文章编号:1001-9014(XXXX)XX-0001-17

DOI:10. 11972/j. issn. 1001-9014. JIMW. 2026139

结合小波感知降采样与多频交叉注意力的红外小目标检测网络

张 磊[1], 刘永贤[2], 宸书琦[2], 陈金宝[1], 郭雅萍[3], 李博扬[2*], 林再平[2], 安 玮[2]

(1. 南京航空航天大学大学 航天学院,南京 210016;
2. 国防科技大学 电子科学学院,长沙 410073;
3. 装备项目管理中心,北京 100064)

摘要:红外小目标检测是国防安全领域的关键技术,同时面临复杂真实场景带来的严峻挑战。红外小目标常因像素占比极低、复杂背景干扰而呈现出形态模糊、信杂比低等检测难点。近年来,基于深度卷积网络的检测方法取得了可观发展,但现有方法多聚焦于空域特征建模,忽视了频域特征的价值以及池化造成的信息损失。为解决上述问题,本文从频域视角开展研究并提出一种结合小波感知降采样与多频交叉注意力的红外小目标检测网络(WDFA-Net)。该网络设计小波感知降采样模块(WAD)替代传统池化,通过空频域双分支协同采样,在降维同时保持特征细节并强化目标响应。此外,在特征融合阶段设计了多频交叉注意力模块(MFCA)。该模块基于多频分量构建具有物理意义的自注意力机制,充分利用高低频特征互补信息,实现跨层特征交互融合与噪声抑制。实验结果表明,WDFA-Net 在 IRSTD-1k、SIRST 和 SIRST-UAVB 数据集上均取得了优异检测性能,具备复杂真实场景下的鲁棒检测能力.

关 键 词:红外小目标检测; 小波感知降采样; 多频交叉注意力; 频域分析; 深度学习

中图分类号:TP753 **文献标识码**: A

刘永贤为本文共同一作

引言

红外成像系统凭借其全天候作业、探测距离远、热源捕获能力强等独特优势[1],现已广泛应用于天基光学监视[2]、高机动目标突防与拦截[3]、无人机探测与侦察[4]、海上舰船目标识别[5]、城市资源规划[6]等军民两用场景,逐步成为各类智能探测体系中至关重要的核心组成部分[7-8],具有极高应用价值[9-10]。远距离成像场景下,红外小目标通常表现为形态模糊、无明确轮廓的点状光斑,存在信杂比、信噪比低的问题。复杂动态环境下,这些固有特性极大提高了红外小目标的检测难度,尤其当目标被强背景杂波淹没并呈现较弱的空间特征时,精准定位识别更为困难。因此,开发鲁棒、泛化的检测方法实现在真实复杂场景中准确定位红外小目标,仍然是本领域内一个极具挑战的难题。

传统的红外小目标检测算法采用模型驱动范式。基于滤波的方法[11]依据整张红外图像的不连续性来抑制背景,从而实现目标检出,主要包括空间域滤波和频域滤波两大类。之后,Chen 等人[12]依据人类视觉系统关注区域显著性的特点提出了基于局部对比度的检测方法。该类方法能够关注图像中的显著性区域并进行目标提取,经典研究包括基于空域局部对比度的方法[13-14]和基于频域谱残差的方法[15]等。接着,基于低秩稀疏分解的方法[16]利用背景的非局部自相关特性和小目标的稀疏特征,将检测问题转化为低秩背景和稀疏目标的恢复问题。尽管取得了不错的检测效果,传统的红外弱小目标检测方法严重依赖于先验知识和超参数调优,限制了其对复杂场景的适应性。

近年来,红外小目标检测领域的研究重点转移到基于深度学习的方法,凭借其强大的建模表征能力提升检测[17-18]、分割[19-20]性能。基于深度学习的检测方法采用数据驱动范式,通过搭建类脑化的深度神经网络架构,采用梯度驱动的反向传播策略来驱使模型学习目标表征,从而实现对多类型场景的鲁棒检测[21]。卷积神经网络是被广泛研究的主流方法。例如,Dai 等人[22]设计一个自顶向下和自底

收稿日期:2026-05-22,**Received date**:2026-05-22,
基金项目:国家自然科学基金(62501618)
Foundation items:Supported by the National Natural Science Foundation of China(62501618)
*通讯作者(Corresponding author):李博扬,男,河南三门峡人,助理研究员,博士。E-mail: liboyang20@nudt. edu. cn。

向上的双向通道注意力结构，交互融合浅层网络提取的轮廓特征和深层网络提取到的语义特征。后来，Dai 等人[23]在前述工作的基础上引入循环位移模块实现了长距离空域上下文信息关联，进一步提升了目标检测的性能。受到 U 型网络[24]启发，Li 等人[25]设计密集嵌套结构来解决池化层引起的小目标信息丢失问题。进一步，Wu 等人[26]提出嵌套式 U-Net 架构，强化多尺度特征学习能力。Liu 等人[27]构建了循环迭代结构，通过反复精炼目标特征实现高效检测。此外，Liu 等人[28]为普通 U 型网络设计了多尺度检测头，来提高定位精度。为了克服卷积神经网络感受野受限的缺陷，Liu 等人[29]提出了第一个基于 UNet-Transformer 混合结构的红外弱小目标检测方法，利用自注意力机制使网络具备更强的长距信息关联能力。针对天基红外舰船检测问题，Wu 等人[5]设计了多级的视觉 Transformer 模块。之后，为了感知全局上下文，Yuan 等人[30]在跳跃连接中设计了空间通道交叉 Transformer 模块，有效增强目标与背景的特征差异。

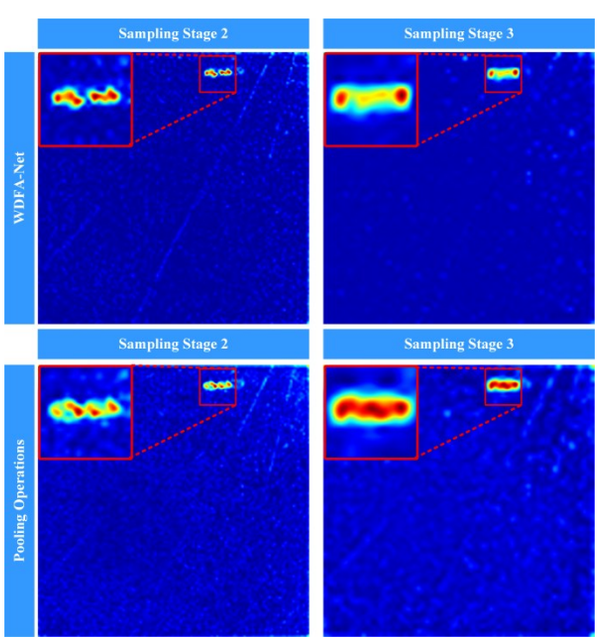

图1 网络层中红外小目标的特征可视化：(a)本文方法；(b)池化操作

Fig. 1 Feature visualization of infrared small targets in network layers: (a) The proposed method; (b) Pooling operation.

以上方法大多聚焦于挖掘图像的空间域特征，采用更强力的网络架构和注意力机制来提取语义信息，但未充分挖掘频域中蕴含的有效特征线索。许多研究指明频域处理技术对各类图像处理任务具有增益效果[31-35]。受该思路启发，Zhu 等人[36]设计了图像分块的频域增强注意力机制，有效增强目标显著性。Liu 等人[37]和 Deng 等人[38]分别设计了频率调制掩码注意力模块、分层门控内核注意力模块，在编解码器间深度融合全局空频域特征。然后，Li 等人[39]认为检测关键在于区分高频成分中的目标与干扰，提出了动态高频卷积算子，在保证计算效率的同时提高检测精度。Yuan 等人[40]则发掘低频特征引导的净化模块，有效抑制噪声干扰引发的虚警。

最近，基于表征学习的检测方法取得了研究进展。例如，Lin 等人[41]引入形状偏置先验，有效弥补红外小目标纹理信息不足的缺陷。接着，他们进一步提出基于对比度增强的形状偏置表征学习方法[42]，通过强化模型对目标形状特征的判别能力来提升检测性能。此外，各类新型神经网络范式也已经应用于红外小目标检测任务。比如，Zhang 等人[43]改进了通用的分割任意模型来解决低信杂比问题。Du 等人[44]设计了条件去噪生成模型迭代地生成目标掩码。而 Jia 等人[45]采用图神经网络架构提取图像结构信息来辅助检测。

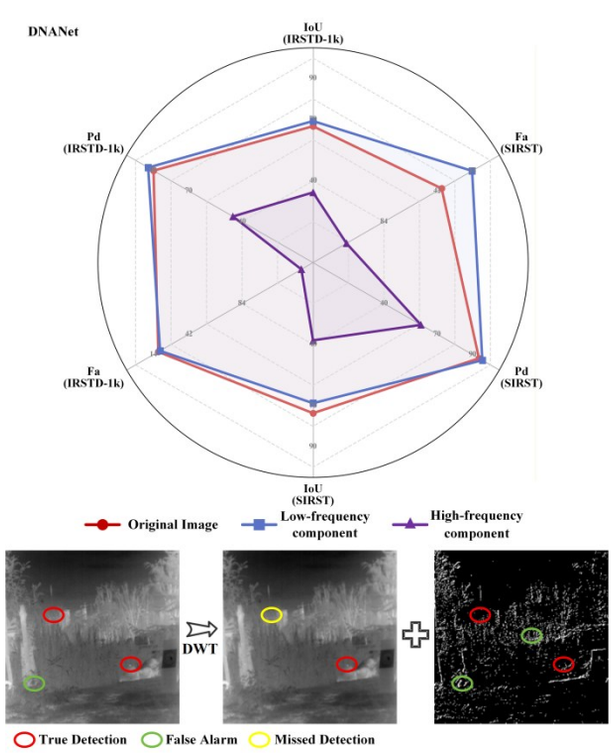

图2 DNANet对数据集不同频率分量的检测性能

Fig. 2 Detection performance of DNANet on different frequency components of datasets

然而,尽管上述基于深度学习的红外小目标检测方法已经取得了可观的性能,但仍面临着亟待解决的挑战。由于红外目标像素占比低,深层网络的空域池化下采样操作存在较为严重的信息损耗,极易导致弱小目标的语义特征混淆或重叠,甚至造成深层网络的响应丢失。如图 1 所示,池化操作致使邻近斑目标的语义特征发生混合畸变,难以实现精细区分。此外,当前方法侧重于搭建复杂的网络结构来增强特征表达能力,但轻视了与目标特性相似的噪声易于引发虚警的问题。如图 2 所示,借助离散小波变换将原始红外图像分解为低频分量和高频分量,并分别将三者用于 DNANet 网络[25]的训练和测试。实验结果表明,高频分量对准确检测至关重要,但其内部包含的干扰造成了虚警上升,这一结论也得到领域相关研究的证实[39]。另一方面,低频分量对提高目标定位能力和抑制虚警具有重要作用。低频分量在 SIRST 数据集上实现了最佳的虚警抑制性能,突显了其作为引导信息来抑制高频分量中噪声的应用潜力。现有的基于频域的红外小目标检测方法大多依托多尺度特征融合策略缓解小目标特征响应易衰减、丢失的问题,但它们未考虑从采样机制入手,解决下采样中固有信息损耗的根本问题。比如,FM-Net[37]设计多级的小波残差块,借助频域特征弥补主干网络确实的目标语义信息,却无法抵消池化操作带来的不可逆特征损失问题。另一方面,现有方法侧重对全频段频域信息进行一体化统一建模,未充分利用不同频域成分对检测能力的差异化贡献,缺乏针对性的频域增强和抑制机制。比如,GSFA-Net[38]采用分层门控内核注意力融合空间与频域全局上下文特征,忽略了高频特征中的背景噪声干扰。综上,图像结构中的不同频率分量具有差异化的功能特性,合理统筹高低频分量各自优势、规避其缺陷,能够为进一步提升红外小目标检测整体性能提供可行的研究思路。

受到以上存在问题和实验现象启发,本文利用离散小波变换充分挖掘多类频域分量特征来提升网络检测性能,提出了结合小波感知降采样与多频交叉注意力的红外小目标检测网络(WDFA-Net)。具体地,针对池化下采样造成信息损失、特征混叠的弊端,本文设计了小波感知降采样模块(WAD),采用空频域双分支结构协同保留细节信息,同步提取空频域互补特征并完成自适应增强。接着,为提高定位精度和降低虚警,本文设计了多频交叉注意力模块(MFCA),在特征融合中基于多类频域特征构建自注意力机制,有效推动相邻层级特征完成空频域信息深度交互。同时,该模块捕获全局上下文关联来建模并剔除背景干扰。本文方法设计的“采样-融合”双阶段协同处理是区别于现有频域检测方法的核心创新。综上,本文的主要创新和研究贡献总结如下:

(1)本文提出了检测网络 WDFA-Net,通过小波变换提取多类频域特征,有效拓展模型对红外小目标的特征表征能力。小波感知降采样可替代传统池化完成空间尺度压缩,在保留特征细节的同时,依托空频域信息实现特征增强与噪声抑制。

(2)特征融合阶段,本文提出了多频交叉注意力模块,促进层级特征交互融合与增强,并感知全局上下文关联来精准区分目标与背景语义,有效抑制干扰。

(3)在三个真实场景数据集(IRSTD-1k、SIRST、SIRST-UAVB)上开展全面的对比实验,结果验证了本文方法具备优异的检测性能。

1　本文方法

1.1　整体结构

图 3 展示了本文提出的 WDFA-Net 网络的整体架构,总体上由编码器、特征融合、解码器、损失计算组成。1.2 节详细介绍本文提出的小波感知降采样模块,1.3 节详细介绍本文提出的多频交叉注意力模块。首先,输入一张红外图像,依次经过小波残差块[37]和小波感知降采样组成的编码器来提取不同粒度的多尺度特征,其中小波残差块模块发挥了挖掘空频融合特征的作用,小波感知降采样实现维度减少的同时有效维持了目标细节特征。然后,编码特征送入分块嵌入模块进行压缩,邻层的特征输入多频交叉注意力模型获取全局结构和特征融合,最后特征重构并送入解码器逐步融合恢复分辨率,对阈值分割生成的像素级预测掩码进行聚类输出。损失计算采用深度检测策略来稳定训练,有效促使网络关注大小、形态变化的小目标。

预处理阶段首先对所有输入图像进行归一化、尺度变换、翻转等数据增强操作。然后,图像被送入编码器提取多尺度特征。WDFA-Net 采用嵌入小波残差块的 18 层 ResNet[46]作为主干网络,其中的下采样阶段使用本文提出的 WAD 模块。该主干网络提取到的不同尺寸的深层编码特征,表示如下:

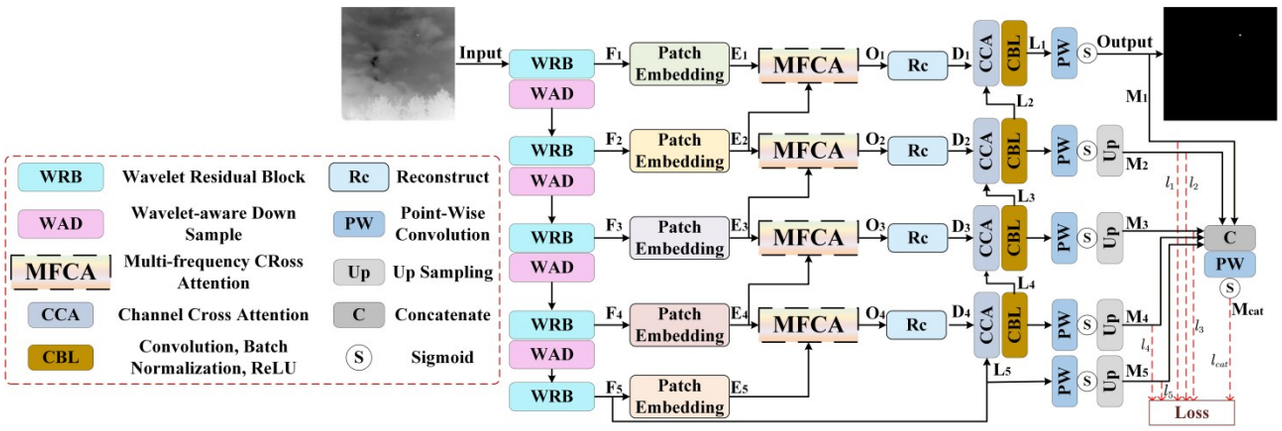

图3 结合小波感知降采样和多频交叉注意力的检测网络示意图

Fig. 3 An illustration of the detection network combining wavelet-aware down sample and multi-frequency cross attention

$$F_i \in \mathbb{R}^{C_i \times \frac{H}{2^{i-1}} \times \frac{W}{2^{i-1}}}, (i = 1,2,3,4,5). \tag{1}$$

其中，C_i是每层特征的通道数目，本文方法取$C_1 = 16, C_2 = 32, C_3 = 64, C_4 = 128, C_5 = 128$。特别地，编码器第五层获取的编码特征通道数为128。该设置兼顾了计算效率和特征提取需求，一方面能够避免因通道数目过大导致网络参数量激增的问题，另一方面，本文的通道设置对有效提取深层特征提供充足的支撑。接着，编码特征由分块嵌入模块进行处理并得到嵌入特征，表示如下：

$$E_i \in \mathbb{R}^{C_i \times \frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}}}. \tag{2}$$

本质上，通过步长等于卷积核子的卷积操作对特征进行图像分块嵌入处理。该处理是Transformer结构中的常规操作，利于降低计算量和提高训练速度。本文方法对编码特征进行8倍压缩来生成图像块嵌入。

然后，嵌入特征被输入多频交叉注意力模块，实现逐层地全局空频信息感知和细粒度特征融合。深层与浅层的嵌入特征被同时送入多频交叉注意力模块，利用小波变换来生成多种频率成分的自注意力，强化低频信息的虚警抑制作用和高频的目标响应增强作用。相应的输出O_i与同层的输入E_i在特征维度、尺寸上保持一致。同时，该阶段还加入了跳跃连接将同层的输入与输出进行相加，来维持多粒度的语义信息。具体的处理流程可以用以下公式表示：

$$O_j = MFCA(E_j + E_{j+1}), (j = 1,2,3,4). \tag{3}$$

其中，$MFCA(\cdot)$表示多频交叉注意力模块。

然后，每层中的融合输出特征被执行重构模块来恢复与编码特征相同的尺寸。重构模块采用的是双线性上采样操作。进一步，卷积块再被用于获取解码特征，表示如下：

$$D_j \in \mathbb{R}^{C_j \times \frac{H}{2^{j-1}} \times \frac{W}{2^{j-1}}}. \tag{4}$$

跳跃连接再次被用于融合编码器和解码器之间的特征，发挥稳定训练的作用。该阶段的处理可以用以下公式进行数学表示：

$$D_j = F_j + Rc_j(O_j). \tag{5}$$

最后，遵循经典范式的解码结构[47]，本文使用通道交叉注意力(Channel cross attention，CAA)[48]来辅助聚合深层和浅层特征。深层特征首先上采样来提高空间分辨率，接着CCA聚合解码特征和上采样的特征。然后，上采样特征与注意力增强后的特征沿通道维度拼接在一起再由两个卷积块进行融合解码。该处理过程可以数学表示为：

$$L_j = CBL([up(L_{j+1}), CCA(up(L_{j+1}), D_j)]). \tag{6}$$

其中，L_j表示第j层的融合输出特征。$up(\cdot)$表示上采样操作。$[\cdot,\cdot]$表示通道维的拼接操作。$CBL(\cdot)$块表示级联的卷积、批归一化和ReLU激活函数。特别说明的是，本文中$L_5 = F_5$。

为了辅助网络训练过程中高效收敛，本文方法的架构采用了多级深度监督策略[30]来进行损失函数计算，优化训练效率。具体地，解码器的输出L_j由逐点卷积进行聚合，使得通道数变为1；再由Sigmoid激活函数来生成预测结果M_i。该处理过程数学表示为：

$$M_i = S(PW(L_i)). \tag{7}$$

其中，$S(\cdot)$表示Sigmoid激活函数，$PW(\cdot)$表示逐点卷积。

然后，多尺度的预测结果被上采样至与原始图像相同的空间分辨率。这些结果再被拼接在一起并由逐点卷积聚合来获得联合预测结果。数学表

示如下:

$$M_{cat} = S(PW([up(M_1),up(M_2),up(M_3),up(M_4),up(M_5)])). \tag{8}$$

最终,本文方法分别对所有预测结果和标签真值计算损失函数。用于网络训练的综合损失为全部损失计算值的和。数学表示如下:

$$l_1 = loss(M_1,G). \tag{9}$$

$$l_i = loss(up(M_i),G), i = 2,3,4,5. \tag{10}$$

$$l_c = loss(M_{cat},G). \tag{11}$$

$$l_{all} = \sum_{i=1}^{5} l_i + l_c. \tag{12}$$

其中,$loss(\cdot)$表示损失函数,G表示标签真值。本文采用的损失函数为 DiceTopK 损失[49],结合了 Dice 损失[50]和 TopK 损失[51]。前者用于描述区域重叠程度,并且侧重关注目标区域和轮廓。后者是交叉熵损失[52]的变体之一,只使用预测最糟糕的 k%像素计算损失值,侧重关注困难样本。因此,复合损失函数 DiceTopK 兼顾了两种损失的优势,具备形状和位置敏感度,适用于小目标检测任务。

在阐述本文所设计的网络模块结构前,我们首先对后续方法中多次使用的离散小波变换(Discrete Wavelet Transform,DWT)进行原理介绍。针对二维图像信号,离散小波变换主要通过可分离的行、列逐级滤波与下采样操作实现多尺度特征分解。设输入图像为$I(m,n)$,DWT 先对图像进行行方向低通滤波、高通滤波及下采样,然后再进行列方向低通滤波、高通滤波及下采样,最终将原始图像分解为四个频率子带,分别为近似子带 LL、水平细节子带 LH、垂直细节子带 HL 和对角细节子带 HH。上述二维离散小波的分解过程可以表示为:

$$LL = \sum_k \sum_l h(k)h(l)I(2m+k,2n+l). \tag{13}$$

$$LH = \sum_k \sum_l h(k)g(l)I(2m+k,2n+l). \tag{14}$$

$$HL = \sum_k \sum_l g(k)h(l)I(2m+k,2n+l). \tag{15}$$

$$HH = \sum_k \sum_l g(k)g(l)I(2m+k,2n+l). \tag{16}$$

其中,$h(k)$和$g(k)$分别为行方向的低通和高通滤波器,$h(l)$和$g(l)$分别为列方向的低通和高通滤波器。四个子带具备差异化的图像特征表征能力,近似子带 LL 包含图像的主要低频结构信息,水平细节子带 LH 捕捉水平方向的边缘特征,垂直细节子带 HL 捕捉垂直方向的边缘特征,对角细节子带 HH 则包含对角方向的高频纹理信息。

本文采用 Haar 小波基完成图像分解,小波基的选择需综合考虑计算效率、边缘检测能力和信号表征效果。Haar 小波具有最短的支撑长度,可依托低计算复杂度实现快速分解,满足红外小目标检测任务的实时性需求。同时,Haar 小波的不连续表征特性与红外小目标的点状突变特征高度契合,能够有效捕捉目标边缘的灰度跳变信息,提高特征提取能力。相比之下,Daubechies 系列小波虽具有更好平滑性,但运算复杂度更高,且边缘响应敏感度偏低和非严格对称性影响目标定位精度。需要指出的是,Haar 小波处理红外图像平滑区域时可能产生高频干扰。为此,在小波分解后,本文利用低频特征作为引导信息来抑制高频分量中的噪声,减少虚警干扰,实现了计算效率与检测性能的有效平衡。

1.2　小波感知降采样

为了将小目标从复杂背景中分离,网络需要充分挖掘并利用语义特征中蕴含的全面信息。现有的基于卷积神经网络的检测方法为了获得较大的感受野来感知更大范围的上下文信息,通常在提取过程中会采用池化操作。由于红外目标本身尺寸小,呈现为空域特征较稀疏,大量的池化操作极易造成信息的损失,甚至区域丢失,对提高检测率造成不利影响。如图 1 所示,传统的最大池化下采样方式易造成空间邻近目标语义特征混淆和重叠的

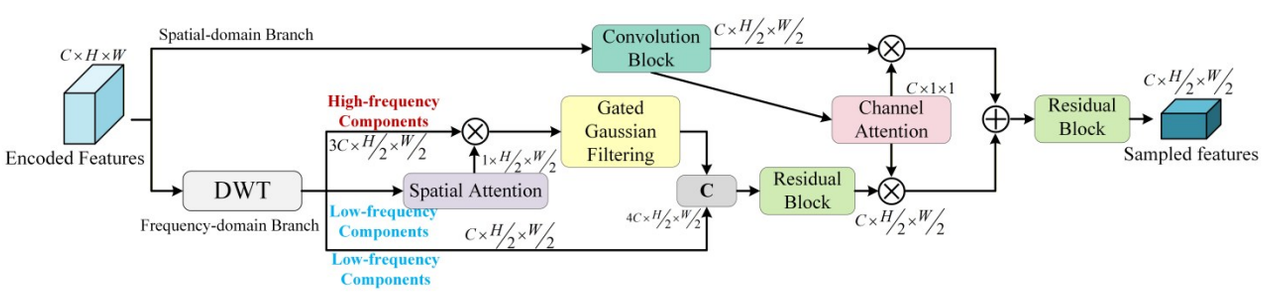

图 4　小波感知降采样示意图

Fig. 4　An illustration of the wavelet-aware down sample (WAD)

问题，还存在弱化响应强度和损失小目标细节信息的不良影响。为了解决以上问题，本文设计了小波感知降采样方法，依托小波变换的多尺度分析特性，将频域信息建模引入空间采样过程，实现空域和频域双分支特征采样的深度融合。该方法在完成特征尺度降维的同时，能够精准捕捉并保留具有判别性的关键细节特征。

图4展示了本文提出的小波感知降采样模块的结构细节。整体上，WAD模块由空域分支和频域分支两部分组成。空域分支上，先采用卷积块进行采样处理。该卷积块由卷积、批归一化和ReLU激活函数级联而成。卷积核为2，步长为2。该分支通过图像分块的方式实现了特征空间降维，可以有效保留局部对比度和空间互补性。该分支数学表示如下：

$$A_c = Ca(ConvB(X_i)). \tag{17}$$

$$Ca(\cdot) = S(PW(P_{\max}(\cdot)) + PW(P_{avg}(\cdot))). \tag{18}$$

$$F_s = ConvB(X_i) \otimes A_c. \tag{19}$$

其中，A_c表示空域分支特征生成的通道注意力。F_s表示空域分支的输出特征。$Ca(\cdot)$表示通道注意力，其融合全局平均和最大池化两路通道信息，经共享的逐点卷积层学习后相加，通过Sigmoid输出各通道重要性权重。$P_{\max}(\cdot)$表示全局最大池化。$P_{avg}(\cdot)$表示全局平均池化。$ConvB(\cdot)$表示卷积块。X_i表示输入的编码特征。$\otimes$表示相乘。

频域分支上，先通过离散小波变换对编码特征开展多尺度分解，将其分解为低频分量和高频分量。其中，低频分量对应近似子带(LL)，高频分量包含水平细节子带(LH)、垂直细节子带(HL)和对角细节子带(HH)。离散小波变换的分解方式具有完备的理论可解释性，整个分解过程能够实现信息无损保留，可由逆变换完成特征完整恢复。同时，其多尺度特性实现了分解后特征空间维度的自然降维。如图2分析所示，高频分量在红外小目标检测任务中具有重要作用，但其易引入虚警。而低频分量虽然会降低目标定位精度，却可以辅助抑制高频信息中的噪声干扰，利于降低误报概率。基于此，该模块的设计充分发挥了不同频率分量的固有优势。具体操作流程为：先使用低频分量生成空间注意力，并通过该注意力对高频分量进行加权。低频分量提供了潜在目标的位置信息，以此引导高频分量中目标相关特征的增强，同时抑制无关干扰信息。然后，通过门控高斯滤波对加权后的高频分量进行细化，进一步抑制并滤除其中的噪声特征，实现高频特征的净化。接着，净化后的高频分量和低频分量拼接在一起，再通过残差块完成两种分量融合。最后，融合后的特征经空域分支生成的通道注意力进行加权，从而强化对关键特征通道的关注。该分支数学表示如下：

$$[F_l, F_h] = DWT(X_i). \tag{20}$$

$$F_f = RB([F_l, GGF(F_h \otimes Sa(F_l))]) \otimes A_c. \tag{21}$$

$$Sa(\cdot) = S(Conv^{7\times7}([P_{\max}(\cdot), P_{avg}(\cdot)])). \tag{22}$$

其中，$DWT(\cdot)$表示离散小波变换，本文方法采用Haar小波，兼顾直观的局部对比度特性和计算效率。F_l表示低频分量。F_h表示高频分量。F_f表示频域分支的输出特征。$RB(\cdot)$表示残差块。$Sa(\cdot)$表示空间注意力，其聚合均值、最大值得到两张空间描述图并拼接，通过大核卷积融合局部空间上下文，经Sigmoid生成逐像素的空间重要性权重。$Conv^{7\times7}(\cdot)$表示核子为7×7卷积操作。$GGF(\cdot)$表示门控高斯滤波，实现自适应地对置信度低的特征施加更强的抑制。具体地，该设计只对绝对值低于阈值的高频分量应用高斯平滑。假设输入特征为X，数学表达式如下：

$$GGF(X) = \mathcal{G}(X) \cdot \psi_{<\tau}(|X|) + X \cdot \psi_{\geqslant\tau}(|X|). \tag{23}$$

其中，τ为阈值，本文设置为0.5。$\psi(\cdot)$表示用于施加门控约束的指示函数，当特征值满足预设阈值关系时取值为1。$\mathcal{G}(\cdot)$表示高斯滤波器，定义为：

$$\mathcal{G}(i,j;\sigma) = \frac{1}{Z}\exp\left(-\frac{(i-c)^2+(j-c)^2}{2\sigma^2}\right) \tag{24}$$

其中，σ表示可学习的标准差。c表示内核中心，$c = \lfloor k/2 \rfloor$。Z表示归一化因子。

空域分支和频域分支是严格的通道对齐。为了发挥空频双分支的特征互补优势，采用相加对两者进行交互，其输出再使用残差卷积块进行深度融合，最终得到采样特征。该过程数学表示如下：

$$F_d = RB(F_s + F_f). \tag{25}$$

其中，F_d表示输出的采样特征。

1.3 多频交叉注意力

有效的网络架构在构建全局上下文信息时，需综合考虑空域、通道与频域特征的协同效用，如典型结构代表FMM模块[35]。该方法侧重于对多尺度特征的空频语义信息进行同步建模，但未能充分考虑图2分析中所揭示的现象，不同频域成分对检测性能提升的贡献存在显著差异。为了充分挖掘多种频域信息的表征潜力来提升检测性能，本文方法

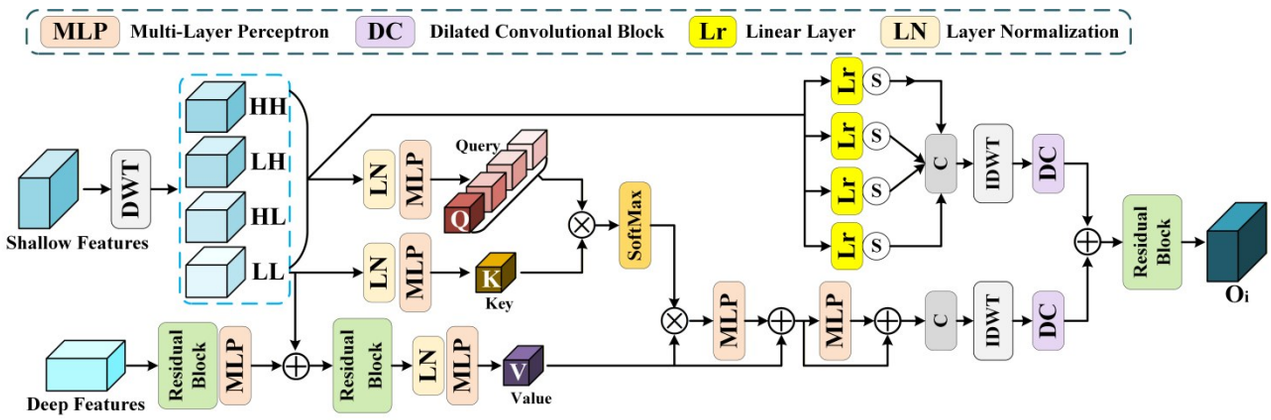

图5　多频交叉注意力示意图

Fig. 5　An illustration of the multi-frequency cross attention (MFCA)

设计了一个多频交叉注意力(Multi-Frequency Cross Attention, MFCA)模块,实现相邻层级编码特征的渐进式深度聚合。如图5所示,该模块采用并行结构同步处理深层与浅层特征,并依托两者构建自注意力机制完成特征增强与交互融合。

查询、键、值是经典自注意力机制[53]的三个核心组成要素。查询用于定位关键特征,键充当相关特征的索引,值是特征对应的具体内容。传统的自注意力仅局限于空域上构建三者的关系,忽略了挖掘低频与高频分量的特性优势。基于此,本文所提的 MFCA 模块构建了基于多频分量的差异化自注意力机制。具体地,该模块选取浅层特征的频域分量作为定位的查询特征,以全局空间结构特征作为索引键,同时将深层特征与空间结构信息融合后的特征作为具体内容值,完成跨层级、跨频域的注意力增强融合。

浅层特征通过离散小波变换分解为不同频率子带 LL、LH、HL 和 HH。该操作也使得浅层特征在空间维度上与深层特征对齐,其数学表达式如下:

$$[F_{LL},F_{LH},F_{HL},F_{HH}] = DWT(E_i). \tag{26}$$

其中,F_{LL}, F_{LH}, F_{HL}, F_{HH}分别表示近似子带、水平细节子带、垂直细节子带和对角细节子带。通过离散小波变换获取频率特征后,该方法直接对其进行层归一化处理,随后输入多层感知机来生成具有频域特性的查询。该处理过程表示如下:

$$Q_x = MLP(LN(F_x)). \tag{27}$$

其中,$LN(\cdot)$表示层归一化。为了方便,下标表示为 $x = LL, LH, HL, HH$。Q_x表示对应不同频率分量的查询。$MLP(\cdot)$表示多层感知机。

近似子带属于低频部分,包含了整体轮廓、灰度分布等全局结构特征。为了将整体空间特征作为注意力机制的键,该方法直接对近似子带进行层归一化处理,随后输入多层感知机来生成具有空间结构信息的键。该处理过程表示如下:

$$K = MLP(LN(F_{LL})). \tag{28}$$

为使值蕴含深浅层特征的完整细节信息,该方法构建了融合浅层特征低频分量与深层特征的聚合型值表征。首先,利用残差块对深层特征进行初步处理,同步完成匹配浅层特征维度的通道对齐,再经由多层感知机完成语义特征精细化建模;随后将优化后的深层特征与浅层低频分量 LL 逐元素相加,并借助残差块实现深层特征融合。最后,融合特征直接执行层归一化处理,输入多层感知机映射生成承载丰富特征信息的值。该处理流程表示如下:

$$F_v = MLP(RB(E_{i+1})). \tag{29}$$

$$V = MLP(LN(RB(F_v + F_{LL}))). \tag{30}$$

其中,F_v表示处理过程中的特征。

在完成多频注意力机制下查询、键、值三元组的构建后,MFCA 模块沿用自注意力经典计算范式完成特征运算,实现深浅层特征间多频信息的交互与深度融合。首先,计算多频查询和空间键之间的关联度,有效感知各类频率分量同全局空间结构间的内在关联特征。随后通过 SoftMax 激活函数来获得多频注意力权重图,其中,高频分量注意力重点关注于增强目标有效响应区域,而低频分量注意力能够精准表征目标空间位置信息并具备良好的噪声抑制能力。该过程表示为:

$$Af_x = SoftMax(Q_x \otimes K). \tag{31}$$

其中,Af_x表示多频分量注意力图。

下一步，该模块将多频分量注意力图分别与值相乘，对深浅层特征相互引导完成目标增强与虚警抑制。然后，特征经过两个带残差连接的多层感知机进行处理，获得融合后的低频与高频的特征。该过程表示如下：

$$S_x = Af_x \otimes V. \tag{32}$$

$$I_x = V + MLP(S_x) + MLP(V + MLP(S_x)). \tag{33}$$

其中，S_x表示处理过程中的特征。I_x表示注意力图提取的多频特征。

最后阶段，MFCA模块提取并融合浅层特征的多尺度频域语义信息，同时融合以上多频注意力提取的全局特征信息，再将深浅层特征进行语义聚合。具体操作为，多频特征拼接后经由逆离散小波变换与空洞卷积块进行融合，充分发挥低频特征对高频特征的净化作用。另一分支，浅层特征分解的多频分量分别经由线性层和Sigmoid函数来挖掘空频融合特征。随后将全部特征拼接起来，经由逆离散小波变换与空洞卷积块进行融合，充分感知浅层特征中目标与背景的语义差异。两分支的融合特征相加后由残差块进行聚合，得到MFCA模块的输出特征O_i。以上处理表示如下：

$$O_h = DC(IDWT([I_x])). \tag{34}$$

$$O_l = DC(IDWT([S(Lr((F_x)))])). \tag{35}$$

$$O_i = RB(O_h + O_l). \tag{36}$$

其中，$IDWT(\cdot)$表示逆离散小波变换。$DC(\cdot)$表示空洞卷积块。O_h表示深层分支输出的融合特征。O_l表示浅层分支输出的融合特征。

2 实验结果与分析

2.1 评价指标

精确的目标定位和精准的目标分割是评价红外小目标检测的基础和前提。为了验证所提出方法的有效性，本文采用了领域内通用的评价指标[25]，包括交并比（Intersection of Union，IoU），检测率（Probability of detection，Pd），虚警率（False-alarm rate，Fa），受试者工作特征曲线（receiver operating characteristic curve，ROC），参数量（Params），浮点运算量（FLOPs）。各个评价指标的定义和公式如下所示：

（1）交并比：交并比是像素级的评价指标，主要用于评估算法对目标轮廓的预测与实际轮廓的拟合度。该指标计算预测结果与真值标签之间的交集区域与并集区域的比值，其计算公式如下：

$$IoU = \frac{G \cap P}{G \cup P}. \tag{37}$$

其中，G表示真值标签，P表示预测结果。

（2）检测率：检测率是目标级的评价指标，用于评估全局像素聚合之后对每个目标的定位精度，计算正确检测到的目标数与总目标数的比值。计算公式如下：

$$P_d = \frac{T_{correct}}{T_{all}}. \tag{38}$$

其中，T_{all}表示图中所有目标总数。$T_{correct}$表示正确检测到的目标。当预测目标的质心与真实标签的质心之间的距离小于预定阈值时，认为该目标被正确检测到。本文实验中设定默认阈值为3。

（3）虚警率：虚警率用于评估预测错误的像素占所有图像像素的比值，其计算公式如下：

$$F_a = \frac{P_{false}}{P_{all}}. \tag{39}$$

其中，P_{all}表示所有图像像素数目，P_{false}表示被错误预测的像素数。当预测目标的质心与真实标签的质心之间的距离大于预定阈值时，认为该目标被错误检测。

（4）受试者工作特征曲线：ROC用于描述在不同虚警率下检测率的变化趋势。其表示在滑动阈值下的总体检测性能，全面反映了网络的综合表现。

（5）参数量：参数量指模型中可学习参数的总数，反映了模型的规模。以常规的二维卷积为例，假设输入特征图为$\mathrm{F}_{\mathrm{in}} \in \mathbb{R}^{C_{in} \times H_{in} \times W_{in}}$，二维卷积核子为$C \in \mathbb{R}^{C_{in} \times K_h \times K_w}$并且输出特征图为$O \in \mathbb{R}^{C_{out} \times H_{out} \times W_{out}}$。参数量与卷积核子的尺寸相关。计算公式如下：

$$Params = K_h \times K_w \times C_{in} \times C_{out}. \tag{40}$$

其中，$C_{out}, C_{in}, K_h, K_w$分别表示输出通道数目，输入通道数目，二维卷积核子的高度，二维卷积核子的宽度。

（6）浮点运算数：浮点运算数是模型执行推理过程中所进行浮点运算总数量，包括加法、减法、乘法和除法运算。以常规的二维卷积为例，浮点运算数与输入特征和卷积核有关。计算公式为：

$$FLOPs = 2 \times K_h \times K_w \times C_{in} \times C_{out} \times H_{in} \times W_{in}. \tag{41}$$

2.2 实验设置

本文采用公开数据集SIRST[22]、IRSTD-1k[54]和SIRST-UAVB[55]对算法的性能进行评估。本文采用了与本领域内一致的数据集划分方式。SIRST为真实数据集，划分方式为213张用于训练，214张用于

测试。IRSTD-1k 为真实数据集,划分方式为800张用于训练,201张用于测试。SIRST-UAVB 为真实数据集,划分方式为2400张用于训练,600张用于测试。值得指出的是,本文使用的三个数据集在训练样本量上具有显著差异,覆盖了从小样本到大规模的不同数据规模,能够在一定程度上反映模型在不同标注条件下的性能表现。

在网络训练之前,首先将所有输入图像进行归一化处理,然后进行数据增强,其过程包含随机翻转、随机裁剪等操作。本文训练中统一采用了256 × 256的分辨率,测试时模型对完整图像进行推理输出。训练的优化器和学习率策略分别为 Adam 和 MultiStepLR。初始学习率为0.0005,批尺寸为16,训练轮数为1000。所有模型均采用 Pytorch 深度学习框架进行实现、训练和测试。基于深度学习检测算法在 Nvidia GeForce RTX 3090 GPU 设备上运行得到结果。传统的检测算法采用 MATLAB R2023b 实现和测试,计算设备为 Intel Core i9-13900HX CPU。

2.3　对比实验分析

为了评估本文所提出算法的分割精度和检测效率等方面的综合性能。本文引入了基于模型驱动的传统算法进行对比,包括 Top-Hat[11],Max-Median[56],WSLCM[57],TLLCM[58]、IPI[16]和 MSLSTIPT[59]。同时引入了本领域内先进的基于数据驱动的深度学习算法,包括 ACM[22]、DNANet[25]、UIUNet[26]、SCTransNet[30]、FM-Net[37]、NS-FPN[40]。定量和定性结果分析显示了本文所提出的 WDFA-Net 具更有效的检测性能,具体分析见下文。本文遵循了领域内常用的结果阈值设定方式,固定采用0.5阈值。同时,本文遵循了对比算法原文的超参数设置。

2.3.1　定量分析

表1展示了各算法在公开数据集上的定量实验结果。由结果可知,基于深度学习的检测方法在三个基准数据集的整体性能显著优于传统检测方法。原因在于,传统方法采用模型驱动范式,高度依赖于人工的超参数调优,泛化能力较弱,难以适配复杂真实场景。而本文选取的三个真实数据集,涵盖多季节、多背景、多光照条件下的真实红外图像,存在背景扰动剧烈,目标尺寸多样等检测难点。同时,本文提出的 WDFA-Net 在三个基准数据集上均取得了优异的检测性能,相较于其他对比方法具备更强的适应性和鲁棒性。具体地,本文方法在 SIRST-UAVB 数据集上实现了全部性能指标的最优值。在 IRSTD-1k 数据集上,本文方法在交并比和检测率指标上均实现了最优水平,但在虚警率指标上略弱于 SCTransNet。WDFA-Net 相较于 SCTransNet 在交并比上提高4.2%,检测率上提高1.34%,而虚警率上涨9.1×10^{-7}。在 SIRST 数据集上,本文方法同样在交并比和检测率指标上均实现了最优水平,但在虚警率指标上略弱于 UIUNet。WDFA-Net 相较于 UIUNet 在交并比上提高2.46%,检测率上提高4.56%,而虚警率上涨2.4×10^{-6}。再根据表2中模型复杂度比较,UIUNet 具有

表1　不同检测方法在公开数据集上性能比较(红色代表最优结果,蓝色代表次优结果)

Table 1　Performance comparison of different detection methods on public datasets (red represents the optimal results and blue represents suboptimal results).

Detection Methods	IRSTD-1k			SIRST			SIRST-UAVB		
	IoU/(%)	Pd/(%)	Fa/(10^{-6})	IoU/(%)	Pd/(%)	Fa/(10^{-6})	IoU/(%)	Pd/(%)	Fa/(10^{-6})
Top-Hat	10.06	75.11	1432	7.14	79.84	1012	1.57	50.06	100
Max-Median	7.00	65.21	59.73	4.17	69.20	55.33	4.96	11.30	2.86
WSLCM	3.45	72.44	6619	1.16	77.95	5446	1.10	39.37	800
TLLCM	3.31	77.39	6738	1.03	79.09	5899	1.20	21.51	200
IPI	27.92	81.37	16.18	25.67	85.55	11.47	3.68	17.13	5.72
MSLSTIPT	11.43	79.03	1524	10.30	82.13	1131	5.90	9.84	3.58
ACM	60.33	**93.94**	69.31	68.34	93.92	30.12	20.80	51.77	11.18
DNANet	65.74	90.24	13.17	73.60	93.54	38.49	32.49	64.21	2.16
UIUNet	65.69	91.58	14.18	76.21	92.40	**10.70**	28.62	67.13	9.50
SCTransNet	68.03	93.27	**10.74**	76.32	**96.96**	14.41	**36.36**	**73.27**	**1.94**
FM-Net	**71.56**	92.26	12.58	76.75	**96.96**	15.71	35.17	71.32	2.12
NS-FPN	67.58	93.27	22.15	**77.85**	96.20	23.94	34.35	69.87	2.37
WDFA-Net	**72.23**	**94.61**	**11.65**	**78.67**	**96.96**	**13.10**	**36.5**	**73.88**	**1.62**

表2 不同检测方法的模型复杂度比较

Table 2 Comparison of model complexity for different detection methods

Detection Methods	Params(M)	FLOPs(G)
ACM	**0.398**	**1.33**
DNANet	4.697	56.11
UIUNet	50.54	186.3
SCTransNet	11.20	40.43
FM-Net	3.61	14.17
NS-FPN	3.91	27.18
WDFA-Net	10.44	17.02

大量的参数和运算量。相比之下,WDFA-Net 在可接受的计算花销下实现均衡且突出的检测性能。需要指出的是,WDFA-Net 在交并比和检测率指标上取得最优性能的同时,在 IRSTD-1k 和 SIRST 数据集上的虚警率相较于部分对比方法出现了小幅上升。该现象主要源于高频特征增强的双面效应。结合图2分析所知,高频分量对精准检测十分重要,但其内部包含与弱目标响应特征相似的背景噪声。WAD 模块虽然通过门控高斯滤波对高频分量进行净化,仍无法完全规避部分噪声随目标特征同步增强的问题。同样地,MFCA 模块中的高频注意力也存在放大噪声的局限。因此,虚警率小幅上升是增强弱目标响应过程中的附带效应,直观体现了检测精度与虚警抑制之间的平衡取舍。从表1整体性能来看,WDFA-Net 在检测性能上的显著提升远大于虚警率的微小涨幅,且在 SIRST-UAVB 数据集上实现了全部指标的最优结果,表明该方法在绝大多数真实场景下能够实现检测精度与虚警抑制的有效均衡。

综合而言,本文提出的 WDFA-Net 能够高效挖掘多频域特征信息,兼顾了高检测精度和强目标定位能力。该网络结合空频双分支降采样策略,有效缓解了特性信息损失和语义混叠问题,维持了深层网络中的目标特征响应强度。在编码器与解码器间的特征融合阶段,多频交叉注意力从频率解耦角度构建深浅层特征的全局上下文关联,进一步放大了目标与背景间的语义表征差异,同时充分发挥了低频信息抑制高频信息中噪声干扰的优势。此外,模型采用由深到浅的层级注意力交互方式,对融合特征完成精细化优化。因此,WDFA-Net 不仅具有精准的目标定位能力,而且可以实现细节的准确表征。

图6展示了不同检测方法在三个公开数据集上取得的 ROC 性能曲线。对比发现,本文提出的 WDFA-Net 较好地平衡了检测率和虚警率的性能表现。在 SIRST-UAVB 数据集上,本文方法 WDFA-Net 实现了检测率与虚警率的最佳均衡,能够在较低虚警率的同时保持高检测率。在 IRSTD-1k 与 SIRST 数据集上,本文方法虽没有实现最优的虚警抑制效果,但维持了最高的检测率。WDFA-Net 整体上的综合检测性能与 SCTransNet 相近,具备较强的能力优势。

2.3.2 定性分析

为了评估本文所提出算法对具体场景的检测性能,将本文方法 WDFA-Net 与对比检测方法 ACM、DNANet、UIUNet 和 SCTransNet 进行了可视化检测结果分析对比,如图7所示。红外小目标类型、形态多变,且被背景的复杂杂波淹没。本文提出的 WDFA-Net 能够精准定位目标并精细勾勒目标的轮廓细节。在图7(1)中,本文方法、ACM 与 SCTransNet 能够精准定位低对比度的目标,但其他方法漏检了这些暗弱的小目标。在图7(2)中,图像中包含与目标相似特性的亮点。本文方法能够准

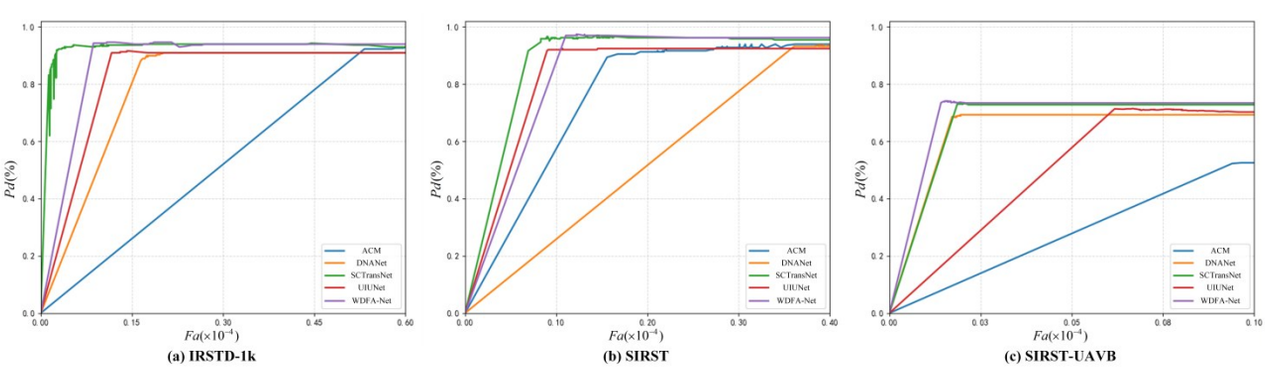

图6 不同检测方法在公开数据集上取得的 ROC 曲线:(a)IRSTD-1k;(b)SIRST;(c)SIRST-UAVB

Fig. 6 The ROC curves obtained by different detection methods on public datasets:(a)IRSTD-1k;(b)SIRST;(c)SIRST-UAVB

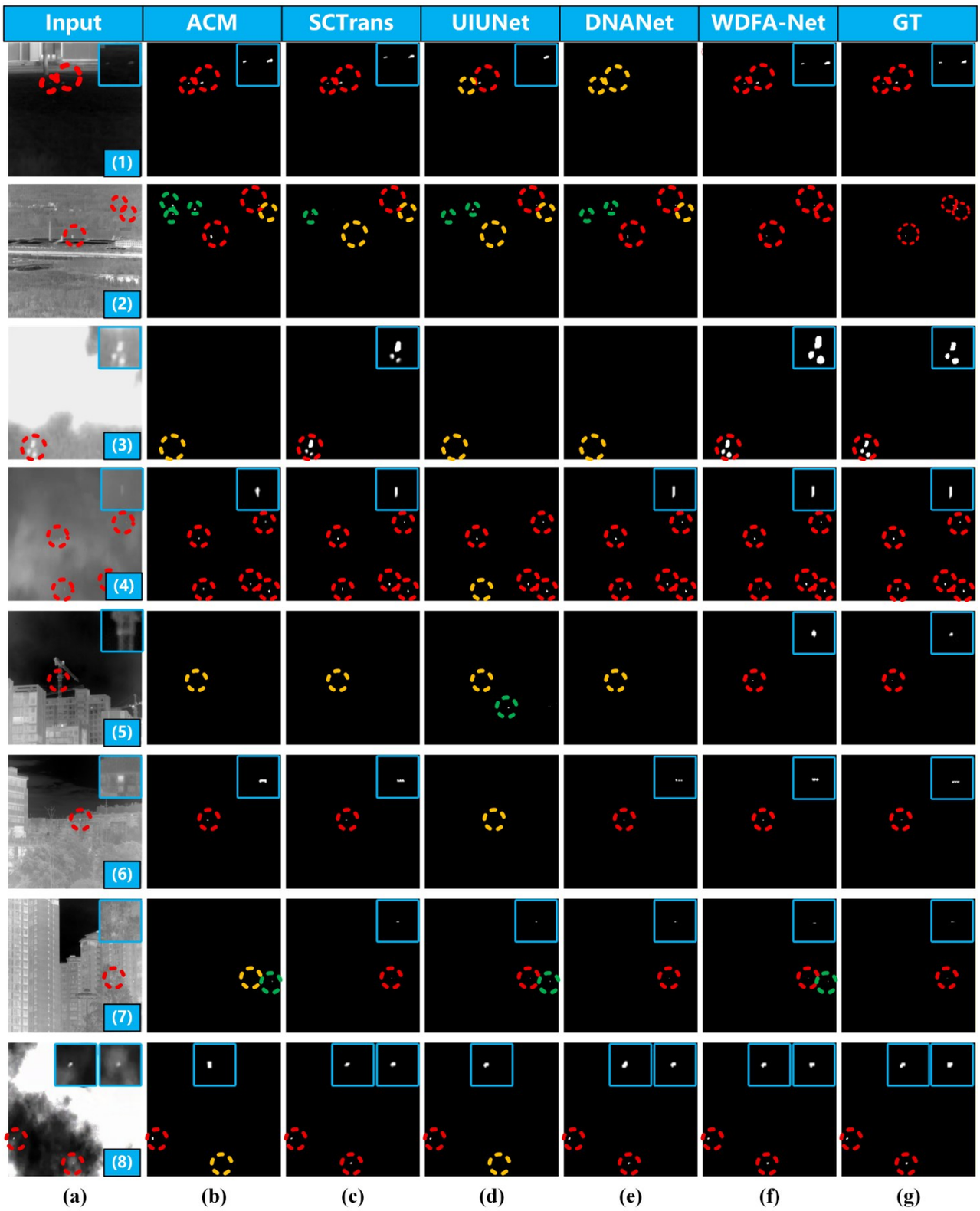

图 7　不同方法取得的检测结果可视化:(1)地面场景;(2)地形场景;(3)高亮场景;(4)天空场景;(5)建筑物场景;(6)植被场景;(7)建筑物场景;(8)高亮场景(红色虚线圆圈表示正确检测,绿色虚线圆圈表示虚警,黄色虚线圆圈表示漏检)

Fig. 7　Visualization of the detection results obtained by different methods: (1) Ground scene; (2) Terrain scene; (3) Highlight scene; (4) Sky scene; (5) Building scene; (6) Vegetation scene; (7) Building scene; (8) Highlight scene (red dotted circles indicate correct detections, green dotted circles indicate false alarms, yellow dotted circles indicate missed detections)

确消除虚警并在密集杂波中定位目标位置，但是其他方法性能表现糟糕，产生了一些虚警和漏检。在图7(3)中，本文方法比SCTransNet能够更准确地匹配目标形状，但其他方法没有检出目标，这是因为高亮背景的强干扰性。在图7(4)与(6)中，只有UIUNet产生了漏检，本文方法比其他方法更好地匹配了目标的轮廓。在图7(5)中，对于目标与建筑物重叠的强干扰场景，本文方法唯一实现了对弱小目标的准确检测。在图7(7)中，该场景包含建筑物、树木、路灯等多重结构性背景杂波，同时画面内存在光源形成的高亮虚假目标干扰，此类干扰特征与真实小目标高度相似。本文方法、ACM和UIUNet均存在误检问题，错误将高亮光源点判别为目标。而DNANet和SCTransNet有效识别干扰并实现了正确检测。在图7(8)中，该场景下小目标被强高亮杂波包围，本文方法有效抑制了干扰并实现了目标精准定位。对比之下，ACM和UIUNet在高亮背景干扰下产生了漏检。在可视化结果对比中，本文所提WDFA-Net在真实复杂场景中实现了鲁棒的检测性能，能够准确定位目标并抑制背景干扰。这得益于WDFA-Net高效利用空频融合信息进行特征表征，达到了提取并维持目标语义细节的效果；利用多类频域特征构建自注意力进行层级特征融合，帮助区分目标与背景的语义差异。此外，本文方法充分发挥了低频与高频特征的独特优势，具备了更强的目标定位能力和虚警抑制能力。但本文方法仍存在局限性，针对与目标特性相似的高亮点状干扰，模型难以完全区分，易产生虚警问题，这也是后续模型优化与性能提升的改进方向。

2.4 消融实验

2.4.1 网络架构

合理选取网络架构有利于算法良好平衡检测性能和运算复杂度。为了确定本文所提模型的最优配置与

最佳性能，本节对网络架构进行了相应的消融实验，包括通道数和特征尺寸两个要素。相应的实验结果在IRSTD-1k数据集上获得。如表3所示，本组实验改变了WDFA-Net编码器的各级通道数。比较发现，当编码器通道数依次设置为16，32，64，128，128时，WDFA-Net整体实现了最佳的检测性能。若进一步缩减通道数，网络虽然可以小幅提升虚警抑制能力，但该提升幅度十分有限，同时会造成交并比、检测率出现明显下降。究其原因，通道数不足会大幅限制网络的特征表征能力，难以学习到具备强区分性的有效语义。然而，过量的通道数引入了大量冗余特征和计算开销，不仅提高模型复杂度，还会造成网络整体检测性能下降。冗余特征极易形成无效干扰，进而引发虚警显著增加。根据公式(32)可知，通道数的改变直接大幅影响网络参数量。综合考虑，确定编码器通道数依次为16，32，64，128，128是最优的架构配置。

表3 WDFA-Net 改变通道数的消融实验结果

Table 3 Results of ablation experiments with WDFA-Net changing the number of channels

Channel Number	IoU	Pd	Fa	Params
8,16,32,64,64	71.65	91.92	**11.33**	**2.64**
16,32,64,128,128	**72.23**	**94.61**	11.65	10.44
32,64,128,256,256	71.87	93.60	33.90	41.54

表4 WDFA-Net 改变特征尺寸的消融实验结果

Table 4 Results of ablation experiments with WDFA-Net changing feature sizes

Feature Size	mIoU	Pd	Fa	Params
16,8,4,2,1	70.83	94.28	35.49	17.76
32,16,8,4,2	**72.23**	**94.61**	**11.65**	10.44
64,32,16,8,4	71.63	90.91	14.01	**8.61**

在WDFA-Net网络的特征融合阶段，编码特征首先经由分块嵌入模块完成预处理，获取压缩感知的特征图。为了探究特征尺寸对检测性能的影响，本组实验改变嵌入特征的尺寸进行对照，实验结果展示在表4中。实验结果表明，当依次采用32，16，8，4，2的特征尺寸时，WDFA-Net网络实现了最佳检测性能。过大的特征尺寸将引入大量的计算开销，拖慢模型训练效率。而尺寸过小极易造成特征过度压缩，丢失关键细节和语义信息，导致检测性能降低。本文中的分块嵌入模块采用卷积操作来实现不同尺寸特征映射，特征尺寸越小需要卷积核尺寸越大，因此表4中参数量指标呈现为特征尺寸越小，网络参数量越高的变化规律。

2.4.2 小波感知降采样模块

编码器中的小波感知降采样模块(WAD)采用空频域双分支结构，替代传统池化实现特征细节感知的下采样。该模块保留特征细节和深层目标响应，同时发挥高频噪声抑制作用。为了验证WAD模块的有效性，本组实验将WDFA-Net与几个相关变体进行了性能比较。

表 5　小波感知降采样模块的消融实验结果

Table 5　Results of ablation experiments with wavelet-aware down sample module

Networks	mIoU	Pd	Fa	Params	FLOPs
N w/o WAD	71.28	90.24	35.66	9.26	**13.34**
N w/o Fdb	71.15	89.90	**7.86**	9.37	13.70
N w/o Sdb	70.46	89.90	14.86	10.35	16.74
N w/o Ca	71.44	90.91	11.22	10.43	17.02
WDFA-Net	**72.23**	**94.61**	11.65	10.44	17.02

N w/o WAD:该变体网络移除了 WDFA-Net 中全部的 WAD 模块,并用传统最大池化操作实现下采样。

N w/o Fdb:该变体网络移除了 WDFA-Net 全部 WAD 模块中的频域分支,仅保留空域分支。

N w/o Sdb:该变体网络移除了 WDFA-Net 全部 WAD 模块中的空域分支,仅保留频域分支。

N w/o Ca:该变体网络移除了 WDFA-Net 全部 WAD 模块中用于两分支交互的通道注意力。

表 5 展示了 WDFA-Net 采用不同 WAD 模块变体的消融实验结果。实验可见,移除 WAD 模块后,检测性能出现了明显下降,尤其是虚警率大幅增加。如图 8 所示,WAD 模块有效辅助 WDFA-Net 实现更精准的特征感知,保留了降采样后目标的细节信息。对比图 8 中采样阶段 2、3 的特征能够发现,移除模块后,传统池化方式导致空间邻近目标发生特征混叠,严重制约后续的目标精准定位。而从采样阶段 4 的特征能够发现,池化操作还造成了特征图边缘出现异常高响应,成为干扰检测精度的潜在隐患。

然后,仅保留单分支结构时,网络检测精度同样大幅下降。这充分说明空频域双分支结构能够有效增强目标响应,提高网络目标定位能力。N w/o Fdb 实现了更低的虚警,既印证频域分支对于提高检测精度十分重要,也说明空域分支采用可学习卷积实现下采样比池化操作更具优势。此外,在移除双分支交互通道注意力后,WDFA-Net 网络取得的交并比下降 0.79%,检测率下降 3.7%,虚警率下降 4.3×10^{-5}。这证明了借助注意力机制完成双分支特征交互融合,可高效整合空频域有效信息,在完整保留目标细节特征的同时实现特征增强,最终实现整体检测精度提升。

此外,由表 5 中模型复杂度结果可知,WAD 模块给 WDFA-Net 方法引入了 1.18 M 的额外参数量和 3.68 G 的额外运算量,分别占模型总参数量的 11.3% 和总运算量的 21.6%。其中,空域分支是计算开销的主要来源,贡献了 1.07 M 参数和 3.32 G 运算量,因其包含下采样的卷积操作。频域分支引入的额外开销较小,参数量仅为 0.09 M,运算量为 0.28 G。而通道注意力的设计较为轻量化,仅引入 0.01 M 额外参数量。这表明 WAD 模块在辅助检测性能提高的同时,引入的计算开销处于可接受范围内。

2.4.3　多频交叉注意力模块

多频交叉注意力模块 MFCA 发挥了跨层特征聚合的重要作用,依据各类频域分量的物理特性生成查询、键、值三元组,再构建自注意力机制来感知

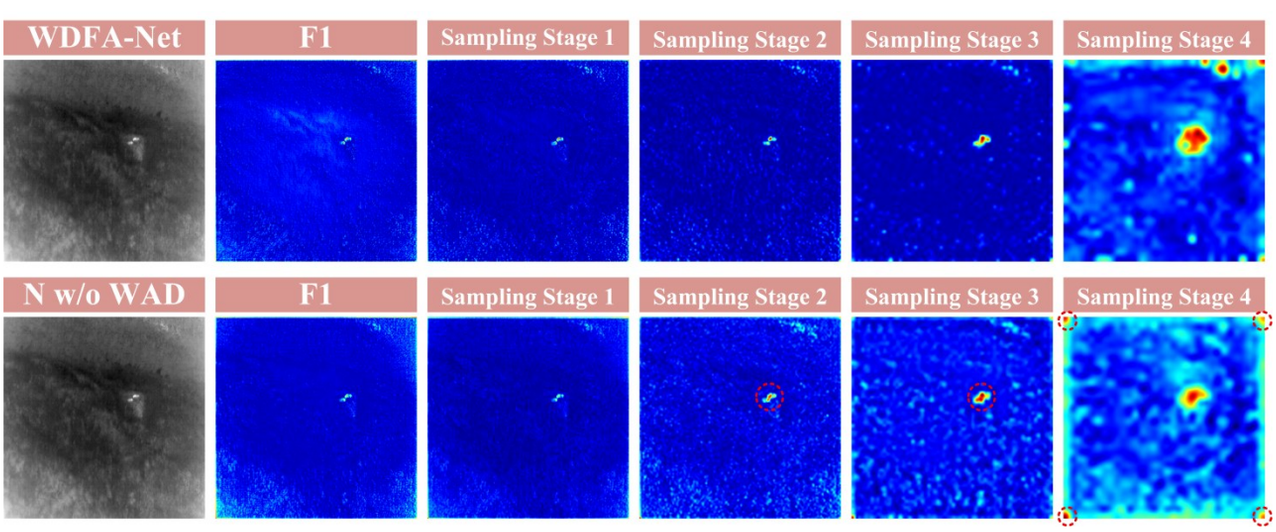

图 8　网络中采样后特征可视化:(a)WDFA-Net;(b)N w/o WAD

Fig. 8　Visualization of sampled features in networks: (a) WDFA-Net; (b) N w/o WAD

表6 多频交叉注意力模块的消融实验结果

Table 6 Results of ablation experiments with multi-frequency cross attention module

Networks	mIoU	Pd	Fa	Params	FLOPs
N w/o MFCA	70.86	91.25	19.93	**4.54**	**16.52**
N w CatRb	71.42	93.60	**11.31**	6.28	16.69
WDFA-Net	**72.23**	**94.61**	11.65	10.44	17.02

全局上下文和增强目标特征响应。MFCA模块能够结合高低频信息优势高效弥补层级特征的语义差异。为了验证MFCA模块的有效性，本组实验将WDFA-Net与几个采用其他聚合方式的变体网络进行了性能比较。

N w/o MFCA：该变体网络移除了WDFA-Net的MFCA模块，并仅用普通跳跃连接来传递每层特征。

N w CatRb：该变体网络移除了WDFA-Net的MFCA模块，仅对深层特征执行上采样操作，再与浅层特征拼接，最后经过残差块完成特征融合。

表6展示了WDFA-Net与两个变体网络在IRSTD-1k数据集上的对比实验结果。实验可见，移除MFCA模块后，WDFA-Net取得了交并比降低1.37%、检测率降低3.36%和虚警率上涨8.28×10^{-6}。这是因

为MFCA利用多类频域信息构建的自注意力来增强特征，同时捕获全局上下文信息。图9中的解码特征可视化验证了MFCA模块的有效性。在两类场景的L1层特征中，N w/o MFCA都产生了一个非目标的高响应区域（红色虚线圆圈标注），极易引发误检虚警。通过两个网络对应特征图进一步比较可以发现，WDFA-Net更精细地建模了全局上下文，既利于抑制背景干扰，又能促进区分目标与背景的特征边界。同时，WDFA-Net在多次解码融合后能够聚焦目标区域，强化目标特征响应。这验证了MFCA的跨层特征交互融合精炼提纯了多尺度语义信息，辅助模型提高检测性能。

N w CatRb相比WDFA-Net检测精度下降，其中交并比降低0.81%，检测率降低1.31%，仅保持了优秀的虚警抑制能力。该结果验证了跨层特征融合对提纯有效特征具备重要作用。但仅依靠残差块完成特征融合无法挖掘频域维度有效信息，还存在感受野受限的不足，难以实现高精度检测和强虚警抑制的双赢局面。而MFCA模块通过高低频特征的互补信息实现了性能提升。

此外，由表6中模型复杂度结果可知，MFCA模块给WDFA-Net方法引入了5.9 M的额外参数量和0.5 G的额外运算量，分别占模型总参数量的56.5%和总运算量的2.9%。该模块参数量增幅较大的主要原因是网络第五层特征的通道维度多，在生成分块嵌入时大核卷积操作引入较多可学习参数。即便相较于简单的通道拼接再融合方式，MFCA模块增加了4.16 M参数和0.33 G运算量，但提高了检测性能。MFCA模块依托适量新增参数完成多频域特征的表征学习，同时保持轻量计算，实现了跨层特征高效融合。

表7 泛化性实验结果

Table 7 Results of generalization experiments

Networks	mIoU	Pd	Fa
SCTransNet	72.04	93.16	35.26
FM-Net	74.80	90.87	42.05
NS-FPN	72.11	91.63	27.99
WDFA-Net	**76.13**	**94.30**	**27.65**

2.5 泛化性实验

为了评估模型面对数据分布偏移时的泛化能力，同时避免对单一数据集产生过拟合的风险，本下节开展跨数据集的泛化性对比实验。本次实验设置严格遵循跨域测试规则，实验中检测模型仅在

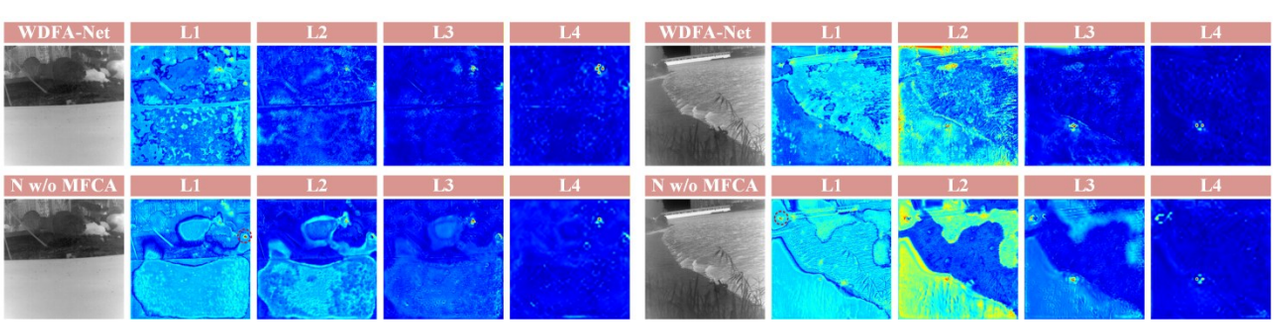

图9 网络中解码特征可视化：(a)WDFA-Net；(b)N w/o MFCA

Fig. 9 Visualization of decoded features in networks: (a) WDFA-Net; (b) N w/o MFCA

IRSTD-1k 数据集的训练集上完成训练,不经过目标数据微调,然后直接在 SIRST 数据集的测试集上进行性能评估。所有超参数和模型配置均保持一致,确保实验公平性。

表 7 展示了各模型的跨数据集泛化实验结果。与主流基线方法相比,WDFA-Net 在测试数据集上实现了更高的交并比和检测率,同时保持了较低的虚警率。实验结果表明,面对不同数据分布的图像差异,本文方法在跨数据集迁移时仍然保持性能优势,具有良好的跨域泛化能力。这是因为本文方法通过多维度频域信息挖掘小目标固有语义特征,摆脱了单一数据集背景纹理的依赖,能够普适性地区分真实目标与背景噪声干扰。此外需要指出的是,虽然本文方法在虚警上相对表现最佳,但虚警率的绝对数值还有待进一步降低,后续可进一步提高模型鲁棒性。

3　结论

针对现有方法池化下采样易导致特征丢失、空域建模为充分挖掘多类频域信息价值等问题,本文提出一种结合小波感知降采样与多频交叉注意力的红外小目标检测网络(WDFA-Net)。

首先,本文设计了小波感知降采样模块(WAD)来替代池化,通过空频域双分支协同采样在降维过程中保留小目标细节特征,同时提取空频域互补信息来增强目标响应。接着,设计了多频交叉注意力模块(MFCA),从频率解耦视角搭建全局上下文关联,以浅层频域特征为查询、全局空间特征为键、深浅融合特征为值,实现跨层特征融合,精准区分目标与背景语义。总之,本文方法充分发挥低频分量抑制噪声、高频分量增强目标响应的互补优势,提高检测性能。

在 IRSTD-1k、SIRST、SIRST-UAVB 三个数据集上的大量实验结果表明,WDFA-Net 在检测精度、虚警抑制与模型鲁棒性取得了优异的检测性能,可在复杂背景、多尺度、多光照等真实场景下实现红外小目标的精准定位与精细轮廓分割。

结合全文实验结果可知,本文方法在极低信噪比场景下依旧存在虚警残留问题,后续将围绕该难题进一步优化模型性能。具体改进方向包括:第一,构建自适应阈值调控机制,根据局部信噪比动态优化门控高斯滤波判定阈值,实现对不同复杂区域虚警的自适应抑制。第二,在 MFCA 模块中嵌入噪声感知分支,感知各区域的噪声强度,在高噪声区域降低高频分量的注意力权重,从而减少噪声引发的误检。上述改进方向有望进一步提升模型在极低信噪比场景下的抗干扰能力与检测稳定性。

面向实际应用,未来工程部署可从以下方面展开探索:第一,通过模型轻量化技术压缩模型规模,适配边缘计算设备;第二,利用推理加速框架优化模型推理速度;第三,针对特定应用场景(如暗目标)进行模型定制化优化。上述工作将作为本文的后续研究重点。

References

[1] Lin Zai-Ping. Research on weak target track-before-detect technologies for space-based infrared image[D]. Changsha: National University of Defense Technology, 2012.
(林再平.天基红外图像弱目标检测前跟踪技术研究[D].长沙:国防科技大学, 2012.

[2] Li Miao. Research on key technologies of detection and recognition via space-based optical surveillance systems [D]. Changsha: National University of Defense Technology, 2017.
(李淼.天基光学监视系统目标检测与识别关键技术研究[D].长沙:国防科技大学, 2017.

[3] Chen Juan. Infrared small target detection under complex background [D]. Beijing: University of Chinese Academy of Science, 2023.
(陈娟.复杂背景红外小目标检测技术[D].北京:中国科学院大学, 2023

[4] Chen Yu-Wen. Research on infrared small target detection algorithm based on visual saliency[D]. Xi'an, Xidian University, 2022.
(陈玉文.基于视觉显著性的红外小目标算法研究[D].西安:西安电子科技大学, 2022.

[5] Wu T H, Li B Y, Luo Y H, et al. MTU-Net: Multilevel TransU-Net for Space-Based Infrared Tiny Ship Detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61(3): 1 - 15.

[6] Wang Xiao-Yang. Research on infrared dim and small target detection theory and methodology based on sparse dynamic inversion[D]. Chengdu: University of Electronic Science and Technology of China, 2018.
(王晓阳.基于稀疏动态反演的红外弱小目标检测理论及方法研究[D].成都:电子科技大学, 2018.

[7] Sun Yang. Infrared small target detection under complex background[D]. Changsha: National University of Defense Technology, 2020.
(孙扬.复杂背景下的红外弱小目标检测技术研究[D].长沙: 国防科技大学, 2020

[8] Ying X, Liu L, Lin Z, et al. Infrared Small Target Detection in Satellite Videos: A New Dataset and A Novel Recurrent Feature Refinement Framework[J]. 2024.

[9] Yang Jun-Gang, Liu Ting, Liu Yong-Xian, et al. Infrared small target detection method based on nonconvex low-rank Tuck decomposition[J]. Journal of Infrared and Millimeter Waves, 2025, 44(2):311-325.
(杨俊刚,刘婷,刘永贤,等.基于非凸低秩塔克分解的红外小目标检测方法[J].红外与毫米波学报), 2025, 44(2): 311-325.

[10] Lin Zai-Ping, Luo Yi-Hang, Li Bo-Yang, et al. Gradient-aware channel attention network for infrared small target image de-

noising before detection[J]. Journal of Infrared and Millimeter Waves, 2024, 43(2):254-260.
(林再平,罗伊杭,李博扬,等. 基于梯度可感知通道注意力模块的红外小目标检测前去噪网络[J]. 红外与毫米波学报), 2024, 43(2):254-260.

[11] Rivest J F, Fortin R. Detection of dim targets in digital infrared imagery by morphological image processing[J]. Optical Engineering, 1996, 35(7): 1886-1893.

[12] Chen C L P, Li H, Wei Y T, et al. A local contrast method for small infrared target detection[J]. IEEE transactions on geoscience and remote sensing, 2013, 52(1): 574-581.

[13] Qin Y, Li B. Effective infrared small target detection utilizing a novel local contrast method[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(12): 1890-1894.

[14] Xiangyue Z , Qinghai D , Haibo L , et al. Infrared dim target detection algorithm based on improved LCM[J]. Infrared & Laser Engineering, 2017.

[15] Yi Xiang, Wang Bing-Jian. Fast infrared and dim target detection algorithm based on multi-feature[J]. Acta Photonica Sinica, 2017. 46(06):129-134.
(易翔,王炳健. 基于多特征的快速红外弱小目标检测算法[J]. 光子学报), 2017, 46(06):129-134.

[16] Gao C Q, Meng D Y, Yang Y, et al. Infrared patch-image model for small target detection in a single image[J]. IEEE transactions on image processing, 2013, 22(12): 4996-5009.

[17] Shi Y, Zhu Y, He R, et al. Auxdet: Auxiliary metadata matters for omni-domain infrared small target detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2026.

[18] Zhang Y H, Dai Z H, Pan C S, et al. NOC-YOLO: An exploration to enhance small-target vehicle detection accuracy in aerial infrared images [J]. Infrared Physics & Technology, 2025.

[19] Liu T, Liu Y, Yang J, et al. Graph Laplacian regularization for fast infrared small target detection[J]. Pattern Recognition, 2025, 158: 111077.

[20] Yue X B, Liu L W, Du Y. IR-DETR: An efficient detection transformer with multi-layer feature fusion for infrared small targets[J]. Infrared Physics & Technology, 2025.

[21] Li B, Ying X, Li R, et al. ICPR 2024 competition on resource-limited infrared small target detection challenge: Methods and results[C]. International Conference on Pattern Recognition, 2024, 62-77.

[22] Dai Y, Wu Y, Zhou F, et al. Asymmetric contextual modulation for infrared small target detection [C]. Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2021, 950-959.

[23] Dai Y, Wu Y, Zhou F, et al. Attentional local contrast networks for infrared small target detection[J]. IEEE transactions on geoscience and remote sensing, 2021, 59(11): 9813-9824.

[24] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [C]. International Conference on Medical image computing and computer-assisted intervention, 2015, 234-241.

[25] Li B, Xiao C, Wang L, et al. Dense nested attention network for infrared small target detection [J]. IEEE Transactions on Image Processing, 2022, 32: 1745-1758.

[26] Wu X, Hong D, Chanussot J. UIU-Net: U-Net in U-Net for infrared small object detection[J]. IEEE Transactions on Image Processing, 2022, 32: 364-376.

[27] Liu Y, Li B, Liu T, et al. RRCANet: Recurrent reusable-convolution attention network for infrared small target detection [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2025.

[28] Liu Q, Liu R, Zheng B, et al. Infrared small target detection with scale and location sensitivity [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, 17490-17499.

[29] Liu F, Gao C, Chen F, et al. Infrared small and dim target detection with transformer under complex backgrounds [J]. IEEE Transactions on Image Processing, 2023, 32: 5921-5932.

[30] Yuan S, Qin H, Yan X, et al. SCTransNet: Spatial-channel cross transformer network for infrared small target detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-15.

[31] Qin Z, Zhang P, Wu F, et al. Fcanet: Frequency channel attention networks [C]. Proceedings of the IEEE/CVF international conference on computer vision. 2021: 783-792.

[32] Yuan S, Qin H, Yan X, et al. ASCNet: Asymmetric sampling correction network for infrared image destriping [J]. IEEE Transactions on Geoscience and Remote Sensing, 2025.

[33] Chen L, Gu L, Li L, et al. Frequency dynamic convolution for dense image prediction[C]. Proceedings of the Computer Vision and Pattern Recognition Conference, 2025: 30178-30188.

[34] Huang J, Huang R, Xu J, et al. Wavelet-assisted multi-frequency attention network for pansharpening [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(4): 3662-3670.

[35] Cui Y, Ren W, Cao X, et al. Image restoration via frequency selection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 46(2): 1093-1108.

[36] Zhu Y, Ma Y, Fan F, et al. Toward robust infrared small target detection via frequency and spatial feature fusion [J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 1-15.

[37] Liu Y, Lin Z, Li B, et al. FM-Net: Frequency-Aware Masked-Attention Network for Infrared Small Target Detection[J]. Remote Sensing, 2025, 17(13): 2264.

[38] Deng C, Zhao Z, Xu X, et al. Gsfanet: Global spatial-frequency attention network for infrared small target detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2025.

[39] Li R, Xiao C, Yin Q, et al. Dynamic High-frequency Convolution for Infrared Small Target Detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2026.

[40] Yuan M, Meng D, Xi Z, et al. NS-FPN: Improving infrared small target detection and segmentation from noise suppression perspective[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2026, 27783-27792.

[41] Lin F, Ge S, Bao K, et al. Learning shape-biased representations for infrared small target detection[J]. IEEE Transactions on Multimedia, 2023, 26: 4681-4692.

[42] Lin F, Bao K, Li Y, et al. Learning contrast-enhanced shape-biased representations for infrared small target detection[J]. IEEE Transactions on Image Processing, 2024, 33: 3047-3058.

[43] Zhang M, Wang Y, Guo J, et al. IRSAM: Advancing segment anything model for infrared small target detection[C]. European Conference on Computer Vision, 2024, 233-249.

[44] Du N, Gong X, Liu Y. ISTD-diff: Infrared small target detection via conditional diffusion models[J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21: 1-5.

[45] Jia G, Cheng Y, Chen T. Irgraphseg: Infrared small target detection based on hierarchical gnn[J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21: 1-5.

[46] He K, Zhang X, Ren S, et al. Deep residual learning for image

recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, 770-778.
[47] Zhou Z, Rahman Siddiquee M M, Tajbakhsh N, et al. Unet++: A nested u-net architecture for medical image segmentation[C]. International workshop on deep learning in medical image analysis, 2018, 3-11.
[48] Wang H, Cao P, Wang J, et al. Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer[C]. Proceedings of the AAAI conference on artificial intelligence, 2022, 36(3): 2441-2449.
[49] Ma J, Chen J, Ng M, et al. Loss odyssey in medical image segmentation[J]. Medical image analysis, 2021, 71: 102035.
[50] Sudre C H, Li W, Vercauteren T, et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations[C]. International Workshop on Deep Learning in Medical Image Analysis, 2017, 240-248.
[51] Fan Y, Lyu S, Ying Y, et al. Learning with average top-k loss [J]. Advances in neural information processing systems, 2017, 30.
[52] Mao A, Mohri M, Zhong Y. Cross-entropy loss functions: Theoretical analysis and applications[C]. International conference on Machine learning, 2023, 23803-23828.
[53] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
[54] Zhang M, Zhang R, Yang Y, et al. ISNet: Shape matters for infrared small target detection[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 877-886.
[55] Yang J, Liu S, Wu J, et al. Pinwheel-shaped convolution and scale-based dynamic loss for infrared small target detection [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(9): 9202-9210.
[56] Deshpande S D, Er M H, Venkateswarlu R, et al. Max-mean and max-median filters for detection of small targets[C]. Signal and Data Processing of Small Target. International Society for Optics and Photonics, 1999, 3809: 74-83.
[57] Han J, Moradi S, Faramarzi I, et al. Infrared small target detection based on the weighted strengthened local contrast measure [J]. IEEE Geoscience and Remote Sensing Letters, 2020, 18(9): 1670-1674.
[58] Han J, Moradi S, Faramarzi I, et al. A local contrast method for infrared small-target detection utilizing a tri-layer window [J]. IEEE Geoscience and Remote Sensing Letters, 2019, 17 (10): 1822-1826.
[59] Dai Y M, Wu Y Q. Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection [J]. IEEE journal of selected topics in applied earth observations and remote sensing, 2017, 10(8): 3752-3767.

Infrared small target detection network combining wavelet-aware down sample and multi-frequency cross attention

Zhang Lei[1], Liu Yong-Xian[2], Yi Shu-Qi[2], Chen Jin-Bao[1], Guo Ya-Ping[3], Li Bo-Yang[2*], Lin Zai-Ping[2], An Wei[2]

(1. College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China;
2. College of electronic science and technology, National University of Defense Technology, Changsha 410073, China;
3. Equipment Project Management Center, Beijing 100064, China)

Abstract: Infrared small target detection is a key technology in the field of national defense and security, yet it faces severe challenges posed by complex real-world scenarios. Infrared small targets often present detection challenges such as blurred appearance and low signal-to-clutter ratio due to their extremely low pixel occupancy and interference from complex backgrounds. In recent years, detection methods based on deep convolutional networks have made significant progress; however, most existing approaches focus primarily on spatial feature modeling, overlooking the value of frequency-domain features and the information loss caused by pooling operations. To address these issues, this paper conducts research from a frequency-domain perspective and proposes a novel infrared small target detection network called WDFA-Net, which combines wavelet-aware down sample with multi-frequency cross-attention. The network replaces traditional pooling with a wavelet-aware down sample module (WAD), enabling dual-branch sampling in both spatial and frequency domains to preserve feature details while reducing dimensionality and enhancing target responses. Furthermore, a multi-frequency cross attention module (MFCA) is designed for feature fusion stage. This module constructs a physically meaningful self-attention mechanism based on multi-frequency components, fully leveraging the complementary information between high- and low-frequency features to achieve inter-layer feature interaction, fusion, and noise suppression. Experimental results demonstrate that WDFA-Net achieves excellent detection performance on the IRSTD-1k, SIRST, and SIRST-UAVB datasets, exhibiting strong robustness in complex real-world scenarios.

Key words: infrared small target detection, wavelet-aware down sample, multi-frequency cross attention, frequency-domain analysis, deep learning