

DOI:10.19651/j.cnki.emt.2519102

改进 YOLOv10 的复杂场景人体跌倒检测方法^{*}

郭莉 张雪松 李萌萌 金花

(大连交通大学轨道智能工程学院 大连 116028)

摘要: 针对复杂场景中高动态人体运动引发的跌倒特征判别性下降、小目标难以识别、关键部位遮挡等问题,提出了一种基于改进 YOLOv10 的人体跌倒检测算法 ICI-YOLO。通过引入上下文锚点注意力替换骨干网络中的部分自注意力机制,实现了全局上下文依赖与细粒度空间融合表征;融合了迭代注意力机制对骨干网络 C2f 进行重构,强化关键区域语义表达能力;并提出融合交互卷积和跨尺度特征融合的交互式特征融合网络,提升了模型多尺度特征融合能力。实验表明,改进后的 ICI-YOLO 模型在自制人体跌倒行为检测数据集 FALL 上召回率和 mAP@0.5 分别提升了 4.3% 和 2.2%,在公开数据集 DiverseFALL10500 上准确率和 mAP@0.5:0.95 分别提升了 2.0% 和 1.5%,且在与主流实时检测算法的对比中展现出更优的检测性能。

关键词: 人体跌倒检测;YOLOv10;上下文锚点注意力;迭代注意力;交互卷积

中图分类号: TP391.4; TN919.8 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Improved YOLOv10 method for human fall detection in cluttered scenes

Guo Li Zhang Xuesong Li Mengmeng Jin Hua

(School of Railway Intelligent Engineering, Dalian Jiaotong University, Dalian 116028, China)

Abstract: To address the issues of degraded feature discriminability, difficulty in recognizing small-scale targets, and occlusion of key body parts caused by high-dynamic human motion in cluttered scenes, an improved fall detection algorithm ICI-YOLO based on YOLOv10 is proposed. The contextual attention aggregation replaces the partial self-attention, achieving global contextual dependency and fine-grained spatial fusion representation. The iterative attentional feature fusion mechanism is incorporated to restructure the C2f of backbone, strengthening semantic representation capabilities for critical regions. An interactive feature fusion network integrating interactive convolution block and cross-scale convolutional feature fusion module is proposed, to improve multi-scale feature fusion capability. Experimental results demonstrate that the enhanced ICI-YOLO model achieves performance gains of 4.3% in recall and 2.2% in mAP@0.5 on the self-constructed human fall behavior detection dataset FALL, while attaining improvements of 2.0% in precision and 1.5% in mAP@0.5:0.95 on the public dataset DiverseFALL10500. Compared with mainstream real-time detection algorithms, the proposed method exhibits superior detection performance.

Keywords: human fall detection; YOLOv10; contextual attention aggregation; iterative attentional feature fusion; interactive convolution block

0 引言

在全球范围内,跌倒已成为威胁老年人健康的重要隐患。根据美国疾病控制与预防中心发布的数据,每年有超过 1 400 万名老年人出现跌倒,其中 60 岁及以上人群占比最高^[1]。跌倒已成为老年人因伤致死的首要因素,若能及时获得救治可显著降低死亡风险。随着我国人口老龄化进程的加快,构建高效且高精度的跌倒检测与实时响应技术,

对提升老年人跌倒事件的应急救治能力具有重要的研究价值与现实意义。

当前跌倒检测算法主要分为两大类:基于非计算机视觉的方法与基于计算机视觉的方法。其中非计算机视觉的跌倒检测算法多依赖可穿戴传感器进行实现。孙巍伟等^[2]提出一种基于多传感信息融合的跌倒监测系统,通过处理器融合智能手机传感器、压力薄膜传感器实现跌倒监测预警。此类检测方法依赖用户长期正确佩戴设备,且传感器

收稿日期:2025-06-14

^{*} 基金项目:国家自然科学基金(62276042)、辽宁省教育厅科学研究项目(LJKZ0486)资助

佩戴位置和方式的不当会降低检测结果的稳定性与准确性。基于视觉的人体跌倒检测算法具有检测速度快、精度较高,并且不依赖于穿戴设备的特点。Qi 等^[3]针对 YOLOv5 在跌倒检测中存在的小目标漏检、锚框不匹配和模糊标签问题,提出了融合锚框聚类优化、反馈融合机制以及最优锚框分配策略的改进方法。改进后的模型准确率达到 93.48%,较基线模型提升了 2.88%。然而其在复杂室外场景中的鲁棒性表现仍有优化空间。Hwang 等^[4]在 YOLOv9 中集成全局注意力机制和 Swin Transformer 模块,通过增强模型对全局空间依赖关系的建模能力优化了人体跌倒检测性能,假阴性样本数量从 194 降至 134。但全局注意力机制的引入增加了计算复杂度,限制了模型在资源受限设备上的部署潜力。Hu 等^[5]融合了图像增强技术并引入了卷积块注意力与坐标注意力机制,显著强化了特征的空间和通道表达能力。与基线模型相比,改进后的模型 mAP 提高了 14.5%,达到 92.9%。然而该改进模型在检测速度上仍有不足,需探索更轻量级的注意力机制,以平衡检测准确率和速度。Zhao 等^[6]引入 SDI 注意力模块以增强特征提取能力,使用 GSConv 和 VoV-GSCSP 模块减轻计算负担,并采用 DBB 模块提升多层特征融合效果。实验结果表明,该模型在提高 mAP 的同时,减少了 3.5% 的参数量和 5.4% 的计算复杂度。未来可通过丰富数据集及优化模型结构,提升模型在复杂场景下的表现。龙艺文等^[7]针对 YOLOX 模型,在颈部层引入 Ghost 模块以压缩参数量,融合坐标注意力机制增强特征提取能力,并利用知识蒸馏技术通过 AuxHead 提升检测头性能。改进后的模型在 mAP@0.5 上提升了 4.6%,模型大小减少了 25.4%,但推理速度略有下降。

尽管现有研究在一定程度上提升了跌倒检测性能,但在复杂场景中仍面临挑战。跌倒过程中的剧烈姿态变化及背景干扰易导致特征判别性下降、小尺度目标识别困难及关键部位遮挡,进而影响模型判别能力,造成误判与漏检,降低系统的准确性与鲁棒性。为应对上述问题,并满足人体跌倒检测对实时性和响应速度的高要求,本文提出了一种基于改进 YOLOv10^[8] 的人体跌倒检测模型 ICI-YOLO,以提升其在复杂场景下的检测性能。主要研究内容包括:

1) 为提升模型在跌倒姿态变化剧烈等复杂场景下的特征提取能力,本文以上下文锚点注意力 (contextual attention aggregation, CAA) 替换基线模型骨干网络中的部分自注意力 (partial self-attention, PSA)。CAA 通过方向分解的大核卷积与通道独立的空间注意力机制,实现对多尺度局部细节与全局上下文特征的高效融合,有效增强模型对关键区域的感知能力与特征表达的稳定性。

2) 对骨干网络中 C2f 模块进行结构性改进,引入迭代注意力特征融合模块 (iterative attentional feature fusion, IAFF),构建 C2f-IAFF 特征聚合模块。通过协同多尺度通

道注意力引导机制与迭代优化策略,实现对语义关键区域的精准聚焦与冗余背景噪声的有效抑制。

3) 提出了交互式特征融合网络 (interactive feature fusion network, IFFN),通过引入交互卷积 (interactive convolution block, ICB),并集成跨尺度特征融合模块 (cross-scale convolutional feature fusion module, CCFM) 重构颈部网络,以统一通道维度、抑制冗余干扰,提升模型对小目标与复杂结构的感知性能。

4) 构建了多场景行人跌倒行为检测数据集 FALL (<https://gitee.com/two-more-grapefruits/fall-dataset>),并通过在该自制数据集 FALL 和公开数据集 DiverseFALL10500 上进行消融实验和对比实验,验证改进方法的有效性及泛化性。

1 改进 YOLOv10 的 ICI-YOLO 模型

本文提出一种改进 YOLOv10s 的高精度人体跌倒检测模型 ICI-YOLO,如图 1 所示。首先在主干网络中,引入上下文锚点注意力 CAA 替换部分自注意力 PSA,增强了全局依赖感知与局部细节捕捉能力。其次,通过融合迭代注意力特征 IAFF 对骨干网络的 C2f 模块进行重构,提升人体姿态在复杂遮挡场景下的结构感知与语义重建性能。最后构建交互式特征融合网络 IFFN,提升了高低层特征的协同融合效能。

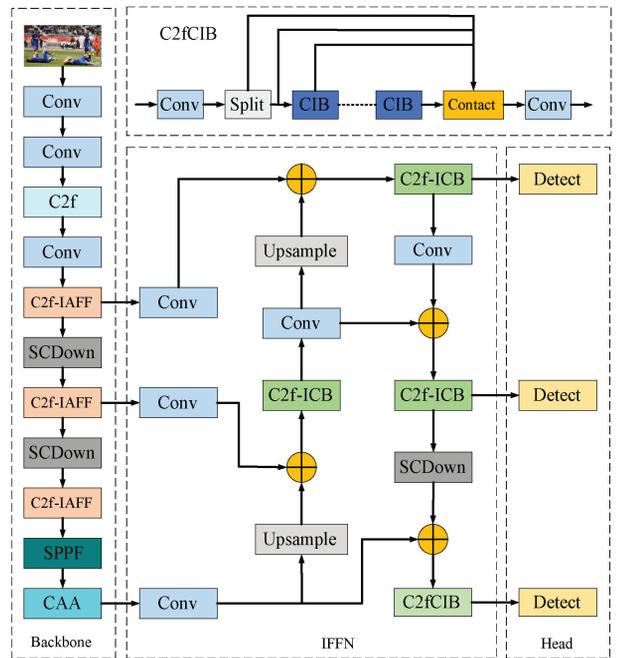


图 1 ICI-YOLO 网络结构

Fig. 1 ICI-YOLO network structure

1.1 CAA 注意力模块

跌倒过程中由于快速下落、翻滚等动态变化引发的跌倒特征判别性下降与局部结构信息损失,易导致特征提取

过程产生表征偏差。尽管 YOLOv10 中引入的 PSA 注意力机制在表征能力与计算效率之间实现了一定的平衡,但其依赖局部信息提取与部分通道处理策略,限制了特征的全局表达能力,导致关键跌倒动作的识别稳定性不足。对此本文采用上下文锚点注意力 CAA^[9] 替代 PSA 自注意力模块,通过方向分解的大核卷积结构联合通道自适应调制机制,实现对长距离依赖与局部空间结构的高效感知,从而有效缓解局部特征退化问题,提升模型在复杂场景中的上下文感知能力与鲁棒性。

在传统卷积过程中,由于感受野的限制,网络难以充分理解不同尺度的空间结构,而 CAA 模块通过多尺度卷积

操作弥补了这一不足。CAA 结构如图 2 所示,首先利用全局平均池化获取输入特征的粗略上下文表达,随后,依次采用水平方向和垂直方向的非对称深度可分离卷积,分别捕捉广域的横向上下文信息与细粒度的纵向局部结构,从而有效增强空间依赖的表达能力。此外,串联的卷积结构设计进一步降低了参数量与计算复杂度。最终,模块通过 Conv1×1 和 Sigmoid 激活生成通道独立的空间注意力图,用于对原始特征图进行逐位加权,强化关键区域响应。CAA 模块在面对特征退化与高动态跌倒等复杂场景时,能够实现稳健且高效的上下文感知,有效增强模型对关键区域的感知能力与特征表达的稳定性。

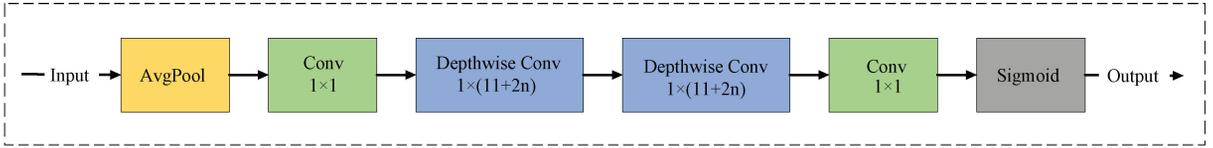


图 2 CAA 模块

Fig. 2 CAA module

1.2 C2f-IAFF 特征聚合模块

在多目标遮挡或背景噪声复杂的情况下,模型难以准确区分关键部位,从而影响对跌倒行为的完整识别与判断。针对此问题本文对主干网络 C2f 模块的 Bottleneck 进行了结构性改进,提出了融合迭代注意力特征 IAFF^[10] 的 C2f-IAFF 模块,以增强特征融合过程中对关键信息的提取,其结构如图 3 所示。

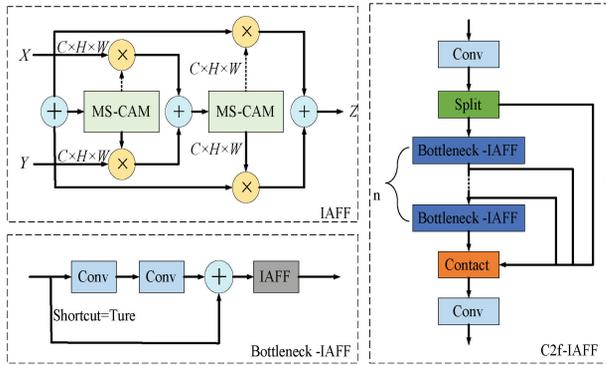


图 3 C2f-IAFF 模块

Fig. 3 C2f-IAFF module

在结构设计方面,IAFF 模块引入双分支输入机制以接收多层次特征图 X 与 Y,并以初始融合表征为基础,构建多阶段的注意力引导特征融合框架,其计算公式如下:

$$F_0 = X \oplus Y \quad (1)$$

首先,模块利用多尺度通道注意力机制 (MS-CAM) 对初始融合特征进行动态加权调制,生成融合权重,其计算公式如下:

$$M_1 = \sigma(L(F_0) + g(F_0)) \quad (2)$$

其中, $L(\cdot)$ 为局部特征, $g(\cdot)$ 为全局特征, σ 为 Sigmoid 函数。在此基础上,模块执行第一阶段融合,输出

为 Z_1 ,其计算公式如下:

$$Z_1 = M_1 \otimes X + (1 - M_1) \otimes Y \quad (3)$$

随后通过迭代优化机制,初步融合结果与原始输入再次输入到第 2 个 MS-CAM 模块,生成精细化融合权重,完成第二阶段融合过程,其计算公式如下:

$$M_2 = \sigma(L(Z_1) + g(Z_1)) \quad (4)$$

$$Z_2 = M_2 \otimes Z_1 + (1 - M_2) \otimes F_0 \quad (5)$$

C2f-IAFF 模块通过动态多尺度通道注意力机制与迭代优化策略,在多目标遮挡场景中显式分离关键部位特征与背景噪声。通过迭代式权重分配抑制背景噪声,强化被遮挡目标的局部结构特征,并融合局部细节与全局语义信息以重建遮挡区域的语义表达。相较于传统静态融合策略,其通过动态权重机制重构层间依赖,高效感知遮挡边界特征,显著增强人体姿态在遮挡条件下的结构完整性表达。

1.3 IFFN 交互式特征融合网络

颈部网络连接主干网络与检测头,通过跨层级融合构建多尺度语义通道,提升模型对不同尺度目标的感知能力。在 YOLOv10 中,所采用的路径聚合网络 (path aggregation network, PAN)^[11] 与在特征金字塔网络 (feature pyramid networks, FPN)^[12] 结构通过融合高层语义特征与低层空间定位特征,实现多尺度特征协同表达。然而高层特征因分辨率较低易丢失细节,低层特征局部感受野有限,难以建立全局语义联系,导致多尺度融合时小目标判别特征易被背景噪声干扰,引发漏检或误检。对此本文提出了交互式特征融合网络 IFFN,如图 4 所示。该模块首先结合跨尺度特征融合模块 CCFM^[13],通过统一多尺度特征的通道维度,缓解尺度间语义歧义,提升模型尺度适应性及小目标检测能力。CCFM 结构如图 5 所示。

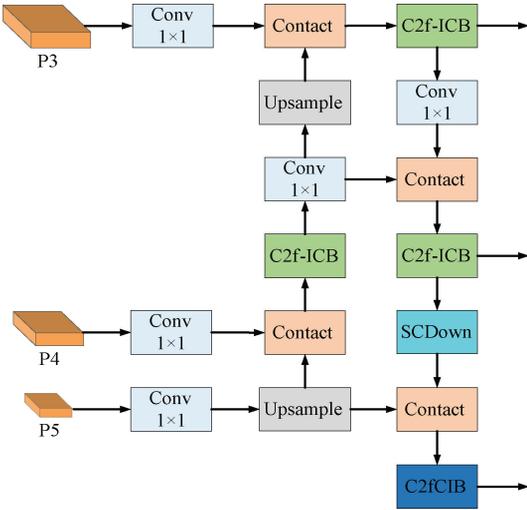


图4 IFFN结构

Fig. 4 IFFN structure

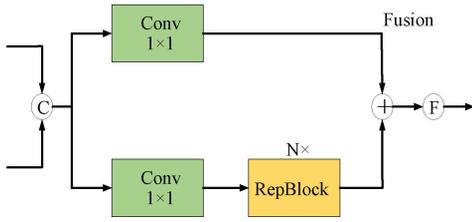


图5 CCFM模块

Fig. 5 CCFM module

此外,通过引入时间序列轻量级自适应网络(time series lightweight adaptive network, TSLANet)^[14]中的交互卷积ICB替换颈部层中C2f的Bottleneck,以增强特征交互与语义融合能力,其结构如图6所示。

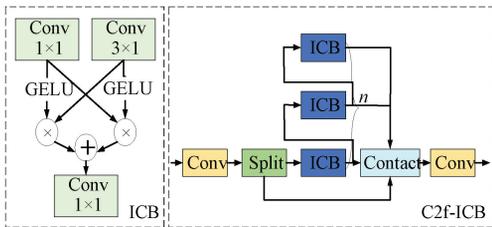


图6 C2f-ICB模块

Fig. 6 C2f-ICB module

ICB通过并行的多尺度卷积路径提取局部和全局特征,并在不同尺度间引入特征交互机制,有效筛选并强化与小目标相关的判别性线索,抑制背景噪声。具体来说,ICB使用两个不同卷积核大小的卷积分别捕捉短期和长期依赖特征,对输入特征 S' 分别应用两次卷积操作,得到 $Conv1(S')$ 和 $Conv2(S')$,再经由激活函数变换后,采用逐元素乘法进行跨通路特征调制,其计算公式如下:

$$A_1 = \phi(Conv1(S')) \odot Conv2(S') \quad (6)$$

$$A_2 = \phi(Conv2(S')) \odot Conv1(S') \quad (7)$$

其中, $\phi(\cdot)$ 是GELU激活函数, \odot 表示逐元素乘法。通过该交叉调制机制,能够显式捕捉多尺度之间的关联关系,增强特征间的动态耦合。最终,交互特征 A_1 与 A_2 被加和并送入融合卷积层Conv3以输出整合后的表征,生成更具判别性和语义一致性的整合特征表示,其计算公式如下:

$$O_{ICB} = Conv3(A_1 + A_2) \quad (8)$$

改进的交互式特征融合网络IFFN实现了高低层特征的有效互补协调,抑制了多尺度特征融合中的语义冲突与信息冗余,增强了小目标的语义关联表征能力,提升了模型在复杂场景下对小目标的判别精度与感知鲁棒性。

1.4 损失函数

本文采用与YOLOv10相同的CIoU^[15]损失函数,用以衡量预测边界框与真实框之间的几何偏差。该损失函数融合了边界框的重叠面积、中心点距离以及长宽比等多项因素,提升了模型对位置误差的敏感度;同时引入了修正因子,有效增强了模型对不同目标形状的适应性与鲁棒性,使模型在边界框回归中能够更准确地捕捉目标的空间结构特征。CIoU计算公式如下:

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (9)$$

$$\alpha = \frac{(1 - IoU) + v}{v} \quad (10)$$

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (11)$$

其中, w^{gt} 和 h^{gt} 分别表示真实框的宽度和高度, w 和 h 分别表示预测边界框的宽度和高度, $\rho(b, b^{gt})$ 表示预测框中心点与真实框中心点之间的欧氏距离, c 表示覆盖预测框和真实框的最小封闭框的对角线长度。

2 实验结果及分析

2.1 实验环境

本研究在PyTorch深度学习框架下进行模型训练,训练总轮数设为200 epoch,批处理大小为16,初始学习率设为0.01,输入图像尺寸为 640×640 ,采用SGD优化器进行参数优化。具体的实验运行环境与配置参数如表1所示。

表1 实验环境

Table 1 Experimental environment

名称	参数
操作系统	Ubuntu9.4.0
GPU	NVIDIA GeForce RTX 3090
CPU	Intel(R) Xeon(R) Gold 5218R CPU @ 2.10 GHz
显存	24 GB
编程语言	Python3.9
深度学习框架	PyTorch 2.0.1+Cuda11.8

2.2 实验数据集

本实验采用了自制人体跌倒行为检测数据集 FALL 和公开数据集 DiverseFALL10500。其中,自制数据集 FALL 是通过融合多源数据构建而成,在基于 PaddlePaddle 平台的公开数据集 FallDown 上筛选获得基础样本,再使用 GoPro 相机进行多视角补拍,通过调整拍摄角度、光照条件和背景环境等方式扩展数据多样性,最终构建的数据集包含 7 721 张图像。所有图像均通过 LabelImg 工具完成标注,根据人体动作在跌倒过程中的特征差异,将动作划分为 4 类:站立(stand)、弯腰(stoop)、蹲下(squat)和跌倒(fall)。最终按照 8 : 1 : 1 的比例将数据划分为训练集、验证集和测试集,用于模型的训练与性能评估。

DiverseFALL10500^[16]是由 Gachon University 研究团队构建的跌倒检测数据集,共包含 10 500 张图像。该数据集覆盖多种典型的室内外场景,包括家庭、公园和街道等,囊括多种光照条件、跌倒角度等复杂因素,共有“跌倒”与“非跌倒”两种类别。

2.3 实验评价指标

本实验主要采用 4 项性能指标对模型进行评估,包括召回率(Recall)、准确率(Precision)、平均精度(mean average precision, mAP)以及浮点运算总数(GFLOPs)。其中,准确率衡量的是模型在预测为正类样本中实际为正类的比例;召回率则反映模型识别出全部正类样本的能力,其计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (12)$$

$$R = \frac{TP}{TP + FN} \quad (13)$$

其中,TP(true positive)表示被正确识别为正类的样本数量,FN(false negative)表示实际为正类但未被模型识别的样本数量,FP(false positive)则表示被误判为正类的负类样本数量。

PR(precision-recall)曲线以召回率作为横轴,精度作为纵轴,曲线下的面积即表示该类别的平均精度(average precision, AP),其数值通过对该曲线下方区域进行积分计算得到。其公式如下:

$$AP = \int_0^1 P(r) dr \quad (14)$$

mAP(mean average precision)是衡量多类别检测性能的综合指标,指所有类别的平均精度 AP 之和除以类别总数,反映模型在各类目标上的整体检测能力。其中, k 表示类别数目,计算公式如下:

$$mAP = \frac{1}{n} \sum_{k=1}^n AP_k \quad (15)$$

此外,采用浮点运算总数(GFLOPs)衡量模型计算复杂度与推理效率。

2.4 消融实验

为验证本文提出的改进方法对 YOLOv10s 性能提升作用,在 FALL 数据集上依次对各改进模块进行消融分析,实验结果如表 2 所示。

表 2 消融实验结果

Table 2 Ablation test results

序号	IAFF	CAA	IFFN	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPs
1				0.793	0.75	0.826	0.516	24.5
2			✓	0.813	0.777	0.841	0.529	23.1
3	✓			0.798	0.789	0.845	0.536	27.8
4		✓		0.786	0.796	0.841	0.529	26.2
5	✓		✓	0.799	0.784	0.847	0.536	24.4
6	✓	✓	✓	0.806	0.793	0.848	0.536	24.1

实验结果表明,各改进模块均在不同程度上优化了模型的精度与性能表现。其中,IAFF 模块的改进效果最为显著,相较于基准模型 mAP@0.5 : 0.95 提升了 2%,mAP@0.5 提升了 1.9%,准确率和召回率分别提升了 0.5% 和 3.9%。IFFN 模块在 mAP@0.5 和 mAP@0.5 : 0.95 指标上分别提升 1.5% 和 1.3%,同时降低了计算量。通过集成 IAFF 与 IFFN 方法,模型检测能力进一步提升,与基准模型相比召回率提升了 3.4%,mAP@0.5 提升了 2.1%,mAP@0.5 : 0.95 提升了 2%,准确率提升了 0.6%,同时模型的计算量有所降低。模型在融合 3 种改进策略后展现出更优的检测

性能,召回率提高了 4.3%,mAP@0.5 提升了 2.2%,mAP@0.5 : 0.95 提升了 2.0%,准确率提升了 1.3%,且计算量进一步下降,表明模型在保持更高检测性能的同时具备更优的计算效率。

综上所述,IAFF、CAA 与 IFFN 模块均对提升跌倒检测性能具有积极作用。三者协同作用不仅提高了模型的检测精度,还有效降低了计算开销,验证了所提出改进策略在复杂场景下人体跌倒检测任务中的有效性。

2.5 对比实验

为了进一步验证 ICI-YOLO 在人体跌倒检测任务中的优越性,分别在自制数据集 FALL 和公开数据集

DiverseFALL10500 上将本文算法与 YOLO 系列各模型及 DETR 系列算法进行对比,如表 3 和 4 所示。

表 3 FALL 数据集对比实验结果

Table 3 Results of comparative experiments on FALL dataset

模型	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPs
YOLOv5s	0.806	0.765	0.834	0.524	23.8
YOLOv5m	0.810	0.768	0.841	0.533	64.0
YOLOv6s ^[17]	0.791	0.789	0.838	0.528	44.0
YOLOv7s ^[18]	0.827	0.779	0.822	0.514	105.2
YOLOv8s	0.789	0.775	0.836	0.528	28.4
YOLOv10s	0.793	0.750	0.826	0.516	24.5
YOLOv10m	0.781	0.781	0.837	0.535	63.4
YOLOv11s	0.824	0.761	0.842	0.532	21.3
YOLOv12s ^[19]	0.807	0.778	0.845	0.532	21.2
rt detr-resnet50	0.795	0.757	0.798	0.513	125.6
rt detr-resnet101	0.807	0.776	0.805	0.510	186.2
本文	0.806	0.793	0.848	0.536	24.1

表 4 DiverseFALL10500 数据集对比实验结果

Table 4 Results of comparative experiments on DiverseFALL10500 dataset

模型	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPs
YOLOv5s	0.880	0.882	0.936	0.747	23.8
YOLOv5m	0.899	0.861	0.934	0.755	64.0
YOLOv6s	0.897	0.866	0.936	0.748	44.0
YOLOv7s	0.861	0.874	0.923	0.706	105.1
YOLOv10s	0.883	0.871	0.932	0.732	24.4
YOLOv10m	0.882	0.864	0.93	0.753	63.4
YOLOv11s	0.884	0.880	0.936	0.752	21.3
YOLOv12s	0.877	0.876	0.933	0.749	21.2
rt detr-resnet50	0.892	0.835	0.903	0.699	125.6
rt detr-resnet101	0.879	0.841	0.904	0.702	186.2
本文	0.903	0.865	0.938	0.747	24.1

实验结果表明,本文提出的 ICI-YOLO 模型在 FALL 数据集上相较于 YOLOv10s,在精确率和召回率上分别提升了 1.3%和 4.3%,说明其在检测准确性方面具有更优表现,漏检与误检数量明显减少。同时,mAP@0.5 和 mAP@0.5:0.95 分别提升了 2.2%和 2.0%,且该性能提升未引入额外的计算成本,有效验证了本文针对 YOLOv10s 所做改进的有效性。与计算量分别为 ICI-YOLO 5 倍和 7 倍的 RT-DETR-ResNet50 和 RT-DETR-ResNet101 相比,ICI-YOLO 在召回率上分别提升了 3.6%和 1.7%,mAP@0.5 分别提升了 5.0%和 4.3%,mAP@0.5:0.95 提升了 2.3%和 2.6%。此外,相较于 YOLO 系列的其他模型,ICI-YOLO 在各项评估指标上也展现出较为优异的性能。

同样从表 4 的实验结果可以看出,ICI-YOLO 在 DiverseFALL10500 数据集上同样保持出色的检测性能。

与 DETR 系列模型相比,其召回率分别提升了 3%和 2.4%,mAP@0.5 提升了 3.5%和 3.4%,mAP@0.5:0.95 提升了 4.8%和 4.5%。相较于 YOLO 系列模型,ICI-YOLO 在召回率和 mAP@0.5 指标上均实现了更高的提升,进一步体现了模型在复杂场景下的检测准确性和鲁棒性。

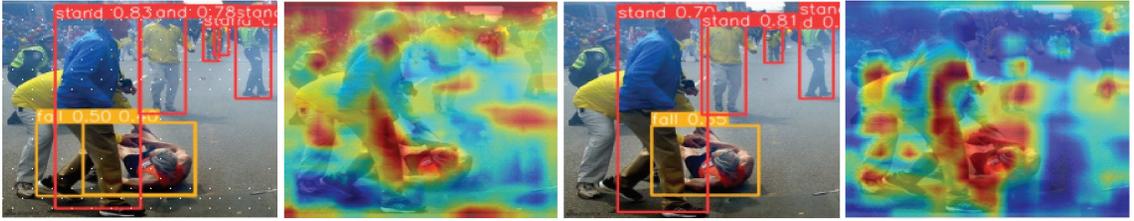
综上所述,本文提出的 ICI-YOLO 在不增加计算复杂度的条件下,显著提升了检测性能,尤其在检测精度方面表现最为优越。综合各项评估指标,所提出的方法在人体跌倒检测任务中展现出更强的整体性能。

2.6 人体跌倒检测结果可视化

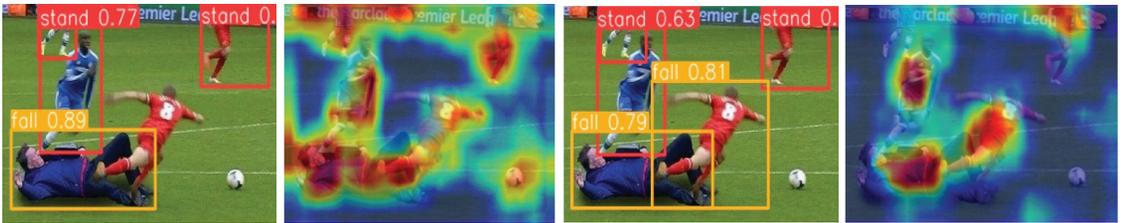
为更直观呈现改进模型 ICI-YOLO 的检测性能,本文对其与基线模型 YOLOv10s 的检测结果进行了可视对照分析。图 7 展示了改进模型 ICI-YOLO 与原始模型

YOLOv10s 在目标检测任务中的效果对比。其中,左侧为 YOLOv10s 的检测结果,右侧为 ICI-YOLO 的结果。为进一步揭示模型在决策过程中的关注区域,采用 Grad-

CAM^[20] 技术生成热力图,以可视化方式呈现模型对关键特征区域的响应。



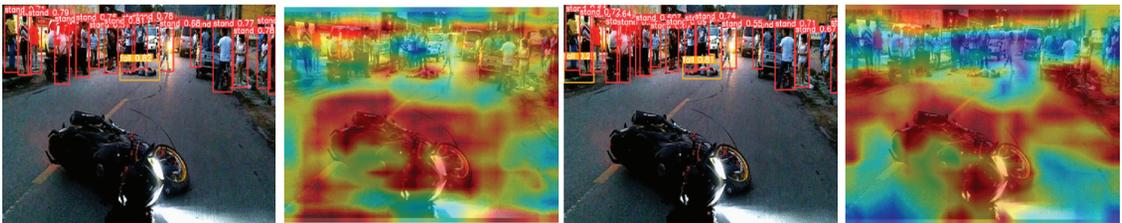
(a) Multi-target occlusion scenario



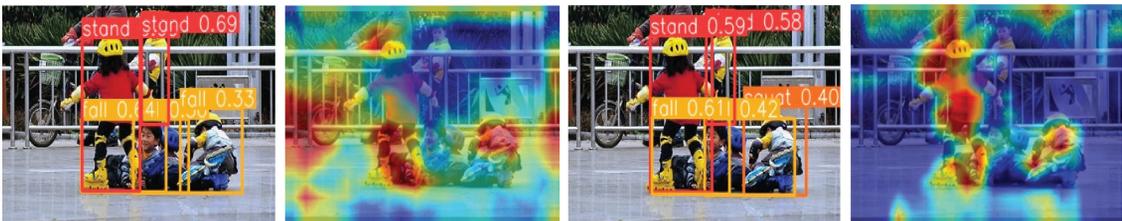
(b) Complex fall posture scenario



(c) Small-target scenario



(d) Low-light scenario



(e) Severe occlusion scenario

图 7 ICI-YOLO 与 YOLOv10s 的检测效果对比可视化

Fig. 7 Visual comparison of detection performance between ICI-YOLO and YOLOv10s

在图 7(a)中, YOLOv10s 在处理多目标遮挡场景时对弯曲的腿部误检为跌倒, ICI-YOLO 实现了对被遮挡目标

的精准识别, 有效规避误检现象。在图 7(b)中, 受限于跌倒过程中人体姿态急剧变化, YOLOv10s 对跌倒行为检测

出现漏检,ICI-YOLO 正确识别出跌倒行为且热力图的对比也显示出 ICI-YOLO 对跌倒目标具有更优的特征聚焦性能。在图 7(c)中,YOLOv10s 对远离镜头的小目标跌倒行为存在漏检问题,ICI-YOLO 成功检测出中间位置的跌倒目标,提升了对小目标的检测性能。在图 7(d)中,由于复杂背景加之低照度环境影响,光照不均和噪声干扰导致 ICI-YOLO 无法有效区分目标与背景,降低了模型对细节的感知能力,从而引发误检。在图 7(e)中,由于极端遮挡导致人体关键部位信息严重丢失,ICI-YOLO 无法提取完整的跌倒空间特征,最终出现漏检情况。

综合对比结果表明,ICI-YOLO 在复杂背景下的人体跌倒检测任务中展现出显著的性能优势。尽管该模型在不同尺度和遮挡条件下的跌倒检测任务中展现了较强的识别能力,但在低照度和极端遮挡环境下仍存在误检和漏检现象。未来的研究将重点探索通过引入多模态传感器数据增强低照度环境下的识别能力,并结合人体关键点检测技术解决极端遮挡下的漏检问题,从而提升跌倒检测技术在复杂场景下的应用效果。

3 结 论

针对复杂场景下人体跌倒检测易受背景干扰、特征信息丢失及目标尺度差异显著等因素影响,导致检测精度受限的问题,本文基于 YOLOv10s 模型进行了改进。首先引入 CAA 注意力替代部分自注意力结构,构建跨层次局部感知与全局上下文依赖的动态聚合机制;其次设计 C2f-IAFF 模块,提高模型对关键区域的细粒度感知与结构化语义表征能力,提升其在密集遮挡与复杂噪声场景中的异构特征处理能力;最后,提出 IFN 颈部网络,用于强化多尺度特征的交互与渐进式聚合,增强模型对小目标与复杂结构的感知能力。在自制数据集 FALL 及公开数据集 DiverseFALL10500 上均体现出良好的检测性能。尽管如此,本文提出的模型仍存在进一步优化的空间。后续研究将致力于在提升跌倒检测精度的同时,进一步降低计算开销,优化实时检测的处理效率,更高效地满足人体跌倒检测对实时性与实用性的双重需求,推动人体跌倒检测技术的实际部署与应用。

参考文献

- [1] KAKARA R. Nonfatal and fatal falls among adults aged ≥ 65 years—united states, 2020-2021 [J]. MMWR. Morbidity and Mortality Weekly Report, 2023, 72(35): 938-943.
- [2] 孙巍伟,梁毅玮,毛亦鹏,等.基于多传感信息融合的跌倒监测系统开发[J].电子测量技术,2024,47(22): 94-103.
SUN W W, LIANG Y W, MAO Y P, et al. Development of a fall detection system based on multi-sensor information fusion[J]. Electronic Measurement Technology, 2024,47(22):94-103.
- [3] QI B, GUAN M, SUN W. A novel fall detection methodology based on feedback-fusion and optimal-allocation model [C]. 2024 IEEE International Conference on Mechatronics and Automation(ICMA). IEEE, 2024: 602-607.
- [4] HWANG H, KIM D, KIM H. FD-YOLO: A YOLO network optimized for fall detection [J]. Applied Sciences, 2025, 15(1): 453.
- [5] HU H, JIANG L. MAM-YOLOv5 based method to detect underground staff falling down accidents[C]. 2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP). IEEE, 2023: 2035-2038.
- [6] ZHAO D, SONG T, GAO J, et al. YOLO-fall: A novel convolutional neural network model for fall detection in open spaces[J]. IEEE Access, 2024, 12: 26137-26149.
- [7] 龙艺文,闵宣霖,陈奕兆,等.改进 YOLOX 的轻量级人体跌倒检测算法[J].电子测量技术,2023,46(10): 109-116.
LONG Y W, MIN X L, CHEN Y ZH, et al. Lightweight human fall detection algorithm of improved YOLOX [J]. Electronic Measurement Technology, 2023,46(10):109-116.
- [8] WANG AO, CHEN H, LIU L, et al. YOLOv10: Real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.
- [9] CAI X, LAI Q, WANG Y, et al. Poly kernel inception network for remote sensing detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition,2024: 27706-27716.
- [10] DAI Y, GIESEKE F, OEHMCKE S, et al. Attentional feature fusion [C]. IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 3560-3569.
- [11] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [12] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [13] ZHAO Y, LYU W, XU S, et al. Detsr beat YOLOs on real-time object detection [C]. IEEE/CVF Conference on Computer Vision and Pattern

- Recognition, 2024: 16965-16974.
- [14] ELDELE E, RAGAB M, CHEN Z, et al. Tslanet: Rethinking transformers for time series representation learning[J]. ArXiv preprint arXiv:2404.08472, 2024.
- [15] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression [C]. AAAI Conference on Artificial Intelligence. 2020, 34(7): 12993-13000.
- [16] KHAN H, ULLAH I, SHABAZ M, et al. Visionary vigilance: Optimized YOLOV8 for fallen person detection with large-scale benchmark dataset [J]. Image and Vision Computing, 2024, 149: 105195.
- [17] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. ArXiv preprint arXiv:2209.02976, 2022.
- [18] 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, 2023: 7464-7475.
- [19] TIAN Y, YE Q, DOERMANN D. YOLOv12: Attention-centric real-time object detectors[J]. ArXiv preprint arXiv:2502.12524, 2025.
- [20] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization [C]. IEEE International Conference on Computer Vision, 2017: 618-626.

作者简介

郭莉, 硕士研究生, 主要研究方向为计算机视觉。

E-mail: 908548454@qq.com

张雪松(通信作者), 博士, 副教授, 硕士生导师, 主要研究方向为计算机视觉、智能优化算法。

E-mail: zhangxuesong@djtu.edu.cn

李萌萌, 硕士研究生, 主要研究方向为计算机视觉。

E-mail: 1697293721@qq.com

金花, 硕士, 副教授, 硕士生导师, 主要研究方向为计算机视觉。

E-mail: zhimingj@126.com