**人体动作识别代码说明文档**

|  |  |
| --- | --- |
| 学院： | 信息工程学院 |
| 专业： | 电子信息工程 |
| 姓名： | 刘艾林 |
| 学号： | 1211002015 |

阅读论文以及ppt说明：

一、基于 OpenPose 的人体动作识别

1.1openpose简介

OpenPose 的行为识别：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。

1.2具体方法

关节亲和场：针对多人姿态估计的时候的几个难点。首先是人数的不确定。其次是物体的遮挡或者是人与人之间的相互遮挡问题。最后一点，人数的增加会严重影响检测效果。

OpenPose为了解决以上的难点，提出了关节点亲和场（Part Affinity Fields, 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。再由关节点拼接成人。

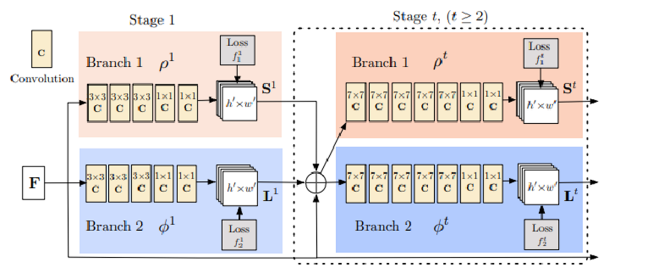
执行过程，首先将图像数据输入到网络，然后分别经过两个网络分支，一个去预测图像中的所有关节点位置信息，另一个去学习预测关节点之间的向量指向信息，输出总的PAFs。然后转化成图论中的匹配问题：边权值最大的二分图匹配。由点拼接出人的骨架。

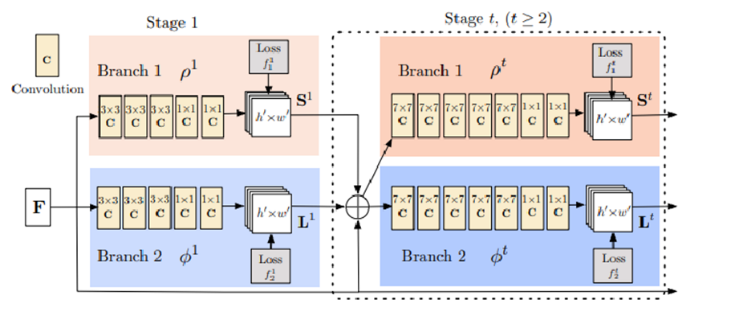
原图首先经过卷积网络，得到特征图谱，记为F，经过网络。该网络分上下两个分支，每个分支都有t个阶段（表示越来越精细），除了起始阶段，后面每个阶段的输入除了该阶段原本的输入f，还包括上一阶段两个分支的输出，具体表示如公式2.1、2.2所示。

其中，S t为t阶段输出的关节点置信图，Lt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，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

部分亲和字段，边权值最大的二分图匹配。（论文）

1.3Openpose的算法原理

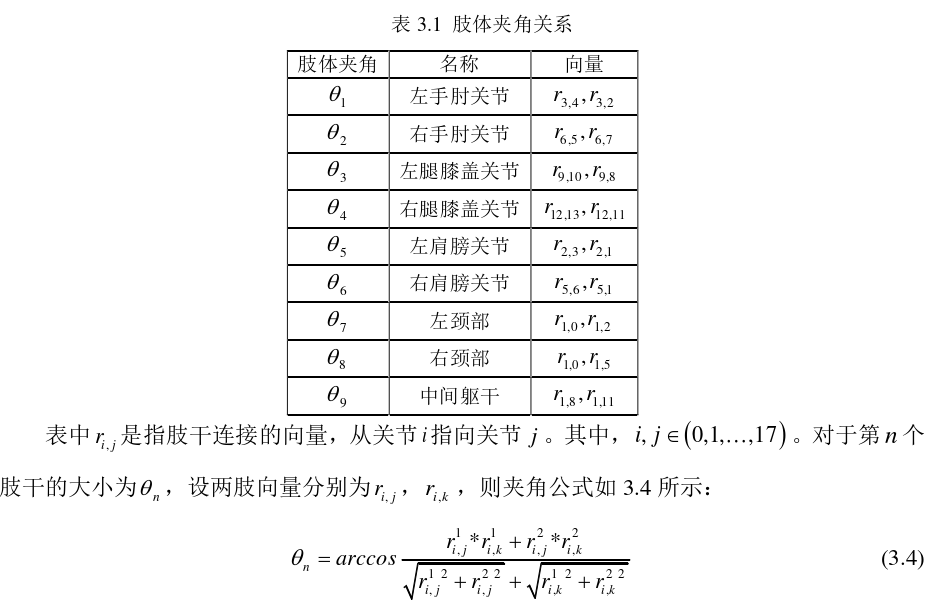
主要介绍人体动作模型构建方法，分为两个部分，骨骼关节点规整化方法及特征设计方法。

1. 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。

1. 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，r1 i,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。



二、人体动作分类算法

2.1 LSTM的选择

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。

特点：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。

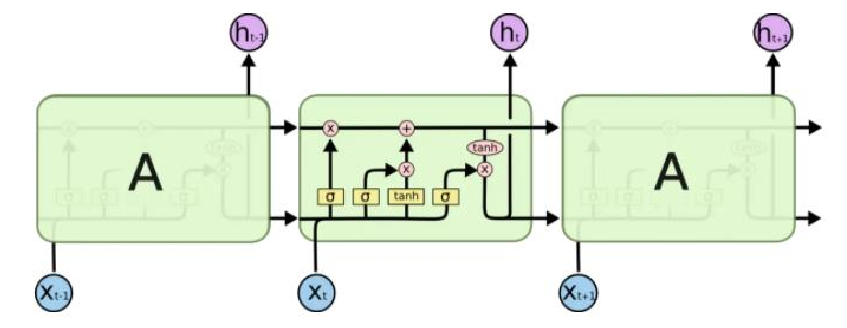
但是循环神经网络有一个很大的缺陷，就是梯度消失问题。就是说，循环神经网络只能记忆较短距离的信息，随着时间间隔的增加，网络学习就会变得非常困难。

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(Long Short-Term Memory, 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。



2.2LSTM 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。（训练过程：前向传播，具体算法不说了）

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

2.3动作分割算法

动作分割：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。

