

一种融合运动特征的行人跌倒检测算法

李 凯, 曹建荣, 尚 硕, 韩发通, 庄 园

(山东建筑大学 信息与电气工程学院, 济南 250101)

摘要: 针对传统跌倒检测算法局限于单帧图像判断而未充分利用行人时序运动特征的问题, 提出了一种融合运动特征的行人跌倒检测算法; 在 YOLOv8n 的主干网络中引入 ECA 注意力机制, 在 Neck 中用 GSConv 代替常规卷积提高行人检测的精度; 为了避免多行人之间的特征发生混淆, 使用 DeepSort 目标跟踪算法进行行人跟踪, 独立提取行人的运动特征; 分类阶段设置一个固定长度的队列来存储行人运动特征值, 利用提前训练好的 1DCNN 进行跌倒识别; 实验证明, 在 VOC2007 测试集上, 改进后的 YOLOv8n 模型比改进前的 mAP 增加 1.5%, 参数量减少 2.97%; 跌倒检测算法在 UR Fall 数据集上实现了 97.8% 的准确率; 在自建的多行人数据集上实现 92.91% 准确率。

关键词: 深度学习; YOLO; 目标检测; 运动特征; 跌倒检测

Pedestrian Fall Detection Algorithm Integrating Motion Features

LI Kai, CAO Jianrong, SHANG Shuo, HAN Fatong, ZHUANG Yuan

(School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, China)

Abstract: Aiming at the problem that traditional fall detection algorithms are limited to single-frame image judgment and do not fully utilize the temporal motion characteristics of pedestrians, a pedestrian fall detection algorithm that integrates motion features is proposed; the efficient channel attention (ECA) attention mechanism is introduced into the backbone network of YOLOv8n, and the GSConv is replaced by conventional convolution in the Neck to improve the accuracy of pedestrian detection; In order to avoid feature confusion among multiple pedestrians, the DeepSort target tracking algorithm is used to track pedestrians and independently extract the motion features of pedestrians; In the classification stage, a fixed-length queue is set to store the feature values of pedestrian motion, and the 1DCNN trained in advance is used for fall recognition. Experiments show that on the VOC2007 test set and compared with the pre-improved model, the improved YOLOv8n model increases the mean average precision (mAP) by 1.5% and reduces the number of parameters by 2.97%; The fall detection algorithm achieved an accuracy of 97.8% on the UR Fall data set, with an accuracy of 92.91% on the self-built multi-pedestrian data set.

Keywords: deep learning; YOLO; target detection; motion feature; fall detection

0 引言

随着全球人口的不断增加和人口老龄化问题的严重加剧, 老年人的健康问题成为社会关注的焦点。根据第七次全国人口普查数据, 我国 60 岁以上的人群共有 2.64 亿人, 占总人口数的 18.70%, 其中 65 岁及以上人口数为 1.90 亿人, 占比 13.50%^[1]。人口老龄化是我国社会的一个重要趋势, 而跌倒则成为危害老年人健康最为严重的风险之一。摔伤后的治疗和跌倒后的急救服务占据了相当大的医疗资源。如果能够在跌倒事件发生后及时、准确地检测, 并迅速提供援助, 将显著降低跌

倒风险及其可能导致的负面影响^[2], 这对于提高老年人的生活质量、减轻医疗系统的负担, 以及降低医疗费用都具有重要意义。因此, 研究跌倒检测算法对这一人群进行监测具有很大的价值。

传统的跌倒检测方法主要分为两类: 基于可穿戴设备的跌倒检测、基于环境式的跌倒检测^[3]。基于可穿戴设备的方法是最初用来检测跌倒的最佳方案, 将加速度、陀螺仪等传感器嵌入到老人手环或者腰带中, 通过传感器来采集人体的运动特征信息, 然后对采集到的数据进行分析进而判断行人是否发生跌倒, 这种方法不需对人体的特征进行采集, 保护了个人的隐私, 但是需要

收稿日期:2023-12-12; 修回日期:2024-01-17。

基金项目:山东省重点研发计划(2019GSF111054);国家自然科学基金(62073196,U1806204)。

作者简介:李 凯(1997-),男,硕士研究生。

引用格式:李 凯,曹建荣,尚 硕,等.一种融合运动特征的行人跌倒检测算法[J].计算机测量与控制,2025,33(2):212-220.

人体长期穿戴, 舒适感很差, 并且在复杂的活动中会有较高的误报率。文献 [4] 使用无线传感器和三轴加速度计构成一个传感器检测单元, 利用传感器单元检测和传输跌倒的情况, 发现误报率发生在躺卧姿势的过渡期间比较多。为了提高检测的准确性, 文献 [5] 将先进的九轴运动传感器与加速度传感器结合, 利用人体跌倒运动学特征, 设计了一套基于可穿戴设备的跌倒检测系统, 利用多阈值检测, 提高了跌倒检测的准确性。基于环境式的检测方法主要是将非视频类传感器, 如加速度、红外传感器和陀螺仪传感器等设备, 嵌入到人体活动的环境里, 例如地板或者天花板, 通过分析传感器采集到的数据来进行跌倒判断。基于环境式的检测方法无需人体长期佩戴设备, 也不需要采集个体特征信息, 但只能采集固定场所的数据, 容易受环境噪音干扰, 且设备配置昂贵。根据行人在跌倒和正常活动时对地板的震动特征不同, 文献 [6] 用一个特殊的压电传感器放在地板表面上, 通过监测地板振动模式来估计正常活动和人体跌倒。文献 [7] 设计了一种基于红外阵列传感器的人体行为识别系统, 该系统通过检测环境中的温度分布和变化情况利用 K 近邻算法实现了 97.5% 的跌倒检测准确率。传统的跌倒检测方法主要是依赖各种传感器设备, 通过传感器记录的运动特征信息来判断人体是否发生跌倒, 但是这类方法准确率有限, 容易出现误报。

1 基于计算机视觉的跌倒检测

随着深度学习和计算机视觉的发展, 用计算机视觉的方法来识别跌倒行为有了较高的准确率。基于计算机视觉的跌倒检测算法主要是通过摄像机对场景中的行人进行连续监控, 然后对采集的视频数据进行特征提取, 并使用相应的算法进行跌倒检测。基于计算机视觉的方法无需穿戴任何设备, 不会影响人体的活动, 具有实时性好、成本低和检测精度高的特点。缺点在于可能会采集到人的隐私信息并且场景中可能存在遮挡等。文献 [8] 利用背景减除法提取人体轮廓的前景信息, 将人体

的轮廓信息作为 CNN (Convolutional Neural Network) 的输入, 通过标识进行分类 (站立、躺下和跌倒), 通过实验证明能达到 96.88% 的跌倒检测准确率, 但是类跌倒行为可能会有误判。文献 [9] 开发了一种基于三维卷积神经网络的跌倒检测方法, 该方法只使用视频运动数据来训练自动特征提取器, 可以绕过深度学习解决方案对大量跌倒数据集的要求, 二维 CNN 只能对空间信息进行编码, 采用的三维卷积神经网络可以从时间序列中提取运动特征, 可以更好地利用跌倒时的空间特征和时序特征信息。文献 [10] 提出了一种基于注意力机制改进的 Yolox 跌倒检测算法, 通过对目标检测算法进行改进, 从而提高跌倒检测的准确率, 但是此类目标检测算法只用单帧静态特征进行检测分类, 不能充分利用跌倒的时序运动信息。

跌倒, 是一种意外且瞬间的活动, 其迅速发生和完结。在本文中, 把跌倒定义为一个特定的动作, 而这个动作通常是需要几秒钟才能发生, 并且本文算法只关注场景中的行人目标, 而不需要对场景进行长时间的分析, 能够减少对整个场景的冗余处理, 使得算法更加高效。

2 跌倒检测算法总框架

2.1 算法总框架

本文算法首先使用改进的 YOLOv8n 模型对视频中的行人进行行人检测, 生成带有矩形框的视频流, 随后将行人框的坐标信息传入 DeepSort^[11] 进行行人跟踪, 独立提取行人运动特征并保存下来, 为了更好地跟踪到跌倒的行人, 并且减少跟踪过程中的 ID 切换次数, 本文 DeepSort 算法中深度外观网络使用跌倒行人数据集混合 Market1501^[12] 数据集进行联合训练, 最后将跟踪提取的行人运动特征传入跌倒检测算法, 利用提前训练好的一维卷积神经网络^[13] 对视频中的每个行人的运动特征轨迹进行跌倒判断。本文的算法总框架如图 1 所示。

2.2 改进的 YOLOv8 目标检测算法

目前, 目标检测算法有许多, 主要代表有两阶段检

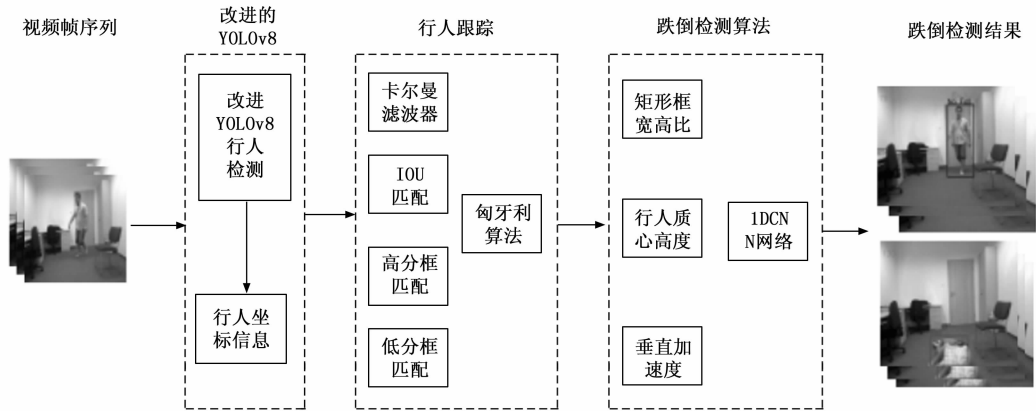


图 1 本文算法总框架

测算法和一阶段检测算法,两阶段检测算法精度高,但是速度慢,一阶段检测算法精度稍低但是检测速度快,其中一阶段检测算法以 YOLO (You Only Look Once)^[14] 算法为代表。YOLOv8 是 YOLO 目标检测模型的最新版本,其架构与 YOLOv5 类似,YOLOv8 使用了与 YOLOv5 类似的主干,只是在 CSPLayer 上做了一些更改,使用了新的特征提取网络,将 YOLOv5 的 C3 模块换成了 C2f 模块。C2f 模块(交叉阶段部分瓶颈与两个卷积)将高级特征与上下文信息相结合,以提高检测的准确性。该结构参考了 YOLOv7^[15] 的 ELAN 结构,梯度流更丰富。在之前的 YOLO 体系结构中,通常有几个输出头负责预测被检测对象的不同方面,比如边界框坐标、类别概率和目标概率。这些输出头通常连接到神经网络的最后几层,并被训练为输出一组可用于定位和分类图像中的对象的值。输出头的数量和类型可以根据具体的目标检测算法和手头任务的要求而变化。YOLOv5 有 3 个输出头,YOLOv8 有 1 个输出头,没有小、中、大的锚框,YOLOv8 使用带有解耦头的无锚模型来独立处理对象性、分类和回归任务。它直接预测对象的中心,而不是与已知锚盒的偏移量,这种设计允许每个分支专注于自己的任务,从而减少了框的预测数量,加快了后处理过程,提高了模型的总体精度。YOLOv8 使用 Ciou^[16] 和 DFL 作为回归损失,使用二元交叉熵进行分类损失。这些损失提高了对象检测性能,特别是在处理较小的对象时。

在跌倒检测过程中,目标检测起到很关键的作用,为了进一步优化 YOLOv8 的检测性能,本文在主干网络中添加 ECA (Efficient Channel Attention)^[17] 注意力机制,提高主干网络的特征提取能力,在 Neck 中使用 GSconv^[18] 代替常规卷积,提高计算效率,降低计算成

本。改进后的网络结构如图 2 所示。

2.2.1 ECA 注意力机制

在深度学习中,引入注意力机制可以使模型更关注任务本身,降低冗余信息的关注度。注意力机制不仅能即插即用而且还能提高网络的表达能力,为了增强 YOLOv8 的特征提取能力,在主干网络中添加 ECA 注意力机制。ECA 注意力机制是一种轻量级的注意力机制,在增加少量参数的情况下可以实现模型的性能增加。首先输入特征经过全局平均池化(GAP, global average pooling)获得每个通道的全局平均值,ECA 通过执行大小为 k 的快速 1D 卷积来生成通道权重,其中 k 的大小可以自适应确定,最后通过一维卷积来获得不同通道之间的权重。ECA 注意力机制的结构如图 3 所示。

2.2.2 GSconv

在神经网络中平衡好模型参数和性能是至关重要的,为了提高模型的性能,引入其他模块往往会增加模型的参数量。在常规卷积中需要卷积核和每个输入通道进行卷积操作,而深度可分离卷积^[19] 将一个卷积操作分为逐通道卷积和逐点卷积两部分,较常规卷积可有效降低模型参数量,但是其通道信息被分离,会使精度降低。GSConv 是一种轻量级的卷积,较深度可分离卷积可以提升精度,较常规卷积可以降低计算成本,并且性能与普通卷积不相上下。GSConv 的结构如图 4 所示,首先输入特征经过一个普通卷积进行下采样操作,然后使用 DWConv 深度卷积对下采样后的特征图进行卷积,然后将两个卷积的结果沿通道维度进行拼接,最后进行 shuffle 操作,让之前两个卷积对应的通道数均匀混合在一起。

2.3 行人跟踪算法

当视频中有多个行人时,为避免行人之间的运动特

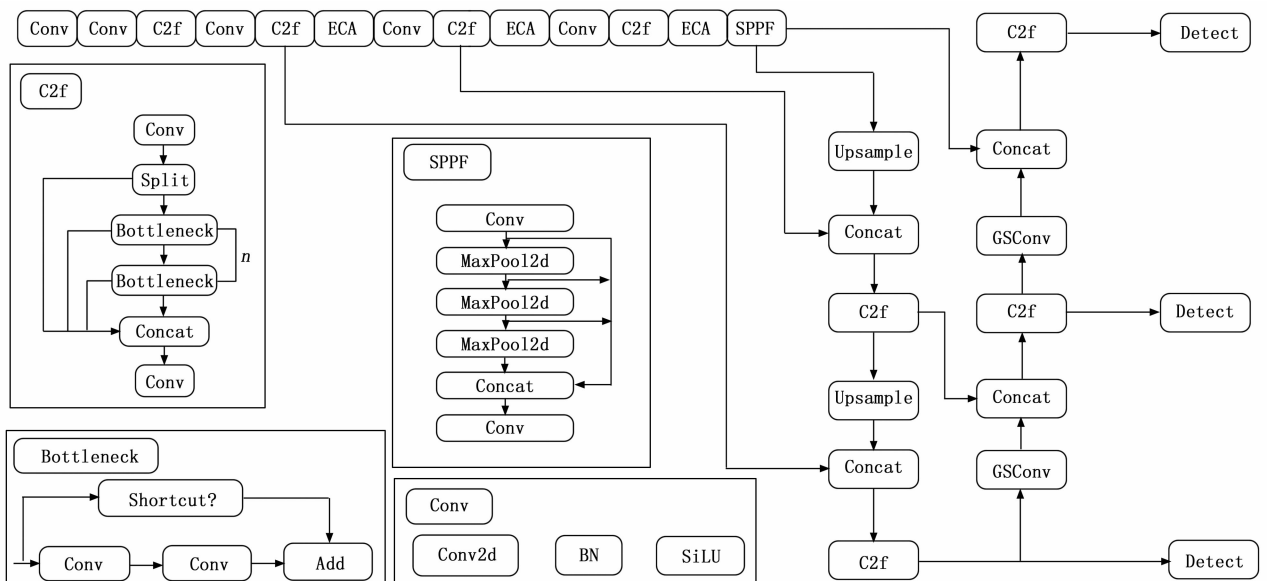


图 2 改进后的 YOLOv8 网络结构图

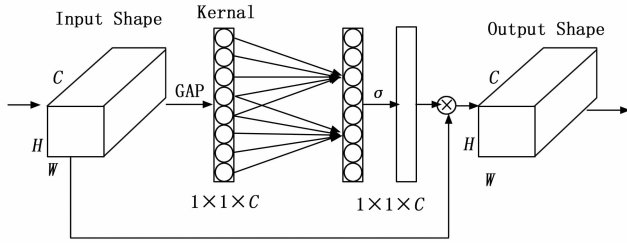


图 3 ECA 网络结构图

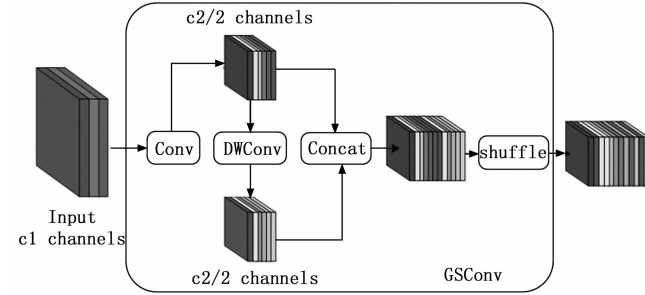


图 4 GSConv 网络结构图

征发生混淆,此时需要对行人目标进行跟踪,提取每个行人的运动特征进行独立分析。经过目标检测后,得到行人的坐标信息,将行人的坐标信息传入 Deepsort 进行目标跟踪。Deepsort 是一种基于检测器的跟踪算法,算法首先接收目标检测器传入的目标,例如 Faster RCNN^[20]、YOLO 等。Deepsort 目标跟踪算法是在 Sort^[21] 目标跟踪算法的基础上提出来的,在 Sort 跟踪算法的基础上增加了级联匹配和深度外观提取网络,目的是在遮挡等环境下降低行人的 ID 切换次数。Deepsort 利用卡尔曼滤波器^[22] 预测该行人框的轨迹,用 8 维空间去预测和更新行人某时刻的状态。首先检测到对象后,执行轨迹预测,将检测对象和预测对象进行匹配,匹配成功后执行状态更新,如果匹配失败,通过阈值来决定是否删除此轨迹,如果检测没有匹配到对象,则会初始化该对象为新的轨迹,新的轨迹需要连续匹配到 3 帧后才能成为确定的轨迹。只有当前帧轨迹和当前帧检测匹配成功才能完成跟踪任务。Deepsort 相较于 Sort 算法在匈牙利算法分配方面使用了运动特征和外观特征完成分配。运动特征使用马氏距离的平方来衡量轨迹和检测的距离。这种改进在匈牙利算法分配方面更为精准,能够更有效地应对复杂场景,更准确、独立地跟踪视频中的多个行人,为后续的跌倒检测提供了可靠的基础。马氏距离如公式 (1) 所示:

$$d^{(1)}(i, j) = (d_j - y_i)^T S^{-1} (d_j - y_i) \quad (1)$$

其中: d_j 为第 j 个检测框, y_i 为第 i 个轨迹, S^{-1} 为第 j 个检测框和第 i 个轨迹之间的协方差矩阵。

外观特征使用余弦距离来度量检测框和轨迹之间的相似度,当行人在长时间检测不到或者遭受遮挡时,需要使用行人的外观特征完成匹配,计算检测框和轨迹之

间的余弦距离,余弦距离公式如下:

$$d^{(2)}(i, j) = \min \{ 1 - r_j^T r_k^{(i)} \mid r_k^{(i)} \in R_i \} \quad (2)$$

$$b_{i,j}^{(2)} = d^{(2)}(i, j) \leq t^{(2)} \quad (3)$$

式中, $r_j^T r_k^{(i)}$ 计算的是余弦相似度^[23],而余弦距离等于 1 减去余弦相似度,通过计算行人的外观特征进行匹配,极大地降低了行人的 ID 切换次数。经过行人跟踪后,此时生成了带有行人 ID 号的视频流,通过 ID 号可以区分每一个行人,并且根据行人 ID 号,提取行人不同帧之间的运动特征。

2.4 行人跌倒检测算法

本文的跌倒检测算法旨在根据行人检测和跟踪的结果,对每个矩形框中的行人完成跌倒与否的二分类任务。传统的跌倒检测算法往往存在特征提取较为单一的问题,有时难以准确判断跌倒,而且大多数算法主要依赖于单帧图像的特征。为了更充分地利用行人在跌倒过程中的时序特征和运动特征^[24],本文提出了使用跟踪后行人的多维运动特征来进行跌倒判断。具体而言,相较于传统方法,本文的创新点在于将行人跟踪的多维运动特征纳入跌倒检测的判定过程,通过跟踪算法获得的连续帧行人轨迹,因此可以更全面地捕捉行人在时序上的运动变化。这种时序信息对于判断跌倒行为至关重要,因为跌倒往往伴随着特定的运动模式和姿态。通过结合运动特征的多维信息,本文的算法能够更准确地区分行人的正常行走和潜在的跌倒状态,这种方法相较于单一帧图像的传统特征提取方式,更具有对跌倒行为的敏感性和可靠性。同时,通过跟踪后的运动特征,算法对于在多个连续帧中发生的跌倒行为能够更加稳健地作出判断,从而提高了跌倒检测的准确度和鲁棒性。

2.4.1 行人运动特征

在视频帧中根据行人跌倒过程中的跌倒方向,可将跌倒分为横向跌倒和纵向跌倒两种情况,图 5 和图 6 分别为人体横向和纵向跌倒过程的运动特征示意图,图中水平方向定义为 x 方向,垂直方向定义为 y 方向,通过对跌倒过程中人体的不同姿态进行分析得到跟踪框宽高比、行人质心高度、行人垂直加速度 3 个运动状态特征,用于描述跌倒过程中的运动特性。

2.4.1.1 跟踪框宽高比

在视频帧中,根据行人跌倒时的倾斜方向可以分为横向跌倒和纵向跌倒,无论是横向跌倒还是纵向跌倒,在跌倒过程中行人矩形框的宽度是逐渐增大、高度是逐渐减小,这一特征对于捕捉跌倒行为的动态变化非常关键,因此可以利用跟踪矩形框宽高比作为跌倒检测的一个特征。设行人跟踪框左上角坐标为 (x_1, y_1) ,右下角坐标为 (x_4, y_4) ,则跟踪框的宽 w 高 h 计算公式为:

$$w = x_4 - x_1 \quad (4)$$

$$h = y_1 - y_4 \quad (5)$$

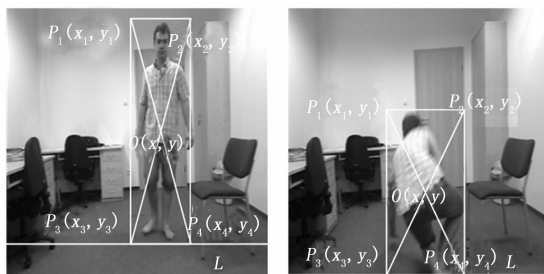


图 5 横向跌倒示意图

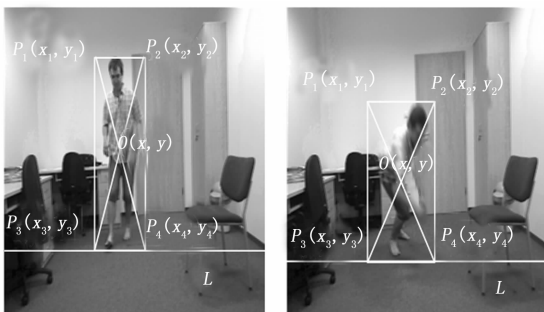


图 6 纵向跌倒示意图

$$r = w/h \quad (6)$$

视频中的行人，无论是横向跌倒还是纵向跌倒，在跌倒前其 $w < h$ ，当行人从站立状态逐渐跌倒时 w 会逐渐增大， h 会逐渐减小。图 7、图 8 是在 URFall^[25] 数据集上对横向跌倒和纵向跌倒宽高比等变化过程示意图，无论是横向跌倒还是纵向跌倒，在跌倒过程中矩形框的宽高比是逐渐增大的。由图可以看出，当行人在站立状态下，其宽高比在 0.5 左右，由站立到跌倒状态时，其宽高比不断增加。

2.4.1.2 行人质心高度

设行人框 3 个顶点的坐标自上到下分别为 $P_1(x_1, y_1)$ ， $P_2(x_2, y_2)$ ， $P_3(x_3, y_3)$ ， $P_4(x_4, y_4)$ 将行人框对角

线进行连接，则交点为行人的质心 $C(x, y)$ ，并且沿 P_3P_4 水平方向画一条直线 L 作为地面相对高度，当行人由站立状态到跌倒状态时，人体的质心高度是逐渐发生变化的，将质心高度定义为质心 C 距离 P_3P_4 水平线的垂直距离，记为 z ，则：

$$z = y_1 - y \quad (7)$$

当行人由站立状态逐渐跌倒时质心高度 z 逐渐减小。这个特征的引入充分考虑了跌倒状态下行人整体的垂直位置变化。因此，可以选择行人质心高度作为一个跌倒检测特征。

2.4.1.3 行人垂直加速度

行人在正常行走、慢跑、蹲下、坐下等过程中，其垂直方向加速度变化幅度较小，而在跌倒这种意外过程中，在这一瞬间，垂直方向加速度会突然变大。设相邻两帧行人的质心为 $C_i(x_i, y_i)$ ， $C_{i+1}(x_{i+1}, y_{i+1})$ ，那么第 $i+1$ 帧行人垂直方向的速度 v_{i+1} 为：

$$v_{i+1} = \frac{|y_{i+1} - y_i|}{t} \quad (8)$$

其中： t 为相邻两帧的时间间隔，此时垂直方向的加速度为：

$$a_{i+1} = \frac{v_{i+1} - v_i}{t} \quad (9)$$

可以利用加速度的瞬时变化，作为跌倒检测的一个特征。

由图 7、图 8 可以看出，横向跌倒在 170 帧后发生跌倒，其宽高比从 0.5 左右不断增加到 2 左右，质心高度不断下降，垂直方向加速度会瞬间增大。纵向跌倒在 90 帧后发生跌倒，宽高比也会逐渐增大，质心高度会逐渐下降，垂直方向加速度会瞬间增大。非跌倒行为与跌倒行为在行人跟踪框宽高比、行人质心高度和垂直方向加速度方面有明显的区分度，并且每一个特征都是相互独立的，具有可分别性。

综上所述，用跟踪框宽高比，行人质心高度和垂直方向加速度可以作为区分人体是否跌倒的行人运动特征。这些特征的组合能够更全面地捕捉到跌倒行为的动态特征，为跌倒检测算法提供了更多的上下文信息，增强了算法对跌倒行为的敏感性和准确性。

2.4.2 特征窗口序列

在传统的跌倒检测算法中，通常仅考虑单帧特征，而跌倒作为一个动作过程，需要结合跌倒过程多帧图像来进行判断^[26]。在本文中使用一个固定大小的队列来实现滑动窗口的思想，以更好地模拟跌倒过程的运动状态变化。首先，初始化一个空队列，队列大小为 m ，将每一帧中的每个行人的特征值分别存入不同的队列中，当队列未滿时，持续将特征值入队至队滿状态，队滿时执行一维卷积判断，接着，出队一个特征值，再将新的

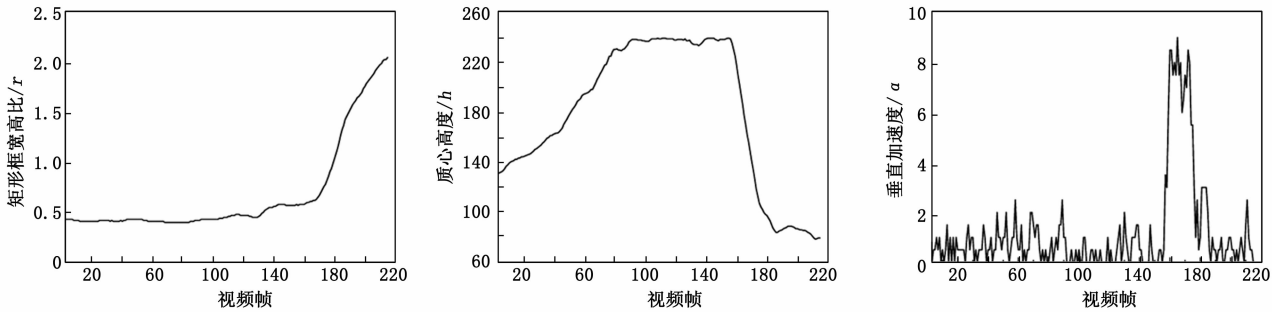


图 7 横向跌倒视频序列

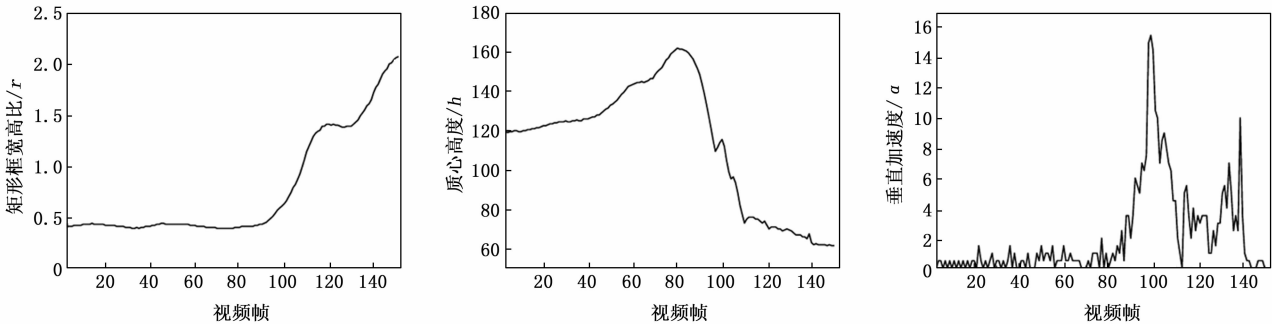


图 8 纵向跌倒视频序列

特征值入队, 继续执行判断。这样, 队列中始终保持 m 个特征值, 整个过程持续进行, 直至视频流结束。这一结合队列和滑动窗口的设计, 有效地提高了对跌倒行为的检测效果。通过综合考虑多帧的信息, 算法更全面地捕捉到跌倒的时序特征。队列如图 9 所示。

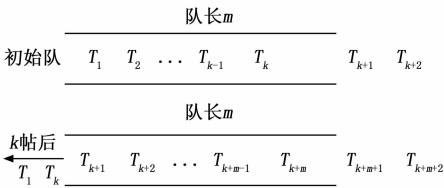


图 9 队列示意图

如图 9 所示, 设队列 Q 的长度为固定大小 m , 在本文实验中, 通过验证使用的队列大小为 20, 将跟踪获得的帧序列特征值依次传入队列中, 当队列满时将特征值传到 1DCNN 卷积神经网络进行跌倒判断。

2.4.3 1DCNN 与跌倒判断

卷积神经网络 (CNN) 早在 20 世纪 90 年代就应用于图像分类等领域。在深度学习中二维卷积神经网络常用来对图像进行卷积操作, 而对于序列类数据使用一维卷积更容易捕获序列之间的关系。一维卷积的卷积方向是一个维度, 只对宽度进行卷积, 而不对高度进行卷积, 但其输入特征可以有多维。一维卷积就是在一维空间上对元素进行卷积。在本文中经过行人检测和跟踪后在每一帧中每一个行人获得了 3 个运动特征值 r_i, z_i, a_i , 其中 i 表示第 i 个行人。将每连续 20 帧的特征值保存成 20×3

的一组向量, 将此向量经过预处理转换维度后输入到一维卷积神经网络中, 经过一系列卷积和池化操作进行特征提取, 最后展平输出为 Fall 和 No fall 两个类别。

3 实验与结果分析

3.1 实验环境

本文实验使用 Amax 服务器进行训练, 该服务器操作系统为 Ubuntu16.04; 硬件平台为 CPUE5-2630 V4@2.2 GHz, 搭载两个显卡为 GTX1080TI (11 GB×2), 内存容量 128 GB; 软件平台采用 Pytorch1.10、CUDA11.3、Pycharm, 使用 Pytorch 深度学习框架进行训练测试。

3.2 目标检测实验

为了验证改进后的 YOLOv8 目标检测算法的性能, 本文选择在公开数据集 VOC2007 和 VOC2012 组成的目标检测数据集上进行实验, 其中 VOC2007 和 VOC2012 数据集共有 20 类, 两个数据集独立, 本文在 VOC2007 和 VOC2012 的 train+val (16 551 张) 上进行训练和验证, 然后使用 VOC2007 test (4 952 张) 进行测试。将 VOC2007 和 VOC2012 的 train+val 训练集和验证集按照 8 : 2 的比例划分, 其中训练集 13 241 张, 验证集 3 310 张, 测试集 4 952 张。最终在测试集上进行算法测试, 各种轻量级的目标检测算法对比实验如表 1 所示。

由表 1 可以看出, 在对比了多个轻量级的目标检测网络后, 改进的 YOLOv8n 目标检测算法在最少的参数量下能够取得最高的 mAP , 并且相对于改进前的

YOLOv8n 目标检测算法 mAP 提高了 1.5%，参数量下降 2.97%。由于改进后的模型网络层数有所增加，改进后模型的推理速度相对于原始模型稍微变慢。

表 1 目标检测算法网络对比

算法	mAP	$Params/M$	t/ms
YOLOv4-Tiny	57.3	5.91	6.0
YOLOv5s	76.6	7.07	6.9
YOLOv7-Tiny	75.2	6.06	9.1
YOLOv8n	76.0	3.01	2.6
改进 YOLOv8n	77.5	2.92	2.9

3.3 跌倒评价指标

在本文的跌倒检测算法评价过程中，使用准确率 (Accuracy) 作为评价指标^[27]，具体计算公式如下所示：

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (10)$$

其中： TP (True Positive) 为视频帧中跌倒类别检测为跌倒类别； FP (False Positive) 为视频帧中非跌倒类别错误地检测为跌倒类别； FN (False Negative) 为视频帧中跌倒类别错误地检测为非跌倒类别； TN (True Negative) 为视频帧中非跌倒类别检测为非跌倒类别。

3.4 跌倒算法公开数据集测试

为了测试本文跌倒检测算法的性能，本文选择在公开数据集上进行测试。在跌倒检测领域，常用的公开数据集几乎都是单人实验室场景，本文选择比较常用的数据集 UR Fall Detection Dataset 数据集，该数据集包含 30 个跌倒视频序列，由于这些跌倒序列大多场景与动作类似，本文选择了 4 个不同的行人，分别选取由站立到跌倒和由坐着到跌倒的八段视频来评价算法。在站立与跌倒（摔倒在地）之间会有一个跌倒中状态，无法具体到某一帧为跌倒状态，本文定义只要在跌倒后能检测到跌倒状态而不再发生误判，则为检测正确。部分实验结果如图 10 所示。

由图 10 可以看出，本文算法在该数据集上能够满足对跌倒的检测，但也有误检的发生，人物 2 在由坐立到跌倒的过程中发生了一帧误检。由于该数据集较为简单，并且场景中只有一个人，算法的平均准确率也比较高，能达到 97.8% 的准确率，如表 2 所示。

为了与其他算法进行对比，本文选择在 UR Fall 数据集上与其他算法进行比较，比较结果如表 3 所示，文献 [28] 是通过计算集成特征向量来构建鲁棒的特征向量，使用传感器获得深度轮廓，通过应用 PCA 降低组合特征向量维度，并应用 SVM 对活动进行分类。文献 [29] 提出了一种基于轮廓变形的跌倒检测算法，该方法使用了一种基于轮廓变形的描述符，并提出了一种新的图像序列表示来捕获不同姿势之间的变化，利用形

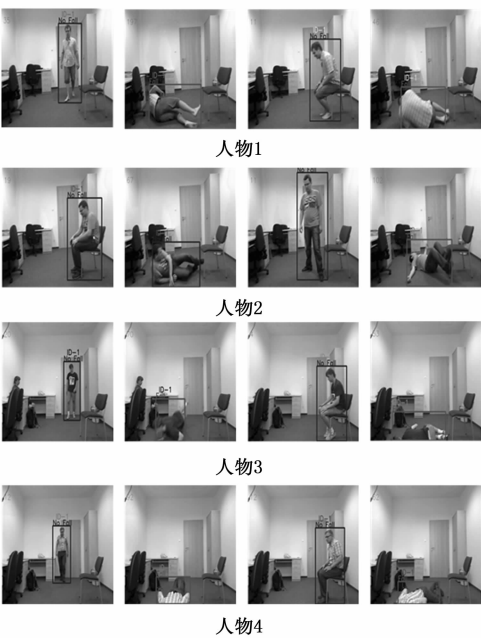


图 10 公开数据集部分测试结果

表 2 算法在 UR Fall 数据集准确率

人物编号	视频序号	检测跌倒帧数	实际跌倒帧数	误检帧数	准确率/%
1	03	29	28	1	96.5
	04	53	53	0	100
2	08	33	32	1	96.9
	11	49	49	0	100
3	13	15	15	0	97.9
	20	49	48	1	100
4	27	12	11	0	91.6
	28	6	6	0	100
平均	—	—	—	—	97.8

状变形和运动信息来区分正常活动和跌倒。文献 [30] 是通过计算平均能量轮廓图像上的 R 变换和 Zernike 矩来构造鲁棒的特征向量，最后进行跌倒与否的分类。本文算法使用同一个数据集对比了以上 3 种算法，在准确率上要高于其他 3 种算法，有更好的检测准确性，主要是本文融合了行人运动特征和行人时序特征，保证了跌倒检测准确率。

表 3 本文算法与其他算法比较

算法	准确率/%
文献[28]	94.00
文献[29]	95.45
文献[30]	96.50
本文	97.80

3.5 跌倒算法自建数据集测试

跌倒，在生活中是一个概率较小的异常行为事件，由于这种行为的罕见性，真实场景下的跌倒行为数据集

的获取非常困难。而公开的数据集大多是基于实验室场景下跌倒, 跌倒画面中仅含有一个行人, 跌倒行为都是刻意模仿的动作, 并且现有跌倒检测算法在实验室场景下有较高的准确率。为了弥补跌倒数据集的不足, 并且验证本文的算法在复杂场景、多行人场景下的跌倒准确性, 本文自建了一个数据集, 采用固定摄像头进行拍摄和网上搜索, 建立了一个复杂场景下的数据集, 该数据集中包含单人跌倒、多人跌倒等多种行为。具体数据集信息如表 4 所示。

表 4 自建数据集类型和数量

视频序列	数量/个
横向跌倒	50
纵向跌倒	50
行走	20
蹲下、坐下、弯腰等	30

自建数据集的部分截图如图 11 所示, 本文自建的数据集画面中不止有一个人, 包含单人站立、单人跌倒、多人站立和多人跌倒等多种行为。



图 11 自建数据集部分截图

在自建数据集的实验结果如表 5 所示, 算法最终取得了 92.91% 的平均准确率。

表 5 自建数据集实验结果

视频序列	数量	平均准确率/%
横向跌倒	50	94
纵向跌倒	50	91
行走	20	100
蹲下、坐下、弯腰等	30	86.66
平均	—	92.91

由表 5 可以看出, 在复杂场景下跌倒检测的准确率会明显低于简单单人场景, 原因是复杂场景下可能会存在遮挡等情况, 行人特征提取相对复杂。在横向和纵向跌倒情况下算法的检测准确率还相对较高, 但是在一些

类跌倒情况下, 算法的准确率会相对较低一些。部分实验结果截图如图 12 所示。



图 12 自建数据集实验结果图

由图 12 可以看出, 在单人场景下算法可以准确检测出跌倒的行人, 在多人场景下算法也可以检测出跌倒的行人, 在场景中有多个行人跌倒时, 算法可以检测出跌倒的行人和非跌倒的行人, 可以满足场景中行人的跌倒检测。

4 结束语

本文在改进的 YOLOv8 目标检测算法的基础上, 提出了一种基于人体运动特征的行人跌倒检测算法, 算法使用了多维运动特征结合一维卷积神经网络实现跌倒检测。在公开数据集 UR Fall 上实现了 97.8% 的跌倒检测准确率, 为了进一步在真实场景下评估算法, 在自行创建的多行人数据集上进行实验, 算法实现了 92.91% 的准确率, 基本可以满足行人跌倒检测, 尽管算法取得了良好的效果, 后期会进一步优化算法, 减少在类跌倒下的误报情况, 提高算法的准确率和鲁棒性。

参考文献:

[1] 宁吉喆. 第七次全国人口普查主要数据情况 [J]. 中国统计, 2021, 5: 4-5.

[2] 赵珍珍, 董彦如, 曹 慧, 等. 老年人跌倒检测算法的研究现状 [J]. 计算机工程与应用, 2022, 58 (5): 50-65.

[3] GOMES M E N, MACêDO D, ZANCHETTIN C, et al. Multi-human fall detection and localization in videos [J]. Computer Vision and Image Understanding, 2022, 220: 103442-103455.

[4] PANNURAT N, THIEMJARUS S, NANTAJEEWARAWAT E. A hybrid temporal reasoning framework for fall monitoring [J]. IEEE Sensors Journal, 2017, 17 (6):

1749–1759.

- [5] 周 坤. 基于可穿戴设备的独居老人跌倒检测系统研究 [D]. 兰州: 兰州交通大学, 2019.
- [6] ALWAN M, RAJENDRAN P J, KELL S, et al. A smart and passive floor-vibration based fall detector for elderly [C] //2006 2nd International Conference on Information & Communication Technologies. IEEE, 2006, 1: 1003–1007.
- [7] 王召军, 许志猛, 陈良琴. 基于红外阵列传感器的人体行为识别系统研究 [J]. 红外技术, 2020, 42 (3): 231–237.
- [8] YU M, GONG L, KOLLIAS S. Computer vision based fall detection by a convolutional neural network [C] //Proceedings of the 19th ACM International Conference on Multimodal Interaction, 2017: 416–420.
- [9] LU N, WU Y, FENG L, et al. Deep learning for fall detection: Three-dimensional CNN combined with LSTM on video kinematic data [J]. IEEE Journal of Biomedical and Health Informatics, 2018, 23 (1): 314–323.
- [10] 周 蕾, 钟海莲, 陈冠宇. 基于注意力机制的改进 YOLOX 行人跌倒检测方法 [J]. 电子器件, 2023, 46 (2): 404–413.
- [11] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric [C] //2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017: 3645–3649.
- [12] ZHENG L, SHEN L, TIAN L, et al. Scalable person re-identification: a benchmark [C] //Proceedings of the IEEE International Conference on Computer Vision, 2015: 1116–1124.
- [13] LIU Z, WANG H, LIU J, et al. Multitask learning based on lightweight 1DCNN for fault diagnosis of wheelset bearings [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 70: 1–11.
- [14] 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.
- [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] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression [C] //Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34 (7): 12993–13000.
- [17] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534–11542.
- [18] LI H, LI J, WEI H, et al. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles [J]. Arxiv Preprint ArXiv: 2206.02424, 2022.
- [19] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [J]. Arxiv Preprint Arxiv: 1704.04861, 2017.
- [20] 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 & Machine Intelligence, 2017, 39 (6): 1137–1149.
- [21] BEWLEY A, GE Z, OTT L, et al. Simple online and realtime tracking [C] //2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016: 3464–3468.
- [22] 王广玉, 窦 磊, 窦 杰. 基于自适应卡尔曼滤波的多目标跟踪算法 [J]. 计算机应用, 2022, 42 (s1): 271–275.
- [23] 成 怡, 朱伟康, 徐国伟. 基于余弦相似度的改进 ORB 匹配算法 [J]. 天津工业大学学报, 2021, 40 (1): 60–66.
- [24] 曹建荣, 吕俊杰, 武欣莹, 等. 融合运动特征和深度学习的跌倒检测算法 [J]. 计算机应用, 2021, 41 (2): 583–589.
- [25] KWOLEK B, KEPSKI M. Human fall detection on embedded platform using depth maps and wireless accelerometer [J]. Computer Methods and Programs in Biomedicine, 2014, 117 (3): 489–501.
- [26] 谢 辉, 师后勤, 齐云霄, 等. 基于注意力机制子网络的时空跌倒检测算法 [J]. 计算机与现代化, 2022 (3): 70–75.
- [27] 曹建荣, 朱亚琴, 张玉婷, 等. 基于关节点特征的跌倒检测算法 [J]. 计算机应用, 2022, 42 (2): 622–630.
- [28] DHIMAN C, VISHWAKARMA D K. High dimensional abnormal human activity recognition using histogram oriented gradients and Zernike moments [C] //2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC). IEEE, 2017: 1–4.
- [29] MERROUCHE F, BAH A N. Fall detection based on shape deformation [J]. Multimedia Tools and Applications, 2020, 79: 30489–30508.
- [30] DHIMAN C, VISHWAKARMA D K. A robust framework for abnormal human action recognition using Transform and Zernike moments in depth videos [J]. IEEE Sensors Journal, 2019, 19 (13): 5195–5203.