引用格式：QIU Zhaobing, GUAN Liwen, WU Yu, et al. Lightweight Infrared Small Target Detection Method Based on Point-to-point Regression[J]. Acta Photonica Sinica, 2026, 55(2):0210002
裘兆炳,官丽雯,伍煜,等.基于点对点回归的轻量级红外小目标检测[J].光子学报,2026,55(2):0210002

基于点对点回归的轻量级红外小目标检测

裘兆炳[1],官丽雯[1],伍煜[2],陈丽琼[1],刘宇皓[1],吴光瑞[1],黄峰[1]
(1 福州大学 机械工程及其自动化学院,福州 350108)
(2 中国资源卫星应用中心,北京 100094)

摘 要:提出一种基于点对点回归的轻量级红外小目标检测算法,将红外小目标检测建模为简单的点对点回归任务,实现了目标检测性能和数据标注效率的最佳平衡。首先,提出了非对称全维度动态特征提取模块,构建通道混洗和深度卷积等结构,高效地增强目标关键信息,同时抑制背景干扰。然后,设计高分辨率跨特征增强模块,利用多路高分辨率特征提取分支的交叉融合,简化了传统的多层级卷积堆叠结构,有效挖掘高层语义特征的同时保留目标细节信息。最后,构建自适应点回归检测头,准确定位目标中心位置。大量实验结果表明,所提方法仅需单点数据标注,就能实现优于多种前沿算法的检测精度,同时模型参数量仅为0.24 M。

关键词:红外小目标检测;点对点回归;轻量级;深度学习;动态特征提取
中图分类号:TP391.4 **文献标识码**:A **doi**:10.3788/gzxb20265502.0210002

0 引言

红外小目标检测(Infrared Small Target Detection, IRSTD)作为红外搜索与跟踪系统的核心技术,在军事和民用领域均具有广泛的应用价值[1]。在军事领域,它可用于精确制导、战场侦察等关键任务;在民用领域,也被广泛应用于安防监控、智慧医疗等任务中[2-3]。然而,红外小目标在图像中所占比例极其微小,常常不足0.15%[4]。并且,这类目标缺乏清晰可辨的纹理和轮廓特征,在复杂背景的干扰下,它们极易被掩盖[5]。这些特性无疑给目标检测网络的设计带来了巨大的挑战。

过去几十年间,许多研究人员致力于红外小目标检测技术研究,并取得了令人瞩目的进展。目前,红外小目标检测方法可大致分为两类。第一类是基于模型驱动的算法,如核构面与随机游走(Facet Kernel and Random Walker, FKPW)[6]、红外图像块(Infrared Patch-Image, IPI)模型[7]、部分张量核范数和(Partial Sum of Tensor Nuclear Norm, PSTNN)[8]、全局稀疏度加权的局部对比度测量(Global Sparsity-Weighted Local Contrast Measure, GSWLCM)[9]、鲁棒无监督多特征表示(Robust Unsupervised Multifeature Representation, RUMFR)[10]等,主要通过建模目标和背景的特性差异,有效增强目标和抑制背景干扰。然而,这类方法通常依赖于人工经验设计,难以有效挖掘目标的深层语义信息,因此在复杂场景的表现不佳。

第二类是基于深度学习(Deep Learning, DL)的算法,例如注意力引导金字塔上下文网络(Attention-guided Pyramid Context Network, AGPCNet)[11]、非对称上下文调制(Asymmetric Contextual Modulation, ACM)网络[12]、高分辨率网络(High Resolution Network, HRNet)[13]、基于多尺度头部的U-Net网络(Multi-Scale Head to the plain U-Net, MSHNet)[14]、可学习的局部显著核网络(Learnable Local Saliency Kernel Network, L^{2}SKNet)[15]等。这类方法凭借其强大的表征能力,在检测准确性方面取得了显著的提升。根据检测方式不同,基于深度学习的算法可主要分为基于掩膜到掩膜回归(Mask-to-Mask Regression, M2MR)

基金项目:国家自然科学基金(62405060),福建省自然科学基金(2022J05113, 2024J01245),福建省中青年教师教育科研项目(JAT210035)
第一作者:裘兆炳,qiuzhaobing@fzu.edu.cn
通讯作者:陈丽琼,liqiongchen@fzu.edu.cn
收稿日期:2025-09-08;**录用日期**:2025-11-12

的方法和基于锚框到锚框回归(Box-to-Box Regression, B2BR)的方法。在M2MR的方法中,由于红外小目标尺寸小、边缘模糊,使得背景像素与目标像素的区分变得困难,从而增加了背景像素误分类的问题[16]。为增强特征区分能力,LI Boyang等[17]提出了密集嵌套注意力网络(Dense Nested Attention Network, DNANet),通过密集连接和空间金字塔融合来增强特征提取。WU Xin等[18]在U-Net骨干中嵌入一个小的U-Net,提出嵌套型U-Net(U-Net in U-Net, UIUNet)网络以实现目标的多级和多尺度学习。在B2BR的方法中,交并比(Intersection over Union, IoU)作为评估边界框回归精度的度量指标,其对边界框的位置变化极为敏感,即使发生轻微的边界框扰动,也可能引起IoU值的显著波动[19],不利于红外小目标稳定回归。为此,YANG Bo等人[20]提出增强特征学习网络(Enhancing Feature Learning Network, EFLNet),通过归一化高斯-沃瑟斯坦距离(Normalized Gaussian Wasserstein Distance, NWD)取代IoU作为新的度量指标,该方法有效缓解了边界框的敏感性问题,从而提升了红外小目标的检测性能。

综上所述,基于深度学习的方法虽然在复杂场景中的检测性能有显著提升,但仍面临着以下问题:1)需要依赖大量精确的标注数据,这会产生高昂的人工标注成本,而且标注过程中存在不确定性,影响实际目标检测性能;2)依赖更加复杂的模型结构提高检测精度,导致模型复杂度高、检测速度慢等问题[21],限制了其更加广泛的应用。

为了缓解传统标注方式所导致的高昂人工成本问题,研究人员提出了基于弱监督的深度学习(Weakly Supervised Deep Learning, WSDL)方法,这类方法通过单点标注生成伪掩码进行网络训练,从而降低标注成本。2024年,YING Xinyi等[22]提出了单点监督标签进化(Label Evolution with Single Point Supervision, LESPS)框架,在训练阶段过程中,通过迭代预测更新生成伪掩膜标签。2023年,LI Boyang等[23]提出了一种基于蒙特卡罗聚类(Monte Carlo Line Clustering, MCLC)的方法,重构出像素级的伪掩码。这些基于弱监督的方法可以与M2MR的方法结合,提升红外小目标检测的性能,如LESPS+DNANet[22]和MCLC+DNANet[23]。然而,伪掩膜的质量仍然不如精确的手动标注,导致检测精度显著降低,而且现有的基于M2MR的方式为了提高准确性,引入了不容忽视的参数开销[24]。大量的参数不仅极大地增加了存储需求,还显著提升了计算复杂度,使得模型在推理过程中需要消耗更多的计算资源与时间。

虽然基于弱监督的方法能降低人工标注成本,但其生成的标签精度有限,难以取得令人满意的检测性能。此外,弱监督技术的引入并不能简化模型结构,仍难以有效兼顾检测精度和模型复杂度。因此,如何实现高性能、低标注成本的轻量级红外小目标检测,仍是个亟待解决的技术难题[25]。

针对上述问题,本文提出了一种基于点对点回归的轻量级红外小目标检测网络(Point-to-Point Regression Lightweight Network, PPRL-Net),仅利用单点标注和单点预测就能实现高精度的目标检测。值得注意的是,提出的PPRL-Net将红外小目标检测建模为点对点回归(Point-to-Point Regression, P2PR)任务,它采用单点标注进行训练,通过回归高斯热图预测目标中心位置。不仅大幅度降低人工标注成本,而且实现了高效、高精度的红外小目标检测,并兼具参数量少、运算效率高等优点。

本文的主要内容包括以下三方面:1)提出了一种基于点对点回归的轻量级模型PPRL-Net,通过将红外小目标检测建模为简单的点对点回归任务,不仅能提高数据标注效率,而且能简化模型结构,实现目标检测精度、模型复杂度和数据标注效率的最佳平衡;2)设计了一种非对称全维度动态特征提取模块(Asymmetric Full-dimensional Dynamic Feature Extraction Module, AFDFEM),结合通道混洗、深度卷积和全维度动态卷积(Omni-dimensional Dynamic Convolution, ODConv)[26]技术,大幅降低模型复杂度的同时,有效增强目标关键信息。3)提出了一种高分辨率跨特征增强模块(High-resolution Cross-feature Enhancement Module, HRCFEM),通过多路高分辨率特征提取分支的交叉融合,替换传统冗余的卷积堆叠结构,有效挖掘高层语义特征的同时更好地保留目标空间细节信息。

在多个公开数据集上的实验结果验证了所提算法的有效性和先进性。大量实验表明,本文提出的PPRL-Net算法仅需单点标注进行训练,取得了优于多种前沿算法的检测性能。此外,该算法的参数量仅为0.24 M,相较于其他算法大幅降低了模型参数量和计算复杂度,且推理速度达到36 FPS,具有良好的实时性和广泛的应用场景。

1 本文算法

1.1 整体结构

图1展示了提出的PPRL-Net检测模型的整体架构。对于输入的红外图像，本文首先利用AFDFEM，有效挖掘红外小目标关键信息。随后，构建HRCFEM，通过多路高分辨率特征提取分支的交叉融合，进一步增强目标和抑制背景干扰。最后，构建自适应点回归检测头（Adaptive Point Regression Head，APRH），实现对红外小目标的精确定位。

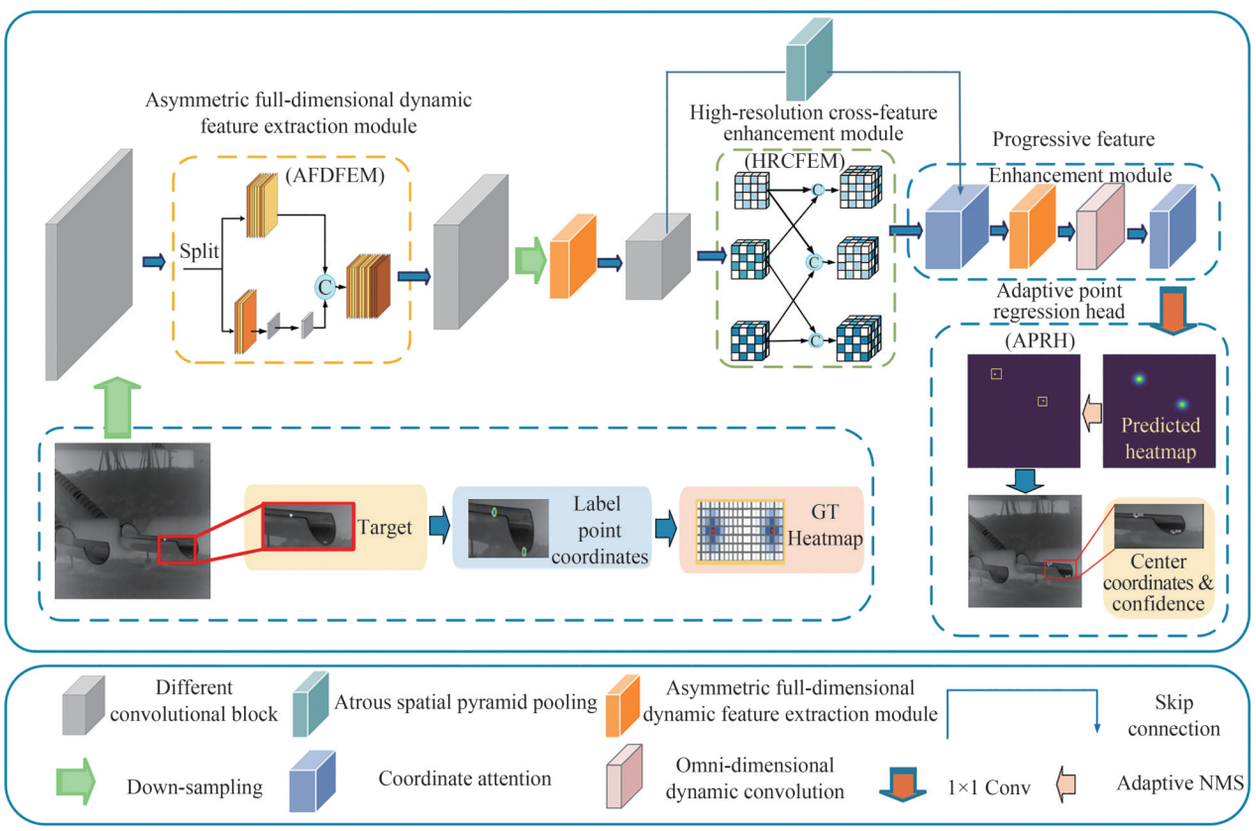


图1 PPRL-Net网络结构

Fig. 1 The structure of the PPRL-Net

首先，PPRL-Net模型通过两组“卷积层＋AFDFEM”作为颈部进行特征提取；其次，利用HRCFEM和空洞空间金字塔池化（Atrous Spatial Pyramid Pooling，ASPP）[27]模块，充分挖掘目标的高层语义和空间细节特征。然后，构建渐进式特征强化单元进一步增强目标和抑制背景干扰，该特征强化单元由坐标注意力（Coordinate Attention，CA）[28]模块、AFDFEM和ODConv组成。其中前置CA模块用于通道压缩与对齐，AFDFEM实现对局部细节的选择性增强，ODConv通过动态核机制提升特征表达灵活性，实现注意力引导和特征动态增强的渐进式特征优化。最后，输出特征经1×1卷积生成预测热图，并通过APRH实现目标的精确定位与置信度估计。

1.2 非对称全维度动态特征提取模块

为了在模型轻量化的同时有效增强目标关键信息，本文提出AFDFEM，其结构如图2所示。首先将输入特征图$\boldsymbol{X}\in \mathrm{R}^{B\times C\times H\times W}$经过通道混洗，按照比例$r$划分为两部分$\boldsymbol{X}_p$和$\boldsymbol{X}_i$

$$[\boldsymbol{X}_p,\boldsymbol{X}_i]=\mathrm{ChannelShuffle}(\boldsymbol{X}) \tag{1}$$

式中，ChannelShuffle(•)表示通道混洗操作，$\boldsymbol{X}_p\in \mathrm{R}^{B\times[C/r]\times H\times W}$表示特征增强部分，$\boldsymbol{X}_i\in \mathrm{R}^{B\times[C-C/r]\times H\times W}$表示恒等分支部分。

其次，特征增强通道$\boldsymbol{X}_p$依次通过深度卷积与空洞卷积，以获取重要通道局部特征和全局上下文信息，丰富特征表达能力

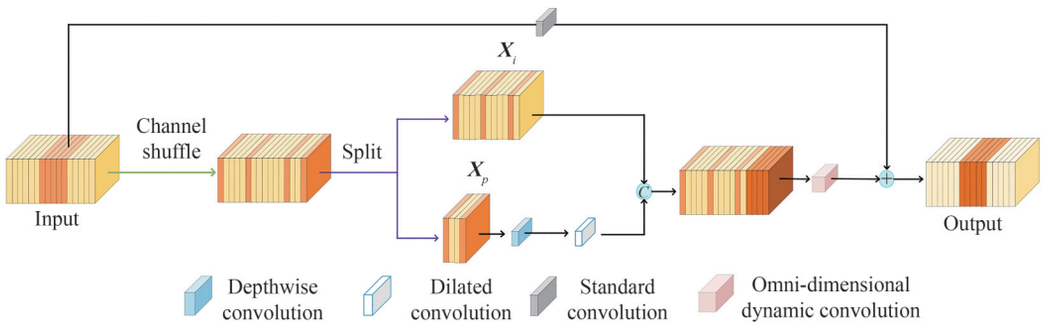


图 2 非对称全维度动态特征提取模块

Fig. 2 Asymmetric full-dimensional dynamic feature extraction module

$$\hat{\boldsymbol{X}}_p=\sigma(\mathrm{DConv}(\sigma(\mathrm{DWConv}(\boldsymbol{X}_p)))) \tag{2}$$

式中，DWConv(•)表示深度卷积，DConv(•)表示空洞卷积，其空洞率设置为2，σ表示ReLU激活函数。

通过上述操作，仅对部分通道进行特征增强，恒等分支保留原始语义信息，这样能有效兼顾特征提取精度和效率。然后得处理后的特征图为

$$\boldsymbol{U}=\mathrm{Concat}[\hat{\boldsymbol{X}}_p;\boldsymbol{X}_i](\boldsymbol{U}\in\mathrm{R}^{B\times C\times H\times W}) \tag{3}$$

式中，Concat[•]表示通道拼接操作。

然后，利用ODConv进一步增强目标多维度特征：

$$\boldsymbol{V}=\mathrm{ODConv}(\boldsymbol{U}) \tag{4}$$

式中，$\boldsymbol{V}$表示处理后的特征图，O DConv(•)表示全维度动态卷积操作。

最后，在输出端引入残差连接：

$$\boldsymbol{Y}=\sigma(\mathrm{BN}(\boldsymbol{V}+\mathrm{Res}(\boldsymbol{X}))) \tag{5}$$

式中，$\boldsymbol{Y}$表示最终输出的特征图，BN为批归一化操作，σ表示ReLU函数，Res(•)表示残差连接，定义为

$$\mathrm{Res}(\boldsymbol{X})=\begin{cases}\boldsymbol{X} & C_{\mathrm{input}}=C_{\mathrm{output}}\\ \mathrm{Conv}_2(\boldsymbol{X}) & \mathrm{Otherwise}\end{cases} \tag{6}$$

式中，Conv_2(•)表示具有BN与ReLU的卷积块，当输入与输出维度不一致时，通过1×1投影对齐。该设计在增强特征的同时保留输入信息，降低细节丢失风险，并改善梯度传播的稳定性。

1.3 高分辨率跨特征增强模块

为了更好地在挖掘高层语义特征的同时保留目标空间细节，本文提出HRCFEM。如图3所示，HRCFEM的特征图包含三路高分辨率特征提取分支，每一路分支在特征提取过程中保持高分辨率，以保留目标细节信息。

具体来说，设输入的特征为$\boldsymbol{X}\in\mathrm{R}^{B\times C\times H\times W}$。首先，将输入特征映射为三路多尺度特征$\{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3\}$。第一路经过ODConv[26]操作，提取细粒度空间特征$\boldsymbol{X}_1$：

$$\boldsymbol{X}_1=\mathrm{ODConv}(\boldsymbol{X}) \tag{7}$$

第二路通过分组卷积进行下采样，增强语义抽象能力，得到中层语义特征$\boldsymbol{X}_2$：

$$\boldsymbol{X}_2=\mathrm{GConv}_2(\boldsymbol{X}) \tag{8}$$

式中，GCon v_2(•)表示组别为1/2 C，步长为2，具有BN与ReLU的组卷积块。第三路则采用深度可分离空洞卷积(Depth-wise Separable and Dilated Convolutions，DSDC)[29]进行下采样，为后续融合提供高层语义特征，最终得到$\boldsymbol{X}_3$：

$$\boldsymbol{X}_3=\mathrm{Conv}_2(\mathrm{DSDC}(\mathrm{DSDC}(\boldsymbol{X}))) \tag{9}$$

式中，DSDC表示空洞率为2或4，具有BN与ReLU的深度可分离空洞卷积块。

然后，构建跨层特征融合结构，第一条分支$\boldsymbol{Y}_1$通过浅层特征$\boldsymbol{X}_1$下采样与中层特征$\boldsymbol{X}_2$在通道维度拼接，再结合全维度动态卷积提取多维度信息

$$\boldsymbol{Y}_1=\mathrm{ODConv}\left(\mathrm{Conv}_2\left(\mathrm{Concat}\left(\left[\mathrm{Down}(\boldsymbol{X}_1);\boldsymbol{X}_2\right]\right)\right)\right) \tag{10}$$

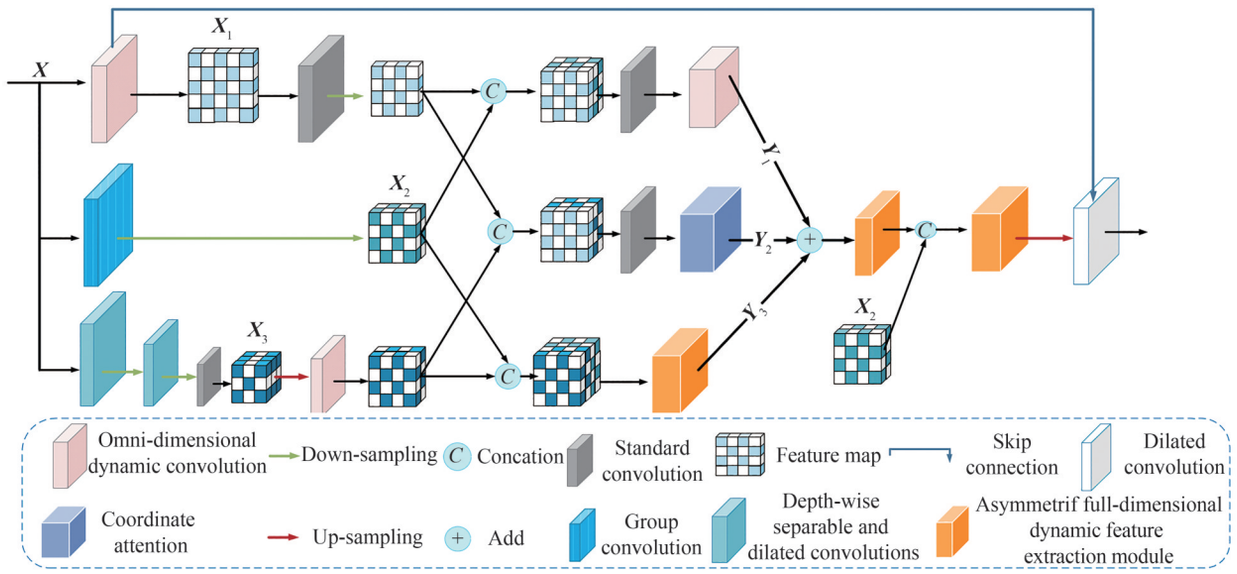


图3　高分辨率跨特征增强模块

Fig. 3　High-resolution cross-feature enhancement module

式中，Down(•)表示步长为2的卷积下采样操作。

分支$\boldsymbol{Y}_2$通过浅层特征$\boldsymbol{X}_1$下采样与深层特征$\boldsymbol{X}_3$上采样在通道维度拼接，并经卷积进行通道对齐，随后引入CA[28]注意力，提升模型对目标区域的精确定位与全局上下文建模能力

$$\boldsymbol{Y}_2=\mathrm{CA}(\mathrm{Conv}_2(\mathrm{Concat}[\mathrm{Down}(\boldsymbol{X}_1);\ \mathrm{ODConv}(\mathrm{Up}(\boldsymbol{X}_3))])) \tag{11}$$

式中，CA(•)表示坐标注意力模块，Up(•) 表示双插值上采样操作。

分支$\boldsymbol{Y}_3$通过浅层特征$\boldsymbol{X}_3$上采样与中层特征$\boldsymbol{X}_2$在通道维度拼接，使用AFDFEM增强关键信息并调整通道

$$\boldsymbol{Y}_3=\mathrm{AFDFEM}(\mathrm{Concat}[\boldsymbol{X}_2;\ \mathrm{ODConv}(\mathrm{Up}(\boldsymbol{X}_3))]) \tag{12}$$

此外，为减少简单加权带来的冗余，实现不同分支贡献的动态平衡，引入一组可学习的融合权重α_i，得到最终融合特征

$$\boldsymbol{F}=\alpha_1\boldsymbol{Y}_1+\alpha_2\boldsymbol{Y}_2+\alpha_3\boldsymbol{Y}_3 \tag{13}$$

最后，融合特征首先输入至AFDFEM，以抑制冗余并增强判别性；随后与中层特征$\boldsymbol{X}_1$拼接，并通过另一个AFDFEM完成通道对齐与重组，从而保证融合的维度一致性与表达充分性；接着，特征上采样至高分辨率，并与浅层特征$\boldsymbol{X}_0$残差相加，在增强高层语义表征的同时保留细粒度空间细节；最后，引入空洞率为2的卷积，在保持分辨率不变的前提下有效扩大感受野，进一步整合局部与上下文信息，获得最终的高分辨率输出特征。

1.4　自适应点回归检测头

检测头部分使用3×3最大池化生成预测热图，通过预测热图的局部极大值特性，结合非极大值抑制和置信度阈值约束，逐步筛选出最具代表性的候选点，并映射回原图坐标，实现小目标的准确定位。

自适应点回归检测头包括三个步骤：最大池化识别局部最大值、更新热图仅保留局部最大值以及选择自适应点。首先使用3×3最大池化识别预测热图的局部最大值，通过保留下的局部最大值更新每个坐标(x_i,y_i)处的热图，更新后的热图$\boldsymbol{H}_{\mathrm{NMS}}$表示为

$$\boldsymbol{H}_{\mathrm{NMS}}=\begin{cases}H(x_i,y_i), & \mathrm{Maxpool}(H)=H(x_i,y_i)\\ 0, & \mathrm{Otherwise}\end{cases} \tag{14}$$

选取最高的置信度分数$(x_{\max},\ y_{\max})$作为基准点，并且计算其他候选点的置信度是否超过一定阈值，如果置信度超过一定阈值，则将该点作为背景点进行抑制；否则，即认为有效目标点。其过程可表示为

$$H_{\mathrm{NMS}}(x_i,y_i)\geqslant\lambda\times H_{\mathrm{NMS}}(x_{\max},y_{\max}) \tag{15}$$

式中，λ表示用于调整背景抑制能力的超参数。

1.5　训练和损失函数

为了降低人工标注成本，本文设计了PPRL-Net，使用单点标注训练，并使用均方误差(Mean Squared

Error, MSE)作为损失函数,实现目标的点对点回归与定位。

(1)训练过程:PPRL-Net的训练过程是通过单点标注作为监督信号,为了增强PPRL-Net在训练过程的模型收敛和目标特征学习能力,以单点标注为中心并向外辐射,生成高斯分布的目标概率热图,目标质心位置附近的概率趋于1,而距离较远的概率趋于0,其中高斯分布定义为

$$G(x,y)=\mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}} \tag{16}$$

式中,x和y分别表示距分布中心的水平与垂直距离,σ表示控制高斯核扩散的标准差。将坐标为(x_i,y_i)处的像素点热图概率定义为

$$\hat{H}(x_i,y_i)=G(x_i-kx_k,y_i-ky_k) \tag{17}$$

式中,$G(x,y)$为式(16)定义的高斯核函数,(kx_k,ky_k)表示第k个目标点的坐标(单个目标$k=1$),$\hat{H}(x_i,y_i)$表示真实目标热图在(x_i,y_i)处的值。

(2)损失函数:为了引导网络专注于目标的中心并提升位置学习能力,本文使用预测热图与目标概率热图之间的MSE作为损失,其中L_{MSE}损失定义为

$$L_{\mathrm{MSE}}=\frac{1}{n}\sum_{x_i=1}^{W}\sum_{y_i=1}^{H}(H(x_i,y_i)-\hat{H}(x_i,y_i))^2 \tag{18}$$

式中,$H(x_i,y_i)$表示(x_i,y_i)处的预测热图,$\hat{H}(x_i,y_i)$表示(x_i,y_i)处的真实目标概率热图,W和H分别表示热图的宽度与高度,$H(x_i,y_i)$的定义式与式(14)一致。

2 实验及结果

2.1 实验细节

实验选取两个公开数据集NUDT-STRST[17]和IRSTD-1K[30]。为提供公平的实验环境,将每个数据集重新划分为训练集、验证集与测试集,划分比例为6∶2∶2。在实验过程中,所有方法均使用PyTorch深度学习框架,实验平台为英特尔酷睿i9-14900HX CPU和NVIDIA GeForce RTX 4070 GPU的计算机上进行评估。

在训练过程中采用Adam优化器,初始学习率设为0.001,并使用ReduceLROnPlateau策略进行动态调整。衰减因子为0.1,耐心阈值(Patience)设为3,批量大小(Batch Size)设为5,训练轮次(Epoch)设为300。为充分验证本文方法的有效性与先进性,选取多种先进算法进行对比,包括模型驱动方法,如FKPW[6]、IPI[7]、PSTNN[8]、GSWLCM[9]、RUMFR[10];以及近年来提出的深度学习方法,如AGPCNet[11]、ACM[12]、DNANet[17]、UIUNet[18]、EFLNet[20]、HRNet[13]、MSHNet[14]、L^{2}SKNet[15];同时为了验证点对点回归的有效性,还选取了基于弱监督的深度学习方法进行比较,如LESPS+DNANet[22]和MCLC+DNANet[23]。对比算法中所有参数和配置均采用对应论文中推荐或最优值,具体设置参数如表1所示。

表1 对比算法的实验参数设置

Table 1 Experimental parameter settings of the comparison algorithms

Category	Methods	Experimental parameter settings
Model-Based	FKPW[6](2019)	$K=5$, $p=6$, $\beta=200$, Window size: 11×11
	IPI[7] (2013)	Patch size: 50×50, stride: 20
	PSTNN[8] (2019)	Patch size: 40×40, step: 40
	GSWLCM[9] (2022)	$\delta=0.01$, $\lambda=0.6$, $k=20$, Window size: 19×19
	RUMFR[10] (2024)	$\delta_1=0.8$, $\delta_2=0.3$, $N_\varepsilon=30$, $k=30$, $\sigma=0.6$, window size: 23×23
WSDL-Based	LESPS+DNANet[22] (2024)	Learning rate: 0.05, batch size: 16, epoch: 1 500
	MCLC+DNANet[23] (2023)	Learning rate: 0.05, batch size: 16, epoch:1 500
DL-Based(B2BR)	EFLNet[20] (2024)	Learning rate: 0.001, batch size: 8, epoch:600
	UIUNet[18](2023)	Learning rate: 0.05, batch size: 3, epoch:1 000
DL-Based(M2MR)	AGPCNet[11] (2023)	Learning rate: 0.05, batch size: 8, epoch:10
	DNANet[17](2023)	Learning rate: 0.05, batch size: 16, epoch:1 500
	MSHNet[14] (2024)	Learning rate: 0.05, batch size: 4, epoch:400
	L^{2}SKNet[15] (2025)	Learning rate: 0.000 5, batch size: 8, epoch:400
DL-Based(P2PR)	ACM[12] (2021)	Learning rate: 0.05, batch size: 8, epoch:500
	HRNet[13] (2019)	Learning rate: 0.001, batch size: 16, epoch:140

2.2 评价指标

由于红外搜索和跟踪系统最终需要获得IRSTD精确的坐标[19]，采用目标级别评估指标，包括精确率（Precision，Pre）、召回率（Recall，Rec）、F1-measure（F1）、每秒检测图像的帧数（Frame Per Second，FPS）、运算浮点数（FLOPs）和参数量（Params），其中准确率表示所有被预测为阳性的样本中，真正为阳性的样本所占的比例

$$\mathrm{Pre}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{19}$$

式中，TP表示真阳性，FP表示假阳性（即被错误地识别为阳性的阴性样本的数量）。召回率表示正确识别的样本在所有实际阳性样本中的比例

$$\mathrm{Rec}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{20}$$

式中，FN表示假阴性（即未被模型识别的阳性样本的数量）。

F1是对目标检测性能的综合评价，定义为

$$\mathrm{F1}=2\times\frac{\mathrm{Pre}\times\mathrm{Rec}}{\mathrm{Pre}+\mathrm{Rec}} \tag{21}$$

式中，F1∈[0，1]，F1越高，表示算法的综合检测性能越好。此外，选择FPS衡量算法的检测速度，FLOPs用于衡量模型的计算复杂度，Params用于衡量模型存储复杂度。

2.3 消融实验

为验证各模块对红外小目标检测性能的作用，本文在IRSTD-1K数据集上，选择同时去除HRCFEM与AFDFEM的网络作为基线，进行消融实验，结果如表2所示。由实验结果可见：在仅加入HRCFEM时，模型参数量增加0.18 M，但精度Pre提升2.53%，召回率Rec提升1.38%，F1值提升1.92%；在仅加入AFDFEM时，参数量几乎保持不变，Pre提升1.04%。当两者同时引入时，Pre、Rec和F1分别提升2.57%、2.07%和2.31%。虽然此时参数量较基线有少量增加，但整体仍为轻量级的网络。综上所述，HRCFEM与AFDFEM在保证轻量级的同时均能有效提升检测性能，并且二者的联合使用能进一步增强模型的综合表现，从而验证了所提模块对整体模型性能的提升。

表2 IRSTD-1K数据集消融实验结果

Table 2 IRSTD-1K dataset ablation experiment result

Baseline	HRCFEM	AFDFEM	Pre/%	Rec/%	F1/%	Params(M)	FLOPs(G)
√			92.62	86.55	89.48	**0.05**	**0.71**
√	√		95.15	87.93	91.40	0.23	1.48
√		√	93.66	86.55	89.96	0.06	0.83
√	√	√	**95.19**	**88.62**	**91.79**	0.24	1.53

2.4 对比实验

为验证本文方法的有效性与先进性，分别在NUDT-STRST和IRSTD-1K两个数据集上与15种代表性算法进行对比实验。表3展示了不同算法检测性能的定量对比结果，包括Pre、Rec、F1、FLOPs、Params和FPS，表中最好的指标用加粗表示，次优的指标用下划线表示。

1）计算复杂度分析。表3中给出了不同算法输入相同尺寸（512×512）的模型规模与推理速度方面的比较结果。PPRL-Net与其他15种前沿算法相比，参数量最低，仅有0.24 M，且保持较低的计算复杂度（1.53 GFLOPs），在保持轻量级设计的同时提高计算效率；在推理速度方面，本文方法达到了36 FPS，满足实时的检测效果。

2）检测性能分析。表3给出了不同算法在两个公开数据集上的性能对比结果。从表3中可知，模型驱动的算法普遍依赖手工特征设计与人工参数调节，在复杂检测场景中容易产生大量误检与漏检，整体检测性能显著低于基于深度学习的方法。

与现有基于深度学习的方法相比，PPRL-Net在检测精度与模型复杂度之间取得了最佳平衡。具体而言，模型参数量最低的PPRL-Net，在IRSTD-1K数据集上取得了最优的检测精度（95.19%）和次优的F1指

表3 不同算法检测性能的定量对比

Table 3 Quantitative comparison of detection performance across different algorithms

Category	Methods	IRSTD-1K			NUDT-STRST			Params(M)	FLOPs(G)	FPS
		Pre/%	Rec/%	F1/%	Pre/%	Rec/%	F1/%			
Model-Based	FKPW[6]	33.22	63.64	43.65	38.45	67.61	49.02	—	—	7.49
	IPI[7]	44.76	65.31	53.11	35.28	80.20	49.00	—	—	1.09
	PSTNN[8]	44.53	60.20	51.19	39.44	70.88	50.68	—	—	3.33
	GSWLCM[9]	62.68	61.38	62.02	55.87	62.63	59.06	—	—	4.06
	RUMFR[10]	45.00	58.56	50.89	58.82	67.36	62.80	—	—	3.25
WSDL-Based	LESPS+DNANet[22]	91.05	80.69	85.56	95.49	96.77	96.13	4.70	56.33	4.70
	MCLC+DNANet[23]	88.05	82.96	85.43	97.61	95.84	96.71	4.70	56.33	4.70
DL-Based(B2BR)	EFLNet[20]	92.12	**92.79**	**92.44**	94.33	96.32	95.31	33.20	33.29	20.48
DL-Based(M2MR)	UIUNet[18]	85.91	87.37	86.63	91.97	97.93	94.86	50.54	217.84	6.76
	AGPCNet[11]	77.48	87.46	82.17	95.95	97.68	96.81	12.43	173.02	3.04
	DNANet[17]	87.76	88.97	88.36	98.16	96.14	97.14	4.70	56.33	4.70
	MSHNet[14]	87.93	88.54	88.23	91.60	94.74	93.14	4.07	24.45	22.25
	L^2SKNet[15]	81.17	89.76	85.25	95.15	98.16	96.63	0.899	27.58	**54.76**
DL-Based(P2PR)	ACM[12]	88.03	85.32	86.66	89.38	68.50	77.65	0.39	**1.13**	27.8
	HRNet[13]	92.54	84.35	88.26	96.32	93.04	94.65	28.54	41.18	16.66
	PPRL-Net(Ours)	**95.19**	88.62	91.79	**99.72**	95.26	**97.44**	**0.24**	1.53	36.35

标(91.79%),意味着良好的检测性能。虽然PPRL-Net在召回率指标上有所下降,但仍优于MSHNet、UIUNet等大部分算法。其主要原因是PPRL-Net只利用单点标签进行训练,虽然能大幅提升标注效率,但由于无法提供更多目标尺寸和外观轮廓信息,使得目标表征和检测困难。此外,PPRL-Net在NUDT-SIRST数据集上虽然在召回率上略低于最高的L^2SKNet,但具有最优的检测精度(99.72%)、F1指标(97.44%)和最低的模型参数量(0.24 M),表明算法取得了更好的综合性能。

值得强调的是,PPRL-Net是一个极轻量的网络,与现有的轻量级模型ACM[12]相比,PPRL-Net在NUDT-SIRST数据集和IRSTD-1K数据集上F1分数分别提升了19.79%和5.13%。这些结果充分证明,PPRL-Net能够在保持低参数量的前提下,显著提升模型的检测精度,实现了参数量与性能之间的更优平衡。

此外,PPRL-Net采用单点标注进行训练,与LESP+DNANet[22]和MCLC+DNANet[23]等依赖复杂前处理生成近似伪掩码的弱监督方法不同,PPEL-Net通过回归生成高斯热图作为监督信号进行训练的方式,避免了因掩膜生成不准确而导致的性能上限限制。因此,相同的标注成本下,PPRL-Net具有更强的特征学习能力并取得了更高的检测精度。

3)综合性分析。为进一步直观展现不同方法在检测精度与计算复杂度之间的平衡关系,本文基于IRSTD-1K数据集的实验结果,分别绘制F1 Vs FLOPs、F1 Vs Params、Pre Vs Params和Rec Vs Params的二维对比散点图,如图4所示。结果表明,本文方法在保证检测精度的同时,显著降低了计算与存储开销,取得了最优的精度和效率平衡。

4)可视化对比。为了直观的比较不同算法的检测情况,本文分别在IRSTD-1K和NUDT-STRST两个数据集上与其他方法进行对比实验,检测结果如图5所示。在Groundtruth中,绿色十字点表示单点标注的目标位置;在检测结果中,红色框表示正确检测的目标,黄色圆圈表示漏检,蓝色框表示虚警。从图中可以看出,深度学习方法的检测结果普遍优于传统方法,取得了更好的检测性能,并有效降低了漏检与虚警。但在目标边缘模糊或对比度较低时(如图5(b)所示),仍存在一定的漏检。相比之下,本文提出的方法在不同场景下均能够准确检测出小目标,表现出更强的稳定性与准确性,不仅能精准检测目标,还能输出较高的置信度结果。

综上,所提算法仅需单点标注,就实现了优于多种前沿算法的检测性能,而且参数量仅为0.24 M,推理速度达到了36 FPS,取得了检测精度和效率的最佳平衡。

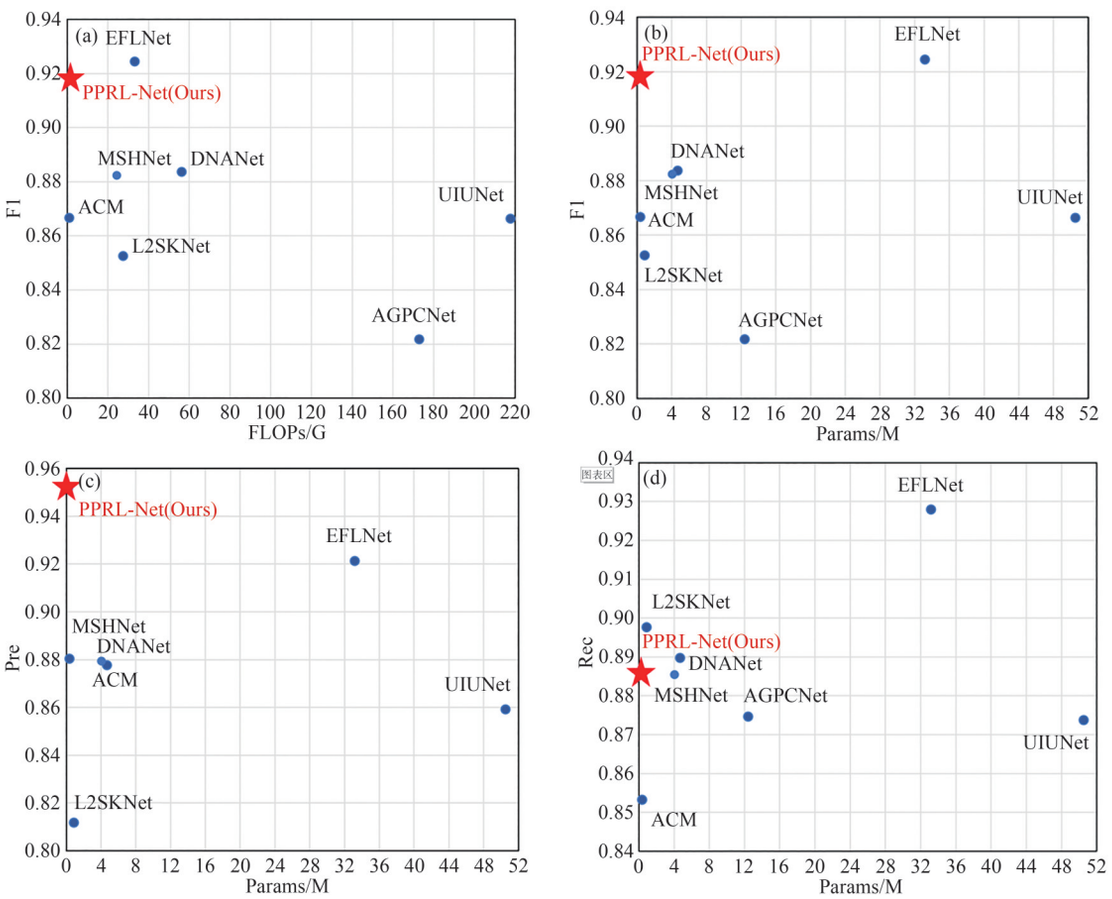
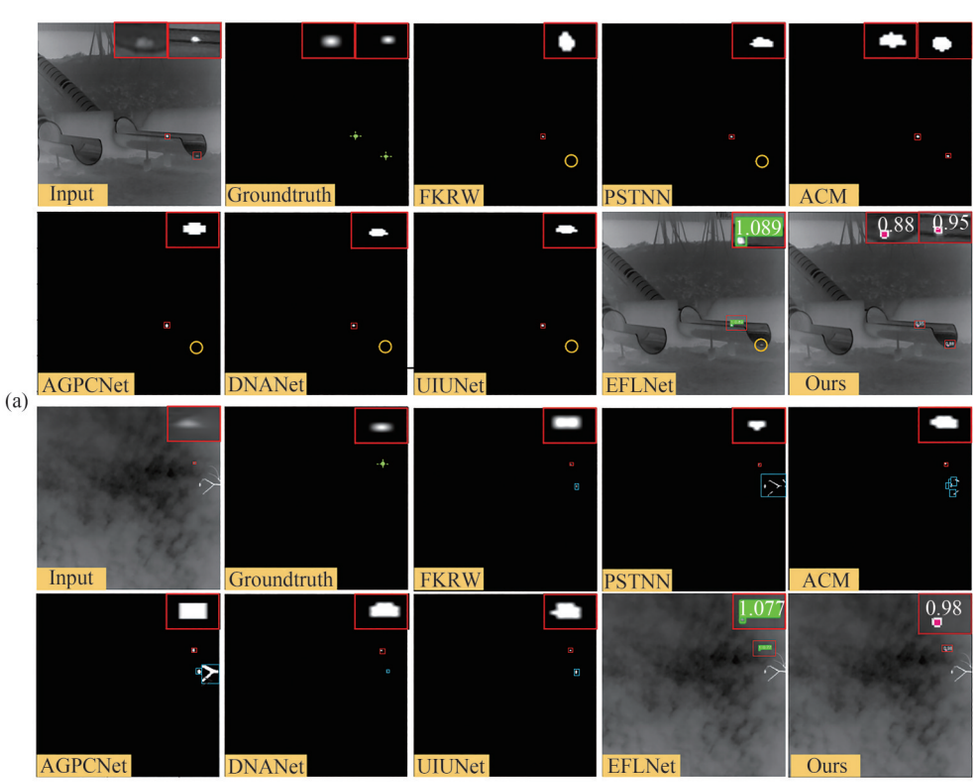


图4　不同算法在IRSTD-1K上的检测精度和计算复杂度对比。(a) F1 vs FLOPs;(b) F1 vs Params;(c) Pre vs Params;(d) Rec vs Params

Fig. 4　Comparison of detection accuracy and computational complexity of different algorithms on IRSTD-1K. (a) F1 vs FLOPs;(b) F1 vs Params;(c) Pre vs Params;(d) Rec vs Params



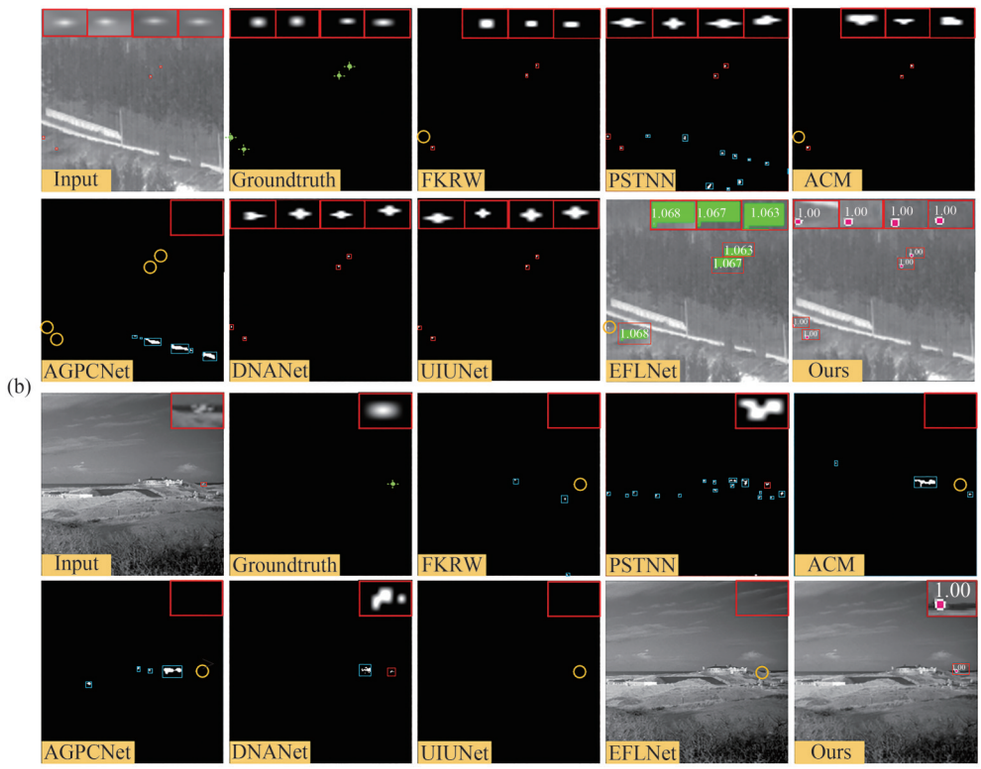


图5 不同算法的检测结果对比(绿色十字表示Groundtruth中的单点标注,红色框表示正确检测的目标,黄色圆圈表示漏检,蓝色框表示虚警)。(a)IRSTD-1K dataset;(b)NUDT-SIRST dataset

Fig. 5 Comparison of detection results across different algorithms(Green crosses indicate ground-truth single-point annotations; red boxes denote correct detections; yellow circles mark missed detections; blue boxes show false alarms). (a) IRSTD-1K dataset; (b)NUDT-SIRST dataset

3 结论

针对红外小目标像素数量少,纹理细节稀缺且依赖高昂的人工标注等问题,本文提出了一种基于点对点回归轻量级红外小目标检测算法PPRL-Net,仅需单点标注就能实现高精度的目标检测。首先,提出了AFDFEM模块从多维度动态提取重要特征,有效增强红外小目标;然后设计HRCFEM模块,通过多路高分辨率特征提取分支的交叉融合,充分挖掘目标高层语义和空间细节特征;最后,构建APRH,准确定位目标位置。大量实验结果表明,所提算法实现了目标检测精度、模型复杂度和数据标注效率的最佳平衡。目前PPRL-Net未充分考虑集群多目标检测中的解混问题,导致一些空间邻近目标的漏检。未来的工作将进一步研究基于点对点回归的集群多目标检测模型,进一步提升复杂场景检测性能。

参考文献

[1] ZHANG Ke, NI Shuyan, YAN Dashuang, et al. Review of dim small target detection algorithms in single-frame infrared images[C]. IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), IEEE, 2021, 4: 2115-2120.

[2] WANG Guogang, SUN Zhaojin, LIU Yunpeng. J-MSF: a new infrared dim and small target detection algorithm based on multi-channel and multiscale[J]. Infrared and Laser Engineering, 2022, 51(5): 136-145.
王国刚,孙召进,刘云鹏. J-MSF:一种新的多通道多尺度红外弱小目标检测算法[J]. 红外与激光工程, 2022, 51(5): 136-145.

[3] WEI Shuigen, WANG Chengwei, CHEN Zhen, et al. Infrared dim target detection based on human visual mechanism[J]. Acta Photonica Sinica, 2021, 50(1): 0110001.
危水根, 王程伟, 陈震, 等. 基于视觉注意机制的红外弱小目标检测[J]. 光子学报, 2021, 50(1): 0110001.

[4] WEI Zhang, CONG Mingyu, WANG Liping. Algorithms for optical weak small targets detection and tracking: review[C]. International Conference on Neural Networks and Signal Processing, 2003, 1: 643-647.

[5] LI Hui, LI Zhengzhou, YANG Yuxin, et al. Infrared weak target detection method based on cross connection and fusion attention mechanism[J]. Acta Photonica Sinica, 2024, 53(9): 0910002.
李慧, 李正周, 杨雨昕, 等. 基于跨越连接与融合注意力机制的红外弱小目标检测方法[J]. 光子学报, 2024, 53(9):

0910002.
[6] YAO Qin, LORENZO B, GAO Chengqing, et al. Infrared small target detection based on facet kernel and random walker [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(9): 7104-7118.
[7] GAO Chenqing, MENG Deyu, YANG Yi, 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 Landan, PENG Zhenming. Infrared small target detection based on partial sum of the tensor nuclear norm [J]. Remote Sensing, 2019, 11(4): 382-382.
[9] QIU Zhaobing, MA Yong, FAN Fan, et al. Global sparsity-weighted local contrast measure for infrared small target detection[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.
[10] CHEN Liqiong, WU Tong, ZHENG Shuyuan, et al. Robust unsupervised multifeature representation for infrared small target detection [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17: 10306-10323.
[11] ZHANG Tianfang, LI Li, CAO Siying, et al. Attention-guided pyramid context networks for detecting infrared small target under complex background[J]. IEEE Transactions on Aerospace and Electronic Systems, 2023, 59(4): 4250-4261.
[12] DAI Yimian, WU Yiquan, ZHOU Fei, et al. Asymmetric contextual modulation for infrared small target detection [C]. IEEE/CVF Winter Conference on Computer Vision and Pattern Recognition(CVPR), 2021: 950-959.
[13] SUN Ke, XIAO Bin, LIU Dong, et al. Deep high-resolution representation learning for human pose estimation [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2019: 5693-5703.
[14] LIU Qiankun, LIU Rui, ZHENG Bolun, et al. Infrared small target detection with scale and location sensitivity [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2024: 17490-17499.
[15] WU Fengyi, LIU Anran, ZHANG Tianfang, et al. Saliency at the helm: steering Infrared small target detection with learnable kernels[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 1-14.
[16] BEARMAN A, RUSSAKOVSKY O, FERRARI V, et al. What's the point: semantic segmentation with point supervision[C]. European Conference on Computer Vision. Cham: Springer International Publishing, 2016: 549-565.
[17] LI Boyang, XIAO Chao, WANG Longguang, et al. Dense nested attention network for infrared small target detection[J]. IEEE Transactions on Image Processing, 2023, 32: 1745-1758.
[18] WU Xin, HONG Danfeng, CHANUSSOT J. UIU-Net: U-Net in U-Net for infrared small object detection[J]. IEEE Transactions on Image Processing, 2023, 32: 364-376.
[19] DAI Yimian, LI Xiang, ZHOU Fei, et al. One-stage cascade refinement networks for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-17.
[20] YANG Bo, ZHANG Xinyu, ZHANG Jian, et al. EFLNet: enhancing feature learning network for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-11.
[21] CHEN Weicong, ZHANG Chenghong, LIU Yuan. PSHNet: hybrid supervision and feature enhancement for accurate infrared small target detection[J]. Applied Sciences, 2025, 15(14): 7629-7629.
[22] YING Xinyi, LIU Li, WANG Yingqian, et al. Mapping degeneration meets label evolution: learning infrared small target detection with single point supervision [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 15528-15538.
[23] LI Boyang, Wang Yingqian, Wang Longguang, et al. Monte carlo linear clustering with single-point supervision is enough for infrared small target detection[C]. IEEE/CVF International Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 1009-1019.
[24] MAN Yanmao, MULLER R, LI Ming, et al. That person moves like a car: misclassification attack detection for autonomous systems using spatiotemporal consistency[C]. USENIX Security Symposium (USENIX Security 23), 2023: 6929-6946.
[25] CHEN Tianxiang, YE Zi, TAN Zhentao, et al. MiM-ISTD: mamba-in-mamba for efficient infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-13.
[26] CHEN Yinpeng, DAI Xiyang, LIU Mengchen, et al. Dynamic convolution: attention over convolution kernels [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020. 11027-11036.
[27] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.
[28] HOU Qibin, ZHOU Daquan, FENG Jiashi. Coordinate attention for efficient mobile network design [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2021: 13713-13722.
[29] GUO Zhishun, WU Lin, HUANG Yabo, et al. Water-body segmentation for SAR images: past, current, and future[J]. Remote Sensing, 2022, 14(7): 1752-1752.
[30] ZHANG Mingjin, ZHANG Rui, YANG Yuxiang, et al. ISNet: shape matters for infrared small target detection [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2022: 877-886.

Lightweight Infrared Small Target Detection Method Based on Point-to-point Regression

QIU Zhaobing[1], GUAN Liwen[1], WU Yu[2], CHEN Liqiong[1], LIU Yuhao[1], WU Guangrui[1], HUANG Feng[1]

(1 *School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou* 350108, *China*)

(2 *China Centre for Resources Satellite Data and Applications, Beijing* 100094, *China*)

Abstract: Infrared small target detection, as the core technology of infrared search and tracking systems, plays a crucial role in military precision guidance, low-altitude economic monitoring, border surveillance, and space situational awareness. Its primary objective is to achieve rapid and accurate localization of small targets with low signal-to-noise ratios and weak contrast under long-distance conditions. In recent years, detection methods based on deep learning have gradually become a research hotspot and have achieved remarkable progress in detection accuracy. Most mainstream approaches adopt mask-to-mask or box-to-box regression training strategies, which can effectively enhance model robustness with the support of large-scale annotated datasets. However, these methods suffer from two major drawbacks. On the one hand, they heavily rely on high-quality, large-scale annotations, which are costly to obtain. On the other hand, in order to further improve accuracy, these methods often increase the depth and width of the network to expand model capacity, leading to massive parameters and complex architectures. To mitigate these issues, some studies have attempted to use weak supervision deep learning as a replacement for traditional mask or bounding-box annotations to reduce annotation costs. However, due to the limited granularity of supervision, the detection accuracy is often significantly constrained, making it difficult to balance practicality and precision.

To address the above challenges, this paper proposes a lightweight infrared small target detection algorithm based on point-to-point regression. The goal is to achieve high-accuracy detection while significantly reducing annotation and computational costs. The main contributions of this work are as follows. First, an asymmetric full-dimensional dynamic feature extraction module is designed. This module integrates channel shuffle with omni-dimensional dynamic convolution, which effectively models fine-grained dependencies across channel and spatial dimensions while maintaining network lightweightness. As a result, it enhances the discriminative features of the target and suppresses background interference. Second, a high-resolution cross-feature enhancement module is proposed. This module aligns multi-level features via upsampling, downsampling, and convolution mapping, and builds cross-layer interaction branches to dynamically fuse low-level detail features with high-level semantic features. In this way, it strengthens semantic representation while preserving spatial detail information, making it particularly suitable for weak and small target detection scenarios. Finally, an adaptive point regression detection head is introduced. This head models the distribution of target centers using Gaussian heatmaps, and combines local maxima filtering with a scale-adaptive mechanism to achieve precise prediction and sub-pixel localization. In this process, only single-point annotations are required for training, which substantially reduces the cost of data labeling.

Extensive experiments conducted on two public datasets, IRSTD-1K and NUDT-SIRST, demonstrate the effectiveness of the proposed method. With an extremely small number of parameters (only 0.24 M) and low computational complexity (1.53 GFLOPs), the proposed algorithm achieves outstanding performance: F1 score of 91.79% on IRSTD-1K and 97.44% on NUDT-SIRST, which are significantly superior to existing state-of-the-art methods. In addition, the model reaches an inference speed of 36 FPS on a single GPU, fully demonstrating its potential for real-time infrared small target detection in practical applications.

Key words: Infrared small target detection; Point-to-point regression; Lightweight; Deep learning; Dynamic feature extraction

OCIS Codes: 100.4996; 040.3060; 150.1135

CSTR: 32255.14.gzxb20265502.0210002

Foundation item: National Natural Science Foundation of China (62405060), Nature Science Foundation of Fujian Province (2022J05113, 2024J01245), Educational Research Program for Young and Middle-aged Teachers of Fujian Province (JAT210035)