

利用openpose轻松搞定手势识别



gangste...

饲养员

33 人赞同了该文章

更新一下：

实在抱歉，本人比较忙，没空回复许多好友申请。本以为没有人会看，不想研究 姿态估计，手势识别的小伙伴还挺多。

大家不要加我的QQ了，加群吧，各路大神可以一起交流，我就是个菜鸡，为大家提供一个交流平台。

群号563459490。



这次更新一个视屏，美丽的长腿小姐姐跳青春有你，包含**简单的手势识别**和**姿态识别**。并不很准确

身体动作：

1handup:一手高举

2handflat:两臂平伸

2handtoge:两手合十

2handup:两手高举

手部动作

手势 1,2,3 , good(伸大拇指),ok

以下是原答案

-----怎么打分割线-----

0 openpose介绍

OpenPose人体姿态识别项目是美国卡耐基梅隆大学（CMU）基于卷积神经网络和监督学习并以caffe为框架开发的开源库。可以实现人体动作、面部表情、手指运动等姿态估计。适用于单人和多人，具有极好的鲁棒性。是世界上首个基于深度学习的实时多人二维姿态估计应用，基于它的实例如雨后春笋般涌现。人体姿态估计技术在体育健身、动作采集、3D试衣、舆情监测等领域具有广阔的应用前景，人们更加熟悉的应用就是抖音尬舞机。



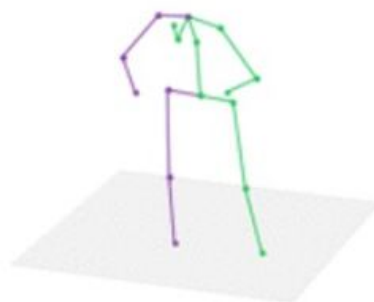
1前言

本人目前正在学习openpose的相关应用，在学习过程中笔者发现，网上大多数demo都是openpose的安装配置，或者是基于节点关系的简单计算及位置判断，例如，若手腕节点超过头部，就判定为“举起了手”。或者是将节点信息在虚拟空间进行位置关系映射，结合unity等游戏引擎操纵虚拟人物，如下图。

source video



online display video



知乎 @gangsterless

而实际上，可以除了这些简单（指思路，实现也不简单）的应用，还有很多可以挖掘的地方。笔者学习openpose的初衷是利用这些节点信息实现动作估计，最终的应用场景可以是健身，舞蹈，体

测等方面。动作估计大致可以分为两种

1 静态动作估计：对某个具体动作进行估计或者与标准动作进行匹配。不考虑动作时间上的连续性关系。直观上感觉不是特别难，无非是计算动作间相似性，可以考虑为分类问题（从预设动作中选择一个）或回归问题（与某个标准动作的相似度）。

2动态动作估计：考虑时间上的连贯性，目前比较优秀的框架是：基于动态骨骼的动作识别方法（ST-GCN）。在著名数据集Kinetics-Skeleton上取得了不错的结果。这种利用时空卷积的框架在实际应用方面有很大的局限性，模型不易部署，难以落地。学习难度较大。当然，也可以不基于神经网络，而是利用关键帧等手段进行模式匹配，这本质上是静态动作估计的延伸。

本着由易到难的原则，笔者决定首先尝试静态动作估计。

2openpose输出分析

Openpose命令介绍：

1. 读取视频，识别135个关键点，保存为图片。

```
OpenposeDemo.exe bin\OpenPoseDemo.exe --video examples\media\video.avi --write_images examples/media_out/
```

2. 手部关键点：--hand 脸部关键点：--face

3. 保存为json。

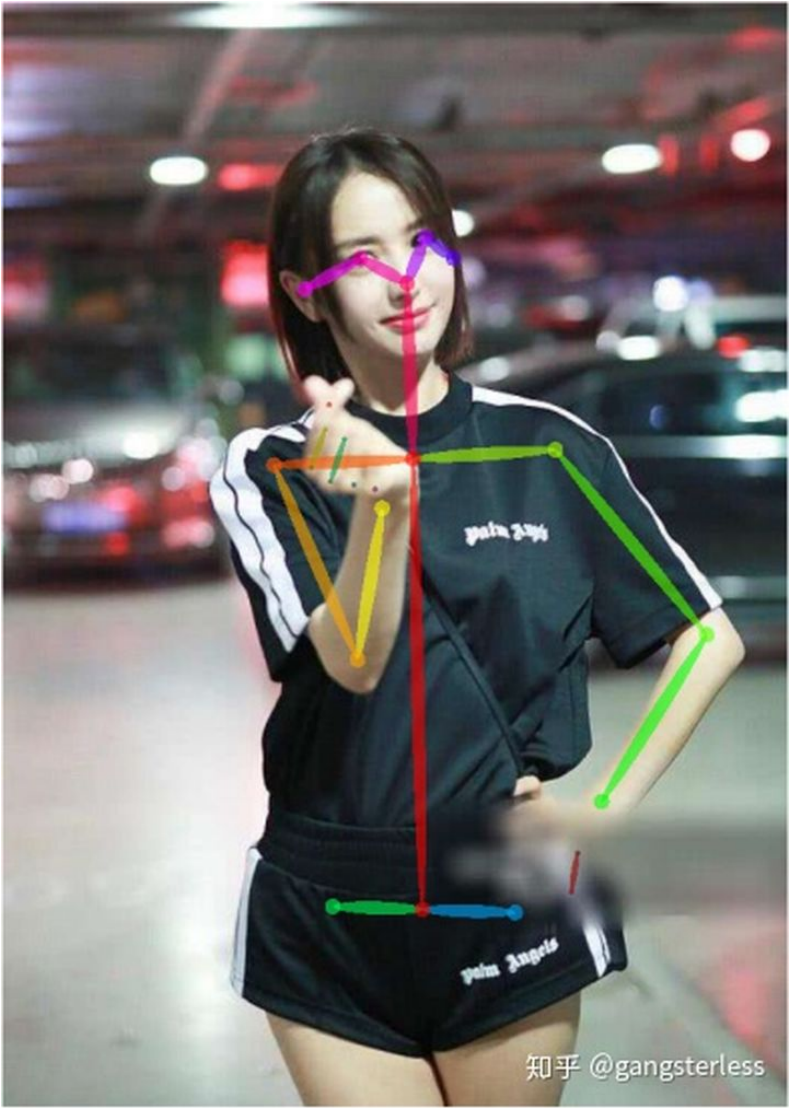
```
OpenposeDemo.exe write_json examples/media_out/
```

4.保存为视频。

```
OpenposeDemo.exe write_video examples/media_out/
```

Json格式分析：

首先分析一下，一个识别较好的动作输出的json文件，这是佟丽娅小姐姐比心图片的json：




```
376.921,348.175,0.0687145
379.086,347.093,0.0978719
366.096,322.737,0.178257
356.895,295.675,0.304902
361.225,278.355,0.524856
369.344,322.196,0.553208
357.436,297.298,0.563665
348.235,285.391,0.753163
339.575,275.107,0.834267
359.06,328.149,0.478438
348.235,305.958,0.518616
342.282,309.747,0.334648
344.988,314.618,0.285777
346.07,335.727,0.444789
339.034,314.077,0.526782
332.539,317.866,0.397358
330.374,322.737,0.244629
333.622,337.351,0.448096
331.998,323.278,0.419656
327.668,321.113,0.526877
```

知乎@gangsterless

由于我们只关注手部动作所以 我们只需要特别关注json(读到python里面就是字典)的这两个键

hand_left_keypoints_2d

hand_right_keypoints_2d

每个键对应21个手部关键节点，节点有三个值(x,y坐标，置信度)，如上图。

3 数据准备

首先，笔者没有发现利用openpose完美标注的免费数据集。遂决定上网找视频、图片自行标注，又经过一番苦苦搜寻，笔者并没有找到令人满意的静态姿态图片。一方面是openpose对人物离镜头的距离，角度有一些要求，很多图片并不能准确识别，另一方面某一类图片实在是不好找。例如我找了比心，经过一个多小时的寻找，只找到了5张左右能用的，也爬了一些视频，想从中挑取，终究不尽如人意。

那没办法了，本人只好为科学献身，尴尬出境。

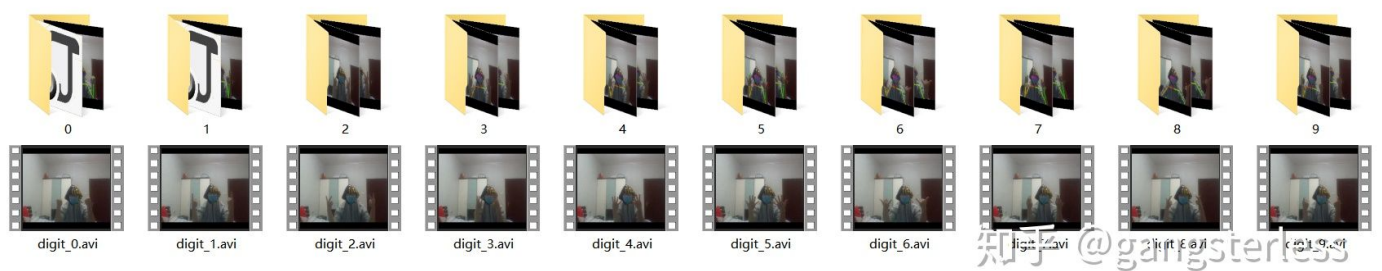
本文的主要工作就是基于openpose输出的节点进行静态动作的识别（手势识别）。我们把这个问题简化为一个分类问题。

要识别的的对象：我的手

具体动作：手势0-9+比心，共11个动作。下图是8.



我剪了11段视频大概每种动作8s左右，需要注意的是，不要离镜头太近，尽量要把关节漏出来，这样识别效果好。



录完视频后，利用openpose命令获取节点信息，用python进行过滤分析，我把平均置信度小于0.3的图片都过滤掉了，因为置信度太小的节点没有意义。

经过过滤，最终留下了大约1400张图片。样本量不是很够，由于左右手其实区别不大，我决定进行一下数据增强。具体的就是将原图片做镜像处理。我并没有直接对图片进行操作，而是对某个节点来说 保持y坐标不变，x坐标变化为 图像宽度-x，这样样本量就增加了一倍。大约有2800个样本。

为了方便读取，我将最终的样本整合为一个csv文件

该文件有65列（21（关键点）*3（x,y,置信度）+左手还是右手+动作标志）

4模型训练

采用sklearn，尝试了线性模型，knn，GBDT。2,8划分测试集，训练集，最终GBDT的准确率在测试集上达到了98.5%，可以说相当可观。要注意数据位置信息要进行归一化处理。

```
warnings.warn(msg, category=DeprecationWarning)
precision: 98.50%, recall: 98.50%
accuracy: 98.50%
real res:
[4, 6, 4, 4, 5, 2, 6, 0, 1, 1, 2, 8, 1, 1, 9, 6, 9, 5, 4,
predict res:
[4 6 4 4 5 2 6 0 1 1 2 8 1 1 9 6 9 5 4 5 4 9 8 4 9 0 0 3
 9 3 5 5 8 0 8 5 4 9 6 1 9 6 8 1 6 5 1 6 知乎@gangsterless
 2 2 7 1 8 8 8 7 5 1 5 1 6 4 5 5 8 5 4 1 5 6 6 8 9 9 6 9
```

5总结及展望

由于模型参数较少，达到近似实时（插帧检测）并不很困难，准确率如此之高跟样本相似性大也有一定关系，由于只有我自己的手，模型的鲁棒性也无法检测。但至少说明这个思路是可行的。

估计也没人看这文章，如果需要代码、节点数据，或者大神不吝赐教欢迎加我

qq 935309780，验证信息请写openpose交流。

数据集和一些清洗数据的代码在这里

代码十分混乱！！！！没时间好好维护，各位见谅！

Baidu AI Studio - 一站式AI开发平台

aistudio.baidu.com



编辑于 2020-05-04