

基于改进 YOLOv5 算法的行人与非机动车实时识别方法

崔晋豪¹ 许博阳²

1 郑州工业应用技术学院

2 河南省智能交通视频图像感知与识别工程技术研究中心

DOI:10.12238/acair.v3i3.15603

[摘要] 随着智慧交通与自动驾驶技术的发展,对行人与非机动车的实时精准感知需求日益迫切。针对 YOLO 等算法在复杂场景下存在小目标漏检、遮挡目标误检等问题,本文提出一种改进的 YOLOv5 实时识别算法。通过引入注意力机制强化特征提取、优化特征融合网络结构以及改进损失函数,显著提升了模型性能。实验结果表明,该算法在自制与公开数据集上均取得了更高的平均精度(mAP),同时保持了较高的检测速度(FPS),有效满足了实际应用中对精度与实时性的双重需求,为智能交通系统提供了有效的解决方案。

[关键词] 目标检测; YOLOv5; 实时识别; 智慧交通

中图分类号: U412.37+9 **文献标识码:** A

Real-Time Recognition Method for Pedestrians and Non-Motor Vehicles Based on Improved YOLOv5 Algorithm

Jinhao Cui¹ Boyang Xu²

1 Zhengzhou University of Industrial Technology

2 Henan Province Intelligent Transportation Video Image Perception and Recognition Engineering Technology Research Center

[Abstract] With the advancement of intelligent transportation and autonomous driving technologies, there is an increasing demand for real-time and accurate perception of pedestrians and non-motor vehicles. Addressing issues such as missed detection of small targets and false detection of occluded objects in complex scenarios with algorithms like YOLO, this paper proposes an improved YOLOv5 real-time recognition algorithm. By incorporating an attention mechanism to enhance feature extraction, optimizing the feature fusion network structure, and improving the loss function, the model's performance is significantly enhanced. Experimental results demonstrate that the proposed algorithm achieves higher mean average precision (mAP) on both self-collected and public datasets while maintaining high detection speed (FPS). It effectively meets the dual requirements of accuracy and real-time performance in practical applications, providing an efficient solution for intelligent transportation systems.

[Key words] Object detection; YOLOv5; Real-time recognition; Intelligent transportation

前言

本文旨在对 YOLOv5 模型进行针对性改进。主要贡献包括:引入 CBAM 注意力机制以增强特征提取能力并抑制背景噪声;设计双向特征金字塔网络(BiFPN)加强多尺度特征融合,提升小目标检测性能;采用 Focal-EIOU 损失函数替换原损失,以加快收敛和提高定位精度。通过在自建与公开数据集上的系统实验,验证了改进模型在保持高帧率的同时显著提升检测精度。

1 相关工作与理论基础

1.1 目标检测算法概述

目标检测是计算机视觉的核心任务,旨在定位图像中特定目标的位置并识别其类别。主流算法可分为两阶段与单阶段两类,其核心思想与性能特点有显著差异(如表1所示)。两阶段检测器以 Faster R-CNN 为代表,首先生成候选区域(Region Proposals),再对每个候选区域进行分类与边界框回归。该类方法结构复杂、计算量大,导致检测速度较低,难以满足实时场景需求,但其设计机制通常带来较高的检测精度。相比之下,单阶段算法如 YOLO 系列和 SSD 摒弃了区域提议步骤,直接在特征图上以密集预测的方式同时完成分类与定位,极大提升了推理速度,更适合实时系

统, 尽管早期版本在精度上略有逊色, 但后续版本通过结构改进不断缩小了差距。

表1 两阶段与单阶段检测算法对比

特性	两阶段算法 (如 Faster R-CNN)	单阶段算法 (如 YOLO, SSD)
检测流程	先产生候选区域, 再分类回归	直接在特征图上预测目标
速度	慢	快
精度	高	较高 (不断追赶)
代表模型	R-CNN 系列	YOLO 系列, SSD

1.2 YOLOv5算法原理

YOLOv5是YOLO系列的重要代表, 以其优异的速度-精度平衡和工程友好性被广泛采用。其网络结构主要包含Backbone、Neck和Head三个部分。Backbone采用CSPDarknet, 通过Focus模块对输入图像进行切片操作, 减少计算量的同时保留丰富特征信息, 并结合CSP (Cross Stage Partial) 结构增强梯度流, 提升网络的学习能力与效率。Neck部分使用PANet (Path Aggregation Network), 融合了FPN的自顶向下语义传递和PAN的自底向上定位信息传递, 实现多层次的特征融合, 有效聚合不同尺度的特征, 提升对不同大小目标的检测能力。Head则负责在融合后的特征图上进行最终预测, 输出目标类别、置信度及边界框坐标。

1.3 注意力机制

注意力机制源于人类视觉系统, 其核心思想是让神经网络能够自主地关注输入中更重要的信息, 抑制无关部分, 从而高效利用计算资源并提升模型表达能力。在卷积神经网络中, 注意力通常以即插即用的模块形式被嵌入到不同深度, 以增强特征的表征质量。卷积块注意力模块 (CBAM) 是一种轻量且有效的注意力模块, 它依次从通道和空间两个维度推导注意力图。通道注意力模块首先对输入特征层分别进行全局平均池化和最大池化, 之后通过共享参数的多层感知机 (MLP) 处理, 将两路结果相加后经Sigmoid激活函数生成通道权重向量, 其简化表达为:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)))$$

空间注意力则侧重于特征图中哪些位置具有重要性。它首先在通道维度上对同一位置进行平均池化和最大池化并完成拼接, 随后利用一个7×7卷积层和Sigmoid函数生成空间权重矩阵, 其计算可简化为:

$$M_s(F) = \sigma(f_{7 \times 7}([AvgPool(F); MaxPool(F)]))$$

最终, 输入特征会依次与通道和空间注意力图相乘, 实现自适应特征优化。将CBAM嵌入主干网络可显著提升模型对行人和非机动车等关键目标的特征提取能力。

2 改进的YOLOv5算法设计

2.1 整体改进框架

针对复杂道路场景下行人与非机动车检测所面临的挑战, 本文对YOLOv5s模型进行了多方位改进。改进后的整体网络结构如表2所示, 其核心Backbone、Neck和Head框架得以保留, 以保证算法的主体优势。主要的创新点体现在三个关键模块的引入与替换: 一是在Backbone网络中每个CSP模块后嵌入了CBAM注意力机制模块, 以增强模型对关键特征的提取能力; 二是将Neck部分的特征融合网络由原PANet替换为更高效的加权双向特征金字塔网络 (BiFPN), 以提升多尺度特别是小目标的融合效果; 三是将Bounding Box回归的损失函数由CIoU Loss替换为Focal-EIoU Loss, 以优化训练过程并提高定位精度。

表2 改进的YOLOv5网络结构

模块	改进点	作用	适用场景
Backbone	CSP 模块+CBAM	增强关键特征提取, 抑制背景干扰	复杂道路场景
Neck	PANet→BiFPN	优化多尺度特征融合, 提升小目标检测	行人、非机动车等小目标
Head	CIoU→Focal-EIoU	改善定位精度, 缓解样本不平衡	密集目标检测

2.2 引入CBAM注意力机制

原YOLOv5模型对特征图上的所有信息进行平等处理, 在背景复杂、目标密集的交通场景中, 容易受到大量无关信息的干扰, 导致对行人和非机动车的特征学习不足。为解决此问题, 可引入卷积块注意力模块 (CBAM), 其动机是让网络学会“关注”更重要的特征通道和空间位置, 自适应地强调行人及非机动车的关键特征并抑制背景噪声。

具体融入方式是: 在Backbone网络的每个CSP_1、CSP_2、CSP_3模块之后, 即完成初步特征提取的位置, 依次嵌入CBAM模块。该模块会接收上一层的输出特征图, 并依次通过通道注意力模块和空间注意力模块进行计算, 预期效果是使模型在面对部分遮挡、光照变化及复杂背景时, 具备更强的特征分辨能力和鲁棒性。

2.3 双向特征金字塔网络 (BiFPN) 的融合

原模型采用的PANet实现了良好的多尺度特征融合, 但其在融合不同尺度的输入特征时, 并未区分它们的重要性, 默认所有输入对输出的贡献是相等的。这可能导致某些对当前预测更重要的特征 (如更精细的小目标特征) 在融合过程中被稀释。

BiFPN通过引入可学习的权重来解决这一问题, 其核心是加权双向跨尺度连接机制。该方式允许网络为不同输入特征分配不同的重要性权重。

用BiFPN结构全面替换原Neck中的PANet。这种替换实现了从“无差别融合”到“加权融合”的转变, 在保证高运行效率的基础上, 显著提升了对小尺度行人与非机动车的检测效果。

2.4 Focal-EIoU Loss损失函数的优化

CIoU Loss是YOLOv5原生的定位损失函数,虽然综合考虑了重叠面积、中心点距离和长宽比,但其长宽比项 v 的定义有时会导致训练后期收敛速度变慢甚至回归不准。此外,目标检测任务中简单样本数量远多于困难样本,CIoU Loss会赋予简单样本过大的损失贡献,从而淹没困难样本的梯度,不利于模型优化。

为解决这些问题,可以采用Focal-EIoU Loss作为新的定位损失函数。该损失函数首先基于EIou Loss,将其CIoU Loss中的长宽比项拆解为对目标宽和高分别进行回归,形式化定义为:

$$\begin{aligned} LEIoU &= LIoU + Ldis + Lasp \\ &= 1 - IoU + \frac{\rho^2(\mathbf{b}, \mathbf{b}_{gt})}{(c_w)^2 + (c_h)^2} + \frac{\rho^2(w, w_{gt})}{(c_w)^2} + \frac{\rho^2(h, h_{gt})}{(c_h)^2} \end{aligned}$$

此定义使得收敛过程更加平滑稳定,定位精度更高。

3 实验与结果分析

3.1 实验环境与数据集

实验在Ubuntu 20.04操作系统上进行,使用Python 3.8和PyTorch 1.10.0深度学习框架,硬件平台为NVIDIA GeForce RTX 3090 GPU(24GB显存)并搭配CUDA 11.3加速环境。研究采用MSCOCO数据集进行模型预训练,并使用一个大规模自制数据集进行主要训练与测试。该自制数据集采集自多个城市道路的监控画面,包含10,000张高质量图像,共计标注了超过35,000个实例,目标类别涵盖行人、自行车和电动车。所有图像均使用LabelImg工具进行精细标注,并经过了严格的清洗和校验流程以确保数据质量。训练过程中采用了Mosaic、随机翻转、色彩抖动等数据增强策略,有效提升了模型的泛化能力和鲁棒性。

3.2 评价指标

本研究采用多项权威指标对模型性能进行全面评估。精度方面,使用平均精度均值mAP进行评估,其中mAP@0.5表示交并阈值为0.5时的结果,mAP@0.5:0.95则表示在不同IoU阈值下的平均精度。召回率用于衡量模型检测出真实正例的能力。速度性能通过FPS进行量化,即模型每秒能够处理的图像帧数。

3.3 消融实验

为验证各改进模块的有效性,我们设计了系统的消融实验。实验结果表明,基线模型YOLOv5s的mAP@0.5为86.2%。单独添加CBAM模块后,mAP提升至87.8%,证明注意力机制有效增强了特征选择性。引入BiFPN带来了最显著的性能提升,mAP达到88.5%,凸显了加权特征融合的重要性。采用Focal-EIoU损失函数也使mAP提高了0.8%。当全部改进模块集成后,完整模型达到了90.5%的mAP,相比基线提升4.3%,且FPS仍保持在115的高水平,完全满足实时检测需求。

3.4 对比实验

将本文提出的改进模型与当前主流目标检测算法进行对比实验,对比实验表明,传统两阶段算法Faster R-CNN精度较高但速度缓慢,仅达到15 FPS。单阶段算法中,SSD300和YOLOv3-tiny在速度上有优势但检测精度不足。YOLOv4取得了88.9%的mAP,

但计算成本较高。本文改进的YOLOv5s模型在mAP@0.5指标上达到90.5%,显著优于所有对比模型,同时保持了115 FPS的实时处理速度,在参数量仅轻微增加的情况下,实现了精度与速度的最佳平衡。

3.5 可视化结果分析

通过可视化对比分析,直观展示了改进模型的有效性。在复杂场景测试图像中,原始YOLOv5模型存在明显的小目标漏检和遮挡目标识别困难问题。改进后的模型不仅能够准确检测出远处的小尺度行人和非机动车,对部分遮挡目标也表现出更强的识别能力。同时,改进模型显著降低了将背景干扰误检为目标的情况,检测结果更加可靠。

4 结论与展望

4.1 研究工作总结

本研究针对复杂道路场景下行人与非机动车检测的实际需求,对YOLOv5模型进行了系统性改进。通过引入CBAM注意力机制、BiFPN特征融合网络以及Focal-EIoU损失函数,有效提升了模型的检测性能。实验结果表明,改进后的模型在自制数据集上的mAP@0.5达到90.5%,较原始YOLOv5s基线模型显著提升了4.3%,同时保持了115FPS的高处理速度,实现了精度与速度的良好平衡,验证了所提改进策略的有效性。

4.2 主要创新点

本文的主要创新点体现在三个方面:首先,将CBAM注意力机制与YOLOv5模型相结合,增强了模型对关键特征的聚焦能力,有效抑制了复杂背景的干扰。其次,采用BiFPN结构替代原特征融合网络,通过加权双向融合机制显著提升了对小尺度目标的检测性能。第三,引入Focal-EIoU损失函数优化边界框回归过程,不仅加快了模型收敛速度,同时提高了定位精度。

4.3 未来工作展望

未来的研究工作将从以下几个方向展开:一是探索模型剪枝、量化等轻量化技术,进一步提升模型效率,以适应车载终端等嵌入式设备的部署需求。二是将现有的2D检测扩展至3D空间感知和多目标跟踪,获取更丰富的环境信息。三是深入研究模型在极端天气和低光照等恶劣条件下的鲁棒性提升方案。最后,将探索领域自适应方法,使模型具备快速适应新环境的能力。

[项目支持]

基于深度学习的行人与非机动车识别算法(ETRC-240204)。

[参考文献]

- [1]李洋.基于改进YOLOv5s算法的人流量识别技术[J].电子设计工程,2025,33(14):51-56.
- [2]温福新,许钢,凌成.基于改进YOLOv5s的红外图像行人检测算法[J].贵州大学学报(自然科学版),2025,42(3):65-73+80.
- [3]周龙刚,魏本昌,魏鸿奥,等.改进YOLOv5的密集行人检测算法研究[J].软件导刊,2024,23(12):249-254.

作者简介:

崔晋豪(2005--),女,汉族,河南安阳人,就读于郑州工业应用技术学院,研究专业:人工智能专业。