老人跌倒检测系统设计与实现

靳冲
重庆工贸职业技术学院，人工智能学院，中国·重庆 408000

摘要：国内人口老龄化程度持续加深，居家养老成为主流养老模式，老年人在室内发生跌倒极易造成骨折、颅内损伤等严重继发性伤害，是老年群体意外伤害的首要诱因。当前主流跌倒检测方案分为纯视频视觉识别与单一穿戴惯性传感检测两类，单一视觉方案易受画面遮挡、光照变化干扰产生漏报，仅依靠加速度传感器则无法区分弯腰、快速落座、下蹲等日常动作，误报、漏报问题难以同时兼顾。针对上述痛点，本文设计并实现一套融合三维人体姿态与腰部惯性传感信息的老人跌倒智能检测系统，依托深度学习与 Transformer 注意力机制完成多模态特征融合判别。系统采集两路异构数据作为模型输入：视觉支路通过 RTMPose3D 提取 COCO17 标准三维骨骼关键点序列，依靠深度维度信息消除平面成像带来的姿态歧义，精准捕捉人字失衡、侧向倾倒、后仰跌倒等典型失稳姿态；穿戴设备支路借助 BeltAccelerometer 腰带式三轴加速度传感器采集人体质心运动时序数据，捕捉跌倒瞬间的运动冲击特征。在数据预处理阶段采用 Trial ID 匹配完成样本粗对齐，再通过时间戳插值实现帧级时序配准，同时对视频、传感数据执行清洗、采样、归一化等操作，剔除无效噪声样本。预处理后分别提取具备物理语义的姿态特征与时域运动特征，搭建双分支编码器完成特征升维映射，再经由 Transformer 模块实现跨模态特征交互融合，最后通过多层感知机输出跌倒分类结果，采用交叉熵损失完成模型迭代训练。消融实验结果表明，相较于单模态基线模型，双模态融合方案整体识别准确率可达 94.28%，跌倒召回率显著提升，能够有效过滤日常活动引发的误报警。整套系统可同时完成本地存储视频离线解析与实时监控流在线推理，运行稳定可靠，能够为居家养老环境下老年人安全监测提供可行的技术方案。

关键词：老人跌倒；人体姿态；数据融合；深度学习

Design and Implementation of an Elderly Fall Detection System

Jin Chong
School of Artificial Intelligence, Chongqing Industry & Trade Polytechnic, China Chongqing 408000

Abstract: China is experiencing continuous deepening of population aging, and home-based elderly care has become the dominant elderly care model. Falls of the elderly indoors may easily cause severe secondary injuries such as fractures and intracranial injuries, constituting the primary cause of accidental injuries among the elderly population. Current mainstream fall detection schemes fall into two categories: pure video visual recognition and standalone wearable inertial sensing detection. Vision-only solutions are susceptible to missed alarms due to occlusions and illumination variations, while acceleration sensor-based approaches cannot distinguish routine movements including bending, rapid sitting, and squatting, making it difficult to simultaneously mitigate false alarms and missed alarms. To address the aforementioned challenges, this paper designs and implements an intelligent fall detection system for the elderly that fuses 3D human pose and waist inertial sensing information. Deep learning and the Transformer attention mechanism are adopted to realize discriminative multi-modal feature fusion. Two heterogeneous data streams are collected as model inputs. For the visual branch, RTMPose3D is employed to extract sequential 3D skeletal keypoints conforming to the COCO17 standard. Depth information is leveraged to resolve pose ambiguity arising from planar imaging, enabling accurate capture of typical unstable postures such as forward imbalance, lateral collapse, and backward falls. For the wearable device branch, a BeltAccelerometer waist-worn triaxial accelerometer collects time-series motion data of the human center of mass to capture impact characteristics at the moment of a fall. In the data preprocessing stage, Trial ID matching is adopted for coarse sample alignment, followed by frame-level temporal registration via timestamp interpolation. Meanwhile, video and sensory data undergo cleansing, sampling, normalization and other

operations to remove invalid noisy samples. After preprocessing, pose features with physical semantics and temporal motion features are extracted separately. A two-branch encoder is constructed for feature dimension-ascending mapping, and a Transformer module is utilized for cross-modal feature interaction and fusion. Finally, a multilayer perceptron outputs fall classification results, and cross-entropy loss is applied for iterative model training. Ablation experiments demonstrate that the proposed bimodal fusion scheme achieves an overall recognition accuracy of 94.28%, with markedly improved fall recall compared with unimodal baseline models, and can effectively filter false alarms triggered by activities of daily living. The entire system supports both offline analysis of locally stored videos and online inference for real-time surveillance streams with stable and reliable operation, providing a feasible technical solution for safety monitoring of the elderly in home care scenarios.

Keywords: Elderly fall; Human pose; Data fusion; Deep learning

0 引言

国内老年人口持续增多，居家养老成为主流模式，室内跌倒是威胁老人健康的常见隐患。老年人骨质脆弱，摔倒后易引发骨折、颅脑损伤，若未能及时发现救治，极易引发各类严重并发症[1]。目前主流监测手段分为视频识别、穿戴加速度传感两类，二者均存在明显短板：纯视觉方案易受遮挡、光照干扰造成跌倒漏检；单一传感设备仅能捕捉运动幅度，难以区分日常动作与真实跌倒，频繁误报降低看护人员使用体验，均难以满足家庭长期监护需求。

视觉三维姿态与腰部惯性传感具备天然互补性，三维骨骼算法可还原人体空间姿态，提前识别失衡前倾、双腿分开等跌倒前兆；腰带加速度传感器紧贴人体重心，能够捕捉跌倒瞬间失重冲击[2]。但两类数据采集频率、存储格式存在差异，原始数据还夹杂画面残缺、传感数值异常、标注错误等噪声，如何完成多源数据规整、挖掘模态关联，并保障设备异常时系统持续预警，是落地应用亟待解决的问题[3]。

针对上述痛点，本文设计融合三维骨骼与腰带加速度的跌倒检测系统。利用 RTMPose3D 提取 17 个人体三维关键点，结合三轴加速度构建双模态数据；通过 Trial ID 分组与时戳插值完成时序对齐，筛选物理语义特征后借助 Transformer 实现跨模态融合判别。消融实验表明，融合模型检测精度优于单模态基线，摄像头、传感器任一信号中断时均可降级工作，对直立、平躺等日常行为识别精准，有效平衡漏报与误报问题，适用于居家养老安全监测场景。

1 人体姿态数据预处理

原始视觉、传感与标注数据格式杂乱、噪声繁杂，维度和时间轴无法直接匹配，直接投入训练会降低特征有效性，削弱模型检测效果。本章通过清洗、时序对齐、归一化等全套预处理流程，搭建规范统一的多模态数据集，支撑后续特征融合与网络训练。

数据集包含图像、视频、传感文本三类素材，覆盖日常活动、各类跌倒姿态。视频文件共 20GB，涵盖行走、落座、平躺与失稳跌倒，用于提取连续时序特征；静态图像补充各类典型姿态。传感数据以 CSV 存储，共计十万条，记录腰带三轴加速度数值，为跌倒判定提供运动量化指标。

数据对齐分为两级处理。依托 URL 携带的 Trial ID 完成样本分组，保证同一段试验的画面与传感数据一一对应，杜绝样本交叉错乱；再通过时间戳插值实现帧级精准匹配。对齐完成后分别提取特征：基于三维骨骼计算躯干倾角、髋膝跨度等人字失衡相关指标，由加速度序列提取幅值、倾角变化、时域波动等运动特征。两类特征经编码器映射后，依靠跨模态注意力完成融合，生成时序统一特征送入分类网络。

本研究选用特征层拼接融合方案，将处理后的视觉特征 F_v 与传感特征 F_s 组合为 $F_f=Concat(F_v,F_s)$。该方式完整保留人体空间姿态与瞬时运动变化，相比决策层融合，可让网络在训练阶段自主挖掘两种模态内在关联，弥补单一检测方式的短板。

数据清洗针对画面残缺、关键点识别失效、传感数值跳变、标签错误四类问题。设置有效关键点最低阈值 12，识别点数不足则剔除该帧；连续三帧以内的关键点缺失，使用线性插值补齐坐标。传感器异常极值借助前后时序窗口平滑过滤，标签错乱片段人工复核校正，保证全部 134 组样本标注准确。降噪处理后，训练过程能够专注学习人体失稳姿态变化规律。

原始图像规格不统一，统一缩放至 640×640 正方形尺寸，缩放维持原图比例，空白区域填充灰度，避免人体形变干扰三维坐标计算。该尺寸兼顾识别精度与边缘设备运算速度。缩放完成后依照 RTMPose3D 预训练参数做像

素标准化，统一像素分布区间，缓解明暗、光照变化对关键点提取造成的负面影响。

2 模型架构设计

网络整体采用分段预训练搭配 Transformer 注意力融合的结构，输入包含监控图像与腰带加速度两类数据。图像先经 RTMDet-m 完成人体检测，截取人物画面边界框，再由 RTMPose3D 解算出 17 组三维骨骼关键点，送入视觉编码器提取表征。

视觉分支 Pose3Dencoder 接收 51 维展平骨骼坐标，搭建两层 MLP 完成特征映射：第一层 Linear 将 51 维升维至 256 维，搭配批归一化与 ReLU 激活；第二层输出 128 维标准化视觉隐特征。传感支路 CSV_MLP_Backbone 以加速度时域特征 Fs 为输入，设计三层 MLP 强化运动特征挖掘，中间层增设 0.2 比例 Dropout 抑制过拟合，最终同样输出 128 维传感特征。

MultimodalFusionTransformer 作为跨模态融合核心，兼容单、双模态两种输入工况。当图像或传感数据中断时，采用可学习 Missing Token 填充空缺序列，保证输入维度稳定。两路 128 维特征堆叠形成序列，叠加模态编码区分视觉、传感来源，送入单层 Transformer Encoder 完成自注意力交互，多头注意力设置为 4 头。融合后的序列展平后，经由两层线性层与 ReLU 组成的 MLP 输出分类 Logits。

模型以二分类交叉熵作为唯一损失，损失表达式为 $L_{total}=L_{CE}$(Fall Detection)。训练阶段不断最小化该损失，持续更新网络全部参数，缩小跌倒、日常动作预测结果与真实标签之间的误差。

3 系统设计与实现

3.1 结构功能设计

面向居家老人长期监护场景，本系统以高检测精度、持续稳定运行为核心设计目标，整体拆分为用户鉴权、多源数据预处理、多模态智能推理三大解耦模块，各模块职责划分清晰，可单独迭代升级，能够承载多路监控视频与穿戴传感数据并发传输压力。

用户鉴权模块负责全平台访问权限管控，划分管理员、看护人员多级使用身份。依托账号密码校验与会话令牌机制隔离功能操作、数据查看权限，拦截外部非法访问，守护监控画面、人体运动传感记录等隐私信息；搭配会话超时机制，平衡数据安全与日常操作便捷性。

数据接入预处理模块作为所有异构数据的统一入口，同步接收实时视频流与腰带加速度传感器文本数据。模块内部完成数据解析后，依次执行 Trial ID 样本分组、时间戳时序配准、无效帧过滤、传感数据异常平滑、图像尺寸统一缩放与像素标准化等操作。模块内置多线程缓冲队列缓解多设备同时接入带来的数据冲击，输出时序规整、格式统一的标准化特征，为 AI 推理提供合规输入。

多模态智能推理模块是系统核心运算单元，接收预处理后的视觉与惯性特征，依靠双分支编码器、Transformer 跨注意力融合网络、分类层完成人体行为判别。模块兼容实时监控流在线推理、本地视频离线解析两种工作模式；当摄像头遮挡、传感器断连导致单路数据丢失时，自动启用降级推理策略，保障监护服务不会直接中断。

三大模块借助消息队列搭建流水线式数据链路，单个模块故障不会造成整套系统瘫痪，既适配居家环境 7×24 小时不间断监测需求，也为后续算法优化、新增监测功能预留拓展空间。

3.2 业务流程设计

新用户进入系统页面后，填写账号、两组一致密码完成注册申请。后端检索数据库判断账号是否重复，同时校验两次输入密码是否匹配；校验全部通过后将加密用户信息存入数据表并返回注册成功提示，存在信息冲突则引导用户修改后重新提交[4]。

平台设置强制登录准入规则，所有包含监控画面、检测数据的核心页面均需完成身份核验。登录时提交的账号密码与库内加密信息比对，鉴权通过方可进入监护操作主页，未通过的访问请求直接拦截，从访问源头保护居家影像隐私。

账号登录成功后即可启动跌倒识别功能，整套运算流程分为四步：

（1）多源数据同步采集：系统同步读取监控图像帧与腰带三轴加速度时序数据，依靠统一时间戳完成两组数据一一配对；

（2）双分支特征提取：视觉分支采用 RTMPose3D 提取人体三维骨骼关键点，通过 MLP 完成特征升维编码；传感分支先对加速度原始序列滤波降噪，再经独立多层感知机编码运动特征；

（3）跨模态特征融合运算：若运行中出现画面遮挡、传感信号中断，系统通过 Missing Token 机制补全缺失序列；完整双模态特征输入 Transformer 自注意力结构，完成跨模态信息交互并自适应分配特征权重；

（4）行为分类与告警输出：模型输出直立、静坐、平躺、跌倒四类行为判定结果，一旦识别跌倒事件，系统即时向前端推送告警提示，提醒看护人员及时介入处理。

3.3 系统实现

整套系统分为前端交互界面、Web 后端服务、AI 推理核心、数据持久存储四大部分，各模块分层解耦，兼顾轻量化部署与实时跌倒检测需求。

前端承担人机交互工作，负责页面展示、监测画面渲染与检测请求发送。本项目摒弃 Vue、React 这类重型单页框架，选用服务端渲染搭配原 JS 的轻量化开发模式。页面依托 Jinja2 模板引擎完成渲染，登录、监控面板等页面模板由 FastAPI 后端直接下发。页面视觉采用原生 CSS3 开发，运用玻璃拟态设计风格，搭配 FontAwesome 图标优化视觉表现。交互逻辑全部依靠原生 JavaScript 实现，DOM 渲染、页面事件、AJAX 数据传输均调用 fetch 原生接口，全程不引入 Webpack、Vite 等打包工具，降低部署资源占用。同时借助媒体查询完成响应式布局，电脑、平板等不同设备均可正常访问。

后端划分为 Web 服务层与 AI 算法调度层，承担请求接收、身份校验、业务流转工作，整体基于 Python 生态搭建。Web 框架选用异步高性能框架 FastAPI，搭配 Uvicorn ASGI 服务器承载并发访问。账号安全采用 JWT 鉴权方案，用户密码通过 passlib 完成哈希加密处理，身份令牌存放于 HttpOnly Cookie，有效规避跨站脚本窃取凭证的风险。

AI 推理模块是系统运算核心，单独封装为 Python 程序包供后端调用，底层依托 PyTorch 深度学习框架，结合 OpenCV、MMPose、MMDetection 视觉工具库完成图像处理与姿态解析。模型流水线分为三级，RTMDet 目标检测框选出画面内人体区域，RTMPose3D 提取 17 个标准三维骨骼关键点，多模态融合模型结合骨骼坐标与腰部加速度传感数据，区分跌倒、行走、静坐、平躺四类人体动作。程序内部封装模型初始化加载、单帧完整推理、动作分类预测三类核心函数，简化后端调用流程。

系统完整业务数据流如下：用户在前端上传图片、视频或开启实时摄像头，通过 AJAX 将图像数据流提交至后端接口。服务端先校验 Cookie 内 JWT 令牌完成身份放行，区分上传文件与实时视频流两类数据解析方式，再将图像送入 AI 推理包运算。推理流程依次完成人体框选、三维关键点提取、行为判定，输出动作标签、预测置信度与骨骼三维坐标。后端调用可视化工具在原图绘制骨骼线条与预警标识，编码为图像流返回前端渲染；若识别结果为跌倒，同步向数据库写入告警记录。

数据存储选用轻量化嵌入式数据库 SQLite，借助 SQLAlchemy 实现 Python 对象与数据表的映射。数据库设置两张核心表单，User 表存放账号、邮箱与加密密码，FallDetectLog 记录表存储每条监测任务的类型、识别结果与时间戳。数据库配置文件独立管理，通过开启多线程兼容参数，适配 FastAPI 异步并发运行环境。如图 3.1 老人跌倒检测界面图所示。



图3.1 老人跌倒检测界面图

4 结语

本文面向居家养老场景下老人跌倒监测需求，搭建一套融合三维人体姿态与腰带惯性传感的多模态检测系统。文中完成多源异构数据标准化预处理，通过两级时序对齐、噪声清洗、图像归一化等操作构建高质量数据集；同时设计双分支 MLP 编码器搭配 Transformer 跨模态融合网络，依靠可学习缺失 Token 实现模态失效下的稳定推理。整套系统采用轻量化前后端架构，划分鉴权、数据预处理、智能推理三大模块，依托 FastAPI 完成异步数据交互，搭配嵌入式数据库存储用户信息与跌倒告警记录。对比实验证实，融合模型相比单一视觉、单一传感方案识别效果更佳，既能降低日常动作带来的误报，也可在摄像头遮挡、传感器断连时降级运行。本方案仍存在一定局限，针对缓慢失衡类跌倒样本识别能力尚有提升空间。后续研究可扩充多样化跌倒数据集，优化跨模态注意力交互逻辑，并完成模型轻量化改造，部署至低性能边缘设备，进一步完善居家环境下老人安全监护方案。

参考文献：

[1] 王建霞，田楠楠，李璇等. 基于改进 YOLOv12 的多场景老人跌倒检测[J]. 河北工业科技，2026, 43 (4): 319-327.

[2] 殷贤涛. 基于 YOLOv8-pose 与时空图卷积网络的跌倒检测算法研究[D]. 贵州师范大学，2026.

[3] 邹一唱，王雅婷，赵贞莲等. 基于 YOLOv10 和 OpenPose 的老人跌倒检测综述[J]. 数字技术与应用，2026, 44 (1): 178-180.

[4] 吴宗胜，黄奕，王芝怡等. 基于 YOLOv11 跌倒检测系统的设计与实现[J]. 咸阳师范学院学报，2025, 40 (6): 24-29+35.

基金项目：重庆工贸职业技术学院校级科研项目《基于多模态数据融合与注意力机制的 Transformer 模型在 3D 人体姿态估计中的应用研究》(ZR202430)。

作者简介：靳冲（1989.10-），男，湖北松滋人，硕士研究生，高级工程师、讲师，研究方向：计算机视觉，机器学习。