

# 控制与决策

Control and Decision

## 基于深度学习的公交客流检测算法

武慧荣, 张敬宜, 郭春敏

引用本文:

武慧荣, 张敬宜, 郭春敏. 基于深度学习的公交客流检测算法[J]. *控制与决策*, 2025, 40(6): 1827-1837.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2024.1245>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### [行人重识别中度量学习方法研究进展](#)

A survey on metric learning in person re-identification

*控制与决策*. 2021, 36(7): 1547-1557 <https://doi.org/10.13195/j.kzyjc.2020.0801>

#### [基于混合整数规划的智能网联车冲突区时序优化模型](#)

Mixed integer programming model of scheduling for connected automated vehicles in a conflict zone

*控制与决策*. 2021, 36(3): 705-710 <https://doi.org/10.13195/j.kzyjc.2019.0886>

#### [无人系统视觉SLAM技术发展现状简析](#)

A survey of visual SLAM in unmanned systems

*控制与决策*. 2021, 36(3): 513-522 <https://doi.org/10.13195/j.kzyjc.2019.1149>

#### [基于卷积神经网络的云雾遮挡舰船目标识别](#)

Obscured ship target recognition based on convolutional neural network

*控制与决策*. 2021, 36(3): 661-668 <https://doi.org/10.13195/j.kzyjc.2019.0781>

#### [机器视觉在轨道交通系统状态检测中的应用综述](#)

A survey of the application of machine vision in rail transit system inspection

*控制与决策*. 2021, 36(2): 257-282 <https://doi.org/10.13195/j.kzyjc.2020.1199>

# 基于深度学习的公交客流检测算法

武慧荣<sup>†</sup>, 张敬宜, 郭春敏

(东北林业大学 土木与交通学院, 哈尔滨 150040)

**摘要:** 针对公交客流检测因忽略边缘计算而导致的数据处理延迟和准确性问题, 基于“云-边-端协同架构”提出一种实时且轻量级的公交客流检测算法 BPF-DETR. 首先, 采用 RT-DETR-r18 作为基线算法来提高实时处理能力; 然后, 引入轻量级 iRMB 模块更新 ResNet-18 作为特征提取主干, 通过倒置残差结构充分学习乘客目标的长距离特征交互以及小目标的局部特征交互, 在提高轻量性和精度的同时增强算法的边缘适用性; 接着, 引入 ASF 架构中的 SSFF 模块和 TFE 模块构建多尺度特征融合模块 MSFM, 以进一步提升所提出算法在多尺度和复杂环境下的检测精度; 最后, 为了验证所提出算法的有效性, 采用基于 ROI 的图像拼接方法, 以提高数据集的代表性和多样性, 构建公交客流监控数据集进行训练验证. 实验结果表明, BPF-DETR 的 mAP@0.5 为 96.4%, 模型大小为 32.6 MB, 均优于目前主流的 YOLO 系列模型, 相较于基线算法, mAP@0.5 提升了 1.1%, 模型大小下降了 16%, 能够满足公交客流检测准确率以及边缘部署轻量化要求.

**关键词:** 交通运输规划与管理; 公交客流; 深度学习; 目标检测; 计算机视觉; 边缘部署

中图分类号: U121 文献标志码: A

DOI: 10.13195/j.kzyjc.2024.1245

引用格式: 武慧荣, 张敬宜, 郭春敏. 基于深度学习的公交客流检测算法 [J]. 控制与决策, 2025, 40(6): 1827-1837.

## Deep learning-based bus passenger flow detection algorithm

WU Hui-rong<sup>†</sup>, ZHANG Jing-yi, GUO Chun-min

(School of Civil Engineering and Transportation, Northeast Forestry University, Harbin 150040, China)

**Abstract:** To address the issues of data processing latency and accuracy in bus passenger flow detection due to the neglect of edge computing, a real-time and lightweight bus passenger flow detection algorithm based on a “cloud-edge-end collaborative architecture”, bus passenger flow-DETR (BPF-DETR), is proposed. Firstly, RT-DETR-r18 is adopted as the baseline algorithm to enhance real-time processing capabilities. Secondly, a lightweight iRMB module is introduced to update ResNet-18 as the feature extraction backbone. Through the inverted residual structure, the model effectively captures long-range feature interactions of passenger targets and local feature interactions of small object targets, enhancing both lightness and accuracy while improving edge applicability. Thirdly, the SSFF module and TFE module components from the ASF architecture are integrated to construct the multi-scale feature fusion module (MSFM), further enhancing the detection accuracy under multi-scale and complex environments. Finally, to validate the effectiveness of the algorithm, a ROI-based image stitching method is used to enhance the representativeness and diversity of the dataset, constructing a bus passenger monitoring dataset for training and validation. Experimental results show that the mAP@0.5 of the BPF-DETR is 96.4%, and the model size is 32.6 MB, which are both superior to current mainstream YOLO series models. In comparison with the baseline algorithm, the mAP@0.5 is improved by 1.1%, and the model size decreases by 16%, meeting the accuracy and lightweight deployment requirements for bus passenger flow detection.

**Keywords:** transportation planning and management; bus passenger flow; deep learning; object detection; computer vision; edge deployment

## 0 引言

智慧公交是智慧交通系统的核心组成部分, 通过实时精确的公交乘客流量检测<sup>[1]</sup>, 交通管理部门可

实施动态控制策略, 能够满足乘客出行需求, 提升公交运营效率, 并鼓励更多市民选择公交出行. 这不仅便于车辆管理和调度, 还能利用收集到的数据深入

收稿日期: 2024-10-24; 录用日期: 2025-01-14.

基金项目: 黑龙江省自然科学基金项目 (YQ2022E003).

责任编委: 龙建成.

<sup>†</sup>通信作者. E-mail: wuhuirong@126.com.

分析公交拥挤度等关键运营指标,为乘客提供决策支持<sup>[2]</sup>.实现这一功能需要实时精准统计车内乘客数,并确保数据与站台以及云端低延迟互联,从而为城市交通规划和管理优化提供实际应用价值<sup>[3]</sup>.

传统公交通过乘客刷卡可统计上车人数,但是无法追踪下车人数,导致无法准确掌握车内实时人数.尽管全威等<sup>[4]</sup>提出了一种在线乘客识别算法来解决 MAC 地址冗余干扰问题,但是,仍然存在隐私保护不足和数据准确性问题.因此,现有公交客流检测技术在隐私和数据精度方面仍然需要改进.

近年来,随着深度学习和计算机视觉技术的快速发展,基于神经网络算法进行目标检测已成为主流趋势<sup>[5]</sup>.该技术可利用车内监控数据进行有效地乘客检测,主要分为二阶段算法和单阶段算法:二阶段算法先定位候选目标,再分类确认类型,代表算法包括 R-CNN、Fast R-CNN 和 Mask R-CNN<sup>[6-8]</sup>;单阶段算法则直接生成目标的类别和位置信息,代表算法主要以 YOLO 系列<sup>[9-12]</sup>为主.尽管 YOLO 算法在实时检测中具有速度优势,但是,在处理复杂场景的精度和鲁棒性方面存在局限,促使研究者对其进行针对性改进来满足特定需求.刘晓雷等<sup>[13]</sup>改进了 YOLOX-SORT 算法,提升了公交客流检测的精确度和实时性,并实现了移动端部署,但是,在极端拥挤和低光照条件下仍然存在挑战;黄丽等<sup>[14]</sup>通过 BiFPN 结构和 Mixup 数据增强技术优化了 YOLOv5 模型,提升了遮挡条件下的检测准确性,但是,面临过拟合风险以及混合样本生成的泛化问题.

Transformer 模型<sup>[15]</sup>最初用于自然语言处理,近年来,逐步应用于计算机视觉领域;2020年,Facebook AI<sup>[16]</sup>提出了基于 Transformer 的端到端目标检测模型 DETR,通过全局预测生成唯一预测框,省去了 NMS 处理,简化了流程并取得了优异性能,但是训练时间较长,限制了其在时效性要求高的场景中的应用;2023年,百度飞桨团队发布了 RT-DETR (real-

time detection transformer)<sup>[17]</sup>,其兼具了实时性和高精度,性能优于大部分 YOLO 系列模型,然而,RT-DETR 对于计算资源需求较高,限制了其在资源受限的边缘设备上的部署,此外,RT-DETR 在小尺寸目标检测中表现不足,影响了其在复杂场景中的实用性.

针对复杂场景下目标检测算法精度和实时性不足的问题,本文提出一种实时且轻量级的公交客流检测算法 (BPF-DETR).首先,选择无需 NMS 后处理的轻量级 RT-DETR-r18 作为基线算法,相较于常见的使用 NMS 后处理的 YOLO 系列算法减少了计算复杂度并增强了算法的鲁棒性;然后,采用轻量级 iRMB<sup>[18]</sup>特征提取主干来实现复杂条件下目标特征的快速提取,以提升模型整体的运行速度并降低内存占用,满足边缘部署的需求;接着,构建多尺度特征融合模块优化不同尺度特征的结合,弥补基线算法在高密度、遮挡和低光照等复杂场景中的局限性,平衡检测效率与准确性;最后,为了验证 BPF-DETR 算法在公交客流监控视频检测的有效性,构建公交客流监控数据集 (BPF-Dataset) 进行训练.在构建数据集过程中,本文提出一种基于感兴趣区域 (ROI)<sup>[19]</sup>的图像拼接方法,能够显著提升算法在实际公交监控场景中的应用效果,同时降低存储需求、节省带宽和传输成本,为实际部署提供了更有力的支持.

## 1 RT-DETR 算法

RT-DETR 是基于 Transformer 结构的实时端到端目标检测网络,通过高效混合编码器和 IoU 感知查询选择,在保证速度的同时显著提升检测精度,性能超越了传统的 YOLO 系列模型.通过省略后处理步骤,实现了无延迟的推理速度和稳定性能.在 RT-DETR 的多个分支模型中,官方提供了以 ResNet 和 HGNet 等为主干的不同模型配置.针对运行速度的需求,本文选用轻量级 RT-DETR-r18 作为基线算法,其网络结构如图 1 所示.

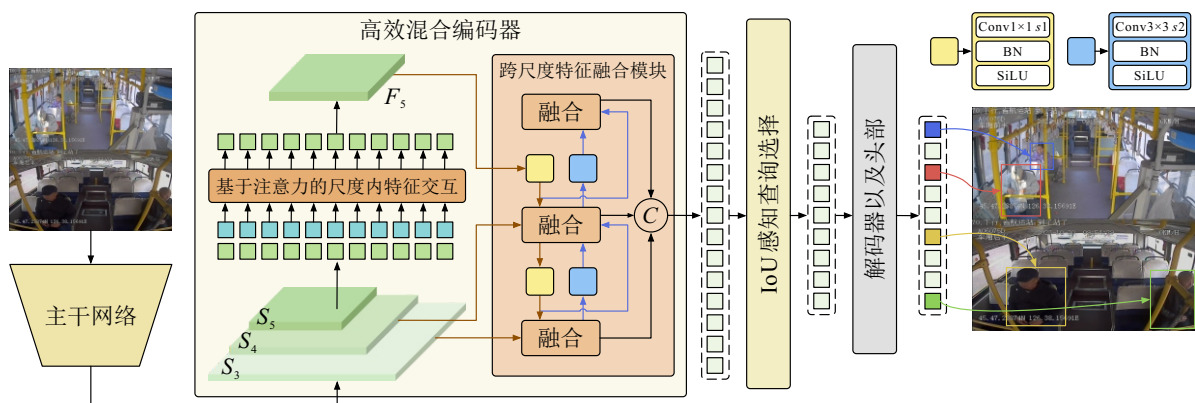


图1 RT-DETR 网络结构

由图1可见,模型由主干网络、混合编码器和带有辅助预测头的解码器组成.首先,主干网络使用ResNet-18网络提取特征,将最后3个阶段的输出作为混合编码器的输入;然后,混合编码器中基于注意力的尺度内特征交互模块(AIFI)用于编码最高层级的特征,跨尺度特征融合模块(CCFM)实现跨尺度特征融合,生成目标检测所需的图像特征;最后,通过IoU感知查询模块从这些特征中选取初始目标查询,生成预测框和置信度评分<sup>[20]</sup>.RT-DETR的编码器结合尺度内交互与跨尺度融合的解耦策略,显著降低了计算成本,实现了实时目标检测.其解码器采用多层Transformer结构,支持在推理过程中灵活调整解码器层数,无需重新训练,确保了推理速度和精确性.

## 2 实时且轻量级 BPF-DETR 算法

在实际公交客流检测过程中,理想检测模式通过公交车监控视频采集客流图像数据,采用“云-边-端协同架构”(如图2所示),由边缘设备就地处理数据并协同云端完成任务.该架构降低了数据传输延迟,提高了处理速度和资源利用效率,保障了复杂环境下的高效运行和检测准确性.此外,边缘设备的数

据处理模式有效保护了用户隐私.通过在边缘端处理图像数据,仅传输处理后的或匿名化数据至云端,避免原始图像直接上传,从而降低隐私泄露风险,符合相关隐私法规和数据保护要求,同时增强了系统的安全性和合规性.

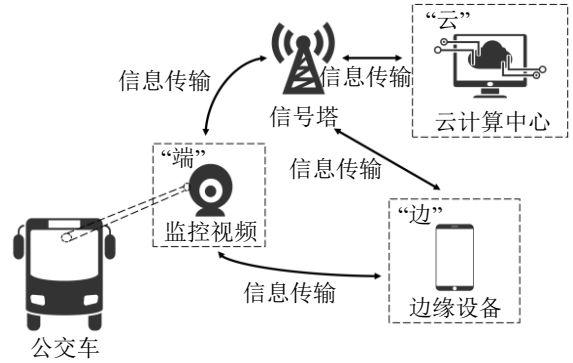


图2 “云-边-端协同架构”示意图

为了确保“云-边-端协同架构”在实际部署中能够发挥最大的效率和准确性,需要开发既高效又轻量级的算法来适应边缘设备的处理能力限制.因此,本文提出实时且轻量级 BPF-DETR 算法以支持在边缘设备上的快速部署和高效运行,网络结构如图3所示:从左至右依次为 iRMB 特征提取主干、多尺度融

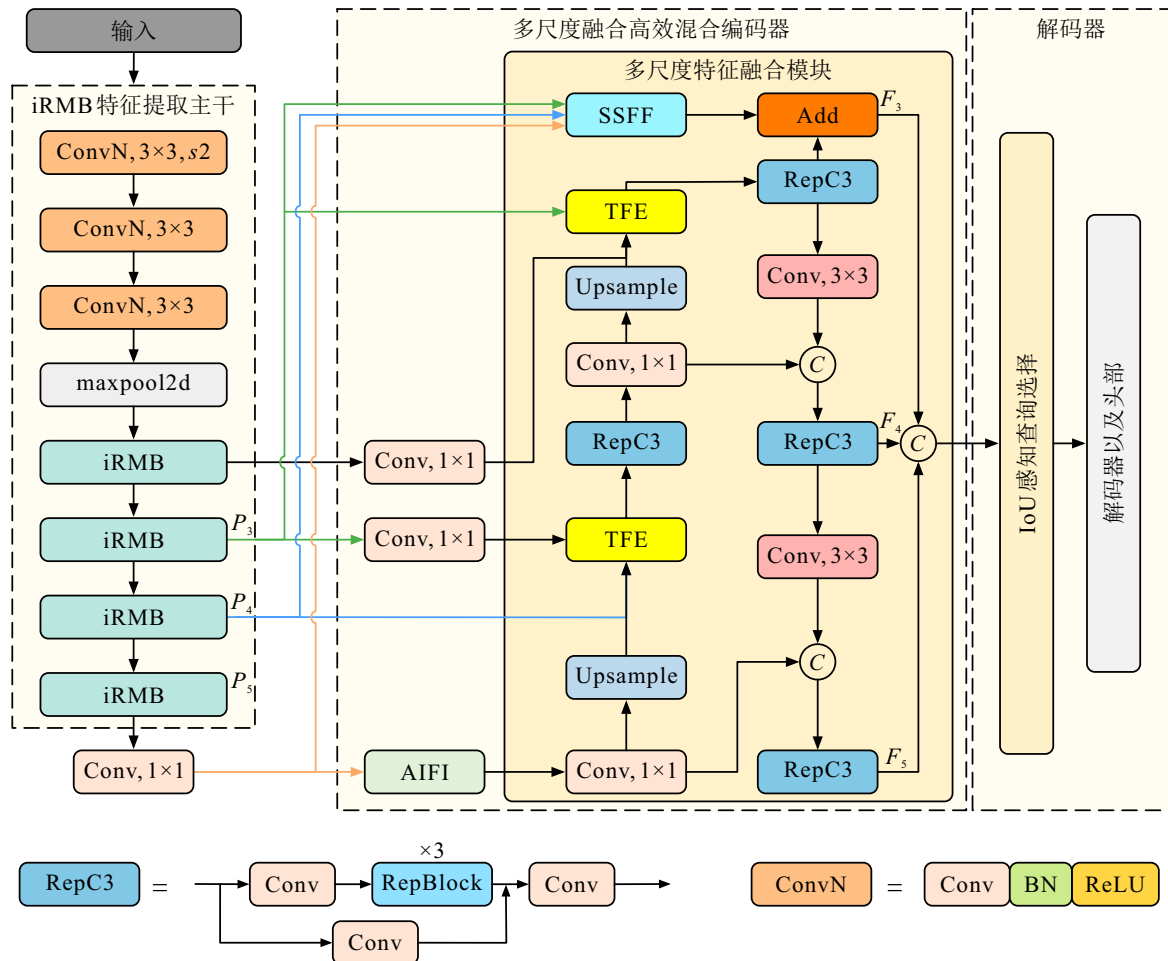


图3 BPF-DETR 网络结构

合高效混合编码器 (MFEHE) 以及解码器 3 个部分。

首先, iRMB 特征提取主干从公交客流监控图像中提取乘客特征映射  $\{P_3, P_4, P_5\}$ ; 然后, 多尺度融合高效混合编码器通过基于注意力的尺度内特征交互模块 (AIFI) 和多尺度特征融合模块 (MSFM) 生成包含丰富定位和语义信息的特征映射  $\{F_3, F_4, F_5\}$ ; 最后, 将  $F_3 \sim F_5$  拼接后输入解码器来获取乘客检测结果, 拼接前后的特征如表 1 所示。

表1  $F_3 \sim F_5$  拼接前后特征

层名称	描述	特征尺寸	特征通道数
$F_3$	特征层 $P_3$ 处理后	$64 \times 64$	256
$F_4$	特征层 $P_4$ 处理后	$32 \times 32$	256
$F_5$	特征层 $P_5$ 处理后	$16 \times 16$	256
Concat	$F_3, F_4, F_5$ 拼接后	$16 \times 16$	768

## 2.1 iRMB 特征提取主干

轻量级算法 ResNet-18 因其较少的参数量和低

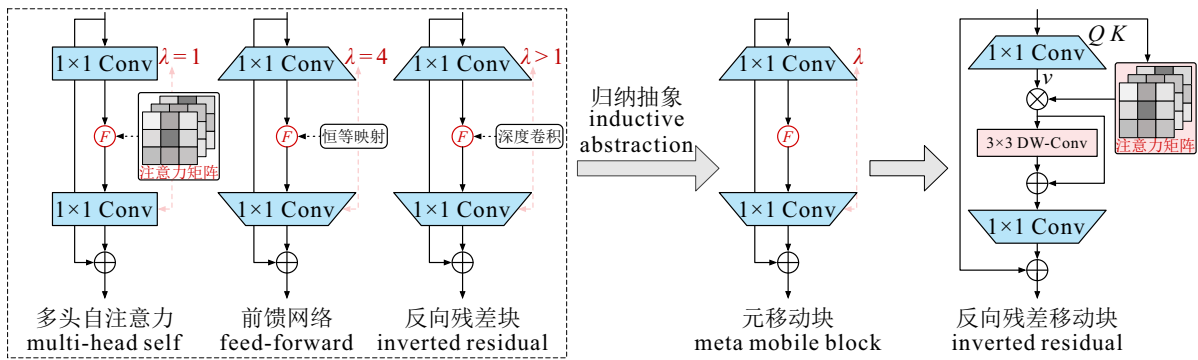


图4 反向残差移动块构建流程

iRMB 先从 MobileNetV2<sup>[21]</sup> 中的倒置残差块 (IRB) 和 Transformer 中的多头自注意力 (MHSA) 以及前馈网络 (FFN) 模块中抽象出共同的设计特征, 归纳并简化为一个包含扩展和收缩操作的单一残差结构——元移动块 (MMB)。MMB 包含一个或多个高效操作单元, 如深度可分离卷积, 旨在通过调节参数  $\lambda$  (扩展比例) 和  $F$  (高效算子) 动态生成具有不同特性的网络模块。

假设给定输入图像为  $X (\in R^{C \times H \times W})$ ,  $R$  为实数集,  $C$ 、 $H$ 、 $W$  分别为图像的通道数、高度、宽度。MMB 首先通过扩展多层感知机 (MLPe), 以输出与输入比  $\lambda$  增加通道维度, 从而扩展输入特征, 得到扩展后的图像  $X_e$ , 如下所示:

$$X_e = \text{MLPe}(X) (\in R^{\lambda C \times H \times W}). \quad (1)$$

然后, 使用高效算子  $F$  进一步增强图像特征, 得到增强后的图像  $X_f$  为

$$X_f = F(X_e) (\in R^{\lambda C \times H \times W}). \quad (2)$$

计算需求而被广泛应用于多个模型的特征提取主干。然而, 在 RT-DETR 算法中, 因其较高的参数数量和计算量未能满足边缘部署的需求, 特别是在公交监控视频应用中, 随着拍摄距离的增加, 画面中小目标乘客的数量也随之增多, 导致算法在远距离提取乘客相关特征的能力下降, 进而引起漏检率上升和检测精度降低。

为了解决这一问题, 并进一步增强算法对小目标乘客特征提取的能力, 采用 iRMB 特征提取主干更新 ResNet-18 作为 BPF-DETR 的特征提取框架。其核心在于引用一种结合了 Transformer 的全局理解能力与卷积神经网络 (CNN) 的局部分析能力的倒置残差移动块 (iRMB) 来替换 ResNet-18 中的 BasicBlock, 从而使得网络在特征提取阶段迅速聚焦关键信息, 进而提升检测精度, 设计构建流程如图 4 所示。

接着, 使用收缩的多层感知机 (MLPs), 以输出与输入比  $\lambda$  收缩通道维度, 得到收缩后的图像  $X_s$  为

$$X_s = \text{MLPs}(X_f) (\in R^{C \times H \times W}). \quad (3)$$

最后, 使用残差连接获得最终输出  $Y = X + X_s$  ( $\in R^{C \times H \times W}$ )。

此设计引入了元移动模块, 整合了 CNN 用于处理静态局部信息的能力与 Transformer 在全局范围内的动态交互学习能力, 开发出了反向残差移动块, 其高效算子  $F$  如下所示:

$$F(\cdot) = \text{Conv}(\text{MHSA}(\cdot)). \quad (4)$$

不同于传统的 MHSA, EW-MHSA 在计算注意力矩阵时, 直接使用未展开的输入  $X$  作为  $Q$  和  $K$ , 即  $Q = K = X (\in R^{C \times H \times W})$ , 而将展开的输入  $X_e$  用作  $V = X_e (\in R^{\lambda C \times H \times W})$ 。iRMB 的这种改进采用了一种改进的扩展窗口多头自注意力 (EW-MHSA), Skip 为跳跃连接操作, 如下所示:

$$F(\cdot) = (\text{DW-Conv}, \text{Skip})(\text{EW-MHSA}(\cdot)). \quad (5)$$

该机制通过计算输入特征间的关系来捕捉图像的全局内容与远距离的依赖关系. 通过结合深度可分离卷积 (DW-Conv) 与一种改进的扩展窗口多头自注意力机制 (EW-MHSA), 优化了模型处理短距离与长距离依赖关系的能力, 提升了模型的计算效率和性能, 同时降低了计算复杂度和参数数量. iRMB 采用深度可分离卷积技术, 允许每个通道进行独立卷积运算, 从而减少运算资源的消耗. 此外, iRMB 通过其多重瓶颈结构, 将输入图像转换为特征图的初始层, 再将数据进行多轮特征处理下采样, 生成不同尺度的特征映射, 从而有效提取公交监控图像中尺寸较小的乘客目标. 倒置残差移动块融合了捕捉动态全局上下文和整合静态局部特征的功能, 从而增强了特征信息的流动性和有效性, 因此, 选择倒置残差移动块来替换 ResNet-18 中的基本块<sup>[22]</sup>.

### 2.2 多尺度特征融合模块 (MSFM)

考虑到基线算法在卷积和下采样过程中丢失了少量的目标信息. 为了解决这一问题, 本文引入 ASF (attentional scale sequence fusion)<sup>[23]</sup> 网络中的 SSFF 和 TFE 模块集成至 RT-DETR 的 CCFM 模块, 提出一种新的多尺度特征融合模块 (MSFM). 多尺度融合高效混合编码器 (MFEHE) 通过结合 AIFI 和 MSFM 来优化特征处理. AIFI 模块加强模型在同一尺度内的空间特征相关性和精确度; 而 MSFM 模块则融合不同尺度的全局和高层特征信息, 增强对目标尺寸变化的适应性. 该结合不仅提升了特征的整体表达力, 也优化了小型目标在复杂环境中的检测精确度. 局部和全局特征信息的有效融合进一步确保了检测结果的准确性. 图 3 中的多尺度融合高效混合编码器展示了 MSFM 框架的概览, 该框架结合了空间和多尺度特征. 三重特征编码模块 (TFE) 捕获小目标乘客的细节信息, 而尺度序列特征融合模

块 (SSFF) 融合来自不同尺度图像的全局或高级特征信息, 为小目标乘客检测提供互补支持.

最初, SSFF 模块有效地融合来自 BPF-DETR 主干网络  $P_3 \sim P_5$  的特征信息. 这 3 层提取了不同空间尺度下乘客在尺寸和形状上的特征信息. SSFF 模块将  $P_3 \sim P_5$  的特征图调整至相同的尺度, 通过上采样后连接, 作为 3D 卷积操作的输入数据, 该卷积有效地结合了来自不同尺度的特征; RepC3 模块通过多次的卷积和非线性激活操作, 有效地扩展特征表达能力, 增强特征的非线性表达, 有助于复杂场景下的对象检测; TFE 模块通过结合空间维度中 3 种不同大小的特征图来捕获关于小目标乘客的详细信息以提高网络识别拥挤公交环境中小目标乘客的能力. 然后, TFE 模块的数据通过路径聚合网络 (PANet) 结构被纳入每个特征分支, 与来自 SSFF 模块的多尺度数据合并, 创建  $F_3$  分支. 最后,  $F_3$  分支与 RepC3 模块输出的  $F_4$ 、 $F_5$  分支整合作为解码器的输入. 这样的特征处理策略极大地优化了模型在复杂环境中对小目标的检测精确度.

#### 2.2.1 尺度序列特征融合模块 (SSFF)

SSFF 模块基于高分辨率特征图的  $P_3$  层设计, 因为它在检测过程中能够提供更全面、更准确的信息, 尤其适用于小乘客目标的识别和检测, 具体结构如图 5 所示. 首先, 使用  $1 \times 1$  的二维卷积调整  $P_4$  层和  $P_5$  层的通道数以匹配  $P_3$  层; 然后, 采用最近邻插值法调整其特征图尺寸以匹配  $P_3$  层, 通过使用 Unsqueeze 方法扩展每个特征层的维度, 将其从三维张量 [高度, 宽度, 通道] 转换为四维张量 [深度, 高度, 宽度, 通道]; 接着, 通过沿深度维度连接四维特征图, 创建三维特征图, 为后续的卷积操作做准备; 最后, 使用 SiLU 激活函数、三维卷积和三维批量归一化来完成尺度序列特征提取过程<sup>[24]</sup>.

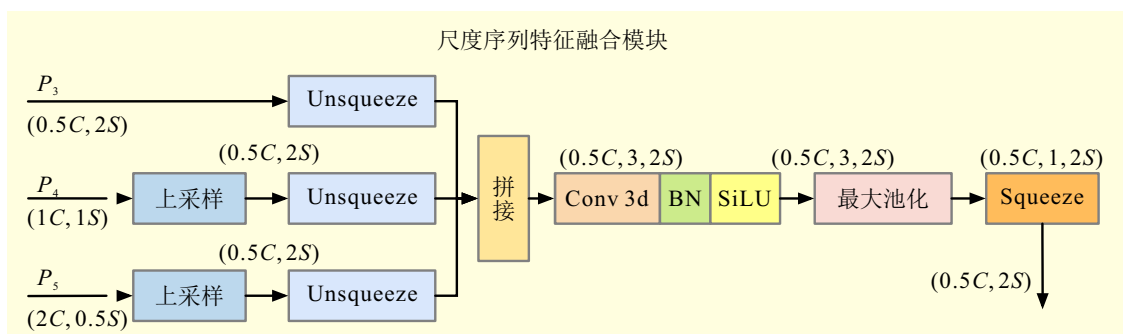


图5 尺度序列特征融合模块细节

#### 2.2.2 三重特征编码模块 (TFE)

放大图像并比较不同尺度下的形状或外观变化是识别密集重叠的小物体的一种有用方法. 然而, 传

统的特征金字塔网络 (FPN) 仅对小尺寸特征图上采样并与前一层特征融合, 忽略了大尺寸特征图中的细节信息, 影响了特征提取的完整性. 为了解决这一

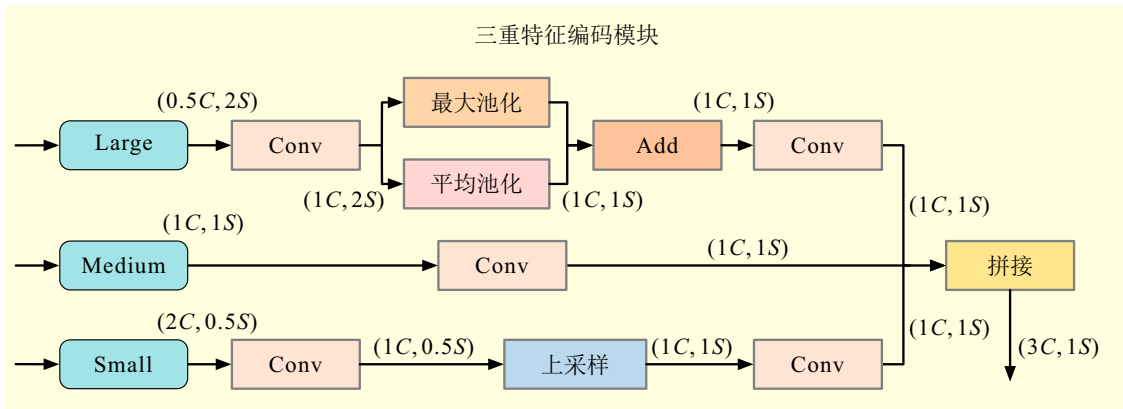


图6 三重特征编码模块细节

问题, TFE 模块通过划分并处理大、中、小尺寸的特征图, 强化不同尺度特征间的关系, 从而提升细节信息的表达和融合精度. TFE 模块的结构如图 6 所示.

特征通道数首先被更改来对应中等尺寸的特征通道, 以便于特征编码. 在处理大尺寸特征图时, 卷积模块将其通道数设定为  $1C$ . 然后, 通过结合最大池化和平均池化的混合结构实现下采样, 既能保留高分辨率特征, 又能有效减少信息丢失, 确保特征的多样性和有效性; 对于小尺寸特征图, 卷积模块也对其通道数进行调整, 并使用最近邻插值法进行上采样, 以避免小目标特征信息的丢失, 同时保持低分辨率图像中的局部特征丰富性<sup>[24]</sup>. 最后, 将大、中、小尺寸特征图调整至相同尺寸后, 分别进行另一轮卷积. 通过这些卷积操作, TFE 模块能够有效地融合来自不同尺度的特征信息, 并最终生成一个包含不同尺度细节的统一特征图, 如下所示:

$$F_{\text{TFE}} = \text{Concat}(F_l, F_m, F_s). \quad (6)$$

其中:  $F_{\text{TFE}}$  为由 TFE 模块生成的特征图,  $F_l$ 、 $F_m$  和  $F_s$  分别为大、中、小尺寸的特征图.  $F_{\text{TFE}}$  是通过连接  $F_l$ 、 $F_m$  和  $F_s$  生成的.

## 3 实验与结果分析

### 3.1 数据集构建

鉴于公共交通监控视频的公开数据集较为稀缺, 本文采用自建数据集来支持公交客流检测的实验需求. 数据来源于哈尔滨市某公交线路的真实车厢场景, 确保了实验数据的真实性和相关性. 为了避免监控视频相邻帧的高度相似性, 采用间隔抽帧法提取样本图像构建数据集, 同时对乘客面部信息进行遮蔽以保护隐私. 原始视频由公交车顶部前段和中段的两个摄像头录制, 覆盖除驾驶员区域外的所有乘客区域, 记录白天和夜晚的不同光照条件. 为了提高数据集的完整性和实用性, 本文针对遇到的挑战提出了解决策略, 具体如下.

### 3.1.1 视频边界字幕遮挡

在公交车监控系统中, 前摄像头拍摄的图像边界包含日期、时间和车站等字幕信息, 遮挡了部分后排乘客, 导致图像边缘信息丢失. 为了解决这一问题, 本文将前段和中段摄像头拍摄的图像垂直拼接, 有效补充了字幕遮挡区域的乘客信息.

同时, 为了避免双摄像头拍摄区域重叠导致的双摄像头拍摄区域重叠导致的双摄像头拍摄区域重叠导致的后排乘客重复检测问题, 本文采用感兴趣区域 (ROI) 方法来精确定义目标检测范围. 该方法通过在前摄像头图像中标定仅由前摄像头覆盖而后摄像头不涉及的区域作为 ROI, 从而避免后摄像头覆盖区域的数据干扰. 基于 ROI 的图像拼接应用前后的检测效



(a) 传统单摄像头图像检测



(b) 基于ROI的双摄像头图像拼接检测

图7 基于 ROI 的图像拼接应用前后检测效果对比

果对比如图7所示。图7(a)中因字幕遮挡未检测出部分后排乘客;而图7(b)中通过ROI方法拼接双摄像头数据后,成功补充了遮挡区域的乘客信息,显著提高了检测的准确性和完整性。这一方法不仅解决了重复检测问题,还提升了监控系统的检测效率和精度。

### 3.1.2 车厢内部遮挡问题

公交车内基础设施(如座椅背部和支撑杆等)常导致前后排乘客的下半身或头部被遮挡,使得依靠头部或局部的检测方式易出现漏检。然而,乘客通常未完全被遮挡,可通过衣物、头发、手臂或随身物品等特征进行识别。为此,本文标注乘客全身作为检测目标,简化模型训练并减少因局部遮挡导致的漏检问题。拼接后的数据集包含2990张图像,尺寸大小为640×640像素。本文使用LabelImg对数据集进行标注,标签统计如图8所示。数据集按照7:1:2的比例划分为训练集、验证集和测试集,分别包含2093张、299张和598张图像,对应的标注实例数量分别为28853个、3917个和8405个。公交在白天行驶时使用高清摄像头,拍摄图片为RGB高清图像;夜晚运行时启用红外摄像头,拍摄的是黑白图像。

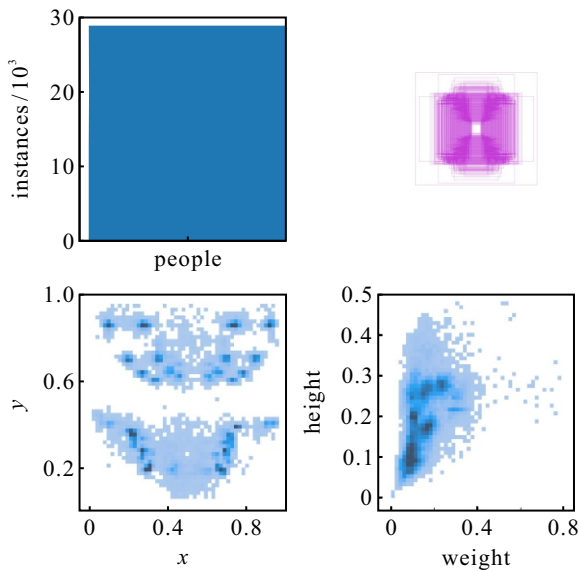


图8 数据集标签详情统计

### 3.2 实验环境与评价指标

本文实验环境为Ubuntu22.04.3操作系统,

Pytorch2.1.2, Python3.10.8和Cuda12.1。实验平台配置如下:CPU(Intel(R)Xeon(R)Platinum 8474C),GPU(RTX 4090D(24 GB))。训练超参数设置如下:训练轮次为100,批次大小为16,输入图像尺寸为640×640,学习率为0.0001,学习率动量为0.9,并采用AdamW优化器对训练过程进行优化。

为了客观评价算法在公交场景中的检测性能,采用精确率( $P$ )、召回率( $R$ )、均值平均精度(mAP)、检测速度(FPS)、模型大小(size)、参数量(parameters)和计算复杂度(GFLOPs)作为衡量算法性能的指标。其中:mAP@0.5表示在IoU(intersection over union)阈值为0.5时,所有类别的平均精度(AP);mAP@0.5:0.95则是IoU从0.5~0.95(步长为0.05)范围内加权平均得到的平均精度。相关公式如下所示:

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

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

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i, \quad (9)$$

$$mAP@0.5 = \frac{1}{n} \sum_{i=1}^n AP_i. \quad (10)$$

其中:TP表示判定为正类的正样本数量,FP表示判定为正类的负样本数量,FN表示判定为负类的正样本数量, $n$ 为类别个数(本文类别为1)。

### 3.3 对比实验

为了进一步验证所提出算法在公交客流检测方面的优越性,本文将BPF-DETR与YOLOv5m、YOLOv6m、YOLOv8m、RT-DETR进行对比。所有算法均在同一数据集上进行训练和测试,实验环境和参数设置保持一致,实验结果如表2所示。

实验结果表明:BPF-DETR在关键性能指标上表现突出,其中mAP@0.5达到了96.4%,性能最优,相较于基线模型RT-DETR与次优模型YOLOv8m分别提升了1.1%和1.3%,精度分别提高了1.4%和1.6%,这一性能提升主要得益于优化的多尺度特征融合策略,增强了复杂场景适应性以及细节处理能力,更能够有效识别不同尺度和密度的目标,这对

表2 不同模型的检测性能对比

模型	$P$ /%	$R$ /%	mAP@0.5/%	mAP@0.5:0.95/%	FPS/(f/s)	size/MB	参数量/ $10^6$	GFLOPs
YOLOv5m	91.1	90.9	95.0	65.7	93.1	50.5	25.1	64.4
YOLOv6m	90.7	91.3	94.7	65.1	60.5	104.3	52.0	161.1
YOLOv8m	91.4	91.6	95.1	67.1	78.3	52.0	25.8	78.7
RT-DETR	91.6	90.8	95.3	69.5	71.6	38.6	19.9	56.9
BPF-DETR	93.0	93.1	96.4	72.4	62.8	32.6	16.7	53.6



域的关注程度. 在 RT-DETR-r18 + iRMB 配置中, 乘客头部区域的活跃度显著提高, 通过优化特征提取机制提升了对长距离和小目标特征的学习能力. 同时, 在 RT-DETR-r18 + MSFM 配置中, 通过多尺度特征融合技术在多个尺度上捕捉乘客细节的优势, 显著提升了模型的整体识别和判别能力. 引入 iRMB 和 MSFM 模块后, BPF-DETR 能够提取包含丰富定位和语义信息的特征, 对乘客关键特征区域的关注度显著提升, 尤其在头部和手部等关键部位表现出. 这一改进对于公交车监控系统中精确的乘客行

为分析至关重要. 可视化热力图验证了所提出算法在提升检测精度和语义信息处理能力方面的有效性, 同时, 清晰地展现了模型对关键区域的聚焦能力. 这些改进使得所提出算法在精度和效率上相比于基线算法表现出显著优势, 体现了其在实际应用中的潜力.

### 3.5 检测结果可视化

根据文献 [26] 提出的公共交通中乘客拥挤度的量化标准 (如表 4 所示), 测试 5 种算法在不同拥挤度和光照条件下公交客流的检测效果 (如图 11 所

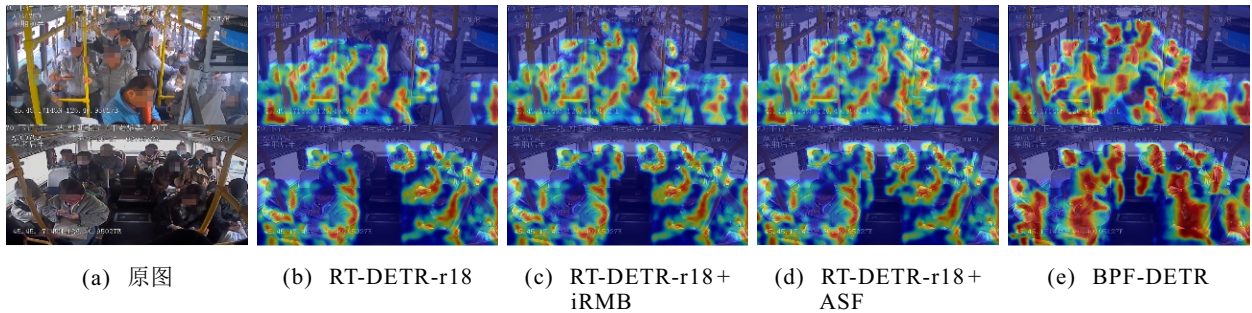


图10 消融实验热力图对比

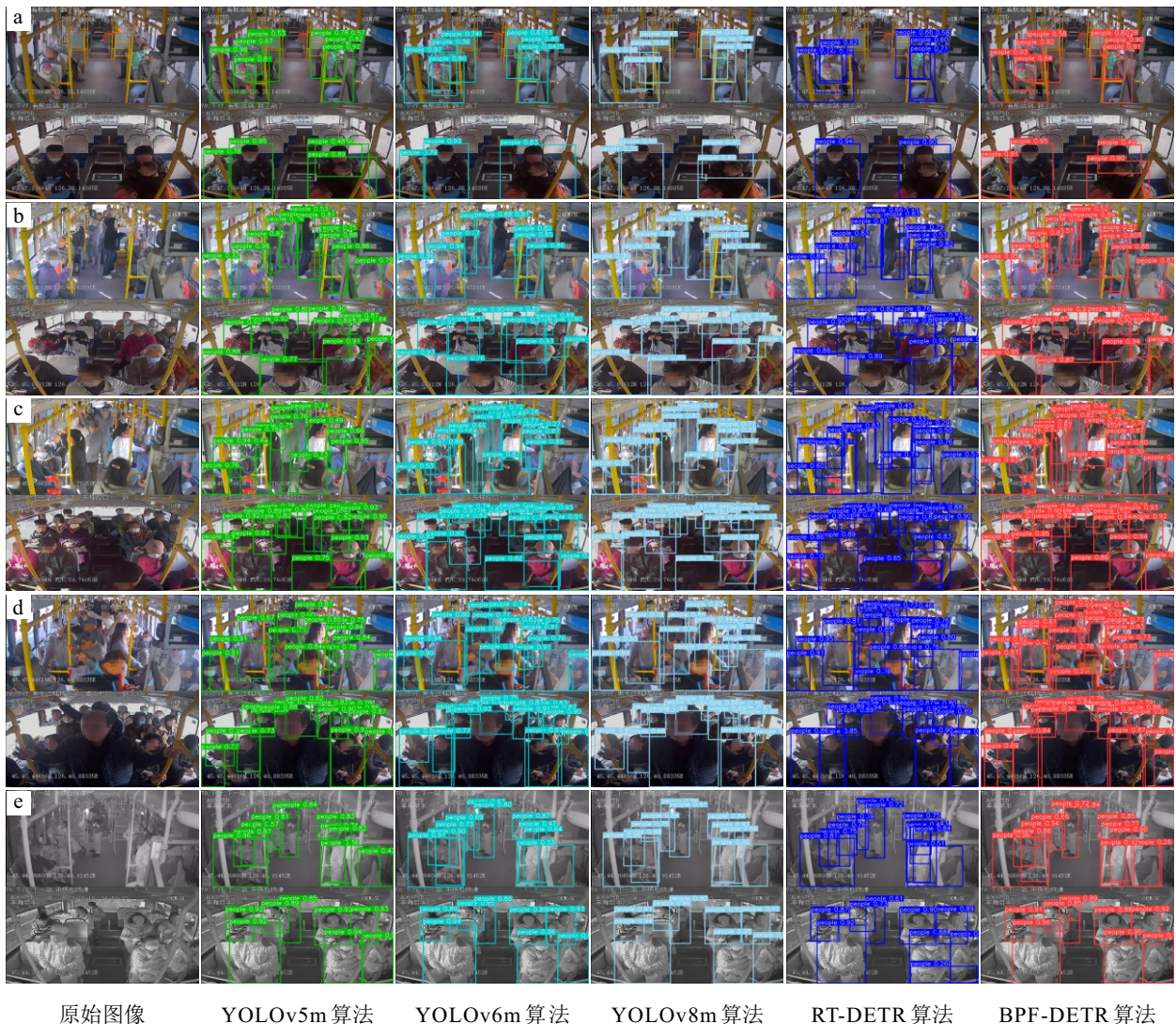


图11 5种算法公交客流检测效果对比

示). 图 11 中的 a ~ d 分别为 4 种不同拥挤度, 即舒适、正常、拥挤、非常拥挤的客流场景的原始图像, 图 11 的 e 为夜视环境下拍摄的黑白图像, 表示光照不足的场景. 这些图像清晰地标示了在白天光照充足和夜晚光照不足两种情况的客流场景, 为评估不同模型在实际应用中的表现提供了直观的参考依据.

表4 公共交通乘客拥挤度量化标准

拥挤度	站立乘客面积 (平方米/人)	乘客拥挤程度描述
舒适	$\geq 1.00$	乘客具备充足空间移动, 大多有座位
正常	[0.40, 1.00)	尚有移动空间, 站立乘客几乎无接触
拥挤	[0.20, 0.40)	站立乘客有身体接触, 明显感到拥挤
非常拥挤	$< 0.20$	站立乘客密集接触, 极度妨碍乘客进出

由图 11 可见, 5 种算法在乘客人数较少时检测精度相差不大. 然而, 当乘客人数增至中等或较多时, 所提出改进 BPF-DETR 算法和 RT-DETR 算法的正确检测人数与 YOLOv8m 算法接近, 但是, 所提出改进 BPF-DETR 算法的漏检和错检人数均少于其他算法, 表明所提出算法在乘客拥挤时检测优势更明显. 同时, 在光照不足的公交场景下, 所提出改进 BPF-DETR 算法的正确检测人数均高于其他算法, 表明所提出算法在光照不足的复杂环境中同样表现出较强的鲁棒性和适应性.

## 4 结论

针对公交客流检测准确率和满足“云-边-端协同架构”的边缘部署要求, 本文基于 RT-DETR 进行了改进, 提出了一种实时且轻量级的 BPF-DETR 算法. 通过采用轻量级 iRMB 网络作为特征提取主干, 捕捉了乘客目标的长距离特征交互以及小目标的局部特征交互, 构建了多尺度特征融合模块高效整合多尺度关键特征信息, 增强了多尺度特征的融合能力. 将 BPF-DETR 在公交客流监控数据集进行了实验验证, 实验结果表明所提出算法在公交客流检测任务中不仅提升了检测的精度和速度, 同时也满足了边缘部署的轻量化需求. 此研究的成果有助于优化公交服务, 提高公交系统的运营效率. 未来研究将在智能边缘设备部署所提出算法, 进一步进行公交实地检测应用研究. 此外, 将进一步优化算法, 减少其参数量和计算需求, 使其能够适应更广泛的公交客流检测任务, 提升模型的鲁棒性和泛化能力.

## 参考文献 (References)

- 温震宇, 胡慧峰, 钱滨, 等. 多端交通视频分析任务卸载决策[J]. 控制与决策, 2024, 39(8): 2773-2782.  
(Wen Z Y, Hu H F, Qian B, et al. Multi-client traffic video analysis task offloading decision[J]. Control and Decision, 2024, 39(8): 2773-2782.)
- Pei Y L, Ran S M, Wang W J, et al. Bus-passenger-flow prediction model based on WPD, attention mechanism, and Bi-LSTM[J]. Sustainability, 2023, 15(20): 14889.
- 周厚盛, 戚建国, 杨立兴, 等. 基于灵活编组的城轨车底运用计划及鲁棒客流控制策略[J]. 控制与决策, 2023, 38(9): 2663-2671.  
(Zhou H S, Qi J G, Yang L X, et al. Joint optimization for rolling stock circulation plan based on flexible train composition mode and robust passenger flow control strategy on urban rail transit lines[J]. Control and Decision, 2023, 38(9): 2663-2671.)
- 全威, 王华, 满永兴, 等. 基于终端定位信息的公交客流采集技术[J]. 重庆交通大学学报: 自然科学版, 2022, 41(11): 52-57.  
(Quan W, Wang H, Man Y X, et al. Bus passenger flow collection technology based on terminal positioning information[J]. Journal of Chongqing Jiaotong University: Natural Science, 2022, 41(11): 52-57.)
- 金沙沙, 龙伟, 胡灵犀, 等. 多目标检测与跟踪算法在智能交通监控系统中的研究进展[J]. 控制与决策, 2023, 38(4): 890-901.  
(Jin S S, Long W, Hu L X, et al. Research progress of detection and multi-object tracking algorithm in intelligent traffic monitoring system[J]. Control and Decision, 2023, 38(4): 890-901.)
- Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, 2014: 580-587.
- 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.
- He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision. Venice, 2017: 2961-2969.
- 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. Las Vegas, 2016: 779-788.
- Redmon J, Farhadi A. Yolov3: An incremental improvement[R]. Computer Vision and Pattern Recognition. 2018.
- Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[J/OL]. 2020, arXiv: 2004.10934.
- 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]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, 2023: 7464-7475.
- 刘晓雷, 张维忠, 张宏峰. 基于改进的 YOLOX-SORT 算法公交客流实时检测方法[J]. 青岛大学学报: 工程技术版, 2022, 37(4): 23-29.

- (Liu X L, Zhang W Z, Zhang H F. Realtime detection method of bus passenger flow based on the improved YOLOX-SORT algorithm[J]. Journal of Qingdao University: Engineering & Technology Edition, 2022, 37(4): 23-29.)
- [14] 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.  
(Huang L, Gong Z P, Liu F F, 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.)
- [15] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30: 5998-6008.
- [16] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]. European Conference on Computer Vision. Glasgow, 2020: 213-229.
- [17] Zhao Y A, Lv W Y, Xu S L, et al. DETRs beat YOLOs on real-time object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, 2024: 16965-16974.
- [18] Zhang J N, Li X T, Li J, et al. Rethinking mobile block for efficient attention-based models[C]. IEEE/CVF International Conference on Computer Vision. Paris, 2023: 1389-1400.
- [19] Aliouat A, Kouadria N, Maimour M, et al. Region-of-interest based video coding strategy for low bitrate surveillance systems[C]. The 19th International Multi-Conference on Systems, Signals & Devices. Sétif, 2022: 1357-1362.
- [20] Rezatofighi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, 2019: 658-666.
- [21] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 4510-4520.
- [22] 胡继文, 张国梁, 沈明哲, 等. 面向松木表面缺陷检测的改进 RT-DETR 模型[J]. 农业工程学报, 2024, 40(7): 210-218.  
(Hu J W, Zhang G L, Shen M Z, et al. Detecting surface defects of pine wood using an improved RT-DETR model[J]. Transactions of the Chinese Society of Agricultural Engineering, 2024, 40(7): 210-218.)
- [23] Kang M, Ting C M, Ting F F, et al. ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation[J]. Image and Vision Computing, 2024, 147: 105057.
- [24] 郎德宝, 周凯红. 改进 YOLOv8 的轻量化轴承缺陷检测算法[J]. 现代电子技术, 2024, 47(19): 115-122.  
(Lang D B, Zhou K H. Lightweight bearing defect detection algorithm combined with improved YOLOv8[J]. Modern Electronics Technique, 2024, 47(19): 115-122.)
- [25] Zhang S S, Yang J, Schiele B. Occluded pedestrian detection through guided attention in CNNs[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 6995-7003.
- [26] Zhao J D, Lei W, Li Z J, et al. Detection of crowdedness in bus compartments based on ResNet algorithm and video images[J]. Multimedia Tools and Applications, 2022, 81(4): 4753-4780.

## 作者简介

武慧荣 (1980-), 女, 副教授, 博士, 主要研究方向为交通运输系统规划与设计、智能交通系统, E-mail: wuhuirong@126.com;

张敬宜 (1999-), 女, 硕士生, 主要研究方向为智能交通系统, E-mail: zjy0@nefu.edu.cn;

郭春敏 (2001-), 女, 硕士生, 主要研究方向为交通运输系统规划与设计, E-mail: 3415868884@qq.com.