

基于多跳深度网络的红外微小目标检测方法

燕舒乐¹, 陈润宇¹, 蔡念^{1*}, 许少秋¹, 陈健²¹广东工业大学信息工程学院, 广东 广州 510006;²中国科学院长春光学精密机械与物理研究所, 吉林 长春 130033

摘要 红外检测作为远程搜索和监视的重要手段,其对微小目标的检测精度影响实际应用价值。为了提高复杂背景下微小目标的测量精度,提出一种基于多跳深度网络的检测框架。首先,为了应对红外微小目标呈现“弱”和“小”的形状特点,利用无锚机制搭建特征金字塔作为骨干网络提取特征图。随后,为了实现渐进式特征交互和自适应的特征融合,在跳跃连接部分设计了由多尺度膨胀卷积组构成的多跳多尺度融合模块。最后,为了降低模型对微小目标位置偏差的敏感性,在训练中使用真实目标与预测目标的 Wasserstein 距离作为两者的相似性度量。实验结果表明,所提测量方法比已有测量方法具有更好的测量精度和效率。

关键词 红外微小目标检测; 多跳深度网络; 无锚点机制; Wasserstein 距离

中图分类号 TP391

文献标志码 A

DOI: 10.3788/LOP240610

Infrared Tiny Target Detection Method Based on a Multi-Hop Deep Network

Yan Shule¹, Chen Runyu¹, Cai Nian^{1*}, Xu Shaoqiu¹, Chen Jian²¹School of Information Engineering, Guangdong University of Technology,
Guangzhou 510006, Guangdong, China;²Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences,
Changchun 130033, Jilin, China

Abstract Infrared target detection is an important means of remote search and monitoring, and the accuracy of infrared tiny target detection determines the practical application value of this method. A detection framework based on a multi-hop deep network is proposed to improve the performance of tiny target detection in complex backgrounds. First, to deal with the “weak” and “small” shape characteristics of tiny targets, an anchor-free mechanism is used to build feature pyramids as the backbone for extracting feature maps. Then, to realize progressive feature interaction and adaptive feature fusion, a multi-hop fusion block composed of multi-scale dilation convolution groups is designed at the connection level. Finally, to reduce the sensitivity to position perturbations of tiny targets, the Wasserstein distance between the real and predicted targets is used as a similarity measure. The experimental results show that compared to existing methods, the proposed method delivers better detection performance in terms of accuracy and efficiency.

Key words infrared tiny target detection; multi-hop deep network; anchor-free; Wasserstein distance

1 引言

红外目标探测不仅在军事情报收集和远程侦察告警系统中发挥着至关重要的作用,而且在地球科学研究和自然灾害的遥感监测中也得到了广泛应用^[1]。红外微小目标的出现受多种因素的影响:由于红外成像固有的物理限制^[2],目标往往表现出低热辐射和微小的空间尺寸特征^[3];异常天气、光照变化等复杂气象因

素严重降低了目标的可视性;此外,军事目标可能会采用高速机动等策略,进一步损害目标的可见性。上述因素使得微小目标极易淹没在复杂背景中。随着探测距离的不断发展,如何将复杂背景下的微小目标精确检出是当前红外目标检测技术发展面临的重要挑战。

近年来,深度学习方法因数据驱动无需手动设计和抽取特征特点,被广泛应用于红外微小目标检测。最初,一些学者基于语义分割的思想对红外微小目标

收稿日期: 2024-01-25; 修回日期: 2024-01-26; 录用日期: 2024-04-03; 网络首发日期: 2024-04-12

基金项目: 国家自然科学基金(62171142)、广东省自然科学基金(2021A1515011908)、茂名市科技计划项目(2022S048)

通信作者: *cainian@gdut.edu.cn

进行建模。Wang 等^[4]引入了一种条件生成对抗网络,并以最小化漏警率和虚警率的平均检测结果作为输出。Dai 等^[5]引入了一种不对称上下文调制模块以增强目标对比度,从而实现有效检测。但是,分割本质是寻求对目标像素集的粗略近似^[6],因此只能作为红外微小目标定位的中间形式。由于红外微小目标可视性较弱,精确像素级标注的成本急剧上升。受限于此,算法精度的进一步提高变得具有挑战性。因此,一些通用目标检测领域解决微小目标漏检问题的策略逐渐被引入。Li 等^[7]提出一种密集嵌套的注意力网络自适应融合特征,并通过八连通邻域聚类确定目标区域。Mou 等^[8]基于 YOLO-v5s 使用特征重构采样策略和 CARAFE 算子减少信息损失,以增强检测质量。

尽管上述方法已经取得了可观的成效,但依然存在一些问题。现有方法主要在简单、均匀的背景背景下检测红外目标,很少有研究关注红外微小目标。它们在图像中以点状出现,这意味着要在更远距离探测目标,对军事情报侦察具有重要意义。由于红外目标所提供的特征微弱,在实际探测时目标经常淹没在云杂波、建筑物等背景中,即使经过图像处理,干扰点的灰度峰值仍然接近目标,容易产生大量虚警。此外,目标检测大多采用交并比(IoU)来进行目标框的回归,然而其对目标框位置偏移的过度敏感会导致 IoU 值急剧下降,造成红外微小目标漏警。

为进一步提高红外微小目标的检测精度,本文提出一种基于多跳深度网络的检测框架,其能有效减少红外微小目标的虚警和漏警,提升目标定位和

尺寸回归的精度。针对红外微小目标特征微弱、容易被淹没的成像特点,改进 CenterNet 的特征金字塔模块并设计尺寸无偏差的目标框回归函数。为了实现渐进式的特征交互和自适应的特征融合,在跳跃连接部分提出由多尺度膨胀卷积组构成的多跳多尺度融合模块(MMFB),在网络深层中维护微小目标的信息,防止其淹没在背景中。为了缓解对微小目标位置偏移的过度敏感性,使用真实目标与预测目标的 Wasserstein 距离衡量两者的相似性,并将其作为损失函数进行微小目标的回归训练,增加模型对微小目标的关注。

2 复杂背景下红外微小目标成像特性分析

图 1 为复杂背景下红外微小目标成像特性,其中,目标用黑框标注。由于远距离成像的色散效应^[9],只有无人机头部的发动机区域表现为明亮区域。因此,在远处的微小无人机目标为点状,只占几个像素,极易淹没在背景中。点状的微小目标不仅容易遗漏,而且存在误检的隐患。而在较近的探测距离下,目标在红外图像中为不同形态的斑点状小目标。所以,以像素级的精度标注红外目标的难度巨大,真实目标框包含了无法避免的模糊性。而这些不确定像素和模糊边界的影响是十分显著的。因此,提出 MMFB,在由浅入深的跳跃层级设计多尺度的膨胀卷积组,以模拟人类视觉系统中感受野(RF)结构^[10],捕捉不同尺度的目标细节和语义信息,进一步提升红外微小目标的检测精度。

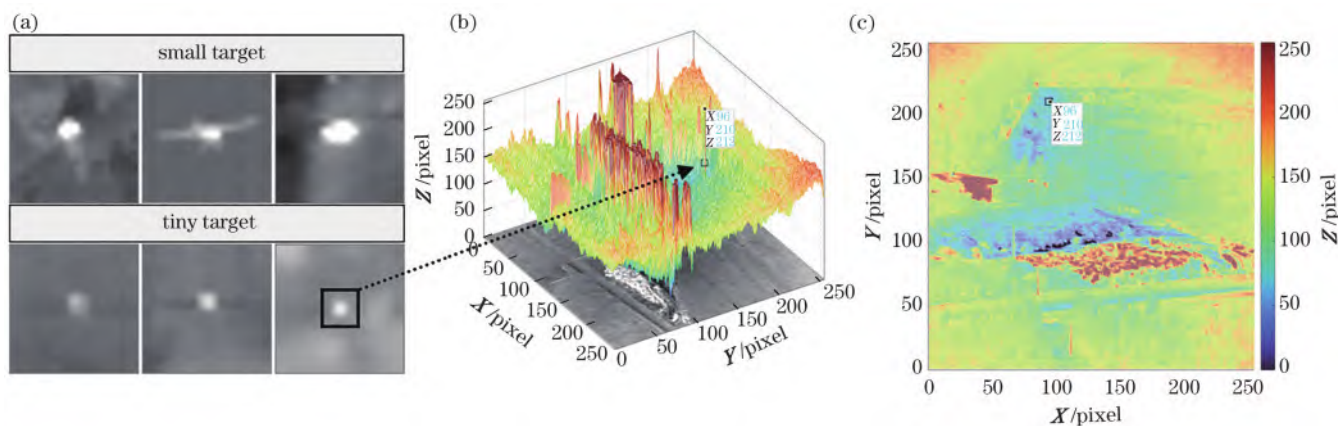


图 1 复杂背景下红外目标成像特性。(a)目标成像表现;(b)红外图像的灰度分布;(c)灰度分布的俯视图

Fig. 1 Imaging characteristics of infrared targets in complex background. (a) Target imaging performance; (b) grayscale distribution of infrared images; (c) top view of grayscale distribution

3 方法与理论

所提用于检测红外微小目标的多跳深度网络的总体框架如图 2 所示,由用于提取红外图像特征的骨干网络和用于目标属性回归的检测头模块组成。骨干网络由编-解码器与快速空间金字塔池化模块(FSPPB)

组成,在其中的跳跃连接层设计了一个 MMFB。为了逐步增加视觉感受野,MMFB 由多尺度的膨胀卷积组组成。该设计使网络能够有效地集成和充分利用红外微小目标的上下文信息,实现渐进式特征交互。随后,在特征金字塔的解码器中实现特征融合,聚合来自 MMFB 分支和 FSPPB 的不同细粒度信息。编码器包

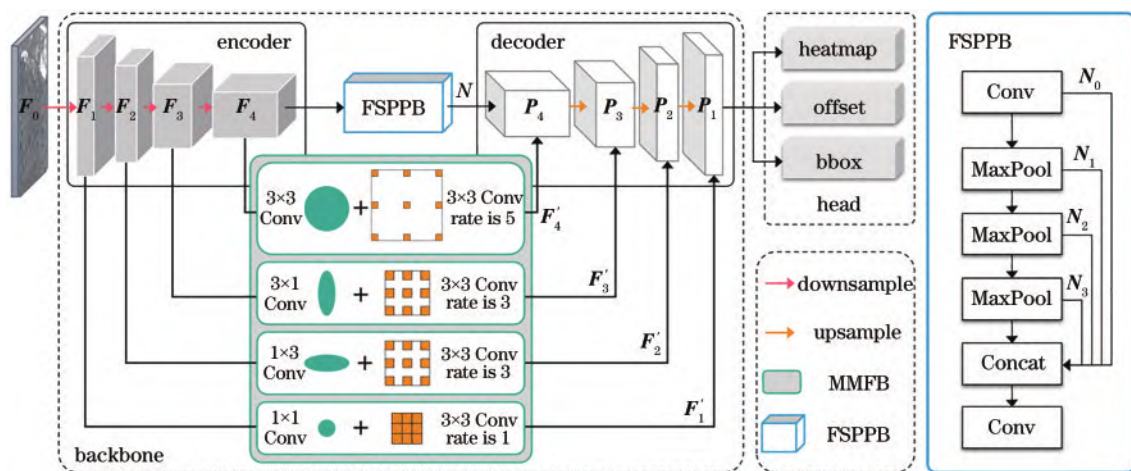


图2 所提方法的总体框架

Fig. 2 Overall framework of proposed method

括4个下采样阶段的若干下采样块,每个块由1个 3×3 卷积层、1个批归一化(BN)层和1个ReLU层组成。解码器包括3个上采样块,每个上采样块由 4×4 反卷积层、BN层和ReLU层组成。在检测头模块中,红外微小目标的中心、偏移量和目标框分别通过中心点热力图、偏移量和目标框分支进行回归训练,从而完成对红外微小目标的高精度检测。

3.1 基于快速空间金字塔池化的无锚骨干网络

考虑到红外目标的尺寸变化很大,例如,在文献[11]数据集中从 2×2 到 33×30 不等,且分布极度不均匀。因此,通用特征金字塔的参数设置并不适用,不适当的卷积核大小和池化层步长将导致细粒度信息的丢失,从而在网络更深处丢失微小目标。为了避免设计烦琐的显式锚点,采用一种无锚机制^[6]来构建特征金字塔,并利用FSPPB来缓解网络中的冗余特征提取问题。通过编码器中多个层级的特征提取模块,由浅入深生成不同尺度的特征图。这些特征图在空间上具有不同的分辨率,可以提供丰富的尺度信息。为了利用FSPPB和MMFB中不同层级的特征图信息,在解码器中通过上采样操作实现多尺度特征对齐融合。

具体来说,首先在编码器中对输入的原始红外图像 $F_0 \in \mathbb{R}^{H \times W \times C}$ 逐步进行下采样,逐渐降低分辨率,得到不同空间尺度的特征图 F_s 。为了增强模型对不同尺度红外目标的感知能力,在不引入额外参数的情况下在颈部引入FSPPB,如图2所示。只需经由池化层提取一次即可获得不同尺度下的特征信息 N_i ,定义如下:

$$N_i = \prod_{t=0}^i \text{MaxPool}^{3 \times 3} [\text{Conv}(F_4)] \quad t = 0, 1, 2, 3, (1)$$

式中: $\prod_{t=0}^i \text{MaxPool}^{3 \times 3} [\cdot]$ 表示特征经过 t 个级联的 3×3 最大池化层。特征信息 N_i 将在后续级联的池化层中多次使用,以降低计算成本和内存消耗。随后,FSPPB将多个尺度的特征信息 N_i 聚合。具体来说,以

Concat操作连接特征 N_i 后,利用一个 1×1 的卷积层将其整合,为来自不同层次的信息提供一个更紧凑的表示 N :

$$N = \text{Conv} \{ \text{Concat} [N_0, N_1, N_2, N_3] \}. \quad (2)$$

骨干网络无需设计和使用先验锚点,这使得训练过程更加简洁,减少了计算量和训练时间,并缓解了正负样本不平衡的问题。FSPPB可以抽取红外图像不同尺度的特征信息,使模型能够更好地处理不同尺寸的红外目标。

3.2 基于多跳结构的特征融合策略

通常情况下,网络随着层级不断加深,能够捕获更高维度的语义信息,进而提升模型的检测能力。但这一过程也伴随着信息的不断损失,影响红外小目标的定位精度,甚至使红外微小目标容易在网络深层中丢失。现有方法中编码器和解码器之间的多尺度融合策略大多采用一个带有不同卷积层的单跳结构,以拼接来自不同卷积层的特征,这表明融合的特征本质上来自单一尺度。因此,为了在由浅入深的网络层级中维护视觉信息,捕获不同层级特征的内在模式,受human visual system(HVS)^[10]的启发,设计了逐级增加视觉感知尺度的MMFB,其结构如图3所示。MMFB以逐级增大的感知尺度为各个层次上不同细粒度级别的特征搭建融合路径。浅层特征得以传递到深层,网络能有效维持微小红外目标的信息,降低漏警率;逐级增大的感知尺度可以包含更广泛的图像区域,有助于理解目标及其周围环境之间的关系,抑制虚警率。

具体来说,MMFB由4个具有不同卷积核和膨胀卷积组分支组成。在人类视觉系统中,群体感受野(PRF)的尺寸作为偏心率的函数,有两个明显趋势:每个视网膜拓扑图中PRF的尺寸与偏心率成正比;不同拓扑图中的PRF尺寸不同。因此,为MMFB不同分支的卷积层设计尺寸不一的卷积核,以模拟PRF的分布。而膨胀卷积层为每个分支分配单独的偏心率,以

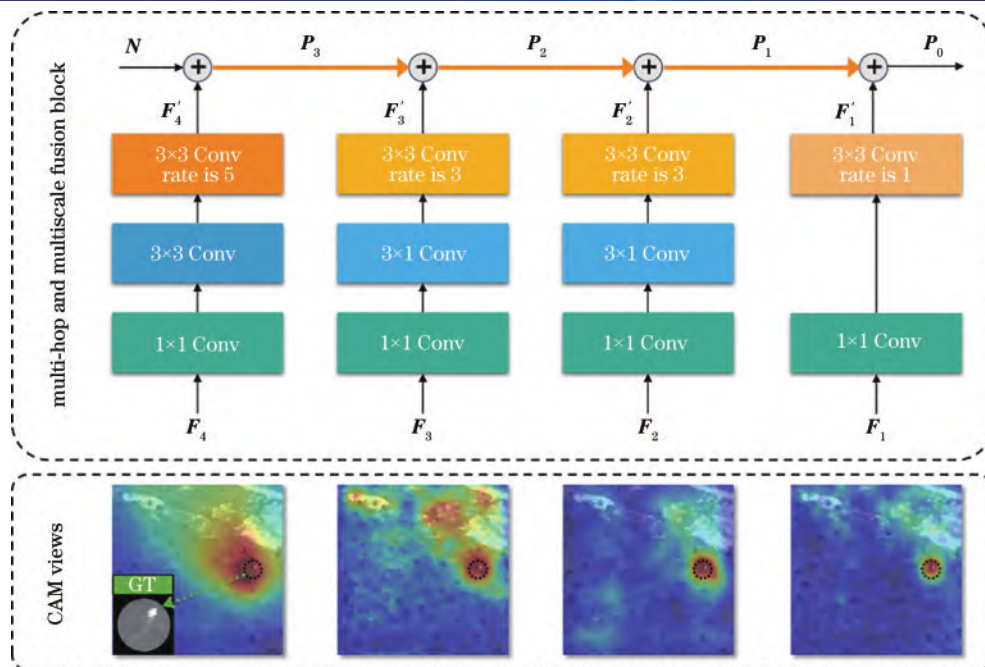


图3 MMFB结构及其对应分支的类激活映射图

Fig. 3 Structure of MMFB and class activation mapping diagram of corresponding branches

此模拟人类视觉系统中感受野结构中PRF的尺寸和偏心率之间的关系^[10]。每个分支中的信息与对应连接

处的深层特征以逐点加和的方式进行融合,融合后的特征为

$$P_l = \begin{cases} \text{smooth}\{F'_{l+1} + \text{ResizeConv}[\text{upsample}(P_{l+1})]\} & l = 0, 1, 2 \\ \text{smooth}(F'_{l+1} + N) & l = 3 \end{cases}, \quad (3)$$

式中:smooth(·)表示平滑操作,用于减少图像中的锯齿状边缘或栅格化导致的不连续性;upsample(·)表示上采样,用于对齐特征图;ResizeConv(·)为逐点卷积操作; N 表示FSPPB输出的聚合特征。

为了直观展示MMFB各分支在红外图像中“看”或关注的位置,使用类激活映射(CAM)图^[12]来可视化MMFB中不同分支的贡献分布。其中,具有较高分数(较暖颜色)的区域表示更强的网络响应和对预测输出的更显著贡献。如图3所示:较大的感受野可以包含更广泛的图像区域,有助于检测目标或捕获全局背景信息;更小的感受野可以提供更多局部和细节信息,例如边界和特定特征,有助于精确定位目标。

MMFB通过特征层之间逐级递增的感受野来获

取不同尺度下目标的细节和语义信息。因此,模型能够更好地应对具有较大尺寸变化的红外目标检测任务,能在理解目标、周围环境整体结构的同时捕捉细节信息,有助于模型回归到红外目标的准确位置。

3.3 损失函数

为了提高红外目标的回归精度,在检测头模块中通过中心点热力图(heatmap)、偏移量(offset)和目标框(bbox)分支进行回归训练,模型的损失函数被设置为

$$L_{\text{det}} = \lambda_{\text{heatmap}} L_{\text{heatmap}} + \lambda_{\text{offset}} L_{\text{offset}} + \lambda_{\text{bbox}} L_{\text{bbox}}, \quad (4)$$

式中: λ_{heatmap} 、 λ_{offset} 、 λ_{bbox} 是调整损失函数中各部分权重的超参数,分别设置为1、0.1、1。热力图损失函数 L_{heatmap} 用于回归到红外微小目标的中心点:

$$L_{\text{heatmap}} = -\frac{1}{N} \sum_{xyc} \begin{cases} (1 - \hat{P}_{xyc})^\alpha \log(\hat{P}_{xyc}), & P_{xyc} = 1 \\ (1 - \hat{P}_{xyc})^\beta (\hat{P}_{xyc})^\alpha \log(1 - \hat{P}_{xyc}), & \text{otherwise} \end{cases}, \quad (5)$$

式中: N 是红外图像中目标的数量; $\hat{P}_{xyc} \in [0, 1]^{W/R \times H/R \times C}$,是模型预测的红外微小目标热力图, W 、 H 和 C 分别表示宽度、长度和目标类型, R 代表下采样率; P_{xyc} 是根据红外微小目标真实标注框生成的

高斯热图; α 和 β 分别设为2和4,以平衡难易样本和正负样本。

通过热力图定位红外微小目标时,通常会得到目标在当前特征图的相对坐标,其位置信息是当前热力

图的整数坐标。在将特征图重新映射回原始输入尺寸时,会出现精度误差。因此,对于红外微小目标的中心点 p ,需要使用附加的局部偏移来补偿中心点的误差。当前特征图上的所有中心点共享相同的预测偏移值。偏移量损失函数 L_{offset} 表示为

$$L_{\text{offset}} = -\frac{1}{N} \sum_p \left| \hat{O} - \left(\frac{p}{R} - \hat{p} \right) \right|, \quad (6)$$

式中: $\hat{O}$ 是模型预测的中心点偏移信息; $\hat{p}$ 是下采样 R 倍后的中心点坐标的整数部分。

与基线方法 CenterNet 不同,为避免陷入独立优化中心位置和尺寸而不考虑其几何关系的误区,目前已有方法^[13]在 bbox 回归分支引入 IoU 计算红外目标的预测目标框和真实目标框的差距。然而,IoU 对微小目标的目标框扰动的容忍度较低^[14],使 IoU 模型对微小目标的关注较少。使用 normalized Wasserstein distance (NDN)^[15]将预测与真实目标框建模为两个二维高斯分布,继而将评估目标框相似度问题转化为计算两个高斯分布的距离,通过 Wasserstein 距离的解算如下:

$$L_{\text{bbox}} = \exp \left[-\sqrt{\text{Wasserstein}_2^2(\hat{N}, N) / C} \right], \quad (7)$$

式中: $\hat{N}$ 和 N 分别是预测目标框和真实目标框的高斯分布; C 为数据集中所有目标绝对尺寸的统计平均值。

在推理时,使用得分最高的中心点 (x, y) 、索引对应的偏移量 (o_x, o_y) 和目标框 $[x_c, y_c, w, h]$,最终的检测结果为 $\left[(x_c + x + o_x) / 2, (y_c + y + o_y) / 2, w + 2o_x, h + 2o_y \right]$ 。

4 实验结果与分析

4.1 实验设置

实验使用由 22 段序列、16177 帧红外图像、16944 个红外目标组成的红外弱小飞机数据集^[11]。该数据集涵盖了复杂多变的场景,如天空、山脉、森林、平原、丘陵、建筑物等。为便于实验分析,综合国际光学工程学会(SPIE)的定义^[16],将数据集中目标尺寸介于 4 pixel×4 pixel 至 9 pixel×9 pixel 之间的称为红外小目标(共计 1912 个,占比 13.4%),将目标尺寸为 4 pixel×4 pixel 的称为红外微小目标(共计 12385 个,占比 86.6%)。待测红外目标视觉特性复杂且差异巨大,采用平均信杂比(SCR)及其标准差(SD)定量描述检测难度,计算方式为

$$R_{\text{SCR}}^{\text{Mean}} = \frac{\sum_{i=1}^k n_i \cdot \mu_i}{\sum_{i=1}^k n_i}, \quad (8)$$

$$D_{\text{SD}} = \sqrt{\frac{\sum_{i=1}^k n_i \cdot \sigma_i^2 + \sum_{i=1}^k n_i \cdot (R_{\text{SCR}}^{\text{Mean}} - \mu_i)^2}{\sum_{i=1}^k n_i}}, \quad (9)$$

式中: n_i 表示第 i 个序列中的目标数量; μ_i 和 σ_i^2 分别表示序列中所有目标信杂比的平均值和方差。经统计,筛选后数据集所有红外目标的平均信杂比为 3.7049,而标准差为 2.2675。参考信杂比小于 5 的红外弱小目标定义^[17],数据集能提供逼近复杂探测环境的红外微小目标样本集。

由于原始数据集只包含中心点标注,为保证实验顺利开展,对数据集进行半自动化的目标框标注。设计程序首先为远距离的微小目标自动生成 4 pixel×4 pixel 的目标框,再对其他更近距离的无人机小目标进行人工标注。

所有实验都是在一台配备 NVIDIA GeForce RTX 3090 24 GB GPU 和 Intel(R) Xeon(R) CPU、内存为 64 GB 的计算机上进行的。在训练期间,初始学习率、批次大小和训练周期分别设置为 0.01、8 和 200。实验使用本领域常用的召回率(recall)、精确率(precision)、F1-score、虚警率(F_a)、漏警率(FNR)、每帧耗时,以及接收者操作特性(ROC)曲线^[21]作为衡量检测性能的评价指标^[18-20]。此外,由于 IoU 对微小目标偏移的容忍度较低^[14],分别使用不同的判定条件:对于小目标使用经典的 IoU^[22]指标评估检测质量;对于红外微小目标,将预测与实际目标中心点之间的距离和尺寸差异作为衡量目标是否正确检出的标准,其中心点距离的可接受阈值为 3 pixel^[20],可接受的尺寸差异阈值为 16 pixel。

4.2 与当前红外目标检测方法的对比实验

为了验证所提红外微小目标检测方法的有效性,将所提方法和当前基于深度学习的红外目标检测方法进行对比。这些方法包括 MDFA-CGAN^[4]、ACM^[5]、DNANet^[7]和 YOLO-FR^[8]。为了确保公平,上述所有方法都使用同一训练集重新训练,实验结果(性能最优标记为蓝色,次优标记为绿色)如表 1 所示。

MDFA-CGAN 和 ACM 将红外目标检测转化为图像翻译和语义分割任务,即目标被视为一个像素集合而非目标框,这更适用于轮廓清晰的显著目标。因此,它们在微小目标上的性能受标签模糊性的影响很大,ACM 的漏警率甚至达到 80.1%。此外,它们还需要高昂的计算成本来恢复像素级细节,每帧耗时分别为 27.1 ms 和 30.3 ms。

DNANet 使用 U 形嵌套结构来提取特征,对小目标实现了 2.6% 的虚警率。但重复上下采样产生了无法修复的精度损失,导致微小目标出现 20.2% 的虚警率和 30.1% 的漏警率。同样,模型的堆叠和嵌套也带来了显著的计算负担,每帧耗时 31.3 ms。

YOLO-FR 受益于轻量级架构和加速机制,以每帧耗时 6.4 ms 的检测速度获得极大优势。且由于采用特征重组采样和小感受野的目标探测头,微小目标漏警率降至 7.9%,小目标漏警率降至 9.4%。但固定的小感受野缺乏全局信息,背景中的干扰点难以与目标区

表1 与不同红外目标检测方法的比较

Table 1 Comparisons of different methods on infrared target detection

Method	Target	Recall / %	Precision / %	F1-score / %	F_a / %	FNR / %	Time / ms
MDFA-CGAN	Tiny	56.8	52.4	54.5	41.6	43.2	27.1
	Small	71.1	45.4	55.4	16.5	28.9	
ACM	Tiny	19.9	26.7	22.8	44.1	80.1	30.3
	Small	54.9	57.4	56.1	7.8	45.1	
DNANet	Tiny	69.9	73.6	71.7	20.2	30.1	31.3
	Small	86.8	86.7	86.8	2.6	13.2	
YOLO-FR	Tiny	92.1	60.6	73.1	48.3	7.9	6.4
	Small	90.6	56.3	69.5	13.5	9.4	
Proposed method	Tiny	97.0	94.5	95.7	4.5	3.0	11.3
	Small	91.1	96.4	93.7	0.7	8.9	

分,导致微小目标出现 48.3% 的高虚警率。

与此相比,所提多跳深度网络在每帧耗时 11.3 ms 的可接受速度下,获得了最佳的检测性能。对于微小目标,虚警率和漏警率分别为 4.5% 和 3.0%;对于小目标,虚警率和漏警率分别为 0.7% 和 8.9%。通过进一步分析可知,尽管小目标相对于微小目标能提供更多信息,但其在数据集中所占比例(13.4%)较少且特征差异更大。为了缓解对微小目标轻微偏移的过度敏感,NWD 以数据集中所有目标绝对尺寸的统计平均值作为约束尺度。因此,在保证模型精准检出微小目标

的前提下,小目标漏警率相对微小目标更高,而两者均取得最优。

图 4 展示了对比实验中各检测方法的 ROC 曲线,横纵坐标分别表示虚警率和检测率。DNANet 和 YOLO-FR 取得了良好的表现。ACM 在很大程度上受限于对目标像素集合的粗略回归,无法避免复杂场景下标签的模糊性,表现出最差的性能。结果表明,所提方法表现出明显高于其他所有对比方法的性能。当虚警率发生变化时,检测率可以迅速响应变化,尤其对微小目标性能最佳。

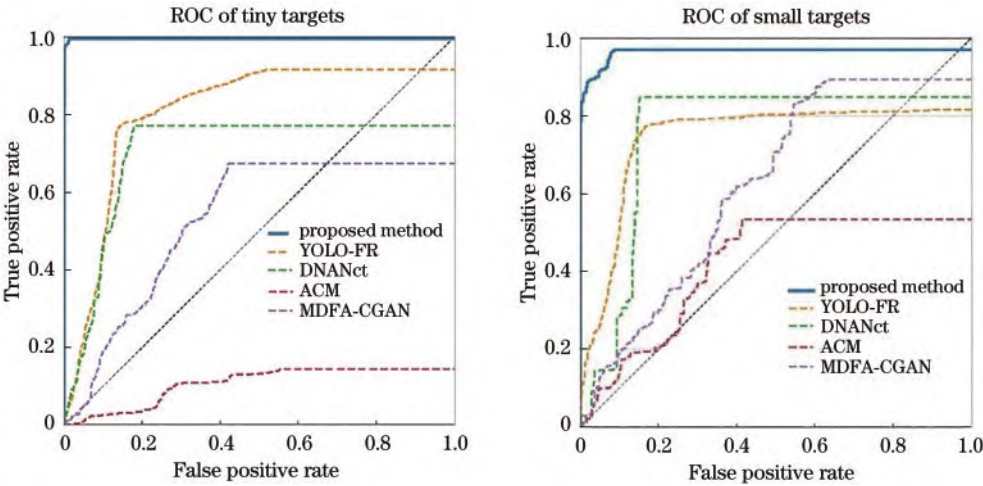


图4 不同对比方法的 ROC 曲线
Fig. 4 ROC curves for different compared methods

两张图像的检测结果的可可视化如图 5 所示,目标区域被放大,正确检测到的目标和误检区域分别用绿色和红色框突出显示,而真实目标用蓝色标识。与现有方法相比,所提多跳深度网络对各种尺寸的目标具有更强的鲁棒性,特别是对微小目标具有显著优势。对于微小目标,MDFA-CGAN 和 ACM 仅能发现目标,难以准确预测目标尺寸,导致虚警。YOLO-FR 在检测到目标的同时产生了虚警。只有 DNANet 和所提方法成功检测到了微小目标。然而, DNANet 错误地将两个小

目标预测为一个目标。这些视觉结果与上述不同方法的统计分析一致,表明所提网络可以适应不同尺寸的目标,检测结果中目标定位和尺寸更加准确。

4.3 消融实验

本小节展开消融实验,以分析所提红外目标检测方法中 MMFB 模块和 NWD 损失函数的有效性。为基线方法^[23]增加 MMFB 和 NWD,以说明这两个各自为检测精度带来的改进,实验结果如表 2 所示(性能最优标记为蓝色)。

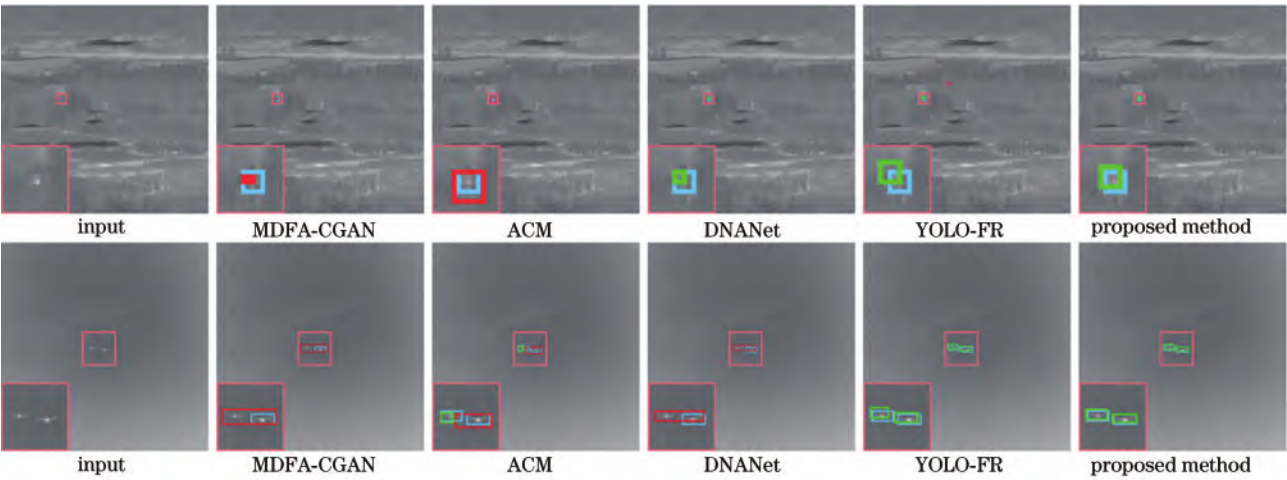


图 5 检测结果可视化
Fig. 5 Visualization of detection results

表 2 消融实验

Table 2 Ablation experiments								unit: %
Baseline	MMFB	NWD	Target	Recall	Precision	F1-score	F_a	FNR
✓			Tiny	95.1	79.0	86.3	20.4	4.9
			Small	46.5	99.6	63.4	19.5	53.5
✓	✓		Tiny	95.9	84.7	90.0	4.2	4.1
			Small	89.4	89.7	89.6	2.0	10.6
✓		✓	Tiny	92.0	90.4	91.2	5.3	8.0
			Small	93.5	82.6	87.7	3.8	6.5
✓	✓	✓	Tiny	97.0	94.5	95.7	4.5	3.0
			Small	91.1	96.4	93.7	0.7	8.9

由于MMFB逐级增大视觉感知尺度,模型能够学习到融合语义和细节信息的特征表示,以帮助网络在深层维护红外目标的弱小特征。而逐级增大的感知尺度可以包含更广泛的图像区域,有助于理解目标及其周围环境之间的关系,从而显著减少了异常检测情况,微小目标的虚警率降低至4.2%。通过引入NWD衡量目标框的相似度,给予红外微小目标更高的相似度,引导网络更加关注微小目标,缓解了IoU对微小目标扰动的敏感问题,进一步提高了红外目标的检测质量,获得了微小目标3.0%的最低漏警率和小目标0.7%的最低虚警率。

4.4 在另一个数据集上的拓展应用

为了更全面地评估所提方法,采用红外弱小目标(IRDST)数据集^[24]进行进一步比较,该数据集涉及85个场景中的40650帧真实红外图像。与弱小飞机数据集^[11]相比,IRDST数据集涉及两种尺寸的采集图像,其目标尺寸集中分布,93.4%为5 pixel×5 pixel~9 pixel×9 pixel之间的红外小目标。在本研究中,IRDST数据集中的红外目标可以被认为小目标。IRDST数据集按照6:2:2的比例划分为训练集、验证集和测试集,图像帧数分别为24393、8131和8131。

如表3所示,所提方法保持了最佳性能,获得了4.3%的虚警率,进一步验证了其对各种场景的适应性。此外,图6展示了不同检测方法在IRDST数据集上的测试结果。目标区域被放大,正确检测到的目标和误检区域分别用绿色和红色框突出显示,而真实目标区域用蓝色标识。MDFA-CGAN在产生2个虚警的情况下,漏警2个真实目标。ACM漏警2个目标,DNANet漏警1个目标。YOLO-FR在产生2个虚警的情况下,不精确地检出3个目标。相较于上述现有方法,所提方法在不产生虚警的情况下,精准检出3个红外目标。因此,上述事实可以充分验证所提方法在各类应用场景的优越性能。

表 3 在IRDST数据集上的对比

Table 3 Comparison on IRDST dataset						unit: %
Method	Recall	Precision	F1-score	F_a	FNR	
MDFA-CGAN	41.6	44.8	43.1	55.2	58.4	
ACM	35.5	35.9	35.7	64.1	64.5	
DNANet	89.2	90.5	89.9	9.5	10.8	
YOLO-FR	83.2	58.5	68.7	41.5	16.8	
Proposed method	93.6	95.7	94.6	4.3	6.4	

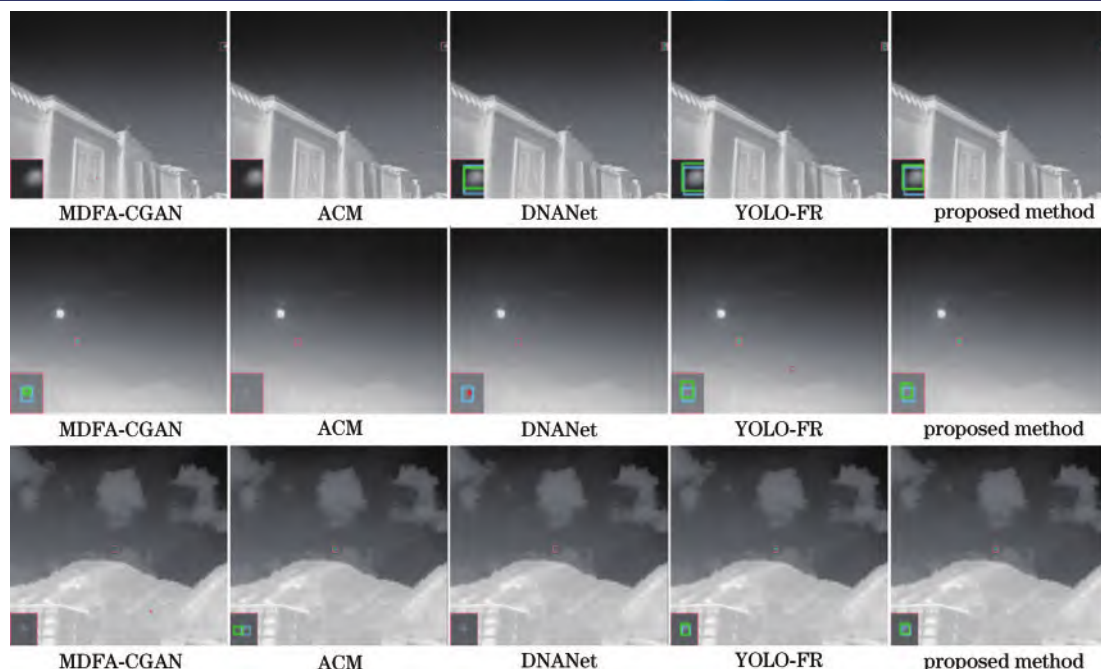


图6 IRDST数据集中红外目标检测结果

Fig. 6 Visualization of infrared targets on IRDST dataset

5 结 论

提出一种端到端的多跳深度网络以检测红外微小目标,利用无锚机制搭建特征金字塔作为骨干网络提取特征图,在跳跃连接部分设计了由多尺度膨胀卷积组构成的MMFB以融合特征,使用多检测头对特征图进行目标各个属性的回归。为了降低模型对微小目标位置偏差的敏感性,在训练中使用真实目标与预测目标的Wasserstein距离作为两者的相似性度量。实验结果表明,相比于现有红外目标检测方式,所提多跳深度网络能够快速、精确地检出红外目标。

相比于现有的红外目标检测方法,所提方法对微小目标已经取得了良好的检测结果,但是仍有不足。当前每帧11.3 ms的检测速度无法很好适应超高速飞行器,在更大尺度目标上,漏检率还有下降的空间。接下来,将在轻量化策略和加速技术方面进行研究,并继续关注漏检的小目标样本,开展能自适应目标尺度的相似度计算研究。

参 考 文 献

- [1] 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.
- [2] 胡伟达, 李庆, 陈效双, 等. 具有变革性特征的红外光电探测器[J]. 物理学报, 2019, 68(12): 120701.
Hu W D, Li Q, Chen X S, et al. Recent progress on advanced infrared photodetectors[J]. Acta Physica Sinica, 2019, 68(12): 120701.
- [3] 侯义锋, 丁畅, 刘海, 等. 逆光海况下低质量红外目标的增强与识别[J]. 光学学报, 2023, 43(6): 0612003.

Hou Y F, Ding C, Liu H, et al. Enhancement and recognition of infrared target with low quality under backlight maritime condition[J]. Acta Optica Sinica, 2023, 43(6): 0612003.

- [4] Wang H, Zhou L P, 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), October 27-November 2, 2019, Seoul, Republic of Korea. New York: IEEE Press, 2019: 8508-8517.
- [5] Dai Y M, Wu Y Q, Zhou F, et al. Asymmetric contextual modulation for infrared small target detection [C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV), January 3-8, 2021, Waikoloa, HI, USA. New York: IEEE Press, 2021: 949-958.
- [6] Dai Y M, Li X, Zhou F, et al. One-stage cascade refinement networks for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5000917.
- [7] Li B Y, Xiao C, Wang L G, et al. Dense nested attention network for infrared small target detection[J]. IEEE Transactions on Image Processing, 2023, 32: 1745-1758.
- [8] Mou X G, Lei S, Zhou X. YOLO-FR: a YOLOv5 infrared small target detection algorithm based on feature reassembly sampling method[J]. Sensors, 2023, 23(5): 2710.
- [9] 王好贤, 董衡, 周志权. 红外单帧图像弱小目标检测技术综述[J]. 激光与光电子学进展, 2019, 56(8): 080001.
Wang H X, Dong H, Zhou Z Q. Review on dim small target detection technologies in infrared single frame images[J]. Laser & Optoelectronics Progress, 2019, 56

- (8): 080001.
- [10] Liu S T, Huang D, Wang Y H. Receptive field block net for accurate and fast object detection[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11215: 404-419.
- [11] 回丙伟, 宋志勇, 范红旗, 等. 地/空背景下红外图像弱小飞机目标检测跟踪数据集[J]. 中国科学数据, 2020, 5(3): 291-302.
Hui B W, Song Z Y, Fan H Q, et al. A dataset for infrared detection and tracking of dim-small aircraft targets under ground/air background[J]. China Scientific Data, 2020, 5(3): 291-302.
- [12] Desai S, Ramaswamy H G. Ablation-CAM: visual explanations for deep convolutional network via gradient-free localization[C]//2020 IEEE Winter Conference on Applications of Computer Vision (WACV), March 1-5, 2020, Snowmass, CO, USA. New York: IEEE Press, 2020: 972-980.
- [13] Shi T J, Gong J N, Hu J M, et al. Feature-enhanced CenterNet for small object detection in remote sensing images[J]. Remote Sensing, 2022, 14(21): 5488.
- [14] Xu C, Wang J W, Yang W, et al. Dot distance for tiny object detection in aerial images[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 19-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 1192-1201.
- [15] Wang J W, Xu C, Yang W, et al. A normalized Gaussian Wasserstein distance for tiny object detection [EB/OL]. (2021-10-26)[2024-05-06]. <https://arxiv.org/abs/2110.13389>.
- [16] 吴洪凯, 董科研, 宋延嵩, 等. 基于加权块对比度的红外小目标检测方法[J]. 激光与光电子学进展, 2023, 60(16): 1628006. Wu H K, Dong K Y, Song Y S, et al. Infrared small target detection method based on weighted patch contrast[J]. Laser & Optoelectronics Progress, 2023, 60(16): 1628006.
- [17] Sun Y, Yang J G, 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, 2021, 59(5): 3737-3752.
- [18] Tong X Z, Su S J, Wu P, et al. MSAFFNet: a multiscale label-supervised attention feature fusion network for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5002616.
- [19] Guo L C, Sun X L, Zhang W L, et al. Small aerial target detection using trajectory hypothesis and verification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5609314.
- [20] Yan P T, Hou R Z, Duan X G, et al. STDMA Net: spatio-temporal differential multiscale attention network for small moving infrared target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5602516.
- [21] Fu K R, Zhao Q J, Gu I Y H, et al. Deepside: a general deep framework for salient object detection[J]. Neurocomputing, 2019, 356: 69-82.
- [22] Yu J H, Jiang Y N, Wang Z Y, et al. UnitBox: an advanced object detection network[C]//Proceedings of the 24th ACM International Conference on Multimedia, October 15-19, 2016, Amsterdam, The Netherlands. New York, NY, USA: ACM Press, 2016: 516-520.
- [23] Zhou X Y, Wang D Q, Krähenbühl P. Objects as points [EB/OL]. (2019-04-16)[2024-05-06]. <https://arxiv.org/abs/1904.07850>.
- [24] Sun H, Bai J X, Yang F, et al. Receptive-field and direction induced attention network for infrared dim small target detection with a large-scale dataset IRDST[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5000513.