验

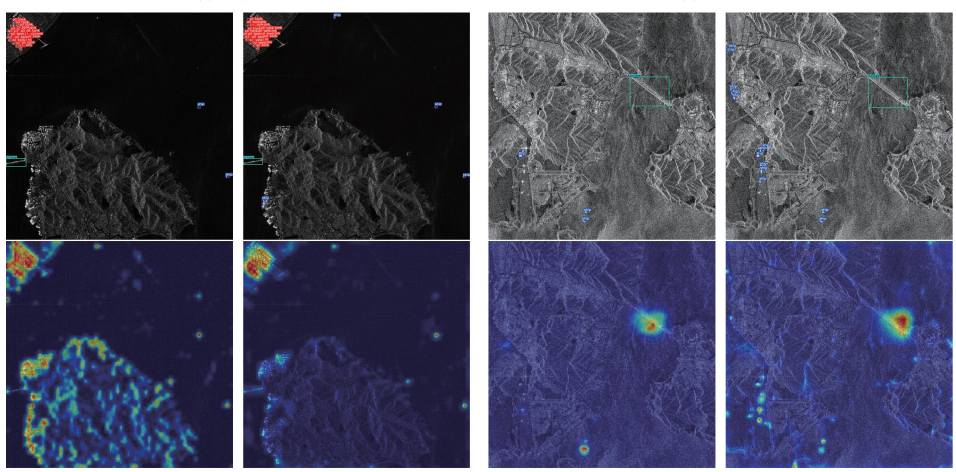
Table 1　Ablation experiment of XMNet

Baseline	+SHViT	+Head	+XMPAN	+DySample	+PMCBlock	mAP50/%	AP50/%				FPS
							船只	油罐	飞机	桥梁	
√						81.7	91.8	82	73.3	79.6	**226.1**
√	√					84.1	92.9	83	77.1	83.3	192.7
√		√				86.5	92.2	86.2	86.9	80.8	202.3
√	√	√				87.9	93.4	83.7	90.3	84.2	188.6
√	√	√	√			89.8	93.9	88.7	92.4	84.1	193.5
√	√	√	√	√		90.1	94.0	88.3	92.5	**85.4**	191.6
√	√	√	√	√	√	**90.4**	**94.1**	**88.8**	**93.4**	85.3	185.4

注：加粗字体为每列最优值。“√”表示添加模块，空白表示未添加模块。



(a) 机场场景　(b) 港口场景
(a) Airport　(b) Harbor



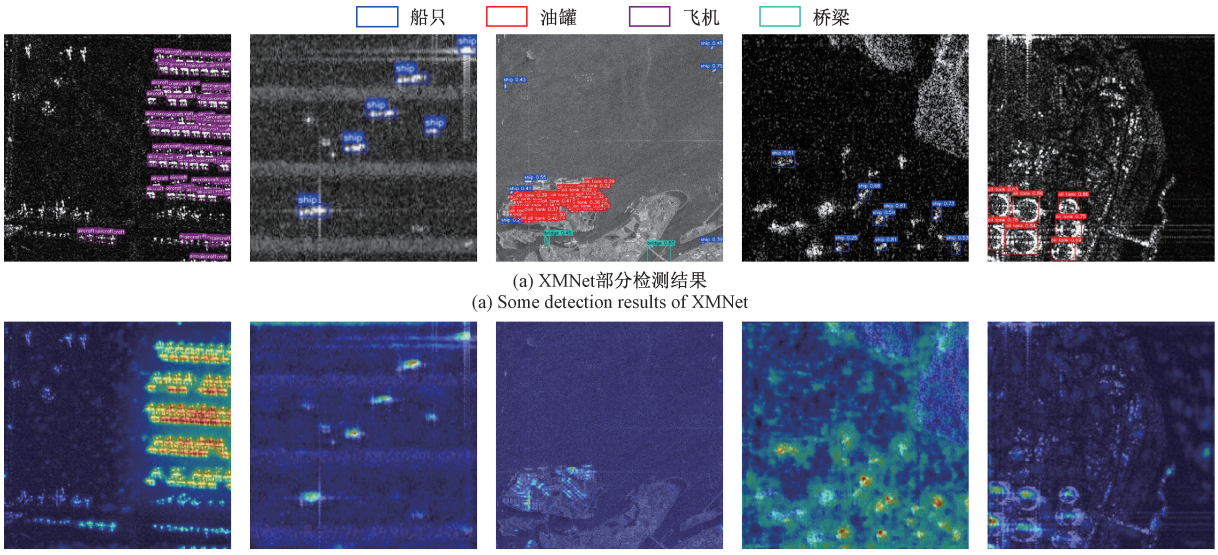
(c) 岛屿场景　(d) 近岸场景
(c) Islands　(d) Nearshore

图 10　XMNet 与基准模型检测结果对比及热力图可视化（左图为基准模型，右图为 XMNet）

Fig. 10　Comparison and heatmap visualization of XMNet and baseline (left: Baseline, right: XMNet)

图可视化效果。实验结果表明,通过引入优化的检测头并实施跨层多尺度特征融合策略,XMNet 在小尺寸目标检测性能上实现了显著提升。具体而言,在图 10(a)～(d)各组场景中,XMNet 不仅能精准定位小尺寸目标,在目标密集分布区域也展现出优异的鉴别能力。特别值得注意的是,在图 10(c)组场景的复杂地形边缘区域,基准模型产生大量虚警信号;而 XMNet 通过增强的特征表征能力,有效抑制了背景干扰,准确区分了地形边缘特征与真实目标。这些结果充分验证了跨层特征融合策略在提升模型区分目标与背景能力方面的有效性。

图 11 综合展示了 XMNet 在 MSAR-1.0 数据集 3 类典型场景下的检测结果与热力图响应特征。在机场密集飞机目标、海洋相干斑噪声干扰下的船只以及海港区域内油罐与桥梁共存的复杂场景中,模型均实现了高精度目标定位与背景抑制。热力图可视化结果表明,目标区域内最高响应强度与标注框内关键点高度吻合,而在背景杂波中呈现显著的低响应特征。该现象印证了 XMNet 对 SAR 图像中目标语义的深层理解能力,同时验证了其在相干斑噪声干扰和复杂背景条件下保持强判别力的鲁棒性优势。



(a) XMNet部分检测结果
(a) Some detection results of XMNet

(b) XMNet检测热力图可视化
(b) XMNet detection heatmap visualization

图 11　XMNet 在 MSAR-1.0 数据集上的部分检测结果与热力图可视化

Fig. 11　Part of detection results and heat map visualization of XMNet on the MSAR-1.0 dataset

实验结果表明,XMNet 在 MSAR-1.0 数据集上展现出卓越的检测性能,其全类别平均精度 mAP50 达 90.4%,较基准模型提升 8.7%,显著超越其他 SAR 图像检测模型 HRLE-SARDet(86.8%)和 MDD-YOLOv8(78.9%)。在极具挑战性的小目标检测任务中,模型针对飞机类别的检测精度实现 20.1%的显著提升,达到了 93.4%,同时全类别的检测性能全面领先,AP50 指标均为所有模型中最高,这验证了本文算法在 SAR 图像微小目标特征提取与定位方面的性能突破。与其他主流检测架构相比,XMNet 在多尺度目标检测适应性上展现出了优势,相较于 FCOS 与 CenterNet 等先进模型,mAP50 分别提升 25%和 24.5%,而 XMNet 参数量仅为 5.17 M,同时图像推理速度可达 185.4 FPS,仍保持与基线模型相近的推理速度,且远高于其他模型,体现了轻量化的优势。

图 12 展示了不同模型在 MSAR-1.0 数据集上的检测结果,图 12(a)为真实标注,图 12(e)为本算法提出的 XMNet 模型,图 12(b)～(d)3 组分别为对比算法 CenterNet、YOLOv5 和 FCOS 的检测结果。在首组典型机场场景的密集目标检测中,CenterNet 与 YOLOv5 均出现检测框冗余现象,尤其在停机坪区域产生显著的重叠检测框;FCOS 算法则存在明显的漏检问题,未能完整识别排列紧密的飞机目标。相较之下,XMNet 通过创新的特征融合机制,在保证召回率的同时实现了精确的目标定位,展现出最优的检测性能。

3.5　对比实验

为进一步评估 XMNet 的检测有效性,将本文方法与其他的检测方法在 MSAR-1.0 数据集上进行了对比实验。选取的对比方法包括:Faster R-CNN,Retinanet[26],FCOS[27],PVT[28],YOLOv5,CenterNet[29],HRLE-SARDet[30],MDD-YOLOv8,以及基线模型 YOLOv8。表 2 中详细对比了上述前沿模型、基准模型(baseline)与 XMNet 在典型 SAR 目标上的检测精度 AP50 和全类平均精度均值 mAP50、各模型参数量以及 FPS 指标。

在第 2、3 组大尺度遥感场景的检测分析中,对比模型对小尺度目标的敏感性不足问题尤为突出。具体而言,YOLOv5 和 CenterNet 在处理细长桥梁结构时存在显著

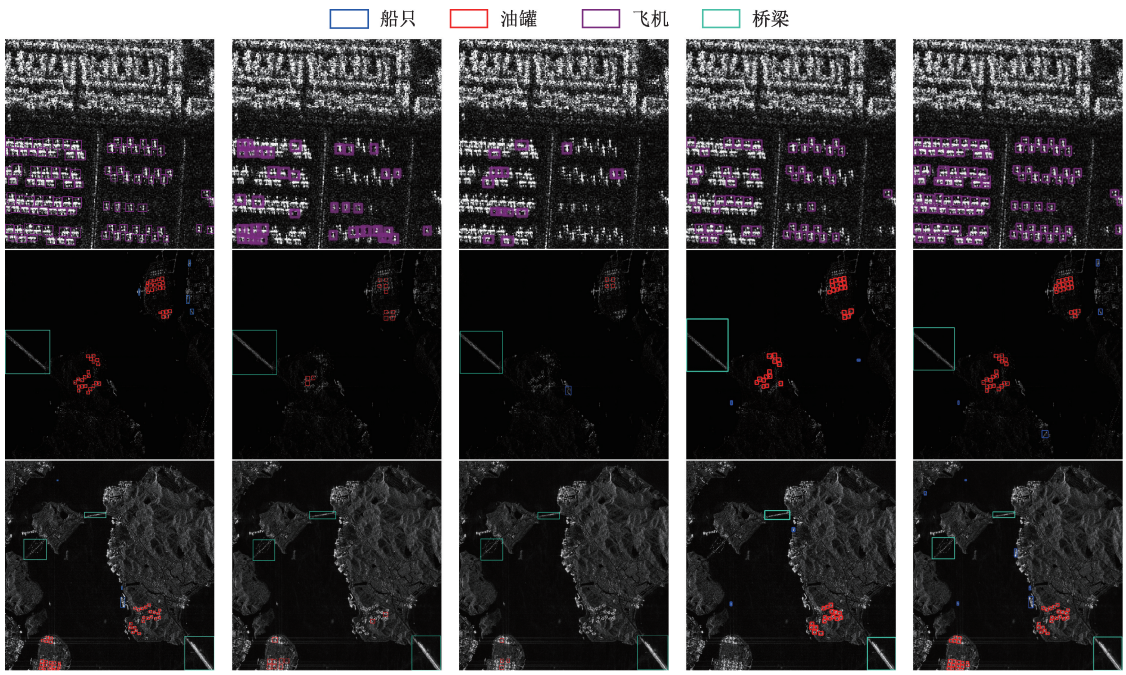
漏检现象，而 FCOS 在船只和油罐检测中则出现了明显虚警问题。这些模型受限于特征融合网络的尺度适应性不足问题，难以有效捕捉细长地物（如桥梁）的形态特征和小型目标（如船只）的判别性信息。相比之下，XMNet 通过引入跨层多尺度特征融合，在充分保留大场景全局上下文的基础上，显著增强了微小目标的特征表达能力，从而在复杂遥感场景中实现了更精准的细长结构定位和更可靠的小目标判别能力。

表 2　不同方法在 MSAR-1.0 上的对比实验

Table 2　Comparative experiments of different methods on MSAR-1.0

方法	mAP50/%	AP50/%				参数量	FPS
		船只	油罐	飞机	桥梁		
Faster R-CNN	58.0	82.4	74.4	2.3	72.7	41.36 M	33.2
CenterNet	65.9	91.1	76.3	19.8	76.2	32.12 M	34.5
Retinanet	58.9	87.4	73.2	3.3	71.8	37.97 M	38.4
FCOS	65.4	90.4	76.7	16.0	78.3	32.3 M	39.8
PVT	57.1	86.0	71.7	2.3	68.5	12.95 M	43.1
HRLE-SARDet	86.8	92.7	85.9	85.4	83.2	8.96 M	112.7
MDD-YOLOv8	78.9	90.1	69.8	88.7	67.0	3.64 M	134.4
YOLOv5	75.1	87.3	80.1	66.9	66.0	**2.51** M	**257.5**
Baseline/YOLOv8	81.7	91.8	82.0	73.3	79.6	3.16 M	226.1
XMNet	**90.4**	**94.1**	**88.8**	**93.4**	**85.3**	5.17 M	185.4

注：加粗字体为每列最优值。



(a) 真实标注
(a) Ground truth　(b) CenterNet　(c) FCOS　(d) YOLO v5　(e) XMNet

图 12　不同模型在 MSAR-1.0 数据集上的检测结果

Fig. 12　Detection results of different models on the MSAR-1.0 dataset

3.6　模型泛化性验证分析

为了评估模型在未见场景下的泛化能力，本研究设计了零样本跨域测试。该测试选择两个具有显著域差异的 SAR 图像数据集——SADD 与 SAR-Ship-Dataset，直接应用训练好的模型进行推理预测，不进行任何针对新域的微调或再训练。这两个数据集在目标语义和场景分布上存在本质区别：SADD 专注于机场场景下的飞机目标检测，其高分辨率图像包含滑行跑道、维修设备等复杂的地面背

景干扰物;而 SAR-Ship-Dataset 则针对海洋环境中的船舶检测,其挑战主要源于多尺度的船体目标及海浪杂波的动态组合。这种直接将未在训练过程中见过的场景用于测试能够更真实地反映模型在面对未知且特性迥异环境时的适应性与鲁棒性。

在严格的零样本迁移设定下,本文方法 XMNet 相较基线方法展现出稳健的跨域适应能力。如表 3 所示,在 SADD 数据集上,XMNet 以 74.3% 的 mAP50 超越基线 YOLOv8 的 63.8%,且召回率达 68.9%;在 SAR-Ship-Dataset 数据集上,如表 4 所示,XMNet 达到了 75.0% 的 mAP50,远超过基线的 50.6%。虽然推理速度略低于基线方法,但是 XMNet 在检测准确率和召回率方面有着更优的表现。

表 3　XMNet 与基线方法在 SADD 数据集上对比

Table 3　Comparison of XMNet and baseline on the SADD

方法	mAP50/%	P/%	R/%	FPS
YOLOv8	63.8	65.5	56.7	**176.46**
XMNet	**74.3**	**73.2**	**68.9**	161.08

表 4　XMNet 与基线方法在 SAR-Ship-Dataset 数据集上对比

Table 4　Comparison of XMNet and baseline on the SAR-Ship-Dataset

方法	mAP50/%	P/%	R/%	FPS
YOLOv8	50.6	61.0	41.1	**177.59**
XMNet	**75.0**	**78.3**	**64.1**	146.51

XMNet 在独立数据集上仍有不错的检测性能表现,实验结果证实本文方法具备目标无关的特征泛化能力,对场景背景干扰与 SAR 典型噪声(如相干斑噪声、海浪杂波)具有强鲁棒性,为跨任务 SAR 目标检测的实际部署提供了有效性支撑。

4　结　　论

针对合成孔径雷达(SAR)图像中多尺度目标检测精度不高,小目标漏检率高的挑战,提出了应用多尺度特征跨层融合策略的 XMNet 模型。通过引入改进型 SHViT 并重新设计 XMPAN,其中引入动态上采样(DySample)以及并行多尺度卷积模块(PMC block),有效解决了传统方法对于小目标以及细长目标的特征表征方面的不足。实验表明,XMNet 模型在 MSAR-1.0 数据集上的全类别平均检测精度达到了 90.4%,相较于基准模型提升了 8.7%,其中飞机类小目标的检测精度提升幅度达 20.1%,验证了跨层特征融合策略的有效性。总之,XMNet 在保持模型轻量化的同时,显著提升了模型在 SAR 图像目标检测的工程实用性,为遥感的实时监测提供了基础。未来的工作可以探索模型在低分辨率 SAR 图像数据集上的表现,针对低分辨率 SAR 数据特性进行结构设计与优化,探究目标特征弱化对于检测稳定性的影响,最终实现更加鲁棒的 SAR 图像检测框架。

参考文献

[1] TSOKAS A, RYSZ M, PARDALOS P M, et al. SAR data applications in earth observation: An overview[J]. Expert Systems with Applications, 2022, 205:117342.

[2] LI T, LIU ZH, XIE R, et al. An improved superpixel-level CFAR detection method for ship targets in high-resolution SAR images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 11(1): 184-194.

[3] WANG C L, BI F K, CHEN L, et al. A novel threshold template algorithm for ship detection in high-resolution SAR images[C]. 2016 IEEE International Geoscience and Remote Sensing Symposium(IGARSS), 2016:100-103.

[4] REDMON J. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016:779-788.

[5] REDMON J. Yolov3: An incremental improvement[J]. ArXiv preprint arXiv:1804.02767, 2018.

[6] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017:7263-7271.

[7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14, 2016:21-37.

[8] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6):1137-1149.

[9] SUN ZH ZH, LENG X G, LEI Y, et al. BiFA-YOLO: A novel YOLO-based method for arbitrary-oriented ship detection in high-resolution SAR images[J]. Remote Sensing, 2021, 13(21):4209.

[10] SU N, HE J Y, YAN Y M, et al. SII-Net: Spatial information integration network for small target detection in SAR images[J]. Remote Sensing, 2022, 14(3):442.

[11] ZHAO K, LU R T, WANG S Y, et al. ST-YOLOA: A swin-transformer-based YOLO model with an

attention mechanism for SAR ship detection under complex background[J]. Frontiers in Neurorobotics, 2023, 17:1170163.

[12] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. IEEE/CVF International Conference on Computer Vision, 2021:10012-10022.

[13] 李波, 李志康, 周钰彬. 结合特征融合和注意力机制的 SAR 舰船检测算法[J]. 电子测量技术, 2024, 47(10):134-140.
LI B, LI ZH K, ZHOU Y B. SAR ship detection algorithm combining feature fusion and attention mechanism[J]. Electronic Measurement Technology, 2024, 47(10):134-140.

[14] LIU J, LIU X, CHEN H X, et al. MDD-YOLOv8: A multi-scale object detection model based on YOLOv8 for synthetic aperture radar images[J]. Applied Sciences(2076-3417), 2025, 15(4):2239.

[15] 胡欣, 马丽军. 基于 YOLOv5 的多分支注意力 SAR 图像舰船检测[J]. 电子测量与仪器学报, 2022, 36(8):141-149.
HU X, MA L J. Multi-branch attention SAR image ship detection based on YOLOv5[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(8):141-149.

[16] YUN S, RO Y. Shvit: Single-head vision transformer with memory efficient macro design[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024:5756-5767.

[17] LIU W Z, LU H, FU H T, et al. Learning to upsample by learning to sample[C]. IEEE/CVF International Conference on Computer Vision, 2023: 6027-6037.

[18] ZHANG P, XU H, TIAN T, et al. SEFEPNet: Scale expansion and feature enhancement pyramid network for SAR aircraft detection with small sample dataset[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15:3365-3375.

[19] WANG Y Y, WANG CH, ZHANG H, et al. A SAR dataset of ship detection for deep learning under complex backgrounds[J]. Remote Sensing, 2019, 11(7):765.

[20] LIU SH, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018:8759-8768.

[21] DOSOVITSKIY A. An image is worth 16x16 words: Transformers for image recognition at scale[J]. ArXiv preprint arXiv:2010.11929, 2020.

[22] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018:7132-7141.

[23] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017:2117-2125.

[24] CAI X H, LAI Q X, WANG Y W, et al. Poly kernel inception network for remote sensing detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024:27706-27716.

[25] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016:2818-2826.

[26] ROSS T Y, DOLLÁR G. Focal loss for dense object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017:2980-2988.

[27] TIAN ZH, SHEN CH H, CHEN H, et al. Fcos: Fully convolutional one-stage object detection[C]. IEEE/CVF International Conference on Computer Vision, 2019:9627-9636.

[28] WANG W H, XIE E, LI X, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions[C]. IEEE/CVF International Conference on Computer Vision, 2021:568-578.

[29] DUAN K W, BAI S, XIE L X, et al. Centernet: Keypoint triplets for object detection[C]. IEEE/CVF International Conference on Computer Vision, 2019: 6569-6578.

[30] ZHOU ZH, CHEN J, HUANG ZH X, 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.

作者简介

赵喆,硕士,主要研究方向为遥感图像处理。
E-mail:zhaoz9e@nuaa.edu.cn

李勃(通信作者),副研究员,主要研究方向为遥测遥控。
E-mail:libo70205830@nuaa.edu.cn

徐文校,博士,主要研究方向为目标检测。
E-mail:971149614@qq.com

李尧,硕士,主要研究方向为无人机路径规划。
E-mail:1044970679@qq.com