首都师范大学
CAPITAL NORMAL UNIVERSITY

人体动作识别汇报总结

汇报人：刘艾林，刘媛媛，毕曼珊，朱睿睿

目录
DIRECTORY

1 研究背景及国内外研究现状

1.1 研究背景

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。

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；

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；

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。

1.2 国内外研究现状

1.2.1 动作特征提取

特征提取算法	取得进展
Yamato 等人通过选取图像轮廓信息的方法表示人体行为特征	首先对输入的 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

存在问题：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 国内外研究现状

1.2.2 动作分类方法

选取模型	取得进展
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。	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。

2

数据集介绍

数据集介绍

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**，如图所示，其数据格式说明如下：

Tools

COCO API

Images

2014 Train images [83K/13GB]
2014 Val images [41K/6GB]
2014 Test images [41K/6GB]
2015 Test images [81K/12GB]
2017 Train images [118K/18GB]
2017 Val images [5K/1GB]
2017 Test images [41K/6GB]
2017 Unlabeled images [123K/19GB]

Annotations

2014 Train/Val annotations [241MB]
2014 Testing Image info [1MB]
2015 Testing Image info [2MB]
2017 Train/Val annotations [241MB]
2017 Stuff Train/Val annotations [1.1GB]
2017 Panoptic Train/Val annotations [821MB]
2017 Testing Image info [1MB]
2017 Unlabeled Image info [4MB]

官网地址：https://cocodataset.org/#download

百度云下载地址：

链接：https://pan.baidu.com/s/1zxvq0eeYM0_rM5lEZ4lcDQ
提取码：cfpl

数据集介绍

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：

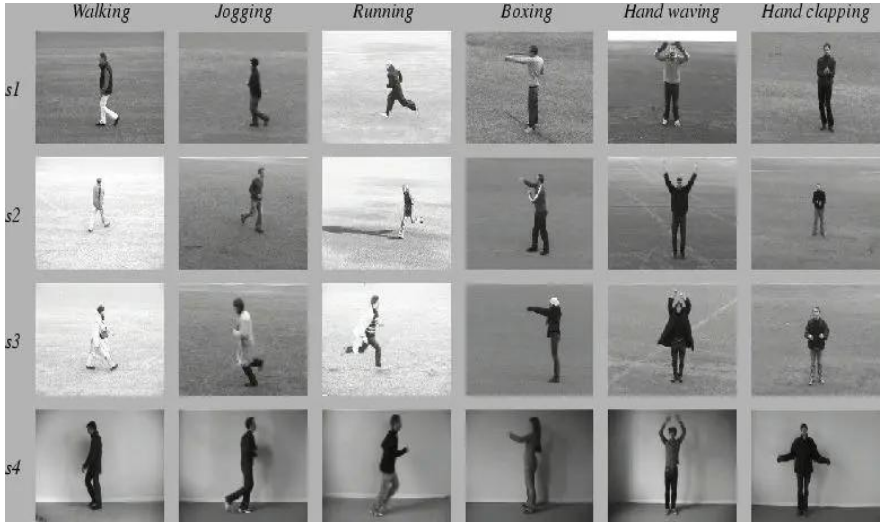
Number of frames of the 9 actions as training data.

Actions	1	2	3	4	5	6	7	8	9	Total
	wave	stand	punch	kick	squat	sit	walk	run	jump	
Number of frames	1239	1703	799	1162	964	1908	1220	1033	1174	11202

数据集介绍

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 25 fps，分 辨率为 **160×120**，平均长度为 4 秒。



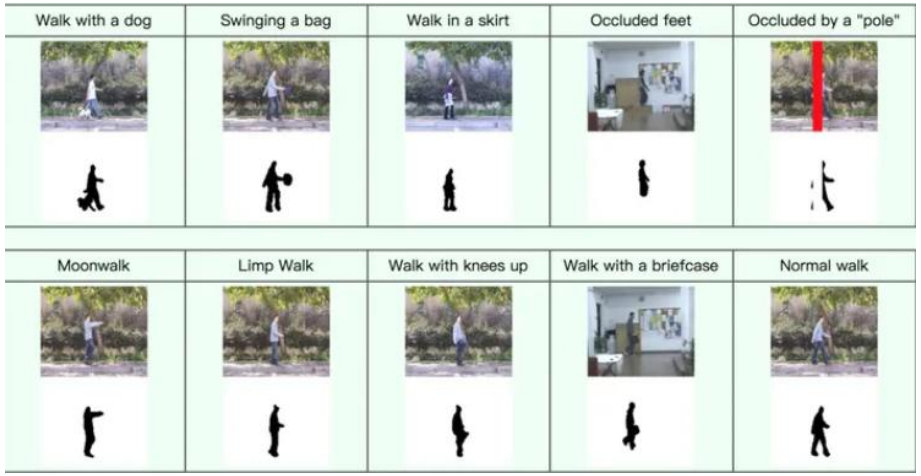
官网： https://www.nada.kth.se/cvap/actions/

数据集介绍

数据集相关获取方式:

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-short Learning来说或许是适合的数据集。



官网：

http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html

数据集介绍

数据集相关获取方式:

HMDB51动作检测数据集：

51个类别可以被分为如下5个大类：

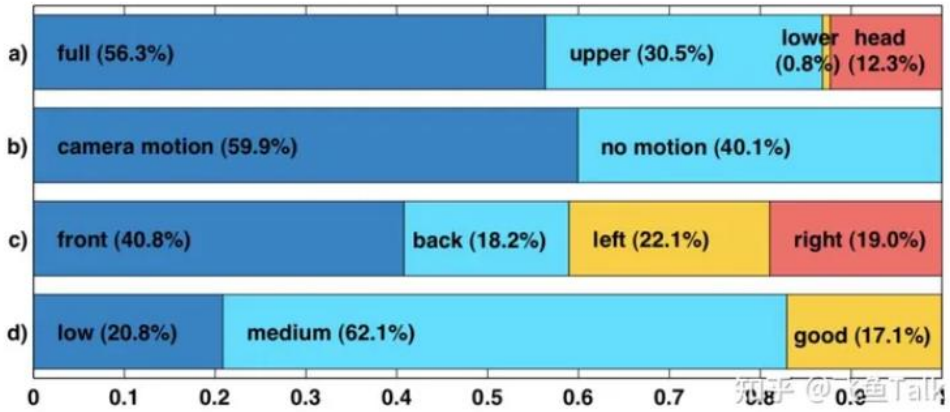
常见的面部动作 (smile，laugh，chew，talk)

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 (brush hair，catch，draw sword)

多人交互肢体动作 (hug，kiss，shake hands)



HMDB51数据集元信息分布

https://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database/#dataset

3

基于 OpenPose 的人体动作识别

基于 OpenPose 的人体动作识别

3.1 OpenPose

基于卡耐基梅隆大学于 2016 年提出的 **OpenPose 骨骼关节点提取算法**，详细说明该算法的执行过程与方法。先是**采用自底向上的方法回归出视频帧中所有的关节点**，然后创造性的提出了关节点亲和场这一概念和方法，获取关节点配对的置信度，最后再利用解决**二分图匹配**的思路完成人体骨架的拼接。

3.2 人体姿态关节点获取

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，如图 3.1 所示，其数据格式说明如下：

如图所示，其数据格式说明如下：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。

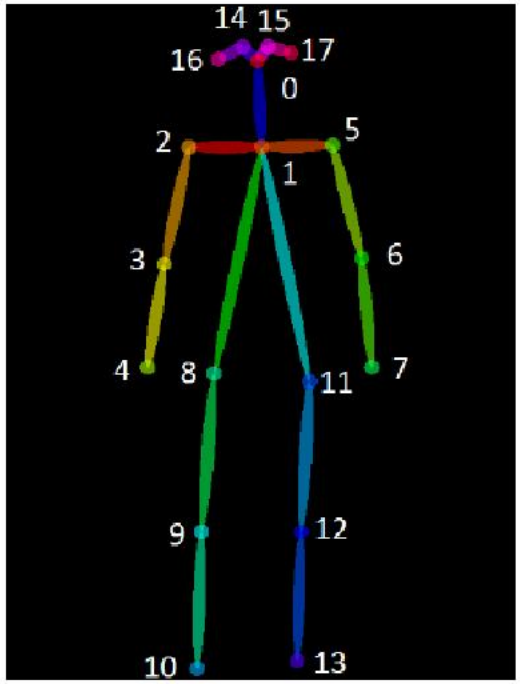
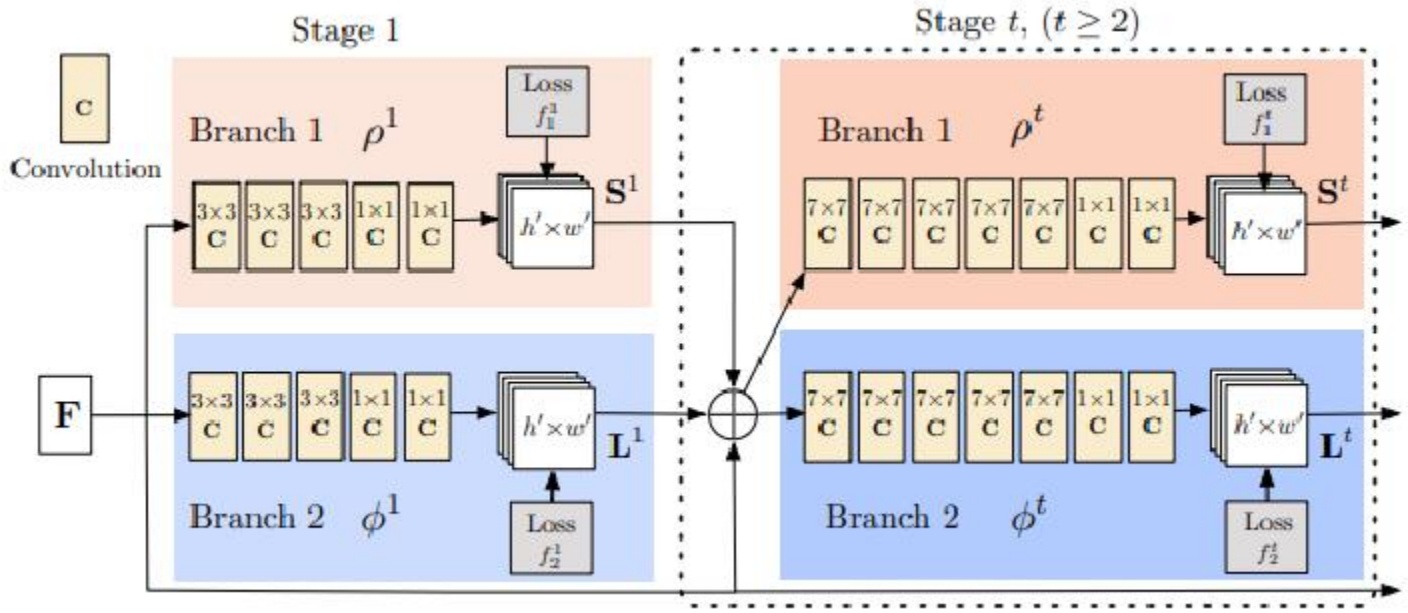


图 3.1 COCO 人体姿态骨骼模型

3.3 OpenPose网络结构



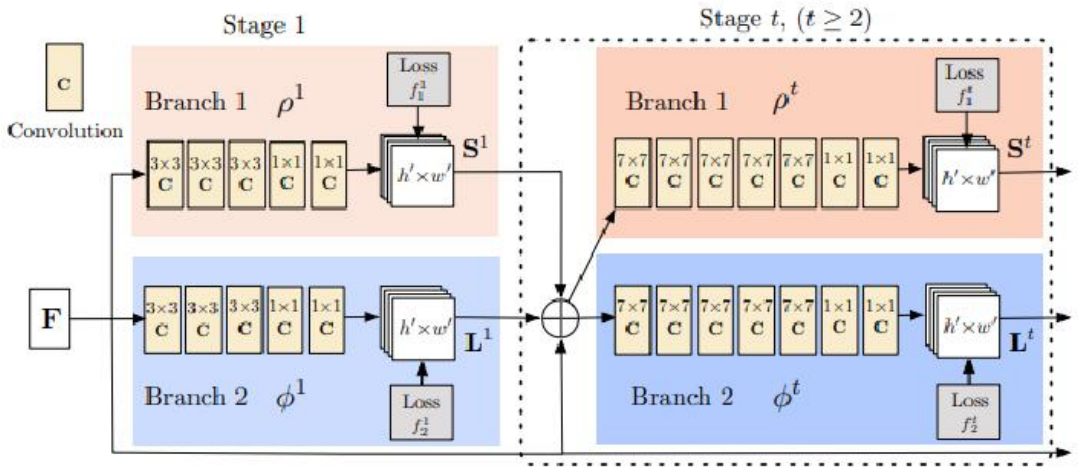
$$S^t = \rho^t\left(F, S^{t-1}, L^{t-1}\right), \forall t \geq 2 \qquad (2.1)$$

$$L^t = \varphi^t\left(F, S^{t-1}, L^{t-1}\right), \forall t \geq 2 \qquad (2.2)$$

其中，S^t 为 t 阶段输出的关节点置信图，L^t 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 2.3、2.4 所示。

3.3 OpenPose网络结构



$$S^t = \rho^t\left(F, S^{t-1}, L^{t-1}\right), \forall t \geq 2 \tag{2.1}$$

$$L^t = \varphi^t\left(F, S^{t-1}, L^{t-1}\right), \forall t \geq 2 \tag{2.2}$$

$$f_S^t = \sum_{j=1}^{J}\sum_{p} W(p) \cdot \| S_j^t(p) - S_j^*(p) \|_2^2 \tag{2.3}$$

$$f_L^t = \sum_{c=1}^{C}\sum_{p} W(p) \cdot \| L_c^t(p) - L_c^*(p) \|_2^2 \tag{2.4}$$

$$f = \sum_{t=1}^{T}\left(f_S^t + f_L^t\right) \tag{2.5}$$

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。总体的损失值公式如公式2.5所示，为各个阶段的损失值之和。

3.4 人体模型构建

3.4.1骨架数据规整化

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。在 RBG 图像中人体姿态关节点示例如图 3.2 所示。

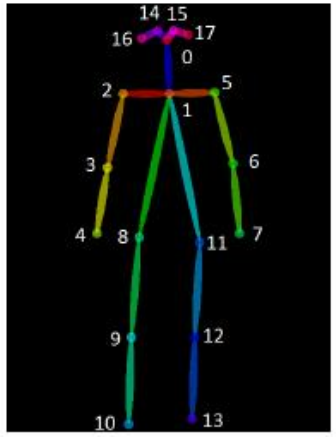


图 3.1 COCO 人体姿态骨骼模型

$$c_0 = (c_x, c_y) \tag{3.1}$$

则第 j 个关节点的平面坐标如公式 3.2 所示：

$$c_j = (c_x^j - c_x, c_y^j - c_y) \tag{3.2}$$

那么，人体骨架信息可以用 S 来表示，如公式 3.3 所示：

$$S = (C, E),\ C = \{c_0, c_1, \ldots, c_{17}\}, E = \{e_0, e_1, \ldots, e_{17}\} \tag{3.3}$$

其中， 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，在第 t 时刻，也就是第 t 帧视频中，第 j 个关节点的位置可以定义为 $c_j(t) = (x_{ij}, y_{ij})$，其中， $j \in \{0,1,\ldots 17\}$。

3.4 人体模型构建

3.4.2 全量肢体夹角特征设计

全量肢体夹角共有九组。其中包括：左手肘关节：2-3-4；右手肘关节：5-6-7；左腿膝盖关节：8-9-10；右腿膝盖关节 11-12-13；左肩膀关节：3-2-1；右肩膀关节：1-5-6；左颈部：0-1-2；右颈部：0-1-5；中间躯干：1-8-11。肢体的数学表示就是点与点连接而成的向量，肢体夹角的具体表示关系如表 3.1，骨架图如 3.3 所示：

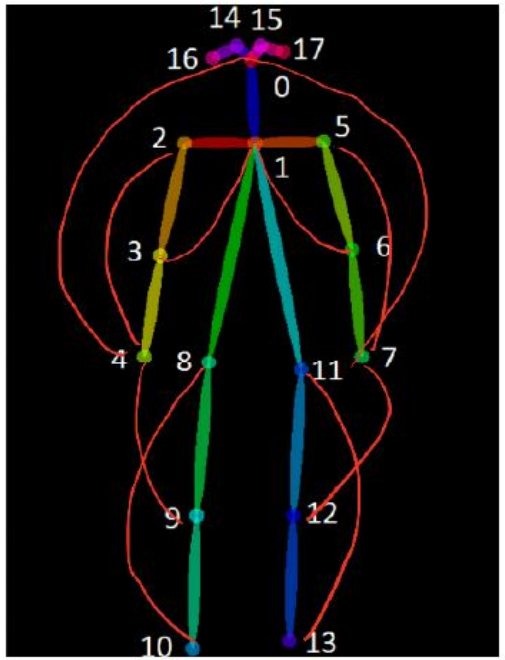


图 3.3 特征设计图

表 3.1 肢体夹角关系

肢体夹角	名称	向量
θ_1	左手肘关节	$r_{3,4}, r_{3,2}$
θ_2	右手肘关节	$r_{6,5}, r_{6,7}$
θ_3	左腿膝盖关节	$r_{9,10}, r_{9,8}$
θ_4	右腿膝盖关节	$r_{12,13}, r_{12,11}$
θ_5	左肩膀关节	$r_{2,3}, r_{2,1}$
θ_6	右肩膀关节	$r_{5,6}, r_{5,1}$
θ_7	左颈部	$r_{1,0}, r_{1,2}$
θ_8	右颈部	$r_{1,0}, r_{1,5}$
θ_9	中间躯干	$r_{1,8}, r_{1,11}$

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，$i, j \in (0,1,\ldots,17)$。对于第 n 个肢干的大小为 θ_n，设两肢向量分别为 $r_{i,j}$，$r_{i,k}$，则夹角公式如 3.4 所示：

$$\theta_n = arccos \frac{r_{i,j}^1 * r_{i,k}^1 + r_{i,j}^2 * r_{i,k}^2}{\sqrt{r_{i,j}^{1\,2} + r_{i,j}^{2\,2}} + \sqrt{r_{i,k}^{1\,2} + r_{i,k}^{2\,2}}} \tag{3.4}$$

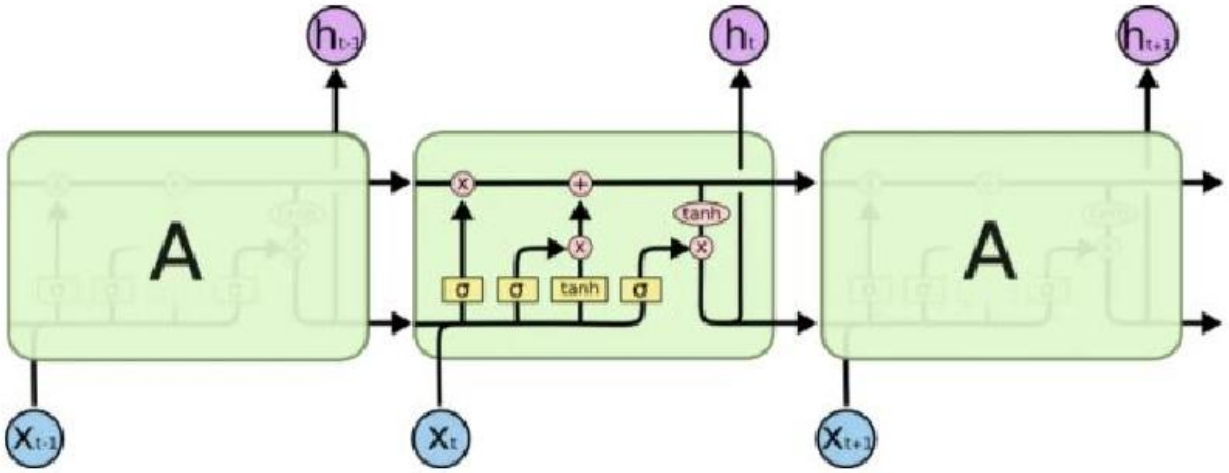
4

动作分类算法及程序演示

4. 1 LSTM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，但是当数据是诸如语音音阶、视频帧这样的**时序数据**的时候，效果往往不是很好。

循环神经网络的结构设计类似于**自动化领域内的反馈机制**，即：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。



LSTM 结构图

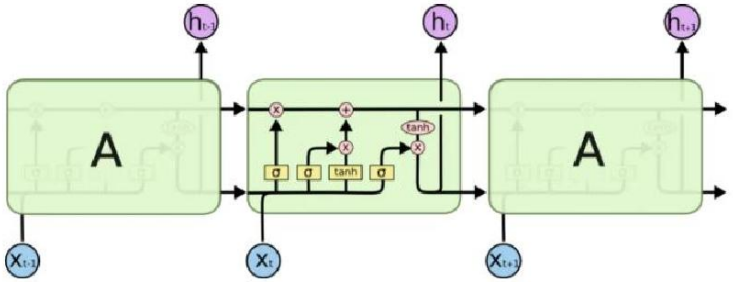
4. 2 基于 LSTM 的人体动作识别

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。接下来主要借助该网络的前向传播来介绍网络的计算过程



LSTM 结构图

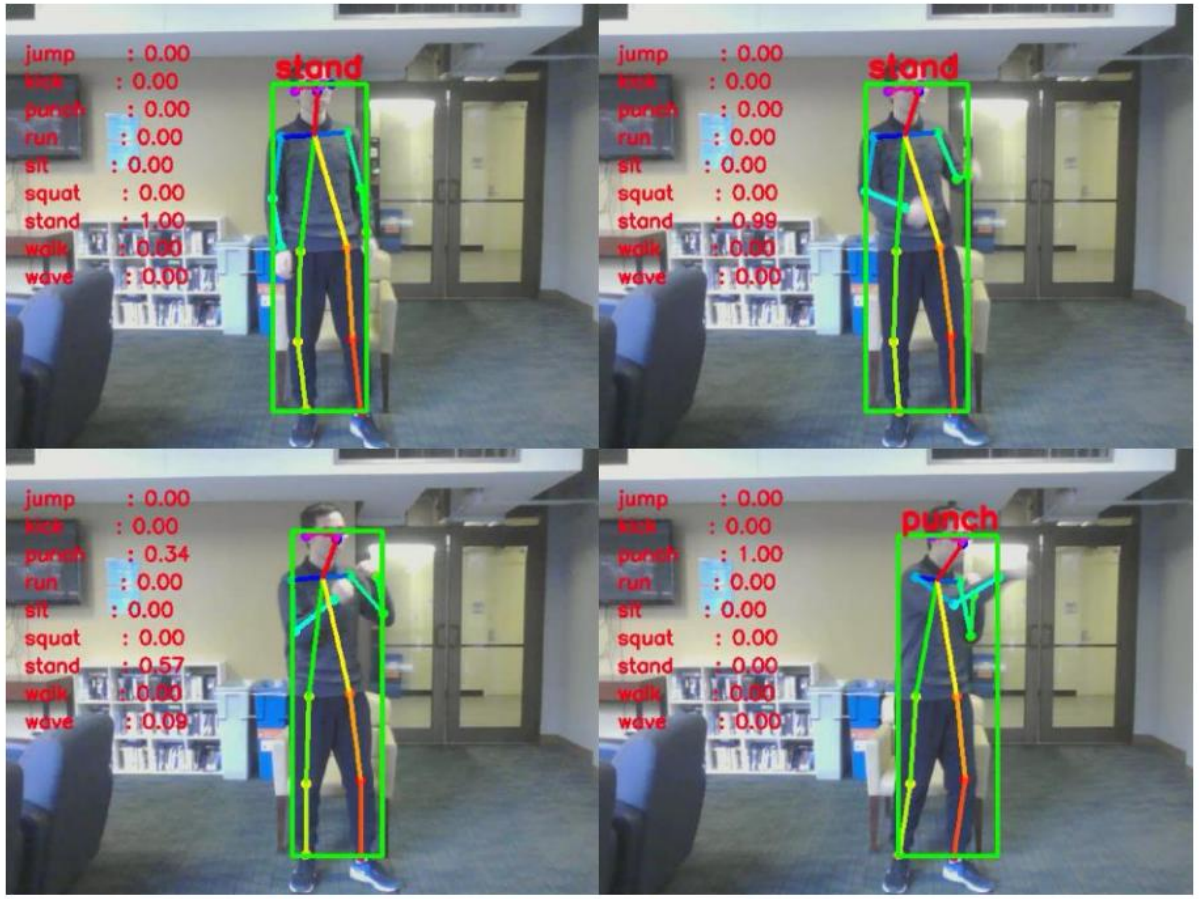
4. 3 基于 OpenPose与LSTM 的人体动作识别系统架构

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。

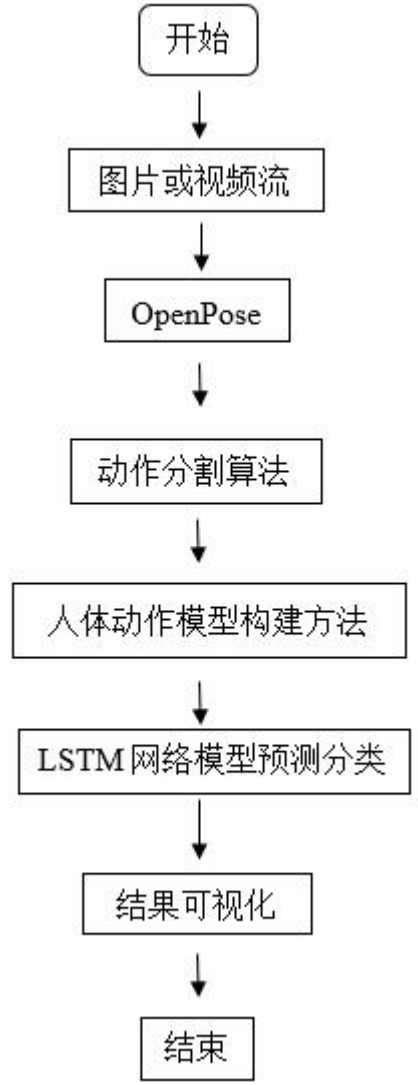
在数据获取模块中，通过 OpenPose对视频帧数据处理拿到关节点位置信息，输出关节点时空信息矩阵。动作分割模块用于对关节点时空矩阵的裁剪过滤操作。动作模型构造模块主要是对关节点信息进行处理，获得网络所需要的特征信息。

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

4. 3 基于 OpenPose与LSTM 的人体动作识别系统架构



基于 OpenPose 的行为识别系统工作流程



4.4 实验程序演示

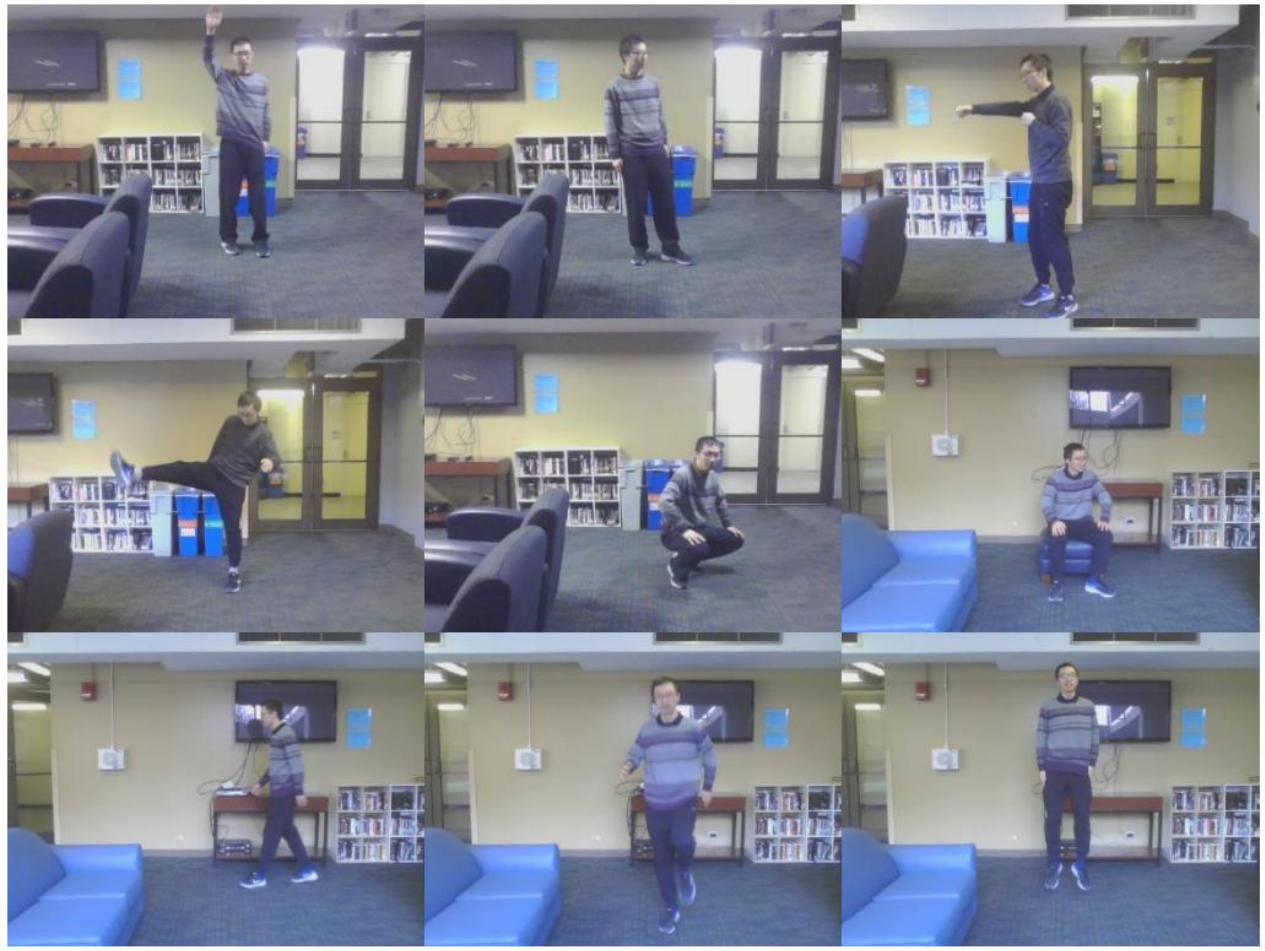
实验环境配置如下表所示，基于 Tensorflow 搭建的网络模型，编程语言 Python 3.7 实现。

表 4.1 训练与测试的环境配置

软件硬件实验平台	具体参数型号
操作系统	Windows 10 专业版
机器学习框架	Tensorflow
实验软件	Pycharm
CPU	Intel Core i7-7700HQ @ 2.80GHz 四核
GPU	NVIDIA GEFORCE GTX1060

4.4 实验程序演示

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']



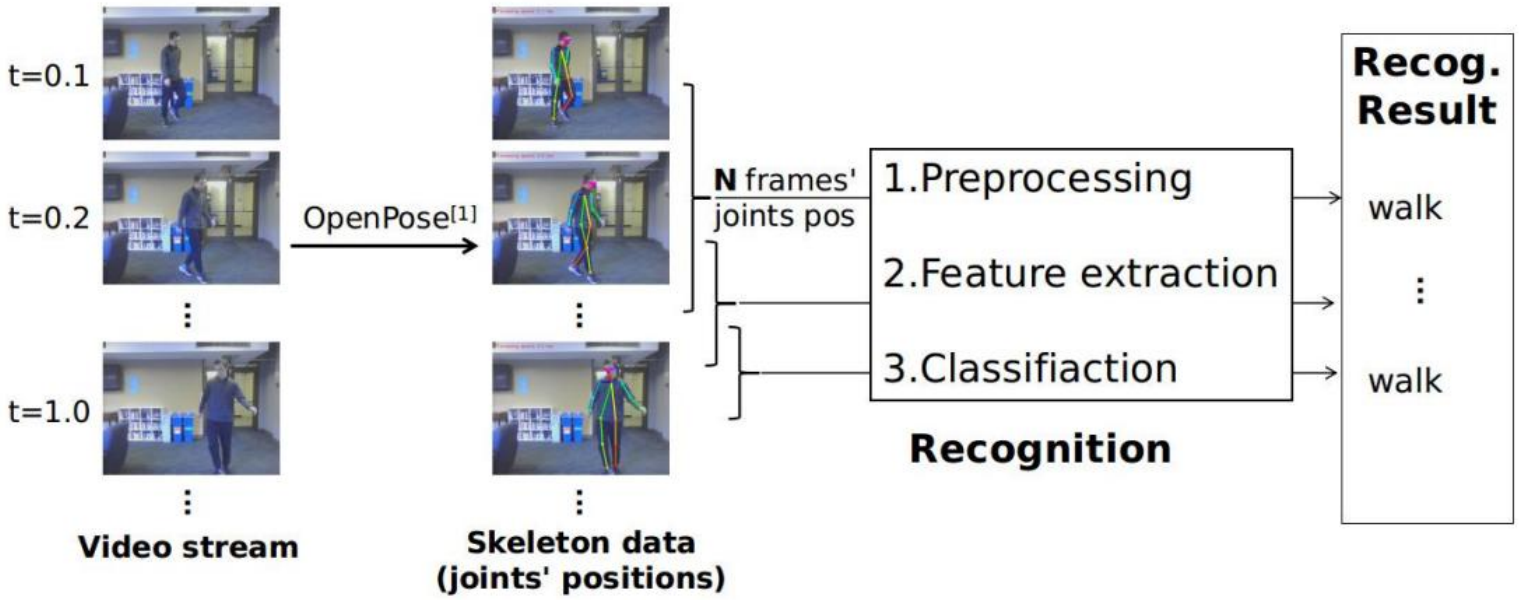
4.4 实验程序演示

数据集处理流程:

系统的输入是来自摄像机或视频文件的视频流。

然后采用**OpenPose**算法处理每个帧。接下来，大小为N的滑动窗口聚合前**N个帧的骨骼数据**。

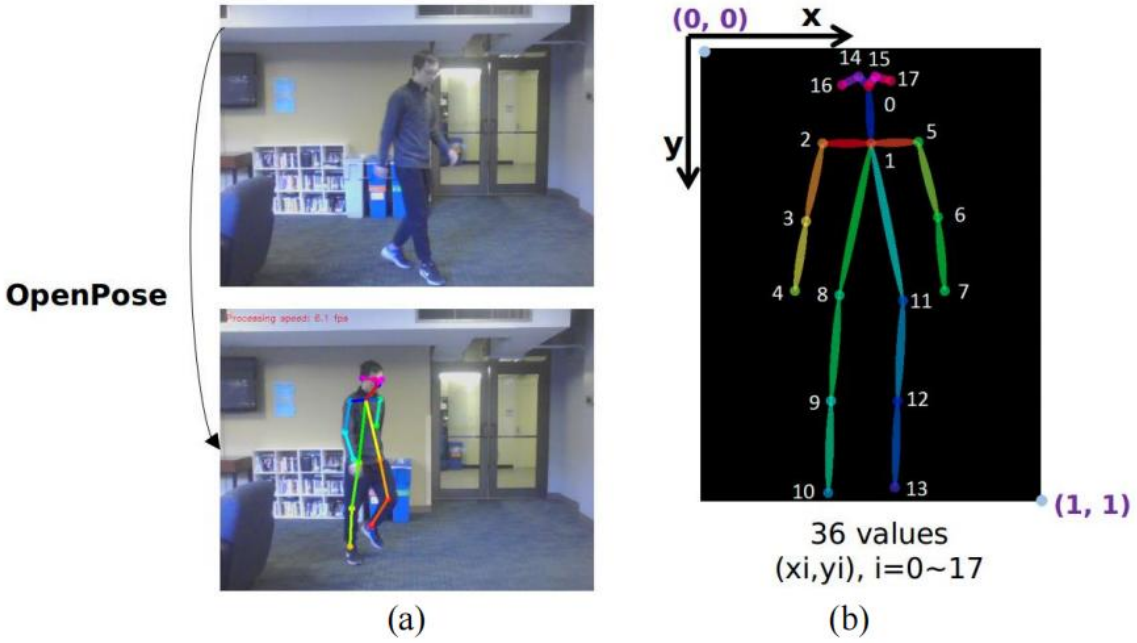
这些骨骼数据经过预处理并用于**特征提取**，然后将其**输入分类器**以获得（该窗口的）最终识别结果。类似地，为了实现实时识别框架，窗口沿着的时间维度逐帧滑动并输出每个视频帧的标签。



4.4 实验程序演示

数据集处理流程:

采用OpenPose算法从图像中检测人体骨骼。OpenPose的输入是一个图像，输出是所有人类的骨骼算法检测。每个**骨骼有18个**关节，包括**头部、颈部、手臂和腿部**，如图所示如图3.2所示。每个关节位置在图像坐标中表示，坐标值为x和y，所以每个骨架总共有36个值。



4.4 实验程序演示

数据归一化操作:

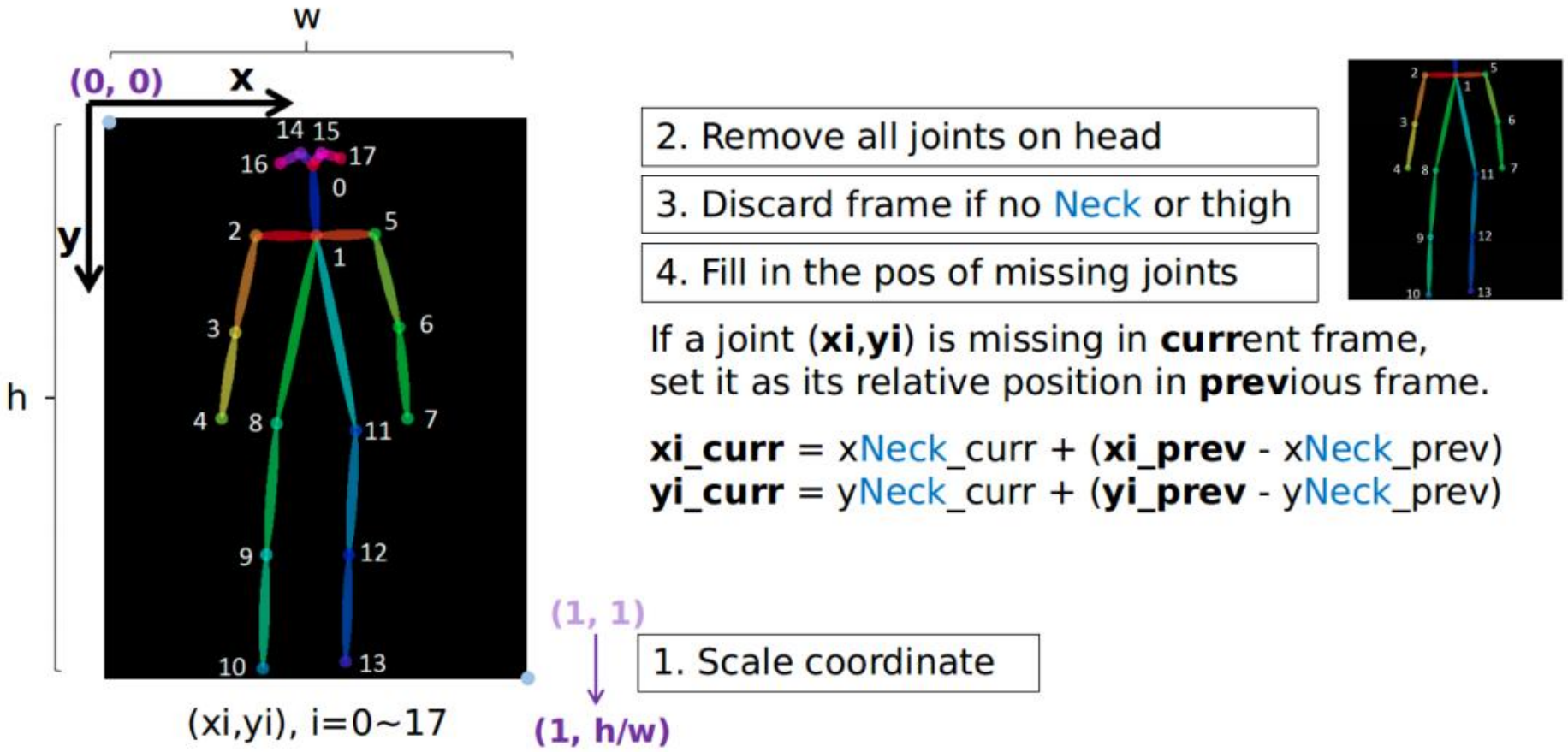
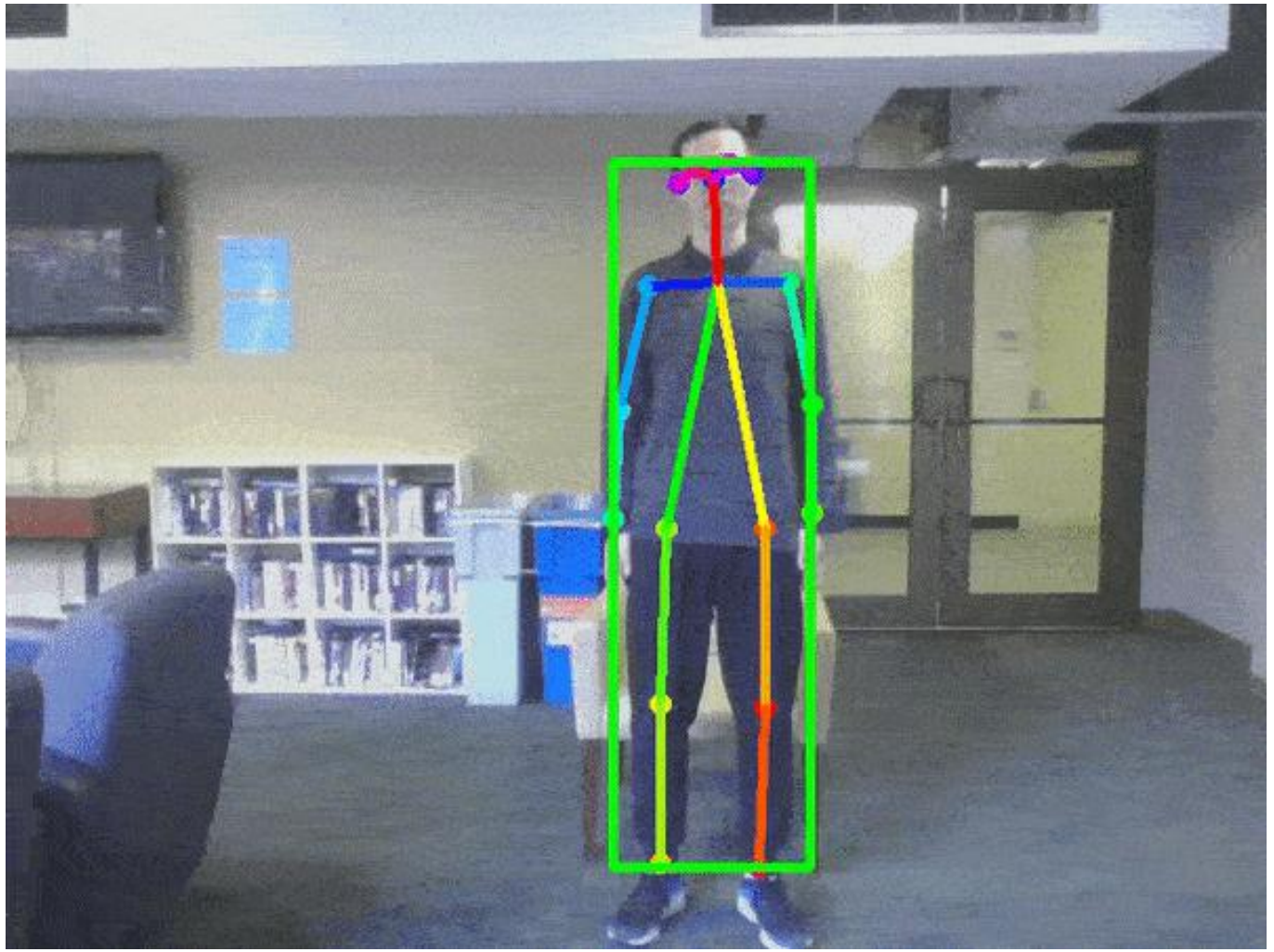
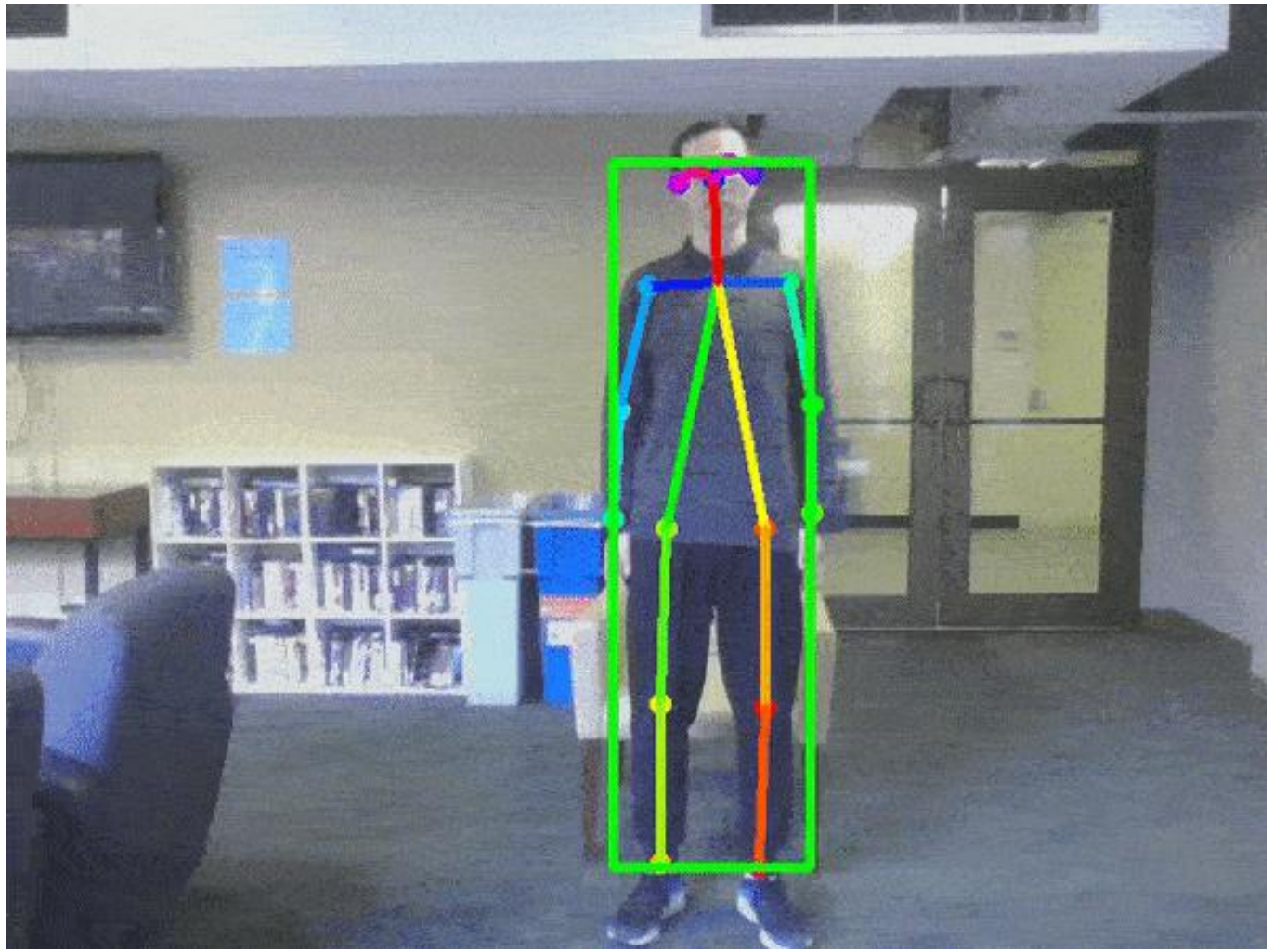


Fig. 3.3. Preprocessing joint positions in 4 steps.

4.4 实验程序演示



4.4 实验程序演示



组内分工

刘艾林：查找数据集，搜寻相关资料论文并分析算法，算法PPT演示，项目文档整合书写。
刘媛媛：查找数据集及算法资料，系统环境搭建，PPT汇报总结，项目文档整合，上传相关资料到github。
朱睿睿：查找数据集，查找算法资料并分析代码，系统PPT演示，项目文档整合。
毕曼珊：查找数据集，查找算法论文资料，数据集PPT演示，项目文档整合并进行格式排版。

谢谢观看