



基于 OpenPifPaf 的多摄像头多人实时跌倒等  
异常行为识别预警应用研究  
技术报告

参赛队名：\_\_\_\_\_ 华南“感”知 \_\_\_\_\_

队 长：\_\_\_\_\_ 陈冠名 \_\_\_\_\_

联系方式：\_\_\_\_\_ 15815509164; chenantony@163.com \_\_\_\_\_

组员名单：\_\_\_\_\_ 王洋、陈铭泓 \_\_\_\_\_

指导老师：\_\_\_\_\_ 刘寿强、汤庸、陈卫东 \_\_\_\_\_

参赛单位：\_\_\_\_\_ 华南师范大学人工智能学院 \_\_\_\_\_

\_\_\_\_\_ 物理与电信工程学院、计算机学院 \_\_\_\_\_

作品类别：应用

## 摘要

近年来，公众安全事故频发，社区医院等区域老人小孩跌倒摔伤屡见不鲜，公共场合踩踏，暴力恐怖袭击等异常行为时有发生，此类跌倒等异常行为对民众生命财产安全构成了极大的威胁，急需及时监测预警。本项目旨在用智能技术手段对医院社区或其他公共场合跌倒等异常行为进行监测预警，及时发现救助，并防患于未然。

本项目在研究近年来人体行为识别实际应用案例的基础上，通过案例展示相关技术和 AI 识别人体行为的价值，之后对该项目的需求进行分析，提出设计方案、实现方案和预期目标等内容，并详细的介绍所用技术的理论、流程和展示。首先使用 OpenPifPaf 对多摄像头输入视频进行姿势估计，通过 LSTM 提取姿势信息中的时间和空间特征以预测跌倒等异常行为并进行预警，支持多摄像头和多人实时检测。同时此项目致力于运用步态识别技术确认异常行为主体的身份信息，基于 Gaitset 进行身份辨别，并在已有数据集上取得较好的结果，达到了对跌倒等异常行为进行监测预警的目的。最后，总结了该项目的特色、创新和转换前景，展望了人体行为识别技术未来可能的发展趋势及其应用。

**关键词：**跌倒等异常行为；行为识别；步态识别；监测预警

# 目录

- 1.选题的目的及意义 .....1
- 2.国内外研究现状 .....1
  - 2.1 人体动作捕捉与识别 .....1
  - 2.2 行为识别.....1
  - 2.3 步态识别.....2
  - 2.4 实际案例.....2
- 3.应用价值 .....4
  - 3.1 行为识别.....4
  - 3.2 步态识别.....4
- 4.设计方案 .....5
  - 4.1 需求分析.....5
  - 4.2 预期目标.....5
- 5.实现方案 .....6
  - 5.1 系统功能图 .....6
  - 5.2 跌倒检测流程图 .....6
  - 5.3 多摄像头中的主体辨别及跟踪.....7
  - 5.4 LSTM 网络模型 .....8

5.5 Gaitset 网络模型 .....	10
6.运行效果.....	14
6.1 视频输入检测实例 .....	14
6.2 摄像头检测实例 .....	15
7.特色及创新 .....	16
7.1 创新性 .....	16
7.2 优越性 .....	16
7.3 市场价值.....	17
7.4 转化前景.....	17
8.总结与展望 .....	17
8.1 总结项目已有成果.....	17
8.2 展望.....	18
8.2.1 近期项目规划.....	18
8.2.2 长远项目规划.....	19
9.参考文献.....	19

# 1.选题的目的及意义

伴随着信息技术的快速发展，人体行为识别技术逐渐被引入到各领域，如安防监控、运动分析、医学辅助诊断和智能人机交互等方面。而人的姿态信息，即人的动作种类的信息，是人体行为的一种。而近年来，恐怖袭击频发，对民众生命安全造成了极大的威胁，为有效预防恐怖袭击的发生，提前检测出是否存在危险行为，有效预防可能发生的危险，提升公众安全，我们提出了能识别并预警人体摔倒等异常行为的项目。

## 2.国内外研究现状

### 2.1 人体动作捕捉与识别

在 2019 年，冲击香港立法会、暴力袭击警察的激进分子几乎都将口罩、头盔作为隐藏身份的工具，这就为香港警方执法带来不小的难度。在当时有港媒建议，应从内地引入仅通过行走步态，就能辨识个人身份的技术，令企图犯罪的人失去“心理保护伞”。

为此，一家内地人工智能企业在北京发布全球首个步态识别互联系统，即使目标人物将脸遮住，该系统也可以通过走路姿态辨认出来。内地科技公司这一步态识别系统通过分析一个人步行时，从身体轮廓到手臂摆动，再到脚尖朝内还是朝外等数以千计的特征，建立数据库，可以识别 50 米以外的人，即使这个人遮挡脸部或背对镜头，一样可以被认出。据这些科技公司介绍，和人脸识别相比，步态识别拥有远距离、全视角等诸多独特优势，准确率逐渐提高。该系统能基本满足公共安全领域的基础需求。据了解，这套系统已陆续在湖北、广东、上海等地应用。

### 2.2 行为识别

人体行为识别是计算机视觉领域中极为重要的研究方向之一，可以定义为计算机采用算法对含有人体动作的视频进行分析和理解。

人体行为识别本身有着广泛的应用场景：在校园里，人体行为识别可以根据人员的聚集情况及时发现意外的发生，还可以在运动场上监测学生的步态提供专业性建议；在医疗机构，可以用于监测病患的身体情况，对摔伤跌倒等情况进行监控；在公共场合，可以用于识别是否有

打架等暴力行为的发生。除了在生活工作中充当隐形保护者，人体行为识别还可以用于人机交互，如体感摄像头利用人体行为识别算法实现了游戏中虚拟人物跟随现实玩家的动作做运动的功能。用于视频理解方面，借助人体行为识别根据视频中人物的行为类别对视频进行切分和分类，解决海量运动视频的检索问题。

视频中人体行为识别是计算机视觉领域一个具有挑战性的热门课题，这个问题包含了视频中人体定位和人体行为分类两个部分。行为识别在视频理解中起着重要的作用，为视频分割和视频检索提供了新的方法。与图片相比，视频数据增加了时序上的信息，本质上视频是一组有序的视频帧，每一帧都可以当做是一张图片。如何有效把握视频中包含的时域信息是研究的重点。

## 2.3 步态识别

生物识别技术是指利用生物内在的物理或行为特征来进行身份鉴别，除了人脸、指纹、虹膜、DNA 等常规特征以外，步态作为一种远距离非接触式的生物特征也逐渐引发大量关注。目前常见的研究，根据数据类型的差异可以大致分为：基于步态轮廓的识别方法，基于人体骨骼点的识别方法和基于可穿戴传感器的识别方法。

当前步态识别通常采用的是基于步态轮廓的方法，为解决角度这一最大的影响因素，通常有两类解决方法：1) 角度无关特征的提取，2) 基于学习的匹配方法。角度无关特征的提取可以利用角度转换模型（VTM）等方法将不同角度下的步态特征投影到同一角度空间，也可以直接利用深度学习的方法，通过训练来提取角度无关的步态特征；另一方面，基于学习的匹配方法旨在通过优化模型的辨别能力及最大化类间距离、最小化类内距离，包括：基于排序的支持向量机，线性判别分析，多视角判别分析，联合贝叶斯模型等机器学习的方法及生成的多视角费舍尔分析（GMMFA）。

## 2.4 实际案例

### 2.4.1 行为识别

纽约警察局开发了一种新的模式检测计算机软件，以帮助抓捕犯罪分子。纽约警察局一直在使用一种新的内部开发的软件，它识别某些行为模式，并将它们与数千个报告的盗窃，盗窃罪和抢劫案的数据库进行比较。该软件，更好地称为“Patterizr”，是一组学习算法，取自

10 年的警方数据，扫描犯罪信息。Patternizr 从诸如时间，进入方法和使用的武力类型等犯罪中获取信息，并试图发现可用于帮助识别嫌疑人的行为模式。

确定犯罪模式是警方调查工作的重要组成部分。识别模式的传统方法涉及大量的工作和人力，这会占用分析师和侦探的宝贵时间。由于计算机算法，Patternizr 在很短的时间内为侦探提供了一份潜在嫌疑人名单。这大大节省了侦探的时间，这可以让侦探更好地利用他们的时间捕捉罪犯。

在 Patternizr 的帮助下，纽约警察局能够抓住一个试图抢劫 Home Depot 的注射器的小偷。该算法设法获取数据，表明在一个不同的 Home Depot 之前几周被一名挥舞着注射器的男子抢走。然后，官员们能够将这两个案件放在一起。“因为 Patternizr 在算法中获得了这些关键细节，它带回了我不知道的其他区域的投诉，”负责 Home Depot 案件的布朗克斯犯罪分析师 Rebecca Shutt 说。

## 2.2.4 步态识别

2018 年，甘肃省一家金店被抢劫，监控摄像头距离嫌犯大概有 40-50 米远的距离，而且他刻意挡住了面部。但在步态识别技术面前也纤毫毕现，很快被系统检索出身份。

案例使用了银河水滴步态识别互联系统——水滴慧眼，集步态建库、步态识别、步态检索、步态比对、大范围追踪等功能于一体，支持海量历史视频与实时视频瞬间检索与定位，支持事前预警、事中报警和事后追踪，支持地图布控及地图轨迹追踪。水滴慧眼使用的步态底库建构方法为步态检索智能一体机——水滴神鉴，这是一款便携式的步态识别安防专用设备，集步态识别算法、软件和硬件为一体，主要依靠步态识别迅速检索目标嫌疑人。现在模拟一下它的工作流程：一个案件中，犯罪现场的监控摄像头只捕捉到了犯罪嫌疑人的作案过程，视频比较模糊，看不清人脸，或人脸被刻意遮挡，但是嫌疑人的步态信息是清晰的，那么就可以将嫌疑人案发现场的步态信息进行注册，产生步态底库，之后再调取相关的视频，系统对相关视频中的人物步态信息进行快速分析，并与底库进行快速的步态信息比对，从而检索到与嫌疑人相似的目标人物，并通过多种手段进一步确认嫌疑人信息。目前，以步态信息检索嫌疑人，一小时视频，水滴神鉴最快可在 1 分钟内检索完毕，大大改变了此前人工办案的现状，节省了大量的人力和时间，并且，步态识别技术比普通的办案人员更加精准，识人的效率和准确度更高。

目前，步态识别技术已广泛应用于平安城市(安防、车站、机场、博物馆、学校、景区等)、重要基础设施(核电站、发电站、石油石化基地等)等场景。步态识别技术弥补了传统生

物识别技术怕伪装、识别距离近等缺点，成为安防领域最高效的生物识别技术。

据了解，银河水滴发布了全球首个步态识别互联系统，该系统集步态建库、步态识别、步态检索、大范围追踪等功能于一体，实现海量摄像机下步态识别的实时智能互联：支持上百万路摄像机实时并发，支持海量历史视频与实时视频瞬间检索与定位，支持事前预警、事中报警和事后追踪，支持地图布控及地图轨迹追踪，为平安城市建设提供有力保障。

## 3.应用价值

### 3.1 行为识别

医院、养老院以及十字路口等人员密集场合经常出现残障人士及老人或其他特殊人群等突然跌倒无人发现及时救助的问题。针对这一问题，我们提出的方案可以及时发现异常情况并上报工作人员，可以很好解决这一痛点。本项目的人员监护系统在多场所都有着中有着广阔的应用前景，如在医院门诊、住院大楼的安防监控，病房、监护室的患者 24 小时监护；养老院等一些需要长时间关注人员情况的场所；以及交通监控下对一些异常情况的检测。

通过以上人员信息，本项目有以下的应用场景：

（1）分析需要监护场所的人员密集程度，以减少疫情期间人员的密集接触。

（2）分析监护场所的人员行为，如是否有人跌倒或者躺倒，如果发现该情况，该系统可以及时报警，监护人员可以根据视频迅速定位到人员位置进行救助。

（3）分析监护场所的人员行为，如行走和坐立状态，判断人员的行为是否正常合理。

对于老人、儿童、病人这一类特定用户，一个好的产品的推广首先要求其操作简单，而本产品即将引入的移动端 app 基本不需要老人进行任何操作，只需要用户开启相应的软件即可。而对于室内场景（如医院、养老院），常常会发现一些用户的异常行为，而一旦发生异常行为，所带来的后果是非常严重的。因而对于在发生异常行为第一时刻通过监控系统的采集和后台分析对事件进行响应是至关重要的。

### 3.2 步态识别

对于检测目标，利用手机或监控摄像头进行拍摄，并对画面中的目标进行步态识别，识



别目标对象，并对其步态进行分析，提出科学的指导意见，可以在多方面良好应用。例如，在医疗机构，对患有腿部疾病的病患进行分析，帮助医生诊断病情，如偏瘫、截瘫、小脑性共济失调都有其对应的特殊步态；在运动场，对运动员的步态进行分析，提出训练指导意见；在社会场景，如对罪犯进行步态识别，标记逃犯位置并实时进行监控。步态识别可以良好的应用于日常生活，提高工作生活质量。

## 4.设计方案

### 4.1 需求分析

随着人类社会的不断发展与生活质量的不断提升，人们生活空间在不断的扩大和延展，公共及私人场所的领域也在随之发展，遇到各种突发情况的概率在不断增加，尤其是在公共场所，由于其监控难度较大，人口密集。通过简单的监控和简单依靠值班人员的坚守，想要真正做到全时段及时发现危险或不正常行为并行动还有较大难度因此，选用智能监控系统已经成为当前社会解决这一根本问题的必由之路。

### 4.2 预期目标

形式：一个基于多个摄像头的多人实时行为检测模型，可连接至常规摄像头（公共），对异常行为进行识别并发出预警，有效防范踩踏、自杀等威胁生命安全的事件。

效益：对各种异常行为进行检测，可让当事人亲属及医院、公安等相关人群及时得到事件信息，尽快处理突发事件，保护生命财产安全。

## 5.实现方案

### 5.1 系统功能图

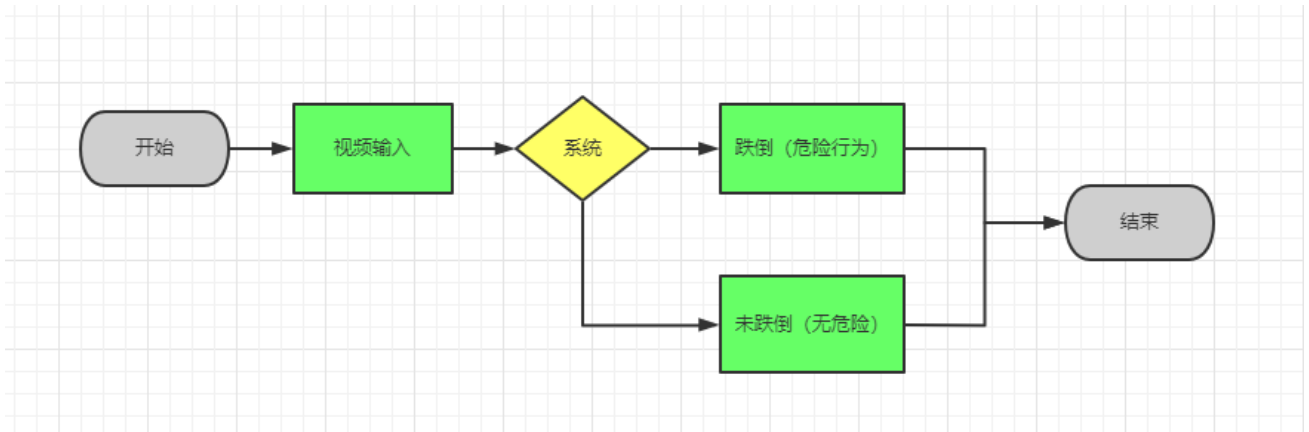


图 1 系统功能图

系统功能思路：以视频流作为输入，经过系统处理后，输出危险行为预警。

### 5.2 跌倒检测流程图

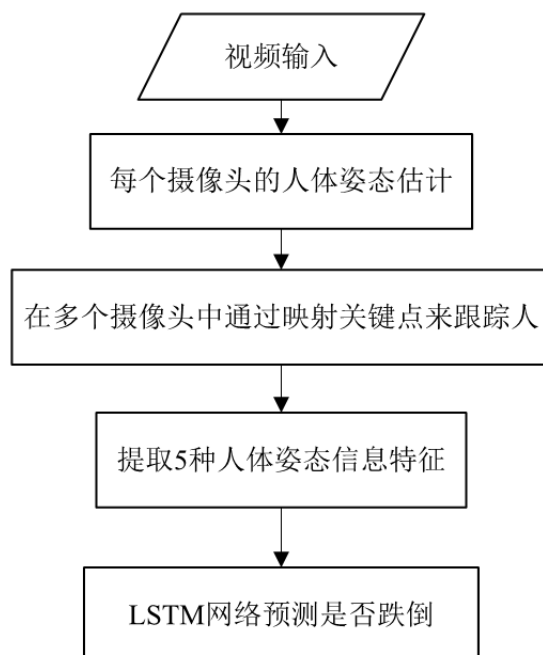


图 2 跌倒检测流程图

此流程的输入是来自每个摄像机的视频。我们将每个视频分析为一系列图像并依次处理帧。使用公开可用的自底向上人体姿势检测器 **OpenPifPaf** 从每一帧中提取人体姿势的关键点。它的优势在于与其他最先进的算法相比，它有着低分辨率图像的性能以及更低的硬件要求。对于从每个相机获得的每一帧，都会获得一个关键点集列表，每个关键点集代表一个人的关节位置。

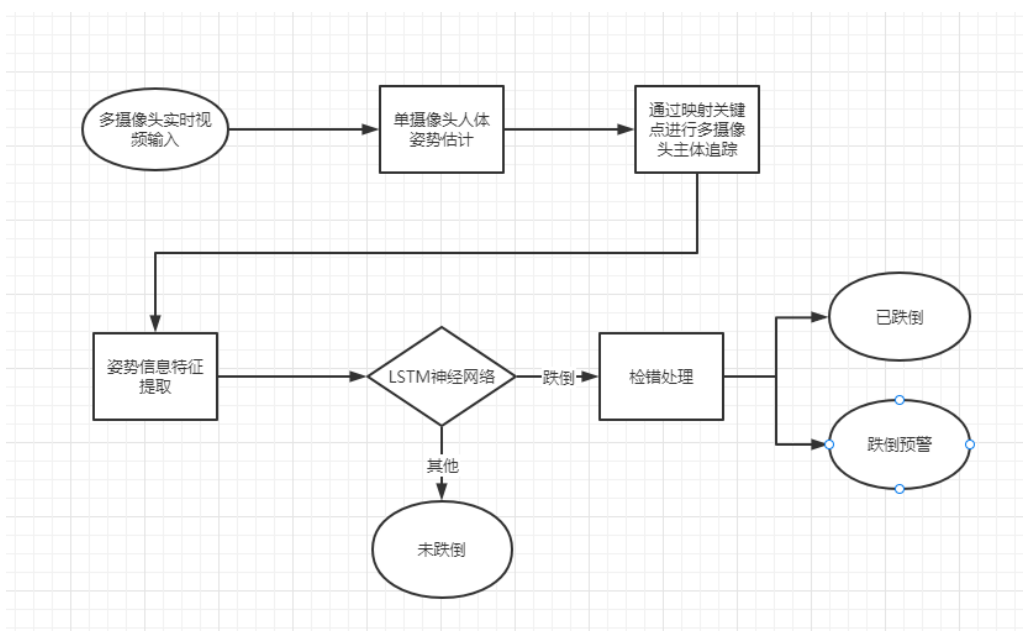


图 3 算法流程概述

从上数姿态信息从中进行特征提取，经过 LSTM 神经网络处理，会对信息主体进行一个大致的判断（跌倒或未跌倒）。紧接着，有一个对跌倒情况的跟踪处理，判断信息主体所处状态为已跌倒或即将跌倒，最终进行跌倒等危险行为预警。

### 5.3 多摄像头中的主体辨别及跟踪

下图描述了用于对象识别和跟踪的算法。在这项工作中，我们只使用两个摄像头。然而，该方法可以推广到多个相机。在第一阶段结束时，我们获得每个相机的关键点集列表，每个关键点集对应不同的人。在帧中有多个人的情况下，如果给我们一个摄像机的任何关键点集，我们不知道其他摄像机的哪个关键点集对应于同一个人。在这个阶段，我们使用两步骤程序创建关键点集之间的映射。首先，得到关键点集列表，每个集都映射到上一个人的关键点集框架。我们使用连续帧中关键点之间的空间欧几里得距离作为距离度量。随后，我们使

用定制的 Gale-Shapley 算法将第一个摄像机中的特定于个人的关键点集映射到第二个摄像机中具有最高相似度的关键点。这种匹配方法与 HSV（色调、饱和度，值）颜色直方图相关。

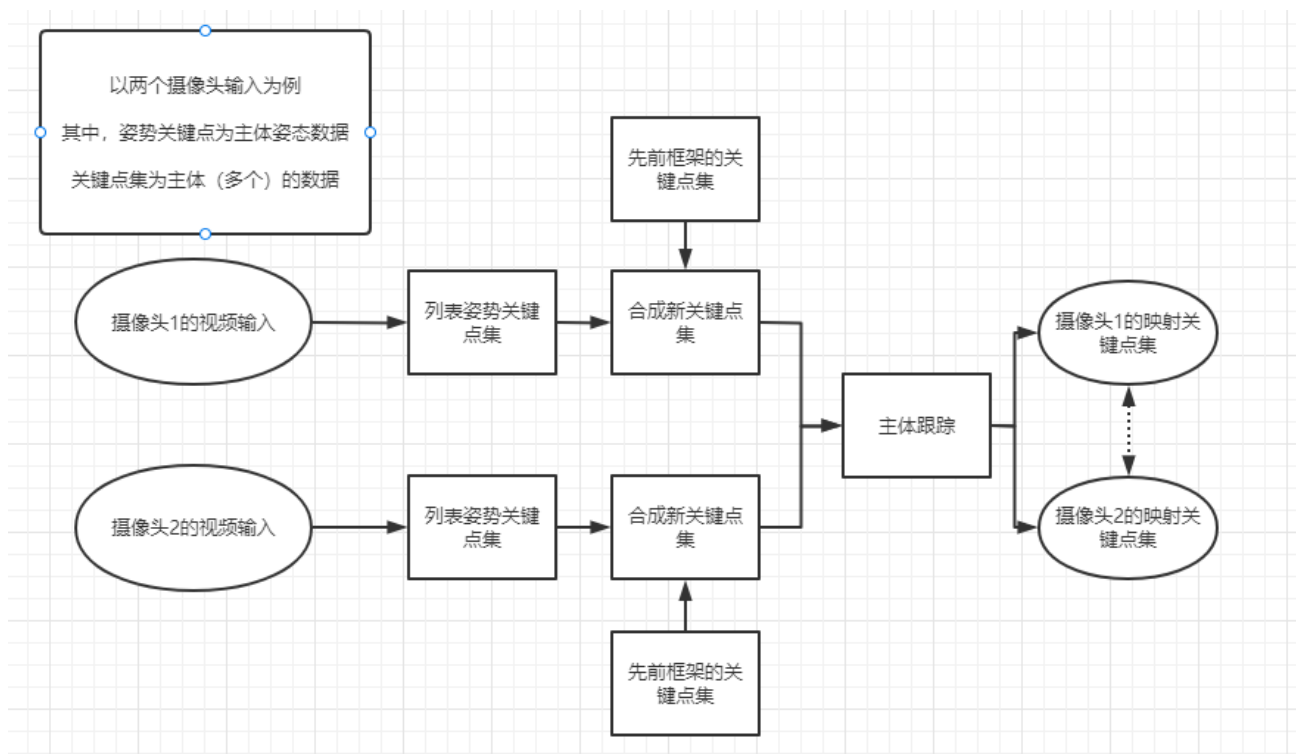


图 4 多台摄像机中对象识别和跟踪的概念（以两台摄像机为例）

## 5.4 LSTM 网络模型

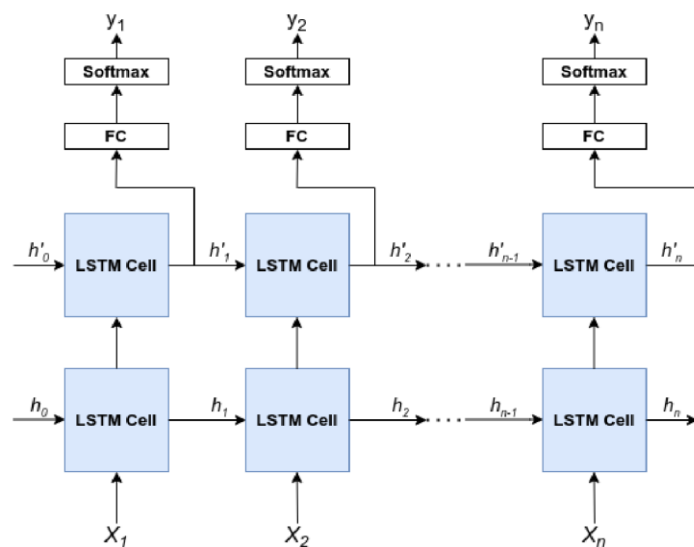


图 5 LSTM 神经网络

一旦我们从两个摄像头获得每个人的特征，我们就可以使用如图所示的 LSTM 神经网络架构来预测一个人是否跌倒。在每个时间步长  $n$ ，LSTM 网络使用从当前帧中提取的特征  $X_n$  预测此人是否跌倒。在部署期间，我们为每个观察到的人存储隐藏状态。当观察到一个新帧时，对于每个人，他们的特征向量  $X_n$  连同他们的隐藏状态  $(h_{n-1}, h0_{n-1})$  被馈送到 LSTM 网络，然后 LSTM 网络给出更新的隐藏状态  $(h_n, h0_n)$  以及包含每个活动的概率向量  $y_n$ ，即该人正在执行该活动的概率。如果“跌倒”活动的概率在预测中最高，则该帧检测到跌倒。

网络以  $(X, Y)$  形式的数据进行训练，其中  $X$  是从数据集中在两秒内获取的连续帧中的一个人的特征集， $Y$  是活动标签 (Tbl. 1) 最后一帧。因此，该模型采用人的特征向量  $X_n$  并学习预测 7 个不同类别的概率分布，一个是“Fall”类别，其余是分别表示的“No Fall”类别。当模型从视频剪辑的第一帧接收特征时，模型的隐藏状态用零向量初始化。

代码实现：

```
class LSTMModel(nn.Module):
    def __init__(self, input_dim=5, h_RNN_layers=2, h_RNN=256, drop_p=0.2, num_classes=1):
        super(LSTMModel, self).__init__()
        self.input_dim = input_dim
        self.h_RNN_layers = h_RNN_layers    # RNN hidden layers
        self.h_RNN = h_RNN                  # RNN hidden nodes
        self.drop_p = drop_p
        if h_RNN_layers < 2:
            drop_p = 0
        self.num_classes = num_classes
        self.LSTM = nn.LSTM(
            input_size=self.input_dim,
            hidden_size=self.h_RNN,
            num_layers=h_RNN_layers,
            dropout=drop_p,
            batch_first=True,    # input & output will has batch size as 1s dimension. e.g. (batch, time_step, input_size)
        )
        self.fc1 = nn.Linear(self.h_RNN, self.num_classes)

    def forward(self, x, h_s=None):
        # print('forward started')
```

```

self.LSTM.flatten_parameters()
RNN_out, h_s = self.LSTM(x, h_s)
""" h_n shape (n_layers, batch, hidden_size), h_c shape (n_layers, batch, hidden_size) """
""" None represents zero initial hidden state. RNN_out has shape=(batch, time_step, output_size) """

# FC layers
out = self.fc1(RNN_out[:, -1, :]) # choose RNN_out at the last time step
return out, h_s

```

## 5.5 Gaitset 网络模型

网络结构:

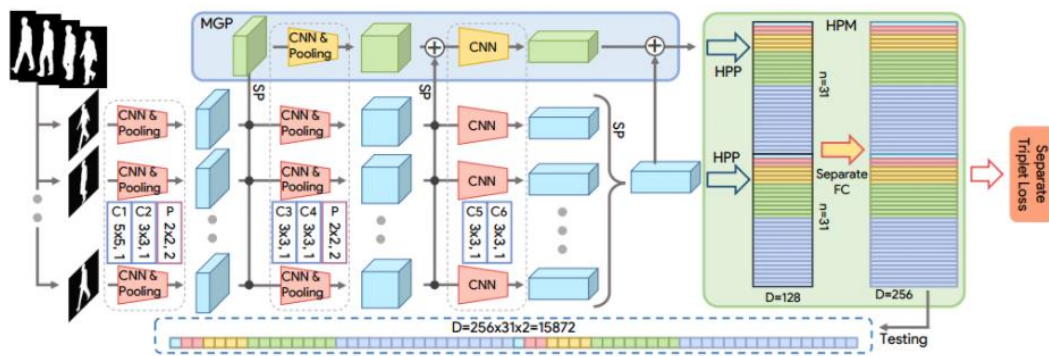


图 6 Gaitset 网络结构

网络的输入为 5 个维度  $\text{num\_of\_people} * \text{frame\_num} * \text{channel} * \text{height} * \text{width}$ 。其中， $\text{batch\_size} = \text{num\_of\_people} * \text{frame\_num}$ 。训练的时候可以输入多个人的多张图，测试的时候，输入一个人的多张图，也就是说测试的时候， $\text{num\_of\_people} = 1$ 。

上图中连续输入的图片是表现在 batch 维度的操作，虽然图中画了 3 行红色和蓝色相间的这样的操作，但是，实际上，只有一个分支，通过 batch 维度实现，也就是权值共享的意思。其中，黄色部分分别表示，卷积+卷积+pooling 操作，卷积+卷积操作。粉色部分每行分别表示，卷积+卷积+pooling 操作，卷积+卷积+pooling 操作，卷积+卷积操作。

MGP 模块表示 Multilayer Global Pipeline (MGP)，主要融合不同层的输出特征，增加模型的不同感受野的信息。

SP 表示 Set Pooling

$$\begin{aligned} z &= G(V) \\ V &= \{v^j | j = 1, 2, \dots, n\} \\ G(\{v^j | j = 1, 2, \dots, n\}) &= G(\{v^{\pi(j)} | j = 1, 2, \dots, n\}) \end{aligned}$$

Z 表示集合维度的特征，一个人表示一个集合，

V 表示帧维度的特征

$\Pi$  表示任意的组合

$$\begin{aligned} G(\cdot) &= \max(\cdot) + \text{mean}(\cdot) + \text{median}(\cdot) \\ G(\cdot) &= 1\_1C(\text{cat}(\max(\cdot), \text{mean}(\cdot), \text{median}(\cdot))) \end{aligned}$$

SP 操作就表示为在 set 维度，也就是第二维 frame\_num 维度进行全局的 max，mean，median 操作，然后再进行 1\*1 卷积操作。后续又对 SP 模块引入 attention 机制，

$$f'_{s,t} = \text{maxpool}(z_{s,t}) + \text{avgpool}(z_{s,t}),$$

HPM 模块表示最终输出的特征向量，MGP 模块和主网络模块分别经过 HPP 操作，宽度和高度这两个维度被压缩成 1, 2, 4, 8, 16 这样的 1 个维度，像下图的第一列的蓝色的  $h*w*c$  变成了第二列的蓝色的  $1*c$ ，红色的  $h*w*c$  变成第二列的  $2*c$ 。两个这样的 31 维度 concat 成 62 维度。channel 维度不变，为  $D=128$ ，然后再经过一个全连接操作，变成  $D=256$  维度。

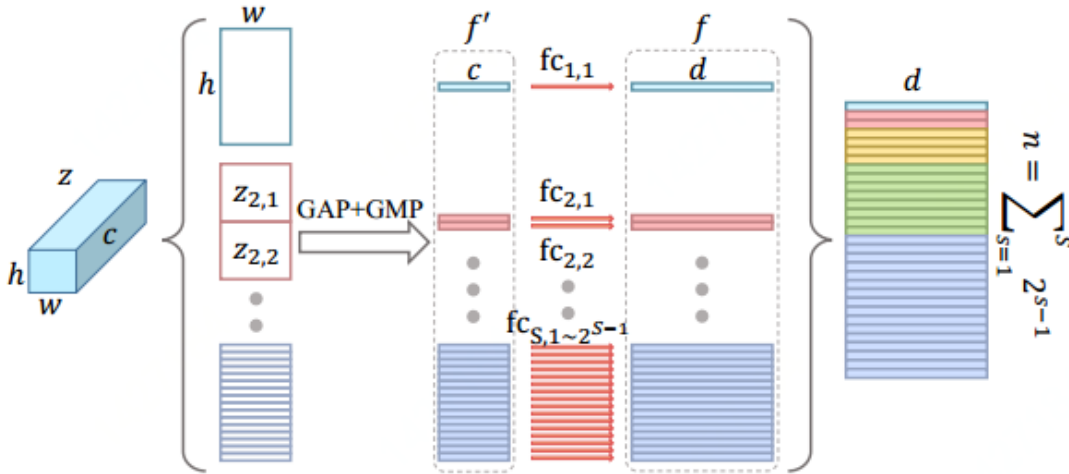


图 7 全链接 HPP 示意图

Batch 维度原本为  $\text{num\_of\_people} * \text{frame\_num}$ ， $\text{frame\_num}$  维度经过 SP 操作变为 1，被挤掉后，batch 维度就是  $\text{num\_of\_people}$ ，也就是说训练的 batch 中有几个人，就输出几个对

应的  $62 \times 256$  的特征向量。测试的时候，因为只有 1 个人输入，所以输出 1 个  $62 \times 256$  的特征向量。这个特征向量可以理解为，1 个人对应 62 个全连接层的特征向量，每一个特征向量的维度为 256 维。

对于训练部分，则计算每个特征向量组成的矩阵和自己的转置的自相关矩阵，也就会得到  $\text{anchor}$  和  $p$ ， $\text{anchor}$  和  $n$ ，相当于对  $\text{batch}$  内的所有样本都进行了组合。

代码实现：

```
class BasicConv2d(nn.Module):
    def __init__(self, in_channels, out_channels, kernel_size, **kwargs):
        super(BasicConv2d, self).__init__()
        self.conv = nn.Conv2d(in_channels, out_channels, kernel_size, bias=False, **
kwargs)

    def forward(self, x):
        x = self.conv(x)
        return F.leaky_relu(x, inplace=True)

class SetBlock(nn.Module):
    def __init__(self, forward_block, pooling=False):
        super(SetBlock, self).__init__()
        self.forward_block = forward_block
        self.pooling = pooling
        if pooling:
            self.pool2d = nn.MaxPool2d(2)

    def forward(self, x):
        n, s, c, h, w = x.size()
        x = self.forward_block(x.view(-1,c,h,w))
        if self.pooling:
            x = self.pool2d(x)
        _, c, h, w = x.size()
        return x.view(n, s, c, h ,w)
```



```

class HPM(nn.Module):
    def __init__(self, in_dim, out_dim, bin_level_num=5):
        super(HPM, self).__init__()
        self.bin_num = [2**i for i in range(bin_level_num)]
        self.fc_bin = nn.ParameterList([
            nn.Parameter(
                nn.init.xavier_uniform(
                    torch.zeros(sum(self.bin_num), in_dim, out_dim)))])

    def forward(self, x):
        feature = list()
        n, c, h, w = x.size()
        for num_bin in self.bin_num:
            z = x.view(n, c, num_bin, -1)
            z = z.mean(3)+z.max(3)[0]
            feature.append(z)
        feature = torch.cat(feature, 2).permute(2, 0, 1).contiguous()

        feature = feature.matmul(self.fc_bin[0])
        return feature.permute(1, 0, 2).contiguous()

```

## 6. 运行效果

### 6.1 视频输入检测实例



图 8 输入视频检测实例 1



图 9 输入视频检测实例 2

以网上一外国男子演示的跌倒过程作为系统输入，如图，在男子站立时，系统输出“Normal”即无危险，当跌倒时，系统经历了两个过程，一个是短暂的“FALL WARNING”阶段，由于人跌落经历了一个短暂的倾斜过程，此项目对这个过程进行了及时的捕捉并输出了预警，最终男子完全跌落后则是“FALL”阶段。

## 6.2 摄像头检测实例

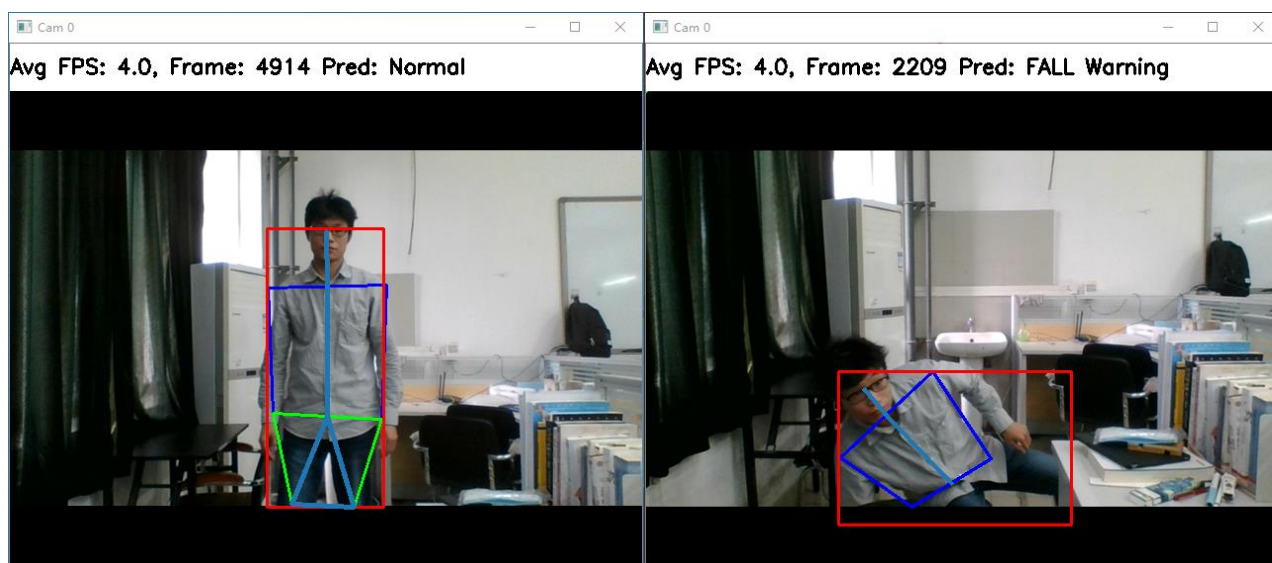


图 10 摄像头检测实例 1

以实验室配备的摄像头作为输入，系统在判断信息主体正常站立时输出“NORMAL”即无危险，在信息主体即将跌落时输出“FALL WARNING”给出预警。

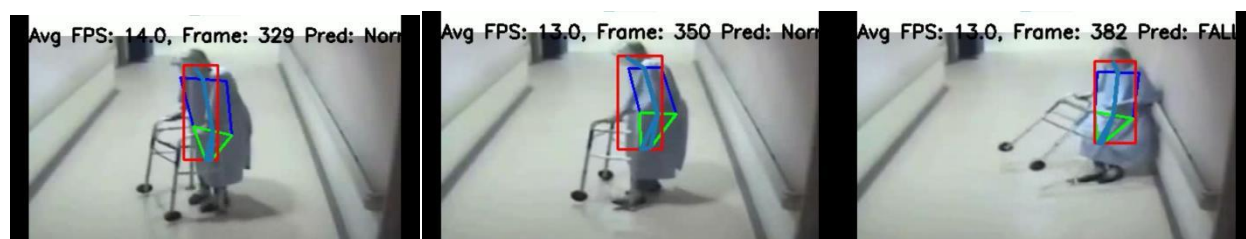


图 11 摄像头检测实例 2

以常规医院摄像头作为输入，可见，在有一定距离且画面中有杂物和遮挡（拐杖）的情况下对主体的识别和跟踪不会出错，且会在信息主体即将跌倒时有效地进行“FALL WARNING”的跌倒预警。

## 7. 特色及创新

### 7.1 创新性

我们新增的多摄像头、主体识别以及跟踪功能扩展了近年来的研究成果。另外，实现多人姿态估计方法，通过将 LSTM 神经网络与人体姿势估计获得的特征相结合，项目支持多摄像机输入及多人实时坠落检测，使移动群智感知的部署具备可行性，可为运动场地、医院等场景采集实时监控数据提供服务。

### 7.2 优越性

(1) 由于我们的网络可以全天候不间断工作，并能第一时间发现异常并自动发出预警，该项目可以大大减少视频监护中投入的人力物力，能够 24 小时实时地提供监护预警的帮助。不仅降低了监护过程中的人力成本，还能给被监护对象提供更为可靠的帮助，为使用方切实带来一定的效益。

(2) 该项目由于可迁移性很强，可以应用于各大视频监护场所，只要读取视频监控摄像头的视频流，就可以实现对该区域的人员信息的监控。所以该项目普及的难度和成本都比较低。需要做的是继续进行一步在模型压缩、模型推理速度、计算平台性能等方面上的优化。

(3) 由于我国目前越来越重视医疗行业发展，以及目前新冠疫情的影响，医院等人流密集场所急需自动化人员行为分析，能快速发现并救助异常行为人员。另外像各种人员密集单位:银行，机场，养老院等，均需及时高效的人员分析设备。另外加上我国人口老龄化日益严重，空巢老人数量剧增。因此对于独居老人进行看护也显得尤为迫切。目前智能监控领域市场份额巨大，有很大的需求并且能切实为社会带来实际效益。本项目的市场前景广阔，推广并实施的可行性强。当今处于一个老龄化的社会，同时也是一个社会压力倍增的社会，年轻人由于迫于社会压力对于老人和儿童的关照很少，老年人出现摔倒受伤甚至猝死,儿童由于好奇出现攀爬导致高空坠落、以及由于追逐打闹出现受伤的报道时有发生。社会上由于对于老人和儿童的实时行为检测和看护的系统还没有出现，而本产品填补了这一领域的空白。

## 7.3 市场价值

(1) 本系统使用简单，能够自动实时地对老人、儿童、残障人士等实现看护，因而容易在这一目标用户进行推广。

(2) 当今处于老龄化的社会，老年人数量基数大，同时由于独生子女政策和社会压力的剧增，导致子女很难抽出时间去照顾老人和儿童，而本产品能够在保障被看护人员安全的前提下最大限度地减少看护人员的精力投入，易于被年轻人所接受。

(3) 养老院、医院数量的增多、交通的日益发达，这些应用场景均可作为本产品的受众对象，能够帮助其解决人员优化调度的问题。

## 7.4 转化前景

(1) 本项目可以向着与 3D 动画人体运动模拟软件相结合的方向发展，通过识别出人体的动作从而让 3D 动画中的人物也做出相同的动作；

(2) 本项目可以与监控摄像头相结合，对监控摄像头画面中的人体行为进行识别分析，在画面中的人做出危险或不正常的动作时发出警告让用户得知，及时处理；

(3) 本项目可以与家用监控结合，当监控发现家里人出现危险行为时，可以立即通知用户并让用户调查监控查看发生了什么事，从而避免事态更加恶化，这样一来哪怕用户不在家里也能及时得知家里出现了什么状况；

(4) 目前医院病人众多，有时候医院的工作人员来不及照顾那么多病人，所以病人出现问题却没有及时得到解决的事情时有发生，而本项目的预警危险或不正常动作的功能可以让医院的工作人员立即得知病人情况并得到及时处置。

# 8.总结与展望

## 8.1 总结

本研究首先对近年来对人体行为识别实际应用案例进行了举例，并通过案例展示相关技术和 AI 识别人体行为的价值，通过以上案例可以看出，人体行为识别有很多很大的应用价值，

只要我们继续研究发展人体行为识别，就能造福更多不同领域的人民。

之后对该项目的需求进行分析，提出设计方案、实现方案和预期目标等内容，并详细的介绍所用技术的理论、流程和展示。本研究是以“基于 OpenPifPaf 的跌倒检测模型：支持多摄像头、多人实时监测”项目为主，“Alphapose”和“Gaitset”为辅构成的，这些项目的特点和进程如下所示：

（1）基于 OpenPifPaf 的跌倒检测模型：支持多摄像头、多人实时监测，OpenPifPaf 对输入视频进行人体姿势估计，通过 LSTM 从姿势信息中提取时间和空间特征进行“跌倒”等异常行为检测；我们虽然已经实现了利用 LSTM 识别人类的跌倒、将要跌倒和正常状态，但是这个项目无法识别其他异常动作，所以这个项目目前只能识别人类是否跌倒，功能相对比较单薄。

（2）Alphapose：一种图像预处理方法，可以从有人物的视频中将人物的骨架信息标记并输出带有骨架信息的 json 文件或.avi 格式的视频文件，其输出的文件将人的躯干以骨架表示，为下一步识别预测人体动作或动作捕捉做准备。然而，由于目前没有做到与 Alphapose 结合的好的后续项目，所以目前 Alphapose 没有做到和其他项目联动配套。

（3）Gaitset：一种步态识别方法，能够对步态轮廓图（光流图）进行识别和验证，对已有个体步态信息进行学习，并对验证目标进行分类，目前已在 CASIA-B 和 OU-MVLP 数据集上进行训练，支持非时序性输入。目前该项目的功能尚在开发当中。

## 8.2 展望

### 8.2.1 近期项目规划

（1）将基于 OpenPifPaf 的跌倒检测模型与其他人体识别模型相结合，以提高动作检测的准确性。

（2）增加检测人体异常行为的数量，丰富基于 OpenPifPaf 的跌倒检测模型：支持多摄像头、多人实时监测项目的识别人体异常行为的能力。

（3）找到能对 AlphaPose 预处理后的数据进行处理并识别的项目，并最终做到输入一段视频就能实时检测识别人体动作。

(4) 继续完善 Gaitset 算法，最终要让这个算法能识别输入图片的主体身份，进而做到识别视频中人物身份的能力。

## 8.2.2 长远项目规划

(1) 预期成果形式：一个基于多个摄像头，多人实时行为检测模型。

成果形式 1:集成为小型 APP,通过手机连接摄像头对老人进行实时跌倒等异常行为检测，在意外发生时有效提醒亲属或直接拨打 120，尽可能保护老人健康。

成果形式 2:连接公共场合摄像头，对异常行为进行识别并发出预警，有效防范暴恐等犯罪事件以及跳河等自杀事件。

(2) 预期效益:对各种异常行为的检测，能让当事人亲属和医院，公安等相关部门及时得到预警，更快更好处理突发事件，保护生命和财产安全。

## 9.参考文献

- [1]Lei Wang, Du Q. Huynh, Piotr Koniusz. A Comparative Review of Recent Kinect-based Action Recognition Algorithms[J]. IEEE TRANSACTIONS ON IMAGE PROCESSING,2019
- [2]Nusrat Tasnim , Mohammad Khairul Islam and Joong-Hwan Baek. Deep Learning Based Human Activity Recognition Using Spatio-Temporal Image Formation of Skeleton Joints[J].applied sciences
- [3]Mickael Delamare, Cyril Laville, Adnane Cabani and Houcine Chafouk. Graph Convolutional Networks Skeleton-based Action Recognition for Continuous Data Stream: A Sliding Window Approach[J]. 16th International Conference on Computer Vision Theory and Applications
- [4]Tasweer Ahmad, Lianwen Jin, Xin Zhang, Songxuan Lai, Guozhi Tang, and LuoJun Lin. Graph Convolutional Neural Network for Human Action Recognition: A Comprehensive Survey[J]
- [5]Zehua Sun, Jun Liu, Qiuhong Ke, Hossein Rahmani, Mohammed Bennamoun, and Gang Wang. Human Action Recognition from Various Data Modalities: A Review[J]
- [6]Zehua Sun, Jun Liu, Qiuhong Ke, Hossein Rahmani, Mohammed Bennamoun, and Gang Wang. Human Action Recognition from Various Data Modalities: A Review[J]

- [7] Jinmiao Cai, Nianjuan Jiang, Xiaoguang Han, Kui Jia, Jiangbo Lu. JOLO-GCN: Mining Joint-Centered Light-Weight Information for Skeleton-Based Action Recognition[J]
- [8]Chao Li, Qiaoyong Zhong, Di Xie, Shiliang Pu. SKELETON-BASED ACTION RECOGNITION WITH CONVOLUTIONAL NEURAL NETWORKS[J]
- [9]Xianshan Li, Fengchan Meng, Fengda Zhao, Dingding Guo, Fengwei Lou, Rong Jing. Two-stream adaptive-attentional subgraph convolution networks for skeleton-based action recognition[J]. Multimedia Tools and Applications2021. PP 1-18
- [10] 《GaitSet:Cross-view Gait Recognition through Utilizing Gait as a Deep Set》
- [11] 《Gait Lateral Network : Learning Discriminative and Compact Representations for Gait Recognition》
- [12] Kim Woojoo, Sung Jaeho, Saakes Daniel, Huang Chunxi, Xiong Shuping. Ergonomic postural assessment using a new open-source human pose estimation technology (OpenPose) [J]. International Journal of Industrial ErgonomicsVolume 84, 2021
- [13] Lin ChuanBi, Dong Ziqian, Kuan WeiKai, Huang YungFa. A Framework for Fall Detection Based on OpenPose Skeleton and LSTM/GRU Models[J]. Applied SciencesVolume 11, Issue 1. 2020. PP 329-329
- [14] Nakano Nobuyasu, Sakura Tetsuro, Ueda Kazuhiro, Omura Leon, Kimura Arata, Iino Yoichi, Fukushima Senshi, Yoshioka Shinsuke. Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras[J]. Frontiers in Sports and Active Living 2020
- [15]Weiming Chen, Zijie Jiang, Hailin Guo, Xiaoyang Ni. Fall Detection Based on Key Points of Human-Skeleton Using OpenPose[J]. Symmetry Volume 12, Issue 5. 2020
- [16]OpenPifPaf ( <https://github.com/openpifpaf/openpifpaf> )
- [17]UP-fall detection Dataset ( <https://dx.doi.org/10.3390/s19091988> )
- [18]Multi-camera, multi-person, and real-time fall detection using long short term memory ( <https://doi.org/10.1117/12.2580700> )