

基于改进YOLOv5模型的车载红外目标检测

杨新祺 刘洋 白雪琼 刘皓皎 陈亚洲

Vehicle-mounted infrared target detection based on improved YOLOv5 model

YANG Xinqi, LIU Yang, BAI Xueqiong, LIU Haojiao, CHEN Yazhou

引用本文:

杨新祺, 刘洋, 白雪琼, 等. 基于改进YOLOv5模型的车载红外目标检测[J]. 应用光学, 2025, 46(5): 1044–1053. DOI: 10.5768/JAO202546.0502004

YANG Xinqi, LIU Yang, BAI Xueqiong, et al. Vehicle-mounted infrared target detection based on improved YOLOv5 model[J]. Journal of Applied Optics, 2025, 46(5): 1044–1053. DOI: 10.5768/JAO202546.0502004

在线阅读 View online: <https://doi.org/10.5768/JAO202546.0502004>

您可能感兴趣的其他文章

Articles you may be interested in

基于改进SSD的车辆小目标检测方法

Detecting method of small vehicle targets based on improved SSD

应用光学. 2020, 41(1): 150–155 <https://doi.org/10.5768/JAO202041.0103004>

基于改进YOLOv5s模型的红外弱小目标检测方法

Infrared dim target detection method based on improved YOLOv5s model

应用光学. 2024, 45(5): 975–981 <https://doi.org/10.5768/JAO202445.0502005>

基于改进YOLOv5s的曲面光学镜片缺陷检测

Defect detection of curved optical lenses based on improved YOLOv5s

应用光学. 2024, 45(4): 781–789 <https://doi.org/10.5768/JAO202445.0403003>

基于深度学习的粒子场数字全息成像研究进展

Research progress of particle field digital holography based on deep learning

应用光学. 2020, 41(4): 662–674 <https://doi.org/10.5768/JAO202041.0409003>

基于IDOU-YOLO的红外图像无人机目标检测算法

Infrared image UAV target detection algorithm based on IDOU-YOLO

应用光学. 2024, 45(4): 723–731 <https://doi.org/10.5768/JAO202445.0402001>

复杂背景下红外图像弱小目标检测

Dim-small targets detection of infrared images in complex backgrounds

应用光学. 2021, 42(4): 643–650 <https://doi.org/10.5768/JAO202142.0402002>



关注微信公众号, 获得更多资讯信息

文章编号: 1002-2082 (2025) 05-1044-10

引用格式: 杨新祺, 刘洋, 白雪琼, 等. 基于改进 YOLOv5 模型的车载红外目标检测 [J]. 应用光学, 2025, 46(5): 1044-1053.

YANG Xinqi, LIU Yang, BAI Xueqiong, et al. Vehicle-mounted infrared target detection based on improved YOLOv5 model[J]. Journal of Applied Optics, 2025, 46(5): 1044-1053.



在线阅读

基于改进 YOLOv5 模型的车载红外目标检测

杨新祺¹, 刘 洋¹, 白雪琼¹, 刘皓皎¹, 陈亚洲²

(1. 北京信息科技大学 仪器科学与光电工程学院, 北京 100192; 2. 中国人民解放军 95821 部, 甘肃 酒泉 735000)

摘 要: 针对车载红外目标检测算法中, 因红外小目标特征少、对比度低导致的检测精度低的问题, 提出了 FE-YOLOv5s 车载红外目标检测模型。首先, 在主干网络中引入 PPA 模块以弥补下采样操作带来的特征信息损失; 其次, 基于 AFPN 网络和 DASI 模块, 构建特征融合网络 DAFPN, 直接融合非临近尺度层的特征信息, 同时将五种不同尺度层的特征信息都纳入最终的特征输出, 进一步增强目标的特征表达能力, 并重构目标预测层, 以提高模型对小目标的识别率; 最后, 采用 MPDIoU 作为边界框损失函数以提升目标的定位精度。实验结果表明, 本文所提出的模型在 FLIR 数据集和艾睿车载红外数据集上的 mAP@0.5 分别达到了 88.1% 和 84.7%, 相比 YOLOv5s 提升了 7.2% 和 3.2%, 缓解了模型对红外小目标错检漏检的问题, 且检测速度为 100 帧/s, 为嵌入式部署提供了可行性。

关键词: 目标检测; 深度学习; 特征融合; 神经网络; 红外图像

中图分类号: TN219; TP391

文献标志码: A

DOI: 10.5768/JAO202546.0502004

Vehicle-mounted infrared target detection based on improved YOLOv5 model

YANG Xinqi¹, LIU Yang¹, BAI Xueqiong¹, LIU Haojiao¹, CHEN Yazhou²

(1. School of Instrumentation Science and Opto-electronics Engineering, Beijing Information Science and Technology University, Beijing 100192, China; 2. No. 95821 Unit of PLA, Jiuquan 735000, China)

Abstract: To address the issue of low detection accuracy in vehicle-mounted infrared target detection algorithms, caused by the limited features and low contrast of small targets, this paper proposed the FE-YOLOv5s vehicle-mounted infrared target detection model. Firstly, the PPA module was integrated into the backbone network to mitigate the feature information loss caused by downsampling operations. Secondly, based on the AFPN network and the DASI module, we developed a feature fusion network termed DAFPN. This network directly fuses feature information from non-adjacent scale layers and incorporates features from five different scale layers into the final feature output, further enhancing the feature representation capability of the targets. The target prediction layer was reconstructed to improve the model's recognition rate of small targets. Finally, MPDIoU was adopted as the bounding box loss function to enhance target localization accuracy. Experimental results demonstrated that the proposed model achieves mAP@0.5 scores of 88.1% on the FLIR dataset and 84.7% on the iRay vehicle-mounted infrared dataset, representing improvements of 7.2% and 3.2% over the YOLOv5s model. It effectively alleviates the problem of false and missed detections of infrared small targets while operating at 100 FPS, which confirms its potential for embedded deployment.

Key words: target detection; deep learning; feature fusion; neural network; infrared image

收稿日期: 2025-04-03; 修回日期: 2025-05-24

基金项目: 北京市自然科学基金 (4244105)

第一作者: 杨新祺, E-mail: 1059534115@qq.com

通信作者: 刘洋, E-mail: 20141891@bistu.edu.cn

引言

自动驾驶技术^[1]作为未来交通领域的重要发展方向,不仅为人们的生活带来了极大的便利,同时也对车辆的安全性和可靠性提出了更高的要求。许多新能源车企采用“多传感器融合”方案,通过多个激光雷达^[2]辅助摄像头感知路况。然而,相较于纯视觉方案,多传感器融合方案需要更高的成本,并且不同传感器获取的数据可能存在冲突,从而导致算法在决策时出现误判,引发潜在危险。纯视觉方案^[3]能够更好地模拟人类驾驶,通过视觉信号传递到深度神经网络进行实时处理分析,进而做出驾驶决策。但是,在恶劣天气或光照条件不佳的情况下,基于可见光的视觉目标检测方法^[4]的感知能力会受到限制。相比之下,红外图像目标检测不仅探测距离远,而且抗干扰能力强。因此,在上述情况下,以车载红外系统^[5]作为一种补充视觉检测方案尤为重要。

目标检测算法主要分为传统算法和基于深度学习的算法。传统方法如方向梯度直方图(histogram of oriented gradients, HOG)^[6]和尺度不变特征变换(scale-invariant feature transform, SIFT)^[7],它们依赖手工设计目标特征来提取信息,计算成本高且泛化能力差。深度学习方法中,基于区域的卷积神经网络(regions with convolutional neural networks, R-CNN)^[8]通过卷积神经网络(convolutional neural networks, CNN)提取特征并结合候选区域机制,显著提升了检测精度,但计算效率较低。后续的 Fast R-CNN^[9]和 Faster R-CNN^[10]虽然优化了速度,但仍难以满足实时性的需求。以 YOLO^[11]系列为代表的单阶段目标检测算法将目标检测任务视为回归问题,通过单次向前传播直接预测边界框和类别,大幅提高了检测效率,更适合实时应用。张建君^[12]等人基于 YOLOv5s 提出改进红外弱小目标检测方法,通过 TCC 模块增强特征提取并引入 Ghost 模块轻量化网络,提高了检测精度;刘皓皎^[13]等人在 YOLOv5 模型中引入坐标注意力机制,增强对红外目标特征的提取能力,同时添加小目标检测层增强红外小目标的检测精度;娄树理^[14]等人提出改进 YOLOX-S 的红外舰船检测算法,采用深度可分离卷积、ECANet 注意力机制和 CIoU 损失函数,使平均精度达到 98%,较原模型提升了 3%,检测速度为 56 帧/s,实现了轻量化检测。尽管以上改进模型在红外目标检测中取得进展,但仍存在有待完善之处。一方面,在红外目标特征信息提取过程中,下采样操作会不可避免地导致信息丢

失,而这些模型均未对该问题予以考量。另一方面,在特征融合阶段,它们并未将所有特征层的信息全面纳入融合范畴。这使得模型在红外目标检测的特征信息增强与精度提升方面仍存在较大的优化空间。

本文针对上述问题以及红外小目标检测的难点,提出了改进 YOLOv5 红外目标检测模型。主要研究内容包括:针对特征提取过程中特征信息丢失的问题,在主干网络中引入并行化感知注意力机制(parallelized patch-aware attention, PPA)^[15];基于渐进特征金字塔网络(asymptotic feature pyramid network, AFPN)^[16]和多维度感知选择性集成模块(dimension-aware selective integration, DASI)^[15]构建了维度感知渐进式特征金字塔网络(dimension-aware asymptotic feature pyramid net, DAFPN),直接融合非相邻特征层的特征信息,将五种不同尺度的特征信息都纳入最终的特征表达,进一步增强目标的特征信息表达能力,并重构多尺度目标预测层^[17],提升对小目标的检测精度;采用最小点距离交并比(minimum points distance intersection over union, MPDIoU)^[18]进一步提升复杂道路场景中红外目标的检测精度。

1 改进的 YOLOv5s 算法模型

1.1 FE-YOLOv5s 算法

本文采用 YOLOv5 系列算法作为基础框架,其核心优势体现在高效率推理与轻量化部署的平衡性。YOLOv5 包含 5 种不同型号,分别为 YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l 以及 YOLOv5x,这些型号在参数量、模型体积和检测精度方面呈现依次递增的趋势,而检测速度则依次递减。具体而言,YOLOv5n 因特征提取过浅,难以满足实际需求,而其他几种模型虽在精度上有优势,但检测速度较慢。最终选定 YOLOv5s 作为基础模型,并对其展开了优化改进,提出了 FE-YOLOv5s(feature enhancement YOLOv5s)模型,如图 1 所示。首先,为在下采样过程中保留更多的目标特征细节,在主干网络中的第 4、8、16 倍下采样前引入 PPA 模块。随后,将主干网络提取的不同尺度特征信息送入 DAFPN 特征融合网络,从低层特征信息开始逐步融合高层信息,通过 DASI 模块将最高层和最低层特征信息整合,进一步增强三种不同尺度预测层特征的表达能力。最后,引入了新的边界框损失函数 MPDIoU,以加快收敛并提高边界框定位精度。

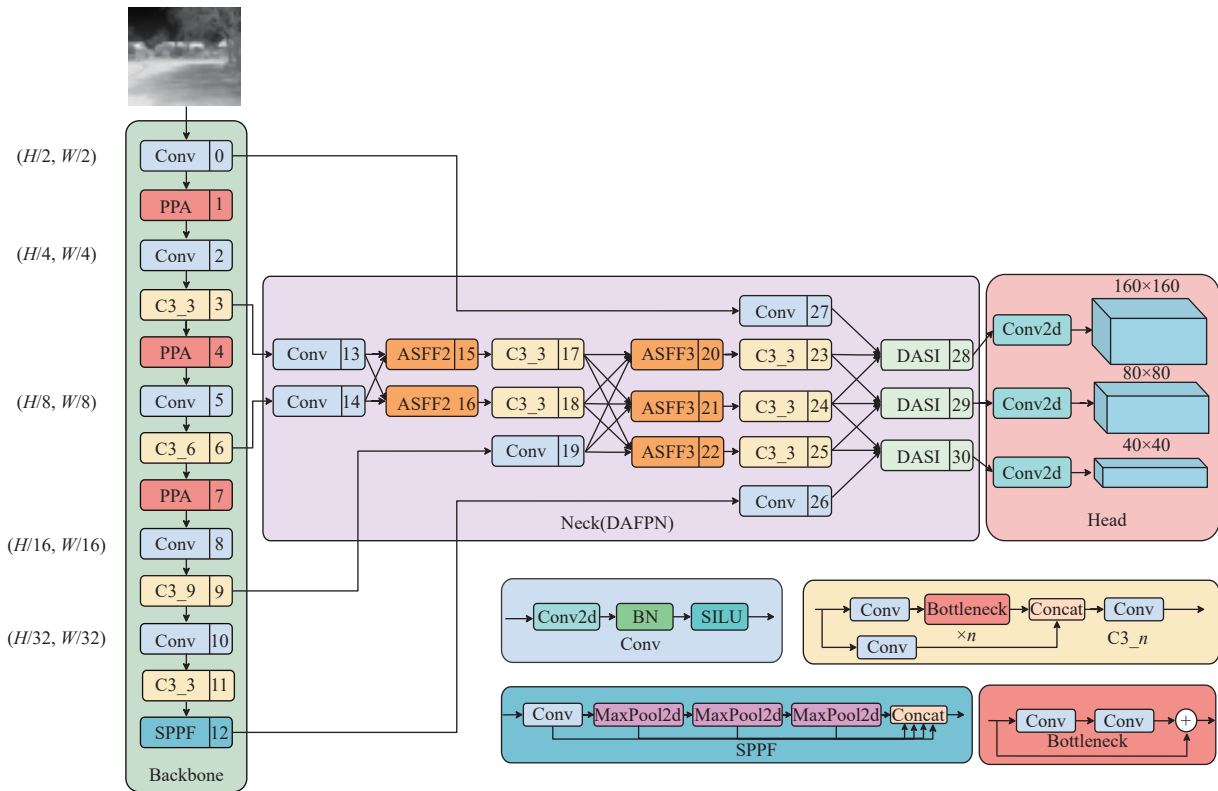


图1 FE-YOLOv5s 模型结构

Fig. 1 Structure of the FE-YOLOv5s model

1.2 PPA 模块

PPA 模块采用并行支路特征提取策略和注意力机制,有效地保留了目标在下采样过程中可能丢失的特征信息,同时增强了目标的特征表达能力,从而提升目标检测的精度。首先,给定输入特征张量 $F \in \mathbb{R}^{H' \times W' \times C}$,其中 H' 、 W' 、 C 分别代表输入张量的高、宽以及通道维度。经过逐点卷积 (point-wise

convolution, PWConv) 调整输入通道数,生成新的张量 $F' \in \mathbb{R}^{H' \times W' \times C'}$ 。然后,通过三个并行支路来进行多尺度特征提取,分别为全局卷积支路、串行卷积支路、局部卷积支路。经过这三个支路处理后,分别生成的特征张量为 $F_{\text{global}} \in \mathbb{R}^{H' \times W' \times C'}$ 、 $F_{\text{conv}} \in \mathbb{R}^{H' \times W' \times C'}$ 和 $F_{\text{local}} \in \mathbb{R}^{H' \times W' \times C'}$ 。最后,将这三个支路的特征张量相加,得到张量 $\tilde{F} \in \mathbb{R}^{H' \times W' \times C'}$ 。模块结构如图2所示。

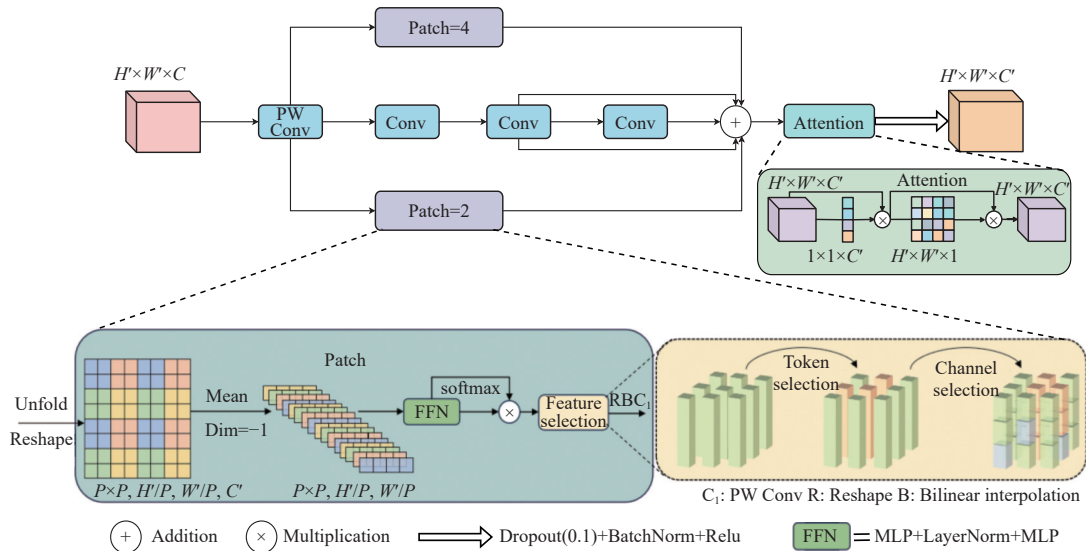


图2 PPA 模块结构图

Fig. 2 Structure diagram of the PPA module

利用展开和重塑操作, 将 $F' \in \mathbb{R}^{H' \times W' \times C'}$ 划分为一组在空间维度非重叠的补丁, 之后采用前反馈神经网络 FFN 进行线性计算, 生成更复杂的特征表示, 通过激活函数得到线性计算特征在空间维度上的概率分布, 并相应地调整其权重。

将加权结果表示为 $(t_i)_{i=1}^d$, 其中 $t_i \in \mathbb{R}^C$ 表示第 i 个单元, $d = (H' \times W') / (p \times p)$ 表示一共有 d 个单元。接下来进行特征选择, 对每个单元的操作分为两步: 单元选择和通道选择。单元选择的输出为 $\text{sim}(t_i, \xi) \cdot t_i$, 其中 $\xi \in \mathbb{R}^C$ 为任务嵌入参数, 表示哪些单元与任务相关, 并通过余弦相似度 $\text{sim}(\cdot, \cdot)$ 计算与 t_i 的相似性, 获得与哪些单元更相关后, 每个 t_i 都会根据其任务的相关性重新加权。第二步是对每个单元的通道进行线性变换 P , $P \in \mathbb{R}^{C' \times C'}$ 是与任务相关的可学习参数矩阵, 得到输出 $\hat{t} = P \cdot \text{sim}(t_i, \xi) \cdot t_i$, 最终生成 $F_{\text{global}} \in \mathbb{R}^{H' \times W' \times C'}$ 、 $F_{\text{local}} \in \mathbb{R}^{H' \times W' \times C'}$ 。串行支路则由三个 3×3 卷积层组成的串行卷积, 卷积后得到三个不同的输出: $F_{\text{conv1}} \in \mathbb{R}^{H' \times W' \times C'}$ 、 $F_{\text{conv2}} \in \mathbb{R}^{H' \times W' \times C'}$ 、 $F_{\text{conv3}} \in \mathbb{R}^{H' \times W' \times C'}$ 。将三个输出相加得到 $F_{\text{conv}} \in \mathbb{R}^{H' \times W' \times C'}$, 最后将三个支路相加的特征张量先后通过通道注意力和空间注意力进行自适应特征增强, 可以让模型聚焦于对当前任务有帮助的区域和特征。

1.3 DAFPN 网络

特征融合层的核心作用是整合来自不同层次和尺度的特征信息, 增强模型的表达能力, 提升模型的性能和准确性。低层特征包含丰富的细节信息, 如边缘、纹理等, 对于精确定位目标和细节描述至关重要; 高层特征则通常具有更强的语义信息, 有助于理解目标的整体形态和类别。因此, 直接融合低层和高层特征, 可以同时利用细节特征和语义特征, 形成更加丰富、全面的特征表达, 从而提升模型对目标的理解和检测能力。

传统特征融合网络主要是自顶向下和自下向上的融合方式, 比如特征金字塔 (feature pyramid network, FPN) 和路径聚合网络 (path aggregation network, PAN), 这些方法依赖于相邻层的特征交互, 难以实现非相邻特征层的直接信息融合, 从而导致特征信息的丢失和退化, 影响融合效果。为解决该问题, 本文基于 AFPN 网络和 DASI 模块, 设计了 DAFPN 网络, 通过 AFPN 实现了非相邻特征层之间的直接交互, 与此同时渐进式的融合方式避免了非相邻层语义差异过大; 通过 DASI 将最高层和最低层特征也并入整合, 根据目标尺寸自适应地选择高低层特征进行融合, 进一步提高三种尺度预测层中目标的特征表达能力。DAFPN 融合网络

实现了五个不同尺度特征层都能够贡献于最终的特征表达, 从而更全面地提升检测任务的性能。

1.3.1 AFPN 网络

AFPN 网络采用了渐进式特征融合策略, 在主干网络自上而下提取特征时, 生成一组从低到高的不同尺度的特征 $\{a_1, a_2, a_3\}$, 首先从 a_1, a_2 低层特征开始输入到 AFPN 网络中进行融合, 然后添加 a_3 , 最终输出三种融合后的不同尺度的特征图。这种渐进式融合方式有效避免了非邻近层之间过大的语义差距, 使得非邻近层之间直接融合更充分, 其网络结构如图 3 所示。

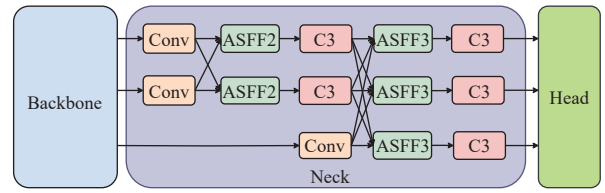


图 3 AFPN 结构图

Fig. 3 Structure diagram of the AFPN

在 AFPN 网络融合过程中, 自适应空间特征融合 (adaptive spatial feature fusion, ASFF)^[19] 起到了关键作用, 实现了非相邻层级信息的直接融合, 并能够自适应调整不同层级的重要程度进行融合, ASFF 的算法模型如图 4 所示。ASFF 算法分为两步, 维度对齐和自适应融合。维度对齐是通过上采样或下采样, 将不同尺度的特征层调整为相同的通道和尺度, 为后续的特征融合做准备。在自适应融合阶段, ASFF 对三个层级特征图自适应分配不同权重比例, 来决定各层级特征的共享程度, 并据此进行特征融合。ASFF3 的运算公式为

$$y^k = \lambda^k \cdot x_1^{1 \rightarrow k} + \beta^k \cdot x_2^{2 \rightarrow k} + \gamma^k \cdot x_3^{3 \rightarrow k} \quad (1)$$

式中: k 为 ASFF3 输出的当前层级; y^k 为输出结果; $\lambda^k, \beta^k, \gamma^k$ 分别为不同尺寸特征图尺度对齐后自适应选择的权重参数且 $\lambda^k + \beta^k + \gamma^k = 1$; x_1, x_2, x_3 则分别为来自 level1、level2、level3 的特征。

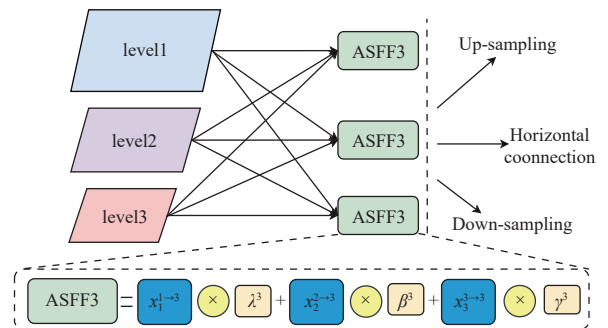


图 4 ASFF 算法模型

Fig. 4 ASFF algorithm model

1.3.2 DASI 模块

在多级下采样过程中,高层特征可能导致小目标特征信息丢失,而低层特征缺乏足够的上下文信息。为了更好地结合两者的优点,根据目标尺寸大小自适应选择合适的特征进行融合,特别是在处理小目标任务时,DASI 模块能够有效地在高层和低层特征之间进行选择并融合,从而提升目标的显著性,其结构如图5所示。

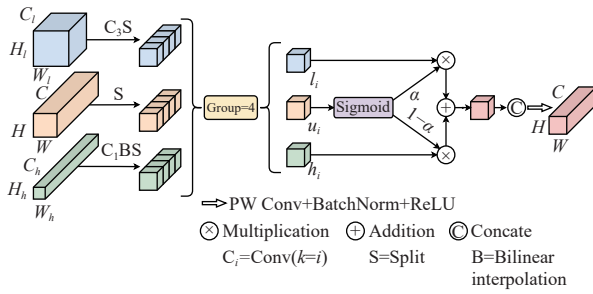


图5 DASI 模块结构图

Fig. 5 Structure diagram of the DASI module

DASI 模块通过卷积和双线性插值等方式将高层特征 (\$F_h \in \mathbb{R}^{H_h \times W_h \times C_h}\$)、低层特征 (\$F_l \in \mathbb{R}^{H_l \times W_l \times C_l}\$)、当前层特征 (\$F_u \in \mathbb{R}^{H_u \times W_u \times C_u}\$) 进行空间维度和通道维度上的对齐。之后,将三个维度特征在通道维度上平均分成四部分,得到 \$(h_i)_{i=1}^{(4)} \in \mathbb{R}^{H \times W \times \frac{C}{4}}\$、\$(l_i)_{i=1}^{(4)} \in \mathbb{R}^{H \times W \times \frac{C}{4}}\$、\$(u_i)_{i=1}^{(4)} \in \mathbb{R}^{H \times W \times \frac{C}{4}}\$, 其中, \$h_i\$、\$l_i\$、\$u_i\$ 分别表示高层特征、低层特征和当前层特征的第 \$i\$ 个划分区域。过程如式(2)~式(5)所示。首先,通过 Sigmoid 函数对当前层特征的第 \$i\$ 个分区 \$u_i\$ 生成一个权重系数 \$\alpha\$:

$$\alpha = S(u_i) \quad (2)$$

该系数用于控制高层特征和低层特征的融合程度,当 \$\alpha > 0.5\$ 时,模型倾向于使用低层特征,当 \$\alpha < 0.5\$ 时,模型则倾向于使用高层特征。根据权重 \$\alpha\$ 融合高层特征 \$h_i\$ 和低层特征 \$l_i\$ 的输出结果 \$u'_i\$ 可表示为

$$u'_i = \alpha l_i + (1 - \alpha) h_i \quad (3)$$

根据上式可得到四个融合结果 \$u'_1\$、\$u'_2\$、\$u'_3\$、\$u'_4\$, 合并可得特征图 \$F'_u\$:

$$F'_u = [u'_1, u'_2, u'_3, u'_4] \quad (4)$$

经卷积、批量归一化和 ReLU 激活函数处理后生成最终融合输出 \$\hat{F}_u \in \mathbb{R}^{H \times W \times C}\$:

$$\hat{F}_u = \delta(\varphi(\text{Conv}(F'_u))) \quad (5)$$

式中: \$\delta(\cdot)\$ 为 ReLU 激活函数; \$\varphi(\cdot)\$ 为批量归一化; Conv 为卷积。

1.4 重构预测层

由于多次下采样会导致更多小目标特征信息的丢失,为了使模型更好地适应对红外小目标的检测需求,模型引入了小目标检测层,进一步保留高分辨率下小目标的特征信息。YOLOv5s 原模型的预测层输出三种尺度的特征图,分别为 \$20 \times 20\$、\$40 \times 40\$、\$80 \times 80\$, 用来检测三种目标尺寸。根据红外小目标的检测需求控制参数量,FE-YOLOv5s 删减了 \$20 \times 20\$ 大尺寸感受野的输出,并新增了 \$160 \times 160\$ 小尺寸感受野的输出,通过在更高分辨率的特征图上提取小目标的详细特征信息,能够更有效地提高检测精度。

1.5 改进损失函数

YOLOv5 的损失函数包含三部分:分类损失、边界框回归损失和置信度损失。其中,分类损失和置信度损失采用二值交叉熵 (binary cross-entropy, BCE), 边界框回归损失使用完全交并比 (complete intersection over union, CIoU), 与传统的 IoU 相比, CIoU 在计算中增加了中心点距离和宽高比的惩罚项,从而优化边界框的重叠、位置和形状,提升回归精度。然而, CIoU 在宽高比相同时无法区分框的大小差异,并且在中心点重合时仅等同于 IoU。因此,本文在模型中采用 MPDIoU 损失函数,引入角点的最小欧式距离,利用左上角和右下角的距离来避免宽高比一致时存在的问题,且在中心重合时优化预测位置,从而提升损失函数的精度。该损失函数表达式为

$$d_1^2 = (x_1^p - x_1^g)^2 + (y_1^p - y_1^g)^2 \quad (6)$$

$$d_2^2 = (x_2^p - x_2^g)^2 + (y_2^p - y_2^g)^2 \quad (7)$$

$$L_{\text{MPDIoU}} = 1 - L_{\text{IoU}} - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \quad (8)$$

式中: \$L_{\text{IoU}}\$ 为 IoU 损失,表示预测边界框和真实边界框的交集和并集之比; \$(x_1^p, y_1^p)\$ 和 \$(x_2^p, y_2^p)\$ 分别为预测框的左上角和右下角的坐标; \$(x_1^g, y_1^g)\$ 和 \$(x_2^g, y_2^g)\$ 分别为真实框的左上角和右下角坐标; p 和 g 分别为预测框和真实框; w 和 h 分别为输入图像的宽度和高度,用于将距离进行归一化以消除尺度上的影响。

2 实验结果及分析

2.1 实验环境及参数设置

模型实验采用 Pytorch 深度学习框架,训练轮次为 150 轮,批大小为 64,采用 Adam 优化器作为

模型的权重更新策略, 初始学习率为 0.001, 一阶矩估计的指数衰减率为 0.937。输入端的图像尺寸为 640×640。权重衰减 (weight_decay) 为 0.000 5。为增强数据多样性和提高模型的泛化能力, 启用了 mosaic 数据增强技术。实验环境如表 1 所示。

表 1 实验环境

Table 1 Experimental environment

Experimental platform	Model
OS	Ubuntu20.04 64bit
CPU	Intel(R)Corei9-10900X
GPU	NVIDIA RTX 3090×2
Python	3.10.13
Pytorch	1.12.1
CUDA	CUDA 11.6

2.2 数据集

本文使用自动驾驶环境感知的 FLIR^[20] 红外热成像数据集, 该数据集是由安装在车上的热成像相机获取, 数据采集于街道和高速公路环境, 具有真实性、复杂性以及挑战性等特点。该数据集包含 15 个类别, 其中小尺度(行人)目标和中尺度(车辆)目标在数据集中占比最多, 同时也是道路环境感知较为密集的两类目标。数据集一共有 11 886 张图片, 将其按照比例 8 : 1 : 1 划分为训练集、验证集和测试集, 训练集 9 509 张, 验证集和测试集分别为 1 189 张和 1 188 张。数据集部分图片如图 6 所示。



图 6 FLIR 数据集

Fig. 6 FLIR dataset

2.3 评价指标

在目标检测任务当中, 评价指标是评估模型性能的重要工具, 通常以精确度 (precision, P), 召回率 (recall, R), 平均精确度 (average precision, AP), 平均精确度均值 (mean average precision, mAP), PR 曲线等作为模型的评价指标。精确度和召回率的计算公式分别为

$$\eta_P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (9)$$

$$\eta_R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (10)$$

式中: 精确度 η_P 表示模型预测为正例的样本中实际为正例的比例; 召回率 η_R 衡量实际为正例的样本中模型预测为正例的比例; N_{TP} 为模型正确预测为正例的样本数量; N_{FP} 为模型错误预测正例的样本数量; N_{FN} 为模型错误预测为负例的样本数。

AP 是单个类别的评价指标, 表示在给定 IoU 阈值下, 精确度-召回率曲线下的面积, 其计算公式为

$$P_{AP} = \int_0^1 \eta_P(\eta_R) d\eta_R \quad (11)$$

式中: P_{AP} 为单个类别的 AP 值, 通常介于 0 和 1 之间, 值越大表明模型的性能越好,

mAP 是对所有类别的 AP 值取平均数, 是整个模型在所有类别上的综合评价指标, 其计算公式为

$$P_{mAP} = \frac{1}{C} \sum_{i=1}^C P_{AP_i} \quad (12)$$

式中: P_{mAP} 为 mAP 值; C 为类别总数; P_{AP_i} 为第 i 类的 AP 值。AP 和 mAP 是目标检测模型评价的核心指标, AP 衡量单个类别的检测精度, mAP 通常用于比较不同模型的总体检测能力, 是衡量模型性能的标准之一。

2.4 消融实验

以 YOLOv5s 为基准, 进行了 7 种消融实验 (A~G)。其中, 前四组实验主要评估单个改进点添加的效果, 后三组则评估组合对模型性能的优化。如表 2 所示, 在 A、B、C、D 实验中, 每个单独添加的模块均有效提升了模型性能, 当改进特征融合层为 DAFPN 时, 性能提升最为显著, mAP 相比 YOLOv5s 提高了 4.4%, 表明非相邻特征层的直接融合策略和将五种不同尺度特征层纳入特征融合, 能够将目标的特征信息和语义信息更有效地进行融合, 增强了目标的特征表达能力, 从而提高了检测精度。引入 PPA 模块的改进模型相比于原始模型, 平均精度提升了 2.5%, 证明 PPA 模块通过并行特征提取的策略和注意力机制, 有效缓解了下采样导致的目标特征丢失问题。实验 D 重构目标预测层, 较原模型提高了 3.1%, 表明优化特征输出结构, 保留更多小目标特征信息, 对检测小目标更有利。改进损失函数 MPDIoU 使模型的平均精度提升了 1.6%, 进一步证明 MPDIoU 通过引入角点的最小欧式距离能更精确地匹配边界框的位置

和尺寸,从而提高定位精度。在 E、F、G 实验中,逐步添加模块组合进一步验证了每个优化方向对平均精度均值的积极贡献。

表 2 消融实验结果
Table 2 Ablation experimental results

	Dectet	PPA	DAFPN	MPDIoU	mAP@0.5%
YOLOv5s					80.9
A				√	82.5
B			√		85.3
C		√			83.4
D	√				84.0
E	√	√			86.4
F	√	√	√		87.5
G	√	√	√	√	88.1

2.5 多尺度目标检测性能评估

为了评估改进后的模型对红外小目标的检测性能,使用 cocotools 的评估指标对逐步引入改进点的改进模型和基准模型 YOLOv5s 进行了比较。根据目标所占图像的像素数进行划分,小目标所占像素小于 32×32 像素,中目标所占像素为 32×32 到 96×96 像素,大目标所占像素数大于 96×96 像素。从表 3 可以看出,逐步优化的模型对小目标的检测精度持续提升,对中目标的检测性能相对基准模型保持稳定,对大目标的检测性能出现轻微波动。其中,引入 PPA 模块后,由于注意力权重向小目标特征信息偏移,模型的中大目标检测性能有所下降。加入 DAFPNet 网络后,其渐进式的融合五种不同层特征的策略,使模型对中目标和大目标的检测性能有所回升。FE-YOLOv5s 相较于基准模型,对小目标和中目标的检测性能分别提升了 9.7% 和 0.8%,保持模型对中大目标检测能力的同时,显著提升模型对红外小目标的感知性能。

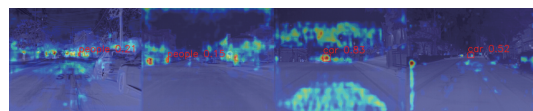
表 3 不同尺度的目标检测评估

Table 3 Evaluation of object detection at different scales

Model	mAP@0.5%		
	Small/%	Medium/%	Large/%
YOLOv5s	65.5	93.0	90.0
+Dectet	70.3	93.2	89.2
+Dectet+PPA	72.0	91.8	88.6
+Dectet+PPA+DAFPN	74.7	93.2	89.0
+Dectet+PPA+DAFPN+MPDIoU	75.2	93.8	89.3

2.6 模型特征信息关注度分析

为了更好地验证模型对于小目标的检测性能,本文对 YOLOv5s 和 FE-YOLOv5s 模型在同一场景下对同一检测目标的热力图进行对比分析,如图 7 所示。图 7(a) 为 YOLOv5s 的目标关注度热力图,图 7(b) 为 FE-YOLOv5s 的目标关注度热力图。从图 7 中可看出,YOLOv5s 在检测到目标时,其热力分布较为分散,说明关注度分散在图像的多个区域,并未聚焦于目标本身。相比之下,FE-YOLOv5s 则对目标特征给予了更高的关注度,检测信息聚焦于目标特征,受到来自背景噪声的影响较弱,从而有效提升了目标检测定位精度和抗干扰能力。同时,可以观察到 FE-YOLOv5s 在检测同一目标时的置信度高于 YOLOv5s,反映了改进模型对小目标的特征信息更敏感。



(a) YOLOv5s 模型热力图



(b) FE-YOLOv5s 模型热力图

图 7 YOLOv5s 和 FE-YOLOv5s 模型小目标热力对比图
Fig. 7 Heatmap comparison of small target between YOLOv5s and FE-YOLOv5s models

2.7 模型检测能力对比分析

为验证模型检测能力,本文选取了目标检测模型 Faster R-CNN、YOLOv3、YOLOv3-tiny、YOLOv5s、YOLOv6s、YOLOv8n 及针对小目标检测模型 HIC-YOLOv5s^[21] 和 STASPPNet^[22] 与 FE-YOLOv5s 在 FLIR 数据集上的性能表现进行了对比分析,主要评价指标为精确度、召回率、平均精度、参数量 (Parameters) 以及计算量 (GFLOPS) 和帧率 (Frames Per Second, FPS)。实验结果如表 4 所示。Faster R-CNN 是经典的两阶段目标检测模型,它也是所有对比模型中参数量和计算量最高的,但检测速度最低,其 mAP 值仅仅高于 YOLOv3-tiny 和 YOLOv8n 两个轻量级模型。YOLOv3 是所对比的 YOLO 基础模型版本中 mAP 值最高的,达到 84.7%,但其参数量和计算量仅次于 Faster R-CNN。YOLOv8n 因其轻量化的结构在所有对比模型中参数量和计算量最低,所以它的检测速度达到最高为 184 帧/s,但 mAP 相较于改进模型低 9.1%。小目标检测模

型 HIC-YOLOv5s 和 STASPPNet 表现出优异的检测性能, mAP 分别达到了 85.3% 和 85.5%, 但相较于本文的改进模型分别低 2.8% 和 2.5%。改进模型相较于原始 YOLOv5s 模型, 虽然在检测速度上有所降低, 但其精确度、召回率、mAP 分别提升了 2.1%、8%、7.2%, 改进模型综合考量了参数量、模型大小、检测速度后, 提升了模型均值平均精度 mAP, 检测速度为 100 帧/s。在 640×512 分辨率的红外图像输入下, 改进模型在 Hi3559A 开发板上的图像处理时间延时为 9 ms, 具备车载平台嵌入式部署的可行性。

表 4 多模型对比实验

Table 4 Comparative experiments of multiple models							
Model	P /%	R /%	mAP@ 0.5%	Parameters /M	Size /MB	GFLOPS	FPS
Faster R-CNN	64.0	53.8	80.4	99.2	330.6	440.0	33
YOLOv3-tiny	76.8	55.1	61.6	8.6	17.4	12.9	181
YOLOv3	86.3	76.2	84.7	61.5	123.5	154.6	82
YOLOv5s	85.0	70.8	80.9	7.0	14.5	16.0	126
YOLOv6s	85.6	72.7	82.3	16.3	32.8	44.0	112
YOLOv8n	84.5	69.4	79.0	3.0	6.3	8.1	184
HIC-YOLOv5	85.6	74.9	85.3	9.3	19.6	30.6	120
STASPPNet	86.3	77.5	85.5	8.1	17.3	47.1	116
FE-YOLOv5s	87.1	78.8	88.1	7.1	14.5	69.5	100

如图 8 所示通过可视化图像对比分析了 A、B、C、D 四组不同的场景下, YOLOv5s 和 FE-YOLOv5s 模型的检测效果图, 图 8(a) 为 YOLOv5s 模型的检测效果图, 图 8(b) 为 FE-YOLOv5s 的模型检测效果图。为更好地展示模型误检漏检情况, 将图 8(b) 进行放大展示, 如图 8(c) 所示, 其中绿色框代表原始模型与 FE-YOLOv5s 模型相比错检漏检的示意框。FE-YOLOv5s 模型在图 8(c) 中 A、B 场景分别检测到了远处密集的行人红外小目标以及远处稀疏的行人小目标。从 C、D 场景检测效果来看, 改进后的模型能够检测出较远处红外车辆目标以及多个车辆目标遮挡情况下特征信息不足时的目标。综合四种场景, 原始模型在检测小目标的行人和车辆以及多目标重叠时, 会出现错检漏检的情况。而 FE-YOLOv5s 对遮挡目标以及远距离小目标都能更好地定位其在图像中的位置, 并且准确分类, 展现了改进后模型良好的道路环境感知能力。

表 5 艾睿红外数据集检测性能对比表

Table 5 IRay Infrared dataset detection performance comparison							
Model	P/%	R/%	mAP@0.5%	Parameters/M	Size/MB	GFLOPS	FPS
YOLOv5s	84.2	73.2	81.5	7.0	14.5	16.0	126
FE-YOLOv5s	84.6	78.3	84.7	7.1	14.5	69.5	100

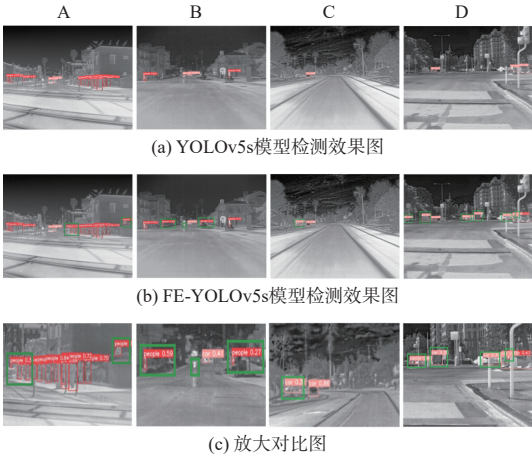


图 8 FE-YOLOv5s 模型和 YOLOv5s 模型检测对比图
Fig. 8 Comparison of detection results between FE-YOLOv5s model and YOLOv5s model

2.8 艾睿光电红外数据集验证

为了验证改进的算法针对不同数据集依旧有很好的检测性能, 采用在道路行驶条件下获取的艾睿光电车载红外数据集进行验证。该数据集图像分辨率为 640×512, 共计 10 815 张图像, 数据集含 5 类目标标签, 包括行人、小汽车、公交车、自行车、卡车。将数据集按照 8 : 1 : 1 进行划分, 用测试集进行验证。数据集部分图像如图 9 所示。

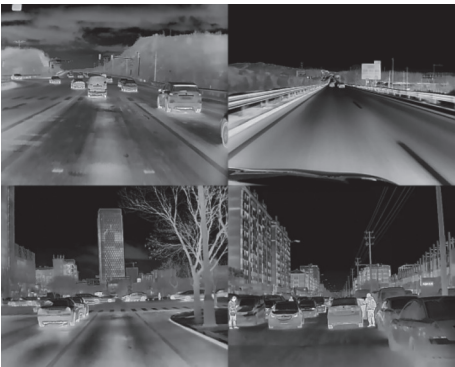


图 9 艾睿红外数据集
Fig. 9 IRay infrared dataset

文中将改进后的 FE-YOLOv5s 算法和 YOLOv5s 算法在艾睿车载红外数据集上训练后的性能对比如表 5 所示, 实验结果表明, 改进后的模型较原模型 YOLOv5s, 在平均精度上高了 3.2%, 召回率提高了 5.1%, 证明了文中改进算法的有效性和泛化能力。

3 结论

本文以 YOLOv5 算法模型为基准,提出了改进的 FE-YOLOv5s 模型,改进模型分别从主干、特征融合网络、检测头、损失函数四个方面进行改进研究。实验结果表明,四个改进方面都对红外目标检测的精度有积极贡献,改进模型对红外小目标检测精度更高且抗干扰能力更强。在 FLIR 车载红外数据集的测试中,改进模型相较于其他算法的平均精度更高,与原 YOLOv5s 模型相比改进的 FE-YOLOv5s 精度提升 2.1%,召回率提升 8%,mAP 提升 7.2%,改善了原模型检测红外目标时存在错检漏检的问题,同时检测速度为 100 帧/s,最后,在艾睿车载红外数据集上进行对比实验,实验结果表明,改进后的模型相较于原模型仍有较好的检测性能,改进算法可为自动驾驶中的环境感知方案提供参考。

参考文献:

- [1] 王聪, 胡文, 李文博, 等. 社会认知自动驾驶[J]. *机械工程学报*, 2023, 59(20): 304-324.
WANG Cong, HU Wen, LI Wenbo, et al. Social cognitive autopilot[J]. *Journal of Mechanical Engineering*, 2023, 59(20): 304-324.
- [2] 许立, 凌铭, 王若铭. 基于改进 complex-YOLO 激光雷达的行人检测算法研究[J]. *激光技术*, 2025, 49(1): 106-112.
XU Li, LING Ming, WANG Ruoming. Research on improved complex-YOLO LiDAR pedestrian detection algorithm[J]. *Laser Technology*, 2025, 49(1): 106-112.
- [3] 曹家乐, 李亚利, 孙汉卿, 等. 基于深度学习的视觉目标检测技术综述[J]. *中国图象图形学报*, 2022, 27(6): 1697-1722.
CAO Jiale, LI Yali, SUN Hanqing, et al. A survey on deep learning based visual object detection[J]. *Journal of Image and Graphics*, 2022, 27(6): 1697-1722.
- [4] 宋存利, 柴伟琴, 张雪松. 基于改进 YOLO v5 算法的道路小目标检测[J]. *系统工程与电子技术*, 2024, 46(10): 3271-3278.
SONG Cunli, CHAI Weiqin, ZHANG Xuesong. Road small target detection based on improved YOLOv5 algorithm[J]. *Systems Engineering and Electronics*, 2024, 46(10): 3271-3278.
- [5] 刘彦磊, 李孟喆, 王宣宣. 轻量型 YOLOv5s 车载红外图像目标检测[J]. *中国光学*, 2023, 16(5): 1045-1055.
LIU Yanlei, LI Mengzhe, WANG Xuanxuan. Lightweight YOLOv5s vehicle infrared image target detection[J]. *Chinese Optics*, 2023, 16(5): 1045-1055.
- [6] BILAL M, HANIF M S. Benchmark revision for HOG-SVM pedestrian detector through reinvigorated training and evaluation methodologies[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 21(3): 1277-1287.
- [7] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [8] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 580-587.
- [9] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015: 1440-1448.
- [10] REN S Q, HE K M, 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.
- [11] 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). Las Vegas, NV, USA: IEEE, 2016: 779-788.
- [12] 张建君, 陈玉丹, 刘玉玲, 等. 基于改进 YOLOv5s 模型的红外弱小目标检测方法[J]. *应用光学*, 2024, 45(5): 975-981.
ZHANG Jianjun, CHEN Yudan, LIU Yuling, et al. Infrared dim target detection method based on improved YOLOv5s model[J]. *Journal of Applied Optics*, 2024, 45(5): 975-981.
- [13] 刘皓皎, 刘力双, 张明淳. 基于 YOLOv5 改进的红外目标检测算法[J]. *激光技术*, 2024, 48(4): 534-541.
LIU Haojiao, LIU Lishuang, ZHANG Mingchun. An improved infrared object detection algorithm based on YOLOv5[J]. *Laser Technology*, 2024, 48(4): 534-541.
- [14] 娄树理, 王岩, 郭建勤, 等. 改进 YOLOX-S 的红外舰船目标检测算法[J]. *应用光学*, 2023, 44(5): 1054-1060.
LOU Shuli, WANG Yan, GUO Jianqin, et al. Infrared ship target detection algorithm based on improved

- YOLOX-S[J]. *Journal of Applied Optics*, 2023, 44(5): 1054-1060.
- [15] XU S B, ZHENG S C, XU W H, et al. HCF-Net: hierarchical context fusion network for infrared small object detection[C]//2024 IEEE International Conference on Multimedia and Expo (ICME). Niagara Falls, ON, Canada: IEEE, 2024: 1-6.
- [16] YANG G Y, LEI J, ZHU Z K, et al. AFPN: asymptotic feature pyramid network for object detection[C]//2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Honolulu, Oahu, HI, USA: IEEE, 2023: 2184-2189.
- [17] 何宇豪, 易明发, 周先存, 等. 基于改进的 YOLOv5 的无人机图像小目标检测[J]. *智能系统学报*, 2024, 19(3): 635-645.
- HE Yuhao, YI Mingfa, ZHOU Xiancun, et al. UAV image small-target detection based on improved YOLOv5[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(3): 635-645.
- [18] MA S L, XU Y. MPDIoU: aLoss for efficient and accurate bounding box regression[J]. *ArXiv*, 2023: 2307.07662.
- [19] LIU S T, HUANG D, WANG Y H. Learning spatial fusion for single-shot object detection[J]. *ArXiv*, 2019: 1911.09516 .
- [20] TELEDYNE FLIR. Free FLIR thermal dataset for algorithm training[EB/OL]. [2025-03-06] <https://www.flir.com/oem/adas/adasdataset-form>
- [21] TANG S Y, ZHANG S, FANG Y N. HIC-YOLOv5: improved YOLOv5 for small object detection[C]//2024 IEEE International Conference on Robotics and Automation (ICRA). Yokohama, Japan: IEEE, 2024: 6614-6619.
- [22] WU H, HUANG X, HE C H, et al. Infrared small target detection with swin transformer-based multiscale atrous spatial pyramid pooling network[J]. *IEEE Transactions on Instrumentation and Measurement*, 2024, 74: 5003914.