

**人体动作识别**

**Human Action Recognition**

**Computer Vision**

目录

研究意义 1

算法研究 2

基于单幅图像的姿态估计与动作分类 2

算法创新 2

算法优势 2

算法公式 2

基于卷积神经网络的人体动作识别 3

算法原理 3

实验结果 4

GUI设计 5

个人分工讲解 6

个人总结 7

1. 研究意义

机器在许多方面有着人们所不能比较的准确性，随着人工智能技术的兴起，许多相对于人类来讲简单低能的脑力识别、判断以及分析工作由机器来替代会有更好的效果。人体动作识别一直在体育、刑侦等多个领域的可视数据处理方面有着极高的需求，比如以下这些方面：

1. 行为分析和监控：人体动作识别可以应用于视频监控系统，通过识别和分析人体动作来检测异常行为、实时跟踪目标、提供智能安防等功能。这对于公共场所的安全和人员管理具有重要意义。

2. 人机交互和虚拟现实：通过人体动作识别技术，可以实现更自然、直观的人机交互方式，例如手势识别、身体姿态追踪等。这对于增强现实、虚拟现实和游戏等领域具有很大的潜力。

3. 健康与运动监测：人体动作识别可应用于健康管理和运动监测领域。通过分析人体动作，可以实时监测和评估人体的运动状态、姿态和运动技巧，为康复训练、体育训练和健身提供支持。

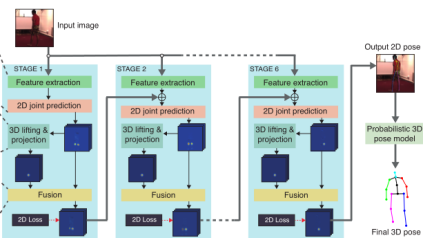
4. 视频内容分析和检索：人体动作识别可用于视频内容的自动标注、分析和检索。通过识别视频中的人体动作，可以帮助建立更智能的视频搜索和推荐系统，提高视频管理和分析的效率。

5. 人体行为理解与建模：人体动作识别可用于研究人类行为、社交互动和情感表达等。通过对人体动作的分析和建模，可以了解人类行为背后的含义和目的，进一步推动人工智能在心理学、社会学和人类行为学等领域的应用。

所以人体动作的识别是十分具有研究意义的，本小组将结合两种算法，从经典的骨骼结构标记方面对人体动作识别进行研究

1. 算法研究

* 基于单幅图像的姿态估计与动作分类（使用openpose识别动作）

首先，必须在图像中找到人体关节或地标的二维位置，由于不同的摄像机视点、外部和自身遮挡或服装、体型或照明的变化导致视觉外观的巨大变化，这是一个困扰模糊性的问题。其次，将2D地标的坐标从单个图像提升到3D仍然是一个病态问题——与人类的2D地标位置一致的可能的3D姿态空间是无限的。找到与图像匹配的正确3D姿态需要注入额外的信息，通常以3D几何姿态先验和时间或结构约束的形式。

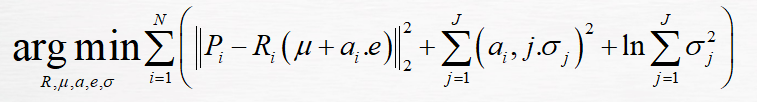
人体姿态模型是独立训练的，专门从3D动作捕捉数据。

在这项工作中，我们展示了如何将预学习的3D人体姿势模型直接集成到一个新的CNN架构中。（如左图）

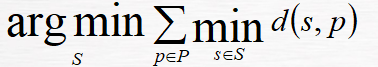
1. 算法创新

我们的算法在训练过程中包括一个基于人体姿势的概率3D模型的新层，负责将2D姿势提升到3D，并将有关骨骼结构的3D信息传播到2D卷积层。这样能够利用三维层的信息影响二位层信息使2D位置与姿态信息更加精确。

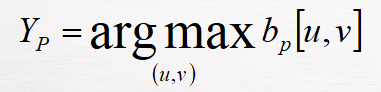
1. 算法优势
2. 算法是端到端的训练模式，大大节省了云端训练上传下载所需要耗费的时间以及可能出现的错误
3. 算法是2D和3D层面的联合训练，数据集独立性很强，进行训练集的独立增强时并不会影响到另外一个层面的训练集
4. 算法基于3D层面的动作捕捉以及2D层面的人体关键关节部位的标注训练而成，所以一方面识别过程中能够更清晰的定位到人体关节的二维位置，另一方面避免了二维图像向三维转换时出现的单对多映射模糊问题。
5. 算法公式
6. 因为图像的特征会因图像旋转而改变，所以在找3D姿态模型时要考虑旋转不变形，为每个姿势寻找最佳旋转公式是将给定一组n个训练3d姿势，寻求平均三维姿态u的全局变量，并通过正交基矩阵和噪声方差及旋转因子找出旋转最小化值



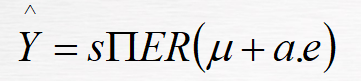
1. 为了使3d估计模型密度估计较强，采用正确初始化算法，首先对对齐的姿势进行子采样，然后计算对之间的欧几里德距离d寻找一组k个样本s，使得点和他们最近的样本之间的距离最小化



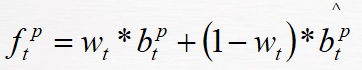
1. 寻找地标位置



1. 将3D姿势投影到2D模型上



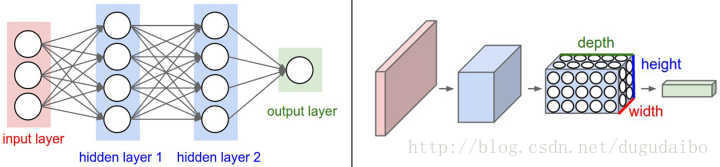
1. 概率3D姿态模型预测的2D置信图根据以下方程与基于CNN的置信图bp融合



* 基于卷积神经网络的人体动作识别（使用mediapipe识别动作）

算法原理

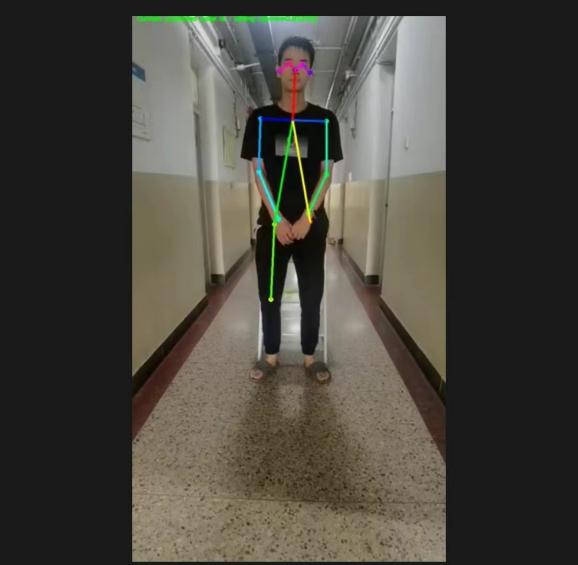
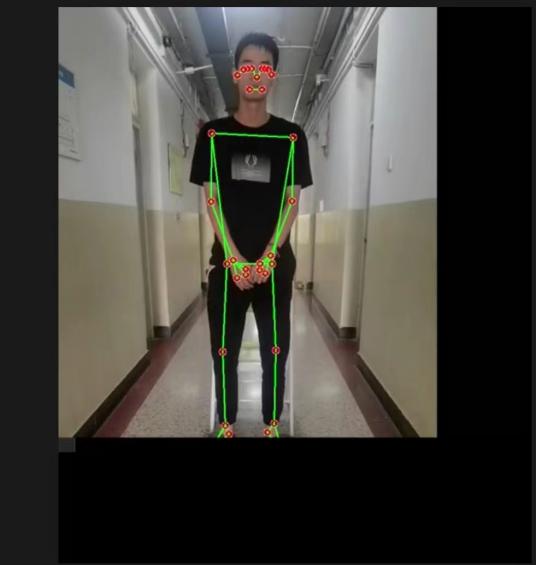
新的算法采用mediapipe库的内置动作标记方法，在图像识别骨节构判断上拥有着更多的点位（33）个，通过卷积模型的优化，带来了骨节构点稳定性，以及处理后视频的流畅性，解决了上面算法存在的骨节构点标记不准，生成视频不流畅以及动作识别准确性差的缺陷，但是本算法仅仅采用了单姿态估计，所以仅仅只能识别单人的动作模式。

图中左侧是一个3层的神经网络；右侧是一个卷积神经网络，将它的神经元在成3个维度（宽、高和深度）进行排列。卷积神经网络的每一层都将3D的输入数据变化为神经元3D的激活数据并输出。在图的右侧，红色的输入层代表输入图像，所以它的宽度和高度就是图像的宽度和高度，它的深度是3（代表了红、绿、蓝3种颜色通道），与红色相邻的蓝色部分是经过卷积和池化之后的激活值（也可以看做是神经元） ，后面是接着的卷积池化层。

我们从每个训练视频中采用50-200张图片不等，通过打标签的形式分类动作，用以上卷积神经网络训练，得出我们的模型，然后在卷积的深度方面采用2个数据体的选择，从而可以基于两个训练集去判断动作，提高判断的准确性。

1. 实验结果

* 上文已经提到，老算法由于图像识别的方式研究出的时代有些久远，即使功能方面更加的全面但是在处理准确性和稳定性方面都有着很大的不足，对训练集有着较高的要求，所以我们选择了相对来说更加适配我们所能构建到的训练集的基于mediapipe库的人体动作识别方式。

如上图所示，左侧基于openpose的处理方法，骨骼结构识别易受环境光暗的影响，存在部分未识别标记的情况。而且在我们的视频测试中，其标记点线存在闪动的问题，而右侧的mediapipe处理方式就很好的解决了这一点，点位稳定，且识别不易受影响。

* 本次实验我们构建了8组不同的训练集，对八个组合动作进行了识别，算法会对动作与原图进行了匹配度测试，匹配度较高的一方会呈现在屏幕左上角，并展现出匹配度。我们分别用训练视频和测试视频一同放置到训练好的模型中进行匹配，测试结果如下：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 训练视频1 | 训练视频2 | 测试视频 |
| 篮球运球+投 | 99%-100%（运） | 99%-100%（投） | 二者均99%-100% |
| 仰卧起做 | 70%-99%（仰） | 70%-92%(卧) | 仰是70%-99%卧是50%-92% |
| 跑步-走路 | 99%-100%（跑） | 99%-100%（走） | 二者均99%-100% |
| 坐-站 | 99%-100%（坐） | 99%-100%（站） | 二者均99%-100% |
| 挥舞-摆腿 | 99%-100%（挥舞） | 99%-100%（摆拳） | 二者均99%-100% |
| 勾拳-直拳 | 99%-100%（勾拳） | 99%-100%（直拳） | 勾拳100% |
| 翘-坐 | 60%-90%（翘） | 70%-95%（坐） | 坐的比例更大一些 |
| 波比跳 | 80%（俯卧） | 98%（跳） | 98%（跳）-80%（俯卧） |

* 98%（跳

由实验结果可以看出动作区别相对来讲越大，实验模型的识别会更加准确，比如篮球的两个动作，投篮和运球，骨骼结构的方向完全不同，识别效果最好，区分度很高，而且识别匹配度也非常高。而波比跳这一块，显然由于跳的过程中手臂前倾，类似于俯卧撑较大幅度的上抬，所以训练时会认为俯卧中某些动作属于起跳，从而会造成识别匹配度相对较低的情况出现。

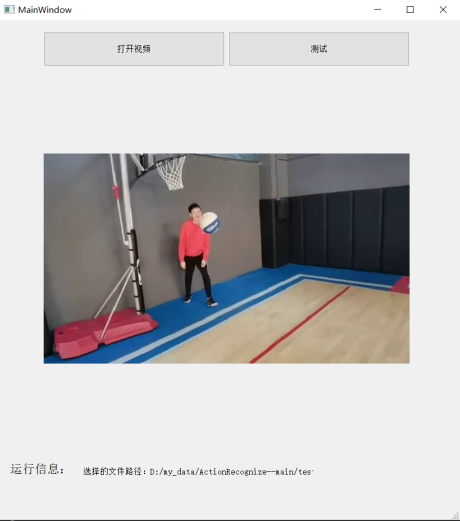
通过实验我们大致总结出了算法存在的如下优点：

1、 单人检测框架，检测速度特别快，同样的的一段检测对象，同样的使用CPU进行人体姿态检测，会检测更快。

1. 我们可以在图像或视频中准确地检测人体的关键关节位置，从而识别人体的姿态。能运用在很多场景，比如选用素材中的运动健身，动作捕捉。

3、 基于机器学习技术，结合了深度学习和计算机视觉算法。 它能够实时地从图像或视频中提取人体的姿态信息，并给出关键关节的坐标和置信度。

1. GUI设计



我们的GUI界面大概如上图所示，采用python中所带的QT相关扩展包所设计，大致分为如下几个模块：

1. 命令模块，通过“打开视频”按键可以定位到测试文件夹层中的测试视频，通过cv相关函数读取视频并传送给可视化模块。通过“测试”指令训练模块的代码在训练集中提取视频进行训练，并与测试视频进行对比分析，标记处人体骨骼结构，将分析结果呈现在视频左上角。
2. 可视化模块，位于界面正中间，可以将命令模块传输过来的视频播放出来。
3. 提示模块，即为图中的运行信息这一部分，运行测试视频时会呈现其路径，处理视频时会出提示语“处理中请稍等”，处理结束后会显示“处理完成”

动作判断和动作标准度（与训练集的匹配程度）会呈现在处理视频的左上角。

1. 个人分工讲解

* 代码与算法层面

1. 环境配置与搭建：因为我电脑的环境是python3.7，和代码本身自带的虚拟环境python3.10不符合，所以我通过下载了一个anaconda然后在设置中利用它给代码创造了虚拟环境，并安装了代码所需要的包，中间不小心把环境建成了conda环境，了解到虚拟和conda环境不一样，会出现解释器路径寻找的相关问题，最终也是成功解决，跑通了代码。
2. 训练集的创建和模型的优化：为加深我们对代码的理解，训练集其实是我们自己所拍摄并设计的动作视频总共8组，我负责其中五组的创建以及模型训练，在过程中，我发现了几个问题：

* 训练算法对于骨架朝向非常的敏感，有一组战立坐下的视频，我们双手捂裆拍摄由于手臂关节会在体前交叉，所以即使我们后面拍摄的双手下垂也属于坐的范畴，但是识别出的动作匹配度就不高，甚至会出现将坐识别成为站的情况出现。所以我选择了用捂裆或者不捂裆两个坐姿去训练，增加模型的数据类别提高准确度，或者用和训练集相近的视频测试，相对来说得到了更好的效果。
* 训练算法需要足够多的图片，才能够识别出一个动作，尤其是作为训练的两个动作不同但骨骼结构相近的情况下。简而言之就是训练集每个动作最好持续时间稍长一些，至少能达到一秒，这样的话才能够判断出不同的两个动作。

3、8个数据集测试匹配度的归纳与整理：在我们训练集构建完成后，我们对其进行了动作识别的测试，通过模型的优化调整，以及训练集、测试集内容的更换大致得出了较为可靠的训练数据，也就是系统所带的动作匹配度，了解到模型对于骨节构朝向多元化的复杂动作具有更高的识别能力，加入动作分明停顿较长的训练集会加强识别效果。

1. 最后小组代码的整合（因为后期训练测试以及GUI制作是同步进行的）

* 非代码层面

1. 每次小组阶段性的成果展示都是由我来汇报的，汇报前需要沟通了解组内成员所查找的资料，并作出一定的实现，比如跑代码或者是搭环境。
2. 第三次小组ppt关于算法介绍内部分的制作是我通过论文归纳总结出来的。
3. 最后大作业的成品文档，除了最后两个部分，即“个人分工讲解”和“个人总结”以外的公共部分都是我完成的写作。
4. 个人总结

在完成这次大作业的过程中，我感觉我个人得到了非常大的提升。

首先我负责一部分代码方面的工作，所以需要去做环境方面的搭建，之前我们的python课由于疫情原因没有对我们进行项目工程能力的培养，所以我可以说对于pycharm环境搭建所知甚少，唯一的了解可能还是源于大一的时候需要在pycharm跑几段基础代码积累的经验，但是面对许多要装的包以及解释器的更新换代，我遇到了许多的困难，好在最后通过网上查找以及某些编程能力强大的同学的点拨，我解决了环境搭建的问题并收获了许多关于python语法还有系统性能的知识。

其次来谈一谈代码方面的收获，由于我们本次大作业前后总共用了两种算法去测试，我了解到了两种算法的异同以及优点和缺点。在动作识别方面OpenPose 使用卷积神经网络（CNN）作为其主要的姿势估计算法，并将多个阶段的网络结构组合在一起。相比之下，MediaPipe 采用了一种称为全身关键点（Whole Body Pose）推断的方法，通过联合估计三维关键点坐标和姿态方向来实现。稳定性相对来说是后者更佳，准确性的话后者算法适合处理一些轻量级的视频数据，比如我们这次用于实验的这些数据，但是就网上查到的资料和实验数据而言，其实对于大量的数据，在有足够的训练数据做基础的情况下，其实openpose的准确性会更高一些。

作为数据测试者，我认识到构建一个有效的训练集对于人体动作识别的成功至关重要。在构建训练集时，我明确了研究目标并确定了需要收集的数据类型。我从多个来源搜集了大量的人体动作数据，包括各种不同的动作姿势和角度。通过对数据的筛选和处理，我确保了训练集的多样性和代表性。我们也能够根据训练需要，实时去调整模型以符合我们的需求。

在代码之外，通过小组的沟通合作，我的交流能力得到了极大的提升，因为所有人的代码能力有所区别，而且各自任务完成进度也会有区分，任务与任务之间的配合，以及我向某些同学求教解决代码的问题，这些都非常的锻炼交流能力。在这次大作业合作中我的拖延症也基本消失，因为小组的组员相比我更喜欢早些完成任务，所以我很多事情都不会在拖，会早一些去完成。由于我主要承担组内的汇报演讲内容，这也提升了我在公共场合发言的逻辑能力。

总而言之，通过这次人体动作识别的大作业，我了解到了关于python许多的语法方面的知识，也了解了人体动作识别目前两种比较厉害的算法包，以及算法的优劣特性。同时，我也锻炼了我的组织和演讲能力，使我更加成熟和自信地应对未来的挑战。我期待能将这些经验应用于更广泛的领域，并持续不断地提升自己的能力。