

三 维 动 画 和 交 互 设 计 课 程 读 书 报 告



作者姓名 苏迪宇

作者学号 21851402

年级专业 2018级软件工程

所在学院 软件学院

完成日期 2018.12.25

Research on Computer Animation and Interaction Design

A Report Submitted to

Zhejiang University

Major Subject: Software Engineering

Advisor: Li Qilei

By

Su Diyu

College of Software Technology, Zhejiang University

2018

# 论文题目

论文题目：Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields

作者：Zhe Cao Tomas Simon Shih-En Wei Yaser Sheikh

发表于：CVPR，2017

# 论文概述

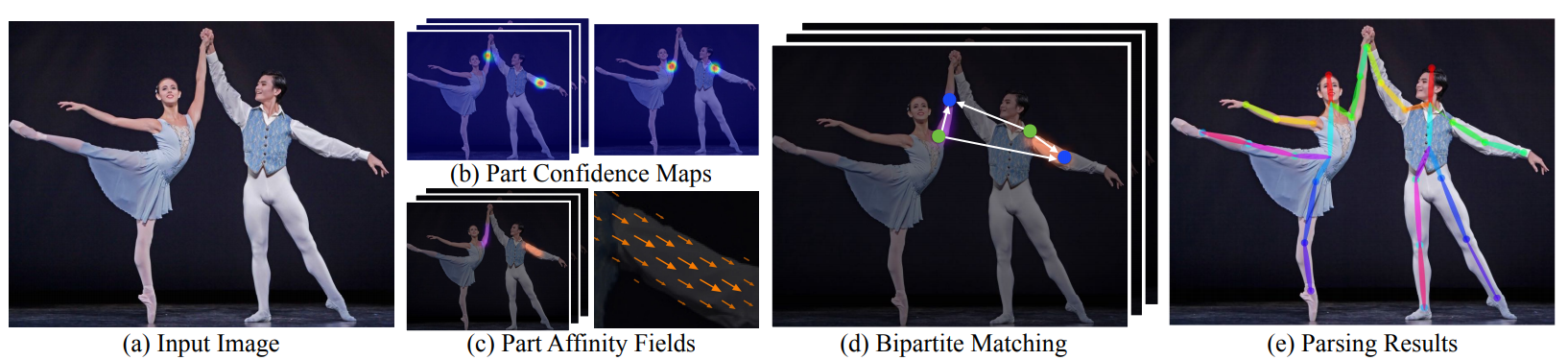
 论文提出了一个可以高效率地检测多人姿态的方法——Part Affinity Fields（PAFs）。它的特点是可以学习到人体两个关节点之间的方向信息，从而将不同人的关节区分开来。因此，这篇论文所提出的方法在多人关节点检测的任务中表现非常好。因为cnn广阔的感知域，模型其实encode了相当部分的全局信息，因此允许文章中使用的贪心的从下至上的解析方式来对检测到的关节点做解析（以确定哪些关节点是同一个人的），正因为这种贪心的解析方式，使论文中的方法效率非常高，可以做到实时检测。这篇论文的效果非常好，在COCO 2016关节点挑战赛上取得了第一名。

值得一提的是，著名开源库openpose就是以这篇论文为基石所开发的。

# 模型架构与实现细节

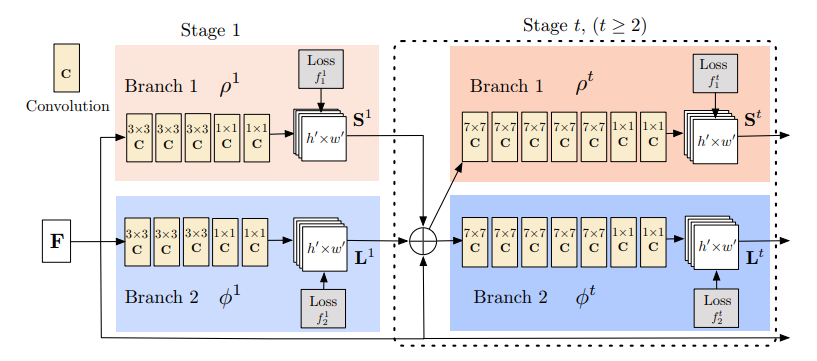
模型分为两部分：第一部分是通过深度神经网络（CNN）抽取图片特征，获得1.关节点坐标 2.关节点之间的方向信息；第二部分是利用CNN提取出来的信息，找到属于同一个人的关节。

模型整体pipeline如下图：



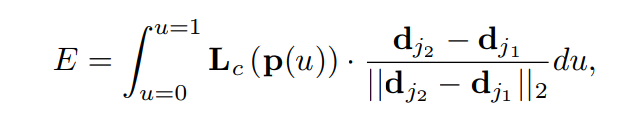
第一部分：

第一部分是利用神经网络提取图像的特征。神经网络分成两个分支：上半个分支预测关节点的位置，下半个分支预测肢体的方向。并且第t帧提取到的特征，会作为神经网络输入的一部分，用于下一帧的预测。



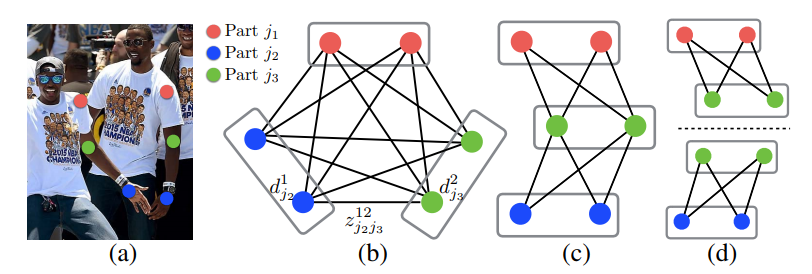
第二部分：

截止第一部分，模型以及可以很好地对单人的图片做出姿态估计了，但是当图片中有许多人时，这个模型会出现问题：如何把这些关节点结合起来，形成多人姿态估计呢？这篇论文提出了一个非常漂亮的做法。利用神经网络提取到的方向特征，我们可以求得两个关节点之间的相关性强弱。具体做法如下：

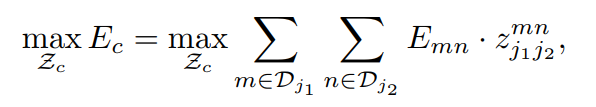


其中，E表示两个关节相关性大小，u为两个关节位置的内插值，L为在某位置的方向向量（第一部分的神经网络预测结果），d是关节点的位置。

但是，当图片中人很多时，如果仅仅是简单的对所有关节点两两计算相关性，这个算法的性能将会非常差。



先考虑二分图（bipartite graph，即图d）的情况：在二分图的限定下，寻找最优方案并不难，具体来说，就是优化下面的表达式：



找到两组使总E值最大的关节，即找到相关性最大的两组关节。可以用Hungarian算法来解决这个问题。

对于多人姿态这个问题，论文作者进行了贪心的优化：首先，并不是将所有节点两两进行比较（如图b），作者选择了最少的边（如图c）；接着，把图c进行切割，形成图d，这样就把问题转化成为我们可以解决的二分图问题了。

作者对贪心算法的合理性进行了解释：因为CNN有很大的感知域，所以不相邻关节之间的关系，已经被建模了，只需关注相邻的节点，就可以达到很好的准确度和速度。

# 参考文献

[1]. Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields", CVPR, 2017.