2021年易方达资产杯“人工智能+”大学生创新技能挑战大赛暨2022年江苏省大学生计算机设计大赛校内选拔赛

人工智能实践赛作品报告

作品名称：基于人体骨骼关键点识别的摔倒检测技术

作　　者：一发密苏里

填写日期：　2021.10.16

填写说明：

1. 本文档适用于人工智能实践赛；
2. 正文一律用小四号宋体，1.3倍行距，0.5行段后距；一级标题为二号黑体，其他级别标题如有需要，可根据需要设置，标题格式为阿拉伯数字，如第1章的标题写为1.1，1.1.1等；
3. 本文档应结构清晰，突出重点，适当配合图表，描述准确，不易冗长拖沓；
4. 提交文档时，**命名为“人工智能实践赛作品报告-队伍名称”以PDF格式**提交；
5. 本文档内容是正式参赛内容组成部分，务必真实填写。如不属实，将导致奖项等级降低甚至终止本作品参加比赛。

**目 录**

[第1章 作品概述 1](#_Toc70698312)

[第2章 问题描述 1](#_Toc70698313)

[第3章 技术方案 1](#_Toc70698314)

[第4章 系统实现 1](#_Toc70698315)

[第5章 分析验证 2](#_Toc70698316)

[第6章 作品总结 2](#_Toc70698317)

[参考文献 2](#_Toc70698318)

# 作品概述

【填写说明：重点介绍本作品的主题创意来源，产生背景，作品的用户群体、主要功能与特色、应用价值、推广前景等。如果有同类竞品，建议从多个维度对本作品与竞品进行比较，建议不超过2页】

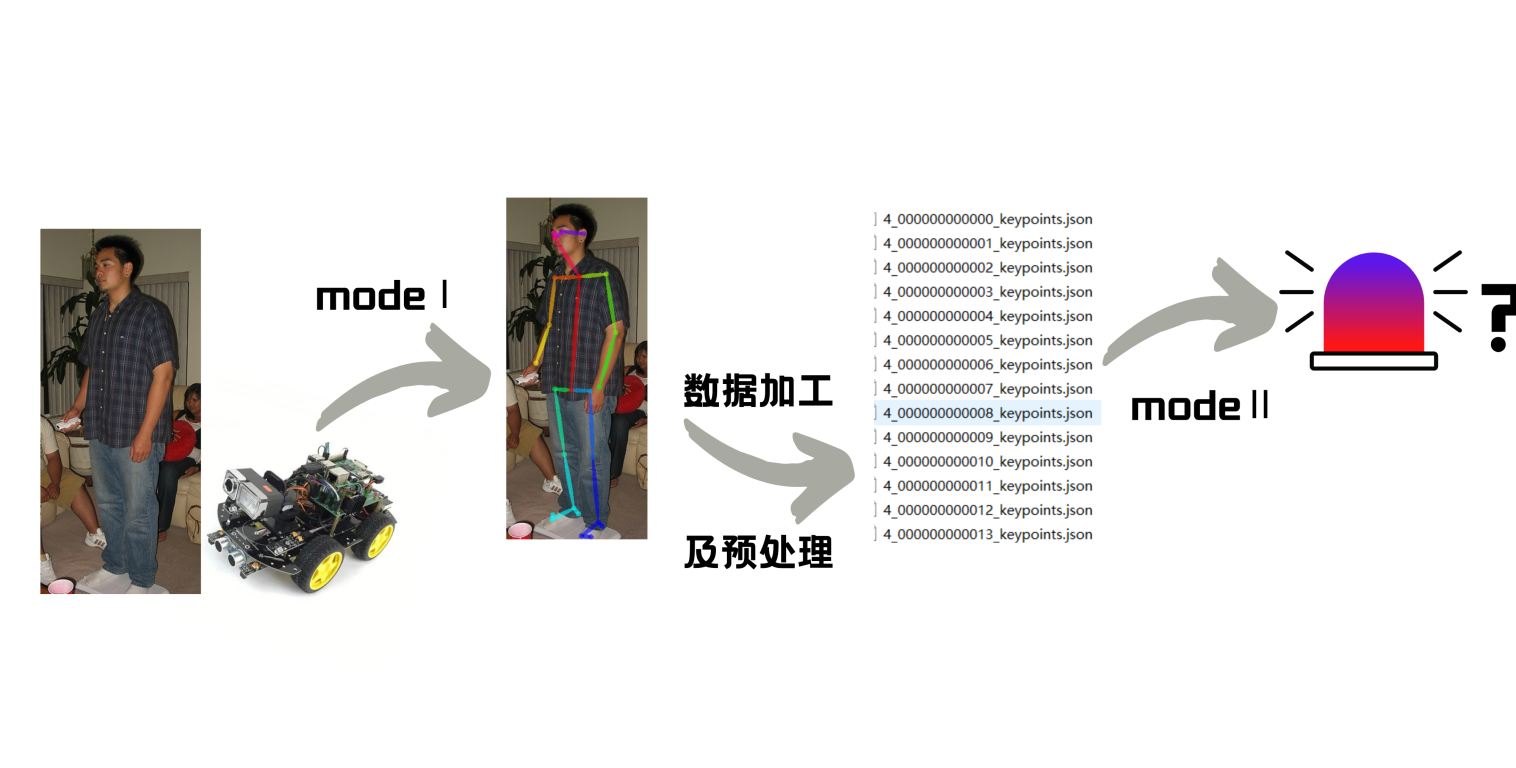
**1.1 设计背景**

根据联合国人口署2020年公布的数据中国人口年龄中位数已达到38.4，并且预计在2050年达到48。未来30年国内最大的投资主题应该是老龄化带来健康和医疗服务机遇，而对老年人健康最大的威胁就是独居时的意外摔倒。糖尿病，心血管疾病，骨骼疾病引发的摔倒或晕厥很可能由于得不到及时的处理而恶化。

市场上上已经出现多款家用智能监控，小米等对家居安全深入的企业已经推出了支持双向语音通话，24h2k高清红外摄像机，同时搭配简单人形检测AI的家用摄像头。但这些固定的摄像头普遍有着覆盖范围小，需操作者查看，不能做到实时响应危险等问题，不能提供广泛存在的独生子女家庭需要的服务。

**1.2作品功能**

本作品基于以上不足与潜在需求，我们使用了CMU开源的人体姿态模型与识别技术openpose[1]，以及MSCOCO数据库和深度神经网络。期望完成以下工作构想：

1. 从可移动的摄像头接受输入
2. 以高帧率处理图像，识别图像中人物
3. 解析目标人体人体姿态以及骨骼关节点，保存数据
4. 利用时段内关节点信息重建模型判断目标状况
5. 目标处于摔倒状态超过阈值发出警报（紧急联系人）

整体工作流程可以参考下图

**1.3突出特点**

面对上面提出的已存在产品的不足，我们期望打造产品的主要优势和攻关方向有以下几点：

速度与相应时效，基于现有的技术加以改进取舍，在不影响识别正确率的基础上降低系统流程的反应延迟，做到即时检测；

可靠性，依靠嵌入式开发，将系统集成到类似树莓派的硬件之上，搭配可移动平台做到全方位监控同时保证隐私；

舒适性，避免误报产生的打扰，将在训练和检测数据中加入模糊数据重采样，同时对一个时段内的数据集中分析。

**1.4应用前景**

未来面向高龄人群的家居与健康服务市场潜力相当之大，肯定会有配套的系列产品出现。加之许多企业提出了智能家居的理念构想，我们的作品完全可以嵌入式部署在扫地机器人之类的产品上，配套监测心率健康手环，以实现更准确的摔倒判断以及定位目标位置等功能，成为家居物联网中不可缺少的一部分。而且这套技术应用范围不仅限于此，在同一套硬件设备的基础上，还可以附加婴幼儿，宠物监护，防盗防火等功能，以供用户个性化选择。

# 问题描述

【填写说明：详细描述作品拟解决的实际问题，作品的功能和性能需求；使用的数据集，包括数据格式，数据来源，数据获取方式，数据特点，数据规模等，并给出具体的数据样例。所提出的指标点必须等在第5章得到印证】

**3.1 样本不均衡与属性缺失**

模型一人体姿态分析使用的是开源的MSCOCO数据[3]（url=cocodataset.org 图[2]）进行训练和测试，该数据集提供了超过两万张包含所有人体关键点信息标注的图片，可以直接通过官网打包下载图片与标签数据，能完全满足openpose模型对于一般环境之下的正确率的需求[1]。但从二级输入以及应用背景的角度来看，仅使用该数据集是不够的。

其一是因为本系统的最终目标是检测摔倒，而该数据集中此类标签之下的样例太少，交给模型二的数据正反比例太过悬殊，摔倒这一行为太难准确地进行采样；

其二性是由于环境遮挡，图片截取等多方面的因素，模型一的输出结果（关键点信息）常存在属值缺失的情况，其中有些样例进行填充之后可继续使用（如左耳），而有些关键的属性丢失会导致整个样例失效。必须对模型一的输出加以判断和处理。对此我们的策略是：

1. 扩大数据集，遇到类别不均衡问题时，可以再增加数据更多的数据，更多的数据往往能够得到更多的分布信息，新增加的样本尽量调和原有的比例，即多增加小类的样例。具体来说，我们是在不同环境下多人额外录制视频按照 10帧/s的频率采样，其中反例的比率较大。
2. 数据集重采样，可以使用一些策略该减轻数据的不平衡程度。对小类的数据样本进行过采样（over-sampling）来增加小类的数据样本个数，对大类的数据样本进行欠采样（under-sampling）来减少该类数据样本的个数。
3. 数据增强，对较少的正例样本图片进行适当的剪切，平移，旋转得到更多的少量样本，从而实现正负类样本的均衡。
4. 数据预处理，分析不同关键点的重要权值，对于缺失较少的样例可用均值或者特征填充，缺失较多的样例直接舍弃。

**3.2 快速实时响应**

在设想的应用环境里，系统响应时长和处理速度都至关重要。在出现意外后能够及时做出相应，以及处理运行的速度应与视频输入的速度保持一致，是衡量系统性能的重要指标。为了减少整个系统的运行用时，以达到即时对视频输入作出反应，我们仅对图片中置信度（confidence）最高的部分所组成的人体进行判断分析。同时减少相关性较低的face feature 和 hand feature 的模型关键点匹配。从原有的 22(face) + 12(hand) + 25(body)减少为25关键点位。同时降低输出的帧速率以提高模型二的吞吐量。

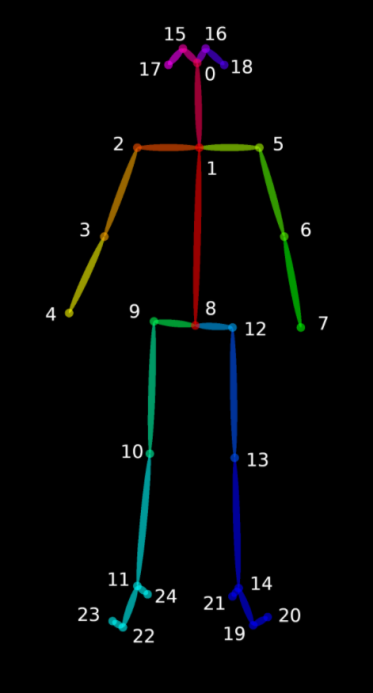
**3.1 假反例率高**

不仅样本中正例(正常)反例(摔倒)比例过大，正反例的先验概率之比也很大。直观来想，生活中真正摔倒的情况是比较罕见的，而且存在很多容易被误判为摔倒的动作：躺下，睡觉，坐下等。对于这些敏感易混淆的数据应重点分析处理。不然就会频繁出现虚假警报的情况，降低真正危险受到的注意程度。同时也不能降低整体反例的预测阈值——误报固然会形成骚扰，但漏报会影响产品整体可靠性的评估。最好是能从多方面测量提高模型的准确度。

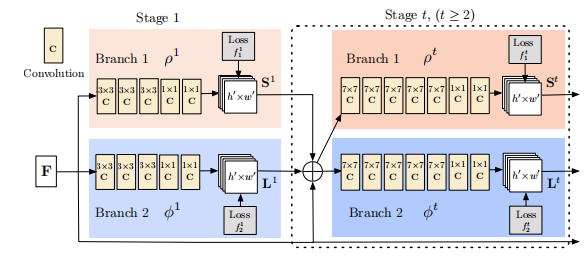
# 技术方案

【填写说明：从原理层面，详细介绍系统所采用的技术方案，先总体介绍，给出技术路线框架图，然后分模块详细介绍。着重介绍解决问题的思路，以及所涉及的模型、协议、算法等，以及可能的对算法的改进；原创工作详述，非原创工作简述，并尽可能标注引用文献】

**3.1 openpose关键点检测.**

我们使用了基于CMU团队开发的openpose检测架构的bottom-up算法实现模型一检测人体关键点的功能[1] 。其继承了已有的 Convolutional Pose Machines[2]与传统Pose Machines[4]的优势,使用mscoco数据集提供的body匹配模型与关键点信息。输入指定尺寸的图片，输出检测到的关键点的位置信息。

模型整体由两部分卷积神经网络构成，分别是匹配不同关键点的belief map与将不同关键点之间连接起来构成完整人体的limb vector map ，依次经过卷积池化线性激发等层提取特征。其工作原理如下：

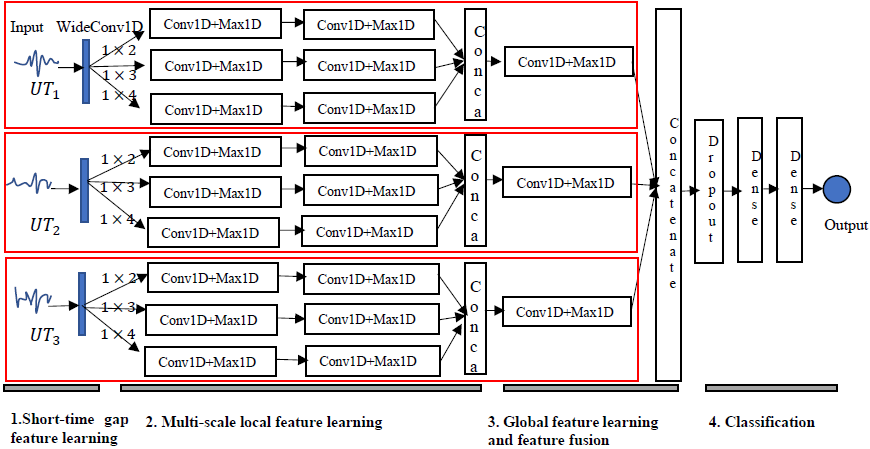


整个训练过程会被分成多个效果渐强的stage，在最初的stage，原图作为输入由两个网络做出初步的分析，其结果会与原图的信息一起作为下一阶段的输入。网络会不断提取相关特征，在限制时间内以及损失函数值给出最终的检测信息。

训练数据全部含有标记，使用传统的监督学习方法，损失函数为生成的map与ground truth之间的距离差。

**3.2跌倒时间序列检测**

通过3.1的模型，我们能够在视频输入的每一帧中提取人体关键点（共２５个）的坐标．我们每次选取从３０帧中提取的数据作为一个时间序列作为输入，这是一个多变量时间序列分类问题(Multivariate Time Series Classification, MTSC).由于每个时间切片本身没有类别标签，类比文本分类问题，我们采用一维卷积神经网络的方法．然而，传统的基于CNN的MTSC算法存在着无法适应较小数据集的问题．同时，对于该问题，若人体关键点坐标和置信度以不同通道输入到模型中时，存在互相干扰中和的问题．釜山大学Xiaorui Shao, Chang Soo Kim提出了一种多规模特征提取的TSC模型[5],该模型将时间序列的不同变量分开，分别送入本地的卷积神经网络，该网络拥有GoogLeNet[6]的Inception结构，多个拥有不同大小的卷积核的卷积层能够同时识别不同长度的特征．最后将这些本地网络输出的特征向量拼接，再送入一个卷积层进行特征提取．



Xiaorui Shao, Chang Soo Kim的工作将时间序列的不同变量完全分开提取特征再连接，然而考虑到本问题的特殊性，我们对算法做出了一些改进．不同关键点的坐标若完全分开送入卷积神经网络提取特征可能导致丢失与关键点之间相对位置有关的信息，导致卷积核过拟合．所以，在我们的模型中，我们将关键点的横坐标，纵坐标和置信度作为三个变量分别送入三个本地卷积神经网络，在本地神经网络中，第一层卷积层我们用三种大小的核来识别不同长度的特征（对应于摔倒的时长），同时将步长设计为一帧的数据量．此外，考虑到三种变量时间序列的联合分布与类别较强的相关性，我们在连接三个本地神经网络输出特征图之后，再采用一层卷积层进行特征提取．

# 系统实现

【填写说明：从工程实现的角度，详细阐述第3章提出的技术方案的具体实现过程，包括且不限于软件设计实现，用户界面，数据来源，数据训练，改进过程，以及系统部署方法等，以及其中所遇到的困难，解决的方法等】

**4.1 数据采集与标注**

对于模型1，可以使用已有的带有标注的MSCOCO图片数据集进行训练，鉴于数据集过于庞大，我们使用随机采样选取部分数据集，并根据图片上像素点位的标记分别生成最大值高斯分布的map作为两部分神经网络的groundtruth损失相对参考。





而对于模型二，需要网络上较少的视频数据集，而且网络上几乎没有摔倒相关的数据集，因此除了搜集补充数据集之外我们还需要手动进行标记，我们除了使用，通过网络上搜索收集、自行拍摄等方法补充了部分样例（主要是较少的正例）并把我们认为比较典型的样例做了变换之后添加入数据集作为加权重处理。例如，我们使用了University of Rzeszow公开的动作数据集[5]，但该数据集只含有70段10s左右的视频参考，我们在此基础上对这些视频处理后产生的中间数据

**4.2 关键点检测模型参数调整与训练**

关键点识别模型默认会使用最大运行设备GPU默认最大输出帧速率，为了能够以稳定的速率产生输出，同时保证数据动作的连贯性以及后续模型的处理响应速度，我们固定了frame的频数，同时在进行bottom-up的时仅保留构建出的完整个体中置信度最大的一个(默认使用环境中只会出现一人)，最大化提高处理速度。

**4.3 数据的加工与预处理**

模型1的输出结果是pattern中每一关键点相对于图像起点的二维偏移坐标，以浮点数格式表示，未识别到的部分会以空值代替，为了避免人物所处的位置这一无关特征以及缺失值的影响，将对数据做以下处理：

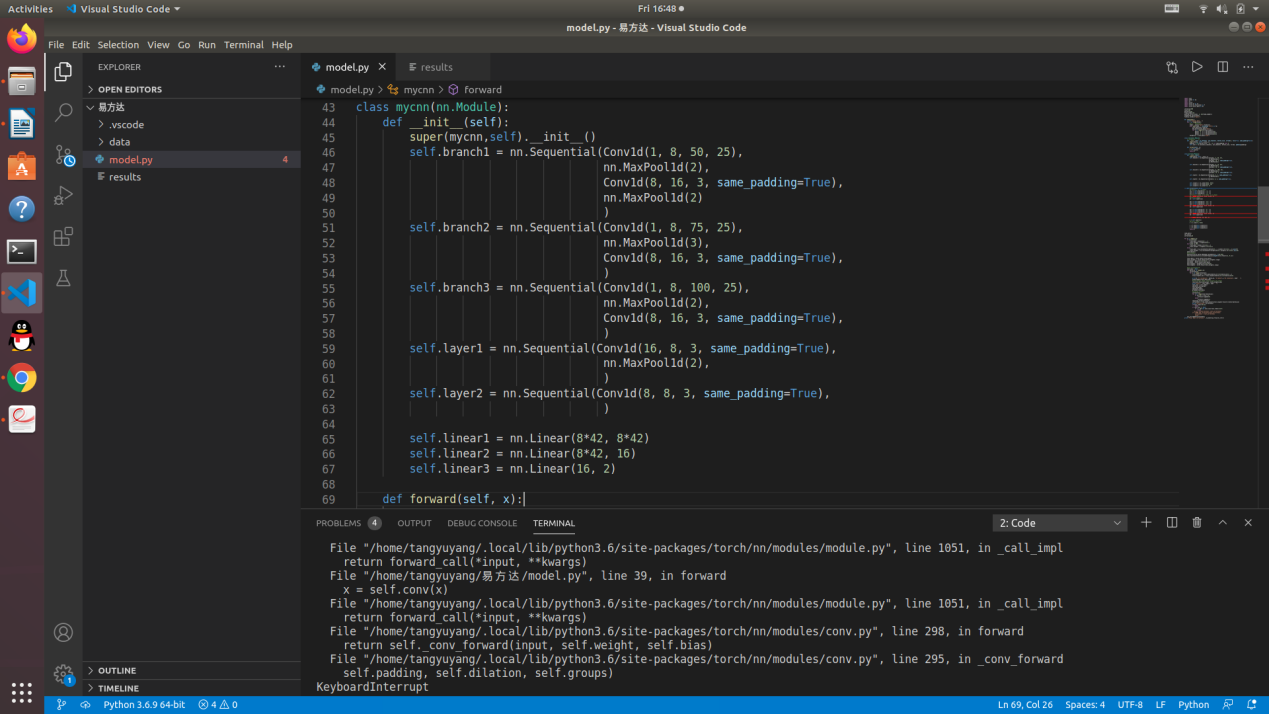
缺失值填充：有时可能会由于遮挡，画面虚化，人物离开摄像范围导致部分关键点信息缺失，对此我们将试图读取该frame前后其他时刻的采样值。由于视频中人物的动作应该是连续的，我们可以使用一段时间内的该关键点的均值来填补缺失值，当然对于重要关键点的缺失（如躯干、肩、腿、髋），我们会判别为该帧信息无效。

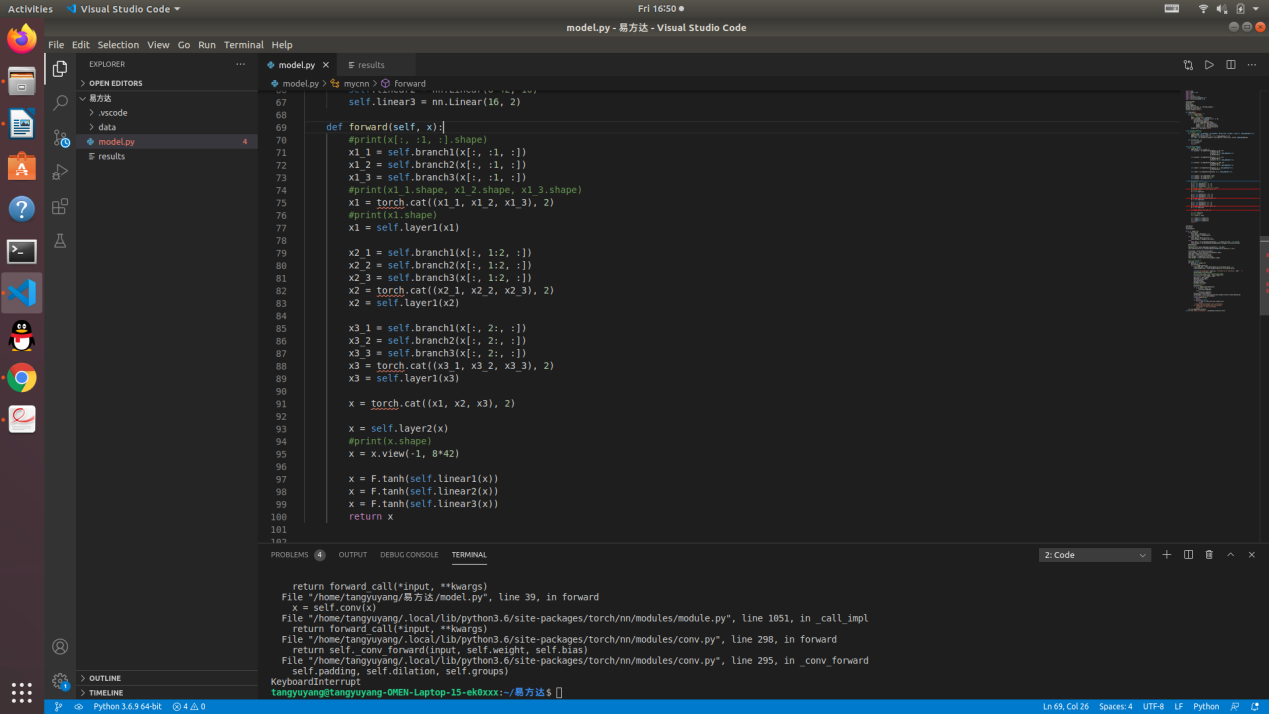
样本中心化：最大程度提炼数据中的姿态信息，将对json文件中关键点信息进行修正，基于躯干关键点(point=1),重新修改各关键点位置为相对于躯干位置偏移。

将输出的连续帧json数据按0.5s一个样例切割聚合，并标注相应类别标签，作为下一步的训练和测试数据。

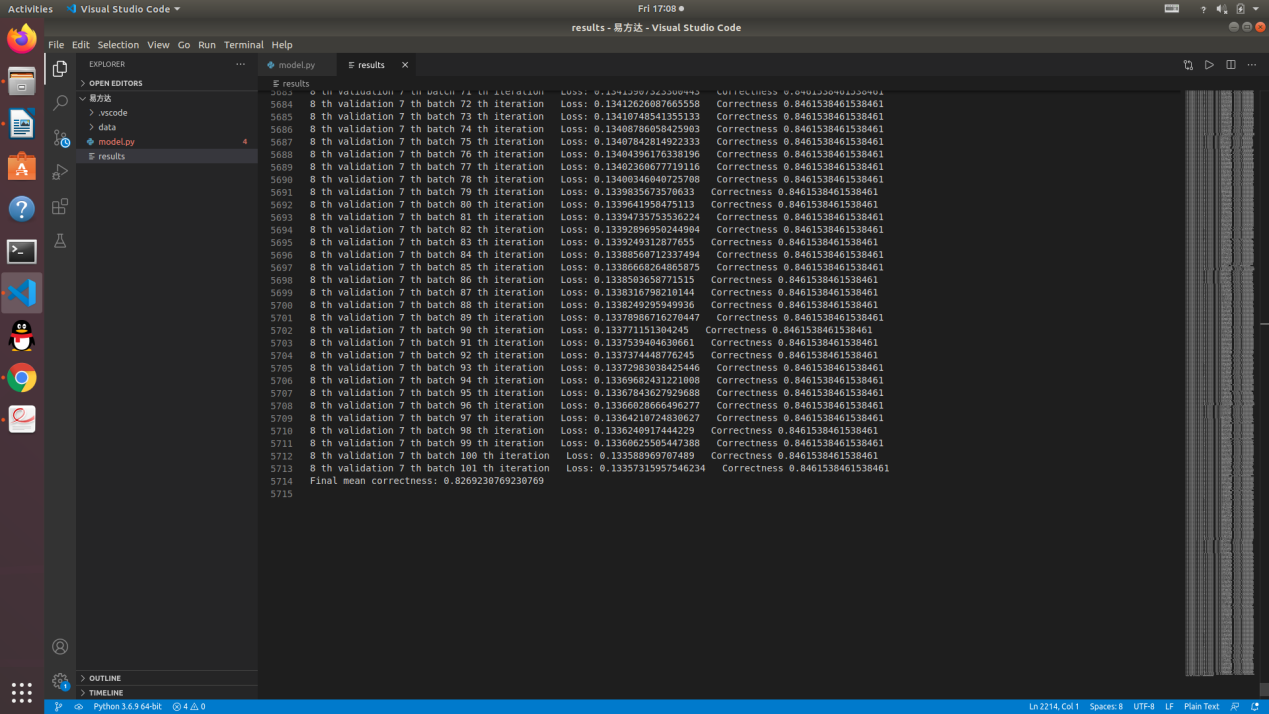
**4.4　模型的训练和评估**

我们利用Pytorch库对3.2所述模型进行了搭建．在本地神经网络中我们采用50，75和100作为第一层卷积核大小得到三个分支．在一层池化层和第二层卷积层之后将三个特征向量连接，再通过一个卷积层layer1得到此变量的时间序列特征向量．

　　在所有变量的特征向量提取完毕后，我们将它们连接并用一个新的卷积层提取与它们联合分布相关的信息，最后通过全连接层收敛并输出一个一维的类别向量．对于该二分类问题，我们采用交叉熵作为损失函数，SGD优化器并用早停法防止过拟合．

****

进行模型评估时，我们采用k折交叉验证法，将数据平均分为k份，每次取其中一份作为测试集，其余为训练集．最后将k次训练结果在测试机上的泛化误差求平均得到模型的性能．



**4.5 系统的嵌入式部署**

我们最初的配置是一块raspberry4b(4G)+2wd平台车+L298N电机驱动模块+800万像素的索尼IMX219 的套件，期望将整个系统移植部署到这个移动系统上。但在实现过程中发现Nivida未发布相应的硬件驱动，使用cudatoolkit的模型1无法在raspberry上运行，于是改变计划仅将摄像头远程连接到笔记本电脑上，通过ssh链接仍在电脑上使用模型进行判别和输出。但这样没有利用到树莓派本身的编程和计算能力，我们在上面配置了简单的目标检测功能，以脚本的形式控制摄像头的开关。下一步可能会应用相对简单目标识别以及路径规划的功能来辅助系统提高运行效果。

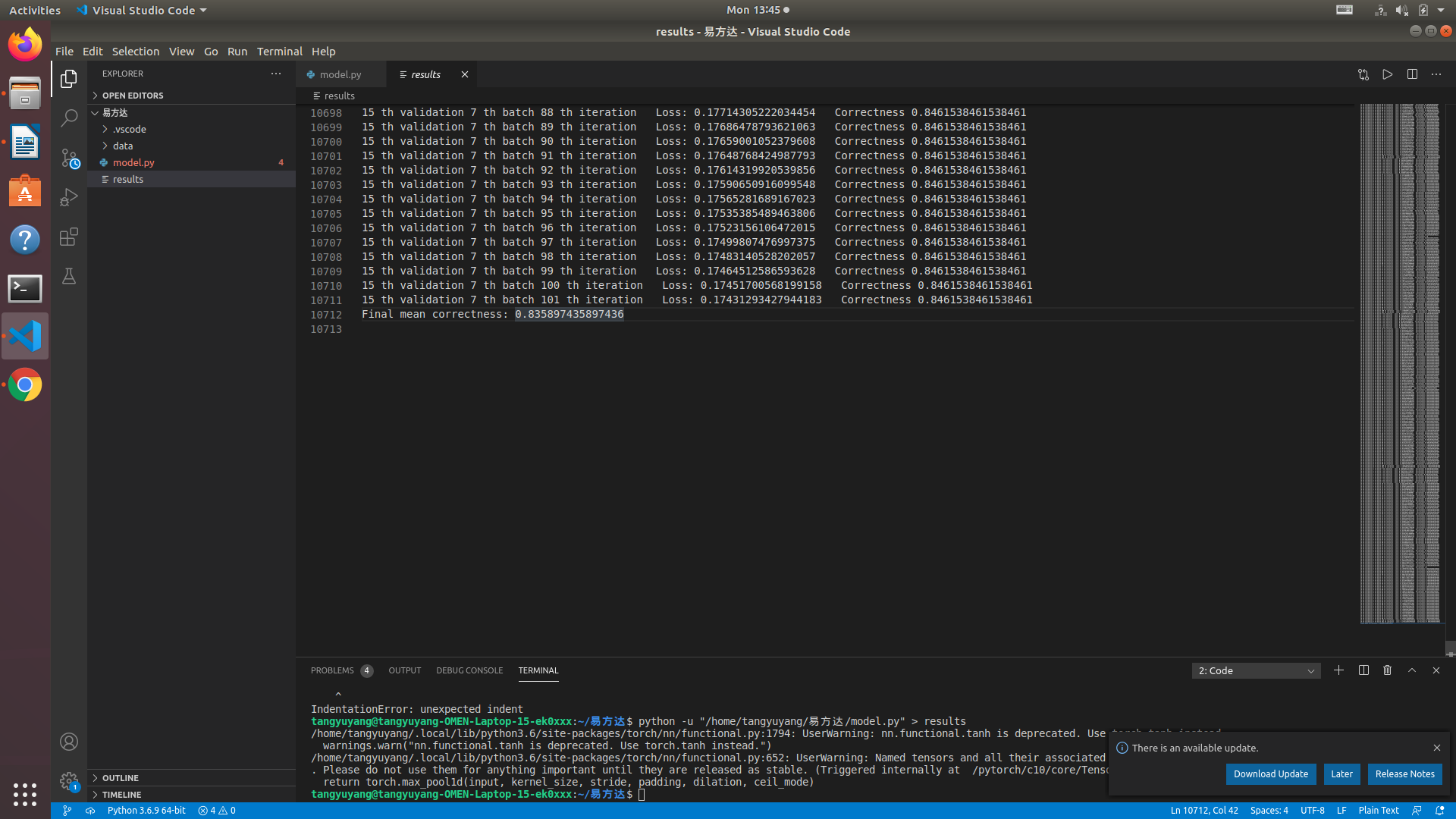
# 分析验证

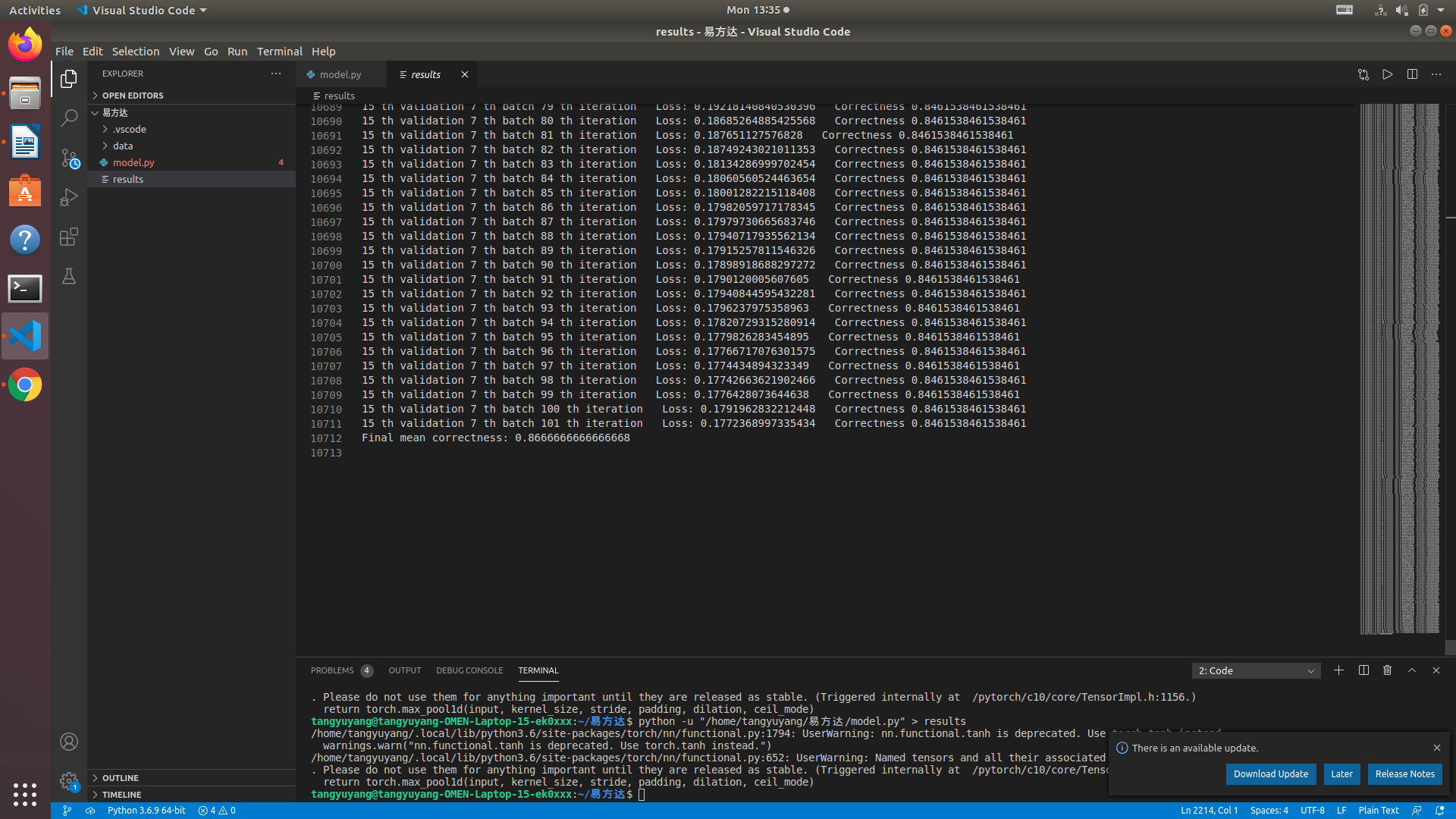
【填写说明：通过测试与对比，论证系统的有效性，可包括验证数据的来源与规模、测试过程、分析与结论等等。各参赛队务必重视数据测试，所有对自己作品准确性、有效性、稳定性，甚至作品受欢迎的程度的宣称，都应该得到数据结果或对比实验的支持，否则评审人有理由怀疑其真实性】

**5.1模型准确率**

针对所研究的时间序列，用将三个通道的信息同时放进同一个卷积神经网络，效果不佳，准确率仅能达到约69％．采用三个信号分别输入三个网络，且用多种卷积核同时识别特征的方法后，准确率上升到约77％．

　　在这基础上，我们在整个网络最后加入一层卷积层，并采用mini-batch梯度下降的方法，调整batch大小，同时，运用拉伸，镜像等方法扩充训练集，最终将精度提高到约87％．（以上精度均由k折交叉验证产生）



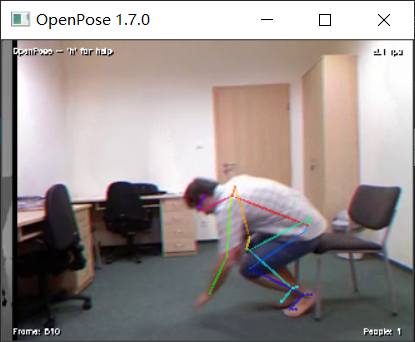


（添加尾端卷积层之前与之后的精度对比）

**5.2系统运行速度**

从终端拍照-上传电脑-输出结果这一全部过程中主要的时间消耗在模型1的处理部分，可以调整处理frame的频率以达到实时响应。下面进行分析验证，在多段视频上先使用默认设置的最高fps以及处理速率进行测试，使用的硬件以及平台信息如下：

|  |  |
| --- | --- |
| CPU | Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz |
| GPU | NVIDIA GeForce GTX 1660 Ti |
| RAM & OS | 8.00 GB Windows 10 21H1 |





由于在进行处理会首先将输入流拉伸处理到默认尺寸比例（w:400h:300)所以视频的尺寸影响不大。默认最大fps会在10左右之间波动，同时每秒产生100个frame的json文件，如果要兼顾后续的处理和判断，必须对frames/s 进行调整。

另外，我们通过树莓派摄像头使用Vlc向计算机推流，在现阶段使用的局域无线网存在大概1~2s波动的画面延迟。

综合考虑以上，我们固定接受的fps为10，处理frame的步长为3，拍摄1280x720像素的画面，系统整体响应延迟能够控制在5s以内

# 作品总结

【填写说明：从创意、技术路线、工作量、数据和测试效果等方面对作品进行自我评价和总结，并对作品的进一步提升和应用拓展提出展望】

在本作品中，我们通过不断地尝试和调整，基本完成了预期的目标需求，还在开发过程中加入了许多新的创新点。我们围绕预定的应用环境在现有的成熟模型与技术的基础上进行改进和调整。我们首先完成了快速识别人体主要关键点的模型Ⅰ。接下来对于问题一开始存在的缺少数据以及类别不平衡，我们通过各种渠道重新收集了大量数据，并对原有数据中不同类别的样例进行了一些处理。同时对视频通过模型Ⅰ的结果进行中心化、填补、切割打包等操作，作为模型Ⅱ的训练和测试数据。

最后将全部的模型部署为一个系统，进行初步的产品开发尝试。以树莓派避障车上的摄像头作为视频源输入，在电脑上进行处理以及结果的预测。

从最终测试结果来看，我们的系统还存在反应速率比较慢，仍有一定的错误率，处理过程过于复杂等问题。我们下一步改进的工作重心是继续完善相应数据集的多样性，以提高模型泛化能力，同时进行一些更专业的市场需求调研，在系统上部署被需要的功能，完善系统提高整体用户体验。

我们设计的初衷是让技术更好地服务于社会，解决实际的民生问题。希望我们的作品能够在未来真正进入家庭，为人们排忧解难！

# 参考文献

【请按照标准参考文件格式填写】

[1]Zhe Cao ,Tomas Simon.Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields.The Robotics Institute of CMU,In CVPR, 2016.

[2]Shih-En Wei, Varun Ramakrishna,Yaser Sheikh .Convolutional Pose Machines.

The Robotics Institute of CMU,In CVPR, 2016.

1. MSCOCO keypoint evaluation metric. http://mscoco.org/dataset/#keypoints-eval.
2. V. Ramakrishna, D. Munoz, M. Hebert, J. Bagnell, and Y. Sheikh. Pose Machines: Articulated Pose Estimation via Inference Machines. In ECCV, 2014.
3. http://fenix.univ.rzeszow.pl/~mkepski/ds/uf.html