



引文格式:黄丽,龚志鹏,刘凡凡,等.图像跨尺度特征融合与数据增强的公交客流检测模型[J].武汉大学学报(信息科学版),
2024,49(5):700-708.DOI:10.13203/j.whugis20220690

Citation: HUANG Li, GONG Zhipeng, LIU Fanfan, et al. Bus Passenger Flow Detection Model Based on Image Cross-Scale
Feature Fusion and Data Augmentation[J]. Geomatics and Information Science of Wuhan University, 2024, 49(5): 700-708. DOI:
10.13203/j.whugis20220690

图像跨尺度特征融合与数据增强的公交客流 检测模型

黄丽¹ 龚志鹏¹ 刘凡凡¹ 程起敏²

1 中南民族大学电子信息工程学院,湖北 武汉,430074

2 华中科技大学电子信息与通信学院,湖北 武汉,430074

摘要:针对目前公交客流目标检测中由于前后排乘客图像尺寸差距大、遮挡严重而难以检测的问题,提出一种改进的目标检测模型YOLOv5s_P。该模型基于YOLOv5基本架构,针对乘客图像尺寸差距大的问题,使用重复加权双向特征金字塔网络结构替换YOLOv5中的路径聚合网络结构,引入权重机制双向跨尺度融合不同特征,增强提取复杂目标特征的能力;同时针对遮挡问题,结合Mixup数据增强方法,扩充遮挡和重叠图像的训练样本,提高模型泛化能力,减少客流图像残缺导致的漏检。为了验证YOLOv5s_P模型性能,将其应用于实际公交场景,并与Faster R-CNN、SSD300、RetinaNet、YOLOv5s 4种模型进行对比测试。实验结果表明,YOLOv5s_P模型在不影响检测速度的情况下,平均精度均值达到96.9%,平均漏检率较YOLOv5s模型降低了3.43%,提高了公交客流的检测精度。

关键词:深度学习;YOLOv5s_P;公交客流;目标检测;特征融合;数据增强

中图分类号:P237;TP39

文献标识码:A

收稿日期:2022-10-24

DOI:10.13203/j.whugis20220690

文章编号:1671-8860(2024)05-0700-09

Bus Passenger Flow Detection Model Based on Image Cross-Scale Feature Fusion and Data Augmentation

HUANG Li¹ GONG Zhipeng¹ LIU Fanfan¹ CHENG Qimin²

1 College of Electronics and Information, South-Central Minzu University, Wuhan 430074, China

2 School of Electronic Information and Communication, Huazhong University of Science and Technology, Wuhan 430074, China

Abstract: Objectives: In order to resolve the issues of large size gap, serious occlusion and overlapping images in the objection detection research of bus passenger flow, an improved object detection model YOLOv5s_P is proposed. **Methods:** The PANet structure in YOLOv5 model is replaced with the BiFPN structure to strengthen features information of different scales maps in order to extract complex target features. At the same time, Mixup data augmentation method is used to increase the training samples of occlusion and overlapping images to improve model generalization capabilities and to reduce the detection errors caused by the fragmentation of passenger flow images. In order to verify the performance of the YOLOv5s_P model, it is compared with four other models: Faster R-CNN, SSD300, RetinaNet and YOLOv5s to detect bus passenger flow in real bus scenario, in which the image sets are labeled to detect the upper body of the human instead of the head. **Results:** Experimental results show that the average accuracy of YOLOv5s_P model reached 96.9% without affecting the detection speed, and the average missed detection rate was reduced by 3.43% compared with YOLOv5s model, which improved the detection accuracy of bus passenger flow. **Conclusions:** In the future, research will be integrated with the attention mechanism to further improve the detection accuracy, and combined with the tracking algorithm to

基金项目:国家自然科学基金(42271352);国家民委高等教育教学改革研究项目(23093)。

第一作者:黄丽,博士,讲师,主要从事通信信息处理等方面的研究。szemmah@hotmail.com

solve the problem of the bus passengers number fluctuation.

Key words: YOLOv5s_P; bus passenger flow; object detection; feature fusion; data augmentation

公交客流是智慧公交网络设计与优化的重要依据^[1],及时准确获得乘客人数信息、分析客流量,不仅可以合理调度车辆,而且有助于解决交通拥堵,提高经济效益,实现城市智慧交通。随着信息技术发展,智能检测技术替代了人工调查方法^[2],如读取公交 IC 卡数据^[3]和获取手机支付数据^[4]是最普遍的公交客流数据采集技术。但由于乘客下车不刷卡和扫码,也无法获取投币人员数量,因此上述方法不能准确统计乘客数量。此外,红外^[5]等传感器计数方法在客流拥挤时数据采集准确率不高,且传感器易受环境温度和光线的影响。由此可见,传统的采集方法无法快速准确地获取客流信息。

目标检测方法利用计算机视觉技术分析处理视频信号^[6],可识别跟踪监控场景变化,广泛应用于人脸检测^[7]、行人检测^[8]、自动驾驶^[9]、监控安防^[10]等领域。为此,可利用部署在公交车内的监控设备采集图像,结合目标检测技术实现客流自动识别与实时统计。早期的目标检测模型需人工提取目标特征,检测准确率低、速度慢。随着深度学习算法的发展和设备算力的提升,将深度学习引入目标检测技术中,可自动提取图像特征,提高模型自适应能力,在小目标^[11]、低光照^[12]等复杂场景下均有广泛应用^[13]。

与商场、客运站、街道等开阔场景不同,公交车内环境具有环境狭窄、相互遮挡、后排乘客图像较小等特点。由于摄像头安装角度固定,乘客图像遮挡严重,比如车内设施遮挡乘客、乘客之间相互遮挡等,导致部分图像重叠残缺,同时后排乘客图像较小,如图 1 中红色圆圈的乘客图像(图像来源于 <https://www.hunantoday.cn/article/202109/202109082043305798.html>)。公交车内乘客图像残缺和尺寸差距大,增加了目标检测的难度,对模型性能要求较高。目前公交客流检测研究中,采用目标跟踪算法在乘客拥挤时易丢失目标,两阶段检测模型速度较慢,单阶段检测模型易漏检遮挡乘客的图像,使得现有客流检测方法的实时性和精度尚不能完全满足应用需求。

为解决乘客图像残缺和尺寸差距大导致检测精度不高的问题,本文提出一种改进的目标检测模型 YOLOv5s_P,基本思路如下:针对乘客图像尺寸差距大以及后排乘客图像较小的问题,采

用重复加权双向特征金字塔网络(bidirectional feature pyramid network, BiFPN)^[14]替换 YOLOv5s 模型中的路径聚合网络(path aggregation network, PANet)^[15],通过高效的双向跨尺度连接和加权特征融合,增强不同尺度图的特征信息,提高复杂目标的检测精度。针对乘客图像遮挡重叠严重问题,结合 Mixup 数据增强方法^[16],扩充重叠图像训练样本,提高模型学习能力,降低遮挡图像的漏检率。为验证 YOLOv5s_P 模型的性能,首先构建公交真实场景数据集 BusUpperBody 共 1 000 张,数据集来源于公交车实际运营视频,包含了全天不同时段和光照条件下的客流高峰和低谷图像;然后在数据集上测试了本文模型与 Faster R-CNN(faster region with convolutional neural networks)模型^[17]、SSD300(single shot multibox detector)模型^[18]、RetinaNet 模型^[19]和 YOLOv5s 模型^[20]的性能,对比和消融实验结果表明,YOLOv5s_P 模型在不影响检测速度的情况下,能够有效提高客流检测的准确性,显著提升对遮挡目标的检测效果。



图 1 公交车内客流图

Fig. 1 Passenger Flow Diagram Inside the Bus

1 相关工作

基于深度学习的目标检测模型主要分为单阶段和两阶段两大类^[21],单阶段模型包括 SSD 系列^[18]、YOLO 系列^[22]等,两阶段模型包括 R-CNN^[23]、Faster R-CNN^[17]等。单阶段模型能一次性完成目标分类和定位两个任务,检测速度较快,但检测精度略有损失。两阶段模型首先根据图像生成候选框,并提取候选检测框的深度特征,再进行分类和回归,检测精度较高,但检测速度较慢。通用目标检测模型研究虽不断改进,但

应用在公交车客流检测场景中,其漏检率仍有待提高。

现有自然场景下人脸检测研究中, Li等^[8]针对不同尺度的行人检测问题,开发了规模感知快速框架,采用不同权重将大尺寸子网络和小尺寸子网络融合到统一架构中,解决行人尺度差异问题。李祥兵等^[24]提出改进的Faster R-CNN模型,采用多尺度特征融合、增加锚框数、改变非极大值抑制算法等方法,提高自然场景下小尺度人脸的检测精度。为解决小尺寸行人图像的误检问题, Hsu等^[25]提出比率尺度感知方法,提高了对目标很小、长宽比差异大的图像检测性能。针对小目标误检问题,郭磊等^[26]使用Mosaic-8方法进行数据增强,修改目标框回归公式,解决训练过程中梯度消失的问题,提升了防护面具佩戴的检测精度。上述自然场景下行人检测研究中, R-CNN系列模型的检测速度较慢, YOLO系列模型易漏检遮挡严重的目标。

目前基于深度学习的公交客流检测研究中, 范伟^[27]利用改进的ResNet101和RFCN(region based fully convolutional networks)网络结构获取头部坐标信息采集人数,检测速度较慢,并且无法识别被遮挡的头部图像。张开生等^[28]在车厢内前中后3个位置部署摄像头,搭建卷积神经网络提取图像特征,可判断车内拥挤情况,但难以准确检测乘客数量。Liu等^[29]通过监测车门处的视频图像,采用YOLOv2模型与MIL跟踪器结合方法,计数上下车的乘客数量。类似研究还有You等^[30]采用Faster R-CNN模型识别乘客头部区域特征,用卡尔曼滤波器跟踪目标运动轨迹,统计上下车客流。上述公交客流检测研究中,检测车门处图像计数的方法在乘客拥挤时易丢失目标,统计误差较大;多通道图像检测算法复杂度高、计数易重复,且人头检测方法当头部图像残缺时易漏检。

针对现有模型中被遮挡乘客图像残缺导致漏检率高的问题,本文提出YOLOv5s_P模型,通过检测人体上半身来替代人头检测,同时改进网络结构,引入数据增强方法,以降低后排和遮挡乘客的漏检率,提高客流检测精度。

2 改进模型YOLOv5s_P

YOLOv5s_P模型针对公交车内场景的图像特点,从跨尺度特征融合和数据增强两方面来提高复杂目标的检测性能。首先,为解决前后排乘

客图像尺寸差距大的技术难点, YOLOv5s_P模型借助加权的双向特征金字塔网络,根据不同尺度的特征在融合时贡献的大小,自适应调整特征融合分支的权重,能够在不增加消耗的情况下融合更多特征,从而加强对不同尺寸乘客图像的特征提取能力。其次,为增强公交车内遮挡乘客图像的检测能力,采用Mixup数据增强方法构建虚拟样本扩充训练集,提升模型的泛化能力和对抗样本的鲁棒性。

2.1 YOLOv5s_P网络架构

YOLOv5s_P模型基于YOLOv5模型框架^[31],网络整体架构包括输入(Input)、骨干网络(Backbone)、颈部(Neck)和头部(Head)模块,如图2所示。Input模块对输入数据集进行预处理, Backbone模块采用CSP-Darknet53跨阶段局部网络提取输入图像特征,可在减少计算量的同时保证检测准确率。Neck模块采用BiFPN结构融合骨干网不同尺度的特征信息,并将图像特征传递到预测层。Head模块利用Backbone提取的特征图或Neck融合的特征图对图像特征进行预测,生成边界框并预测类别。YOLOv5s_P模型与YOLOv5模型的主要区别是使用BiFPN结构替代PANet结构,增强模型对不同尺度对象的特征融合能力,从而在检测精确率和效率之间取得更好的平衡。

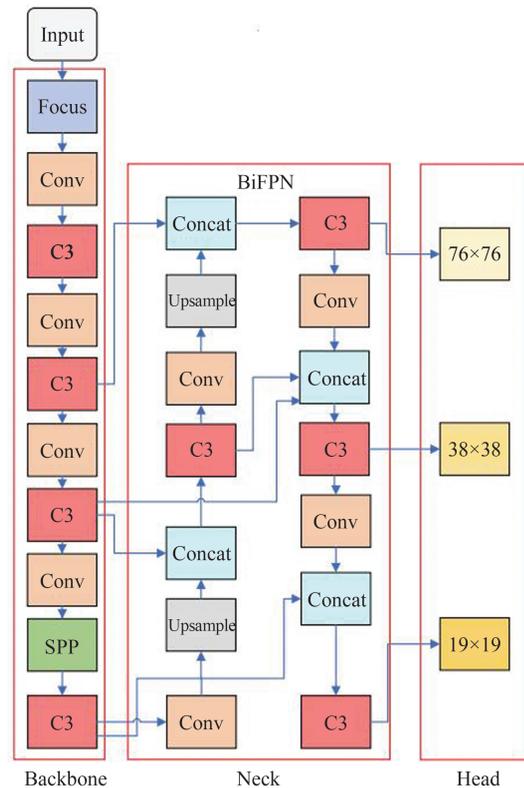


图2 YOLOv5s_P网络架构

Fig. 2 YOLOv5s_P Network Structure

2.2 BiFPN 结构

BiFPN 结构是 Google 公司在 2020 年发表的 EfficientDet 论文^[14]中提出的,该结构借鉴 PANet 结构的多尺度特征融合思想,可以用更少的参数达到更好的检测精度。图 3 对比了 PANet 结构和 BiFPN 结构的不同,BiFPN 结构采用加权双向特征金字塔网络,引入了可学习的权值来调整不同输入特征的贡献,同时重复应用自顶向下和自底向上的多尺度特征融合,集成双向跨尺度连接和快速归一化特点,实现简单快速的多尺度特征融合。

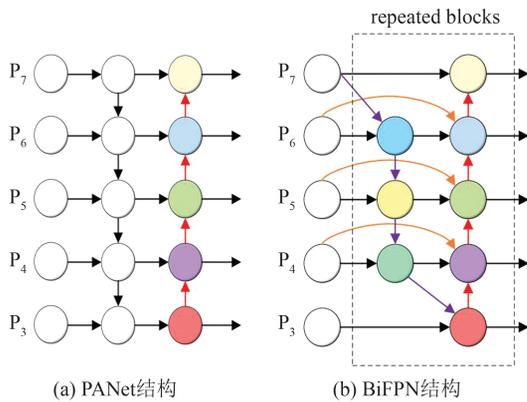


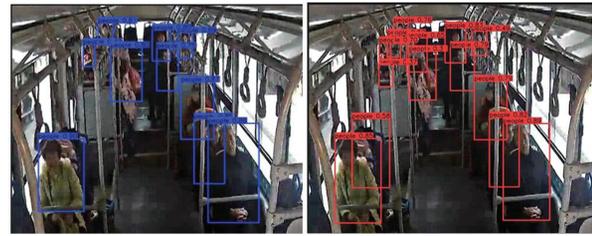
图 3 PANet 和 BiFPN 结构比较图^[14]

Fig. 3 Comparison of PANet and BiFPN Structures^[14]

BiFPN 结构图中,紫色部分为自顶向下的通路,传递的是高层特征语义信息(有利于物体分类),红色部分是自底向上的通路,传递的是低层特征位置信息(有利于物体定位),橙色部分是同一层输入和输出节点间增加的通路。对比 PANet 结构,BiFPN 结构首先删除了只有一条输入或输出边的节点,因为这些节点对特征融合整体效果提升不明显,删减后可减少计算量,简化双向网络。其次,BiFPN 结构在原始输入和输出节点之间添加了额外的通路,可跨层融合更多特征,提高检测精度。由此可见,BiFPN 结构加强了不同尺度特征之间的关联信息,可提升边缘、模糊、遮挡等困难目标的检测效果。图 4 给出了 BiFPN 结构替换 PANet 结构后的效果对比图,结果显示,BiFPN 结构对模糊和遮挡的后排乘客图像检测能力明显提升(图像来源于四川省自贡市 1 路公交车的视频截图)。

2.3 Mixup 数据增强

Mixup 数据增强方法^[16]是 2018 年由 MIT 和 Facebook 人工智能研究院提出,基于邻域风险最小化原则使用线性插值得到新样本数据,增加对抗样本,以提升模型鲁棒性和网络稳定性。为了提高公交车内遮挡目标的检测能



(a) PANet 结构检测结果 (b) BiFPN 结构检测结果

图 4 PANet 与 BiFPN 结构测试对比

Fig. 4 Test Comparison of PANet and BiFPN Structure

力,YOLOv5s_P 模型采用 Mixup 方法将不同图像混合后扩充训练数据集,减少模型在数据集中的泛化误差。Mixup 数据增强的具体做法是随机挑选训练集中的两张图像,将像素值按权重线性相加后生成新的图像,同时将图像样本标签也对应加权求和,计算损失后再反向求导更新参数。Mixup 数据增强原理公式如下:

$$\tilde{x} = \lambda x_i + (1 - \lambda) x_j \quad (1)$$

$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j \quad (2)$$

$$\lambda \sim \text{Beta}(\alpha, \alpha), \alpha \in (0, \infty) \quad (3)$$

式中, x_i 和 x_j 指随机挑选的训练集图像; \tilde{x} 表示经过 Mixup 数据增强后得到的样本; y_i 和 y_j 是 x_i 和 x_j 的标签,即公交车内乘客; \tilde{y} 表示与 \tilde{x} 对应的标签; $\lambda \in [0, 1]$ 表示服从参数为 α 的 Beta 概率分布。

采用 Mixup 数据增强方法融合后的图像如图 5 所示,可见构造出的训练样本图像比原图像增加了噪声,且增加了目标检测乘客的数量(红色框标注),这样用于训练模型的图像样本更复杂。



图 5 Mixup 示例

Fig. 5 Mixup Sample

为验证 Mixup 数据增强方法的效果,图 6 分别对比了使用和未使用数据增强方法的检测效果,结果显示,使用 Mixup 数据增强方法后提升了残缺乘客图像的检测效果,减少了漏检人数,提高了模型的学习能力(图像来源于四川省自贡市 1 路公交车的视频截图)。

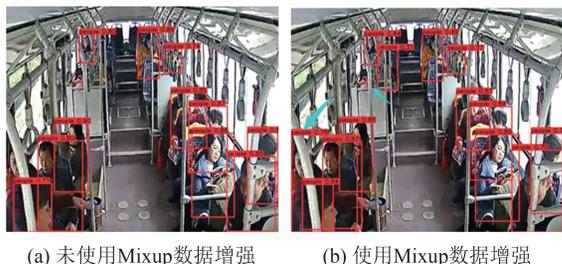


图6 Mixup数据增强检测效果对比

Fig. 6 Test Comparison of Mixup Data Enhancement Effect

3 实验与结果分析

为验证 YOLOv5s_P 模型的性能,将改进后模型应用在实际公交车场景数据集中,并与 Faster R-CNN^[17]、SSD300^[18]、RetinaNet^[19]、YOLOv5s^[20] 4种经典模型进行对比实验,测试不同模型的检测精确度、速度和模型大小等指标,以验证 YOLOv5s_P 模型的性能。

3.1 数据集制作

使用的图像数据集来源于真实公交场景,原始监控视频是由公交车顶前方的摄像头拍摄的,覆盖了除司机位置以外的所有区域,视频采集时间为公交车全天运营时段,涵盖了白天和夜晚不同的光照条件。原始监控视频如图7所示(图像来源于四川省自贡市1路公交车的视频截图),图像边界有日期、路线等字幕,遮挡了公交车内画面。制作数据集时,为消除原始视频边界字幕对网络训练的影响,首先裁剪原始图像周边包含字幕的区域,得到没有字幕干扰的监控图像。图像裁剪后得到1000张数据集,尺寸大小为704×576像素,随机分配获得训练集800张,测试集200张。



图7 原始图像

Fig. 7 Initial Image

观察实验数据集中的图像发现,由于距离摄像头较远,公交车后排座位乘客下半身和头部被完全或部分遮挡,如果仅检测乘客头部或全身容易漏检。而乘客上半身没有被完全遮挡,可由乘

客的衣服、头发、露出的手臂和携带的背包等进行识别。如图8中红色标注部分所示(图像来源于四川省自贡市1路公交车的视频截图),蓝衣和黑衣两名后排乘客的脸部几乎不可见,下半身也被遮挡,但依据上半身的衣服手臂和头发可识别出乘客。因此本文选择标注乘客的上半身作为检测范围,以降低模型训练和检测难度,并减少因头部图像被遮挡造成的漏检。使用 LabelImg 软件标注裁减后的图像,每张图像标注结果保存为 VOC 格式和 YOLO 格式,并将乘客标注类别命名为 person(标签0)。



图8 遮挡示例图像

Fig. 8 Occlusion Sample Image

3.2 评价指标与训练过程

实验评估模型性能的指标包括平均精度均值(mean average precision, mAP_0.5)、检测速度、模型大小、精确率(P)和召回率(R),各指标计算如下:

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

$$P_{mAP_{0.5}} = \frac{\sum_{i=1}^N AP_{i_{0.5}}}{N} \quad (6)$$

$$V_{FPS} = \frac{N_{Figure}}{T_{Total}} \quad (7)$$

式中,TP(true positive)指预测为正样本且实际也是正样本的个数;FP(false positive)指预测为正样本但实际是负样本的个数;FN(false negative)指预测为负样本但实际是正样本的个数; N 表示类别的总数(本文类别为1); N_{Figure} 表示检测图像的总数量; T_{Total} 表示检测使用的总时间。

模型训练过程中,100个迭代周期内平均精度均值、精确率、召回率和验证集损失变化曲线如图9所示。由图9可见,模型参数在warmup后只在训练前期小幅振荡,随着余弦退火算法调整学习率,模型迭代25次后逐渐收敛,收敛速度较快。

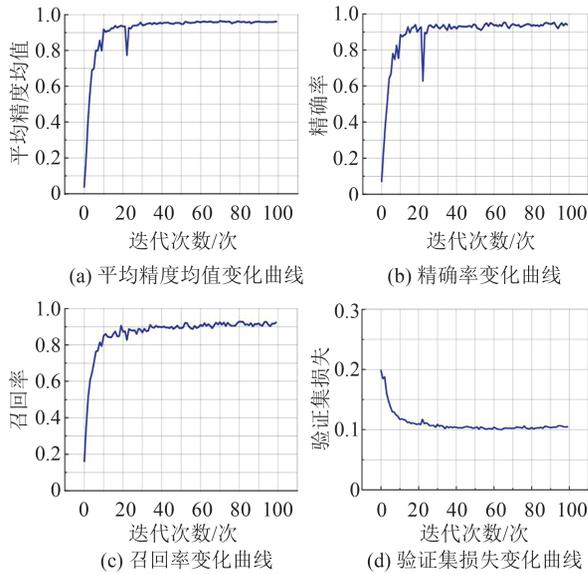


图9 本文模型训练过程

Fig. 9 Training Process of the Proposed Model in this Paper

3.3 实验配置

实验在 Kaggle 平台上测试,采用四核 CPU 和 16 GB RAM, GPU 为 Tesla P100- PCIE- 16 GB, 操作系统为 Linux-5.10.90-x86_64-with-debian- bullseye-sid, 软件版本 Python v3.7.12, 网络框架采用 Pytorch1.7.0。通过观察模型实际的训练效果, 实验中调整合适的超参数, 参数设置如下: 初始学习率设置为 0.01, 使用余弦退火策略调整学习率, 以兼顾学习速度并避免振荡过大。为保证模型收敛速度和泛化能力, Batchsize 设置为 8, 共训练 100 个 epoch。为加快梯度下降和模型参数更新速度, 随机梯度下降 (stochastic gradient descent, SGD) 动量设置为 0.937。权重衰减系数设置为 0.005, 可避免模型过拟合并降低模型的复杂度。Mixup 数据增强的重叠系数设置为 0.05 时, 模型的泛化能力和鲁棒性较好。

3.4 实验结果

在相同实验环境和图像数据集下, 将改进的 YOLOv5s_P 模型与 RetinaNet、Faster R-CNN、SSD300、YOLOv5s 4 种模型的性能进行对比测试, 实验结果如表 1 所示。由表 1 可知, 改进 YOLOv5s_P 模型的 mAP_0.5 达到 96.9%, 性能最优, 模型大小仅为 16.6 MB, 约为 SSD300 模型的 1/6, 同时检测速度达到 96.2 帧/s, 能满足公交客流统计实时性需求。表 1 结果验证了改进模型可提高公交客流检测精度, 同时保证了模型轻量化和检测速度, 具有较高的实用性。

为验证后排和被遮挡乘客的检测效果, 图 10 对

比测试了原始 YOLOv5s 模型和 YOLOv5s_P 模型的检测结果 (图像来源于 <https://www.hunantoday.cn/article/202109/202109082043305798.html>)。根据文献 [32] 对遮挡的定义, 1%~35% 遮挡称为部分遮挡, 36%~80% 遮挡称为严重遮挡。结果显示, YOLOv5s 模型漏检了被严重遮挡的乘客和后排小目标乘客, 而 YOLOv5s_P 模型明显提升了对这些乘客的检测能力。测试结果表明, YOLOv5s_P 模型较原始 YOLOv5s 模型的平均漏检率减少了 3.43%, 乘客截图漏检情况明显改善。

表 1 模型对比实验结果

Tab. 1 Experiment Results of Different Models

模型	mAP_0.5/%	检测速度/ (帧·s ⁻¹)	模型大小/ MB
RetinaNet	90.9	16.4	257.1
Faster R-CNN	91.2	16.7	330.3
SSD300	84.5	66.7	105.1
YOLOv5s	95.6	97.1	14.4
YOLOv5s_P	96.9	96.2	16.6



(a) YOLOv5s模型检测 (b) YOLOv5s_P模型检测

图 10 遮挡和后排乘客检测结果对比

Fig. 10 Detection Results of Occlusion and Rear Passengers

图 11 对比了不同拥挤程度的公交场景下 5 种模型的检测效果 (图像来源于四川省自贡市 1 路公交车的视频截图), 图 11(a)、11(b)、11(c) 分别为乘客数量少、中等、多时的原始图像, 代表 3 种不同拥挤和遮挡程度的客流场景。由图 11 可见, 5 种模型在乘客人数较少和中等时检测精度差异不大, 但在乘客人数较多时, YOLOv5s_P 模型和 YOLOv5s 模型的正确检测人数与双阶段 Faster R-CNN 模型接近, 但 YOLOv5s_P 模型的漏检和错检人数均少于其他模型, 表明本文模型在乘客拥挤时检测优势更明显。

3.5 消融实验

为验证 YOLOv5s_P 模型中各项改进策略提升检测结果的有效性, 进行了消融实验, 分别验证 YOLOv5s 模型中用 BiFPN 结构替换 PANet 结构

的有效性和 Mixup 数据增强效果,以及改进模型 YOLOv5s_P 的整体性能。实验结果如表 2 所示。



图 11 5种模型公交客流检测效果对比

Fig. 11 Detection Effect Comparison of Bus Passenger Flow with Five Models

表 2 消融实验

Tab. 2 Ablation Experiment

模型	mAP _{0.5} /%	R/%	检测速度/ (帧·s ⁻¹)	模型大 小/MB
YOLOv5s	95.6	92.6	97.1	14.4
YOLOv5s+BiFPN	96.0	93.8	96.2	16.6
YOLOv5s+Mixup	95.9	94.3	97.1	14.4
YOLOv5s_P (BiFPN+Mixup)	96.9	94.5	96.2	16.6

由表 2 可知,对比经典 YOLOv5s 模型,引入 BiFPN 结构后模型召回率提升了 1.2%, mAP_{0.5}提升了 0.4%;使用 Mixup 数据增强方法后,召回率提升了 1.7%,且不会引入新参数,对模型大

小和检测速度没有影响;YOLOv5s_P 模型整体性能中 mAP_{0.5}精度提高了 1.3%,召回率提升了 1.9%,可见 YOLOv5s_P 模型在检测速度与模型大小无明显改变情况下,提升了检测准确度,证明了改进策略的有效性。

4 结 语

客流检测是公交智慧调度的关键技术之一,本文针对车内乘客图像遮挡严重、尺寸差距大和后排乘客图像较小等技术难点,提出了改进的目标检测模型 YOLOv5s_P。

首先采用 BiFPN 结构提高多尺度特征图的

融合效率,增强对边缘模糊和遮挡等目标的特征提取;同时联合 Mixup 数据增强方法,提高模型学习和泛化能力。为验证改进模型性能,将公交车内监控图像制作成数据集,对比测试了 YOLOv5s_P 与其他模型的性能。实验结果表明, YOLOv5s_P 模型在保证检测速度和模型轻量化情况下,平均检测精度和召回率均有提升,后排乘客漏检率显著下降,从主观和客观层面验证了模型提升复杂目标检测的有效性。未来希望融入注意力机制和跟踪算法,减少移动乘客导致的人数统计波动。

参 考 文 献

- [1] Xu Meng, Liu Tao, Zhong Shaopeng, et al. Urban Smart Public Transport Studies: Review and Prospect [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2022, 22(2): 91-108. (徐猛,刘涛,钟绍鹏,等.城市智慧公交研究综述与展望[J].*交通运输系统工程与信息*, 2022, 22(2):91-108.)
- [2] Wang Chisheng, Wang Lehan, Zhang Juan, et al. Remote Sensing Video Production and Traffic Information Extraction Based on Urban Skyline [J]. *Geomatics and Information Science of Wuhan University*, 2023, 48(9): 1490-1498. (汪驰升,王乐涵,张涓,等.基于城市天际线的遥感视频生产与交通信息提取[J].*武汉大学学报(信息科学版)*, 2023, 48(9): 1490-1498.)
- [3] Chen Xian, Gong Xiaolin, Wu Lan. Estimation Method of Bus Passengers' Boarding and Alighting Stops in Small and Medium-Sized Cities Based on IC Card and Sparse GPS Data [J]. *Transport Research*, 2022, 8(1): 79-88 (陈娴,龚小林,邬岚.基于IC卡和稀疏GPS数据的中小城市公交乘客上下车站点推算方法[J].*交通运输研究*, 2022, 8(1): 79-88)
- [4] Zhang Lijun. Analysis on the Characteristics of Urban and Rural Public Transportation Based on Mobile Payment Data and Research on Scheduling Optimization [D]. Lanzhou: Lanzhou Jiaotong University, 2020. (张利军.基于移动支付数据的城乡公交出行特征分析及调度优化研究[D].兰州:兰州交通大学,2020.)
- [5] Wang Qinggang, Chen Xuyang. Design and Implementation of Human Body Infrared Counting System Based on SCM [J]. *Information & Communications*, 2020, 1(1): 111-113. (王庆刚,陈旭阳.基于单片机的人体红外计数系统的设计和实现[J].*信息通信*, 2020, 1(1): 111-113.)
- [6] Zhao Z, Zheng P, Xu S, et al. Object Detection with Deep Learning: A Review [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 30(11): 3212-3232.
- [7] Le C, Moh T. Facial Detection in Low Light Environments Using OpenCV [C]// 2022 IEEE World AI IoT Congress, Seattle, WA, USA, 2022.
- [8] Li J, Liang X, Shen S, et al. Scale-Aware Fast R-CNN for Pedestrian Detection [J]. *IEEE Transactions on Multimedia*, 2018, 20(4): 985-996.
- [9] Tian D, Lin C, Zhou J, et al. SA-YOLOv3: An Efficient and Accurate Object Detector Using Self-Attention Mechanism for Autonomous Driving [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(5): 4099-4110.
- [10] Shahbaz A, Jo K. Improved Change Detector Using Dual-Camera Sensors for Intelligent Surveillance Systems [J]. *IEEE Sensors Journal*, 2021, 21(10): 11435-11442.
- [11] Ye T, Zhao Z, Wang S, et al. A Stable Light-Weight and Adaptive Feature Enhanced Convolution Neural Network for Efficient Railway Transit Object Detection [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(10): 17952-17965.
- [12] Wu Y, Guo H, Chakraborty C, et al. Edge Computing Driven Low-Light Image Dynamic Enhancement for Object Detection [J]. *IEEE Transactions on Network Science and Engineering*, 2022, 10(5): 3086-3098.
- [13] Jiao Licheng, Zhang Ruohan, Liu Fang, et al. New Generation Deep Learning for Video Object Detection: A Survey [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(8): 3195-3215.
- [14] Tan M, Pang M, Le Q. EfficientDet: Scalable and Efficient Object Detection [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020.
- [15] Liu S, Qi L, Qin H, et al. Path Aggregation Network for Instance Segmentation [C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018.
- [16] Zhang H, Cisse M, Dauphin Y, et al. Mixup: Beyond Empirical Risk Minimization [C]//International Conference on Learning Representations, Vancouver, BC, Canada, 2018.
- [17] Ren S, He K, 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.
- [18] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector [C]//European Conference on Computer Vision, Amsterdam, The Netherlands, 2016.
- [19] Lin T, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection [C]// IEEE International Conference on Computer Vision, Venice, Italy, 2017.
- [20] Yang G, Feng W, Jin J, et al. Face Mask Recognition System with YOLOv5 Based on Image Recognition [C]//IEEE 6th International Conference on Computer and Communications, Chengdu, China, 2020.
- [21] 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. (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.)
- [22] Wang C, Bochkovskiy A, Liao H. Scaled-YOLOv4: Scaling Cross Stage Partial Network [C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021.
- [23] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014.
- [24] Li Xiangbing, Chen Lian. Face Detection in Natural Scene Based on Improved Faster-RCNN [J]. *Computer Engineering*, 2021, 47(1): 210-216. (李祥兵, 陈炼. 基于改进 Faster-RCNN 的自然场景人脸检测[J]. *计算机工程*, 2021, 47(1): 210-216.)
- [25] Hsu W Y, Wen Y L. Ratio-and-Scale-Aware YOLO for Pedestrian Detection [J]. *IEEE Transactions on Image Processing*, 2021, 30: 934-947.
- [26] Guo Lei, Wang Qiulong, Xue Wei, et al. A Small Object Detection Algorithm Based on Improved YOLOv5 [J]. *Journal of University of Electronic Science and Technology of China*, 2022, 51(2): 251-258. (郭磊, 王邱龙, 薛伟, 等. 基于改进 YOLOv5 的小目标检测算法[J]. *电子科技大学学报*, 2022, 51(2): 251-258.)
- [27] Fan Wei. Research and Implementation of Bus Crowd Counting System [D]. Chengdu: Southwest Jiaotong University, 2018. (范伟. 公交车人群计数系统的研究与实现[D]. 成都: 西南交通大学, 2018.)
- [28] Zhang Kaisheng, Liu Zexin, Guo Bixiao, et al. Research on Bus Passenger Flow Detection System Based on Depth Learning [J]. *Journal of Shihezi University (Natural Science)*, 2019, 37(5): 654-660. (张开生, 刘泽新, 郭碧筱, 等. 基于 CNN 公交客流检测系统的设计[J]. *石河子大学学报(自然科学版)*, 2019, 37(5): 654-660.)
- [29] Liu L, He J, Hou Y, et al. A Technology for Automatically Counting Bus Passenger Based on YOLOv2 and MIL Algorithm [C]// IEEE 5th International Conference on Image, Vision and Computing, Beijing, China, 2020.
- [30] You X, Li G, Zhao Y, et al. Research on Bus Passenger Flow Statistics Based on Video Images [C]// The 2nd International Conference on Electronics, Communications and Information Technology, Sanya, China, 2021.
- [31] Jocher G, Stoken A, Borovec J, et al. Ultralytics/YOLOv5: v3.1-Bug Fixes and Performance Improvements [S]. <https://github.com/ultralytics/yolov5>, 2020.
- [32] Zhou C, Yuan J. Multi-label Learning of Part Detectors for Heavily Occluded Pedestrian Detection [C]// IEEE International Conference on Computer Vision, Venice, Italy, 2017.