

Pose Estimation

Convolutional Pose Machines

将深度学习应用于人体姿态分析，同时用卷积图表达纹理信息和空间信息。主要网络结构分为多个stage，其中第一个stage会产生初步的关键点的检测效果，接下来的几个stage均以前一个stage的预测输出和从原图提取的特征作为输入，进一步提高关键点的检测效果。

用各部件响应图来表达各部件之间的空间约束，响应图和特征图一起作为数据在网络中传递。人体关键点在空间上的先验分布会指导网络的学习，假如stage 1的预测结果中右肩关键点的预测结果是正确的，而右肘关键点的预测是错误的，那么在接下来的stage中肩和肘在空间上的先验分布会指导网络的学习。

Wei S E, Ramakrishna V, Kanade T, et al. Convolutional Pose Machines[J]. 2016:4724-4732.

Question

1. 什么是响应图？是heatmap吗？
2. 什么是关键点的先验分布？是ground truth吗？

Cascaded Pyramid Network

本论文主要关注的是不同类别关键点的检测难度是不一样的，整个结构的思路是先检测比较简单的关键点、然后检测较难的关键点、最后检测更难的或不可见的关键点。

具体实现时，分为两个stage，GlobalNet和RefineNet。其中GlobalNet主要负责检测容易检测和较难检测的关键点，对于较难关键点的检测，主要体现在网络的较深层，通过进一步更高层的语义信息来解决较难检测的关键点问题；RefineNet主要解决更难或者不可见关键点的检测，这里对关键点进行难易程度进行界定主要体现在关键点的训练损失上，使用了常见的Hard Negative Mining策略，在训练时取损失较大的top-K个关键点计算损失，然后进行梯度更新，不考虑损失较小的关键点。

Chen Y, Wang Z, Peng Y, et al. Cascaded Pyramid Network for Multi-Person Pose Estimation[J]. 2017

Question

1. 怎么定义哪些是容易检测，哪些是较难检测的点？在训练集中如何做到区分的？
2. 什么是Hard Negative Mining策略？

RMPE

本论文主要考虑的是自上而下的关键点检测算法在目标检测产生Proposals的过程中，可能会出现检测框定位误差、对同一个物体重复检测等问题。检测框定位误差，会出现裁剪出来的区域没有包含整个人活着目标人体在框

内的比例较小，造成接下来的单人人体骨骼关键点检测错误；对同一个物体重复检测，虽然目标人体是一样的，但是由于裁剪区域的差异可能会造成对同一个人会生成不同的关键点定位结果。本文提出了一种方法来解决目标检测产生的Proposals所存在的问题，即通过空间变换网络将同一个人体的产生的不同裁剪区域（Proposals）都变换到一个较好的结果，如人体在裁剪区域的正中央，这样就不会产生对于一个人体的产生的不同Proposals有不同关键点检测效果。

Fang H S, Xie S, Tai Y W, et al. RMPE: Regional Multi-person Pose Estimation[J]. 2016:2353-2362.

Question

如何进行空间变换，什么叫较好的结果？

Part Segmentation

即对人体进行不同部位分割，而关键点都落在分割区域的特定位置，通过部位分割对关键点之间的关系进行建模，既可以显式的提供人体关键点的空间先验知识，指导网络的学习，同时到最后对不同人体关键点进行聚类时也能起到相应的连接关键点的作用。

Xia F, Wang P, Chen X, et al. Joint Multi-person Pose Estimation and Semantic Part Segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017:6080-6089.

Question

1. 什么叫分割区域的特定位置？

Part Affinity Fields

该方法通过对人体的不同肢体结构进行建模，使用向量场来模拟不同肢体结构，解决了单纯使用中间点是否在肢干上造成的错连问题。

PAFs则不仅使用中间点来建模肢干，而且在中间的位置之外还给每个中间点加上了方向的信息，这样就能解决出现的错连问题。

Cao Z, Simon T, Wei S E, et al. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields[J]. 2016:1302-1310.

Question

方向点是如何获取的？中间点是否就是图片上距离的中间点？

Associative Embedding

该方法通过使用高维空间的向量来编码不同人体的不同关键点之间的关系，即同一个人的不同关键点在空间上是尽可能接近的，不同人的不同关键点在空间上是尽可能远离的，最后可以通过两个关键点在高维空间上的距离来判断两个关键点是否属于同一个人，从而达到聚类的目的。

Newell A, Huang Z, Deng J. Associative Embedding: End-to-End Learning for Joint Detection and Grouping[J]. 2016.

Question

应该是特殊训练集，训练集中就需要给出不同人的关键点的向量编码

Mid-Range Offsets

该方法通过直接回归一个关键点到另一个关键点之间的offset来建模两个关键点之间的关系，这种较长距离的offsets是较难学习的，在回归具体数值时会有较大误差，这里可以用关键点周围的Short-Range Offsets去进一步Refine对应的offsets。

Papandreou G, Zhu T, Chen L C, et al. PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model[J]. 2018.