

DOI: 10.13382/j.jemi.B2508132

改进 YOLOv11 的电梯乘客异常行为检测算法*

刘伟杰^{1,2} 李志星^{1,2} 庞玉东^{1,2}

(1. 北京建筑大学机电与车辆工程学院 北京 100044; 2. 北京建筑大学城市轨道交通车辆服役性能保障重点实验室 北京 100044)

摘要:针对电梯内乘客发生异常行为可能给乘客带来安全隐患的问题,提出一种改进 YOLOv11 的异常行为目标检测模型 YOLO_LP。首先,在骨干网络中引入全新的特征提取组件跨阶段局部 Transformer 模块(CSP-PTM),其拥有强大的局部和全局特征提取能力,能够有效提高模型的检测精度;其次,引入上下文信息融合模块改进特征金字塔网络,此方法基于权重化思想对特征信息进行重组,有效提升了特征图的判别能力;再次,利用损失函数 WIoU 解决目标类别和大小不平衡的问题,进一步提升模型的精度和收敛速度;最后,采用全新设计的轻量化检测头(LDH)替换原有检测头使网络模型实现轻量化。实验结果表明,在检测电梯内乘客的异常行为时,改进的模型精度达到了 90.4%,高于原始模型 3.5%。此外,相比于 YOLOv11n 模型,平均精度均值(mAP)mAP@0.5 和 mAP@0.5:0.95 分别提高了 2.9%和 2.1%,参数量和计算量分别降低了 10%和 17%。可见 YOLO_LP 模型的综合性能更优,满足在电梯轿厢内进行乘客异常行为检测的精度和速度要求。

关键词:目标检测;异常行为;YOLOv11;特征提取;上下文信息

中图分类号: TN762; TP277 **文献标识码:** A **国家标准学科分类代码:** 520.2060

The algorithm for detecting abnormal behaviors of elevator passengers with improved YOLOv11

Liu Weijie^{1,2} Li Zhixing^{1,2} Pang Yudong^{1,2}

(1. School of Electromechanical and Vehicle Engineering, Beijing University of Architecture and Engineering, Beijing 100044, China; 2. Key Laboratory of Service Performance Guarantee of Urban Rail Transit Vehicles, Beijing University of Architecture and Engineering, Beijing 100044, China)

Abstract: In response to the safety risks associated with abnormal passenger behavior in elevators, an enhanced anomaly detection model, YOLO_LP, based on YOLOv11, is proposed. First, a novel feature extraction component, CSP-PTM, is incorporated into the backbone network. This component enables powerful local and global feature extraction, significantly improving the model's detection accuracy. Next, a contextual information fusion module is introduced to enhance the feature pyramid network. This approach reorganizes feature information through a weighting mechanism, effectively improving the discriminative capability of the feature maps. Additionally, the wise intersection over union loss (WIoU) function is employed to address class and size imbalances, further enhancing the model's accuracy and convergence speed. Finally, a newly designed LDH detection head replaces the original, resulting in a lightweight network model. Experimental results demonstrate that the improved model achieves an accuracy of 90.4% in detecting abnormal passenger behavior in elevators, 3.5% higher than the baseline model. Furthermore, compared to the YOLOv11n model, it shows improvements of 2.9% and 2.1% in mAP@0.5 and mAP@0.5:0.95, respectively, while reducing the number of parameters and computational load by 10% and 17%, respectively. These findings highlight the superior performance of the YOLO_LP model, which meets the accuracy and speed requirements for abnormal behavior detection in elevator cabins.

Keywords: object detection; abnormal behavior; YOLOv11; feature extraction; context information

收稿日期: 2025-01-17 Received Date: 2025-01-17

* 基金项目:国家自然科学基金青年基金(51805275)、北京市属高校基本科研业务费项目(X21053)、河南省高等学校重点科研项目(23A460020)资助

0 引言

随着我国工业体系的逐步完善以及城市化进程的加快,电梯的使用量正在逐年上升。电梯给人们生活带来便利的同时,也带来了安全隐患。尤其是电梯内的乘客异常行为,不仅可能引发设备故障,还可能对其他乘客的安全和舒适性产生影响。例如,乘客的打架、摔倒、吸烟等行为,都可能导致电梯事故或乘客伤害。因此,电梯乘客的异常行为检测已经成为电梯安全管理中的一个重要课题。

人体异常行为识别的研究起步较早,传统方法主要依赖于基于规则的行为建模和特征提取,Mabrouk 等^[1]提出了一种利用光流信息和时空兴趣点的时空特征来检测视频内暴力行为的方法。近年来,基于深度学习的异常行为识别逐渐成为主流。Ko 等^[2]创新性地提出一种深度卷积框架,能够有效地提取目标特征,同时引入长短时记忆网络(long short term memory, LSTM)完成对动态目标的特征建模,以此识别人体异常行为。Cao 等^[3]提出了一种多人 2D 姿态估计算法,它能更精确地从图像中提取人体骨架信息,从而提高行为识别的准确率。Santoro 等^[4]设计的关系网络(relation networks, RNs)从根本上解决了关系推理的问题,通过推断对象之间的关系和属性来识别别人的交互行为。建中华等^[5]根据人体行为空间复杂及时间差异的特点,通过构建相邻帧时空特征张量和自适应注意力机制来完成时空特征融合,提高了行为识别的精度。张海超等^[6]提出了一种融合注意力的轻量级人体行为识别模型,旨在提高模型的检测速度。基于视频帧聚类^[7]的关键帧提取方法逐渐成为行为识别的热点,刘雨萌等^[8]根据标准化流及时空图卷积的定位网络来学习正常帧的概率分布,然后筛选异常帧输入进网络模型,解决了长视频和多人异常动作识别。目标检测作为一种新兴且高效的技术也被引入到了这一领域。与上述方法相比,目标检测能够处理更复杂人体行为动作和吸烟检测。基于深度学习的目标检测算法主要分为两类^[9]:一类是单阶段算法,常见的有 RetinaNet^[10]、SSD^[11](single shot multibox detector)、EfficientDet^[12]、YOLO^[13-15](you only look once)、DETR^[16](detection transformer)等;另一类是两阶段算法,如 Faster R-CNN^[17](faster region-based convolutional neural network)、Mask R-CNN^[18](mask region-based convolutional neural network)。两阶段方法的主要思路是先生成候选区域,搜索出图片中可能包含目标的位置,然后在对其分类并获得目标的精确位置。两阶段算法的精度较高,但计算量偏大。单阶段算法将检测任务视为一个回归问题,不再需要生成候选区域。王新等^[19]利用 MobileNetV3 网络

和 Slim Neck 模块对 YOLOv5 模型进行轻量化,并改进质心跟踪,提高了跌倒检测的准确性。朱胜豪等^[20]通过引入新型特征提取模块和 Swin Transformer block 有效解决了不同尺度目标的跌倒检测。娄文等^[21]通过替换 YOLOv7 主干网络使模型轻量化,同时加强特征金字塔网络的特征层融合以提高吸烟检测的精度。朱强军等^[22]通过引入压缩和激励网络(squeeze-and-excitation networks, SENet)和损失函数来改进 YOLOv8s 模型,能够有效检测老人的摔倒姿态。尽管目标检测算法对异常行为检测取得了较好的效果,但其在电梯等封闭空间的应用仍然存在一些不足。具体来说,电梯内乘客的行为模式复杂,现有的模型在面对高密度乘客、背景干扰以及不规则的姿态变化时,准确性和实时性之间难以平衡。

针对上述算法在异常行为检测领域的不足,本文提出了一种基于改进 YOLOv11n 的全新网络模型 YOLO_LP(you only look once_lightweight and precision)。为满足复杂背景下的异常行为检测要求,提高模型的检测精度,提出了跨阶段局部 Transformer 模块(cross stage partially transformer module, CSP-PTM),增加了模型对全局特征的提取能力。同时,引入上下文信息融合模块(contextual information fusion module, CIFM),利用权重化的特征重组操作,增强其特征表达能力。然后引入 WIoU(weighted intersection over union)损失函数,进一步提升模型的准确性和泛化能力。最后,使用改进的轻量化检测头(lightweight detection head, LDH),降低模型的参数量,减少计算开销。

综上,通过上述改进,YOLO_LP 模型展现出了显著的性能提升。该模型在处理电梯复杂背景时,有效地减少了对乘客异常行为的漏检情况。同时,在准确性、实时性和轻量化这 3 个关键性能指标之间实现了平衡,形成了一种高效的解决方案,为电梯环境下乘客异常行为的检测任务提供了强有力的支持。

1 YOLOv11 算法原理

YOLOv11 算法相较于前几代 YOLO 系列算法在精度和速度上都有显著提升。YOLOv11 的模型架构主要由骨干网络(backbone)、颈部网络(neck)和检测头(head)三部分组成。

YOLOv11 的骨干网络通过 5 次下采样生成不同尺度的特征图,然后使用 C3k2 模块替换了原来的 C2f 模块,C3k2 模块利用其内部更小的核卷积,在保留图像特征提取能力的同时,提高处理速度,降低计算成本,优化网络中的特征信息。除此之外,引入快速-空间金字塔池化(spatial pyramid pooling-fast, SPPF)模块和通道和位置空间注意力(channel and position spatial attention,

C2PSA) 模块。其中, SPPF 模块可进行多尺度特征提取, 融合特征信息, 旨在提高模型捕获不同大小物体的能力。C2PSA 机制实质上是 C2 机制与多头注意力机制的结合, 它对物体特征有更准确的感知, 能够优化特征图。YOLOv11 的颈部网络采用路径聚合网络 (path aggregation network, PAN) 并集成了 C3k2 单元, 通过自下而上的路径融合多尺度特征并优化特征传递效率, 以此提升模型对不同尺度目标的检测能力和整体性能。

YOLOv11 的检测头采用解耦结构, 包含分类分支和回归分支, 分类分支负责预测目标所属的类别概率, 回归分支则聚焦于确定目标的位置信息。与此同时, 在分类检测头中额外添加了两个深度可分离卷积 (depthwise separable convolution, DSConv) 模块, 有效地降低了模型的参数量和所需的计算量, 从而在一定程度上优化了模型的整体性能。检测头最终输出 3 个特征图的预测结果, 能够保证在检测过程中凭借更精细的细节信息将目标准确检测出来。

2 模型改进

本文主要对电梯轿厢内乘客的异常行为进行目标检测, 包括乘客打架、摔倒和吸烟 3 种行为。由于打架、摔倒行为的目标尺寸和吸烟行为的目标尺寸差距较大, 同时又受电梯内环境光的影响, 很容易导致误检、漏检等问题。鉴于此种情况, 一种可行的方案是采用体积更大的模型, 以期提高识别精度。然而, 这种方法存在弊端, 它会导致计算量和参数成本的显著增加。考虑到模型最终需要在嵌入式设备上运行, 既要保证较高的检测精度, 又要实现模型轻量化。因此, 综合考虑模型的性能与资源消耗, 本文提出一种改进 YOLOv11n 的模型 YOLO_LP, 其网络结构如图 1 所示。

本文提出的 YOLO_LP 模型具体有如下改进: 首先, 采用 CSP-PTM 特征提取组件替换 YOLOv11 骨干网络中的 C3k2 模块。该组件结合了卷积神经网络 (convolutional neural networks, CNN) 和 Transformer 二者的优点, 通过分配输入通道来提高模型的特征提取能力和计算效率。其次, 在颈部网络中引入 CIFM 模块改进特征金字塔网络, 充分利用上下文信息, 进一步提升模型精度。然后, 采用改进的 LDH 检测头代替原有检测头, 降低模型的参数量和计算量, 在保证较少的精度损失的前提下, 减少资源消耗。最后, 加入 WIoU 损失函数, 提升模型的检测精度和收敛速度, 解决目标大小和类别不平衡的问题, 进一步提高模型的泛化能力。

2.1 CSP-PTM 模块

在 YOLOv11 模型中, 对检测精度影响最大的是其主干特征提取网络。为了提高模型对异常行为的检测精

度, 本文引入全新的特征提取组件 CSP-PTM 替换 C3k2 模块来改进 YOLOv11 的主干网络, CSP-PTM 的总体结构如图 2 所示。在跨阶段局部网络架构的基础上, 使用局部 Transformer 模块 (partially transformer module, PTM) 代替传统的 Bottleneck。

在计算机视觉任务中, 由于 CNN 的感受野大小较少, 导致其只能提取局部特征, 而 Transformer 结构却拥有强大的全局特征提取能力。但是, 由于 Transformer 结构的计算复杂度较高, 直接将其应用于所有通道会导致显著的计算开销。为了保证高效提取特征的同时降低计算成本, PTM 模块结合了 CNN 和 Transformer 机制, 其结构如图 3 所示。它将输入特征图分为两部分, 分别由 MHSA_CGLU (multi-head self-attention_convolutional gated linear unit) 和 Bottleneck 处理, 旨在结合二者的优势, 融合局部和全局特征。这种通过对输入通道进行部分分配的处理方式可以优化计算效率和特征提取能力。MHSA_CGLU 模块由传统的 Transformer Block 改进而成, 利用多头自注意力机制 (multi-head self-attention, MHSA), 使模型关注输入数据的不同部分之间的关系, 从而更好地捕捉数据的特征和上下文信息。同时, 采用卷积通道混合器 (convolutional gated linear unit, CGLU)^[23] 替换原有的前馈神经网络 (feed-forward neural network, FFN), 弥合了 GLU (gated linear unit) 和 SENet 机制之间的差距, 使每个 Token 能够基于其最近邻图像特征进行通道关注, 增强其局部建模能力和非线性特征表达能力。

综上, CSP-PTM 模块不仅能够提升模型的检测精度, 同时不会增加模型的参数量和计算量, 综合性能得到极大改善。

2.2 CIFM 模块

在 YOLOv11 的颈部网络中, 对上下文信息的利用较少。本文提出一种上下文信息融合模块, 旨在改进 YOLOv11 中的特征金字塔网络, 进一步提高模型精度。该模块的设计考虑了多尺度特征融合过程中上下文信息的引导和自适应调整, 其结构如图 4 所示。

CIFM 模块是将维度相同的特征图进行通道连接, 然后将它们送入通道注意力模块, 该模块利用卷积和池化操作来生成注意力权重。通过通道注意力机制, 模块能够在特征融合过程中捕捉并利用重要的上下文信息, 从而增强特征表示的有效性, 并有效引导模型学习检测目标的信息, 继而提高模型的检测精度。然后, 通过逐元素乘法应用这些权重来对原始特征进行加权, 并将所得特征添加到另一个分支的原始特征中。这种通过权重化的特征重组操作, 能够增强重要特征, 同时抑制不重要特征, 提升特征图的判别能力。

特征强化过程如式 (1)~(3) 所示。

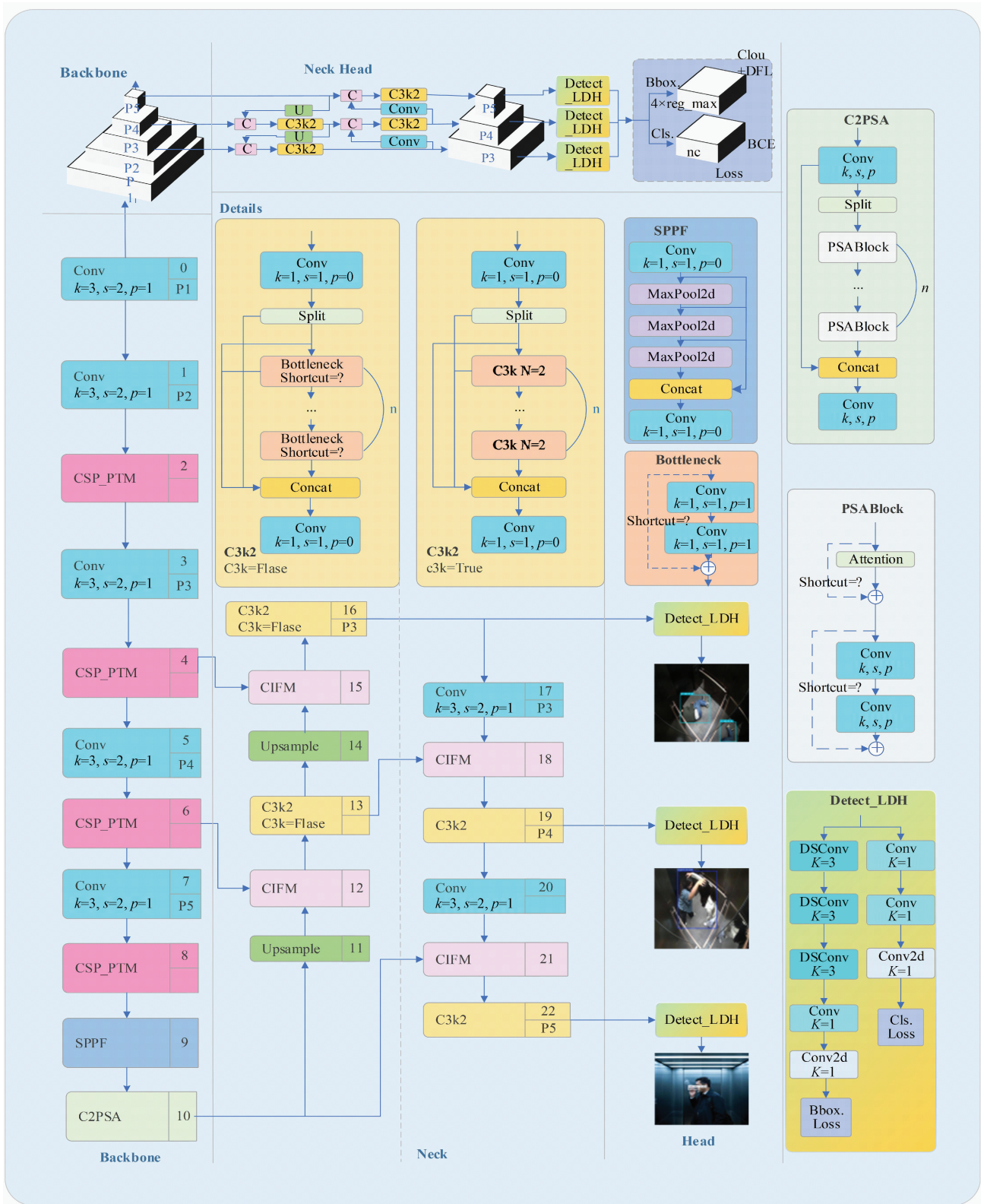


图 1 YOLO_LP 模型结构
Fig. 1 YOLO_LP model structure

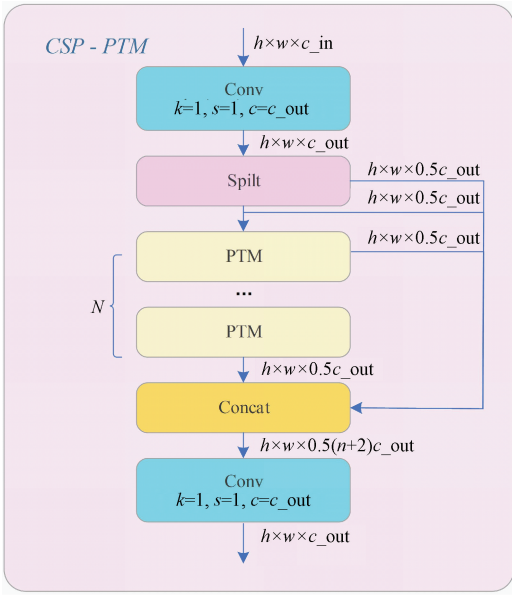


图 2 CSP-PTM 模块结构
Fig. 2 CSP-PTM module structure

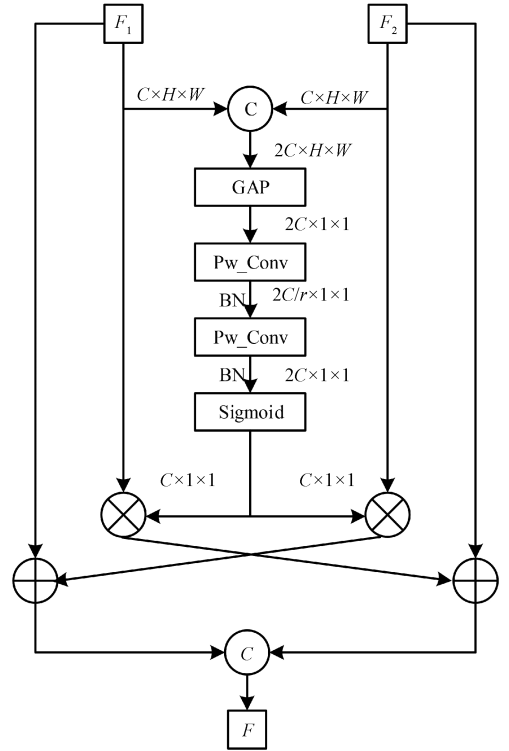


图 4 上下文信息融合模块结构
Fig. 4 Context information fusion module structure

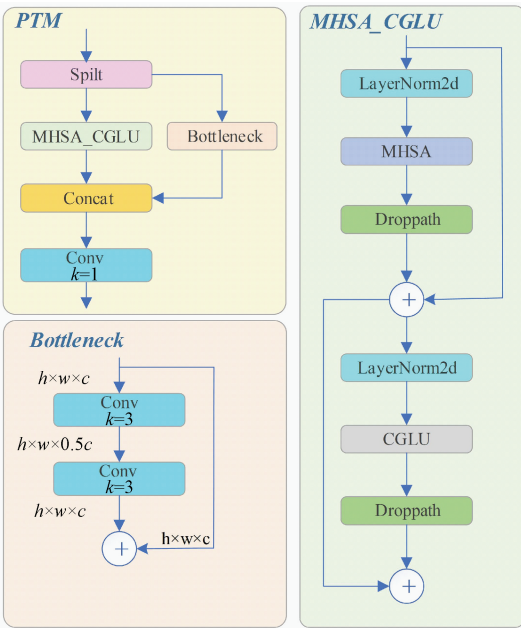


图 3 PTM 模块结构
Fig. 3 PTM module structure

$$\hat{F}_1 = F_1 \oplus (F_2 \otimes \delta(Pw_Conv^n(GAP(C(F_1, F_2)))))) \quad (1)$$

$$\hat{F}_2 = F_2 \oplus (F_1 \otimes \delta(Pw_Conv^n(GAP(C(F_1, F_2)))))) \quad (2)$$

$$F = C(\hat{F}_1, \hat{F}_2) \quad (3)$$

式中: F_1 和 F_2 为输入特征; F 为输出特征; \oplus 表示逐元

素求和; \otimes 表示逐元素乘法; $Pw_Conv^n(\cdot)$ 表示 n 个级联的逐点卷积层; $\delta(\cdot)$ 和 $GAP(\cdot)$ 分别表示 sigmoid 函数和全局平均池化; $C(\cdot)$ 表示通道维度上的串联操作。

实验结果表明,通过对 CIFM 模块的有效设计,提升了电梯内乘客异常行为检测的精度,同时该模块的设计相对简单,不会引入过多的计算开销。

2.3 改进检测头

在 YOLOv11 的检测头中,用于分类和回归任务的卷积层包含了较多的常规卷积操作,在结构设计上并没有充分考虑到轻量化的需求。在分类任务中,原本的卷积结构较为复杂,没有利用更简洁高效的方式来实现类别概率的输出,导致在提取类别相关特征时引入了过多不必要的参数。同样,在回归任务部分,常规的卷积设置使得在预测目标边界框信息(如中心坐标、宽和高)时,也积累了大量的冗余参数,这些参数对于提升检测精度的边际效益并不高,却大大增加了模型的复杂度。针对上述问题,本文设计一种轻量化 LDH 检测头,结构如图 5 所示。

LDH 采用解耦头结构,对 P3、P4、P5 层的特征图分别执行分类(Cls.)和回归(Reg.)操作。由于 YOLOv11 模型的任务的难度相对不高,所以采用两个 1×1 卷积。对于回归任务而言,为了增大感受野并扩充 Reg 分支的任务参数,本文采用了 3×3 DSCConv 作为传统 3×3 卷

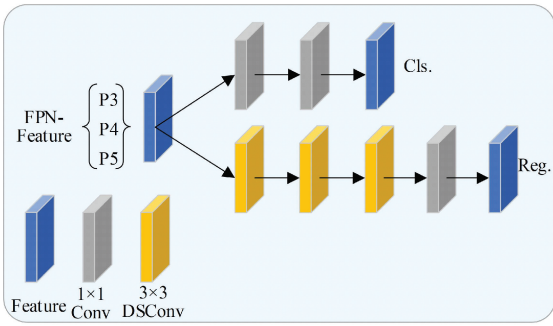


图 5 LDH 检测头结构

Fig. 5 LDH detection header structure

积的替代方案。DSCConv 相对于标准卷积的显著优势在于其独特的卷积分解机制,即将卷积操作拆分为深度卷积和逐点卷积,从而大大减少了参数量。深度卷积之后的特征图数与输入层的通道数相同,并且无法扩展特征图。由于该操作是对输入层中的每个通道独立实施的,因此无法有效地利用同一空间位置上不同通道的特征信息。为解决此问题,逐点卷积得以引入。逐点卷积的操作模式在一定程度上类似于常规卷积操作,其核心功能是借助 1×1 的卷积核对不同通道之间的信息进行融合。它不会改变特征图的空间维度,仅会对通道数产生影响。因此,逐点卷积操作结合了以前的映射在深度方向上加权以生成新的特征图。同时,通过逐点卷积将来自不同通道的信息组合在一起,以保持一定水平的特征提取能力。这种方法进一步降低了模型参数并提高检测速度。

实验结果表明,采用 LDH 检测头可以对网络模型的参数量及计算量产生极为显著的优化效果。

2.4 优化损失函数

在电梯背景下进行人体异常行为检测,很容易受到环境噪声的影响,同时识别人体动作和识别烟头的目标大小有着很明显的差别。YOLOv11 使用传统的 CIoU (complete intersection over union) 来计算边界框的回归损失,所有的样本统一对待,未考虑样本的难易程度;同时,CIoU 损失函数的惩罚因子包括纵横比,如果考虑到实际框和预测框的纵横比相同,但宽度和高度值不同这种特殊情况,那么惩罚项将无法反映实际框与预测框之间的真实差异。鉴于此,本文使用 WIoU 作为 YOLOv11 的回归损失函数,旨在解决异常行为检测中目标类别和大小不平衡的问题,进一步提升检测精度和收敛速度,提升模型的泛化能力。

WIoU v3 引入了离群度 β 来表征锚框的质量,其表达式如式(4)所示。通常情况下,离群度较小意味着锚框具有较高的质量,此时会为其分配较小的梯度增益,以使边界框回归操作侧重于普通质量的锚框。而对于离群度较大的低质量锚框,将为其分配相对较小的梯度增益,

以此防止此类锚框产生有害梯度,进而保障模型的整体效果,确保模型性能的稳定性和可靠性。

$$\beta = \frac{L_{IoU}^*}{L_{IoU}} \in [0, +\infty) \quad (4)$$

式中: $\overline{L_{IoU}}$ 为动量 m 的滑动平均值。鉴于 $\overline{L_{IoU}}$ 具备动态特性,相应地,锚框质量划分标准亦呈动态变化,这使得 WIoU v3 能够实时制定出最契合当下状况的梯度增益分配策略。

利用离群度 β 构造一个非单调聚焦系数 γ 并将其应用于 WIoU v1。WIoU v3 定义如式(5)~(9)所示。

$$L_{WIoUv3} = \gamma L_{WIoUv1} \quad (5)$$

$$\gamma = \frac{\beta}{\delta \alpha^{\beta-\delta}} \quad (6)$$

$$L_{WIoUv1} = R_{WIoU} L_{IoU} \quad (7)$$

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^2}\right) \quad (8)$$

$$L_{IoU} = 1 - IoU \quad (9)$$

式中: α 和 δ 为超参数; W_g 和 H_g 为最小包围框的尺寸; (x, y) 和 (x_{gt}, y_{gt}) 分别为锚框和目标框中心点的坐标。

WIoU v3 融合了 EIoU (expected intersection over union) 和 SIoU (scylla intersection over union) 的优势,契合损失函数的设计理念,同时使用动态非单调机制评估锚框质量,提升模型的定位能力。对于异常行为识别的检测任务,目标大小不同增加了检测难度, WIoU v3 可以动态优化小目标的损失权重,提高模型的检测性能。

3 实验结果与分析

3.1 实验数据集

由于电梯轿厢背景下的人体异常行为的公开数据较少,因此本文使用的 LC 数据集部分选自 Roboflow、Kaggle 等数据集网站,其余部分通过采集实验室人员在电梯中打架、摔倒及抽烟行为构建。图像采集设备选用海康 MV-CS200-10GM 工业相机,具有 $2\,000 \times 10^4$ pixels,分辨率为 $5\,472 \times 3\,648$ 。经过人工筛选低质量图像后,得到高分辨率彩色图像,使用 LabelImg 标注工具进行标注。

为进一步提高模型的泛化能力,本文采用 Albumentations 库中的一系列增强策略对采集的图像进行数据增强。数据增强方法及关键参数如表 1 所示。

该数据集共分为 3 类异常行为,打架、摔倒、吸烟,共计 7 491 个数据图像。同时数据集的异常行为样本覆盖了多类人群,其中打架场景中青少年约占 20%,摔倒案例中老年乘客(60 岁以上)占 35% 以上,吸烟行为包含 20% 左右的孕妇群体,具备跨人群的特征提取能力。实验按

表 1 扩充数据集的数据增强方法
Table 1 Data augmentation methods for enriching datasets

增强方法	作用	关键参数
水平翻转	左右翻转图像	$p = 1.0$
垂直翻转	上下翻转图像	$p = 1.0$
随机平移、缩放、旋转	随机仿射变换	$shift_limit = 0.0625, scale_limit = 0.1, rotate_limit = 15$
亮度调整	亮度增强	$brightness_limit = 0.2$
对比度调整	对比度增强	$contrast_limit = 0.2$
高斯噪声	添加随机噪声	$var_limit = (10.0, 50.0)$
随机 90° 旋转	旋转 90°/180°/270°	$p = 1.0$

注: p 表示变换概率; $shift_limit$ 表示图像平移范围; $scale_limit$ 表示图像缩放范围; $rotate_limit$ 表示图像旋转变换范围; $brightness_limit$ 表示图片亮度随机调整范围; $contrast_limit$ 表示图片对比度随机调整范围; var_limit 表示添加噪声的方差范围

7 : 2 : 1 比例划分为训练集、验证集、测试集, 其中训练集图片 5 243 张, 验证集图片 1 498 张, 测试集图片 750 张。训练集用于让模型学习数据规律以构建预测能力,

验证集辅助在训练中调整超参数及模型结构评估性能, 测试集则用于最终衡量模型对新数据的泛化能力。电梯轿厢内乘客的 3 种异常行为如图 6(a)~(c) 所示。

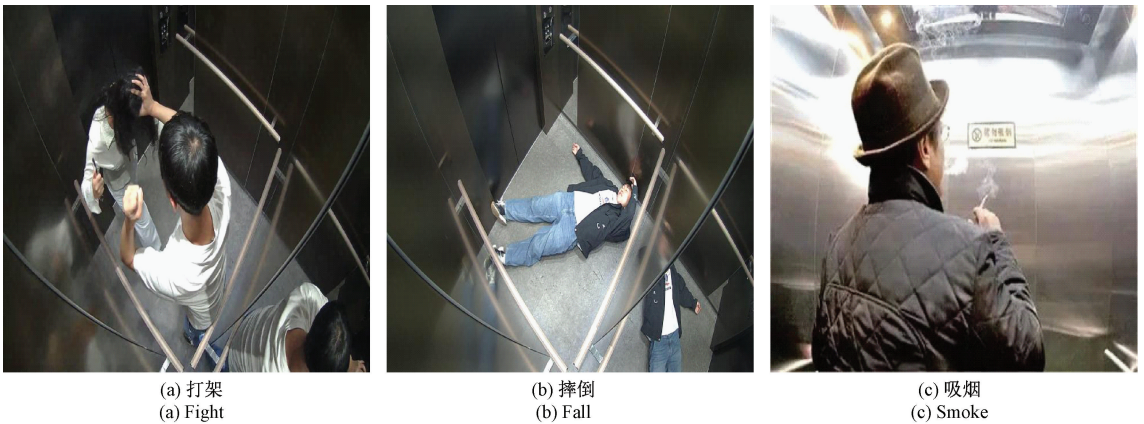


图 6 LC 数据集样图

Fig. 6 Sample images of LC dataset

3.2 实验环境与参数配置

本文实验平台使用 Ubuntu22.04 操作系统, 搭载 16 vCPU Intel® Xeon® Platinum 8352V 处理器, 主频为 2.10 GHz。显卡采用 Nvidia GeForce RTX 4090, 显存容量为 24 GB。PyTorch 框架采用 2.3.0 版本, CUDA 版本为 12.1, Python 版本为 3.12。网络未使用预训练权重, 同时所有实验使用相同环境及统一的超参数, 图像大小设置为 640×640, epoch 设置为 300, Batch size 设置为 16。

3.3 模型评价指标

本文实验采用了目标检测领域常用的评估指标, 包括精度 (precision, P)、召回率 (recall, R)、平均精度值 (mAP), 以此来评价模型的性能表现。与此同时, 对于模型的评估, 模型的大小和计算开销是不容忽视的重要方面。为了对这两个重要维度进行量化评估, 本文采用参数数量和计算量作为衡量指标, 它们能够清晰地反映出模型的规模大小以及在计算资源上的消耗情况, 为准确评价模型提供关键的数据支撑。

1) 精度指在所有检测结果的样本集合中, 正确检测到的样本数量占检测样本总数的比例。其计算公式如式(10)所示。

$$P = \frac{TP}{TP + FP} \quad (10)$$

式中: TP 为真正例 (正确检测的目标数); FP 为假正例 (错误检测的目标数)。

2) 召回率定义为在所有真实目标集合中, 被所构建的模型正确检测出的目标的数量占真实目标总数的比例。其计算公式如式(11)所示。

$$R = \frac{TP}{TP + FN} \quad (11)$$

式中: FN 为假反例 (真实存在但未被检测到的目标数)。

3) 平均精度均值中 $mAP@0.5$ 表示在 IoU 阈值设定为 0.5 时, 所有类别的平均检测精度, $mAP@0.5 : 0.95$ 指在 IoU 阈值从 0.5 开始, 以 0.05 为间隔逐步递增至 0.95 的一系列阈值下, 分别计算每个阈值所对应的平均

精度,随后对不同阈值下计算得到的平均精度进行算术平均运算得到的最终结果。 mAP 的数值越大,该模型的检测性能越优。

4) 参数量通常反映模型的大小和复杂度。一般来说,参数量越大,模型理论上能够拟合数据的能力就越强,可以学习到更复杂的模式和规律。但同时,也导致模型更容易过拟合,并且在存储和计算时需要消耗更多的资源。

5) 计算量主要用来衡量模型在进行前向传播时所需执行的计算操作的数量,通常用 GFLOPs (giga floating-point operations per second) 来表示。计算量体现了模型运行时的计算成本。计算量越大,意味着模型在进行推理或者训练时需要花费更多的计算资源和时间。

3.4 消融实验

本文模型结构的改进是基于 YOLOv11n 模型,为全面验证各改进模块对模型性能的提升效果,在 LC 数据集上对本文提出的改进模型进行消融实验。通过对系统进行模块划分,将各项改进方法逐步集成,以实现每个模块的贡献进行量化评估,从而清晰且精确地呈现出它们各自对于检测精度和效率的贡献情况。首先将骨干网络中的 C3k2 模块替换为 CSP-PTM 模块;然后在颈部网络中引入 CIFM 模块;接着将原有检测头改进为 LDH 轻量化检测头;最后引入 WIoU 损失函数替换原来的 CIoU 损失函数。消融实验结果如表 2 所示, Baseline 为 YOLOv11n 模型。

实验结果表明,各个改进模块在提升模型对异常行为检测的综合性能上都有显著作用。首先,使用 CSP-

PTM 替换原来的 C3k2 模块后,精度提升了 1.6%,召回率提升 1.5%, $mAP@0.5$ 提高 1.8%, $mAP@0.5:0.95$ 提升 1.3%。这说明 CSP-PTM 这一混合结构不仅保留了对局部特征的提取能力,也提高了对全局特征的提取效率。同时还可以看出在引入 CSP-PTM 模块以后,模型的参数量和计算量均略有减少,在保证高效特征提取的同时降低了计算成本。然后引入 CIFM 模块来改进特征金字塔网络,实验结果显示, $mAP@0.5$ 增加了 1.2%, $mAP@0.5:0.95$ 增加了 0.9%。表明 CIFM 模块进一步加强了颈部网络在特征融合过程中对上下文信息的引导和利用,有效提升了模型的检测精度,同时并未引入过多的计算开销。

进一步地,采用改进的 LDH 轻量化检测头代替原有检测头,根据实验结果,其对模型的检测精度影响甚微,但参数量和计算量分别降低了 11% 和 17%。这表明 LDH 检测头合理减少了常规卷积所积累的冗余参数,降低了模型复杂度。在资源有限的环境中完成异常行为检测任务时,这一点尤为关键。接着,引入 WIoU 损失函数解决目标大小和类别不平衡等问题,实验结果显示, $mAP@0.5$ 增加了 1.3%, $mAP@0.5:0.95$ 增加了 0.7%。

最后,综合 4 种改进方法进行实验,结果表明,改进模型的精度高达 90.4%,较原始模型提高了 3.5%。此外,在召回率, $mAP@0.5$ 和 $mAP@0.5:0.95$ 的指标上分别提高了 3%、2.9%、2.1%。同时改进模型的参数量降低了 10%,计算量降低了 17%。综合来看,本文改进的模型在检测精度和检测速度方面都有显著提升,综合性能更好,验证了对模型改进的有效性。

表 2 消融实验结果

Table 2 Ablation experiment results

Baseline	CSP-PTM	CIFM	LDH	WIoU	Precision	Recall	$mAP@0.5$	$mAP@0.5:0.95$	参数量/ $(\times 10^6)$	计算量/GFLOPs
√					0.869	0.823	0.863	0.528	2.58	6.3
√	√				0.885	0.838	0.881	0.541	2.46	6.2
√		√			0.878	0.831	0.875	0.537	2.74	6.5
√			√		0.858	0.816	0.856	0.521	2.28	5.2
√				√	0.883	0.841	0.876	0.535	2.58	6.3
√	√	√			0.892	0.844	0.887	0.545	2.61	6.3
√	√	√	√		0.886	0.836	0.882	0.543	2.31	5.2
√	√	√	√	√	0.904	0.853	0.892	0.549	2.31	5.2

3.5 对比实验

为了验证本文的改进模型在电梯背景下对异常行为检测有着优异的性能,选用目标检测领域常见的几种模型(包括 Faster R-CNN、SSD、YOLOv5s、YOLOv8s、YOLOv10n、YOLOv11n、YOLOv12n、YOLOv12s、RT-DETR-R18、RT-DETR-R34)进行对比实验,实验结果如表 3 所示。由表 3 可知,改进后的 YOLO_LP 模型精度达到了 90.4%,相较于基线网络 YOLOv11n 提升了 3.5%,在

$mAP@0.5$ 和 $mAP@0.5:0.95$ 指标上均优于其他对比模型,从资源消耗方面对比,YOLO_LP 模型的参数量和计算量仅为 2.31×10^6 和 5.2 GFLOPs,较其他模型有着明显优势。表中基于 Transformer 架构的 RT-DETR 模型虽然在检测精度上较接近 YOLO_LP 模型,但其庞大的参数量和计算量将导致显著的计算开销。综上,改进的 YOLO_LP 模型能够有效平衡模型的精度、参数量和计算量,拥有更强的综合性能,更加符合实际应用需求。

表 3 不同算法对比实验结果
Table 3 Comparison experiment results of different algorithms

模型	Precision	Recall	mAP@0.5	mAP@0.5:0.95	参数量/($\times 10^6$)	计算量/GFLOPs
Faster R-CNN	0.825	0.779	0.811	0.476	63.2	370
SSD	0.802	0.758	0.782	0.436	12.3	63.2
YOLOv5s	0.856	0.801	0.842	0.503	7.0	15.8
YOLOv8s	0.875	0.833	0.869	0.529	11.1	28.4
YOLOv10n	0.832	0.774	0.817	0.483	2.7	6.7
YOLOv11n	0.869	0.823	0.863	0.528	2.58	6.3
YOLOv12n	0.871	0.820	0.859	0.517	2.6	6.5
YOLOv12s	0.894	0.831	0.877	0.535	9.3	21.4
RT-DETR-R18	0.891	0.837	0.885	0.533	20.0	57.3
RT-DETR-R34	0.897	0.841	0.887	0.541	30.0	92.6
YOLO_LP	0.904	0.853	0.892	0.549	2.31	5.2

3.6 检测结果可视化对比

为了展示不同模型在 LC 数据集中不同场景下对乘客异常行为的检测效果。选取正常环境、昏暗环境、噪声

环境以及遮挡情景下共 4 张图片,分别采用 YOLOv11n、RT-DETR-R18、YOLOv12s 及改进后的 YOLO_LP 模型进行检测,检测结果如图 7 所示。

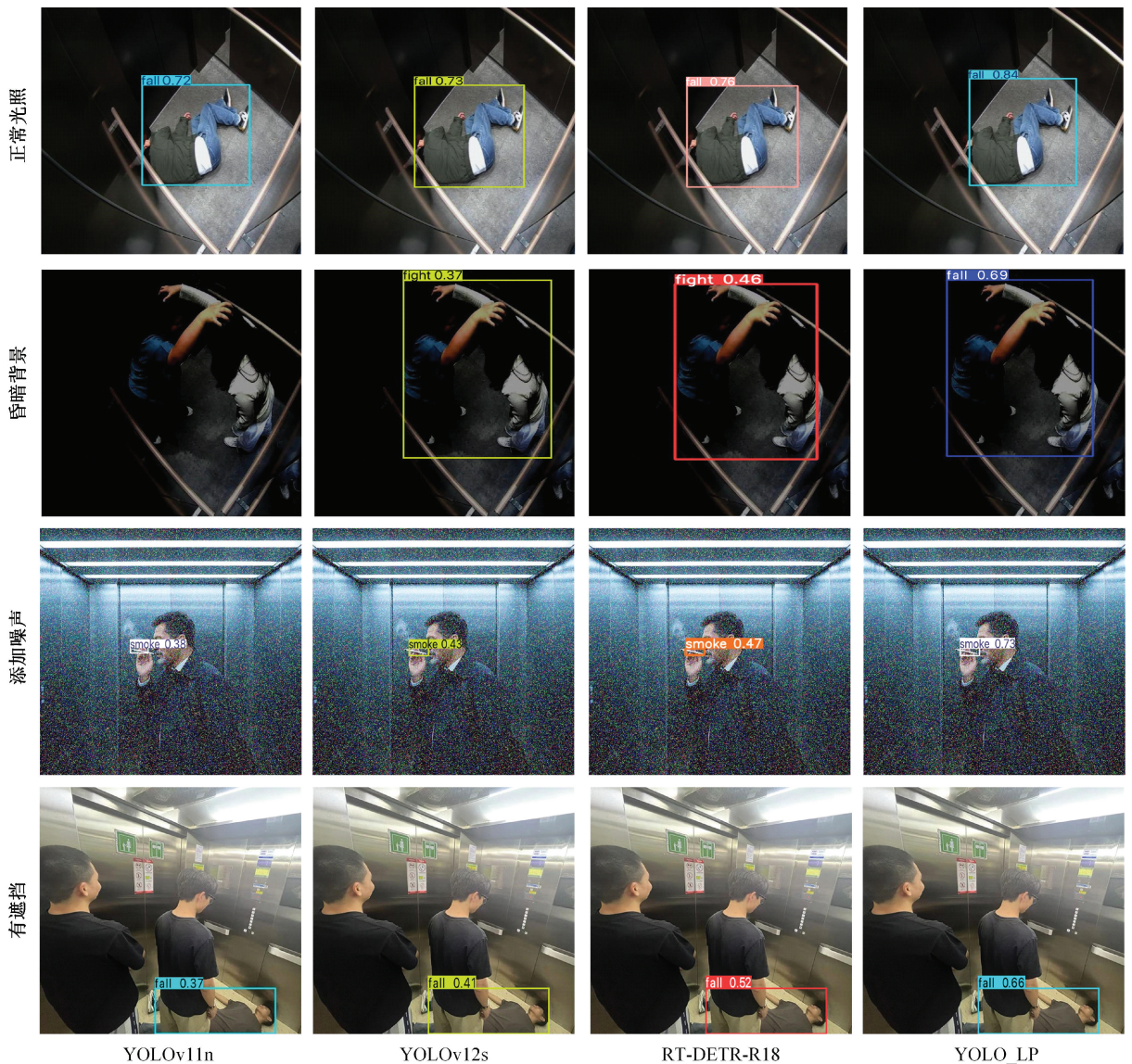


图 7 检测结果可视化

Fig. 7 Visualization of detection results

从图 7 可以看出,在正常环境下,4 种模型均能够检测到乘客的摔倒行为,但 YOLO_LP 模型的置信度更高;而对于光线不足甚至昏暗的场景,YOLOv11n 模型会造成漏检,RT-DETR-R18 及 YOLOv12s 模型的得分也明显低于 YOLO_LP 模型;针对识别添加噪声后的吸烟行为,由于识别目标较小,YOLOv11n、YOLOv12s、RT-DETR-R18 的检测结果相比正常环境均有明显下降,而 YOLO_LP 依然能够保持较高的检测精度;对于遮挡场景下的检测,YOLO_LP 同样表现出更高的精度。综上,改进后的 YOLO_LP 模型能适应各种复杂的环境,同时对小目标的检测也更具鲁棒性。

4 结 论

为解决电梯轿厢内乘客异常行为检测中目标尺度差异大、环境干扰强、计算资源受限等问题,提出了一种基于改进 YOLOv11n 的高效轻量化模型 YOLO_LP。该模型融合了 CSP-PTM 结构、CIFM 模块、WiIoU 损失函数和 LDH 检测头,有效结合了局部细节感知与全局上下文建模能力,显著提升多尺度目标的表征效果。通过自适应特征增强机制抑制背景干扰,增强了在复杂场景下的判别力。基于样本质量评估的优化策略,有效解决了目标尺度与类别不平衡问题。同时,该设计显著降低了模型的计算负载。与其他先进的算法相比,YOLO_LP 算法具有较优的目标识别能力,有效提升了复杂背景下乘客异常行为的检测精度。此外,YOLO_LP 算法参数量和计算量的降低,使其具有更显著的轻量化优势,能够高效地部署于嵌入式设备,在满足实时检测要求的同时保持高精度。尽管本文提出的 YOLO_LP 模型在检测精度和速度方面取得了显著进展,但仍存在优化空间。未来将进一步探索不同网络架构之间的深度融合,增强模型对微小动作的捕捉能力;利用多场景数据训练模型,解决数据稀缺性问题,提升模型泛化能力;研究模型自适应压缩技术,实现资源受限环境下的弹性部署。

参考文献

- [1] MABROUK A B, ZAGROUBA E. Spatio-temporal feature using optical flow based distribution for violence detection [J]. Pattern Recognition Letters, 2017, 92(4): 62-67.
- [2] KO K E, SIM K B. Deep convolutional framework for abnormal behavior detection in a smart surveillance system [J]. Engineering Applications of Artificial Intelligence, 2018, 67(9): 226-234.
- [3] CAO Z, SIMON T, WEI S E, et al. Realtime multi-person 2d pose estimation using part affinity fields [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7291-7299.
- [4] SANTORO A, RAPOSO D, BARRETT D G, et al. A simple neural network module for relational reasoning [J]. Advances in Neural Information Processing Systems, 2017, 30(12): 4967-4976.
- [5] 建中华, 南静, 刘鑫, 等. 基于时空张量融合的人体骨架行为自适应识别方法 [J]. 仪器仪表学报, 2023, 44(6): 74-85.
- [5] JIAN ZH H, NAN J, LIU X, et al. Adaptive recognition method of human skeleton action with spatial-temporal tensor fusion [J]. Chinese Journal of Scientific Instrument, 2023, 44(6): 74-85.
- [6] 张海超, 张闯. 融合注意力的轻量级行为识别网络研究 [J]. 电子测量与仪器学报, 2022, 36(5): 173-179.
- [6] ZHANG H CH, ZHANG CH. Research on lightweight action recognition network integrating attention [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(5): 173-179.
- [7] 陶永辉, 王勇. 基于改进的 K-means 风电异常数据检测 [J]. 国外电子测量技术, 2023, 42(4): 141-148.
- [7] TAO Y H, WANG Y. Abnormal data detection of wind turbine based on improved K-means [J]. Foreign Electronic Measurement Technology, 2023, 42(4): 141-148.
- [8] 刘雨萌, 桑海峰. 基于关键帧定位的人体异常行为识别 [J]. 电子测量与仪器学报, 2024, 38(3): 104-111.
- [8] LIU Y M, SANG H F. Human abnormal behavior recognition based on keyframes localization [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(3): 104-111.
- [9] 徐彦威, 李军, 董元方, 等. YOLO 系列目标检测算法综述 [J]. 计算机科学与探索, 2024, 18(9): 2221-2238.
- [9] XU Y W, LI J, DONG Y F, et al. Survey of development of YOLO object detection algorithms [J]. Journal of Frontiers of Computer Science and Technology, 2024, 18(9): 2221-2238.
- [10] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2980-2988.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14. Berlin: Springer, 2016: 21-37.

- [12] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [13] 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.
- [14] REDMON J, FARHADI A. YOLOv3: An incremental improvement [C]. Computer Vision and Pattern Recognition. Berlin: Springer, 2018, 1804: 1-6.
- [15] 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.
- [16] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision. Berlin: Springer, 2020: 213-229.
- [17] REN SH 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, 2016, 39(6): 1137-1149.
- [18] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [19] 王新, 杨秀梅. 基于 YOLOv5s 和改进质心跟踪的人员跌倒检测 [J]. 电子测量技术, 2023, 46(24): 172-178.
WANG X, YANG X M. Personnel fall detection based on YOLOv5s and improved centroid tracking [J]. Electronic Measurement Technology, 2023, 46(24): 172-178.
- [20] 朱胜豪, 钱承山, 阚希. 改进 YOLOv5 的高精度跌倒检测算法 [J]. 计算机工程与应用, 2024, 60(11): 105-114.
ZHU SH H, QIAN CH SH, KAN X. High-precision fall detection algorithm with improved YOLOv5 [J]. Computer Engineering and Applications, 2024, 60(11): 105-114.
- [21] 娄文, 郭杜杜, 张杰, 等. 基于 YOLOv7 的驾驶人使用手机与抽烟行为识别方法 [J]. 电子测量技术, 2023, 46(21): 123-131.
LOU W, GUO D D, ZHANG J, et al. Identification method of mobile phone use and smoking behavior of drivers based on YOLOv7 [J]. Electronic Measurement Technology, 2023, 46(21): 123-131.

- [22] 朱强军, 程靓靓, 汪慧兰, 等. 改进的 YOLOv8s 摔倒检测算法研究 [J]. 电子测量技术, 2024, 47(19): 190-196.
ZHU Q J, CHENG L L, WANG H L, et al. Research on improved YOLOv8s fall detection algorithm [J]. Electronic Measurement Technology, 2024, 47(19): 190-196.
- [23] SHI D. TransNeXt: Robust foveal visual perception for vision Transformers [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 17773-17783.

作者简介



刘伟杰, 2022 年于成都理工大学获得学士学位, 现为北京建筑大学硕士研究生, 主要研究方向为嵌入式智能安防与在线监测产品开发。

E-mail: 17549634474@163.com

Liu Weijie received his B. Sc. degree from Chengdu University of Technology in 2022. Now he is a M. Sc. candidate at Beijing University of Civil Engineering and Architecture. His main research interests include the development of embedded intelligent security and online monitoring products.



李志星 (通信作者), 2008 年于河北工程大学获得学士学位, 2011 年于内蒙古科技大学获得硕士学位, 2018 年于北京科技大学获得博士学位, 现为北京建筑大学副教授、硕士生导师, 主要研究方向为信号处理和机电装备故障大数据分析 with 智能诊断。

E-mail: onyxlxz@126.com

Li Zhixing (Corresponding author) received his B. Sc. degree from Hebei University of Engineering in 2008, M. Sc. degree from Inner Mongolia University in 2011 and Ph. D. degree from University of Science and Technology Beijing in 2018 respectively. Now he is an associate professor in Beijing University of Civil Engineering and Architecture. His main research interests include signal processing and electromechanical equipment failure big data analysis and intelligent diagnosis.



庞玉东, 2023 年于山东科技大学获得学士学位, 现为北京建筑大学硕士研究生, 主要研究方向为特种设备图像识别与大数据分析。

E-mail: sdzppyd@163.com

Pang Yudong received his B. Sc. degree from Shandong University of Science and Technology in 2023. Now he is a M. Sc. candidate at Beijing University of Civil Engineering and Architecture. His main research interest includes image recognition of special equipment and big data analysis.