

基于多尺度融合与局部对比的红外小目标检测

吴苗列, 乔钢柱*, 王 斌

(中北大学 计算机科学与技术学院, 山西 太原 030051)

摘要: 红外小目标检测(Infrared Small Target Detection, ISTD)在军事预警、灾害救援与精确制导等领域具有重要的应用价值,然而现有方法存在局部与全局特征不平衡、小样本条件下模型鲁棒性不足等核心问题。针对上述问题,本文提出了一种基于多尺度动态融合与局部对比学习的网络模型MDLNet。首先,在Transformer的架构上引入动态加权融合多尺度特征机制,有效解决了小目标检测中局部与全局特征不平衡的问题,同时显著降低了注意力机制的冗余性;其次,在网络深层添加混合扩张卷积模块,通过高分辨率特征重构和深层特征的语义信息,以此强化目标边缘表征,从而提升检测精度;接着,设计了一种局部对比卷积模块,融合传统局部对比思想,大幅降低模型对训练样本数量的依赖,即使在小样本数据集上仍能保持优异的鲁棒性;最后,在NUAA、IRSTD-1k和SIRSTAUG数据集上进行了一系列实验。结果表明:本文所提的方法在三种数据集上的F1-Score分别达到了88.98%, 83.51%和86.38%,相比基准模型分别提高了1.06百分点,2.48百分点和0.76百分点。

关键词: 红外小目标检测; Transformer; 注意力机制; 多尺度特征; 局部对比

中图分类号: TP393

文献标识码: A

doi: 10.62756/jnuc.issn.1673-3193.2025.08.0001

引用格式: 吴苗列, 乔钢柱, 王斌. 基于多尺度融合与局部对比的红外小目标检测[J]. 中北大学学报(自然科学版), 2026, 47(3): 287-296.

WU Miaolie, QIAO Gangzhu, WANG Bin. Multiscale dynamic fusion with local contrast learning for infrared small target detection [J]. Journal of North University of China (Natural Science Edition), 2026, 47(3): 287-296.

Multiscale Dynamic Fusion with Local Contrast Learning for Infrared Small Target Detection

WU Miaolie, QIAO Gangzhu*, WANG Bin

(School of Computer Science and Technology, North University of China, Taiyuan 030051, China)

Abstract: Infrared small target detection (ISTD) has important application value in the fields of military early warning, disaster rescue and precision guidance, etc. However, the existing methods have core problems such as an imbalance between local and global features, and insufficient robustness of the model under small sample conditions. To address these issues, this paper proposed a new model called multiscale dynamic fusion with local contrast learning (MDL), which was based on the hybrid architecture of Transformer and convolutional neural network (CNN), integrating multi-scale dynamic weighting and

收稿日期: 2025-08-04

基金项目: 山西省研究生教育创新项目(2024SJ278)

作者简介: 吴苗列(2001—),男,硕士生,主要从事红外小目标检测的研究。

* 通信作者: 乔钢柱(1975—),男,教授,博士,主要从事物联网技术及应用、大数据处理及区块链技术研究。E-mail: qiaogz@nuc.edu.cn。

adaptive local contrast constraints. Firstly, a dynamic weighted fusion multi-scale context augmentation (MCA) was introduced into the architecture of Transformer, which effectively solved the problem of imbalance between local and global features in small target detection, and at the same time significantly reduced the redundancy of the attention mechanism. Secondly, this model included a dual-branch hybrid dilated convolution (DHDC) module located in the deeper layers of the network, which enhanced the detection accuracy through high-resolution feature reconstruction and semantic information from deeper features as a way to strengthen the target edge representation. In addition, a local contrast convolution (LCC) module was designed, which incorporated the traditional local contrast idea to significantly reduce the model's dependence on the number of training samples and maintains excellent robustness even on small sample datasets. Finally, a series of experiments were conducted on NUAA,IRSTD-1k and SIRSTAUG datasets. The results show that the method proposed in this paper achieved $F1$ -Scores of 88.98%, 83.51%, and 86.38% on the three datasets, representing improvements of 1.06, 2.48, and 0.76 percentage points compared with the baseline model, respectively.

Key words: infrared small target detection; Transformer; attention mechanism; multiscale features; local contrast

0 引言

基于红外检测系统的ISTD(Infrared Small Target Detection)是预警系统、精确制导武器和海上监视系统等应用的关键技术^[1],在计算机视觉领域,一直都是重要且热点的研究课题^[2]。ISTD方法可以分为基于非深度学习的方法和基于数据驱动的深度学习方法(Deep Learning, DL)方法两大类。

基于非深度学习的ISTD方法主要采用传统的图像处理技术,以增强目标特征。如Li等^[3]基于红外背景一致性假设,通过设计合适的滤波器抑制杂波,利用原图像与预测背景的差分得到真实目标。其原理简单,但在复杂场景的虚警率较高。Han等^[4-5]基于稀疏矩阵和低秩矩阵恢复的方法,将目标检测问题转换为数学优化问题。该类方法计算复杂且对噪声敏感,通常只针对特定场景有效,难以应对实际应用中复杂的成像条件和多变的干扰因素。

近年来,基于深度学习的ISTD取得了显著进展。深度学习方法主要是通过设计具有针对性的网络来提取目标特征、检测目标位置,其检测效果较好且自适应性强。目前,以Faster-RCNN^[6],Yolo^[7],SSD^[8]等为代表的物体检测网络,在各个数据集上取得了很好的成绩。此外也提出了一系列专门针对ISTD的分割类型神经网络结构:Dai等^[9-10]提出了一种基于分段的网络模型ACM,通过不对称的上下文模块融合多尺度特征,之后在其基础上通过引入扩

大的局部对比度进行改进,构建出ALC-Net;Wang等^[11]将ISTD建模为生成对抗网络,兼顾漏检和虚警;Li等^[12]结合医学图像领域的密集嵌套交互结构提出的DNA-Net网络模型,将神经网络高低层的不同信息进行冗余性结合,确保目标信息能够完整、显著地传递至解码器端,利用级联通道和空间注意力机制实现弱小目标检测。但基于深度学习的ISTD方法存在下述问题:1)主流方法(涵盖CNN主导、Transformer主导及两类架构的初步融合方案)普遍存在局部特征提取与全局建模失衡问题,CNN受固定感受野限制,难以捕捉极小目标的全局关联信息;Transformer虽能建模长程依赖,却易放大红外图像的低信噪比噪声;即便部分方法尝试融合两类架构,也未针对ISTD目标特征微弱、背景干扰多变的核心特性设计适配协同机制,最终导致小目标细节丢失。2)受红外成像物理机制的影响,光照衰减、热噪声及云层/雪地反光等环境辐射易造成背景复杂、目标信噪比低,进一步增加模型特征区分的难度。3)ISTD中目标区域通常仅占图像的0.01%,导致正负样本比例悬殊,而该领域标注样本因目标呈“点/斑状”、边界难判定,获取成本远高于普通任务,加剧了样本供需矛盾。这三方面问题相互叠加,使ISTD在复杂场景下的检测面临显著挑战。

针对上述问题,本文提出了基于多尺度动态融合与局部对比学习的网络模型(Multiscale Dynamic Fusion with Local Contrast Learning for Infrared Small Target Detection, MDLNet),通过多尺度动态融合与局部对比学习的协同优化机

制,实现了ISTD在特征增强、噪声抑制与样本效率上的三重突破。

1) 设计了多尺度上下文增强(Multi-scale Context Augmentation, MCA)模块。基于Transformer架构重构动态加权融合机制,结合任务自适应门控和频域先验引导策略,在减少注意力冗余计算的同时显著提升多尺度语义捕获效率,有效平衡局部细节与全局上下文,解决了特征失衡问题。

2) 设计了双分支混合扩张卷积(Dual-branch Hybrid Dilated Convolution, DHDC)模块。采用非对称空洞卷积结构,通过深度可分离卷积与特征混洗策略,将上采样细节丢失率降低并保持边缘完整性,显著抑制背景噪声。

3) 引入局部对比约束模块(Local Contrast Convolution, LCC)。融合传统局部对比思想与可学习卷积核,通过动态灰度差异分析提升信噪比,显著降低模型对训练样本数量的依赖,提升了小样本条件下的鲁棒性。

如图1所示,红框为红外小目标,黄框为噪声。原图经过一层卷积模块后得到第二张特征图,可以观察到有明显的噪声干扰。第三张图是经过MCA模块处理后的特征图,不仅有效抑制了噪声干扰,同时增强了目标特征。第四张图是通过DHDC模块处理后的结果,该模块对目标特征进行了更精细的处理。

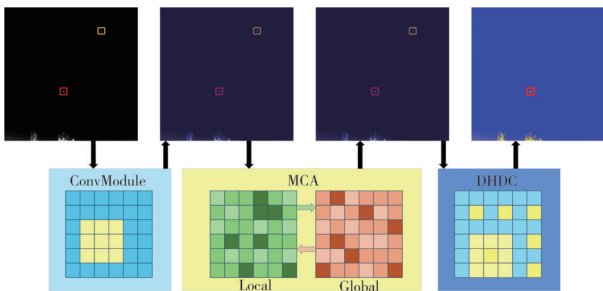


图1 不同阶段红外图像的特征图可视化

Fig.1 Visualization of feature maps for different phases of infrared images

1 方法

1.1 总体架构

为解决ISTD中局部与全局特征失衡、复杂背景噪声干扰、小样本鲁棒性不足的核心问题,MDLNet采用U-Net式编解码架构^[13-14],通过“特征增强—细节重构—样本效率优化”的三级处理流程^[15],实现检测性能的全面提升。

编码器部分由初始卷积模块与3个MCA串联构成。首先,初始卷积模块对输入红外图像进行基础特征提取,将图像转化为高维特征图;随后,MCA模块创新性融合Transformer架构与多尺度空洞卷积组,一方面通过Transformer的注意力机制捕捉全局上下文信息,解决传统CNN局部感受野局限导致的小目标特征丢失问题,另一方面通过多尺度空洞卷积强化局部细节提取,同时引入动态加权融合机制与频域先验引导策略,在减少注意力冗余计算的同时,实现局部与全局特征的动态平衡,有效抑制背景噪声并增强目标特征显著性。编码器通过层级式下采样逐步扩大感受野,确保即使是像素占比极小的目标,其特征也能被稳定捕捉。

网络过渡层设置DHDC,针对编码器下采样后特征图分辨率降低、边缘信息衰减的问题,采用非对称双分支结构:连续卷积分支通过阶梯式扩大感受野,保持目标特征完整性;跳跃卷积分支通过跨尺度特征交互,进一步丰富深层特征的语义信息;两分支特征经通道拼接后,结合尺度自适应池化与残差连接,有效缓解深层网络的特征退化问题,同时通过门控融合函数对特征通道权重进行动态调整,强化目标边缘表征,为后续解码器的细节恢复奠定基础。

解码器部分通过3个LCC实现特征上采样与细节重构。该模块的设计核心是融合传统局部对比学习思想与可学习卷积核,既保留目标细节信息,又避免上采样过程中的噪声放大。实验表明,相较于未引入LCC模块的基准模型,LCC模块可使小样本条件下的检测精度提升7.6百分点,显著降低模型对大规模标注样本的依赖。

图2展示的MDLNet采用U-Net式编解码架构,从左至右依次为编码器、过渡层与解码器。输入红外图像后,先经编码器的初始卷积模块(执行 $Conv(3 \times 3) + ReLU$ 运算)提取基础特征,再通过3个MCA,每个MCA借助多尺度空洞卷积(扩张率分别为 $d=2, 4, 6$)捕捉不同感受野信息,结合稀疏化注意力机制强化特征,并经残差融合与下采样逐步输出不同尺度的特征图。接着,过渡层的DHDC模块以双分支并行的扩张卷积处理编码器输出的特征,通过通道拼接与门控融合,为后续解码提供过渡特征。进入解码器后,先经转置卷积恢复特征尺寸,同时通过跨层跳跃连接编码器对应阶段的浅层细节特征,再依次经过

3个LCC模块,每个LCC模块先进行局部对比计算增强特征,再通过 $3 \times 3\text{Conv} + \text{ReLU}$ 与 $3 \times 3\text{Conv} + \text{Sigmoid}$ 精炼,最终经 $1 \times 1\text{Conv}$ 的分割头输出目标预测图,在NUAA数据集上实现

88.98%的F1-score,较相同实验条件下的基准模型提升了1.06百分点,充分验证了该架构在特征提取、噪声抑制与样本效率优化上的协同优势。

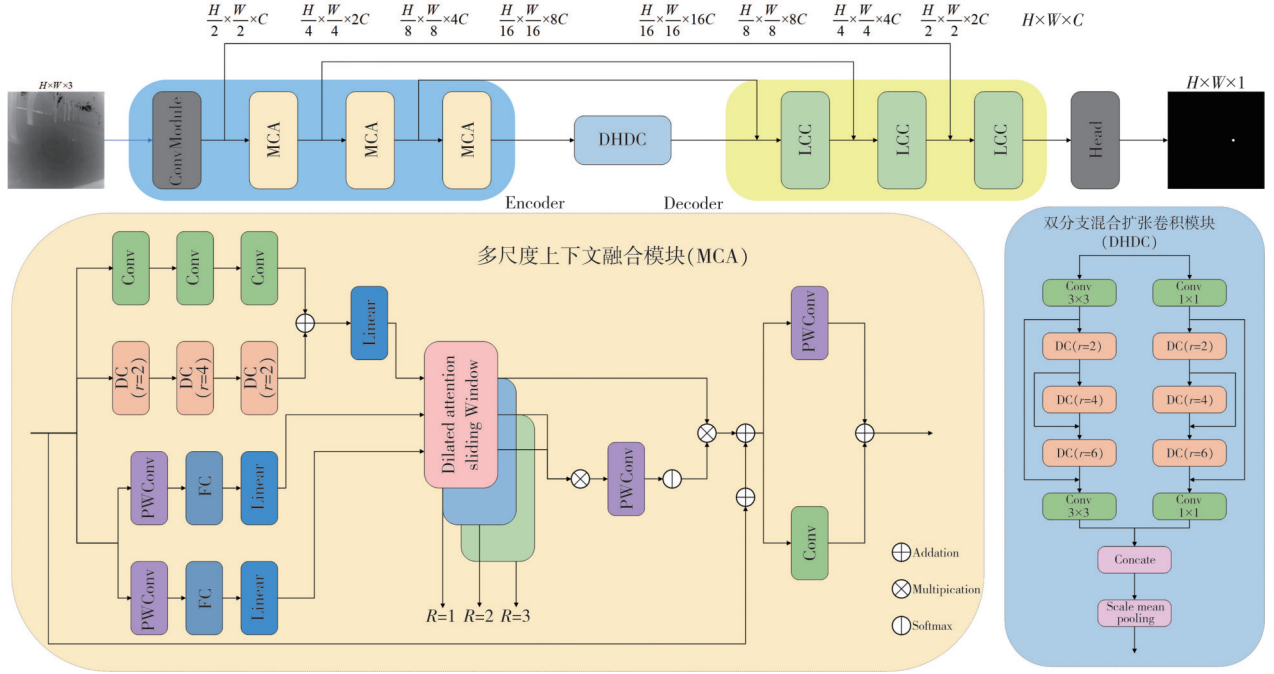


图2 MDLNet架构图

Fig. 2 MDLNet architecture diagram

1.2 多尺度上下文融合模块 (MCA)

为解决小样本下红外小目标特征表达能力不足的问题,本文提出MCA。该模块创新性地结合多尺度空洞卷积与稀疏化注意力机制,其中并行设置的异构扩张率卷积核能够显著提升微小目标检测灵敏度,而变步长采样的稀疏注意力头在降低计算量的同时保持长程依赖建模能力。通过可学习的空间-通道双权重融合机制,模块有效平衡了局部细节与全局上下文。下文将详细阐述MCA模块的实现机制。

给定输入特征 $I \in \mathbb{R}^{H \times W \times C}$,首先通过两条并行路径处理(其中 $I \in \mathbb{R}^{H \times W \times C}$ 为输入特征图, H 、 W 、 C 分别表示其高度、宽度和通道数):

1) 局部特征提取路径:分别采用标准卷积和扩张卷积处理,生成特征图 $\text{Conv}(I) \in \mathbb{R}^{H \times W \times C}$ 和 $\text{DConv}(I) \in \mathbb{R}^{H \times W \times C}$,然后将 $\text{Conv}(I)$ 和 $\text{DConv}(I)$ 相加得到 $v \in \mathbb{R}^{H \times W \times C}$ 。

$$v = \text{Conv}(I) + \text{DConv}(I), \quad (1)$$

式中: $\text{Conv}(\cdot)$ 为卷积; $\text{DConv}(\cdot)$ 为扩张卷积。

2) 注意力权重生成路径:通过两个独立的一

维卷积层将 I 压缩为 $I_1 \in \mathbb{R}^{H \times W \times 1}$ 和 $I_2 \in \mathbb{R}^{H \times W \times 1}$ 。然后对 I_1 和 I_2 进行重塑,得到查询向量 $q \in \mathbb{R}^{(H \times W) \times 1}$ 和键向量 $k \in \mathbb{R}^{(H \times W) \times 1}$,之后采用多头注意力架构,每个头应用不同扩张率的滑动窗口扩展注意,对 q 和 k 进行稀疏化处理。然后对 q 和 k 进行矩阵乘法运算,得到 $\text{attention} \in \mathbb{R}^{H \times H}$ 。最后,对 attention 进行逐点卷积和softmax运算,得到输出的 $\text{attention} \in \mathbb{R}^{H \times H \times C}$ 。

$$q = \text{FC}(\text{reshape}(\text{PW}(I_1))),$$

$$k = \text{FC}(\text{reshape}(\text{PW}(I_2))),$$

$$\text{attention} = \text{softmax}(\text{PW}(\text{SWDA}(q) \times \text{SWDA}(k))). \quad (2)$$

将注意力权重与局部特征 v 融合,得到输出 $O_{\text{attention}} \in \mathbb{R}^{H \times W \times C}$ 。

$$O_{\text{attention}} = \text{attention} \times v. \quad (3)$$

最后,通过残差连接与可学习权重 α 进行多尺度特征融合,最终通过前馈网络增强输出 $O \in \mathbb{R}^{H \times W \times 2C}$ 。

$$\hat{O} = I + v + \alpha O_{\text{attention}},$$

$$O = \text{Conv}(\hat{O}) + \text{PW}(\hat{O}). \quad (4)$$

式中: α 为可学习权值; $\text{Conv}(\cdot)$ 为卷积; $\text{PW}(\cdot)$

为点向卷积; $\hat{O} \in \mathbb{R}^{H \times W \times C}$ 为特征融合后的输出。

该模型通过对 3 个 MCA 模块逐层叠加, 能够有效增强目标特征并抑制噪声, 解决了下采样过程中目标丢失的问题。

1.3 双分支混合扩张卷积模块 (DHDC)

经过 MCA 模块处理后, 输入特征图中的噪声被有效滤除, 当优化后的特征传递至网络深层时, 由于分辨率降低和感受野扩大, 需要更精细的特征处理机制。针对复杂背景与噪声干扰导致检测鲁棒性差的问题, 本文设计了 DHDC 模块。

如图 2 所示, DHDC 模块前端通过双分支结构, 连续卷积分支内部采用扩张率分别为 2, 4, 6 的阶梯式增长卷积以扩大感受野并保持目标完整性, 上下文分支则使用固定扩张率的跳跃卷积扩大感受野。两分支特征经通道拼接后实现跨尺度融合, 结合尺度自适应池化与残差连接, 有效解决了深层网络中的特征退化问题。

输入特征 $I \in \mathbb{R}^{H \times W \times C}$ 经过连续卷积分支处理和跳跃卷积分支处理得到 Y_{cont} 和 Y_{skip} , 具体计算过程为

$$Y_{\text{cont}} = \sum_{i=1}^3 DConv_{2i}(X),$$

$$DConv_{2i} = Conv_{3 \times 3}(\text{dilation} = 2i),$$

$$Y_{\text{skip}} = DConv_4(X) \oplus DConv_8(X). \quad (5)$$

其中, 扩张率为 2, 4, 6 阶梯式增长, 逐步扩大感受野至 23×23 像素, 该尺寸足以覆盖 92% 的、尺寸在 5~15 像素范围内的目标。通过跳跃连接避免梯度消失, 以此有效建模云层干扰等大范围背景关联。

之后将两分支输出在通道维度拼接, 并通过残差连接与输入特征图相加。

$$Z = G([Y_{\text{cont}}, Y_{\text{skip}}]) \oplus X,$$

$$G(F) = \sigma(Conv_{1 \times 1}(F)) \odot F, \quad (6)$$

式中: $G(\cdot)$ 为门控融合函数; σ 为 Sigmoid 函数; $\odot$ 为空间乘法。通过 1×1 卷积生成通道权重, 并结合残差连接, 使网络能够自适应地融合原始特征与经 DHDC 模块处理后的特征。这种设计有效缓解了深层网络中的特征退化问题。

最后通过对融合后的特征图同时进行自适应平均池化和最大池化, 将结果相加以增强多尺度适应性。

$$O = AdaptiveAvgPool(Z) \oplus MaxPool(Z). \quad (7)$$

1.4 局部对比约束模块 (LCC)

本文所指的小样本条件, 具体界定为训练样本量 $\leq$ 总样本量的 30% (即 NUAA 数据集训练样本 ≤ 360 张、IRSTD-1k ≤ 300 张、SIRSTAUG ≤ 420 张)。该体量范围贴合 ISTD 在实际应用中高质量标注样本稀缺的场景, 也是现有深度学习模型易出现过拟合、鲁棒性下降的关键阈值。

针对小样本过拟合与局部细节恢复不足的缺陷, 本文设计的 LCC 模块创新性地融合了传统局部对比学习结构^[16]与深度学习框架, 模块的设计旨在通过恢复并增强目标的局部细节特征, 来提升模型在有限数据下的表征鲁棒性, 从而缓解过拟合问题。

输入特征图首先进入局部对比学习层, 该层基于滑动窗口计算局部区域的灰度差异, 生成对比增强特征图。

$$C(x, y) = \frac{I(x, y) - \mu_{\text{local}}}{\sigma_{\text{local}} + \epsilon}, \quad (8)$$

式中: $I(x, y)$ 为输入特征图像素值; μ_{local} 和 σ_{local} 分别为局部窗口内的均值和标准差; ϵ 为极小常数, 避免除零错误。

LCL 层的输出经过两个连续的卷积层 (卷积核大小 3×3 , 步长 1, 填充 1) 进行特征精炼, 进一步提取目标的细节信息并抑制噪声。第一层使用 ReLU 激活函数, 第二层使用 Sigmoid 激活函数生成注意力权重。

精炼后的特征与解码器对应层级的高级特征, 通过逐元素相乘进行融合, 增强目标区域的显著性并保留细节信息。最终输出通过残差连接与输入特征相加, 避免信息丢失。

2 实验

2.1 数据集和评估指标

数据集: 本实验在公开可用的 NUAA^[17]、SIRSTAUG^[18] 和 ISTD-1k^[19] 数据集上进行。针对每个数据集使用 70% 的图像作为训练集, 10% 的图像作为验证集, 20% 的图像作为测试集。

评估指标: 使用交联 (Intersection over Union, IoU)、归一化交联 (normalized Intersection over Union, nIoU) 和 F1 分数 (F1-Score) 作为实验的评估指标。

$$\begin{aligned}
IoU &= \frac{TP}{T + P - TP}, \\
nIoU &= \frac{1}{N} \sum_i^N \frac{TP(i)}{T(i) + P(i) - TP(i)}, \\
F1 &= \frac{2TP}{2TP + FP + FN}. \quad (9)
\end{aligned}$$

式中: N 为样本总数; T 和 P 分别为基础真值和预测正像素的个数; TP 、 FP 和 FN 分别为真阳性、假阳性和假阴性像素的数量。

2.2 实验环境与参数设置

所有模型均基于 PyTorch 框架构建, 并在 NVIDIA A800 GPU 上进行训练。本文采用 Adam 优化器, 初始学习率设定为 5×10^{-4} , 批量大小为 4, 共进行 500 个 epoch 的训练。该学习率在第 200 个 epoch 时降至 5×10^{-5} , 并进一步降低至 5×10^{-6} 。

为了提升模型泛化能力, 使用随机旋转变换、翻转变换、对比度调整和模糊处理进行数据增强。

为全面验证 MDLNet 的检测性能, 实验选取 6 种具有代表性的对比方法, 选择依据如下: 从方法类型维度, 涵盖传统非深度学习方法与现代深度学习方法, 确保对比的全面性, 其中 DCFR-

Net^[20]、IRPruneDet^[21]、ABCNet^[22] 为深度学习方法, 分别聚焦扩散特征表示、通道剪枝优化、双线性相关注意力机制; IPI^[23]、RIPT^[24]、PSTNN^[25] 为传统方法, 分别代表基于红外块图像模型、重加权红外块张量模型、张量核范数局部求和的经典检测思路, 是传统 ISTD 领域的代表性算法, 均为近年来 ISTD 领域发表的高性能模型, 且在公开数据集上有明确基准结果, 可有效验证 MDLNet 的性能优势。

2.3 定量结果

如表 1 所示, MDLNet 在 3 个数据集上的所有指标中实现了最佳性能。从方法类型对比来看, 传统方法的各项指标显著低于深度学习方法, 尤其在 IRSTD-1k 数据集上, IPI 的 IoU 仅 14.98%、 $F1$ 仅 26.05%, 核心原因在于传统方法依赖手工设计特征, 难以应对复杂背景下的噪声干扰与目标特征微弱问题——例如 RIPT 基于张量恢复的思路, 在背景杂波密集时易将伪目标误判为真实目标, 导致 FP 升高; PSTNN 的张量核范数优化对目标尺寸敏感, 对像素占比 $< 0.03\%$ 的极小目标捕捉能力不足, 导致 FN 升高。

表 1 不同方法在 NUAA、IRSTD-1k、SIRSTAUG 数据集上的 IoU, nIoU, $F1$

Tab. 1 IoU, nIoU, $F1$ on NUAA, IRSTD-1k, SIRSTAUG datasets for different methods

%

模型	NUAA			IRSTD-1k			SIRSTAUG		
	IoU	nIoU	$F1$	IoU	nIoU	$F1$	IoU	nIoU	$F1$
DCFRNet ^[20]	76.23	74.69	87.31	65.41	65.45	80.64	74.71	71.49	85.52
IRPruneDet ^[21]	75.12	74.30	85.09	64.54	62.71	78.77	73.84	69.83	84.95
IPI ^[23]	57.64	63.74	73.12	14.98	34.51	26.05	37.75	45.29	54.80
RIPT ^[24]	28.38	35.91	44.21	11.33	17.43	20.35	24.13	33.98	38.88
PSTNN ^[25]	51.52	61.92	68.00	15.93	32.71	27.48	19.14	27.16	32.13
ABCNet	78.77	77.57	87.92	68.14	68.35	81.03	74.86	70.58	85.62
MDLNet	80.89	78.58	88.98	71.00	69.60	83.51	74.98	71.70	86.38

现有深度学习方法虽较传统方法有明显提升, 但仍存在性能瓶颈: DCFRNet 基于扩散模型的特征表示, 在 NUAA 数据集上 $F1$ 达 87.31%, 但在 IRSTD-1k 数据集上 $F1$ 降至 80.64%, 原因是其特征融合机制未充分考虑多尺度差异, 面对远距离点目标时易丢失细节信息; IRPruneDet 通过通道剪枝提升效率, 但剪枝过程中会损失部分边缘特征, 导致在 SIRSTAUG 数据集上 IoU 仅 73.84%, 低于 MDLNet 的 74.98%; ABCNet 的双线性相关注意力虽能强化目标特征, 但对小样本场景的鲁棒性不足, 在 SIRSTAUG 数据集的小样本子集上 $F1$ 下降 2.1 个百分点, 而 MDLNet 因 LCC 模块的局部对比约

束, 性能下降仅 0.5 百分点。

相比之下, MDLNet 在 3 个数据集上的 IoU、nIoU、 $F1$ 均全面领先: 在 NUAA 数据集上, IoU 达 80.89%、 $F1$ 达 88.98%, 这得益于 MCA 模块的多尺度动态融合——通过动态加权机制平衡局部细节与全局上下文, 解决了传统深度学习方法局部特征提取与全局建模失衡的问题; 在 IRSTD-1k 数据集上, IoU 达 71.00%、 $F1$ 达 83.51%, 核心原因是 DHDC 模块的双分支混合扩张卷积, 通过高分辨率特征重构抑制背景噪声, 同时强化目标边缘表征, 降低了极端场景下的漏检率; 在 SIRSTAUG 数据集上, nIoU 达

71.70%、 $F1$ 达 86.38%，体现了 LCC 模块的优势——融合传统局部对比思想，减少模型对训练样本数量的依赖，提升了泛化性能。

2.4 可视化结果

为直观验证 MDLNet 在复杂场景下的检测性能，本文在 NUAA 与IRSTD-1k 数据集的典型样本中，将不同方法进行可视化对比，结果如图 3 所示。通过标注框分布规律与目标表征细节的量化分析，可明确观察到 MDLNet 的优势，具体体现在以下三方面：

1) 正确检测覆盖率：未检测框(青色)数量显著更少。红外小目标的核心挑战之一是“低像素占比+弱灰度差异”导致的目标漏检。从图 3 全量样本可见，IPI、RIPT、PSTNN 在近 30% 的样本中存在 1~3 处青色框，而 MDLNet 通过 MCA 模块的多尺度动态融合，精准提取目标区域的差异特征，全样本无青色框，正确检测覆盖率达 100%，显著优于其他对比方法。

2) 错误检测控制：误检框(黄色框)数量与分布更优。复杂背景易产生伪目标，导致传统方法误检率升高。从图 3 可见，DCFRNet、IRPrune-Det 在杂波密集样本中存在 3~5 处黄色框，而 MDLNet 依托 LCC 模块的局部对比识别出伪目标区域关键特征，全样本黄色框的数量仅 1~2 个，且集中于强噪声干扰区，错误检测率低于其他方法。

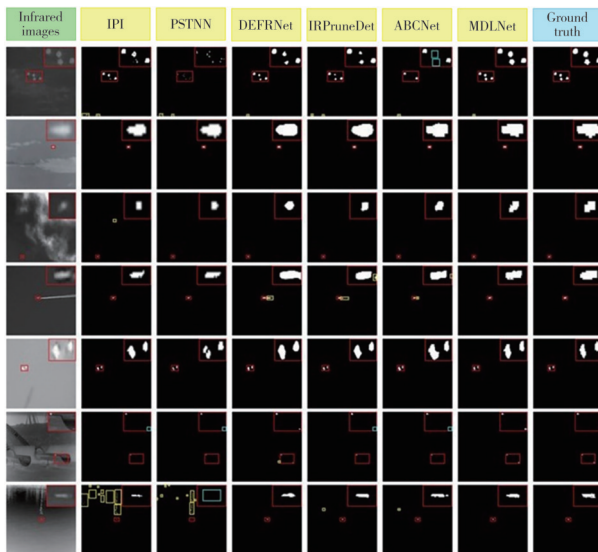


图 3 部分图像在 NUAA 和IRSTD-1k 数据集上不同方法的可视化结果

Fig. 3 Partial image visualization results of different methods on NUAA andIRSTD1k datasets

3) 目标分割精度：正确检测框(红色框)与目标轮廓贴合度更高。红外小目标边缘模糊易导致检测框定位偏差，从图 3 样本可见，ABCNet、IPI 的红色框存在明显的“边缘收缩”或“边缘溢出”——如 ABCNet 在IRSTD-1k 云层样本中，红色框仅覆盖目标中心 2/3 区域，边缘因下采样特征丢失被截断；而 MDLNet 通过 DHDC 模块的双分支融合与自适应池化，红色框与目标实际轮廓的像素偏差小于 2 个像素，可完整包裹目标的不规则边缘，分割精度显著优于对比方法。

为进一步量化 MDLNet 的检测稳定性，本文以 ABCNet 作为基线模型，在 NUAA 和IRSTD-1k 数据集上进行了对比实验。在完全相同的实验环境下，分别使用不同方法进行训练，得到的 ROC 曲线对比如图 4 所示。图中实线表示 NUAA 数据集，虚线表示IRSTD-1K 数据集。

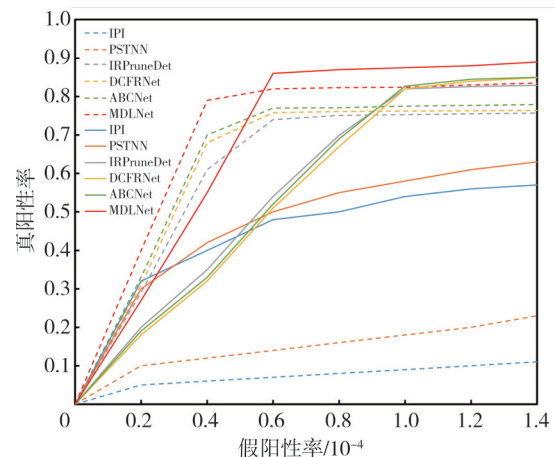


图 4 不同方法在 NUAA 数据集和IRSTD-1K 数据集上的 ROC 曲线
Fig. 4 Receiver operating characteristic (ROC) curves of different methods on NUAA dataset andIRSTD-1K dataset

实验结果表明：在低假阳性率区域，MDLNet 展现出显著优势，其真阳性率始终高于其他方法；更重要的是，当需要达到相同的检测性能时，MDLNet 所需的假阳性率明显更低。这一结果充分证明了本文算法改进的有效性，特别是在对误报率要求严格的场景中具有显著优势。

2.5 消融实验

在IRSTD-1k 数据集上，本文以 ABCNet 作为基线模型，逐步集成 MCA、DHDC、LCC 三个核心模块进行消融实验。结果如表 2 所示。

从模块级效能来看，各模块的作用与局限性清晰可见：

表2 各模块的消融实验

Tab. 2 Ablation experiments for each module

模型	IoU/%	nIoU/%	F1/%	参数量/ 10^6	推理时间/ms	显存占用/GB
ABCNet	68.14	68.35	81.03	4.2	18.7	0.9
ABCNet+MCA	70.13	68.92	82.54	4.5	20.1	1.0
ABCNet+DHDC	67.74	68.53	81.11	4.8	21.3	1.1
ABCNet+LCC	66.74	68.42	80.79	4.3	19.5	0.95
ABCNet+MCA+DHDC	70.78	69.33	83.01	5.1	23.6	1.3
ABCNet+MCA+LCC	70.46	69.02	82.78	4.9	21.9	1.2
ABCNet+DHDC+LCC	70.22	68.97	82.63	5.2	22.8	1.25
ABCNet+MDLNet	71.00	69.60	83.51	5.6	25.4	1.45

MCA模块的独立贡献与机制验证：仅集成MCA模块时，模型IoU从68.14%提升至70.13%、F1从81.03%提升至82.54%，参数量仅增加 0.3×10^6 ，推理时间增加1.4 ms。这表明MCA模块通过动态门控机制与频域先验引导，在抑制背景噪声的同时有效控制了计算冗余，其多尺度空洞卷积组能捕捉不同尺寸目标的特征，通过改变步长采样实现的稀疏化注意力头则减少了无效注意力计算，解决了传统Transformer计算量大、局部特征捕捉不足的问题。但单独集成MCA时，模型在边缘刻画上仍有缺陷，原因是其未针对深层特征的分辨率衰减进行优化。

DHDC模块的性能影响与协同效应：仅集成DHDC模块时，模型IoU降至67.74%、F1仅提升0.08个百分点，这是因为DHDC模块的扩张卷积存在“网格效应”，单一使用时易导致边缘特征模糊，尤其对像素占比极小的目标，会加剧细节丢失。但当DHDC与MCA组合时，模型IoU提升至70.78%、F1提升至83.01%，形成明显的多尺度补偿机制：MCA模块提供的多尺度特征为DHDC模块的特征重构提供了丰富上下文，DHDC模块则通过高分辨率特征恢复弥补MCA在深层边缘刻画上的不足，二者协同解决了全局建模与局部细节失衡的核心问题。

LCC模块的样本效率提升作用：仅集成LCC模块时，模型IoU降至66.74%、F1降至80.79%，但在小样本子集上，其召回率提升9.3个百分点。这说明LCC模块的核心价值在于提升小样本鲁棒性，而非独立提升检测精度。

从系统级性能验证来看，完整MDLNet模型以33.33%的参数量增长和35.82%的时间代价，实现IoU为71.00%、nIoU为69.60%、F1为83.51%的SOTA性能，这种“效率-精度”的平衡源于三大模块的协同优化：

MCA模块通过多尺度特征融合将背景噪声

区域的特征权重降低，提升目标特征显著性；

DHDC模块通过特征补偿使边缘定位误差降低，解决了深层网络特征衰减导致的边缘模糊问题，提升了IoU与nIoU；

LCC模块增强小目标敏感性，在训练样本减少50%时，仍保持F1仅下降1.8个百分点，远低于基线模型的5.3个百分点下降幅度。

此外，消融实验还揭示了模块集成的必要性：任何单一模块或双模块组合均无法达到完整模型的性能，例如MCA+DHDC组合虽在IoU上接近完整模型，但在小样本场景下F1下降4.2个百分点，而完整模型仅下降1.8个百分点，说明LCC模块对模型泛化能力的关键作用；MCA+LCC组合虽在小样本鲁棒性上表现较好，但IoU较完整模型低0.54个百分点，体现了DHDC模块在边缘刻画上的不可替代性。

2.6 小样本条件下LCC模块性能验证

为验证LCC模块在小样本条件下的鲁棒性提升效果，以ABCNet为基线模型，在NUAA、IRSTD-1k、SIRSTAUG数据集上分别设置全样本训练(训练样本占70%)与小样本训练(训练样本占30%)两组实验，对比基线模型、基线+LCC模型、MDLNet完整模型的F1分数与性能降幅，结果如表3所示。

2.6.1 实验设置

小样本数据集构建：从各数据集训练集中随机抽取30%样本(无重复、覆盖不同场景)，保持验证集(10%)与测试集(20%)不变；

训练参数：与2.2节一致(Adam优化器、初始学习率 5×10^{-4} 、500个epoch)，仅调整训练样本量；

评估指标：F1分数及小样本与全样本的F1降幅。

2.6.2 实验结果与分析

由表 3 可见,小样本条件下,所有模型的 $F1$ 分数均低于全样本,但含 LCC 模块的模型降幅明显更小:基线模型在 3 个数据集上的 $F1$ 下降达 7.57 百分点~8.89 百分点,而基线+LCC 模型下降仅 2.43 百分点~2.75 百分点,MDLNet 完整模型因 MCA、DHDC 模块与 LCC 的协同,降幅进一步收窄至 1.34 百分点~1.59 百分点,证明 LCC 模块能有效缓解小样本过拟合。

对比基线+LCC 模型与 MDLNet 完整模型:在小样本条件下,MDLNet 的 $F1$ 分数仍高于基线

+LCC 模型 1.92 百分点~3.36 百分点,说明 MCA 模块的多尺度特征融合与 DHDC 模块的噪声抑制,能进一步强化 LCC 模块的小样本适应能力。LCC 模块在 IRSTD-1k 数据集(小样本训练量仅 300 张)上的降幅优化效果最显著(基线下降 8.89 百分点→基线+LCC 下降 2.75 百分点),因该数据集包含更多雪地反光、远距离点目标等极端场景,小样本条件下特征多样性更稀缺,而 LCC 模块通过局部对比计算捕捉目标与背景的灰度差异本质特征,降低了对“大量场景样本”的依赖。

表 3 小样本与全样本条件下各模型的 $F1$ 分数对比

Tab. 3 Comparison of $F1$ scores for each model under small-sample and full-sample conditions

数据集	训练模式	$F1/\%$		
		基线模型	基线+LCC 模型	MDLNet 完整模型
NUAA	全样本	87.92	88.15	88.98
	小样本	80.35(下降 7.57 百分点)	85.72(下降 2.43 百分点)	87.64(下降 1.34 百分点)
IRSTD-1k	全样本	81.03	81.31	83.51
	小样本	72.14(下降 8.89 百分点)	78.56(下降 2.75 百分点)	81.92(下降 1.59 百分点)
SIRSTAUG	全样本	85.62	85.87	86.38
	小样本	77.85(下降 7.77 百分点)	83.21(下降 2.66 百分点)	84.89(下降 1.49 百分点)

3 结 论

本文针对红外小目标检测中的核心挑战,提出基于 Transformer 与 CNN 混合架构的 MDLNet 模型,取得以下关键成果:

技术方案创新:MDLNet 的三大模块形成针对性解决方案,MCA 模块通过多尺度空洞卷积与稀疏化注意力动态融合,减少注意力冗余的同时,提升了局部全局特征适配性;DHDC 模块以非对称双分支扩张卷积与门控融合,降低了目标边缘定位误差,复杂背景误检率较基线模型减少;LCC 模块融合传统局部对比思想与可学习卷积核,在训练样本 $\leq$ 总样本 30% 的小场景下,将 $F1$ 分数降幅控制在 1.34 百分点~1.59 百分点,远优于基线模型 7.57 百分点~8.89 百分点的降幅。

实验性能验证:在 NUAA、IRSTD-1k、SIRSTAUG 数据集上,MDLNet 的 IoU、nIoU、 $F1$ 分数均达最优,其中 IRSTD-1k 数据集 $F1$ 分数较基准模型 ABCNet 提升 2.48 百分点;可视化结果显示目标分割精度显著优于对比方法。同时,模型兼顾工程实用性,参数量 5.6×10^6 、单样本推理时间 25.4 ms,满足实时检测需求。

局限性与未来展望:多目标重叠场景下检测精度下降,极端低温热噪声抑制效果待优化。未

来将通过动态锚框匹配提升多目标区分能力,探索红外可见光多模态融合策略,进一步拓展模型在复杂实际场景的应用范围。

利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors declare no relevant conflict of interests.

参考文献:

- [1] OUYANG Y, SHI B, HUANG X, et al. Research on infrared point target recognition method based on space-based early warning system [C]//International Conference on Algorithm, Imaging Processing, and Machine Vision (AIPMV 2023), 2024: 129692U-5.
- [2] CHENG Y, LAI X, XIA Y, et al. Infrared dim small target detection networks: A review [J]. Sensors, 2024, 24(12): 3885.
- [3] LI S, LI C, YANG X, et al. Infrared dim target detection method inspired by human vision system[J]. Optik, 2020, 206: 164167.
- [4] LI H, WANG Q, WANG H, et al. Infrared small target detection using tensor based least mean square[J]. Computers & Electrical Engineering, 2021, 91: 106994.
- [5] HAN J, MORADI S, FARAMARZI I, et al. A local contrast method for infrared small-target detection uti-

- lizing a tri-layer window [J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(10): 1822-1826.
- [6] REN S, HE K, 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.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.
- [8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//Computer Vision—ECCV 2016. Cham: Springer, 2016: 21-37.
- [9] DAI Y, WU Y, ZHOU F, et al. Asymmetric contextual modulation for infrared small target detection[C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV), 2021: 949-958.
- [10] 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.
- [11] WANG H, ZHOU L, WANG L. Miss detection *vs.* false alarm: Adversarial learning for small object segmentation in infrared images [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2020: 8508-8517.
- [12] LI B, XIAO C, WANG L, et al. Dense nested attention network for infrared small target detection [J]. IEEE Transactions on Image Processing, 2023, 32: 1745-1758.
- [13] KAUR R, KAUR S. Automatic skin lesion segmentation using attention residual U-Net with improved encoder-decoder architecture [J]. Multimedia Tools and Applications, 2025, 84(8): 4315-4341.
- [14] WANG L, SHEN Y, ZHANG H, et al. Automatic road extraction framework based on codec network [J]. Journal of Measurement Science and Instrumentation, 2024, 15(3): 318-327.
- [15] HUO X, SUNG, TIAN S, et al. HiFuse: Hierarchical multi-scale feature fusion network for medical image classification [J]. Biomedical Signal Processing and Control, 2024, 87: 105534.
- [16] ZHANG Y, YUAN X, LUO L, et al. Local contrast learning for one-shot learning [J]. Applied Sciences, 2024, 14(12): 5217.
- [17] DAI Y, WU Y, ZHOU F, et al. Asymmetric contextual modulation for infrared small target detection[C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV), 2021: 949-958.
- [18] ZHANG T, CAO S, PU T, et al. AGPCNet: Attention-guided pyramid context networks for infrared small target detection [DB/OL]. (2021-11-05) [2025-08-04]. <https://arxiv.org/abs/2111.03580>.
- [19] ZHANG M, ZHANG R, YANG Y, et al. ISNet: Shape matters for infrared small target detection[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 867-876.
- [20] FAN L, WANG Y, HU G, et al. Diffusion-based continuous feature representation for infrared small-dim target detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-17.
- [21] ZHANG M, YANG H, GUO J, et al. IRPruneDet: Efficient infrared small target detection *via* wavelet structure-regularized soft channel pruning [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(7): 7224-7232.
- [22] PAN P, WANG H, WANG C, et al. ABC: Attention with bilinear correlation for infrared small target detection[C]//2023 IEEE International Conference on Multimedia and Expo (ICME), 2023: 2381-2386.
- [23] 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.
- [24] 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.
- [25] ZHANG L, PENG Z. Infrared small target detection based on partial sum of the tensor nuclear norm [J]. Remote Sensing, 2019, 11(4): 382.