# 

# 人体动作识别项目说明文档

**成员：刘媛媛、刘艾林、毕曼珊、朱睿睿**

**指导教师：刘铁**

**目录**

[**人体动作识别项目说明文档 1**](#_Toc1932818974)

[**一、研究现状 2**](#_Toc1541146402)

[1、研究背景 2](#_Toc1963642209)

[2、国内外研究现状 2](#_Toc217836114)

[2.1动作特征提取 2](#_Toc161790113)

[2.2动作分类方法 3](#_Toc695901920)

[**二、问题定义 3**](#_Toc1489731899)

[**三、数据集介绍 4**](#_Toc1145320317)

[1、数据采集 4](#_Toc1465870376)

[2、数据集相关获取方式 4](#_Toc468503644)

[**四、算法介绍 5**](#_Toc1040143211)

[4.1OpenPose网络结构算法 5](#_Toc385331939)

[4.2OpenPose人体动作模型构建算法 6](#_Toc1208014869)

[4.3动作分割算法 7](#_Toc1604672481)

[4.4基于LSTM的行为识别网络 8](#_Toc1065649439)

[**五、代码说明 10**](#_Toc176383776)

[5.1OpenPose 10](#_Toc20563062)

[5.2tf\_pose\_estimation 11](#_Toc1590703788)

[5.3tools 13](#_Toc223904226)

[5.4utils 13](#_Toc2063560746)

[**六、系统演示 13**](#_Toc1755290096)

[**七、参考论文 13**](#_Toc686612625)

**具体分工如下：**

ppt汇总：刘媛媛

运行说明文档：朱睿睿

研究介绍-数据集，参考论文，格式排版：毕曼珊，刘媛媛

算法介绍：刘艾林

代码，系统演示：朱睿睿

# 研究现状

## 1.1研究背景

人体动作识别项目的研究背景根植于对人类行为理解和技术应用的持续追求。人体动作识别技术在智能监控、视频检索、人机交互、运动分析等诸多领域发挥了至关重要的作用。在安防监控领域，人体动作识别技术有助于预防违法犯罪等异常事件的发生，能够准确识别出正在进行的危险行为，从而及时发出预警，有效保障公共场所的安全；在视频检索领域，人体动作识别技术能够协助相关人员高效完成视频检索任务，从大量视频数据中快速定位到特定动作，检索到所需目标；在人机交互领域，人体的动作则是人机交互的桥梁，通过对人体动作的准确识别，能够让机器更加正确地理解人们要表达的意图，从而更精准地完成相应的工作。

随着信息技术和人工智能的迅猛发展，人们对于如何通过技术手段精确捕捉、分析和理解人体动作的需求日益增长。这一研究领域融合了多个学科的知识和技术，包括计算机视觉、机器学习、传感器技术以及人体运动学等。在传感器技术方面，随着传感器的小型化、多样化和成本的降低，惯性传感器、摄像头等能够高精度地捕捉动作数据的设备得到了广泛应用，为动作识别的实时性和准确性提供了坚实的技术基础。随着计算机技术与网络技术的快速发展，多媒体信息呈爆炸式增长，其中视频作为重要的信息载体，越来越受人们青睐更多信息通过视频进行传播。

算法设计是人体动作识别研究的核心，机器学习和深度学习方法在此扮演着重要角色。通过这些方法，研究人员能够处理从传感器获取的大量数据，并从中提取出有意义的动作特征，实现对复杂动作的自动识别和分类。此外，随着深度学习技术的不断进步，特别是卷积神经网络（CNN）和循环神经网络（RNN）的应用，使得对动作模式的识别更加精确和高效。

## 1.2国内外研究现状

## 1.2.1动作特征提取

能否提取出适当特征来充分表达人体行为信息，对后续的识别准确性至关重要。动作特征提取的主要任务是根据人体动作的特征对不同动作进行识别和分类。人体动作识别的研究现状主要集中在动特征提取和分类器设计两个方面。

在动特征提取方面，国内外研究者提出了多种不同的特征提取方法，包括基于颜色、形状、运动和深度等方面的特征。其中，基于运动的特征提取方法被广泛应用于人体动作识别中，其主要包括光流、运动轨迹、骨骼关节点等特征。国内外的研究者通过在实验数据集上的大量实验验证，证明了这些特征在人体动作识别中的有效性。

在分类器设计方面，国内外的研究者主要采用支持向量机、神经网络、决策树等机器学习方法来进行人体动作的分类。近年来，深度学习方法在人体动作识别领域的应用越来越普遍，其中卷积神经网络是最常用的深度学习方法之一。通过深度学习方法，研究者可以自动学习并提取人体动作的高层次特征，从而获得更好的分类性能。

Yamato等人通过选取图像轮廓信息的方法表示人体行为特征，首先对RGB图像二值化处理，再利用边缘检测的算法提取出目标轮廓像素，最后将提取的轮廓信息同标准模板库中的动作模板进行匹配给出最终结果；Carlsson等人则是对边缘检测算法进行了优化，提高了边缘信息提取的准确度；minchisescu设计了对图像信息分块采样的方案，在时间维度上采样图像信息也取得了不错的结果。总体来说，基于RGB图像提取动作特征的方法有一定的优势，但同时也有一个很大的缺陷，就是极其容易受光线、遮挡的影响，使得识别精度降低严重。

## 1.2.2动作分类方法

获取动作特征值信息之后，就需要选择或者设计分类算法完成动作分类的任务。人体动作识别中的动作分类算法是研究重点之一，其主要任务是将提取到的人体动作特征进行有效分类，以识别出不同的动作。国内外的研究者们在动作分类算法方面做出了一系列的研究和探索。

Bobick等人提出了基于运动能量图的方法，其思路是依据视图的变化来实现动作识别；Karpathy等人对帧数据的局部时空信息进行学习，提出了通过在时域上对CNN网络连通性进行扩展方法识别率有提升显著；Srivastava首先将输入数据转化为固定的长度，然后通过堆叠多个LSTM单元完成分类任务，其实质是通过使用多层的LSTM神经网络来实现分类的任务，这使得动作识别效果得到大幅度提升，算力降低。

存在问题：上述研究人员的方法均是倾向于对网络模型进行改进，又或者是帧数据像素级别的处理来提高识别精度。而即使是利用了骨架关节点的信息，也没有考虑关节之间的夹角信息及关节之间的距离的长短对最终动作识别精度的影响。

# 问题定义

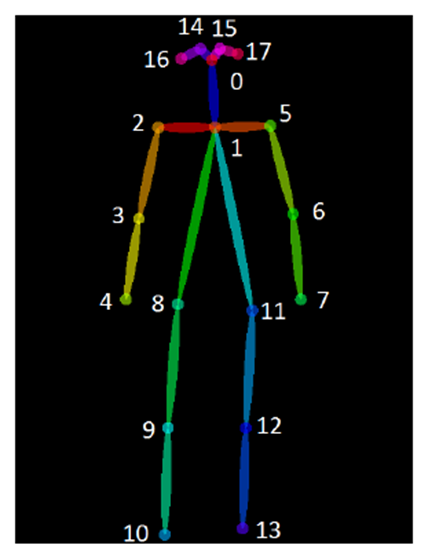
人体动作识别项目是通过计算机视觉技术识别和分析人体在运动中的各种动作，包括但不限于行走、跑步、跳跃、踢腿等。该项目旨在利用人体关键点检测和动作识别算法，实时地监测和识别人类的动作，为健康管理、体育训练、游戏互动等领域提供数据支持和智能应用。在这个项目中，主要涉及到的问题包括数据采集和标注、动作特征提取和建模、模型训练和优化以及实时动作识别和应用等方面。

首先，数据采集和标注是人体动作识别项目中的第一步，需要通过摄像头或传感器等设备获取人体运动的视频或数据，并对其进行标注，标注方式多种多样，如关键点标注、骨架标注等。其次，动作特征提取和建模是关键环节，通过提取人体关键点或骨架的空间位置、运动轨迹、角度变化等特征，建立数学模型描述各种动作。然后，模型训练和优化是项目中的核心任务，通过深度学习、机器学习等方法对大量的标注数据进行训练，优化动作识别模型的准确率和鲁棒性。最后，实时动作识别和应用是项目的最终目标，通过将训练好的模型应用到实际场景中，实现对人体动作的实时监测、分析和识别，为健康管理、运动训练、虚拟现实等应用提供支持。

通过解决以上问题，将能够实现对人体动作的全面识别和分析，为人们的健康管理和运动训练提供更加智能和便捷的支持。

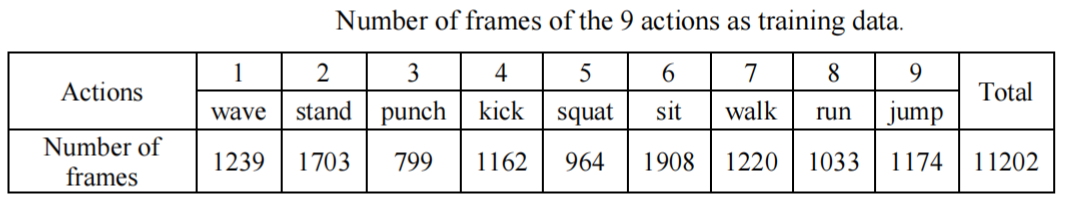
# 数据集介绍

## 3.1数据采集

本文的人体关节点获取部分，直接使用的OpenPose训练好的模型，因此本系统运行时使用笔记电脑本内置摄像头即可。关节点数据模型采用COCO数据集格式，其数据集格式说明如下：

0号位置是鼻子,1号位置是颈部，2号位置是左肩膀，5号位置是右肩膀，3号位置是左胳膊肘，6号位置是右肘，4号位置是左手腕，7号位置是右手手腕，8号位置是左髋关节，11号位置是右髋关节，9号位置是左腿膝盖，12号位置是右腿膝盖，10号位置是左脚踝，13号位置是右脚踝，14号位置是左眼，15号位置是右眼，16号位置是左耳，17号位置是右耳和18号背景信息点。通过OpenPose预先训练好的模型，获取这19个骨骼关节点的x轴信息，y轴信息，还有每一个点的置信度c，代表该关节点的识别准确程度。本系统在特征提取方面只需要x轴信息，y轴信息，没有加入置信度。

本项目收集了9种数据格式的视频数据分别是['stand','walk','run','jump','sit','squat','kick','punch','wave']，每个视频的长度从0.8秒到2分钟不等，并且每个视频仅限包含一种类型的操作。例如，在一个视频中，我踢了0.8秒；在另一个视频中，不停地挥舞着手臂长达2分钟。这些视频是以640x480的大小和10帧/秒的帧速率记录的，数据集分布如下表所示：



## 3.2数据集相关获取方式

KTH数据集：2004年发布，包含6类人体行为：行走、慢跑、奔跑、拳击、挥手和鼓掌，每类行为由25个人在四种不同的场景（室外、伴有尺度变化的室外、伴有衣着变化的室外、室内）执行多次，相机固定。该数据库总共有2391个视频样本。视频帧率为25fps，分辨率为160×120，平均长度为4秒。（官网：<https://www.nada.kth.se/cvap/actions/）>

Weizmann动作检测数据集：数据同样是固定镜头下的10个典型动作的视频，同时数据集提供了一些带有其他物体的动作作为干扰，可以测试模型的鲁棒性。官方同时提供了去除背景的程序，但是数据集的数据量比较少的90组常规数据和21组鲁棒测试数据，对于目前的模型训练来说显得有些不足，不过对于本来就需要用小数据的模型比如迁移学习或者One-shortLearning来说或许是适合的数据集。（官网：<http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html）>

HMDB51动作检测数据集：51个类别可以被分为5个大类：常见的面部动作(smile，laugh，chew，talk)、复杂的面部动作(smoke，eat，drink)、常见的肢体动作(climb，dive，jump)、复杂的肢体动作(brushhair，catch，drawsword)、多人交互肢体动作(hug，kiss，shake,hands)。

# 四、算法介绍

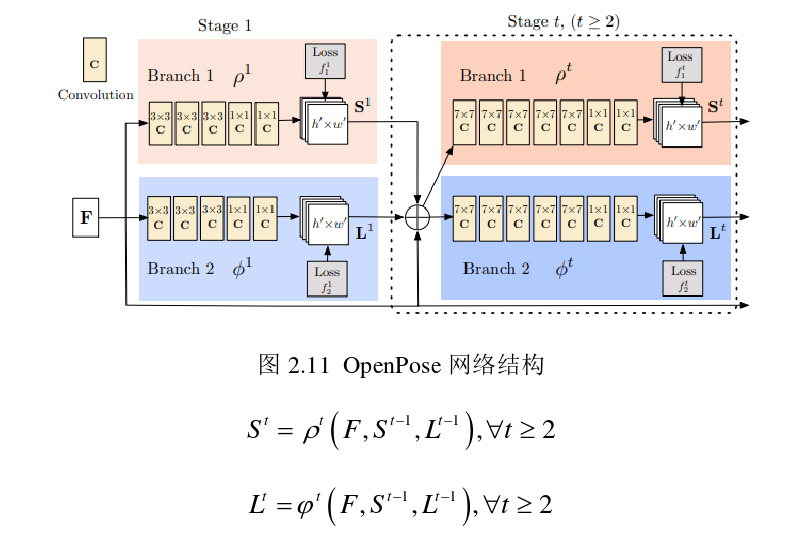
## 4.1OpenPose网络结构算法

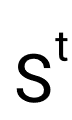
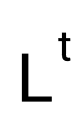
与网络模型法的改进，帧数据像素处理这两种方法相比，openpose不仅利用了骨架关节点的信息，还考虑关节之间的夹角信息及关节之间的距离的长短对最终动作识别精度的影响。OpenPose为了解决以上的难点，提出了关节点亲和场（PartAffinityFields,PAFs）的概念方法。本质上来说，他其实是一个二维矢量场的集合，包含了肢体的位置和方向信息。可以通过骨架各关节点位置信息和PAFs,计算出两个关节点的关联程度。这对后面由关节点拼接成人至关重要。

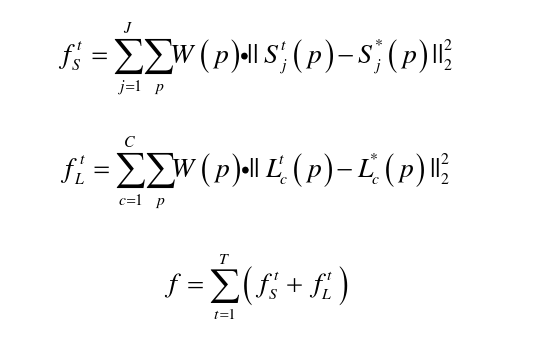
执行过程如下图所示，首先将图像数据输入到网络，然后分别经过两个网络分支，一个去预测图像中的所有关节点位置信息，另一个去学习预测关节点之间的向量指向信息，输出总的PAFs。然后转化成图论中的匹配问题：边权值最大的二分图匹配。由点拼接出人的骨架。



原图首先经过卷积网络，得到特征图谱，记为F，经过如下图所示的网络。该网络分上下两个分支，每个分支都有t个阶段（表示越来越精细），除了起始阶段，后面每个阶段的输入除了该阶段原本的输入f，还包括上一阶段两个分支的输出，具体表示如下图公式所示。



其中，为t阶段输出的关节点置信图，为t阶段输出的两两关节点的亲和度，也就是权重系数。损失函数是保证网络能收敛的最重要的关节点，因此作者对两分支的损失函数均采用L2损失。训练时，每个阶段都会产生损失，为了避免梯度消失，预测时只使用最后一层的输出。阶段损失值公式如下图所示。



其中，带上标\*的表示真值，带上标t的是不同阶段的预测值，p是每一个像素点，W(p)代表示该点缺失标记，只有0和1两个值。若为0，则损失值不予计算。总体的损失值公式见第三行公式，为各个阶段的损失值之和。

## 4.2OpenPose人体动作模型构建算法

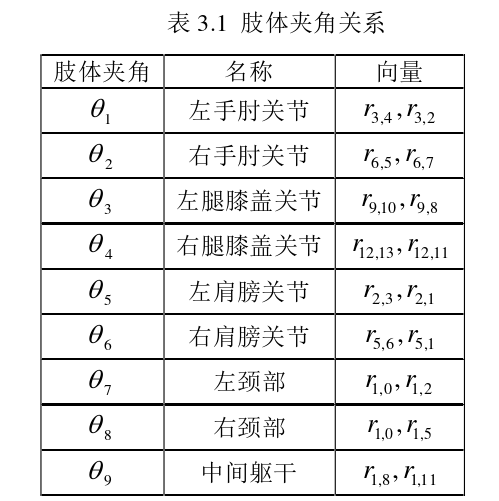
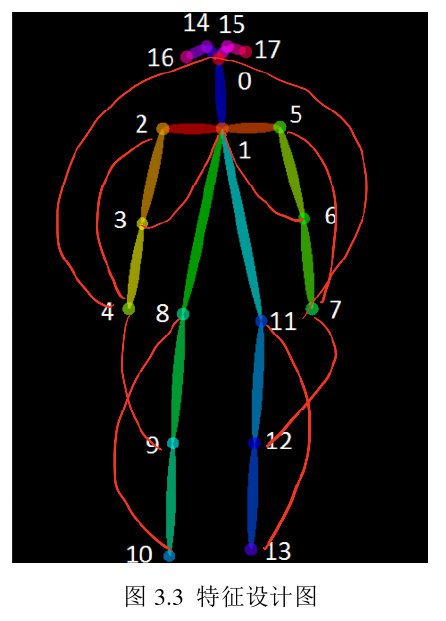
这部分主要介绍人体动作模型构建方法，分为两个部分，骨骼关节点规整化方法及特征设计方法。

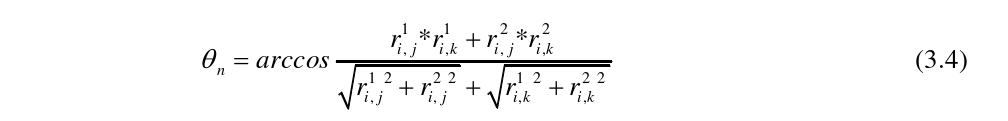
第一，因为人的距离，以及人体高矮胖瘦之间的差异，会严重影响最终的识别准确率。即使是同一个人做出了相同的动作，但因为与摄像头距离的不同，使得获取的关节点位置信息相差很大，导致最终的错误识别。因此,设计了对关节坐标信息进行规整化处理的方案。

首先通过数据获取模块读取单帧视频信息，将其送入OpenPose预先训练好的模型中，获取19个关节点坐标信息，其中因为不需要背景信息，所以去除掉第19号坐标点。然后对每个关节点以颈部坐标点为基准进行坐标规整化操作，假设颈部位置坐标，然后推出各个关节点的坐标。可以看到C表示人体所有关节点位置的集合，E表示肢体向量的集合。在对关节点位置信息规整化处理之后，也就能表示出某个时刻也就是某个帧的关节点信息。

第二，在图片或者视频帧中，人体的两个肢体之间形成的夹角，称之为肢体夹角。当人体在做任何动作的时候，实体表现特征就是身体各个肢体的变化。奔跑动作是随着手臂的摆动和腿部肢干的活动实现的，标准敬礼动作则是由右手臂的弯曲完成的。因此，人体各肢干之间夹角大小的变化，能够充分包含当前人的行为特征信息。所以提出全量肢体夹角特征设计方法，提取出人体所有肢于间的夹角用于表征特征信息。

表中r（i,j）是指肢干连接的向量，从关节i指向关节j。其中,n取值0到8,表示第n个夹角，r1i,j表示i，j两节点组成的肢干向量的第一个值，r2i,j表示i，j肢干向量的第二个值。因此，基于肢体夹角，一张单帧图像的人体动作特征就可以用夹角的值来表示。



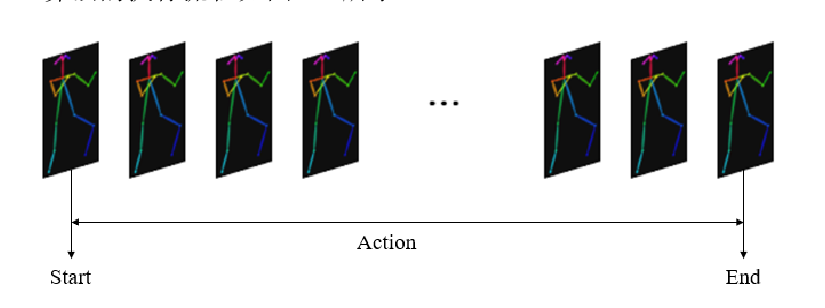


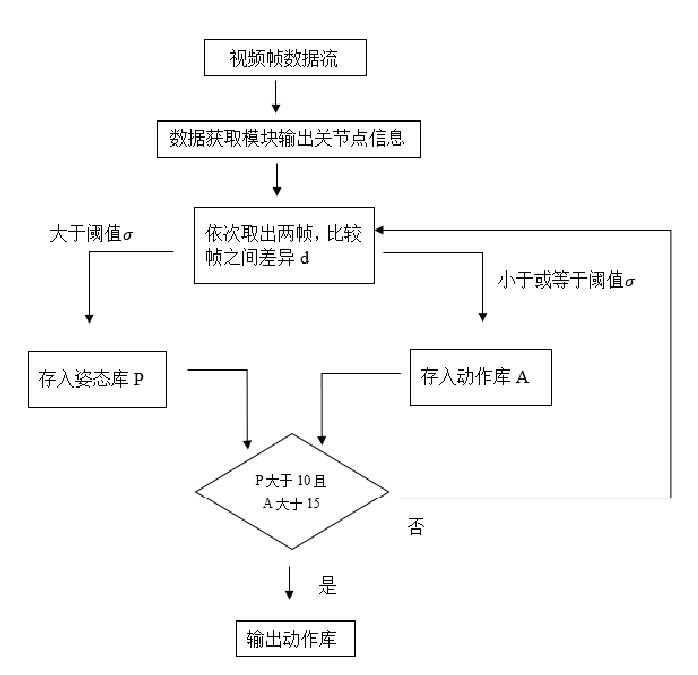
## 4.3动作分割算法

基于一套完整动作由起始动作、执行动作、结束动作组成的思想，提出动作分割算法。

通过对多帧视频流的时空信息矩阵数据进行裁剪，比较相邻帧之间人体各关节点变化情况，确认起始帧、动作执行帧以及动作结束帧。由这三部分构成一个完整动作所需的时空信息矩阵，裁剪出动作的时空信息矩阵，剔除掉冗余视频帧数据，减少计算量，提高识别精确度。

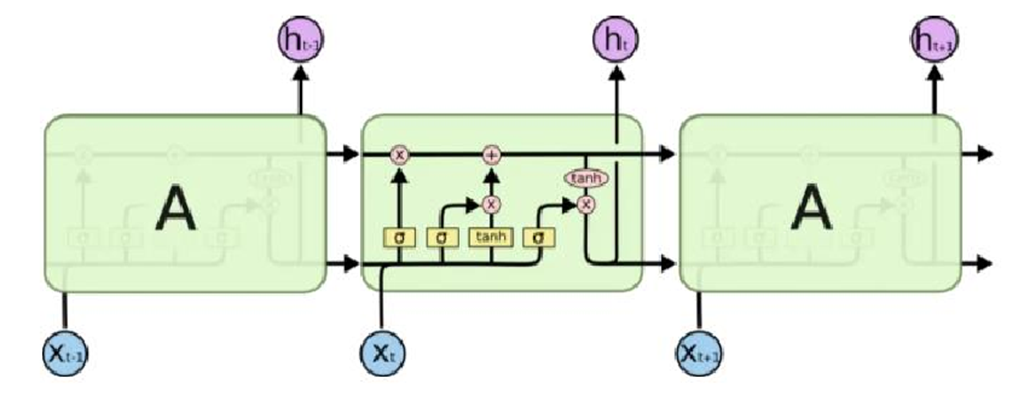
对于输入的连续视频流数据，首先通过OpenPose网络模型获取每一帧的人体关节点坐标信息，得到该输入视频的关节点时空矩阵信息，该矩阵信息作为算法的输入。然后从第二帧开始，依次取出每一帧信息同其前一帧进行比较，计算两帧之间的方差d，即：方差d的大小表示两帧之间的差异程度。然后将该方差同预设的阈值进行比较。其中阈值的取值设定为0.5。如果d超过了阈值，认为该帧为姿势帧，存入姿态库P中，否则认为该帧动作帧，存入动作库A中。一套连续的完整动作会在1秒钟内完成，假设每秒的视频帧按25帧计算，即一套完整的动作用25帧的数据量即可表示。因此当姿态库P所存储的帧数超过预设值十帧且动作库帧数超过预设值十五帧时，认为已经获取到一套完整动作的所有帧信息，输出由该姿态库和动作库组成的时空矩阵信息。该矩阵信息作为后续动作构建方法的输入。





## 4.4基于LSTM的行为识别网络

在处理无时序数据的时候，传统的神经网络能够很好的解决问题，但是当数据是诸如视频帧这样的时序数据的时候，效果往往不是很好。针对视频流的行为识别任务而言，一段几秒钟视频通常是包含几十帧甚至上百帧，我们是对每一帧提取特征数据进行处理，这实际上是每次只处理了一个时刻的视频帧数据，没有考虑到时间维度上，动作之间的变化情况，得出的准确率理论上也不会很高。所以不仅要看当前时刻的数据，还要看之前的数据信息，也就是需要对整个序列分析才能做出准确的判断。所以选用循环神经网络模型。其特点为上一时刻的输出变成下一时刻输入的一部分，以此充分利用获得的信息。但是循环神经网络有一个很大的缺陷，就是梯度消失问题。就是说，循环神经网络只能记忆较短距离的信息，随着时间间隔的增加，网络学习就会变得非常困难。而且，当前的输入包含了上一时刻的全部输出，并没有对其中的信息做筛选之类的操作而全部使用，导致计算量较大，且冗余信息过多。为了解决这些问题，循环神经网络出现了很多变体，其中包括了长短期记忆网络(LongShort-TermMemory,LSTM)，该网络能够很好的解决上述的问题。



LSTM重读的模块拥有一个不同的结构。不同于单一神经网络，这里包括了四个部分，各个部分之间以一种非常特殊的方式进行交互。“门”结构是LSTM网络的一个创新，该结构能够用来控制信息是否向后端传递。其本质就是用一个Sigmoid网络层和一个按位的乘法操作。Sigmoid层的输出是一个数值矩阵，该矩阵中的每个值都在0到1之间。

LSTM网络的搭建与执行过程。

首先，通过划分数据集，使用训练集训练获得网路的各项参数的最优值。

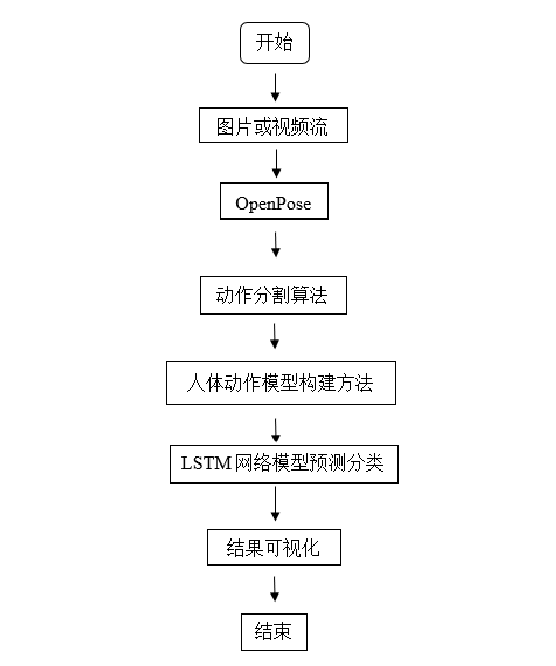
同时，为了优化网络的预测能力与效果，增加注意力机制，使得网络能够对输入的特征数据进行加权运算，进一步提高最终的识别精度。

LSTM神经网络模型的训练过程同大多数深度学习的网络模型一致，主要包含了三部分内容，前向传播、反向传播和梯度更新。

基于OpenPose与LSTM的人体行为识别系统，主要由数据获取、动作分割、动作模型构建、Opencv绘图、神经网络构建识别共五个模块组成。

在数据获取模块中，通过OpenPose对视频帧数据处理拿到关节点位置信息，输出关节点时空信息矩阵。动作分割模块用于对关节点时空矩阵的裁剪过滤操作。动作模型构造模块主要是对关节点信息进行处理，获得网络所需要的特征信息。

神经网络模块完成动作分类识别的任务,Opencv绘图模块完成骨架图的画面和识别结果的展示



# 五、代码说明

项目代码名称：Realtime-Action-Recognition-master

本项目中包含有，人体骨架识别和动作分割训练项目，一个完整的基于tensorflow深度学习框架实现的人体姿态识别项目，以及一些为了更好实现目标而添加的库和函数，除去整体项目各文件的说明，我们也将对后两者进行单独说明。

## 5.1Realtime-Action-Recognition-master

本项目代码的名称。

### 5.1.1config

该文件下属目录仅有一个文件：config.yaml

该文件包含了此项目的配置信息，主要用于处理动作识别相关的图像和骨架数据。

特别指出的是，其中声明了本项目识别的9类动作标签：

classes:['stand','walk','run','jump','sit','squat','kick','punch','wave']

即：站立、行走、奔跑、跳动、坐下、下蹲、踢腿、出拳、挥手。

除此以外，还包含了对各项数据的定义和声明，以及Python脚本文件和相关配置信息，包括骨架数据处理、特征处理、模型训练、测试过程中的配置，同时也包含对各项数据的输入输出路径的指定。

### 5.1.2data

存储训练图像的数据和标注信息。

### 5.1.3data\_test

存储测试用数据信息。

### 5.1.4doc

部分有关项目的说明文件。

### 5.1.5model

存放已训练好的分类模型数据，用以对输入的图像或视频实现已被验证效果较好的识别和分类功能。

### 5.1.6output

进行训练后的数据存放地址，包含标注好的视频和标注框的数据和标签。

### 5.1.7src

包含5个Python源代码文件，涵盖了从原始数据库训练模型到用训练好的模型分析输入的视频的代码和指令，各代码文件的功能说明如下：

1. s1\_get\_skeletons\_from\_training\_imgs.py

读取训练图像，使用OpenPose模型检测图像中的骨架，并将结果按指定格式保存到指定的输出文件夹中。

1. s2\_put\_skeleton\_txts\_to\_a\_single\_txt.py

将多个骨架数据文件合并为一个单一的文本文件，丢弃无效的文件，便于后续处理。

1. s3\_preprocess\_features.py

预处理特征数据，将骨架数据转换为特征向量，准备进行训练。

1. s4\_train.py

从csv文件加载特征和标签，训练模型，保存数据至指定的文件。

1. s5\_test.py

可测试本地视频或摄像头输入的图像数据，进行分析标注并进行可视化展示同时保存训练结果至指定的地址。

## 5.2tf\_pose\_estimation

该文件包含了一个较完整成熟的基于tensorflow深度学习框架实现的人体姿态识别开源项目，其原始项目是采用了caffe框架，使用c++编写的人体姿态识别（检测人体身体部分、手、脸、脚等部位）。

### 5.2.1docker

构建一个Docker容器，快速部署运行tf\_pose\_estimation项目所需的环境。

自动化部署和配置OpenCV、TensorFlow、COCOAPI以及tf-pose-estimation项目所需的环境和依赖。

### 5.2.2etcs

对此项目进行了详细的介绍和说明，包括数据来源，优化过程和灵感来源，参考资料来源等。

### 5.2.3images

存储了部分测试用图像数据。

### 5.2.4launch

包含一个ROS（RobotOperatingSystem）的启动文件XML代码，用于配置和启动多个节点来执行姿态估计、可视化、图像展示、视频录制和视频流处理等任务。

### 5.2.5models

包含调用多种不同用于预下载训练模型图形文件的脚本，本项目使用的是其中的cmu模型，即来自卡内基梅隆大学（CarnegieMellonUniversity）的模型。

### 5.2.6msg

其中含有3个消息类型的文件，分别传输“单个身体部位”“单人”“多人”的身体部位数据，用于传输关于人体各个部位的位置信息和置信度信息，帮助在ROS系统中实现人体动作识别功能。

### 5.2.7scripts

定义了两个ROS节点，分别用于处理图像数据并进行人体动作识别和接受上述数据并将数据叠加在原始图像上进行可视化显示。

### 5.2.8tf\_pose

tf\_pose是一个基于TensorFlow的人体姿态估计库，可以用来检测人体的关节位置和姿势。

## 5.3tools

包含了进行骨骼可视化的文件，不同文件格式转换的文件（图像和视频的互转，png转jpg）

## 5.4utils

lib\_classifier.py：分类器库，用于将骨架数据转换为特征，并进行分类

lib\_commons.py：通用工具库，包含一些常用的辅助函数

lib\_feature\_proc.py：用于从骨骼中提取特征

lib\_images\_io.py：从摄像头读取图像并将图像写入视频文件

lib\_openpose.py：OpenPose骨架检测器的封装，用于从图像中检测人体骨架

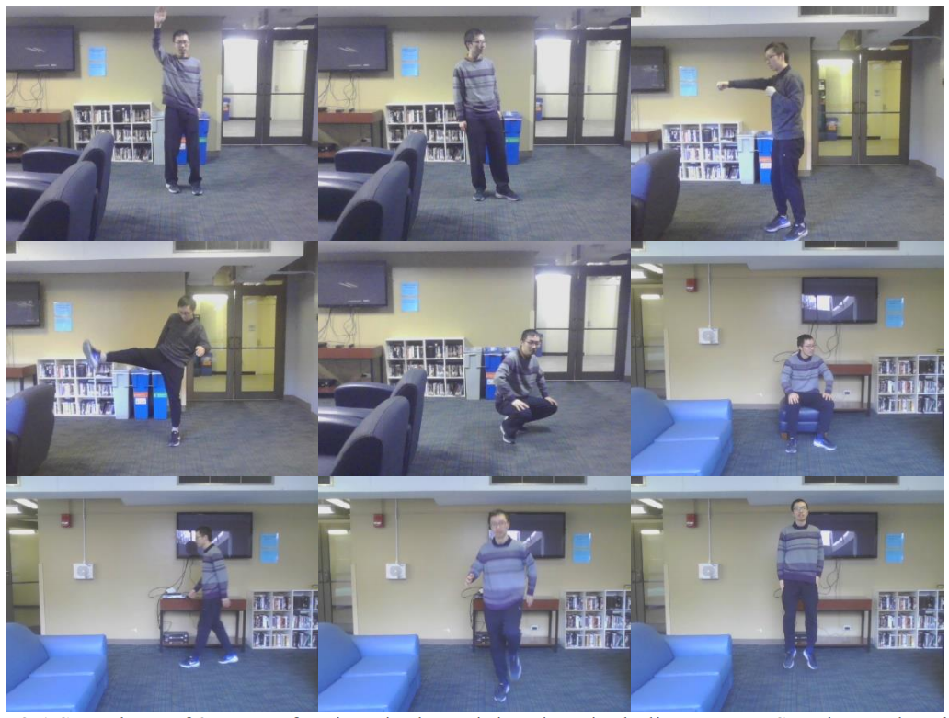
lib\_plot.py：提供绘制混淆矩阵和动作结果的可视化功能。

llib\_skeletons\_io.py：定义了骨架数据的格式和I/O操作，包括读取和写入骨架数据

lib\_tracker.py：跟踪器库，用于跟踪骨架数据

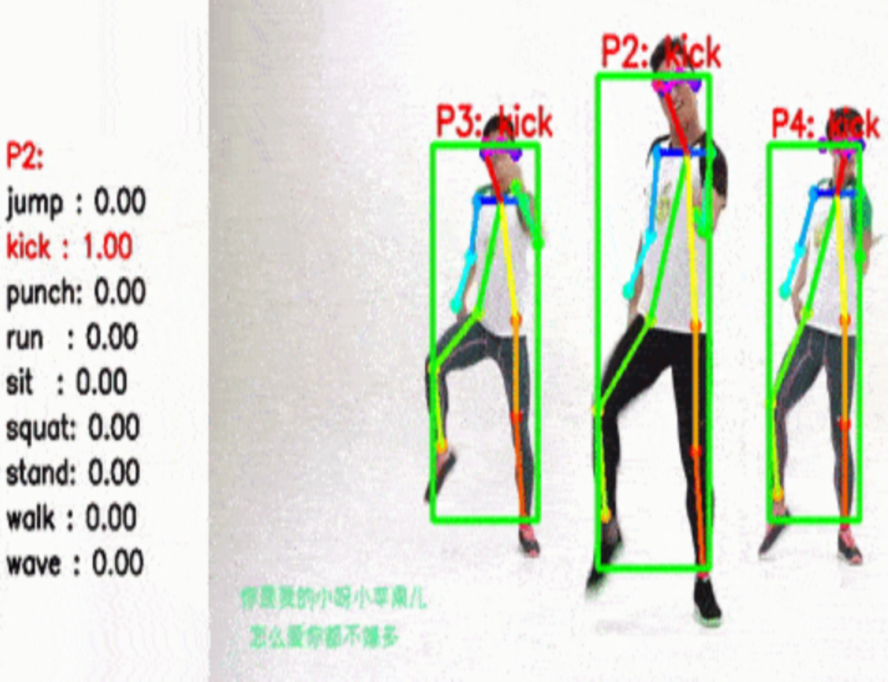
# 系统演示

本项目支持识别的9种动作：['stand','walk','run','jump','sit','squat','kick','punch','wave']



运行效果演示：





# 七、参考论文

[1]王子沁.基于OpenPose的人体行为识别系统研究与实现[D].南京邮电大学,2022.DOI:10.27251/d.cnki.gnjdc.2021.001547.

[2]冯赛楠.基于骨骼关键点的人体动作识别方法研究[D].北方工业大学,2024.DOI:10.26926/d.cnki.gbfgu.2023.000883.

[3]曹光静.基于骨架序列的人体行为识别方法及系统实现[D].中南大学,2023.DOI:10.27661/d.cnki.gzhnu.2023.003816.