



三言两语总结2D关键点检测



Roger

一个想冒险的懒人

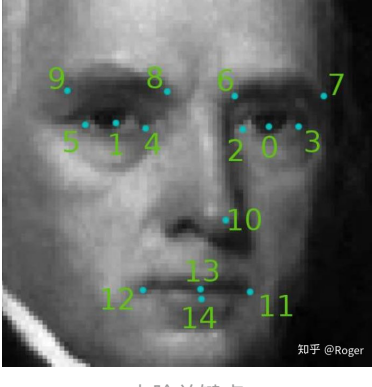
83 人赞同了该文章

这是一篇很简短的总结，主题是2D关键点检测。为什么强调2D，因为还有3D关键点检测，这个笔者小黑暂时没有实操过。

一、关键点有哪些？

关键点本质上就是在图像中用一个点表示物体上特定的部位。（这里吐槽一下，部位当然是有面积的，辣么标注的时候点在部位的哪个地方呢？？这里就包含了标注的一定不确定性...）

常见的关键点有人脸关键点、人体骨骼关键点、车辆关键点等。当然，从关键点本质出发，讲道理你想点哪里就可以定义哪里是一个新的关键点类型。



人脸关键点



人体骨骼关键点



车辆关键点

二、为什么要检测关键点？

人脸关键点涉及人脸识别的相关应用，例如解锁你的iphone手机。人体骨骼关键点可以用于分析人体的行为动作，还可以用于AR等应用等。车辆关键点则一般和车辆模型建模、自动驾驶等相关。（读者如果有补充，欢迎在评论区留言~~）

三、关键点检测怎么做？

(1) 回归派 容易过拟合

回归派主要是发源自人脸关键点检测，用CNN提取特征，然后使用全连接层直接数值回归关键点的坐标。

经典网络：DeepPose（Human Pose Estimation via Deep Neural Networks）、MTCNN(Multi-task Cascaded Convolutional Networks)

(2) heatmap派 擅长单个关键点定位，遮挡、较大转动时偏差大

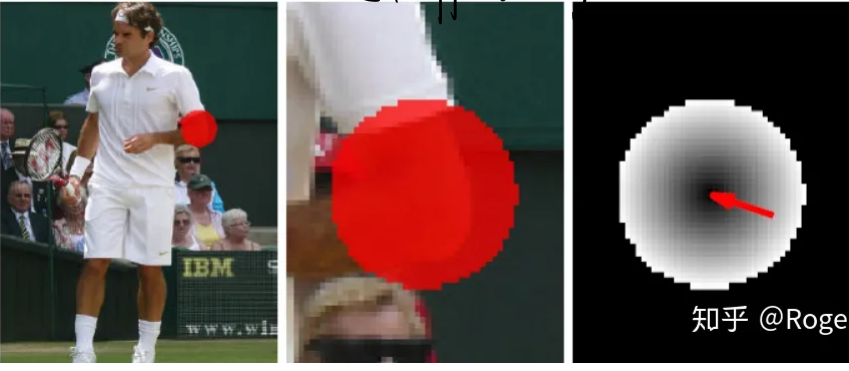
后来在做人体关键点检测时发现，照搬人脸关键点的暴力回归做法效果不太好，原因笔者认为是由于人体的姿态动作变化较多，网络学习难度增大。而人脸的点基本在一个平面上，相对稳定，内在规律比较simple，网络学起来也相对容易。

既然回归走不通，那么前辈们就想到了使用热力图预测的方法。没错，和你想的一样，这里的热力图即标签是一张图，图中每一个通道代表一个类别的关键点，有几个类别的关键点，则有几个通道。一个通道图上关键点位置是一个以其为中心的二维高斯分布，其余位置像素值为0。网络预测的结果也是一张热力图map，提取其中的坐标一般最直接的做法就是提取一个通道中像素响应大于一定阈值的且响应最大的点，该点的坐标就是该类别关键点的坐标。

经典网络：hourglass、openpose、simplebaseline、HRNet

其实在heatmap派里，上面直接使用二维高斯分布的形式作为点标签的，是最常见的做法。也有几个不同的形式，这些形式归纳起来就是“加偏置通道”。例如在谷歌的paper《Towards accurate multi-person pose estimation in the wild》中，每个类别的关键点不仅对应一个通道，而是对应多个通道，一个通道相当于分类的作用，表明一个圆形范围内的点属于这个关键点；两个通道则表示这个属于关键点范围的点，距离真实的关键点位置的坐标偏置数值。

3个通道：1个分类(常规heatmap)
2个 offset [x,y]与真实关键点



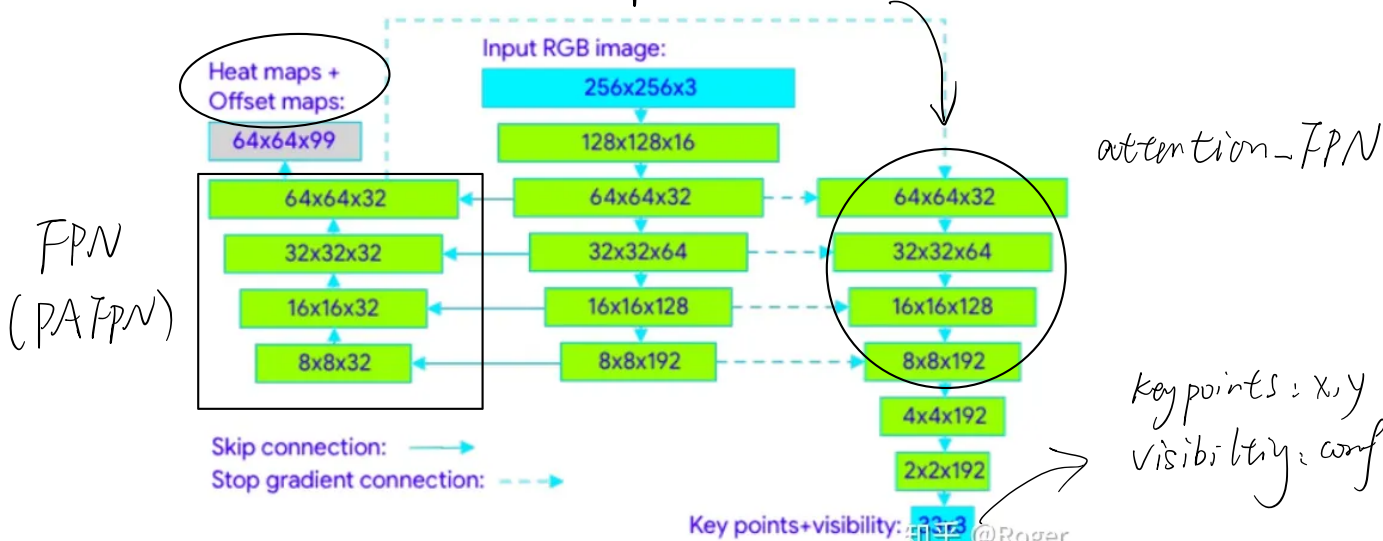
加偏置通道

经典网络：Towards accurate multi-person pose estimation in the wild、Learning to Refine Human Pose Estimation

(3) 混合派

最近还看到一篇paper《BlazePose: On-device Real-time Body Pose tracking》，谷歌出品。同时训练关键点heatmap和坐标回归，融合关键点的heatmap作为embedding信息，有助于提高直接数值回归的效果（感觉有点像先生成一个attention图，然后结合attention信息来做回归）。

可以 heatmap guide attention 相乘



四、关键点变着花样怎么玩？

1、结合结构约束

无论人体还是车辆，关键点之间实际上含有物体的结构信息，因此可以考虑加入结构约束到网络学习中。现在有的加入了条件随机场，有的加入了LSTM，有的结合了图卷积。

推荐阅读：

《Joint Training of CNN and a Graphical Model for Human Pose Estimation》

《Occlusion-Net: 2D/3D Occluded Keypoint Localization Using Graph Networks》

2、向目标检测靠拢

目前很多anchor free的目标检测模型都有关键点检测的影子。

推荐阅读：

《CenterNet—Objects as Points》

《PifPaf: Composite Fields for Human Pose Estimation》

3、蒸馏，提速

关键点检测模型一般比较大，如果想缩小模型提高推理速度，同时保证仍然有大模型的精度，有的学者就向蒸馏出发，让一个小模型去学习一个大模型。

推荐阅读：

《Fast Huamn Pose Estimation》

4、refine错误

不能保证一次推理没有错误的检测点，那就修正。

推荐阅读：

《Learning to Refine Human Pose Estimation》

5、提高定位精度

由于基于热力图的检测方法，在最后的输出feature map大部分网络都会缩小了一定比例，加上预测的偏差，那么最终得到的预测相应图一定不是很规范的二维高斯分布点，因此，有的学者在编码方式上入手，尝试提高定位的精度。（这部分让我联想起了星图定位.....质心法、加权高斯法、二维高斯拟合法.....）

推荐阅读：

《Distribution-Aware Coordinate Representation for Human Pose Estimation》

《The Devil is in the Details: Delving into Unbiased Data Processing for Human Pose Estimation》

6、扩展到单目3Dbbox检测可还行

结合相机参数、投影约束，回归3Dbbox的八个角点在图像中的位置。

推荐阅读：

《RTM3D: Real-time Monocular 3D Detection from Object Keypoints for Autonomous Driving》

