ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 41,No. 1,Jan. 2026,pp. 202－214
DOI:10. 16337/j. 1004-9037. 2026. 01. 014

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

# 联合空间-通道特征及频率选择的SAR目标检测

纪晓平[1], 陶　普[2]

(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:** 针对由于SAR图像存在目标数量和种类多、尺度各异、高度复杂的背景相干斑噪声等特性导致检测精度低的问题,提出了一种联合空间-通道特征及频率选择的SAR目标检测算法。首先,采用经过预训练的ResNet-50网络作为主干网络来提取目标多尺度特征,并通过联合多尺度空间-通道特征增强模块的特征金字塔网络来增强对多尺度特征的表征。随后,在特征域引入频率选择模块来选择性地去除噪声同时保留目标信号以达到增强目标特征的目的。在标准数据集MSAR和SARDet-100K上进行了对比实验,结果表明,该算法在两个数据集上均超越了现有SAR图像目标检测算法Faster R-CNN、ConvNeXt、PVT-T和YOLOF,达到了最优性能。

**关键词:** SAR目标检测;多尺度特征;空间-通道特征;频率选择;特征去噪

**中图分类号:** TP753　　**文献标志码:** A

**引用格式:** 纪晓平,陶普.联合空间-通道特征及频率选择的SAR目标检测[J].数据采集与处理,2026,41(1):202-214. JI Xiaoping,TAO Pu. SAR target detection via joint spatial-channel feature and frequency selection[J].Journal of Data Acquisition and Processing,2026,41(1):202-214.

## 引　　言

合成孔径雷达(Synthetic aperture radar, SAR)作为一种有源微波传感器,具有全时段工作能力且受云、雨、雾等气象条件影响较小[1],在全天候监视方面表现出色。该技术通过搭载于运动载体的雷达设备持续发射电磁波信号,并同步接收地物后向散射回波,借助平台位移形成的等效长孔径实现方位向高分辨率成像。SAR基于独特的相干成像原理和地物电磁响应特性,在环境监测、军事侦察[2-3]及防灾减灾等领域发挥着重要作用,相关应用已在多项研究中得到验证[4-6]。然而,SAR仍然面临着巨大挑战:(1)由于SAR是一种相干成像系统,导致其采集的图像中不可避免地存在相干斑噪声,这些噪声与目标和复杂背景叠加给精准的目标检测带来了极大的难度[7];(2)受限于SAR采集的图像空间分辨率,SAR图像中的目标大多表现为小目标,强噪声导致在检测过程中目标极易受到干扰[8]。以上挑战使得发展兼具高精度与实时性的目标检测算法成为当前遥感信息处理领域的前沿课题,相关研究对提升SAR数据应用效能具有重要价值。

随着深度学习在目标检测领域取得的巨大进展[9],以及MSAR[10]、SARDet-100K[11]等适用于SAR图像目标检测领域的大规模SAR图像数据集的出现,基于深度学习的SAR图像目标检测任务被广泛研究。相较于传统手工设计的图像特征提取模块,基于神经网络设计的特征提取网络可以自适应地从大量数据中学习网络参数,使用训练好的网络不仅可以十分有效地提取图像中的丰富特征,还可以在很大程度上保证网络的准确性与稳定性。现有的基于深度学习的目标检测算法主要可以分为两类:双阶段和单阶段目标检测算法。以Faster R-CNN[12]和ConvNeXt[13]为例,双阶段算法通常需要先定位目

**收稿日期:**2025-08-05;**修订日期:**2025-09-17

标物的位置并生成建议框,再对建议框进行分类和更精准的定位。这种方法虽然保证了较高的精度和召回率,但速度较慢缺乏实时性;以PVT-T[14]和YOLOF[15]为例,单阶段算法通过直接预测物体的类别概率和位置坐标,以端到端的方式在一次检测中给出最终结果。因此,与双阶段算法相比,单阶段算法通常能降低资源消耗并具有更快的检测速度。

为了克服由相干斑噪声导致检测精度降低的问题,研究人员开展了大量的工作。Ai等[16]提出了融合多尺寸特征从而可以利用更多有用的特征信息进行SAR目标分类,其优异的性能证明了多尺度特征提取在目标检测中发挥了重要作用。Si等[17]通过改进YOLOv5中特征融合部分的网络架构,并引入通道注意力机制以增强特征捕获能力,该方法相较于原始YOLOv5网络在性能上有了一定的提升。Li等[18]提出了轻量级SAR目标检测网络SFS-CNet,通过分流处理特征图的空间域和频率域信息并采用特征选择模块融合两部分信息,增强了模型对多尺度特征的提取能力,但算法的检测准确度提升并不显著。肖振久等[19]通过在YOLOv7中引入基于U-Net设计的去噪模块设计了基于Transformer的上下文信息提取模块,这使得该网络有效抑制了相干斑噪声对检测过程的干扰,并在一定程度上减轻了漏检概率,但Transformer的引入带来了较大的计算负担。Zhou等[20]首次提出了基于扩散模型的SAR图像飞机目标检测网络DiffDet4SAR。该方法将飞机目标检测任务映射到边界框的去噪扩散过程,从而有效适应了飞机尺寸的大范围变化。研究者还设计了散射特征增强模块,用于降低推理过程中的杂波强度并增强目标显著性。然而,该算法在不同数据集上的泛化能力仍有待提升。Zhang等[21]提出了频率注意力模块,通过自适应地处理SAR图像中的频域信息抑制图像中存在的海杂波等背景噪声,进而实现目标特征的显著性增强以提升网络检测性能。

针对上述所提到的研究及存在的问题,本文在FCOS算法[22]基础上提出了一种联合空间-通道特征及频率选择的SAR图像目标检测算法。首先,由于经过主干网络初步特征提取后所得到的多尺度特征图中包含了丰富的空间信息和通道间信息,设计了用于特征增强的多尺度特征增强模块(Multi-scale feature enhancement module, MSFEM),并将其作为特征金字塔网络(Feature pyramid network, FPN)的主要组件,这样可以增强对多尺度特征的表征能力。此外,为了能够保留与小尺寸目标特征相关的高频信息,同时优化卷积网络对低频分量的固有偏差,加之考虑到诸如离散小波变换等在本质上是全局性的,这使得变换到频域后的特征能够获取具有更大感受野的上下文信息。最后,采用动态回归损失函数训练所提出的网络,该损失函数的数学表达式为

$$L(\{p_{x,y}\},\{t_{x,y}\})=\frac{1}{N_{\text{pos}}}\sum_{x,y}L_{\text{cls}}(p_{x,y},c^{*}_{x,y})+\frac{\lambda}{N_{\text{pos}}}\sum_{x,y}1_{\{c^{*}_{x,y}>0\}}L_{\text{reg}}(t_{x,y},t^{*}_{x,y}) \tag{1}$$

式中:$L_{\text{cls}}$为Lin等[23]提出的焦点损失;$L_{\text{reg}}$为UnitBox[24]中提出的IoU损失;$N_{\text{pos}}$代表阳性样本(Positive sample)的数量;$\lambda$为平衡权重,本文取为1;$1_{\{c^{*}_{x,y}>0\}}$为指示函数,当$c^{*}_{x,y}>0$时其值取为1,否则为0。

这样不仅确保了算法对目标检测和定位的精准度,还加快了模型的收敛速度并保证了训练的稳定性。

## 1 联合空间-通道特征及频率选择的SAR图像目标检测算法

针对SAR图像中常见的噪声干扰问题,提出了一种基于FCOS检测器的结合空间-通道特征融合与特征维度去噪的高鲁棒性单阶段目标检测算法。FCOS检测器是无锚框的,这使得检测器完全避免了在训练过程中计算与锚框相关的复杂计算,也避免了与锚框相关的超参数。本文整体网络结构参考FCOS的框架进行设计,即经典的金字塔结构,不同点在于增加了特征域去噪操作,具体架构如图1所示,主要由4个子模块组成:主干网络、特征金字塔网络、特征去噪网络和检测头。

在主干网络部分,采用ResNet-50作为骨干特征提取器,通过4个Stage层逐级提取图像的层次化特征。该网络结构结合了残差连接与多次下采样操作,可高效构建具有不同尺度的金字塔式特征表

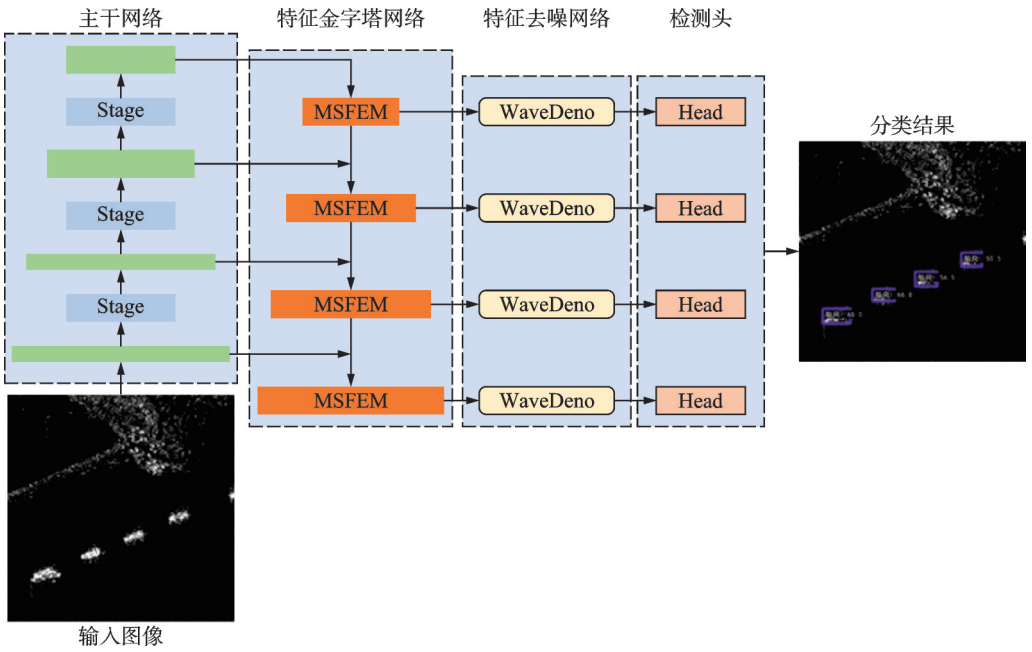


图1 本文网络总体架构

Fig.1 Overall architecture of the proposed network

示,为后续模块提供多尺度语义信息。

在特征金字塔网络中,设计了MSFEM模块,分别嵌入在每一尺度特征路径中。该模块融合空间注意力和通道注意力机制,增强关键目标区域和显著语义通道的响应能力,并采用自顶向下与横向连接结合的方式实现特征融合,有效提升网络对不同尺度目标的检测能力。

考虑到SAR图像中的高频噪声会严重干扰后续目标识别,引入了基于小波变换的WaveDeno模块。该模块在频域上对多尺度特征图进行小波分解,并利用频率选择策略抑制随机噪声和伪影干扰,从而提升特征图的清晰度和可辨识性。在检测头部分,采用Tian等[22]提出的FCOS结构,该方法摒弃传统锚框机制,直接在特征图上进行像素级回归与分类预测。FCOS引入了中心度估计(Centerness)分支,用于过滤边界质量较低的候选框,并结合Focal Loss优化训练过程,显著增强模型对小目标与复杂背景下目标的检测性能。最终,通过多个特征层的Head输出进行密集预测,实现目标的端到端检测与分类。

### 1.1 多尺度空间-通道联合特征增强模块

为了充分处理并融合经过主干网络提取到的特征,同时获取更加丰富的多尺度特征,提出了一个多尺度空间-通道联合特征增强模块。考虑到Transformer近年来在计算机视觉领域中取得的优异表现,同时为了避免由于其中的自注意力机制带来的二次复杂度,参考Transformer的结构设计了多尺度空间-通道联合特征增强模块。其中,根据文献[25]的描述,批归一化(Batch normalization, BN)存在小批量时性能不稳定的情况,引入层归一化(Layer normalization, LN)来保证训练过程中的稳定性。

多尺度特征提取对目标检测任务来说至关重要,为了进一步对特征图进行多尺度深度特征提取及融合,同时不过度增加计算复杂度,采用3个卷积核大小分别为3×3、5×5和7×7的并行的深度可分离卷积来充分获取多尺度信息,在融合阶段采用了对应像素相加来实现,表达式为

$$\mathrm{Spa}(X)=\mathrm{DWConv}_{3\times3}(X)+\mathrm{DWConv}_{5\times5}(X)+\mathrm{DWConv}_{7\times7}(X) \tag{2}$$

式中:$X$表示输入特征图,$\mathrm{DWConv}_{n\times n}$表示卷积核尺寸为$n\times n$的深度可分离卷积(Depth-wise convolu-

tion, DWConv)操作。

现有的绝大多数网络中均需要添加激活函数以提升网络的非线性表示能力,其中ReLU(Rectified linear unit)、GELU(Gaussian error linear unit)等经典激活函数得到了充分的探索和使用。参考现有的一些最优方法[26]所采用的门限线性单元(Gated linear unit, GLU)所取得的优异表现,提出一种简化但性能不显著降低的GLU作为非线性激活函数的替代品,其表达式为

$$\mathrm{Gate}(X,f,g,\sigma)=f(X)\odot\sigma(g(X)) \tag{3}$$

式中:$X$表示特征图,$f$和$g$均表示线性变换,$\sigma$代表例如Sigmoid这样的非线性激活函数,"$\odot$"表示逐像素相乘。由式(3)可知,这一操作相较于直接使用如GELU等非线性激活函数会带来更大的计算复杂度。因此,参考GELU这一激活函数,表达如下

$$\mathrm{GELU}(x)=x\Phi(x) \tag{4}$$

式中:$\Phi$表示标准正态分布的累计分布函数,可进一步近似表达为

$$\mathrm{GELU}(x)=0.5x\left(1+\tanh\left[\sqrt{\frac{2}{\pi}}\,(x+0.044\,715x^3)\right]\right) \tag{5}$$

根据以上描述,GELU可以看作是GLU的一种特殊情况。由式(5)推断可知,GLU本身具有非线性也因此可以不依赖于非线性激活函数,GLU可以简化为:$\mathrm{Gate}(X)=f(X)\odot g(X)$。基于以上描述,直接将特征图沿通道维度均匀分为两组并对两个分组进行逐像素相乘来作为所提出模块的激活函数,这一过程表达为

$$\mathrm{Gate}(X_1,X_2)=X_1\odot X_2 \tag{6}$$

式中:$X_1$和$X_2$为特征图$X$沿通道维度均匀分割而来的特征图。

为了充分利用特征图通道间的信息,在设计这一模块时加入了通道注意力模块,参考Hu等[27]提出的挤压与激励(Squeeze-and-excitation,SE)网络,即通过将空间信息压缩仅保留通道间信息,随后使用多层感知机来计算通道注意力权重与特征图的乘积为特征图实现通道间注意力分配,表达式为

$$\mathrm{SE}(X)=X*\sigma(W_2\mathrm{RELU}(0,\ W_1\mathrm{GAP}(X))) \tag{7}$$

式中:$X$为输入特征图;GAP代表全局平均池化(Global average pooling, GAP);$\sigma$为非线性激活函数,通常为Sigmoid;$W_1$和$W_2$为构成多层感知机的全连接层,两层全连接层之间采用ReLU作为激活函数。为了在不降低性能的前提下尽量降低计算复杂度,通过仅保留SE模块的空间信息压缩和通道信息交互设计了一个更为高效的通道注意力模块,即

$$\mathrm{CA}=X*\mathrm{Conv}_{1\times1}(\mathrm{GAP}(X)) \tag{8}$$

通过将以上层归一化层、多尺度特征提取模块、通道注意力模块和一个简单的前向传播模块按照Transformer的结构组合在一起,得到了所设计的多尺度空间-通道联合特征提取模块,具体结构如图2所示。

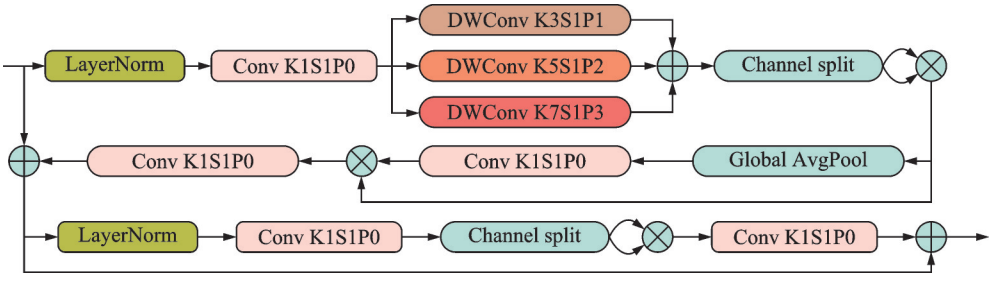


图2 多尺度特征增强模块MSFEM结构

Fig.2 Structure of the proposed multi-scale feature enhancement module (MSFEM)

图2中LayerNorm表示层归一化，Conv K1S1P0表示卷积核大小为1、步长为1、填充为0的卷积操作，DWConv K3S1P1表示卷积核大小为3、步长为1、填充为1的DWConv操作，Channel split表示沿通道维度对特征图进行均匀分割操作。

### 1.2 特征去噪模块

如图3所示，WaveDeno中包含了用于前向变换的二维离散小波变换（Discrete wavelet transform，DWT）模块、用于通道间注意力分配的SE模块和用于频率选择的模块。在频域中，通过DWT将空间信息转换为多尺度的频率表示，揭示了一种层次化的结构模式：低频分量（捕捉图像的全局结构信息）集中于近似子带，而高频分量（包含边缘和纹理等局部细节）则分布于水平、垂直及对角方向的细节子带中。本文通过利用这种具有方向性和多分辨率特征的频谱结构，采用较为常用的Haar小波作为小波基，将频域特征划分并映射到与不同小波子带对应的专属子空间中，随后结合通道注意力模块SE实现对输入数据低频和高频信息的整合，从而强化对相关信号的聚焦与分析。为了促进这种有针对性的处理，将二维特征图展开为一维向量，同时保留各子带间的结构关联性。这种变换确保了低频到高频、粗到细的信息层级得以完整表达，并允许在各独立子空间中进行有效的处理。这一变换的数学表达为

$$\tilde{\boldsymbol{m}}=\mathcal{R}(\boldsymbol{m}):\mathbf{R}^{C\times H\times W}\rightarrow\mathbf{R}^{HW\times C} \tag{9}$$

式中$\mathcal{R}$表示形状重塑操作，将输入的特征图$\boldsymbol{m}$重排为新的矩阵$\tilde{\boldsymbol{m}}$。

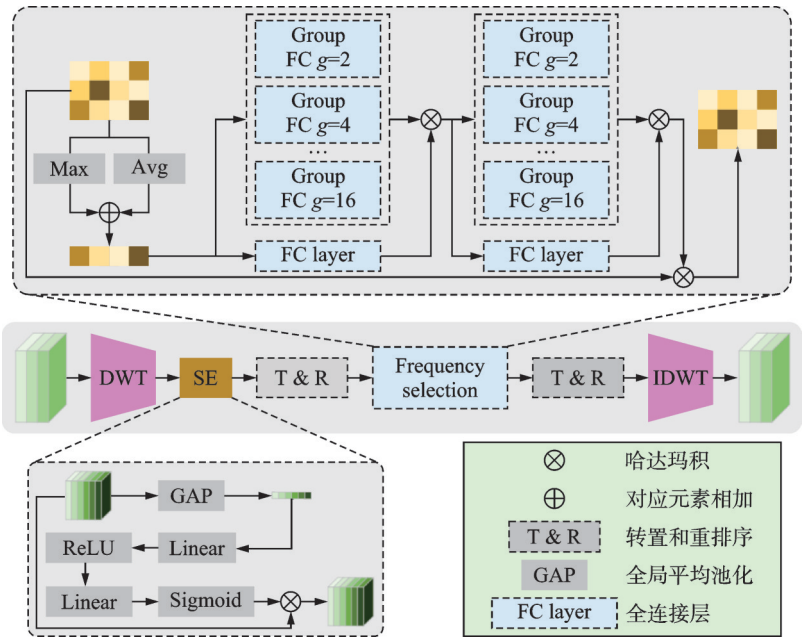


图3 特征去噪模块WaveDeno的流程图

Fig.3 Flow chart of the proposed feature denoising module WaveDeno

频率选择模块类似于一个注意力模块，即通过对全局信息的融合，对输入的频域特征进行调制。对于输入的频域特征$\boldsymbol{m}\in\mathbf{R}^{C\times H\times W}$，沿通道维度对其进行全局平均池化和全局最大池化操作，即

$$\tilde{S}_f=\mathcal{R}(\mathrm{GAP}(\boldsymbol{m})+\mathrm{GMP}(\boldsymbol{m})) \tag{10}$$

式中：GAP($\boldsymbol{m}$), GMP($\boldsymbol{m}$)$\in\mathbf{R}^{1\times H\times W}$，这一操作生成了频谱特征的统计。为了动态地计算阈值$\tilde{\Theta}$，采用将频域特征输入到分组全连接层（Group fully connection layer，Group FC）中。这一过程表达为

$$\tilde{\Theta}(\tilde{\boldsymbol{m}})=\left(1-\sigma\left(\mathcal{D}_{g'}\left(\delta\left(\mathcal{D}_{g}(\tilde{\mathrm{S}}_{f})\right)\right)\right)\right)\odot\tilde{\boldsymbol{m}} \tag{11}$$

式中:$\tilde{\boldsymbol{m}}$表示输入的经过重排后的特征向量;$\mathcal{D}_g$为分组数量为$g$的分组全连接层,在实际使用中通常需要兼顾计算量和模型性能,因此默认采用2、4、8、16四种分组数量;$\delta$代表ReLU激活函数;$\mathcal{D}_{g'}$为第二组分组全连接层,分组情况与第一组分组全连接层一致;$\sigma$代表Sigmoid激活函数。

这样的处理使得网络可以根据输入的数据对其中的低强度组分(噪声)进行衰减,同时保留高强度组分(目标)的特征信息,达到频域去噪与语义提纯的目的。

最后,所指网络对频域特征图的去噪和恢复过程可以表达如下

$$M'=\tilde{\mathcal{W}}\left(\tilde{\mathcal{R}}\left(\mathrm{FS}(\tilde{\boldsymbol{m}},\tilde{\Theta}(\tilde{\boldsymbol{m}}))\right)\right) \tag{12}$$

式中:$M'$表示经过WaveDeno去噪后的输出特征图,$\tilde{\mathcal{R}}$表示重排操作的逆操作,$\tilde{\mathcal{W}}$表示二维小波逆变换。

## 2 实验

### 2.1 数据集及实验环境

为了验证所提算法的可靠性及稳定性,选用MSAR[10]和SARDet-100K数据集[11]进行实验。MSAR数据集图像来自中国商业SAR卫星海丝一号(HISEA-1),包含28 449张尺寸为256×256像素的SAR图像,由4个类别共计60 396个实例组成,包含39 858个船舶目标、1 851个桥梁目标、12 319个油罐目标和6 368个飞机目标,选取19 914张图像作为训练集,8 535张图像作为测试集,且确保两个集合中不包含重复的图像;SARDet-100K数据集来自于10个公开的高质量数据集,包含来自中国的科研部门、欧洲的航天部门和美国的军事部门发布的数据,由6个类别共计245 653个实例组成,选取94 493张图像作为训练集,10 492张图像作为验证集,11 613张图像作为测试集,且确保3个数据集合中不存在图片重复的情况。

实验均采用Windows11操作系统,CPU为Intel(R) Core(TM) i7-12700K,GPU为NVIDIA GeForce RTX4090 24 GB,采用的深度学习框架为PyTorch 2.0.1 + CUDA 11.8。为了确保评估的一致性和公平性,所有算法的输入图像均经过标准化的数据集预处理算法处理。对于MSAR数据集,对所有算法均采用DAdapt-Adam优化器进行训练,训练轮次(epoch)设置为50,训练批量(batch-size)大小设置为64,初始学习率设置为1.0,权重衰减设置为0.05;对于SARDet-100K数据集,训练轮次设置为30,训练批量大小设置为16,其余设置与使用MSAR数据集进行训练时相同。

### 2.2 评价指标

在目标检测领域中,常用于衡量算法性能评价指标包括精度(Precision)和召回率(Recall)[28]。两者的计算基于交并比(Intersection over union, IoU)进行,IoU表示预测框与真实框的重叠度即预测结果与真实标注的匹配程度,其表达式为

$$\mathrm{IoU}=\frac{\sum(\mathrm{Mask1}\ \&\ \mathrm{Mask2})}{\sum(\mathrm{Mask1}\mid\mathrm{Mask2})} \tag{13}$$

由式(13)可知,IoU的取值范围处于0到1之间,取值为0表示完全不重叠,取值为1表示完全重叠。通过将其作为超参数在训练和测试时设置一个固定阈值作为模型预测结果的判别标准。

如果计算所得IoU大于设定的置信度阈值,则认为对应的预测框是真阳性(True positive, TP)的;如果计算所得IoU小于设定的置信度阈值,则认为对应的预测框是假阳性(False positive, FP)的;如果

在真实数据集中没有标记的框但测试时出现了预测框,那么认为它是真阴性(True negative, TN)的;如果在真实数据集中带有标记的框没有与其匹配的预测框,那么认为它是假阴性(False negative, FN)的。根据以上定义,精度可以表达为

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{14}$$

式(14)表达了模型预测为真的所有样本中,与标签匹配的样本的占比,即已检测到的目标的查准率。

召回率的表达式为

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{15}$$

式(15)的含义为在所有为真的样本中,被正确预测出的数量占比,即待检测目标的查全率。

为了综合考量精度和召回率,采用平均精度(Average precision, AP)[28]作为评价指标,为了保证评价的可靠性,采用了AP′07、AP′12、AP@50、AP@75四种评价指标。计算方法描述如下:通过将在设定的阈值下得到的Recall作为横坐标,Precision作为纵坐标,得到Precision-Recall(*PR*)曲线,AP′07采用在横坐标上等间距选取11个Recall值[0.0,0.1,0.2,…,0.9,1.0],将这11个值对应的Precision值取平均作为预测精度,表达式[28]为

$$\text{AP}'07=\frac{1}{11}\times(\text{AP}_\text{r}(0)+\text{AP}_\text{r}(0.1)+\text{AP}_\text{r}(0.2)+\cdots+\text{AP}_\text{r}(1.0)) \tag{16}$$

随后,研究人员开发了基于经过插值处理后的*PR*曲线与*X*轴的包络面积来衡量精度与召回率的综合性能,即AP′12,表达式[28]为

$$\text{AP}'12=\sum_{i=1}^{n-1}(r_{i+1}-r_i)\,p_\text{interp}(r_{i+1}) \tag{17}$$

式中:$r_i$表示第$i$个Recall值,$p_\text{interp}(r_i)$表示$r_i$处对应的Precision值。当$n$趋近于无穷大时即表示包络曲线的积分值,即

$$\text{AP}'12=\int_0^1\text{Precision}(\text{Recall})\,\text{d}(\text{Recall}) \tag{18}$$

在获取数据集中每个类别AP值后,取其平均值即可得到模型在所有类别目标检测上的综合精度,即

$$\text{mAP}=\frac{1}{n}\sum_{i=0}^{n}\text{AP}_i \tag{19}$$

式中:$n$为数据集中目标的种类数量,$\text{AP}_i$表示第$i$个类别的平均精度值。

在目标检测领域中,跟随Lin等[29]的研究通常需要设置IoU阈值定义检测精度的评价指标,当IoU取值为0.5时所得到的平均精度即为“AP@50”,当IoU取值为0.75时所得到的平均精度即为“AP@75”。

### 2.3 对比实验及可视化分析

将设计的算法与当前较为先进的算法在MSAR数据集上进行对比实验,对比算法包含Faster R-CNN[12]、ConvNeXt[13]、PVT-T[14]和YOLOF[15]。对于所提算法以及对比算法在训练和测试时的输入图像,均先采用归一化预处理以保证训练过程中的稳定性以及模型在SAR目标检测任务中的普适性。实验结果如表1所示,所提SAR图像目标检测算法性能超越了4种近年来提出的性能较为优秀的算法。在以AP′07作为评价指标时,mAP达到了70.02%,相较于另外4种算法分别提升了12.95%、13.67%、34.63%和21.90%;在以AP′12作为评价指标时,mAP达到了71.18%,相较于另外4种算法分别提升了15.52%、13.42%、42.06%和25.08%。且对于数据集中不同尺度的目标均实现了最优或次优的检测精度,验证了模型的鲁棒性。

**表1 在MSAR数据集上的对比实验结果**

**Table 1 Comparative experimental results on the MSAR dataset** %

| 算法 | Average precision (AP′07) | | | | | Average precision (AP′12) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | mAP | 船只 | 油罐 | 桥梁 | 飞机 | mAP | 船只 | 油罐 | 桥梁 | 飞机 |
| Faster R-CNN | 57.07 | 84.23 | **66.82** | 76.61 | 0.61 | 55.66 | 83.83 | 66.72 | 71.83 | 0.26 |
| ConvNeXt | 56.35 | 81.18 | 63.60 | 71.53 | 9.09 | 57.76 | 86.02 | 66.89 | 76.58 | 1.55 |
| PVT-T | 35.39 | 60.95 | 49.73 | 30.44 | 0.43 | 29.17 | 56.76 | 33.96 | 25.47 | 0.50 |
| YOLOF | 48.12 | 69.97 | 50.56 | 62.87 | 9.09 | 46.10 | 72.97 | 49.86 | 61.29 | 0.28 |
| Proposed | **70.02** | **89.44** | 63.40 | **78.99** | **48.25** | **71.18** | **90.75** | **68.43** | **79.25** | **46.30** |

为了更直观地对比所提模型在目标检测时的效果,从MSAR数据集的测试结果中随机选取了如图4所示的目标类型,分别为“飞机”、“船只”、“油罐”和“桥梁”的4组检测结果进行可视化对比,其中,蓝色框代表目标被识别为飞机,紫色框代表目标被识别为船只,绿色框代表目标被识别为油罐,红色框代表目标被识别为桥梁。可以看出,所提算法在强干扰、弱目标的情况下仍具备一定的检测和定位精度,相较于其他算法,错检和漏检情况得到了明显改善。

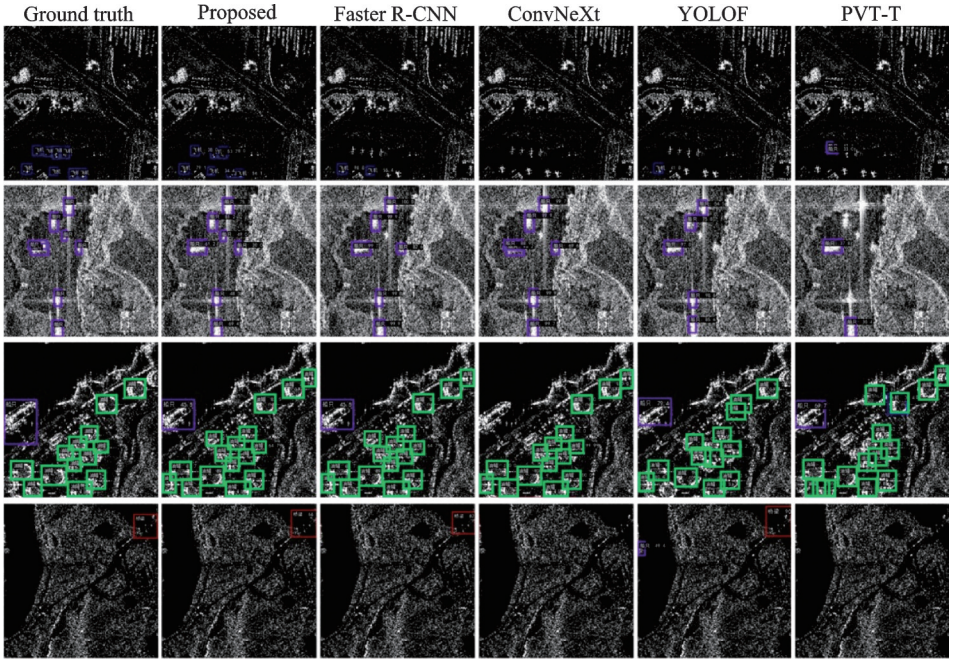


图4 在MSAR数据集上的可视化对比实验结果

Fig.4 Visualization results of comparative experiments on the MSAR dataset

为了增强实验的可信度并验证模型的泛化能力,增加了在SARDet-100K数据集上的对比实验,实验结果如表2所示,所提目标检测算法性能超越了4种对比算法。在以AP@50作为评价指标时,AP达到了85.50%,相较于另外4种算法分别提升了5.90%、0.10%、4.70%和12.60%;在以AP@75作为评价指标时,平均精度达到了60.10%,相较于另外4种算法分别提升了12.50%、1.40%、8.20%和21.90%;在针对各类目标检测的mAP达到了55.90%,相较于另外4种算法分别提升了10.00%、2.70%、7.20%和16.70%。且对于数据集中不同尺度的目标仍达到了最优或次优的检测精度,进一步证明了所提模型的稳定性和可靠性。

**表2 在SARDet-100K数据集上的对比实验结果**

**Table 2 Comparative experimental results on the SARDet-100K dataset** %

| 算法 | mAP | AP@50 | AP@75 | 船只 | 飞机 | 汽车 | 油罐 | 桥梁 | 港口 |
|---|---|---|---|---|---|---|---|---|---|
| Faster R-CNN | 45.90 | 79.60 | 47.60 | 54.42 | 51.46 | 57.80 | 28.83 | 32.70 | 50.19 |
| ConvNeXt | 53.20 | 85.40 | 58.30 | 60.51 | 57.39 | 62.23 | 38.02 | **36.51** | 64.54 |
| PVT-T | 48.70 | 80.80 | 53.94 | 56.15 | 57.06 | 58.63 | 34.20 | 25.11 | 61.05 |
| YOLOF | 39.20 | 72.90 | 38.20 | 50.65 | 49.32 | 49.66 | 19.83 | 20.94 | 44.80 |
| Proposed | **55.90** | **85.50** | **60.10** | **64.90** | **57.41** | **63.46** | **45.77** | 36.41 | **67.45** |

同样，从SARDet-100K数据集的测试结果中随机选取了如图5所示的目标类型，分别为“油罐”、“飞机”、“桥梁”、“船只”、“港口”和“汽车”的6组检测结果进行可视化对比，其中，绿色框代表目标被识别为油罐，深蓝色框代表被识别为飞机，浅蓝色框代表被识别为港口，黄色框代表被识别为桥梁，红色框代表被识别为船只，紫色框代表被识别为汽车。可以看出，所提算法在更大场景下面对强背景干扰、弱小目标的挑战下仍展现出相当的检测和定位精度，相较于其他算法具有更加稳定的表现，但在面临复杂情况时仍然存在错检和漏检的问题，这是后续工作中需要进一步研究并改进的方向。

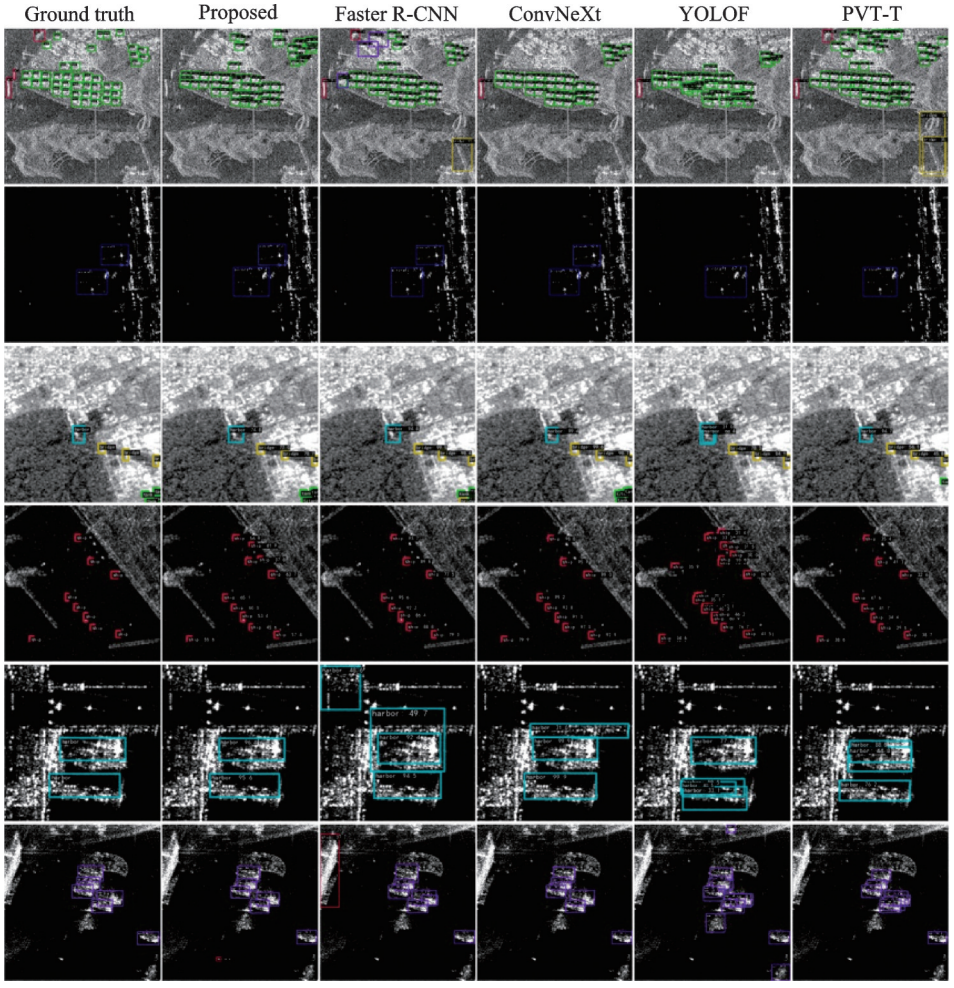


图5 在SARDet-100K数据集上的对比实验结果

Fig.5 Visualization results of comparative experiments on the SARDet-100K dataset

### 2.4 消融实验

为了验证所提算法中的特征去噪模块和多尺度空间-通道联合特征增强模块的有效性,通过逐步将所提出的模块替换为卷积模块进行消融实验。实验采用的网络具体架构如表3所示,其中Case 1表示不去除所提出的模块即本文所提出的初始网络,Case 2表示去除特征去噪模块并保留特征增强模块所设计的网络,Case 3表示去除本文中所提出的所有模块。采用MSAR数据集作为实验数据集,评价指标采用AP′07和AP′12,实验结果如表4所示。其中,加粗标注的结果表示最佳结果,下划线标注的结果表示次优的结果。可以看出,保留特征去噪模块和特征增强模块的网络相较于去除特征去噪模块的网络和去除所有所提出模块的网络,以AP′07作为评价指标时,mAP分别提升了2.32%和4.2%,同时4类目标类型的检测精度分别提升了0.33%、0.13%、4.51%、4.3%和0.23%、0.12%、4.53%、10.94%;在以AP′12作为评价指标时,mAP分别提升了1.57%和4.2%,同时4类目标类型的检测精度分别提升了0.33%、0.13%、4.51%、4.3%和0.23%、0.12%、4.53%、10.94%。以上结果证明了特征去噪模块和特征增强模块在提升SAR目标检测精度上的有效性。

**表3 消融实验策略**

**Table 3 The strategy of ablation studies**

| 网络 | WaveDeno | MSFEM |
|---|---|---|
| Case 1 | √ | √ |
| Case 2 | × | √ |
| Case 3 | × | × |

**表4 在MSAR数据集上的消融实验结果**

**Table 4 Ablation experimental results on the MSAR dataset** %

| 网络 | Average precision (AP′07) | | | | | Average precision (AP′12) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | mAP | 船只 | 油罐 | 桥梁 | 飞机 | mAP | 船只 | 油罐 | 桥梁 | 飞机 |
| Case 1 | **70.02** | **89.44** | **63.40** | **78.99** | **48.25** | **71.18** | **90.75** | **68.43** | **79.25** | **46.30** |
| Case 2 | <u>67.70</u> | 89.11 | 63.27 | <u>74.48</u> | 43.95 | 69.61 | 90.33 | 68.19 | 77.14 | 42.78 |
| Case 3 | 65.82 | <u>89.21</u> | <u>63.28</u> | 73.46 | 37.31 | 67.74 | 90.06 | 68.14 | 76.66 | 36.08 |

## 3 结束语

为充分挖掘SAR图像中目标的多尺度结构特性,并有效抑制复杂相干斑噪声干扰,提出一种融合空间-通道特征与频率选择机制的SAR目标检测算法。该方法从多尺度特征建模与频域去噪两个维度出发,系统提升了检测性能与鲁棒性。具体而言,设计了一个多尺度空间-通道联合特征增强模块,借鉴Transformer的建模思想,引入轻量化卷积以兼顾全局与局部信息建模能力。此外,针对SAR图像中普遍存在的相干斑噪声问题,提出了基于频域映射的特征去噪模块,通过频率选择机制抑制冗余低信噪比成分,增强目标特征的显著性与可分辨性。

在两个公开SAR图像目标检测数据集上的实验结果表明,所提方法在检测精度与鲁棒性方面均达到当前先进水平,展现出良好的适应性与泛化能力。尤其在大场景复杂背景下的小目标检测任务中,表现出更高的检测精度与召回率,具有较强的实际应用潜力。然而,在背景纹理极其复杂或目标尺度极小的场景中,仍存在一定的误检与漏检问题。未来将从动态特征建模、跨尺度上下文融合及小样本增强学习等方面进一步优化,以提升复杂场景下的检测稳定性与准确性。

**参考文献：**


[1] TIRANDAZ Z, AKBARIZADEH G, KAABI H. PolSAR image segmentation based on feature extraction and data compression using weighted neighborhood filter bank and hidden Markov random field-expectation maximization[J]. Measurement, 2020, 153: 107432.

[2] ZALPOUR M, AKBARIZADEH G, ALAEI-SHEINI N. A new approach for oil tank detection using deep learning features with control false alarm rate in high-resolution satellite imagery[J]. International Journal of Remote Sensing, 2020, 41(6): 2239-2262.

[3] 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 &. Astronautics, 2024, 41(6): 725-738.

[4] LIANG Liming, CHEN Kangquan, WANG Chengbin, et al. Remote sensing image detection algorithm integrating visual center mechanism and parallel patch perception[J]. Opto-Electronic Engineering, 2024, 51(7): 1-12.

[5] XIAO Zhenjiu, ZHANG Jiehao, LIN Bohan. Feature coordination and fine-grained perception of small targets in remote sensing images[J]. Opto-Electronic Engineering, 2024, 51(6): 1-13.

[6] 宦若虹, 陶一凡, 陈月, 等. 多线性主成分分析和张量分析的SAR图像目标识别[J]. 数据采集与处理, 2018, 33(5): 872-879.
HUAN Ruohong, TAO Yifan, CHEN Yue, et al. SAR image target recognition based on multi-linear principal component analysis and tensor analysis[J]. Journal of Data Acquisition and Processing, 2018, 33(5): 872-879.

[7] YUE D X, XU F, FRERY A C, et al. Synthetic aperture radar image statistical modeling: Part one-single-pixel statistical models[J]. IEEE Geoscience and Remote Sensing Magazine, 2020, 9(1): 82-114.

[8] WAN Huiyao, CHEN Jie, HUANG Zhixiang, et al. AFSar: An anchor-free SAR target detection algorithm based on multiscale enhancement representation learning[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 5219514.

[9] 徐丹青, 吴一全. 光学遥感图像目标检测的深度学习算法研究进展[J]. 遥感学报, 2024, 28(12): 3045-3073.
XU Danqing, WU Yiquan. Progress of research on deep learning algorithms for object detection in optical remote sensing images [J]. National Remote Sensing Bulletin, 2024, 28(12): 3045-3073.

[10] 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]. 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]. Journal of Radars, 2022, 14: 1488.

[11] LI Yuxuan, LI Xiang, LI Weijie, et al. SARDet-100K: Towards open-source benchmark and toolKit for large-scale SAR object detection[J]. arXiv preprint arXiv:2403.06534, 2024.

[12] REN Shaoqing, HE Kaiming, 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, 2017, 39(6): 1137-1149.

[13] LIU Zhuang, MAO Hanzi, WU Chaoyuan, et al. A ConvNet for the 2020s[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2022: 11976-11986.

[14] 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.

[15] 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.

[16] AI Jiaqiu, MAO Yuxiang, LUO Qiwu, et al. SAR target classification using the multikernel-size feature fusion-based convolutional neural network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 5214313.

[17] SI Jihao, SONG Binbin, WU Jixuan, et al. Maritime ship detection method for satellite images based on multiscale feature fusion[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, 16: 6642-6655.

[18] LI Ke, WANG Di, HU Zhangyuan, et al. Unleashing channel potential: Space-frequency selection convolution for SAR object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2024: 17323-17332.

[19] 肖振久, 林渤翰, 曲海成. 融合多重机制的SAR舰船检测[J]. 中国图象图形学报, 2024, 29 (2): 545-558.
XIAO Zhenjiu, LIN Bohan, QU Haicheng. SAR ship detection with multi-mechanism fusion[J]. Journal of Image and Graphics,

2024, 29(2): 545-558.

[20] ZHOU Jie, XIAO Chao, PENG Bo, et al. DiffDet4SAR: Diffusion-based aircraft target detection network for SAR images[J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21: 4007905.

[21] ZHANG Linping, LIU Yu, ZHAO Wenda, et al. Frequency-adaptive learning for SAR ship detection in clutter scenes[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5215514.

[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] LIN T G, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2017: 2980-2988.

[24] YU Jiahui, JIANG Yuning, WANG Zhangyang, et al. Unitbox: An advanced object detection network[C]//Proceedings of the ACM International Conference Multimedia. [S.l.]: ACM, 2016: 516-520.

[25] NAH S, HYUN K T, MU L K. Deep multi-scale convolutional neural network for dynamic scene deblurring[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2017: 3883-3891.

[26] TU Z, TALEBI H, ZHANG H, et al. MAXIM: Multi-axis MLP for image processing[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2022: 5769-5780.

[27] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 7132-7141.

[28] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88: 303-338.

[29] LIN T G, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context[C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2014: 740-755.

**作者简介:**

**纪晓平**(1982-),男,高级工程师,研究方向:雷达系统设计,E-mail: 52625271@qq.com。

**陶普**(2000-),**通信作者**,男,硕士研究生,研究方向:遥感图像处理,机器视觉,E-mail: tao_pu@nuaa.edu.cn。

(编辑:夏道家)

# SAR Target Detection via Joint Spatial-Channel Feature and Frequency Selection

JI Xiaoping[1], TAO Pu[2*]

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

**Abstract**: Synthetic aperture radar (SAR) imagery is characterized by a large number of targets with diverse categories and significant scale variations, as well as highly complex background clutter caused by coherent speckle noise. These inherent properties substantially degrade detection accuracy and pose significant challenges to reliable target detection. To address the problem of insufficient detection performance under such conditions, this paper proposes a SAR target detection algorithm that jointly exploits spatial-channel feature fusion and frequency selection. Specifically, a ResNet-50 network pre-trained on large-scale datasets is adopted as the backbone to extract hierarchical and multi-scale feature representations from SAR images. On this basis, a feature pyramid network (FPN) augmented with a joint multi-scale spatia-channel feature enhancement module is constructed to strengthen the representation capability of features at different scales. This design enables the network to more effectively capture discriminative target information while alleviating the adverse impact of scale diversity among targets. By jointly modeling spatial and channel-wise dependencies, the proposed enhancement module improves feature expressiveness and robustness, particularly for small and weak targets embedded in cluttered backgrounds. Furthermore, a frequency selection module is introduced in the feature domain to explicitly exploit the frequency characteristics of SAR imagery. This module selectively suppresses noise components while preserving informative target-related signals, thereby enhancing target features and improving the signal-to-noise ratio. Through adaptive frequency-domain filtering, the proposed method effectively mitigates the influence of speckle noise without sacrificing critical structural information, leading to more reliable feature representations for subsequent detection. Extensive comparative experiments are conducted on two widely used benchmark datasets, MSAR and SARDet-100K, to evaluate the effectiveness of the proposed approach. Experimental results demonstrate that the proposed algorithm consistently outperforms several representative and state-of-the-art SAR image target detection methods, including Faster R-CNN, ConvNeXt, PVT-T, and YOLOF, across both datasets. These results indicate that the proposed framework achieves superior detection performance and exhibits strong generalization capability under complex SAR imaging conditions. Overall, the proposed method provides an effective solution for improving SAR target detection accuracy in scenarios involving complex backgrounds, severe speckle noise, and multi-scale target distributions.

**Highlights:**

1. A novel SAR target detection framework is proposed that jointly integrates multi-scale spatial-channel feature enhancement with frequency-aware feature modeling, enabling robust target representation under severe scale variation and speckle noise.
2. A feature-domain frequency selection mechanism is developed to explicitly exploit the spectral characteristics of SAR features, selectively suppressing noise components while preserving discriminative target information.

**Key words:** SAR target detection; multi-scale feature; spatial-channel feature; frequency selection; feature denoising

**Received:** 2025-08-05; **Revised:** 2025-09-17

***Corresponding author, E-mail:** tao_pu@nuaa.edu.cn.