

基于多尺度感知的自适应无人机红外小目标检测^{*}王亦君¹ 张露² 李燕²

(1. 南京信息工程大学计算机学院、网络空间安全学院 南京 210044; 2. 无锡学院物联网工程学院 无锡 214105)

摘要: 无人机红外图像小目标检测是侦察监视、搜索救援等领域的关键技术,其可靠性受到目标尺寸微小、特征表达弱、背景复杂等问题的挑战。针对上述难题,本文提出了一种基于多尺度感知与自适应筛选策略的无人机红外小目标检测算法 PGF-RTDETR。首先,在颈部网络中构建极化通道门控单元,利用极化注意力机制引导多尺度特征进行深度语义融合,从而增强小目标的可辨识度,有效抑制背景干扰。其次,设计了自适应卷积增强模块,利用门控机制增强特征的选择性与精细度,提升细粒度特征的提取效率。最后,骨干网络中引入了融合部分卷积与通道变换的高效残差结构,提升特征表达能力的同时有效降低了模型参数量。实验结果表明,与基准模型 RT-DETR-R18 相比,PGF-RTDETR 在 HIT-UAV 数据集上 mAP50 和 P 分别提高了 2.6% 和 6.6%,模型参数量、计算量分别降低了 20.7%、25%,在 FLIR 数据集上,mAP50 提升 0.6%。改进模型在保持较小参数量和计算量的同时,提高了检测精度,为无人机红外小目标检测提供有效解决方案。

关键词: RT-DETR; 无人机; 红外图像; 小目标检测

中图分类号: TP391.4; TN219 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Adaptive UAV infrared small target detection based on multi-scale perception

Wang Yijun¹ Zhang Lu² Li Yan²

(1. School of Computer Science and School of Cyber-space Security, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. School of Internet of Things Engineering, Wuxi University, Wuxi 214105, China)

Abstract: UAV infrared image small target detection is a key technology in the fields of reconnaissance and surveillance, search and rescue, and its reliability is challenged by problems such as small target size, weak feature expression, and complex background. To solve the above problems, this paper proposes a UAV infrared small target detection algorithm PGF-RTDETR based on multi-scale perception and adaptive screening strategy. Firstly, a polarized channel-gated unit is constructed in the neck network, and the polarized attention mechanism is used to guide multi-scale features for deep semantic fusion, so as to enhance the recognizability of small targets and effectively suppress background interference. Secondly, an adaptive convolution enhancement module is designed to enhance the selectivity and fineness of features by using the gating mechanism to improve the extraction efficiency of fine-grained features. Finally, a high-efficiency residual structure that integrates partial convolution and channel transformation is introduced into the backbone network, which improves the feature expression ability and effectively reduces the number of model parameters. Experimental results show that compared with the benchmark model RT-DETR-R18, PGF-RTDETR increases mAP50 and P by 2.6% and 6.6% on the HIT-UAV dataset, and reduces the number of model parameters and computation by 20.7% and 25%, respectively, and mAP50 increases by 0.6% on the FLIR dataset. The improved model improves the detection accuracy while maintaining a small amount of parameters and computational amount, and provides an effective solution for UAV infrared small target detection.

Keywords: RT-DETR; UAV; infrared image; small target detection

0 引言

无人机技术^[1]快速发展,广泛应用于多领域,推动探测

需求增长。红外成像因被动探测、全天时、抗干扰优势,成为重要探测手段。但受分辨率、背景、尺度、信噪比影响,红外小目标检测仍存挑战^[2]。特别是在云层、地物杂波或热

噪声干扰的复杂场景下,准确有效地检测红外小目标^[3]对提升无人机电势感知与自主导航能力至关重要。因此,设计一种适应复杂场景、结构轻量高效且兼顾检测精度的红外小目标检测算法,成为当前研究的重要方向。

近年来,深度学习技术在无人机红外小目标检测中展现出强大的发展潜力。现有目标检测算法主要分为两类:两阶段检测算法与单阶段检测算法。两阶段算法通常先通过卷积神经网络提取输入红外图像的特征,然后生成候选区域,结合分类器与回归器完成目标的精确识别与定位。典型代表包括 Faster R-CNN^[4]、Mask R-CNN^[5]、Mask R-FCN^[6]等。这类方法精度高,但多阶段处理导致计算复杂度高、推理速度慢等挑战,难以满足无人机平台对实时性与轻量化的双重需求^[7]。相比之下,单阶段算法在同一网络中即可同时完成目标的位置与类别预测,显著提高检测速度并降低计算负担,更适合部署在算力有限的无人机飞行任务中。其主流算法有 SSD^[8] (single shot multibox detector)、YOLO^[9] (you only look once)、DETR^[10] (detection transformer) 和 RT-DETR^[11] (real time detection transformer) 等。

为了提升复杂环境下对红外小目标的检测性能,许多研究者开始针对该领域提出创新的算法设计。目前,YOLO 系列因其实时性和高效性,已成为主流检测框架。例如,Cao 等^[12]提出了 YOLO-TSL,基于 YOLOv8n 构建轻量级检测结构,结合三重注意力机制增强特征提取能力,并引入 Slim-Neck 架构和一次性聚合策略以降低计算复杂度,同时采用 MPDIoU 边框回归损失优化检测精度。Zhao 等^[13]设计了基于 YOLOv7 的红外多尺度目标检测模型 YOLO-ViT,通过引入 MobileViT 骨干网络与 CARAFE 上采样模块增强小目标特征表达,并结合 K-means 聚类优化先验框配置,进一步提升了无人机平台上的检测效果。Mou 等^[14]为解决红外小目标在采样过程中的特征丢失问题,提出了 YOLO-FR 算法,该算法通过引入 STD 模块和 CARAFE 算子分别优化网络的下采样与上采样过程。Zhu 等^[15]提出了 YOLO-SDLUWD 网络,通过引入空间深度卷积(SD-Conv)和改进的特征金字塔网络(LU-PAFPN),以解决红外小目标因下采样导致的信息丢失和特征融合不足的问题。

尽管 YOLO 系列在无人机红外小目标检测中取得成效,但其依赖局部感受野,难以捕捉长距离依赖关系,在低对比度红外图像中易受背景干扰,导致误检与漏检率升高。此外,其依赖固定阈值的非极大值抑制(non-maximum suppression, NMS),在目标密集或尺度差异较大的场景下易出现目标误抑制问题,进一步限制适用范围。为此,DETR 引入 Transformer 来实现端到端检测,不仅省略了 NMS 后处理步骤,还具备全局建模的优势。然而,DETR 存在收敛速度慢、小目标建模能力不足的缺陷,难以满足红外小目标检测对精度与效率的双重需求。为此,百度提出

了实时目标检测算法 RT-DETR,通过引入更加高效的混合编码器、IoU 感知查询机制等技术,保证了检测的实时性。为了提高小目标检测的精度,Yang 等^[16]提出了 Hybrid-DETR 模型,通过设计差分主干网络(HybridNet)与引入小目标检测模块(SODM),有效平衡了特征提取效率与全局感知能力,进而提升了复杂背景下的红外小目标检测性能。Chen 等^[17]提出了 FP-RTDETR,该模型设计了聚焦特征金字塔网络(FFPN),其通过双向融合机制有效捕捉远程依赖,以增强小目标检测。同时,该模型还引入了多尺度并行聚合拼接(MPAS)模块以提升网络对不同目标尺寸的适应性。

尽管上述方法取得了不错的进展,但仍存在一些局限性。首先,现有方法未特别关注解决红外图像中弱小目标与强背景杂波易混淆的特征辨识问题;其次,模型特征提取能力与细粒度建模能力仍有可提升空间。因此,针对上述不足,本文在 RT-DETR 基础上,进一步提出如下改进策略以提升检测性能:

1) 针对 RT-DETR 网络在特征融合阶段对小目标特征表达能力不足、易受复杂背景干扰的问题,该阶段构建极化通道门控单元 PCG。该模块借助极化注意力和门控机制,有效增强了多尺度特征的表达与融合能力,使网络能更好地聚焦与小目标的显著特征,从而增强了复杂背景下小目标区分能力。

2) 为解决主干特征增强阶段特征提取效率有限、参数冗余等问题,设计了自适应卷积增强模块 gConvC3。该模块通过增强特征选择性和细粒度建模能力,进一步提升模型对小目标的检测准确性与鲁棒性。

3) 针对骨干网络中传统残差结构特征表达能力有限、模型冗余的问题,将传统 BasicBlock 替换为融合部分卷积与通道变换的高效残差结构 FasterBlock。该结构通过空间混合与通道变换的协同作用,有效减少了模型的参数量,实现了网络的高效与轻量化。

1 RT-DETR 模型

RT-DETR 是一种基于 Transformer 的单阶段端到端目标检测框架,主要包括以下 3 个模块:骨干网络(Backbone)、高效混合编码器(efficient hybrid encoder)以及 Transformer 解码器(transformer decoder)。RT-DETR 整体结构如图 1 所示。

在本研究中,骨干网络采用 ResNet18 作为基准模型,用于提取输入图像的多尺度特征,生成 3 个特征层{S3, S4, S5},为后续编码器提供丰富的层次化特征表示。

高效混合编码器由两个主要模块构成:基于自注意力的内尺度特征交互模块(attention-based intra-scale feature interaction, AIFI)和基于 CNN 的跨尺度特征融合模块(CNN based cross-scale feature fusion module, CCFM)。其中,AIFI 模块采用单尺度 Transformer 编码器架构,用

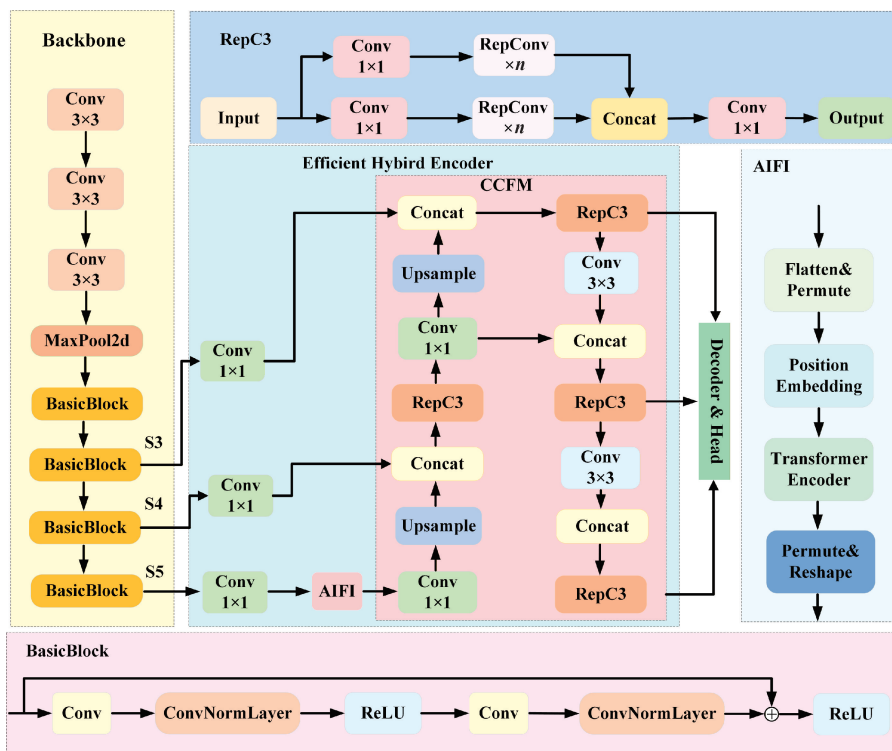


图1 RT-DETR 模型结构

Fig. 1 Network structure of RT-DETR

于处理骨干网络输出的最高层特征图 S5,在保证模型检测精度的同时,有效降低模型复杂度。CCFM 模块则通过级联的融合块(Fusion Blocks)实现多尺度特征整合,每个融合块利用卷积运算对相邻尺度特征进行融合,进而生成具有更强表征能力的特征。

解码器包含不确定性最小化查询选择机制与辅助检测头两部分。解码器接收来自编码器的特征和一组初始查询,基于这些输入通过多层 Transformer 结构对查询特征逐层优化,每一层的优化过程使查询逐渐聚焦于图像中重要的目标区域。最终,解码器输出每个查询对应的物体类别及其边界框坐标,从而完成精确的目标检测。

2 改进的 RT-DETR 网络架构

本文对 RT-DETR 模型进行了一系列改进,提出了无人机红外小目标检测算法 PGF-RTDETR,旨在降低模型复杂度的同时提升小目标检测的精度和模型的整体效率。首先,在特征融合阶段,将原有的 AIFI 模块替换为极化通道门控单元 PCG,以提升多尺度特征的融合与小目标的区分能力。接着,在 FPN-PAN 结构中采用自适应卷积增强模块 gConvC3,增强特征选择性和细粒度信息建模能力。最后,在 Backbone 部分,将传统的 BasicBlock 替换为高效残差结构 FasterBlock,优化特征传递路径,实现网络的高效与轻量化。PGF-RTDETR 结构如图 2 所示。

2.1 PCG 模块

在无人机红外小目标检测任务中,由于目标体积较小、辐射信号微弱以及成像分辨率受限,目标在特征图中往往只占据极少的像素点,细节特征在深层卷积过程中容易被削弱,甚至在某些情况下可能会丢失,从而使检测难度增加^[18]。而传统卷积网络在单尺度卷积下往往只能获取局部的空间信息,对多尺度目标的特征捕捉能力有限,特征表达不足和判别力减弱的问题尤为突出^[19]。

为了解决无人机红外外场中小目标在复杂背景下特征表达不足以及单一尺度特征对多尺度目标适应性有限的问题,本文构建了一种融合高效线性注意力机制(polarity-aware linear attention, Pola)^[20]与门控卷积网络(convolutional gated linear unit, CGLU)^[21]的极化通道门控单元 PCG,替换传统的 AIFI。图 3 展示了 PCG 的结构图,Pola 机制利用极性分解操作,能够显式地区分前景目标与背景杂波的特征,进而在特征层面拉大目标与背景的对比度;而 CGLU 门控机制则能自适应地筛选特征通道,增强关键目标信息,抑制背景噪声。这种设计使网络能更好地聚焦于小目标的显著特征,为后续检测提供了更具判别力的特征表示。

图 4 展示了 Pola 的结构图。针对红外小目标信号微弱、易于背景杂波混淆的问题,Pola 机制的核心思想在于通过极性分解,来同时建模特征间的相关性与对比性。这种设计显式地引导模型学习红外小目标与杂波之间的特

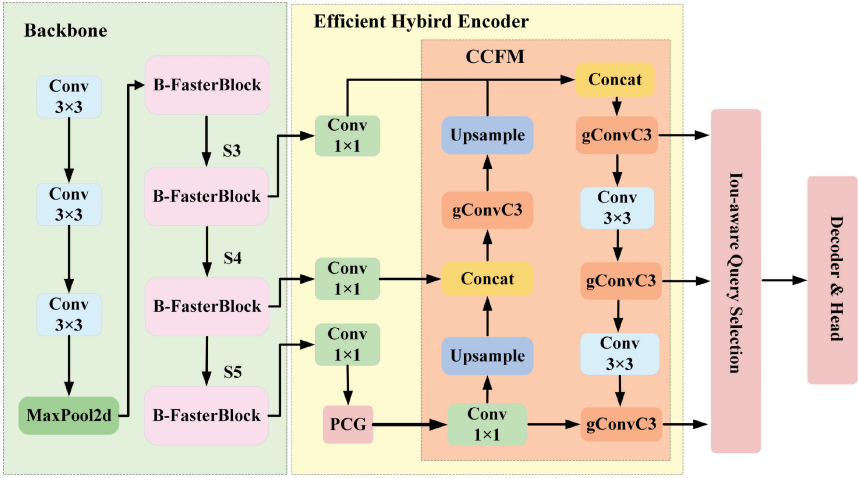


图 2 PGF-RTDETR 网络结构
Fig. 2 Network structure of PGF-RTDETR

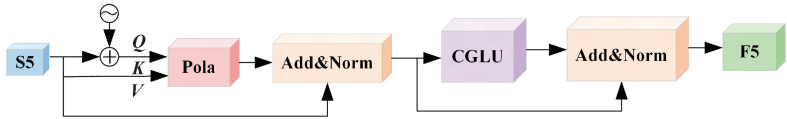


图 3 PCG 结构
Fig. 3 PCG structure

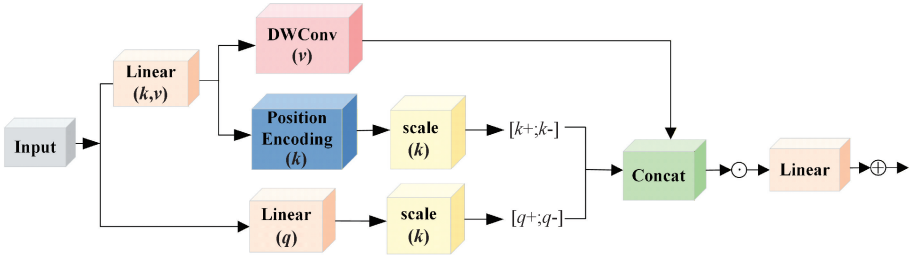


图 4 Pola 结构
Fig. 4 Pola structural

征差异(即对比性),以及背景区域内部的特征一致性(即相关性)。通过在特征层面拉大目标与背景的对比度,从而提升对微弱红外目标的判别能力。其关键的极性分解操作是通过 ReLU 函数及其变体,将查询(Q)和键(K)分别分解为正负两个极性分支。

在此基础上,CGLU 结构通过门控机制对信息流进行自适应调节。如图 5 所示,该门控机制针对红外图像信噪比低的特点,来动态评估每个特征通道的重要性,自动增强包含关键小目标信息的通道权重,同时有效抑制

充满背景噪声和杂波的通道。其核心的门控融合过程可表示为:

$$\mathbf{Y} = \mathbf{F} \odot \mathbf{G} \quad (1)$$

其中, $\mathbf{F}$ 为特征分支, $\mathbf{G}$ 为门控分支, $\odot$ 表示 Hadamard 乘法。通过将 Pola 的极化感知与 CGLU 的通道筛选相结合,PCG 模块能引导浅层细粒度特征与深层语义特征进行更加高效的融合,最终生成更具判别力与鲁棒性的特征表示,从而提升模型在复杂背景干扰下对红外小目标的检测性能。

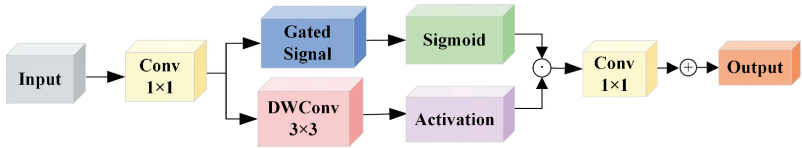


图 5 CGLU 结构
Fig. 5 CGLU structural

2.2 gConvC3 模块

在复杂背景下的无人机红外小目标检测任务中,现有卷积优化模块 RepC3 (repetitive convolutional contextualization with 3×3 kernels) 虽然能够在多特征融合中提升特征表达能力,但在特征选择性和自适应性方面仍存在一定短板。RepC3 模块基于跨阶段局部 (cross stage partial, CSP)^[22] 结构,主干采用重参数化卷积 (re-parameterizable convolution, RepConv) 堆叠,其主要依赖的局部卷积操作无法有效捕捉长距离依赖,这种操作在面对红外图像中信号微弱的小目标和复杂的背景杂波时,缺乏对关键信息的自适应筛选能力,难以有效区分目标与背景^[23]。

为解决上述局限性,本文提出了基于门控卷积单元 (gated convolution block, gConv)^[24] 的 gConvC3 模块,结构如图 6 所示。gConvC3 通过门控机制对特征通道与空间信息进行自适应加权,动态调节不同特征在融合过程中的贡献度,该机制能选择性地增强在复杂背景中呈现为对比度低、尺寸微小、边缘模糊的目标热辐射特征,同时有效抑制来自背景干扰的无关特征。这种对关键目标特征的选择性增强和对背景干扰的抑制,极大地提升了网络在复杂红外场景下的判别能力。

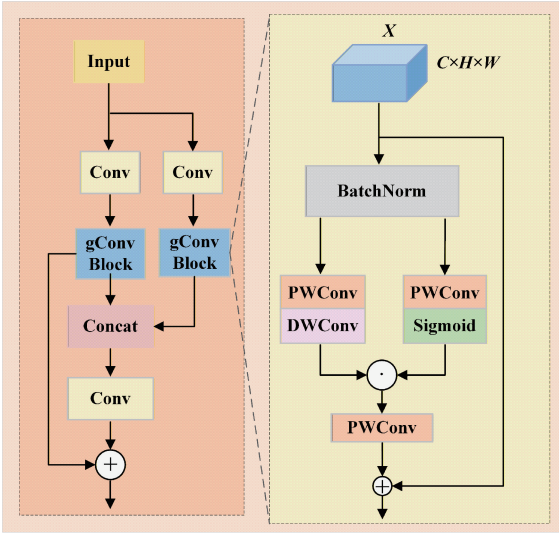


图 6 gConvC3 结构

Fig. 6 gConvC3 structure

gConvC3 是在 CSP 结构基础上,主干采用堆叠的门控卷积单元。如图 6 所示,每个单元将归一化后的特征 $\hat{x}$ 送入两个并行分支:一个分支 x_1 通过 1×1 卷积和激活函数生成自适应的门控权重;另一分支 x_2 则利用深度可分离卷积提取局部空间特征。最后,两个分支通过 Hadamard 乘法进行融合,利用门控权重对局部空间特征进行自适应的加权调节。这种设计使模型能够动态且自适应地增强微弱的小目标热辐射特征,同时抑制非相关的背景杂波信号。调节后的特征再通过 1×1 卷积和残差连接 x 形成最终输出。其核心的门控融合过程可表示为:

$$y = x + PWConv(x_1 \odot x_2) \quad (2)$$

其中, PWConv 表示逐点卷积, x_1 为门控分支, x_2 为特征提取分支, $\odot$ 表示 Hadamard 乘法。

2.3 骨干网络改进

在无人机红外小目标检测任务中,骨干网络的设计直接影响模型在复杂环境下的特征提取效果。由于红外成像常存在分辨率受限、目标信号微弱及背景噪声严重等问题,目标往往表现为低对比度、模糊的热辐射特征,极易在复杂背景中淹没。因此,骨干网络不仅要对小尺度目标细节有敏感捕捉能力,还必须兼顾无人机平台所需的计算效率。然而,传统骨干网络在逐层下采样中易导致小尺度特征信息快速丢失,难以同时满足检测精度与效率的双重需求。

为解决上述问题,本文对骨干网络进行优化,引入了部分卷积 (partial convolution, PConv)^[25] 与残差结构结合的 FasterBlock 模块,对 BasicBlock 进行改进。其核心思想在于通过差异化结构设计减少冗余计算与内存访问,同时保持对空间细节特征的充分提取。FasterBlock 相较于传统 BasicBlock 在保持残差连接优势的基础上,对卷积单元和特征通道的利用方式进行了优化,使得网络能够以更低的计算代价实现对小尺度目标特征的高效捕获。红外图像往往包含大面积信息量低的平坦背景区域,结合 PConv 机制,骨干网络在提取红外小目标关键信息时避免了对无关通道的冗余计算,将计算资源集中用于提取包含关键目标信号的特征。这种设计使得网络能够在降低模型参数量与计算量的同时,保持对小尺度目标 (如低对比度、模糊特征) 的敏感捕捉能力,从而更有效地满足无人机红外小目标检测对精度与计算效率的双重要求。改进后的骨干网络结构如图 7 所示。

标准卷积 (convolution, Conv) 对每个输入通道与卷积核进行全连接操作,如图 8(a) 所示,在红外图像处理中,这意味着模型会对大量信息量低的背景区域 (如天空、地面) 进行同等强度的冗余计算,这不仅导致了计算资源的冗余消耗,同时也增加了过拟合背景噪声的风险。为了缓解这一问题,骨干网络在设计时引入部分卷积 PConv 以降低冗余计算。如图 8(b) 所示, PConv 利用特征图中的空间连续性和通道之间的关联性,选择输入特征图的第一部分或最后一部分连续 c_p 个通道进行卷积运算,而其余通道保持不变,从而将计算力集中在包含关键热信号的通道上,在保持特征表达能力同时优化计算资源的利用率。部分比例 (partial ratio) 是指在卷积过程中,所选通道与总通道数的比例,用公式表示为:

$$r = \frac{c_p}{c} \quad (3)$$

其中, c_p 表示选择的参与计算的通道数, c 是输入特征图的总通道数。当部分比例 $r = 1/4$ 时,由于卷积过程占用内存较少, PConv 的浮点运算量 (FLOPs) 仅为传统卷

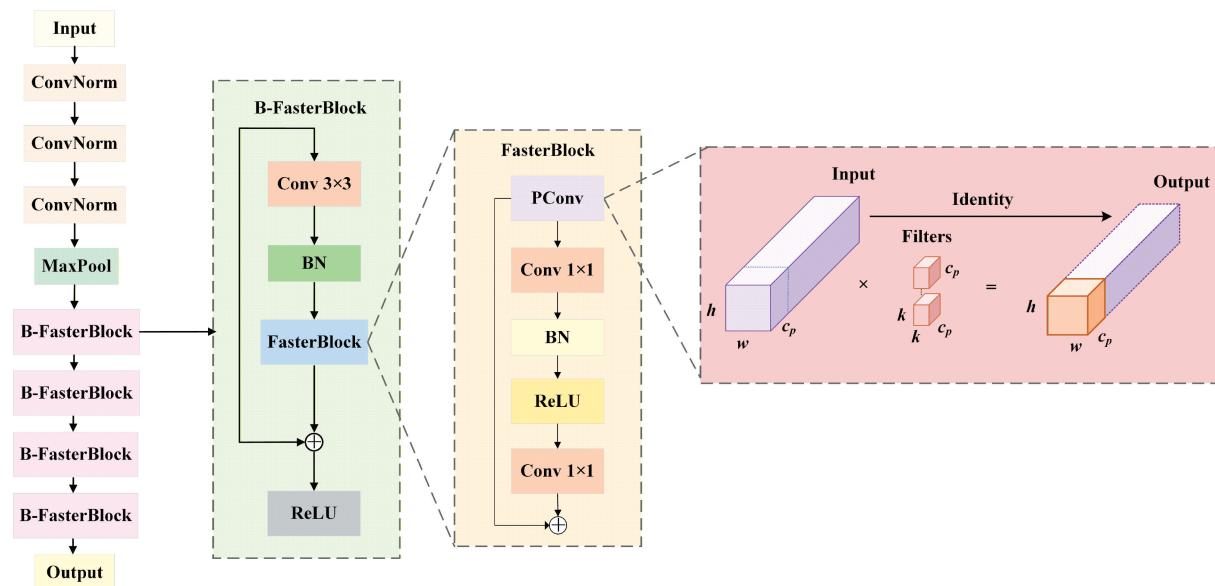


图 7 改进后的骨干网络结构

Fig. 7 Improved backbone network structure

积的 $1/16$, 这一优化减少了冗余的计算负担, 优化了计算资源的使用。计算公式为:

$$FLOPS = h \times w \times k^2 \times c_p^2 \quad (4)$$

其中, h 和 w 表示特征图的维度, k 表示卷积核大小, c_p 表示重参数化分支通道数。内存访问量 MAC 也减少到传统卷积的 $1/4$, 计算公式为:

$$MAC = h \times w \times 2c_p + k^2 \times c_p^2 = h \times w \times 2c_p \quad (5)$$

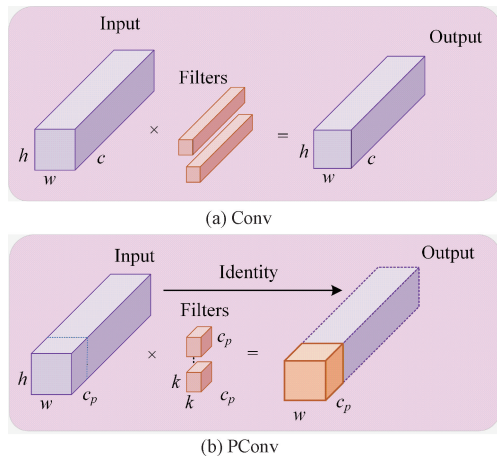


图 8 不同卷积模块结构

Fig. 8 Different convolutional module structures

由此可见, PConv 通过选择性地对输入特征图的部分通道进行卷积计算, 在保持良好特征提取能力的同时, 能够有效减少模型的计算量和参数量^[26]。为了在降低计算冗余的同时保持特征信息的完整性, FasterBlock 在 PConv 之后融入了倒残差连接, 依托其恒等映射路径增强了梯度传播的稳定性。这种设计将计算资源集中于关键热信号, 能够在极大降低冗余计算与内存访问的同时, 保持对红外

小目标存在的低对比度、模糊特征的敏感捕捉能力, 最终使网络在无人机红外小目标检测任务中能够更好地兼顾检测精度与效率。

3 实验设计与结果分析

3.1 数据集

本文实验采用公开的无人机红外图像数据集 HIT-UAV^[27] 进行模型训练与评估。该数据集包含 2 898 张由无人机在高空拍摄的红外热成像图像, 覆盖学校、停车场、道路、操场等多种真实场景环境。图像中涵盖 5 类目标, 包括行人(Person)、自行车(Bicycle)、汽车(Car)、其他车辆(Other Vehicle)以及不确定类别(Don't care), 总目标数量约为 2.5×10^4 个。为方便模型训练与验证, 本文按照 7:2:1 的比例将数据集划分为训练集、验证集和测试集, 对应训练图像 2 029 张, 验证图像 579 张, 测试图像 290 张。

3.2 实验环境

实验在 Windows11 操作系统下进行, CPU 选用 Intel Core i9-14900K 处理器, GPU 为 NVIDIA GeForce RTX 4090。深度学习框架基于 PyTorch 2.5.1, 搭配 CUDA 12.4, 编程语言为 Python3.10。实验所使用的输入图像尺寸统一为 640×640 , 训练轮数设置为 300 轮, Batch Size 设置为 4, 实验中的模型均未使用预训练权重, 训练阶段初始学习率设置为 0.000 1, 优化器选择 AdamW, 所有实验使用一致的超参数进行训练和验证。

3.3 评价指标

为了综合评估 PGF-RTDETR 模型应用于无人机红外小目标检测的有效性, 本文采用了精确率(precision, P)、召回率(recall, R)、平均精度(average precision, AP)、平均精度均值(mean average precision, mAP)、计算量(GFLOPs)、模

型参数量(Params)作为评估指标。其中,GFLOPs 用于衡量模型的时间复杂度,Params 则反映模型的空间复杂度。 P 、 R 和 mAP 用于评估模型的检测精度,值越大表示模型检测准确率越高;GFLOPs 和参数则衡量模型的轻量化程度,值越小表示模型越轻量,对硬件的性能要求也相应降低。性能指标计算公式为:

$$P = \frac{TP}{TP + FP}$$

(6)

$$R = \frac{TP}{TP + FN}$$

(7)

$$AP = \int_0^1 P(R) dR$$

(8)

$$mAP = \frac{\sum_{i=1}^n AP_i}{n}$$

(9)

其中, TP 为真正例,指模型正确预测出正样本的样本实例数量; FP 为假正例,指模型将负样本错误预测为正样本的样本实例数量; FN 为假负例,指将正样本错误预测为负样本的数量; n 表示类别数。

3.4 对比实验

为了验证本文所提算法在无人机红外小目标检测任务中的有效性,将 PGF-RTDETR 与当前主流检测算法进行了性能对比,包括 YOLOv5m、YOLOv8m、YOLOv9m、RTDETR-R34 等,选择了准确率(P)、召回率(R)、 $mAP50$ 、参数量(Params)、计算量(GFLOPs)作为模型性能指标。

实验结果如表 1 所示,在保持模型参数量和计算复杂度相对一致的前提下,PGF-RTDETR 整体性能表现出色。具体而言,在准确率上,PGF-RTDETR 达到 93.8%,相较于性能较好的 YOLOv8m(91.9%)与同系列的 RT-DETR-R34(89.5%),分别实现了 1.9%和 4.3%的显著提升,展现了优异的分类与定位能力^[28]。在 $mAP50$ 方面,PGF-RTDETR 达到了 85.7%,均优于其他主流的检测算法,其中 $mAP50$ 比基线模型 RT-DETR-R18 以及 YOLOv10m 分别高出 2.6%和 5.2%。与 Hybrid-DETR、FP-RTDETR

相比, $mAP50$ 分别高出 3.1%、3.3%。在模型效率上,PGF-RTDETR 的优势尤为突出,其参数量与计算量分别仅为 15.7 M 和 43.7 GFLOPs,适配无人机受限场景,利于快速部署。在召回率方面,PGF-RTDETR 比 RT-DETR-R34 略低 0.7%,这主要归因于 RT-DETR-R34 采用了更庞大的 ResNet-34 骨干网络,该网络具备更大的模型容量和更强的特征表征能力,使其能够捕捉到更广泛、甚至更模糊的目标特征,从而在一定程度上减少了漏检,但其准确率也更低^[29]。总体而言,改进后的 PGF-RTDETR 模型在检测精度以及参数量之间实现了良好的平衡,可有效支撑无人机红外小目标检测任务。

表 1 对比实验结果

Table 1 Comparative experimental results

模型	P/ %	R/ %	mAP50/ %	Params/ M	GFLOPs/ G
YOLOv5m	88.9	76.2	83.1	25.0	64.4
YOLOv8m	91.9	75.1	84.2	25.8	78.7
YOLOv9m	89.5	75.8	83.4	20.0	77.6
YOLOv10m	84.1	75.7	80.5	16.4	63.4
YOLOv11m	88.4	78.5	85.2	20.0	68.2
YOLOv12m	87.7	79.6	84.4	20.1	67.8
RT-DETR-R18	87.2	77.1	83.1	19.8	58.3
RT-DETR-R34	89.5	79.8	85.0	31.1	88.8
RT-DETR-R50	83.4	79.0	82.7	41.9	129.6
Hybrid-DETR ^[16]	91.4	73.3	82.6	—	—
FP-RTDETR ^[17]	89.1	—	82.4	20.0	49.3
本文	93.8	79.1	85.7	15.7	43.7

3.5 消融实验

本文算法使用 RT-DETR-R18 作为基线进行消融实验,通过消融实验验证 PCG、gConvC3 和 FasterBlock 3 个改进方法的有效性。实验结果如表 2 所示。

表 2 消融实验结果

Table 2 Results of ablation experiment

模型	PCG	gConvC3	FasterBlock	P/%	R/%	mAP50/%	GFLOPs/G	Params/M
A	—	—	—	87.2	77.1	83.1	58.3	19.8
B	✓	—	—	88.2	78.2	84.5	58.6	20.0
C	—	✓	—	89.5	78.0	84.1	51.1	18.7
D	—	—	✓	87.6	78.1	84.3	49.5	16.7
E	✓	✓	—	92.1	79.0	85.3	51.7	18.8
F	✓	—	✓	89.9	78.8	84.4	49.8	16.9
G	—	✓	✓	89.7	78.6	84.3	43.9	15.6
H	✓	✓	✓	93.8	79.1	85.7	43.7	15.7

由表 2 消融实验结果可知,PCG、gConvC3 与 FasterBlock 3 个模块的逐步引入均有效提升了模型性能,

并且体现出明显的协同增益作用。首先,用 PCG 替换基线模型的特征融合模块后,mAP50 指标提高了 1.4%,由 83.1% 升至 84.5%。同时,召回率由 77.1% 提升至 78.2%,GFLOPs 与参数量几乎保持不变,表明 PCG 模块的引入能够在几乎不增加计算量的前提下,有效增强模型对微弱红外目标的感知与召回能力。

其次,在 PCG 的基础上,进一步引入门控 gConvC3 模块后,mAP50 达到 85.3%,精确度从 88.2% 提升至 92.1%,性能得到了显著提升。同时,GFLOPs 进一步降低到 51.7 G。gConvC3 的引入通过门控卷积优化了卷积操作,能够在扩展感受野的同时保持高效的特征提取。这

使得模型不仅在精度上有了显著提升,计算量也显著降低,极大地提升了模型的运算效率。

最后,进一步将骨干网络的基础模块替换为 FasterBlock,构建了最终的 PGF-RTDETR 模型。FasterBlock 的最终引入使 mAP 值进一步提升了 0.4%,同时参数量相较基线下降了 20.7%,由 19.8 M 下降到 15.7 M,P 也进一步达到 93.8%。通过 FasterBlock 的引入,模型的计算复杂度进一步降低,同时保留了高效的特征提取能力,最终使得 PGF-RTDETR 在保证较高精度的同时,能够在实时推理任务中更好地应对复杂场景。图 9 直观展示了消融实验的对比结果。

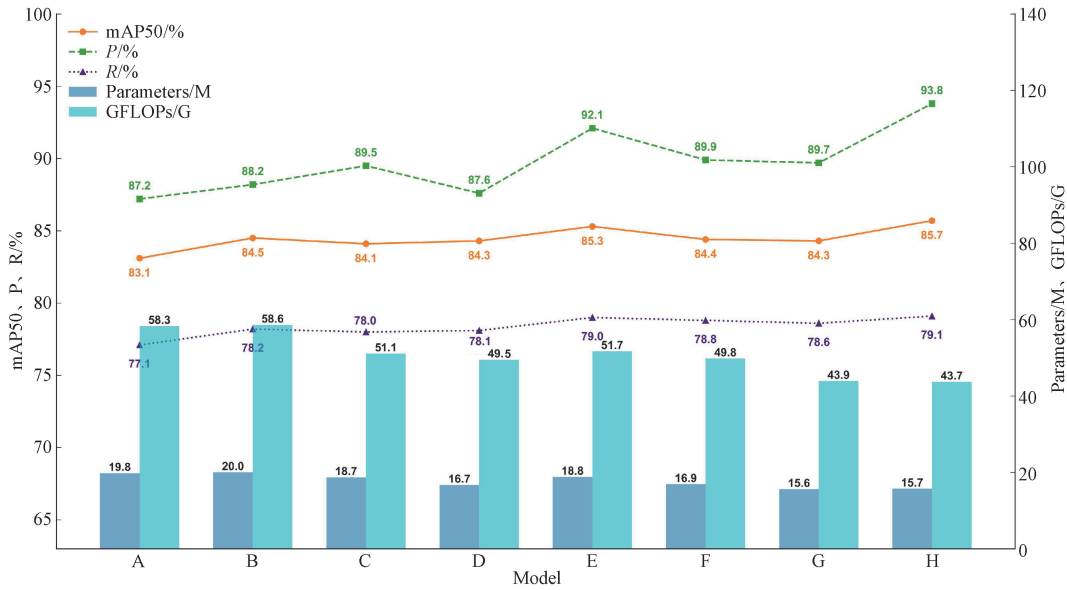


图 9 消融实验检测效果对比

Fig. 9 Ablation experiment detection effect comparison

3.6 稳定性分析

为验证本文所提改进的统计显著性与结果的可靠性,对核心的基线模型(RT-DETR-R18)与本文最终提出的模型(PGF-RTDETR)进行了 5 次独立的重复训练与评估。每次实验均采用不同的随机种子以模拟不同的初始化状态,其他所有超参数与实验设置均保持严格一致。实验结果如表 3 所示,所有性能指标均以“均值±标准差”的形式报告。其中均值反映了模型在多次运行下的平均性能水平,标准差量化了结果的离散程度或随机波动。较高的均值代表更优的平均性能,而较低的标准差则反映了较小的结果方差与性能波动,表明性能更为稳定可靠。

表 3 模型性能稳定性统计

Table 3 Model performance stability statistics %			
模型	P	R	mAP50
RT-DETR-R18	87.2±0.21	77.1±0.18	83.1±0.16
PGF-RTDETR	93.8±0.25	79.1±0.16	85.7±0.14

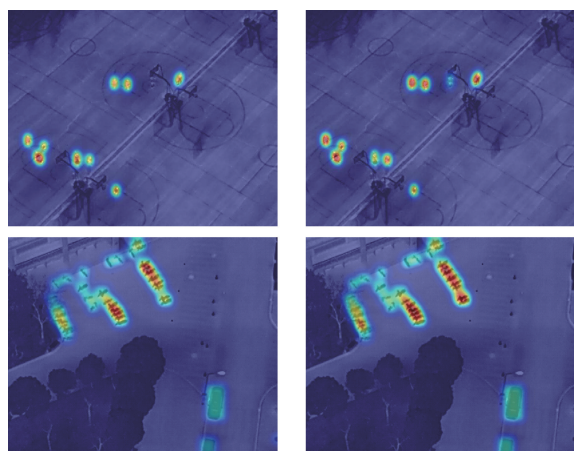
表 3 展示了 5 次独立实验的详细统计结果。从统计结果可以看出,本文提出的 PGF-RTDETR 模型在 5 次重复实验中的 mAP50 达到了 85.7%±0.14% 的平均性能,而基线模型的 mAP50 为 83.1%±0.16%。两个模型所得的标准差均处于较低的水平,表明其性能表现具有良好的稳定性。更关键的是,PGF-RTDETR 模型在每一次独立运行中的 mAP50 指标均稳定且显著地高于对应的基线模型,进一步证明了本文方法的改进是稳定且有效的。

3.7 可视化分析

1) 热力图对比

为了更加直观地评估 PGF-RTDETR 模型在复杂场景下的检测性能,本文采用热力图对基线模型和改进模型进行了可视化对比分析,对比改进前后对红外小目标的关注能力,如图 10 所示。第 1 列为 RT-DETR-R18 模型的可视化结果,第 2 列为 PGF-RTDETR 模型的可视化结果。图中颜色越深,表示模型对该区域的关注度越高。

通过热力图可以清晰观察到,原始模型的注意力缺乏



(a) RT-DETR-R18

(b) PGF-RTDETR

图 10 热力图对比

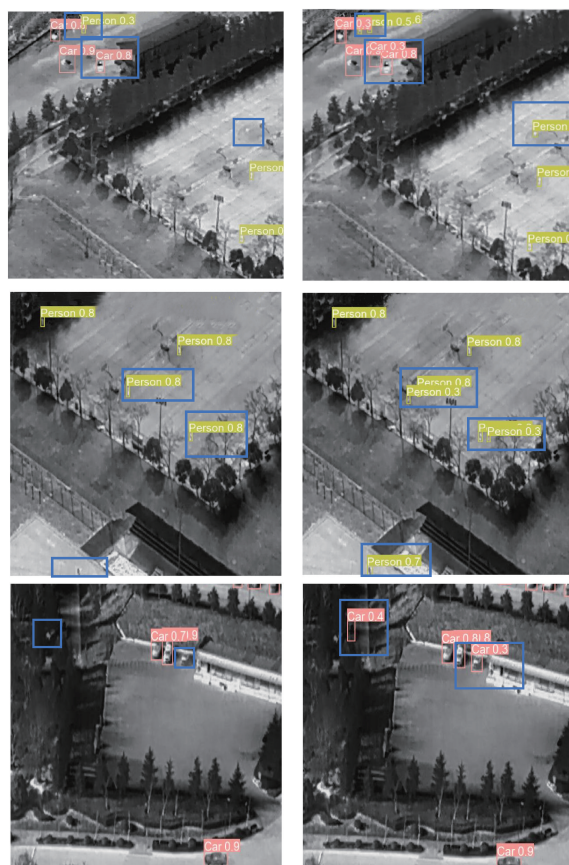
Fig. 10 Heat map comparison

有效的空间约束,其激活区域较为松散。对于行人等小目标的检测,原始模型的关注度相对较弱,难以准确聚焦目标区域,导致目标定位不够精确,特别在复杂背景下,原始模型的关注区域可能会受到周围无关元素的干扰,导致检测性能受到限制。相较而言,改进后的模型展现出更好的特征定位能力,其激活区域高度集中于目标主体,且有效规避了背景中无关元素的干扰。在检测车辆时,模型的关注区域聚焦于目标的核心部分,进一步提升了定位的准确性。整体来看,PGF-RTDETR 模型在复杂环境下的鲁棒性显著增强,并在检测红外小目标方面展现了更高的精度和更强的适应性。

2) 检测结果分析

为了深入分析本文改进算法在复杂无人机红外场景中的检测有效性,本文选取了 HIT-UAV 数据集中 3 组具有代表性的红外图像,进行可视化对比分析,检测效果如图 11 所示。图 11 中左列为原始 RT-DETR-R18 模型的检测结果,右列为 PGF-RTDETR 模型的表现。对比检测图显示,在第 1 组图像中,检测区域存在多辆静止车辆及部分被植被遮挡的行人,原始的 RT-DETR-R18 算法仅检测出部分车辆与个别行人,相比之下,PGF-RTDETR 模型不仅成功识别出更多车辆目标,还能更清晰地区分人物与车辆类别,表现出更强的背景干扰抑制能力。第 2 行场景主要包含多个密集分布的行人目标,在此情形下,PGF-RTDETR 模型显著提高了检测完整性,成功识别出更多目标实例,并准确标注重叠区域中的多个小目标,展现了更强的密集目标分离能力和细粒度特征建模能力。第 3 组图像展示了包含多辆密集停放汽车的复杂场景。原始 RT-DETR-R18 模型虽然识别出了场景中较大、较清晰的车辆目标,但对于中间区域停放的多辆小尺寸汽车则出现了明显的漏检。相比之下,PGF-RTDETR 模型不仅准确识别出了所有清晰目标,还成功检测到了被原始模型

遗漏的多个小型车辆,显著提升了检测的完整性和对小目标的召回能力。这一对比有力地证明了,本文的改进算法增强了模型对小尺寸和密集目标的特征分辨能力,有效降低了漏检率,在复杂场景下的检测性能更为可靠。



(a) RT-DETR-R18

(b) PGF-RTDETR

图 11 检测效果对比

Fig. 11 Detection results comparison

为了进一步说明改进算法的有效性,本文通过对比改进前后算法在各个目标类别上的 mAP50 表现进行分析,对比结果如图 12 所示,横坐标表示目标类别,纵坐标表示 mAP50 的值。结果显示,PGF-RTDETR 算法在检测各种目标时的平均精度比基准算法 RT-DETR-R18 有所提高,尤其在“Don't care”类别上,改进算法的性能优势最为显著。这表明 PGF-RTDETR 模型对于复杂背景、模糊不清或无需关注的目标具有更强的辨别和抑制能力,从而有效降低了背景的干扰,减少了误检,再次验证了改进算法 PGF-RTDETR 的有效性。

3) 局限性分析

本文提出的 PGF-RTDETR 算法在多项关键指标上均优于基准模型 RT-DETR-R18,在无人机红外小目标检测任务上实现了精度与效率的良好平衡。然而,在无人机红外检测的复杂真实环境中,极端场景仍对模型的鲁棒性提出挑战。为客观评估模型的性能边界并为未来的研究工

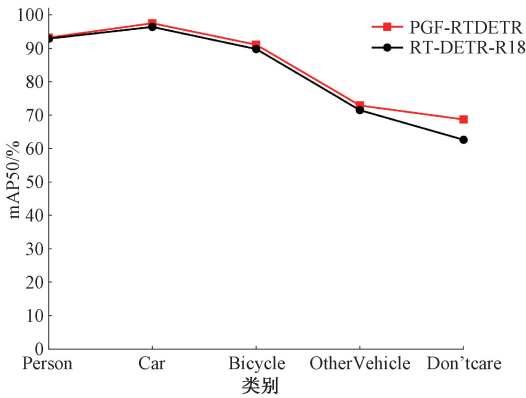


图 12 不同类别上的 mAP50 折线图

Fig. 12 mAP50 line charts on different categories

作指明方向,本节将对测试集中的典型挑战性案例进行可视化分析。如图 13 所示,图 13 中第 1 行为原始图片,第 2 行为其对应的 PGF-RTDETR 模型检测结果。

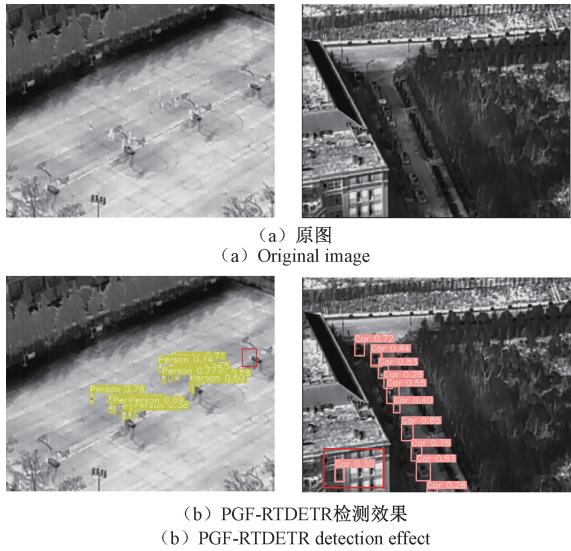


图 13 局限性案例分析

Fig. 13 Case analysis of limitations

左右两组对比检测图显示,在第 1 组图像中,球场右侧存在一个尺度极小、处于孤立状态的“Person”目标,在 PGF-RTDETR 模型的检测结果中,模型能够成功识别了聚集的主体人群,但遗漏了这名微弱的个体。这表明,PGF-RTDETR 模型虽已具备更优的细粒度特征感知能力,但当目标尺度极端微小且信噪比极低时,其特征信号在网络深层传递过程中依然面临被信息洪流淹没的风险。在第 2 组图像中,PGF-RTDETR 的检测结果中,模型将原始图片中建筑物外墙的窗户结构错误地识别为了“Car”目标。其主要原因是该窗户在红外图像下的网格状热特征,在形态和纹理上与高空俯视下的车辆(同样呈现为矩形热斑)产生了高度相似性。这表明,PGF-RTDETR 模型虽已具备更优的背景抑制能力,但在面对这种与目标特征高度

同构的欺骗性背景时,其高级语义辨别能力仍存在局限性。因此,进一步增强模型对极端微弱信号的感知鲁棒性,并提升其在高度同构欺骗性背景下的高级语义辨别能力,将是无人机红外小目标检测领域极具价值的研究方向。

3.8 泛化性实验

本文采用 FLIR 数据集验证 PGF-RTDETR 模型在不同数据集上的泛化性能。FLIR 数据集是一个专为热成像目标检测与分类设计的公开数据集,包含各类场景下的 14 452 张红外图像,涵盖 15 个类别。本实验筛选出 Person、Bicycle 和 Car 3 个关键类别,随机选取训练集 3 405 张、验证集 974 张、测试集 489 张,与主流算法进行对比分析。结果如表 4 所示,改进模型在 FLIR 数据集上的检测精度显著优于主流模型。与基准模型 RT-DETR-R18 相比,PGF-RTDETR 模型的检测准确率从 84.4% 提升至 84.8%,召回率从 74.1% 提升至 74.7%,mAP50 从 82.6% 提升至 83.2%。这些提升表明,PGF-RTDETR 模型在复杂背景和多样化目标的情况下,展现了良好的泛化能力和鲁棒性。

表 4 泛化性实验

Table 4 Generalization experiment			%
算法	P	R	mAP50
YOLOv5m	85.2	68.7	77.8
YOLOv8m	84.6	70.8	78.7
YOLOv10m	83.3	70.5	79.0
RT-DETR-R18	84.4	74.1	82.6
RT-DETR-R34	84.1	75.5	79.6
PGF-RTDETR	84.8	74.7	83.2

4 结 论

本文针对无人机红外图像中目标尺寸小、特征表达弱及尺度变化大等检测难题,提出了一种改进的 PGF-RTDETR 算法。该算法通过引入 PCG 深化了多尺度特征的语义融合,利用 gConvC3 增强了对关键信息的动态筛选,并借助高效的 FasterBlock 结构实现了骨干网络的轻量化。在 HIT-UAV 数据集上的实验结果表示,所提模型 PGF-RTDETR 较于原始模型 RT-DETR-R18,mAP50 和 P 分别提高了 2.6% 和 6.6%,模型参数量、计算量分别降低了 20.7%、25%。此外,在 FLIR 数据集上 mAP50 提高了 0.6%,P 提高了 0.4%。结果表明,该模型在检测精度和模型复杂度之间实现了较好的平衡,为无人机红外小目标检测领域提供了一种新的技术途径。未来的研究将继续提升模型在复杂环境下的检测鲁棒性,特别是强化对极端微弱信号的感知能力,并提升其在高度同构的欺骗性背景下的高级语义辨别能力,以适应更复杂的真实检测环

境。同时,持续优化模型的部署效率,通过探索更先进的模型压缩技术,进一步优化模型复杂度,提升实时检测的处理效率,增强其在资源受限设备中的部署适应性。

参考文献

- [1] 翁俊辉,成乐,黄曼莉,等. 基于 CS-YOLOv5s 的无人机航拍图像小目标检测[J]. 电子测量技术, 2024, 47(7): 157-162.
WENG J H, CHENG L, HUANG M L, et al. Small object detection of UAV aerial images based on CS-YOLOv5s[J]. Electronic Measurement Technology, 2024, 47(7): 157-162.
- [2] 葛诗然,刘睿恺,李娜,等. 智能抑制复杂空间背景下的红外弱小目标检测方法[J]. 红外与激光工程, 2024, 53(11): 275-287.
GE SH R, LIU R K, LI N, et al. Intelligent suppression of infrared weak target detection method in complex spatial background[J]. Infrared and Laser Engineering, 2024, 53(11): 275-287.
- [3] 张瑞芳,刘占占,程小辉,等. 基于改进 YOLOv8 的无人机红外目标检测算法[J]. 电子测量技术, 2025, 48(7): 46-54.
ZHANG R F, LIU ZH ZH, CHENG X H, et al. UAV infrared target detection algorithm based on improved YOLOv8 [J]. Electronic Measurement Technology, 2025, 48(7): 46-54.
- [4] CAO CH Q, WANG B, ZHANG W R, et al. An improved faster R-CNN for small object detection[J]. IEEE Access, 2019, 7: 106838-106846.
- [5] BHARATI P, PRAMANIK A. Deep learning techniques—R-CNN to mask R-CNN: A survey[J]. Computational Intelligence in Pattern Recognition; Proceedings of CIPR 2019, 2019: 657-668.
- [6] ZHANG Y F, CHI M M. Mask-R-FCN: A deep fusion network for semantic segmentation[J]. IEEE Access, 2020, 8: 155753-155765.
- [7] 侯艳丽,王娟. 基于 Efficientnet 的红外目标检测算法[J]. 电子测量技术, 2023, 46(16): 64-72.
HOU Y L, WANG J. Infrared target detection algorithm based on Efficientnet [J]. Electronic Measurement Technology, 2023, 46(16): 64-72.
- [8] KANG S H, PARK J S. Aligned matching: Improving small object detection in ssd[J]. Sensors, 2023, 23(5): 2589.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [10] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 213-229.
- [11] CAI X H, LAI Q X, WANG Y W, et al. Poly kernel inception network for remote sensing detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 27706-27716.
- [12] CAO L, WANG Q, LUO Y H, et al. YOLO-TSL: A lightweight target detection algorithm for UAV infrared images based on Triplet attention and Slim-neck[J]. Infrared Physics & Technology, 2024, 141: 105487.
- [13] ZHAO X F, XIA Y T, ZHANG W W, et al. YOLO-ViT-based method for unmanned aerial vehicle infrared vehicle target detection[J]. Remote Sensing, 2023, 15(15): 3778.
- [14] MOU X G, LEI SH, ZHOU X. YOLO-FR: A YOLOv5 infrared small target detection algorithm based on feature reassembly sampling method [J]. Sensors, 2023, 23(5): 2710.
- [15] ZHU J X, QIN CH, CHOI D. YOLO-SDLWD: YOLOv7-based small target detection network for infrared images in complex backgrounds[J]. Digital Communications and Networks, 2025, 11 (2): 269-279.
- [16] YANG M J, XU R Y, YANG CH Y, et al. Hybrid-DETR: A differentiated module-based model for object detection in remote sensing images [J]. Electronics, 2024, 13(24): 5014.
- [17] CHEN X, WANG G H. FP-RTDETR: Enhancing infrared ship detection with multi-scale feature fusion and lightweight design [J]. The Journal of Supercomputing, 2025, 81(8): 984.
- [18] 吕游,郑茜,齐欣宇,等. 基于改进 EfficientNet 的红外图像光伏组件故障识别研究[J]. 仪器仪表学报, 2024, 45(4): 175-184.
LYU Y, ZHENG Q, QI X Y, et al. A study on fault recognition of photovoltaic module with infrared images based on improved efficientnet [J]. Chinese Journal of Scientific Instrument, 2024, 45(4): 175-184.
- [19] 王慧云,赵俊生,王禹,等. 面向无人边防的复杂环境遮挡小目标检测算法[J]. 电子测量技术, 2024, 47(21): 168-177.
WANG H Y, ZHAO J SH, WANG Y, et al. Detection algorithm for small objects occluding in complex environments for unmanned border defense [J]. Electronic Measurement Technology, 2024, 47 (21): 168-177.

- [20] MENG W K, LUO Y D, LI X, et al. PolaFormer: Polarity-aware linear attention for vision transformers[J]. ArXiv preprint arXiv:2501.15061, 2025.
- [21] DAI SH. Transnext: Robust foveal visual perception for vision transformers[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 17773-17783.
- [22] LUO F, BIAN W X, JIE B, et al. ARBFPN-YOLOv8: Auxiliary reversible bidirectional feature pyramid network for UAV small target detection[J]. Signal, Image and Video Processing, 2024, 19(1): 63.
- [23] 陈枫赞, 李鹏. 基于 STR-DETR 的轻量化 PCB 缺陷检测算法[J]. 电子测量与仪器学报, 2025, 39(6): 30-40.
- CHEN F Y, LI P. Lightweight PCB defect detection algorithm based on STR-DETR [J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(6): 30-40.
- [24] SONG Y D, ZHOU Y, QIAN H, et al. Rethinking performance gains in image dehazing networks[J]. ArXiv preprint arXiv:2209.11448, 2022.
- [25] CHEN J R, KAO S, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.
- [26] 党宏社, 高宛蓉, 孟饶辰, 等. 融合多尺度特征的轻量级航拍目标检测算法[J]. 电子测量技术, 2024, 47(16): 92-99.
- DANG H SH, GAO W R, MENG R CH, et al. Lightweight aerial target detection algorithm integrating multi-scale features [J]. Electronic Measurement Technology, 2024, 47(16): 92-99.
- [27] SUO J SH, WANG T Y, ZHANG X ZH, et al. HIT-UAV: A high-altitude infrared thermal dataset for unmanned aerial vehicle-based object detection[J]. Scientific Data, 2023, 10(1): 227.
- [28] SEN C, SINGH P, GUPTA K, et al. UAV based YOLOV-8 optimization technique to detect the small size and high speed drone in different light conditions[C]. 2024 2nd International Conference on Disruptive Technologies(ICDT), 2024: 1057-1061.
- [29] WANG S, JIANG H P, LI ZH J, et al. PHSI-RTDETR: A lightweight infrared small target detection algorithm based on UAV aerial photography[J]. Drones, 2024, 8(6): 240.

作者简介

王亦君, 硕士研究生, 主要研究方向为图像处理、目标检测。

E-mail: wyjun6680@163.com

张露(通信作者), 博士, 讲师, 主要研究方向为机器学习与大数据处理。

E-mail: luz@cw Xu.edu.cn

李燕, 博士, 教授, 主要研究方向为智能预测与控制。

E-mail: 2225083469@qq.com