

基于双向特征交互与动态尺度感知融合的红红外小目标检测

孙凯文¹, 黄丹², 杨武¹

(1. 重庆理工大学 计算机科学与工程学院, 重庆 400054;

2. 中国兵器科学研究院, 北京 100089)

摘要: 红外小目标检测旨在从复杂背景中识别出尺寸微小、对比度较低的目标, 其性能依赖于模型对弱信号特征的提取、边界区域的精确刻画及背景干扰的抑制能力。现有方法中普遍存在浅层结构信息与深层语义特征交互不足、静态单尺度融合策略适应性差等问题, 导致在复杂背景中目标边界模糊、识别精度下降。在MSHNet模型的基础上, 提出一种双向特征交互和动态尺度感知融合机制的检测模型, 模型通过双向特征交互模块实现编码器与解码器间的高效特征交互, 利用轻量化通道注意力增强目标区域响应与边界判别能力, 后续引入动态尺度感知融合模块, 根据不同尺度特征对目标识别的贡献度自适应分配融合权重, 从而优化特征整合并强化边界表达, 有效提高了目标边界的识别能力。在NUDT-SIRST与IRSTD-1K数据集上的实验结果显示, 该模型比基准算法交并比分别提升了1.7%和2.0%, 验证了模型的有效性。

关键词: 红外小目标检测; 双向特征交互; 动态尺度感知融合; U-Net网络; 深度学习

DOI: 10.11907/rjtk.261013

扫描二维码阅读全文:

中图分类号: TP391.41

文献标识码: A

文章编号: 1672-7800(XXXX)0XX-0001-09

Infrared Small Target Detection with Bidirectional Feature Interaction and Dynamic Scale-Aware Fusion

SUN Kaiwen¹, HUANG Dan², YANG Wu¹

(1. School of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China;

2. China Academy of Ordnance Science, Beijing 100089, China)

Abstract: Infrared small target detection aims to identify small, low-contrast targets from complex backgrounds. Its performance depends on the model's ability to extract weak signal features, accurately characterize boundary regions, and suppress background interference. Existing methods generally suffer from insufficient interaction between shallow structural information and deep semantic features, and poor adaptability of static single-scale fusion strategies, leading to blurred target boundaries and decreased recognition accuracy in complex backgrounds. Based on the MSHNet model, this paper proposes a detection model with a bidirectional feature interaction and dynamic scale-aware fusion mechanism. The model achieves efficient feature interaction between the encoder and decoder through a bidirectional feature interaction module, enhances target region response and boundary discrimination capabilities using lightweight channel attention, and then introduces a dynamic scale-aware fusion module. This module adaptively allocates fusion weights based on the contribution of features at different scales to target recognition, thereby optimizing feature integration and strengthening boundary representation, effectively improving the target boundary recognition capability. Experimental results on the NUDT-SIRST andIRSTD-1K datasets show that the proposed model improves the intersection-union ratio by 1.7% and 2.0% respectively compared to the benchmark algorithm, validating the model's effectiveness.

Key Words: infrared small target detection; bidirectional feature interaction; dynamic scale-aware fusion; U-Net; deep learning

收稿日期: 2026-03-04

作者简介: 孙凯文(2000-), 女, 重庆理工大学计算机科学与工程学院硕士研究生, 研究方向为目标检测; 黄丹(1983-), 女, 博士, 中国兵器科学研究院研究员, 研究方向为目标探测与识别; 杨武(1965-), 男, 硕士, 重庆理工大学计算机科学与工程学院教授, 研究方向为自然语言处理、大数据分析处理技术。本文通讯作者: 黄丹。

0 引言

红外小目标检测 (Infrared Small Target Detection,IRSTD) 是计算机视觉与红外成像领域中的一项重要研究任务,已逐渐在森林搜救、海上监督^[1-2]、火灾预警等应用场景中发挥关键作用。与可见光图像相比,红外成像具有全天候、抗光照变化能力强等优势,但其成像过程通常依赖远距离红外传感设备,受限于成像分辨率和物理成像条件,目标在图像中往往仅占据极少像素,呈现出尺寸小、对比度低、结构信息弱等典型特征。同时,复杂背景杂波、传感器噪声以及环境干扰的叠加效应进一步削弱了目标的可辨识性,使得目标信号极易淹没于背景之中。因此,如何在低信噪比、复杂背景和多尺度变化条件下实现对红外小目标的高效、稳定检测,仍然是一个极具挑战性的研究问题。

为提升红外小目标的检测性能,研究人员提出了众多方法。早期的传统方法主要分为 3 类:基于滤波的方法^[3-4]、基于局部对比的方法^[5-6]和基于低秩建模的方法^[7-11]。基于滤波的方法通过设计不同形式的空间或频域滤波器抑制平滑背景并突出局部亮度异常区域,具有实现简单、计算复杂度较低的优点,但在背景结构复杂或噪声较强时,往往容易放大非目标高亮区域,从而引入大量误检。基于局部对比度的方法借鉴人类视觉感知机制,通过增强目标与其邻域之间的亮度或结构差异以提升目标显著性,但该方法对窗口尺度设置和噪声分布高度敏感,在多尺度目标和复杂场景下检测稳定性较差。低秩建模方法将红外图像建模为低秩背景与稀疏目标的叠加形式,在抑制复杂背景方面具有一定优势,但通常依赖于迭代优化求解,计算开销较大,难以满足实时应用需求。总体而言,这些依赖人工设计特征和先验假设的传统方法在面对目标尺度变化显著、背景结构复杂以及噪声干扰严重的场景时,检测性能仍存在明显局限。

近年来,随着深度学习技术的不断发展,基于深度神经网络的方法已成为红外小目标检测的主流方法^[12-17]。这类方法通过端到端的方式自动学习多层次特征表示,能够在一定程度上同时完成背景抑制与弱目标增强任务。通过引入多层特征表示学习以及自上而下与自下而上的信息传递机制,模型能够在特征空间中同时刻画目标的局部细节与全局上下文,从而提升对复杂背景与弱目标的整体建模能力。相较于传统方法,深度学习方法在特征表达能力、泛化性能和复杂场景适应性方面均具有明显优势,为红外小目标检测提供了新的解决思路。

尽管基于深度学习的方法已经在红外小目标检测领域取得了显著进展,但现有研究仍然面临若干关键问题。首先,大多数方法在特征交互方式上采用较为单一的方式,这导致了浅层特征中的丰富空间细节信息与深层特征

中的高层语义信息之间缺乏充分的协同作用,从而影响小目标的结构信息与上下文语义的全面整合,更容易出现目标边缘模糊或响应不足等问题;其次,现有多尺度特征融合往往采用静态或均匀的权重分配方法,未能充分考虑到不同尺度特征对红外小目标表征能力的差异化贡献,使得部分尺度特征的有效信息被削弱或丢失,甚至引入额外噪声;再次,在复杂背景和多尺度目标共存的场景条件下,上述两部分问题往往会相互叠加,进一步影响网络模型检测结果准确性与稳定性。

针对红外小目标检测在图像对比度低、复杂背景和不同目标尺度变化的条件下,存在红外小目标易受噪声干扰、语义与细节信息缺乏有效协同的问题,本文提出了一种基于双向特征交互 (Bidirectional Feature Interaction, BFI) 与动态尺度感知融合 (Dynamic Scale-Aware Fusion, DSAF) 的红外小目标检测方法。该方法解决了不同尺度特征信息脱节、多尺度预测贡献不均衡和弱小目标显著性不足等关键问题,增强了网络模型在复杂场景下对红外小目标的检测能力。本文主要贡献如下:①提出了一种双向特征交互模块,该模块通过在编码器与解码器之间建立了跨层双向信息流,利用注意力引导机制将高层语义特征对低层空间细节进行语义约束,并反向运用低层空间细节信息对高层语义特征进行有效补充,这缓解了传统红外小目标检测模型框架中语义与细节信息不衔接的问题,增强了红外小目标区域的显著性和边缘表征能力;②设计了一种动态尺度感知融合机制,该模块通过引入针对不同目标匹配程度的自适应权重分配策略,在模型的多尺度预测头内部动态调整各尺度特征融合比例,这避免了固定或均匀融合策略带来的信息弱化与噪声引入问题,提升了模型检测多尺度红外小目标的鲁棒性;③构建了一种包含跨层级特征提取的双向特征交互模块与动态调整权重的动态尺度感知融合模块于一体的网络框架,实现了在低对比度、复杂背景以及不同尺度弱小目标场景下强稳定、高精度的红外小目标检测。在两个公开数据集上的实验结果表明,本文方法展现出了更优的综合检测性能,验证了模型有效性与良好泛化能力。

1 相关工作

针对红外小目标检测中复杂背景和小目标带来的挑战,研究人员提出了多种检测方法 with 建模框架,主要涵盖基于传统图像处理的检测模型和基于深度学习的端到端检测模型。早期传统的红外小目标检测模型主要包括基于经典滤波、局部对比度增强和低秩建模 3 种。

在基于滤波增强的检测方法中,Rivest 等^[3]提出基于数学形态学的 Top-Hat 检测方法,首次将非线性形态学运算系统地应用于微弱红外目标的增强与提取。该方法通过形态学滤波突出特定尺寸的各向同性亮目标,并结合简

单的时序一致性分析与阈值分割以实现检测。然而,该方法依赖固定尺度结构元素,对多尺度目标适应性较差,在复杂背景下易出现误检。随后,Deshpande等^[4]提出Max-Median滤波方法,通过局部窗口内最大值与中值的差分增强亮度突出的微弱目标,同时抑制背景纹理与随机噪声。但由于基于固定窗口进行局部运算,该方法在背景梯度较强或目标对比度较低的场景中容易产生虚警,鲁棒性受限。这类方法具有实现简单、计算效率高等优点,但在复杂背景与强噪声条件下对目标的表征能力不足。

为进一步提升对低对比度小目标的检测能力,研究者提出了一系列基于局部对比度建模的方法。Ha等^[5]提出加权结构化局部对比度量(Weighted Strengthened Local Contrast Measure, WSLCM),通过在局部背景区域引入结构化加权以提升目标与背景之间的亮度差异,从而增强低对比度小目标的显著性。然而,由于依赖固定窗口进行对比度计算,该方法易受噪声扰动影响,在强纹理背景下存在稳定性不足的问题。在此基础上,Han等^[6]进一步提出三层局部对比度量(Three-Layer Local Contrast Measure, TLLCM),通过构建局部、邻域与周边区域的三级对比度特征,实现对背景杂波的更有效抑制。然而,该方法仍受窗口尺度限制,在目标尺寸变化或背景扰动尖锐的场景下表现不够稳定。总体而言,此类方法对弱对比度目标具有良好的响应能力,但对噪声、纹理与尺度变化的适应性有限。

随着低秩与稀疏分解理论被广泛应用于红外小目标检测领域,Zhang等^[7]提出迭代加权稀疏恢复(Infrared Patch-Image, IPI)模型,将红外图像分解为低秩背景与稀疏目标,实现背景抑制与目标独立增强。尽管该方法在复杂背景中具有较强的稳健性,但依赖迭代优化导致计算开销巨大,不适用于实时检测。为提高稀疏建模的表达力,Zhang等^[8]提出非凸鲁棒自治模型(Nonconvex Rank Approximation Minimization, NRAM),利用非凸正则化以增强稀疏表示能力,更精确地从低秩背景中分离微弱目标。但该模型同样依赖迭代求解,计算复杂度较高。Dai等^[9]提出重加权红外图像块张量(Reweighted Infrared Patch-Tensor, RIPT)模型,将红外小目标检测建模为局部与非局部先验约束下的重加权张量表示问题,通过张量块刻画空间—时间或空间—结构关联性,实现对单帧小目标的增强与分离。然而,该方法在优化过程中计算量较大,且对复杂多尺度背景结构的适应能力仍有限。Xu等^[10]提出张量核范数部分和(Partial Sum of Tensor Nuclear Norm, PSTNN)模型,将小目标检测建模为张量空间的低秩—稀疏分解问题,通过多维结构刻画空间—时间关联性,在强噪声与复杂纹理场景下表现出更强稳定性。然而,该方法优化过程耗时,对多尺度结构的表达能力仍有不足。进一步地,Zhang等^[11]提出多子空间学习与时空图像块张量(Multiple Subspace Learning and Spatial-Temporal Patch-Tensor, MSLSTIPT)方法,结合多尺度结构张量与Transformer机

制,对局部结构稳定性与长距离上下文信息进行联合建模,从而提升目标显著性。但在极弱对比度或超小目标场景中,结构特征表达仍可能存在不足,且Transformer引入的计算成本较高。

近年来,基于深度神经网络的检测方法成为研究热点,红外小目标检测逐渐从传统的先验驱动方法转向端到端的特征表示学习。Wang等^[18]提出多域判别与虚警对抗(Multi-Domain versus False Alarm, MDvsFA)模型,通过构建多域判别结构,将检测任务建模为真实目标与伪目标之间的对抗分类问题,并结合空间—频域特征以降低复杂纹理导致的虚警。然而,该方法在超低信噪比条件下对背景扰动仍较为敏感。Dai等^[19]提出非对称局部对比网络(Asymmetric Local Contrast Network, ALCN)模型,引入了注意力引导的低对比度增强模块,通过增强目标区域的显著性特征,显著提升了弱小目标在杂波背景中的检测性能,同时保持了原图对比度的结构完整性,有效缓解了低对比度对检测带来的挑战。Zhang等^[20]提出背景建模与目标增强网络(Image and Structure Network, ISNet)模型,将检测任务建模为背景建模与目标增强的双分支联合优化问题。该方法一方面学习背景平滑先验以去除非平稳纹理,另一方面通过局部显著性聚焦机制增强小目标特征,从而在多数场景中保持稳定的检测性能。但其固定的尺度建模方式对尺寸变化较大的目标适应性较弱,且对极端噪声的泛化能力仍有不足。Li等^[21]提出密集嵌套注意力网络(Dense Nested Attention Network, DNANet)模型,密集嵌套注意力结构将红外小目标检测描述为跨尺度特征聚合与显著性提取的问题。该方法在多尺度路径之间构建细粒度的注意力交互,从而增强弱目标在深层特征中的响应。然而,DNANet的嵌套结构较为复杂,计算开销相对较大,并在背景结构密集区域可能出现注意力漂移现象。Liu等^[22]提出多尺度预测与高层语义融合网络(Multi-Scale Head Network, MSHNet)结构,在U-Net结构基础上进一步引入多尺度预测头,并结合尺度与位置敏感的损失函数,实现了对不同尺寸目标的精准定位。该方法在保证检测精度的同时显著降低了推理复杂度,但主要依托卷积神经网络和注意力机制,从复杂背景中自动学习高层语义特征,从而显著提升检测精度并有效降低虚警率。

尽管现有研究在红外小目标检测的特征建模、多尺度表征与背景抑制等方面取得了显著进展,但仍存在一些亟待解决的问题。一方面,多数方法侧重于单向的信息传递或简单的特征拼接,浅层空间细节与深层语义特征之间缺乏充分而有效的交互,导致小目标的结构信息与上下文语义难以协同表达,易出现边缘模糊与响应弱化现象;另一方面,现有多尺度融合策略通常采用固定或平均的融合方式,未能根据不同尺度特征对目标的实际贡献进行动态调节,在复杂背景或目标尺度变化显著的场景中,容易削弱优势尺度信息,甚至引入冗余噪声,从而影响检测准确性

与稳定性。因此,亟需一种能够强化特征双向交互、具备动态尺度感知与自适应融合能力的红外小目标检测模型,以提升模型在复杂环境下的鲁棒性和泛化性能。

2 本文方法

本文在 MSHNet 模型的基础上提出了一种新模型 BDNet,该模型创新地加入了双向特征交互与动态尺度感知融合模块,实现了不同特征互补与多尺度信息的自适应融合,从而增强了模型对目标的语义信息与空间信息的感知能力。在低对比度、复杂背景和目标尺度不同以及目标微小等场景下,BDNet 模型能够实现更精准的目标定位与边缘刻画,并提高检测鲁棒性和精确度。

2.1 模型架构

BDNet 模型架构如图 1 所示。其整个结构由常用的 U-Net 作为主干网络构成,红外线图像首先通过级联编码

器(ResNet)进行处理,提取到不同层级的特征图,然后通过级联解码器分别将这些特征图逐步恢复到与输入相同的尺寸。由于红外小目标通常只有几个像素大小,如果直接使用解码器的输出进行预测,容易导致目标细节信息丢失,进而影响检测精度。本文研究融合编码器与解码器的特征信息,以弥合两者之间的语义差距,提升模型对小目标的识别能力。BFI 模块在编码器与解码器特征间建立双向注意力引导,实现低层细节与高层语义信息的高效互补,强化小目标的显著性表达与边缘感知能力。DSAF 模块在多尺度预测头上引入自适应权重机制,根据不同尺度特征与真实标签的匹配程度动态分配权重,实现尺度信息的精准融合,提高模型对不同大小目标的感知鲁棒性。通过特征交互和动态融合机制,本文实现了语义与细节信息的高效整合,增强模型对小目标的感知能力,自适应适配复杂场景,同时兼顾小目标的敏感性和背景的鲁棒性。

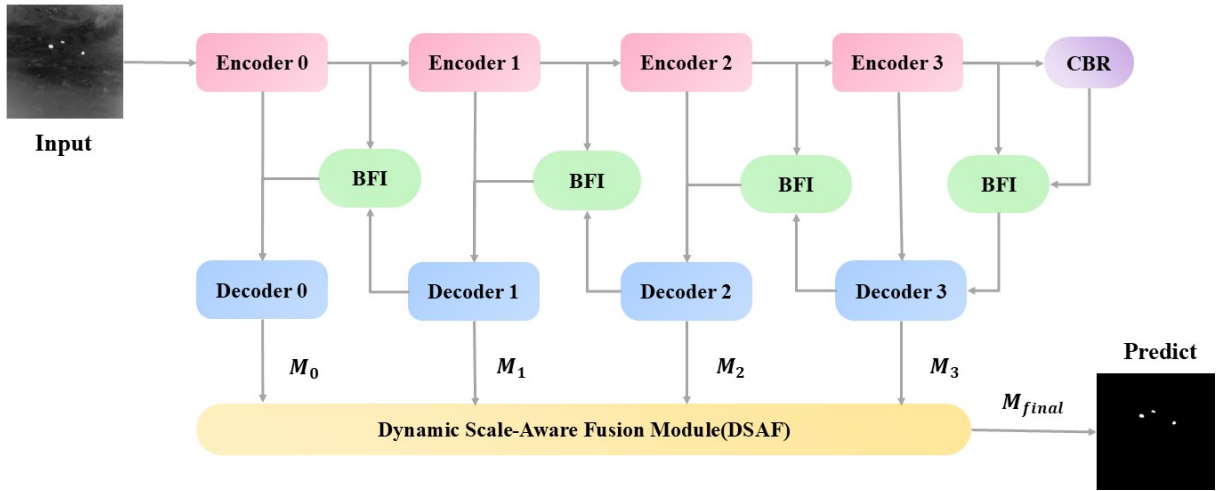


Fig. 1 BDNet model architecture

图 1 BDNet 模型架构

2.2 双向特征交互模块

虽然深层特征蕴含丰富的语义信息,但由于其分辨率较低,往往难以精确刻画目标的边缘与位置;相对地,浅层特征虽然包含更多的空间细节,但语义表达能力有限。因

此,为了实现不同层级特征间的高效融合,增强模型对红外小目标的表征能力,本文提出了双向特征交互模块,其结构如图 2 所示。该模块在结构上采用引导式融合策略,结合通道注意力机制,从而实现边缘信息的有效增强。

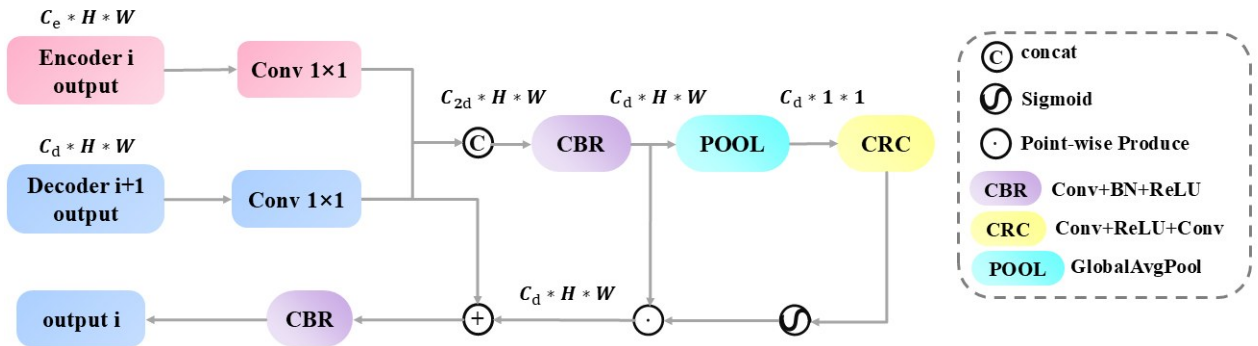


Fig. 2 Bidirectional feature interaction module structure

图 2 双向特征交互模块结构

双向特征交互模块结构主要包括特征对齐、融合交互、注意力增强与残差整合4个阶段。首先,为了实现编码器特征图 $E \in \mathbb{R}^{C_e \times H \times W}$ 与解码器特征图 $D \in \mathbb{R}^{C_d \times H \times W}$ 的有效交互,对二者进行通道对齐,通过 1×1 卷积将编码特征 E' 和 D' 分别映射处理,得到 $E' \in \mathbb{R}^{C_e \times H \times W}$ 和 $D' \in \mathbb{R}^{C_d \times H \times W}$,这一处理确保了后续融合操作中通道维度的一致性,为多层特征间的信息整合奠定基础。将对齐后的特征 E' 与 D' 在通道维度上拼接,并通过卷积融合结构进行处理,以充分挖掘和融合不同层级的语义信息与空间细节,从而获得初步融合特征图 $F \in \mathbb{R}^{C_f \times H \times W}$,如式(1)所示。

$$F = \text{ReLU}(\text{BN})(\text{Conv}_{1 \times 1}(\text{Concat}(E', D')))) \quad (1)$$

其中, BN 为批归一化操作,为了进一步提升重要性特征的响应,引入简化通道注意力机制,其结构包括全局平均池化与两层 1×1 卷积,输出通道注意力权重 $W \in \mathbb{R}^{C_d \times 1 \times 1}$,用于衡量不同通道的重要性,并将权重 W 作用于融合特征图 F ,实现对重要通道的强化,得到加权后的增强特征图 F' ,如式(2)、式(3)所示。

$$W = \sigma(\text{Conv}_{1 \times 1}(\text{ReLU}(\text{Conv}_{1 \times 1}(\text{AvgPool}(F))))) \quad (2)$$

$$F' = F \odot W \quad (3)$$

为兼顾高层特征的语义表达与低层特征的空间感知,模块采用残差增强策略,将增强特征 F' 与解码器特征 D' 相加,并通过 3×3 卷积与非线性变换生成最终输出特征图 $Y \in \mathbb{R}^{C_d \times H \times W}$,实现信息融合与特征优化,如式(4)所示。

$$Y = \text{ReLU}(\text{BN}(\text{Conv}_{3 \times 3}(D' + F'))) \quad (4)$$

双向特征交互模块不仅实现了编码器与解码器特征的有效耦合,还通过注意力机制提升了目标区域的显著性响应,从而更准确地表征红外小目标的边缘与位置信息。

2.3 多尺度贡献融合模块

为进一步提升多尺度目标分割结果的融合质量,特别是在红外小目标检测任务中更充分地发挥不同尺度特征图在局部区域的响应能力,本文提出了动态尺度感知融合模块。该模块能够依据每个尺度掩码对目标区域的响应强度,自适应地分配不同的重要性权重,并以此引导最终融合预测,提高整体检测性能。

设在4个不同尺度下得到的预测掩码分别为 $M_0, M_1, M_2, M_3 \in \mathbb{R}^{B \times 1 \times H \times W}$,其中 M_0 为原始分辨率输出, M_1, M_2, M_3 分别为上采样因子为2、4、8的多尺度预测结果。为了消除不同尺度掩码在数值分布上的差异,首先对各掩码进行归一化处理,使其值域统一,增强后续响应强度评估的可比性,如式(5)所示。

$$\hat{M}_i = \frac{M_i - \min(M_i)}{\max(M_i) - \min(M_i) + \epsilon}, i \in \{0, 1, 2, 3\}, \epsilon = 10^{-6} \quad (5)$$

其中, ϵ 是为了避免数值不稳定和除零操作而在分母中添加的常数。为度量每个尺度掩码对目标区域的响应能力,定义了平均响应强度指标 P_i ,如式(6)所示。在训练阶段使用真实的二值标签 $Y \in \{0, 1\}^{B \times 1 \times H \times W}$,在推理阶段则使用预测的二值掩码 Y ,计算每个尺度在目标区域上的

平均激活程度,该值反映当前尺度下掩码与目标区域的匹配程度,响应越强,代表该尺度对目标刻画越精准。

$$P_i = \frac{\sum_{h,w} \hat{M}_i \cdot Y}{\sum_{h,w} Y + \epsilon}, i \in \{0, 1, 2, 3\} \quad (6)$$

为了防止模型仅依赖单一尺度的输出,引入抑制因子 H_i ,通过对其余尺度的“非响应性”进行乘积累积,实现相对平衡的权重控制,如式(7)所示。

$$H_i = \prod_{j \neq i} (1 - P_j)^\delta, \delta = 0.5 \quad (7)$$

其中, δ 是一个平滑系数,用于控制低响应尺度对最终权重的影响程度。该设计保证即便某一尺度响应特别强,其他尺度也不会完全被忽略。将所有 H_i 进行归一化处理,得到每个尺度的最终融合权重 $w_i \in \mathbb{R}^{B \times 1 \times 1 \times 1}$,如式(8)所示。

$$w_i = \frac{H_i}{\text{mean}(H_i)}, i \in \{0, 1, 2, 3\} \quad (8)$$

最终通过拼接各尺度加权掩码,并使用 1×1 卷积进行特征整合。为保证尺寸一致性,对低分辨率输出进行上采样操作,得到融合预测掩码 M_{final} ,如式(9)所示。

$$M_{\text{final}} = \text{Conv}_{1 \times 1}(\text{Concat}(w_0 \cdot M_0, w_1 \cdot \text{Up}_2(M_1), w_2 \cdot \text{Up}_4(M_2), w_3 \cdot \text{Up}_8(M_3))) \quad (9)$$

其中, Up 表示对应尺度的上采样操作,以对齐分辨率。模块通过响应强度引导的加权机制,实现了多尺度预测结果的动态整合,既保留了小目标的细节信息,又提升了整体预测鲁棒性。

3 实验与结果分析

3.1 实验设置

本文在一台配备单块 NVIDIA GeForce RTX 4090 GPU 的工作站上进行所有实验, CUDA 版本为 12.4, 驱动版本为 550.100。训练环境包括 Python 3.9 和 PyTorch 1.13.1, 优化器采用 Adagrad, 初始学习率设为 5×10^{-2} , 训练批量大小设置为 4。根据不同数据集的需求, 总训练轮数分别设置为 400 和 800。在动态尺度感知融合模块为避免某一尺度在融合过程中权重过大, 抑制其他尺度的有效信息, 引入抑制因子机制。其中, 抑制因子的平滑系数 $\delta=0.5$, 用以在抑制背景干扰与保留有效信息之间取得平衡。

3.2 数据集

实验中使用了两个公开的红外小目标检测数据集: IRSD-1k 和 NUDT-SIRST。IRSD-1k 数据集包含 1 001 张红外图像, 而 NUDT-SIRST 数据集包含 1 327 张红外图像。对于这两个数据集, 图像均按 4:1 的比例分为训练集和测试集。

3.3 训练细节

目前, 在红外小目标检测任务中, 传统损失函数多基于像素级交叉熵或 IoU 等区域重叠度进行优化, 主要关注预测区域与真实区域之间的重合程度, 而对目标尺度差异

和空间位置偏移的敏感性不足。由于红外小目标通常具有尺寸小、对比度低、背景干扰强等特点,不同尺度、不同位置的目标在训练过程中对误差约束的需求并不一致。随着目标尺寸变化和背景复杂度提升,传统损失函数往往难以兼顾尺度一致性与定位精度,容易导致对微小目标不敏感或对复杂背景抑制能力不足的问题。

为了解决这些问题,本文在模型训练过程中沿用 MSHNe 模型中提出的尺度与位置敏感(Scale and Location Sensitive, SLS)损失函数^[22]。该损失函数在 IoU 损失的基础上,引入尺度加权机制和中心位置约束机制,从尺度一致性和空间定位一致性两个维度对预测结果进行联合约束,能够有效缓解传统损失函数对尺度和位置不敏感的问题,在复杂背景下提升目标定位精度并降低虚警率。

SLS 损失函数主要由尺度敏感损失(Scale Sensitive Loss, $\mathcal{L}_S$)和位置敏感损失(Location Sensitive Loss, $\mathcal{L}_L$)两个部分组成,其总体形式为两者之和。

尺度敏感损失定义如式(10)所示。

$$\mathcal{L}_S = 1 - w \cdot \frac{|A_p \cap A_{gt}|}{|A_p \cup A_{gt}|} \quad (10)$$

其中, A_p 表示预测目标的像素集合, A_{gt} 表示真实目标的像素集合, $|A_p \cap A_{gt}|$ 为预测区域与真实区域的交集像素数, $|A_p \cup A_{gt}|$ 为两者的并集像素数, w 为尺度权重因子,将预测区域与真实区域像素数量中的较小值与两者方差之和作为分子,将两者像素数量中的较大值与方差之和作为分母,计算其比值并作为尺度权重。当预测尺度与真实尺度差异较大时,权重 w 减小,从而增大整体损失值,使模型对尺度不匹配情况更加敏感,强化对尺度偏差较大样本的优化能力。

位置敏感损失定义如式(11)所示。

$$\mathcal{L}_L = \left(1 - \frac{\min(d_p, d_{gt})}{\max(d_p, d_{gt})}\right) + \frac{4}{\pi^2} (\theta_p - \theta_{gt})^2 \quad (11)$$

其中,通过对各自像素集合中所有像素坐标取平均获得预测中心与真实中心, d_p, d_{gt} 分别表示预测中心和真实中心在极坐标系下的径向距离, θ_p, θ_{gt} 分别为对应的极角。公式的第一项用于衡量中心点在径向距离上的偏差,公式的第二项用于刻画角度方向上的偏差。通过同时约束径向误差与角度误差,位置敏感损失能够有效区分不同类型的空间偏移,使模型对目标定位误差更加敏感,从而提升定位精度。

3.4 实验评价指标

本文采用交并比(IoU)和虚警率(Fa)作为像素级别的评价指标,并使用检测概率(Pd)以评估目标级别的性能。不同的指标从不同角度揭示了检测器的性能, Fa 和 Pd 分别关注召回率和虚警情况,而 IoU 综合考虑了这两者。

交并比定义如式(12)所示。

$$IoU = \frac{N_{inter}}{N_{union}} \quad (12)$$

其中, N_{inter} 表示预测目标与真实标签的交集像素数量, N_{union} 表示两者并集的像素数量。

虚警率定义如式(13)所示。

$$Fa = \frac{P_{false}}{P_{all}} \quad (13)$$

其中, P_{false} 表示误检的像素数, P_{all} 表示图像中所有像素的总数。

目标检测概率定义如式(14)所示。

$$Pd = \frac{N_{pred}}{N_{all}} \quad (14)$$

其中, N_{pred} 表示被正确检测出的目标数量, N_{all} 表示图像中的真实目标总数。

3.5 比较实验

为验证本文模型的优势,在两个主流红外小目标检测数据集 IRSTD-1k 与 NUDT-SIRST 上,将本文提出的模型与当前多种经典及最新方法进行了系统对比。运用交并比(IoU)、检测率(Pd)和虚警率(Fa)作为评估指标,全面衡量检测精度与鲁棒性。与 14 种代表性方法进行了对比,这些方法涵盖了基于滤波的 Top-Hat^[3]和 Max-Median^[4],基于局部对比的 WSLCM^[5]和 TLLCM^[6],基于低秩的 IPI^[7], NRAM^[8]、RIPT^[9]、PSTNN^[10]和 MSLSTIPT^[11],基于深度学习的方法包括 MDvsFA^[18]、ALCNet^[19]、ISNet^[20]、DNANet^[21]和 MSHNet^[22]。如表 1 和表 2 所示,传统方法在这些具有挑战性的数据集上表现较差,因为手工设计的特征提取先验存在局限性。相比之下,基于深度学习的方法能够自动学习有用的特征,并且比传统方法取得更好的结果。

Table 1 Experimental results on the IRSTD-1k dataset

表 1 IRSTD-1k 数据集上的实验结果 (%)

Method	Description	IoU ↑	Pd ↑	Fa ↓
Top-Hat	Filtering	10.06	75.11	1 432
Max-Median		6.998	65.21	59.73
WSLCM	Local Contrast	3.452	72.44	6 619
TLLCM		3.311	77.39	6 738
IPI	Low Rank	27.92	81.37	16.18
NRAM		15.25	70.68	16.93
RIPT		14.11	77.55	28.31
PSTNN		24.57	71.99	35.26
MSLSTIPT	Deep Learning	11.43	79.03	1 524
MDvsFA		37.34	83.71	88.52
ALCNet		65.68	89.25	27.71
ISNet		62.88	92.59	27.92
DNANet		65.71	91.84	17.61
MSHNet		67.16	93.88	15.03
ours		68.90	94.22	8.88

对于 IRSTD-1k 数据集,所提出的方法模型 IoU 为 68.90%、Pd 为 94.22%, Fa 仅为 8.88%,在各项评价指标上均优于现有对比方法,表现出更加优越且稳定的检测性能。尤其是与当前性能最优的深度学习模型 MSHNet 相比,本文方法在 IoU 上提升了 1.74%,在虚警率方面显著下降了 6.15%,在检测精度与误检控制之间实现了更加平衡的性能表现。

Table 2 Experimental results on the NUDT-SIRST dataset				
表 2 NUDT-SIRST 数据集上的实验结果 (%)				
Method	Description	IoU $\uparrow$	Pd $\uparrow$	Fa $\downarrow$
Top-Hat	Filtering	20.72	78.41	166.7
Max-Median		4.197	58.41	36.89
WSLCM		2.283	56.82	1309
TLLCM		2.176	62.01	1608
IPI	Local Contrast	17.76	74.49	41.23
NRAM		6.927	56.40	19.27
RIPT		29.44	91.85	344.3
PSTNN		14.85	66.13	44.17
MSLSTIPT	Low Rank	8.342	47.40	888.1
MDvsFA		35.86	85.22	95.37
ALCNet		72.89	96.19	30.40
ISNet		67.86	92.59	34.65
DNANet	Deep Learning	79.98	96.93	12.78
MSHNet		80.55	97.99	11.77
ours		82.60	97.19	5.56

在更具挑战性的 NUDT-SIRST 数据集上, 本文方法依然展现出良好的泛化能力和鲁棒性, 取得了 82.60% 的 IoU、97.19% 的 Pd 以及仅 5.56% 的 Fa, 整体性能同样优于大多数对比方法。与 MSHNet 相比, 尽管 Pd 指标略微下降了 0.8%, 但这一变化可能与复杂背景噪声干扰、多尺度融合过程中引入的副作用、较为保守的判定阈值设置, 以及训练与测试数据分布差异等因素有关。

上述性能提升主要得益于本文提出的双向特征交互模块和动态尺度感知融合机制。双向特征交互模块通过在不同尺度与层级之间构建双向信息流, 有效增强了特征

图之间的语义一致性, 使高层语义信息能够对低层细节特征进行指导, 同时低层空间细节也能够对高层特征进行补充, 从而显著提升了目标区域的判别能力。动态尺度感知融合机制能够根据不同尺度特征对目标的实际贡献, 自适应调整融合权重, 有效整合多分辨率下的关键信息, 使模型在面对尺度变化明显的红外小目标时仍能保持良好的响应能力。相较于 DNANet 和 MSHNet 等采用固定或静态多尺度融合策略的方法, 所提出的融合机制能够更加合理地刻画不同尺度特征的重要性, 从而获得更优的整体检测效果。

实验结果表明, 本文方法在不同数据集和复杂场景下均能够在检测精度与虚警控制之间取得良好平衡, 具有较强的稳定性和泛化能力。未来, 可进一步围绕背景噪声建模、多尺度融合策略优化、判定阈值的动态调整以及跨数据集分布适配等方向展开, 以进一步挖掘模型的性能潜力。

3.6 可视化实验

在 IRSTD-1k 数据集上对本文与多种具有代表性的红外小目标检测方法进行了可视化分析, 对比结果如图 3 所示。图中一共包含 8 列, 第 1 列和第 2 列为红外图像和其对应的标注真值, 第 3 列至第 8 列包含了所有对比方法, 其中第 3 列为基于滤波的方法、第 4 列为基于局部对比度的方法、第 5 列为基于低秩建模的方法, 其余列为基于深度学习的方法, 从而能够全面评估不同类别方法在复杂场景下的检测效果。

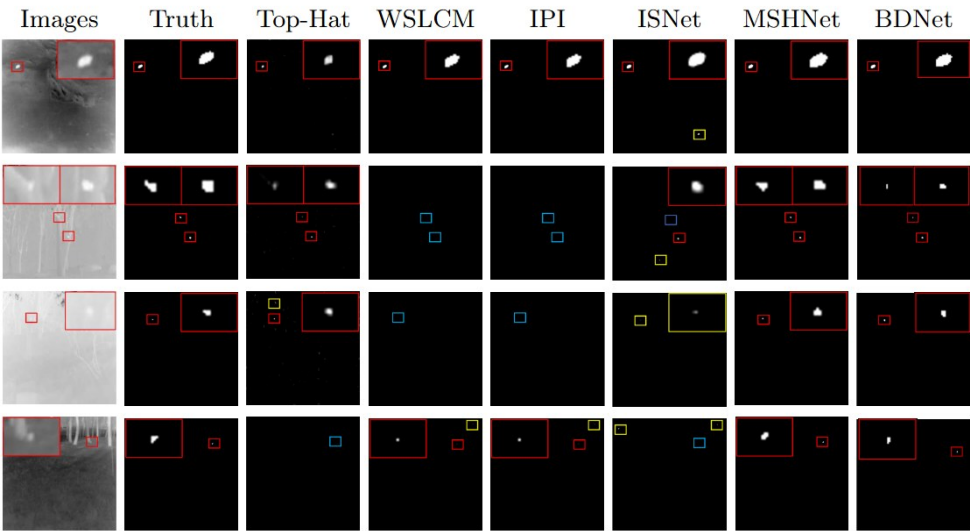


Fig. 3 Visualization results on the IRSTD-1k dataset

图 3 IRSTD-1k 数据集可视化实验结果

观察第 1 行的可视化效果可知, ISNet 方法在背景干扰较小的情况下产生了一定程度的误检, 而其他方法均未出现误检或漏检现象。但对比其他方法可以看出, 本文方法展现出了更优的目标定位精度、更清晰的目标边界以及更标准的形状。第 2 行是包含多个目标的复杂场景, 传统方法都出现了不同程度的问题, WSLCM 和 IPI 方法漏检了所

有目标, ISNet 同时存在误检与漏检问题, Top-Hat 方法检测到的目标存在边缘模糊现象, MSHNet 方法虽能够清晰地检测出所有目标, 但目标边缘的刻画超过了正确答案, 将部分背景识别为目标边界。而本文方法不仅可以准确识别所有目标, 还能够在识别准确的情况下实现最优刻画目标。第 3 行的可视化结果中, 传统方法 Top-Hat 和 ISNet

方法都出现了误检,WSLCM和IPI方法都出现了漏检,深度学习方法MSHNet又将目标周边部分区域错误地纳入目标区域,造成检测出的目标形状比正确目标的形状更大,而本文方法的检测结果在目标定位和形状上保持了更高准确性。在第4行的结果中,Top-Hat方法漏检了目标,WSLCM和IPI方法存在误检,ISNet方法同时出现了误检和漏检问题,MSHNet方法的检测结果依然存在将目标周围区域误判为目标的问题,本文方法保持着准确定位和刻画目标的能力,预测结果与真实标注真值在形状和边界上更接近。

综合可视化分析结果可知,无论是传统方法还是深度学习方法,本文模型不仅能够在多种复杂背景条件下实现对红外小目标的精准检测,而且模型在刻画目标边界和形状方面保持优异表现,这充分证明了本文方法在复杂背景抑制和精确目标检测方面展现出的优势。

3.7 消融实验

为了验证所提出的网络结构以及各模块的有效性,本文在IRSTD-1k和NUDT-SIRST数据集上开展了消融实验,实验结果如表3和表4所示。

Table 3 Ablation experimental results on the IRSTD-1k dataset

表3 IRSTD-1k数据集上的消融实验结果 (%)

BFI	DSAF	IoU ↑	Pd ↑	Fa ↓
		67.16	93.88	15.03
√		68.58	92.86	10.93
	√	67.95	92.52	8.96
√	√	68.90	94.22	8.88

Table 4 Ablation experimental results on the NUDT-SIRST dataset

表4 NUDT-SIRST数据集上的消融实验结果 (%)

BFI	DSAF	IoU ↑	Pd ↑	Fa ↓
		80.55	97.99	11.77
√		80.88	98.03	16.80
	√	81.46	96.91	8.72
√	√	82.60	97.19	5.56

当同时移除双向特征交互模块(BFI)与动态尺度感知融合模块(DSAF)时,模型退化为基础结构,其在IRSTD-1k和NUDT-SIRST数据集上的IoU分别为67.16%和80.55%,Pd分别为93.88%和97.99%,Fa分别为15.03%和11.77%。整体性能明显下降,表明在缺乏有效的跨层特征交互与尺度感知机制的情况下,模型对弱小目标的判别能力和对复杂背景的抑制能力均受到显著限制。

在仅引入双向特征交互模块的情况下(移除动态尺度感知融合模块),模型在IRSTD-1k数据集上的IoU提升至68.58%,Fa降低至10.93%,说明该模块通过增强编码器与解码器之间的跨层信息流,有效提升了目标区域的特征表达能力,并在一定程度上增强了模型对背景干扰的鲁棒性。虽然在NUDT-SIRST数据集上,尽管Pd略有提升(98.03%),Fa却显著上升至16.80%,表明在缺乏动态尺度感知融合的情况下,多尺度特征之间的贡献难以得到合理约束,容易引入额外误检。

在仅引入动态尺度感知融合模块(移除双向特征交互模块)时,模型在IRSTD-1k数据集上的IoU提升至67.95%,Fa降低至8.96%;在NUDT-SIRST数据集上,IoU提升至81.46%,Fa显著下降至8.72%。该结果表明,动态尺度感知融合机制能够有效建模不同尺度特征对检测结果的贡献差异,从而突出更具判别性的尺度信息并抑制虚警。然而,此时Pd出现一定程度的下降,说明缺乏跨层语义与细节的双向协同,可能影响模型对目标整体结构的完整捕捉。

当双向特征交互模块与动态尺度感知融合模块同时引入时,模型在两个数据集上的各项指标均达到更优。在IRSTD-1k数据集上,IoU、Pd和Fa分别达到了68.90%、94.22%和8.88%;在更具挑战性的NUDT-SIRST数据集上,IoU提升至82.60%,Fa进一步降低至5.56%,展现出优异且稳定的检测性能。上述结果充分验证了两种模块在结构上的互补性:双向特征交互模块增强了不同层级特征之间的上下文关联与语义一致性,而动态尺度感知融合模块则自适应调节不同尺度特征的贡献权重。两者的协同作用有效提升了模型在复杂背景和多尺度目标场景下的检测精度与鲁棒性。但同时引入两个模块后,模型在IRSTD-1k和NUDT-SIRST数据集上的性能变化趋势存在差异,这种差异主要源于数据集背景复杂度和目标分布特性的不同。在背景相对简洁、干扰较少的IRSTD-1k数据集上,BFI模块通过强化跨层语义与细节信息交互,有效提升了对弱小目标的表征能力,DSAF模块进一步优化了多尺度特征权重分配,从而实现Pd提升与Fa下降的同步改进。而在背景更为复杂、噪声干扰更强的NUDT-SIRST数据集上,虽然DSAF模块增强了尺度选择性,显著抑制了虚警,使Fa明显降低,但复杂背景下目标与噪声的相似性增加,使模型在权衡误检与漏检时趋于保守,部分微弱目标响应被抑制,导致Pd略有下降。因此,该差异本质上反映了模型在不同背景复杂度条件下对特征增强与噪声抑制之间权衡策略的变化。

4 结语

本文提出了基于双向特征交互与动态尺度感知融合的红外小目标检测方法,其通过引入双向特征交互模块和动态尺度感知融合模块,增强了跨层级特征之间的信息交互能力,有效调整了不同尺度特征对小目标检测的贡献权重,从而显著提高了模型整体性能。消融实验证明了该方法的有效性,总实验和可视化分析展现出该方法优异的检测性能。该方法在红外小目标检测任务中取得了良好的性能提升,但仍有进一步改进的空间。未来,将进一步探索更强的多尺度特征融合方法,适用于云层、海杂波、热噪声等不同场景;利用基于生成模型的数据增强技术提成高质量红外小目标样本,缓解红外小目标数据稀缺问题;研

究弱监督或半监督的方法, 减少对人工标注的依赖, 提高模型可扩展性。

参考文献:

- [1] TEUTSCH M, KRUGER W. Classification of small boats in infrared images for maritime surveillance[C]//Carrara: 2010 International WaterSide Security Conference, 2010.
- [2] ZHAO M J, LI W, LI L, et al. Single-frame infrared small-target detection: a survey[J]. IEEE Geoscience and Remote Sensing Magazine, 2022, 10(2): 87–119.
- [3] 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.
- [4] 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 Targets, 1999: 74–83.
- [5] 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.
- [6] 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.
- [7] GAO C, MENG D, 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.
- [8] ZHANG L, PENG L, ZHANG T, et al. Infrared small target detection via nonconvex rank approximation minimization joint $\ell_2, 1$ norm[J]. Remote Sensing, 2018, 10(11): 1821.
- [9] DAI Y, WU Y. 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.
- [10] ZHANG L, PENG Z. Infrared small target detection based on partial sum of the tensor nuclear norm[J]. Remote Sensing, 2019, 11(4): 382.
- [11] SUN Y, YANG J, AN W. Infrared dim and small target detection via multiple subspace learning and spatial-temporal patch-tensor model[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(5): 3737–3752.
- [12] CHEN L, FU Y, WEI K, et al. Instance segmentation in the dark[J]. International Journal of Computer Vision, 2023, 131(8): 2198–2218.
- [13] WANG H, ZHOU L, WANG L. Miss detection vs. false alarm: adversarial learning for small object segmentation in infrared images[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 8509–8518.
- [14] WEI K, AVILES-RIVERO A I, LIANG J, et al. TFPnP: tuning-free plug-and-play proximal algorithms with applications to inverse imaging problems[J]. Journal of Machine Learning Research, 2022, 23(16): 1–48.
- [15] YANG B, ZHANG X, ZHANG J, et al. EFLNet: enhancing feature learning network for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1–11.
- [16] YING X, XIAO C, AN W, et al. Visible-thermal tiny object detection: A benchmark dataset and baselines[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(7): 6088–6096.
- [17] CIOCARLAN A, LE HEGARAT-MASCLE S, LEFEBVRE S, et al. A contrario paradigm for YOLO-based infrared small target detection[C]//IEEE International Conference on Acoustics, Speech and Signal Processing, 2024: 5630–5634.
- [18] WANG H, ZHOU L, WANG L. Miss detection vs. false alarm: adversarial learning for small object segmentation in infrared images[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 8509–8518.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] LIU Q, LIU R, ZHENG B, et al. Infrared small target detection with scale and location sensitivity[C]//Seattle: The IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024.

(责任编辑: 孙 娟)