

基于改进 YOLOv8n 的红外行人车辆检测算法

秦海洋^{1,2}, 谭功全^{1,2}, 邓豪^{1,2}, 王 峒^{1,2}, 蔡大洋^{1,2}, 文 力^{1,2}

(1. 四川轻化工大学自动化与信息工程学院, 四川 宜宾 644000; 2. 四川轻化工大学人工智能四川省重点实验室, 四川 宜宾 644000)

摘 要: 鉴于红外行人车辆图像分辨率低, 质量不佳, 噪声多等特点, 检测难度较大, 提出一种基于 YOLOV8 的红外图像行人车辆目标检测算法, 即 PSWG-YOLO。针对 YOLOv8n 网络, 增加 160×160 的极大特征图 P2 提高模型对行人小目标的检测能力。同时, 采用 SPD-Conv 部分代替原网络 stride-2 的卷积层, 提升对低分辨率图像的特征提取能力。此外, 将损失函数替换为 WIoU, 优化模型对低质量图像的处理。最后, 引入 Ghost 模块降低模型复杂度。实验结果表明, 改进后的 PSWG-YOLO 算法在保证较高的检测精度的同时, 显著减少了模型体积和参数量。与原 YOLOv8n 算法在公开红外数据集 FLIR_v2 上 P、R、mAP@0.5 分别提升 1.6%、6.3%、7.2%, 且参数量减少 16%, 模型大小减少 15.8%, 提高了红外场景下行人车辆检测的精度并易于部署。

关键词: 红外目标检测; YOLOV8; SPD-Conv; Ghost 模块; 损失函数

中图分类号: TP391.41; TN219 文献标识码: A DOI: 10.3969/j.issn.1001-5078.2025.01.019

Infrared pedestrian vehicle detection algorithm based on improved YOLOV8

QIN Hai-yang^{1,2}, TAN Gong-quan^{1,2}, DENG Hao^{1,2}, WANG Yao^{1,2}, CAI Da-yang^{1,2}, WEN Li^{1,2}

(1. School of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibin 644000, China;
2. Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science & Engineering, Yibin 644000, China)

Abstract: Given that infrared pedestrian-vehicle images are difficult to detect due to their low resolution, poor quality, and high noise, an infrared image pedestrian and vehicle target detection algorithm based on YOLOV8 is proposed, namely PSWG-YOLO. For the YOLOv8n network, a 160×160 maximum feature map P2 is added to improve the model's detection ability of pedestrian small targets. At the same time, the SPD-Conv part is used to replace the stride-2 convolutional layer of the original network to improve the feature extraction capability of low-resolution images. In addition, the loss function is replaced with WIoU to optimize the model's processing of low-quality images. Finally, the Ghost module is introduced to reduce model complexity. The experimental results show that the improved PSWG-YOLO algorithm significantly reduces the model volume and parameter amount while ensuring high detection accuracy. Compared with the original YOLOv8n algorithm, the P, R, and mAP@0.5 on the public infrared data set FLIR_v2 are increased by 1.6%, 6.3%, and 7.2% respectively, and the number of parameters is reduced by 16%, and the model size is reduced by 15.8%, which improves the accuracy of the pedestrian-vehicle detection in infrared scenarios and is easy to deploy.

Keywords: infrared target detection; YOLOV8; SPD-Conv; Ghost module; loss function

基金项目: 人工智能四川省重点实验室科研项目(No. 2019RYJ08)资助。

作者简介: 秦海洋(1997-), 男, 硕士研究生, 研究方向为目标检测、智能驾驶技术。E-mail: 1963019105@qq.com

通讯作者: 谭功全(1970-), 男, 硕士, 教授, 硕士生导师, 研究方向为控制理论与控制工程。E-mail: tgq77@163.com

收稿日期: 2024-04-01; 修订日期: 2024-05-06

1 引言

据相关调查,在雾霾等极端天气条件下,能见度降低,导致驾驶员难以准确辨识道路上的行人和车辆,易发生事故。而红外热成像仪具有对温度敏感、强大的夜视能力和抗天气干扰能力,能够在极端天气条件下高效地检测和分类目标^[1],能为驾驶员提供更清晰的驾驶视野,有效降低交通事故发生率,而目前的红外图像目标检测技术还不够成熟。因此,如何获取更有效的检测方法在当前领域显得尤为迫切。

传统的目标检测使用滑动窗口选取感兴趣区域,然后采用特定的特征提取算法,如 Haar 特征^[2]、HOG 特征^[3]等。接下来,用分类器进行分类,常用的分类器包括 SVM^[4],AdaBoost^[5]等。最后,根据分类结果进行目标检测和定位。这些方法在一定程度上能够实现目标检测,但受限于人工特征提取的准确性和分类器的泛化能力,检测精度有待提高。近年来深度学习方法在目标检测领域取得了显著的进展。基于深度学习的目标检测算法包括 One stage 和 Two stage 两类。相比 Two stage 检测算法,One stage 检测算法不仅能保持较高的检测精度还具有较高的检测速度。代表的算法有 YOLO 系列^[6-9],SSD 系列^[10-11]等。而 YOLO 在实现实时检测和保持较高准确率的同时,更适用于资源受限的平台,目前已成为红外目标检测的主流算法。如胡培帅^[12]等人在 YOLOv5 网络中使用 Mish 激活函数和 Transformer 模块,增加检测层,有效提升了对红外车辆和

行人的检测精度。郭志坚^[13]等人在 YOLOv4 中引入了 CA 注意力机制模块,设计 CSP-DBL 结构,并对 Head 部分进行裁剪,不仅提高了对红外车辆行人的检测精度和速度还轻量化了模型。Luo X^[14]等人提出了一种基于 YOLOv5 的实时性红外行人车辆检测方法,通过集成 FocalGIoU 损失函数提高收敛速度并使用分布式移位卷积 DSCConv 代替传统的卷积,减少了计算负担并在保持精度的同时提高了检测速度。

基于上述研究,本文针对红外图像的检测难点,进一步提出了一种基于 YOLOv8 的红外行人车辆检测算法 - PSWG-YOLO。首先,通过增加特大特征图,加强网络对远景小目标的特征提取能力;其次,引入 SPD-Conv^[15]提升对低分辨率图像的特征提取能力;接着,替换损失函数为 WIoU^[16],优化模型对低质量图像的处理;最后,采用 Ghost 轻量化模块^[17]重构 YOLOv8n 网络,降低模型的复杂度。

2 YOLOv8n 网络结构介绍

YOLOv8 是 Ultralytics 公司在 2023 年推出的最新 SOTA 模型,相比其他 YOLO 模型性能大幅提升。它提供了 N/S/M/L/X 不同尺度模型,以适应各种部署平台和场景。其核心结构包括 Backbone、Neck 和 Head 三部分,分别用于特征提取、特征融合和预测结果。为提高红外行人车辆检测精确度,且满足检测实时性。本文选取 YOLOv8n 进行改进, YOLOv8n 结构和主要模块如图 1 所示。

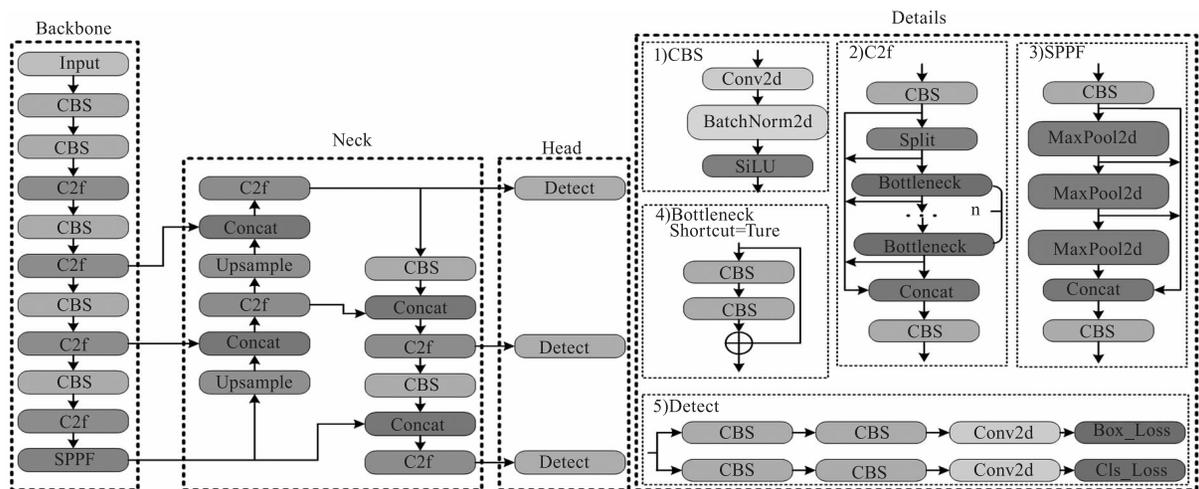


图 1 YOLOv8n 的网络结构及主要模块细节示意图

Fig.1 Schematic diagram of YOLOv8n's network structure and details of main modules

3 PSWG-YOLO 网络结构

为了提高对红外行人车辆的检测效果,本文

提出了一种基于 YOLOv8n 的 PSWG-YOLO 算法,在原有特征图上增加了对小目标检测的特大特征

2) 无步长卷积层: 随后, 将 SPD 层的输出作为带有滤波器 C_2 ($C_2 < \text{scale}^2 C_1$) 的无步长卷积层 ($\text{stride} = 1$) 的输入, 通过这一层的处理, 中间特征图 X' 被转换为新的特征图 X'' , 其尺寸调整为 $(S/\text{scale}, S/\text{scale}, C_2)$ 。在这一过程中, 特征图的宽度和高度保持不变, 仅对通道数进行了调整。总体而言, 特征图 X 在经过 SPD-Conv 层后, 其尺寸由原始的 (S, S, C_1) 转变为 $(S/\text{scale}, S/\text{scale}, C_2)$ 。这种设计可以减少因网络过度下采样可能导致的细节信息损失, 确保特征提取过程的有效性和精确性。

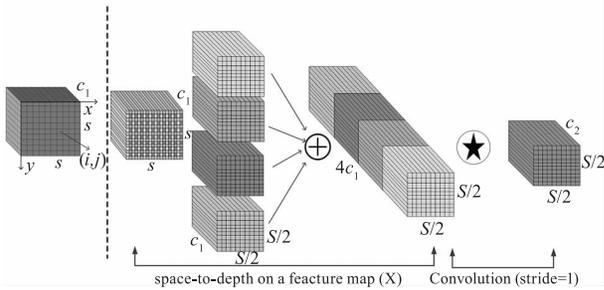


图4 Scale = 2 时的 SPD-Conv 示意图
Fig.4 Schematic of SPD-Conv with scale = 2

3.3 损失函数优化

YOLOv8 采用了 CIOU 损失函数作为边界框的损失函数。然而, 对于低质量图片, 由于目标物体的边缘信息可能不清晰, 可能会导致预测框和真实框之间的位置误差较大, 进而导致预测框与真实框之间的重叠度较小, IOU 值较低, 从而使 CIOU 损失值较大, 在检测低质量图片时效果不佳。针对该问题, 本文引入 WIoU 作为新的边界框损失函数。WIoU 共有 3 个版本 (WIoU-v1、WIoU-v2、WIoU-v3), 经过多次实验, 本文选择 Wiou-v2, Wiou-v1 损失计算公式如式:

$$L_{\text{WIoUv1}} = R_{\text{WIoU}} L_{\text{IoU}} \quad (1)$$

式中, R_{WIoU} 和 L_{IoU} 计算公式:

$$R_{\text{WIoU}} = \exp\left(\frac{(x - xgt)^2 + (y - ygt)^2}{(W_g^2 + H_g^2)^*}\right) \quad (2)$$

$$L_{\text{IoU}} = 1 - IoU \quad (3)$$

其中, x, y 表示预测框的中心点坐标; xgt, ygt 代表真实框的中心点坐标; W_g, H_g , 分别代表由预测框和真实框组成包围框的宽和高; R_{WIoU} 为惩罚项; IoU 为预测框与真实框的交并比; $R_{\text{WIoU}} \in [1, e)$, 作用是将显著放大普通质量锚框的 L_{IoU} ; $L_{\text{IoU}} \in [0, 1]$, 作用是可以显著降低高质量锚框的 R_{WIoU} ,

并在锚框与目标框重合较好的情况下显著降低其对中心点距离的关注。为了防止 R_{WIoU} 产生阻碍收敛的梯度, 将 W_g, H_g 从计算图中分离 (上标 * 表示此操作)。

WIoU-v2 在 Wiou-v1 基础上构造了单调聚焦系数 L_{IoU}^* , 有效降低了简单示例对损失值的贡献。这使得模型能够聚焦于困难示例, 获得分类性能的提升, 其计算公式:

$$L_{\text{WIoUv2}} = L_{\text{IoU}}^* L_{\text{WIoUv1}}, \gamma > 0 \quad (4)$$

但在模型训练过程中, L_{IoU}^* 会随着 L_{IoU} 的减少而减少, 导致后期训练收敛速度较慢。针对此问题, 引入 L_{IoU} 的均值作为归一化因子, 改进后的公式:

$$L_{\text{WIoUv2}} = \left(\frac{L_{\text{IoU}}^*}{\overline{L_{\text{IoU}}}}\right)^\gamma L_{\text{WIoUv1}}, \gamma > 0 \quad (5)$$

其中, $\overline{L_{\text{IoU}}}$ 为动量为 m 的滑动平均值, 使梯度增益 $\left(\frac{L_{\text{IoU}}^*}{\overline{L_{\text{IoU}}}}\right)^\gamma$ 整体保持在较高水平, 解决了训练后期收敛速度慢的问题。动量 m 的计算公式:

$$m = 1 - \sqrt[t]{0.05} \quad (6)$$

其中, t 为 epoch 大小; n 为 batch-size 大小。

3.4 Ghost 轻量化模块

GhostNet^[17] 是一种由华为提出的轻量级神经网络结构, 它通过引入 Ghost 模块可减小模型体积和降低参数量, 同时保持较高的性能。Ghost 模块利用常规卷积对输入特征图进行特征提取并压缩通道。然后, 对压缩通道数后的特征图进行线性变换的廉价操作。最后, 将常规卷积和廉价操作的输出特征图进行拼接在一起, 形成最终的输出。本文将原 YOLOv8n 的 C2f 模块替换为基于 Ghost-Net 的 C3Ghost 模块。这种改进在几乎不影响准确性的情况下拥有更小的参数和体积。Ghost 模块结构如图 5 所示。

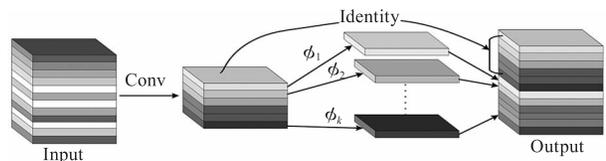


图5 Ghost 模块
Fig.5 Ghost module

图 6 中的 Ghost Bottleneck 主要由两个堆叠的 Ghost Module 组成, 第一个 Ghost Module 负责增加通道数, 起到扩展层的作用; 第二个 Ghost Module 负

责减小通道数,使得通道数与 shortcut 匹配。通过堆叠 Ghost Module 来减少模型的大小和参数量,同时保持较高的精度。

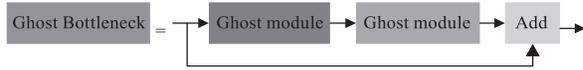


图 6 Ghost Bottleneck 模块

Fig. 6 Ghost Bottleneck module

本文首先使用 Ghost Bottleneck 替换 C3 模块中的 Bottleneck 模块得到基于 GhostNet 的 C3Ghost 模块,其结构如图 7 所示,再将原网络中的 C2f 部分替换为 C3Ghost 模块,显著降低了模型的体积大小,更适用于资源有限的设备。



图 7 C3Ghost 模块

Fig. 7 C3Ghost module

4 实验结果与分析

4.1 实验环境及参数

本实验使用 64 位的 Windows11 操作系统, CPU 为 i9-12900, CUDA12.1, Pytorch 2.1.0, python 3.8, 显卡型号为 NVIDIA GeForce RTX 3080, 在最后 10 轮关闭数据增强。实验参数如表 1 所示。

表 1 实验参数设置

Tab. 1 Experimental parameter settings

参数名称	参数值
输入尺寸	640 × 640
初始学习率	0.01
迭代次数	200
批量大小	16
优化器	SGD
动量	0.937
权重衰减	0.0005

4.2 数据集

本实验数据集采用新版的公开红外数据集 FLIR_v2。包含行人、自行车和汽车等 16 个类别。由于本文聚焦于行人和车辆,因此只保留行人、自行车、汽车三个类别将其他类别删除,并将无效图片删除最后得到 10742 张图像,从如图 8 所示本文将数据集按照 7 : 2 : 1 的比例随机划分,得到训练集、验证集和测试集。数据集样本和边界框大小

如图 8 所示。从目标边界框分布可以看出,数据集包含较多小目标。

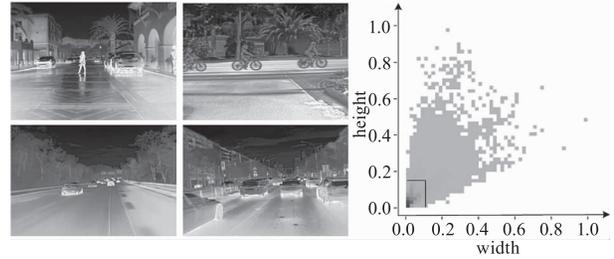


图 8 数据集示例及边界框大小分布

Fig. 8 Example data set and bounding box size distribution

4.3 性能评价指标

本文使用精确率 (Precision, P)、召回率 (Recall, R)、平均精确率 (mAP0.5, mAP0.5 : 0.95)、模型体积大小 (MB)、参数量 (Params) 以及模型每秒检测的帧数 (FPS) 作为模型评价指标。P、R、mAP 越高,表示目标检测越准确; FPS 越高,代表检测速度越快。

其中, mAP0.5 表示 IoU 阈值为 0.5 时所有目标类别的平均检测精度, mAP@0.5 : 0.95 代表在 IoU 阈值为从 0.5 ~ 0.95, 步长为 0.05 时各个 mAP 的平均值。精确率、召回率、平均精度以及 FPS 的计算式如下所示:

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$AP = \int_0^1 P(R) dR \quad (9)$$

$$mAP = \frac{1}{k} \sum_{i=0}^k AP_i \quad (10)$$

$$FPS = \frac{N(p)}{T(p)} \quad (11)$$

其中, TP 为将正类正确地预测为正类的数量; FP 为将负类错误地预测为正类的数量; FN 为将正类错误地预测为负类的数量; K 代表类别数; N(p) 表示处理图像的总数; T(p) 表示处理图像的总时间。

4.4 损失函数对比实验

为了更好的验证本文所用损失函数 WIoU-v2 的优势,设计本次损失函数对比实验,以 YOLOv8-CIoU 为基线,在保证相关配置不变的情况下,将 WIoU-v3 损失函数分别与原 CIoU^[18]、和目前主流损失函数 SIoU^[19]、EIoU^[20]、以及 WIoU-v1 和 WIoU-v2 进行对比,实验结果如表 2 所示。

表 2 不同损失函数在 YOLOv8n 下的对比
Tab.2 Comparison of different loss functions under YOLOv8n

模型	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%
YOLOv8n-CIoU	82.9	67.2	76	45.9
YOLOv8n-SIoU	84.3	66	75.7	45.7
YOLOv8n-EIoU	83	65.9	75.4	46.1
YOLOv8n-WIoU-V1	83.9	66.1	76	46
YOLOv8n-WIoU-V2	83.9	67.1	76.7	46.1
YOLOv8n-WIoU-V3	82.9	67.4	76.5	45.9

综合各项评价指标,可以发现 WIoU-V2 在各项指标中都表现较好,特别是在 mAP@0.5 上表现最好

达到 76.7%,mAP@0.5:0.95 方面也达到了 46.1%,说明其在更高置信度要求下仍然能够保持较好的性能,同时,精确率和召回率也保持在较高水平,与其他模型相比毫不逊色。因此,综合考虑,本文选择 WIoU-V2 损失函数代替原 CIoU 损失函数。

4.5 不同改进模块的消融实验

为了精确评估上述各个改进组件对模型性能的具体影响,以 YOLOv8n 作为基线进行了 8 组消融试验,每组实验均严格控制实验环境配置和超参数的一致性,以确保结果的公正性。结果如表 3 所示。(表中 p2、S、W、G 分别代表 P2 特大特征图、Spd-Conv、WIoU-v2、Ghost 模块,“√”代表在该组实验中使用该模块)。

表 3 消融实验

Tab.3 Ablation experiment

模型	P2	S	W	G	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	参数量	体积/MB	FPS
YOLOv8n					82.9	67.2	76	45.9	3006233	6.3	175.4
A	√				84.3	71.9	81.6	50.3	2921436	6.3	161.3
B		√			83.8	68.7	77.9	47.6	3267353	6.8	153.8
C			√		83.9	67.1	76.7	46.1	3006233	6.3	169.5
D				√	82.3	66.6	75.5	45.5	2293225	4.9	175.4
E	√	√			85.4	72.2	82.4	51.4	3238972	6.6	156.3
F	√	√	√		84.5	73.2	83	51.4	3238972	6.9	151.5
Ours	√	√	√	√	84.5	73.5	83.2	51.6	2525964	5.3	147.1

由表 3 可知,在 A 组实验中,引入特大特征图 P2 后,P、召回率 R、以及 mAP@0.5 和 mAP@0.5:0.95 等关键指标均获得了显著的提升,其中 mAP@0.5 提升了 5.6%,这一改进还减少了参数量,尽管推理速度略有下降,但大幅度提高了检测精度,有效减少了漏检;在第 B 组实验中,添加了 Spd-Conv 后,在引入少量参数和额外体积前提下,模型的 P、R、mAP@0.5 和 mAP@0.5:0.95 指标均获得了不同程度的提升,其中 mAP@0.5 提升了 1.9%。这证明了 Spd-conv 可以有效提高检测精度;在第 C 组实验中,引入了 WIoU-v2 损失函数。结果显示,P 提升了 1%,mAP@0.5 上涨了 0.7%,而参数量等其他指标几乎保持不变。证明了 WIoU-v2 损失函数的有效性;在 D 组实验中,引入 Ghost 模块,mAP@0.5 略有下降 0.5%,但参数量减少了 23.7%,体积下降了 22.2%。这表明在牺牲一定精度的情况下,Ghost 模块可以有效减少模型的大小和参数量;在

E 组实验中,结合了 P2 和 Spd-Conv 两个改进点。结果显示,mAP@0.5 提升了 6.4%,尽管参数量和体积有所增加,但整体性能提升显著;在 F 组实验中,在 E 组的基础上进一步融入了 WIoU-v2 损失函数。实验结果表明,mAP@0.5 进一步提升了 7%,显示出多种改进方法联合使用的优势;最后,将四个改进点全部融合到 YOLOv8n 模型中。结果显示,mAP@0.5 值达到了 83.2%,相比原算法提升了 7.2%,参数量和体积分别减少了 16%和 15.8%,虽然 FPS 略减到 147.1,但考虑到检测精度的大幅提升以及模型复杂度的显著降低,这一微小的 FPS 下降是可以接受的。综上所述,本文通过消融试验验证了各项改进设计的有效性,并展示了它们在提升检测性能、降低模型复杂度方面的积极作用。

4.6 与其他主流算法对比实验

为进一步凸显本文所提算法的优越性,将本文算法与目前主流目标检测算法进行对比实验。实验

结果如表 4 所示,本文所提的 YOLOv8-PSWG 算法在多个关键指标上均表现出色。其精确率高达 84.5%,仅略低于表现最佳的 YOLOV8s 算法 1.7%,而明显高于其他参与对比的算法。在召回率方面,该算法达到了 73.5%,同样超越了其他所有算法。更值得一提的是,在平均精度这一综合评价指标上,无论是 $mAP@0.5$ 还是更为严格的 $mAP@0.5:0.95$,均显著领先于其他算法。此外,在模型大小和复杂度方面,该算法也展现出了极大的优势。其参数量仅为 2525964,模型体积仅为 5.3 MB,

相较于其他主流算法如 yoloV5s,参数量下降了 64%,体积减少了 63.2%。尽管 FPS 略低于部分算法,但其高达 147.1,仍远超实时性检测的基本要求($FPS \geq 30$),充分证明了其在保证高检测精度的同时,也具备出色的实时性能。

综上所述,本文所提出的改进算法 PSWG-YOLO 在精度、速度和模型大小之间实现了良好的平衡,不仅在提高检测精度方面取得了显著成效,而且还在保持检测速度和优化模型大小方面展现出了卓越的性能。

表 4 与主流算法比较

Tab.4 Comparison with mainstream algorithms

算法	P/%	R/%	$mAP@0.5$ /%	$mAP@0.5:0.95$ /%	参数量	体积/MB	帧率/fps
YOLOV5S ^[21]	83.9	69.4	77.9	44.7	7018216	14.4	156.2
YOLOV7-tiny ^[22]	83.9	72.4	80.7	45.4	6083216	12.3	135.1
YOLOV8s	86.2	72.6	81.6	51.3	11126745	22.5	151.5
YOLOV8n	82.9	67.2	76	45.9	3006233	6.3	175.4
PSWG-YOLO	84.5	73.5	83.2	51.6	2525964	5.3	147.1

4.7 检测效果和精度对比

本文在 YOLOv8n 的基础上,提出了 PSWG-YOLO 改进模型,对比原模型,在检测效果上有了明显的提升,其检测效果和精度对比分别为如图 9(a)和(b)所示。

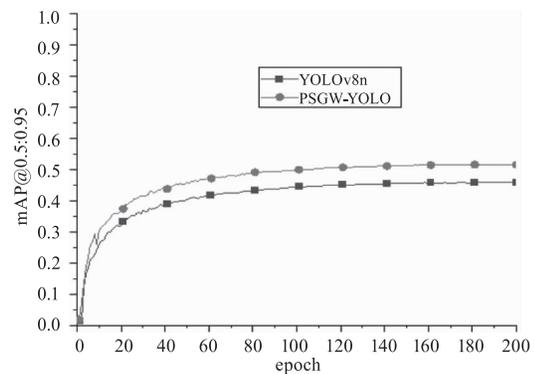
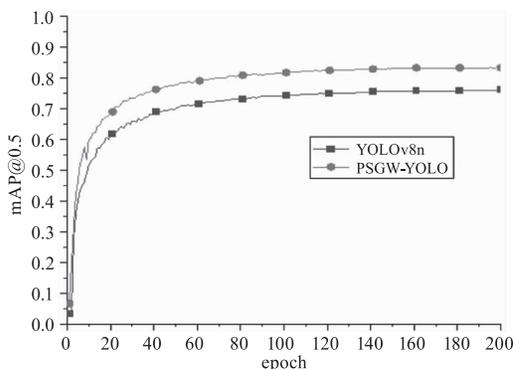
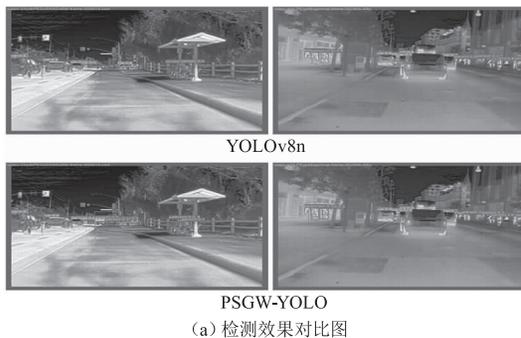


图 9 检测效果和精度对比图

Fig.9 Comparison of detection effect and accuracy

5 结论

本文提出了一种基于 YOLOv8n 的红外行人车辆检测算法 PSWG-YOLO,通过增加极大特征图和采用 SPD-Conv 模块提升了对小目标及低分辨率图像的检测能力。采用动态非单调聚焦的 WIoU 损失函数和 Ghost 模块进一步提高了检测精度并在一定程度上降低了模型复杂度。实验结果显示,在 FLIR-V2 数据集上,该算法相较于原 YOLOv8n 精确率、召回率和 $mAP@0.5$ 分别提升了 1.6%、6.3% 和 7.2%,参数量减少 16%,模型大小减少 15.8%,检测速度达到 147.1FPS。与主流算法相比,该算法在 mAP

等多方面表现出优越性。综合来看,本算法在提高红外行人车辆检测精度的同时,也实现了对模型一定的轻量化,适用于资源受限的平台,并满足实时性检测的需求,可为辅助驾驶中的红外行人车辆检测提供有效的技术支持。

参考文献:

- [1] Zhong Mingxia, Jiang Bojun. Application of infrared thermal imaging technology in ADAS system[J]. *Laser & Infrared*, 2022, 52(5): 721 – 725. (in Chinese)
钟明霞, 姜柏军. 红外热成像技术在 ADAS 系统中的应用[J]. *激光与红外*, 2022, 52(5): 721 – 725.
- [2] Papageorgiou C P, Oren M, Poggio T. A general framework for object detection [C]//Sixth International Conference on Computer Vision. IEEE, 1998: 555 – 562.
- [3] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, 2005, 1: 886 – 893.
- [4] Hearst M A, Dumais S T, Osuna E, et al. Support vector machines [J]. *IEEE Intelligent Systems and their Applications*, 1998, 13(4): 18 – 28.
- [5] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting [J]. *Journal of Computer and System Sciences*, 1997, 55(1): 119 – 139.
- [6] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779 – 788.
- [7] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263 – 7271.
- [8] REDMON J, FARHADI A. Yolov3: an incremental improvement [J]. *arXiv Preprint*, 2018, arXiv:180402767.
- [9] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection [J]. *arXiv Preprint*, 2020, arXiv:2004. 10934.
- [10] Liu W, Anguelov D, Erhan D, et al. Ssd: single shot multi-box detector [C]//Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11 – 14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21 – 37.
- [11] Li Z, Yang L, Zhou F. FSSD: feature fusion single shot multi-box detector [J]. *arXiv Preprint*, 2017, arXiv:171200960.
- [12] Hu Peishuai, Wang Weijiang. Research on pedestrian-vehicle detection technology in infrared scene in ADAS [J]. *Agricultural Equipment and Vehicle Engineering*, 2023, 61(11): 131 – 135. (in Chinese)
胡培帅, 王维强. ADAS 中红外场景下的行人车辆检测技术研究 [J]. *农业装备与车辆工程*, 2023, 61(11): 131 – 135.
- [13] Guo Zhijian, Li Jianguo, Qi Haijun et al. Infrared pedestrian and vehicle Detection algorithm for based on improved YOLOv4 [J]. *Laser & Infrared*, 2023, 53(4): 607 – 614. (in Chinese)
郭志坚, 李江勇, 祁海军, 等. 基于改进 YOLOv4 的红外行人车辆检测算法 [J]. *激光与红外*, 2023, 53(4): 607 – 614.
- [14] Luo X, Zhu H, Zhang Z. IR-YOLO: real-time infrared vehicle and pedestrian detection [J]. *Computers, Materials & Continua*, 2024, 78(2): 2667.
- [15] Sunkara R, Luo T. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects [C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer Nature Switzerland, 2022: 443 – 459.
- [16] Tong Z, Chen Y, Xu Z, et al. Wise-IOU: bounding box regression loss with dynamic focusing mechanism [J]. *arXiv Preprint*, 2023, arXiv:2301. 10051.
- [17] Han K, Wang Y, Tian Q, et al. Ghostnet: more features from cheap operations [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580 – 1589.
- [18] Zheng Z, Wang P, Ren D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation [J]. *IEEE Transactions on Cybernetics*, 2021, 52(8): 8574 – 8586.
- [19] Du S, Zhang B, Zhang P, et al. An improved bounding box regression loss function based on CIOU loss for multi-scale object detection [C]//2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML). IEEE, 2021: 92 – 98.
- [20] Zhang Y F, Ren W, Zhang Z, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. *Neurocomputing*, 2022, 506: 146 – 157.
- [21] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464 – 7475.
- [22] Zhu X, Lyu S, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 2778 – 2788.