

## 计算机视觉在生物力学和运动康复中的应用和研究

近几十年来，在运动生物力学和康复环境中对人体运动的研究取得了长足的进步。基于视觉的运动分析涉及从顺序图像中提取信息以描述运动，可以追溯到 19 世纪后期，Eadward Muybridge 首先开发了捕获屈步态图像序列的技术。此后，运动分析相关技术进步很快，与不断增长的需求相平行，这些技术可以捕获从临床步态评估到视频游戏动画在内的各种运动。在运动生物力学和康复应用中，人体运动学的定量分析是一种功能强大的工具，生物力学工具已经从使用图像的人工注释发展为基于标记的光学跟踪器，基于惯性传感器的系统以及使用复杂的人体模型，计算机视觉和机器学习算法的无标记系统，已经取得了长足的发展。

目前在运动生物力学和康复中最广泛使用的运动分析系统不允许在没有附加标记，控制条件和/或大量处理时间的情况下自动收集运动数据。这些限制可能会妨碍在正常的训练或康复环境中常规使用运动捕捉，因此非常需要开发自动无标记系统，并且结合应用计算机视觉和机器学习中的相关技术。

无标记运动捕捉系统的四个主要组成部分是：(1)使用的摄像头系统；(2)人体表示--人体模型；(3)使用的图像特征；(4)用于确定人体模型的参数（形状，姿势，位置）所使用的算法。用于推断给定图像数据的人体姿势的算法通常分为“生成式”（模型参数可用于生成针对图像数据进行评估的假设，然后对其进行迭代优化以确定最佳拟合）和“判别式”（使用图像数据直接推断模型参数）。通常，无标记运动捕捉系统将具有如下图所示的形式，在此阶段，先前的数据会告知模型设计或基于机器学习的判别算法的训练，然后图像数据将被捕获，处理并输入到可估计人体姿势和形状的计算中。

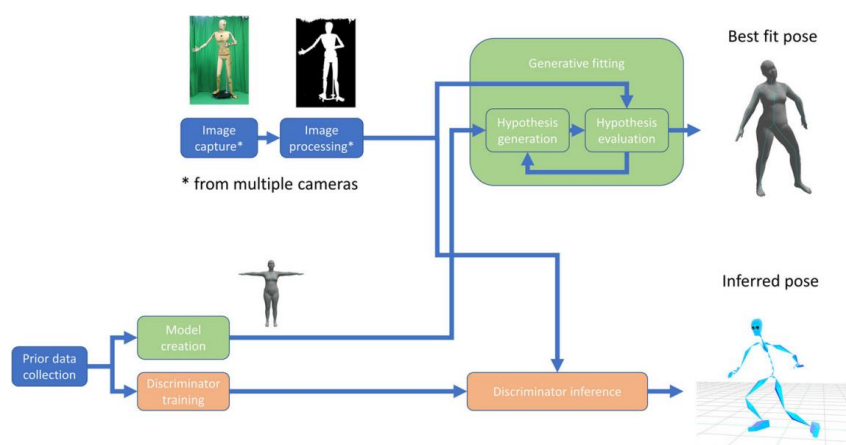


Fig. 1 General structure of a markerless motion capture whether using generative (green) or discriminative (orange) algorithms

### 生成算法

生成运动捕捉方法是通过将人体模型拟合到从图像中提取的信息来确定人的姿势和形状。对于一组给定的模型参数（身体形状，骨骼长度，关节角度），用生成模型表示。然后将该表示形式与从图像中提取的特征进行比较，并计算出一个“误差值”，该误差值表示假设与观察到的数据相差多少。在一种可能性中，可以将由预测参数产生的 3D 三角形网格投影到 2D 图像中，并使网格和人的轮廓的重叠最大化，或者，通过将模型的 3D 顶点与视觉外壳的 3D 点之间的距离最小化，可以通过迭代算法将 3D 人体模型与 3D 重建进行比较最接近的点。

对于生成算法的关键因素是对功能的适当定义，该功能将特定的假设与图像中可用的信息进行比较。如果不仔细考虑，那么搜索最佳模型参数集就很容易失败，从而导致估计结果

的错误。此外,构造一个对图像噪声和不现实模型构型具有鲁棒性的代价函数是比较困难的。在极端情况下,在追踪开始时需要强迫被捕获者摆出一个特定的姿势。如果拟合之后由于遮挡、图像噪声或其他故障而变得混乱,跟踪在没有人工干预的情况下将无法自我纠正。研究人员已经尝试使用改进的搜索算法、身体部位检测器和最近的姿态识别算法中获得的额外信息,或者通过生成方法和鉴别方法的耦合来解决这种情况。

### “生成算法”相关文章解析

#### Multi-view Pictorial Structures for 3D Human Pose Estimation

摘要:

图像结构模型是 2D 人体姿势估计的标准,已提出了许多改进措施,例如,经过区分训练的身体部位检测器,灵活的身体模型以及局部和全局混合。尽管这些技术允许实现 2D 姿态估计的最新性能,但它们尚未扩展为支持 3D 姿态估计。因此,本文提出了一种基于 2D 姿态估计的最新进展的多视图图形结构模型,并结合了跨多个视点的证据以实现可靠的 3D 姿态估计。文中评估了 HumanEva-I 和 MPII Cooking 数据集上的多视图图片结构方法。与 3D 姿态估计的相关工作相比,文中的方法仅在单帧上运行,并且不依赖于特定于运动的运动模型或跟踪,从而获得了相似或更好的结果。对于具有更复杂动作的活动,文中的方法要优于最新技术。

正文:

本文中考虑了来自多个校准摄像机的关节式 3D 人体姿势估计的任务。传统上,此任务使用 3D 人体模型解决,并且涉及 3D 人体配置的高维空间中的复杂推理。为了解决搜索复杂性,已经提出了各种机制,例如退火粒子滤波或非参数置信传播。在本文中,作者认为,通过将 3D 推理问题公式化为每个摄像机视图中姿势的 2D 投影的联合推理,可以显着降低搜索复杂度。为此,文中以成功的 2D 图形结构模型为基础,这些模型被证明对 2D 人体姿态估计有效。仅仅以 2D 推理可以将解决 2D 到 3D 提升歧义的时间延迟到考虑所有图像观测值的时间点,但是这与基于 3D 人体模型的方法相反,该方法需要在推理过程的早期就假设 3D 姿势。

本文提出了一种 2D 姿态估计方法,该方法以色彩特征和更有效的空间术语扩展了作者最新的 2D 图形结构模型。同样,文中将其推广到混合模型,并提出了一种新的混合组分选择方法。同时文中是将 2D 姿势估计模型扩展到多视图模型,该模型对从多个角度看到的人的姿势执行联合推理,然后,此模型的输出将用于恢复 3D 姿势。同时在 HumanEva-I 数据集和 MPII Cooking 数据集上评估该方法,该数据集是多视图 3D 姿态估计的标准基准,在 HumanEva-I 上,文中的方法达到的准确度与别人文献中的最新结果相当,后者依赖于特定活动的运动模型和跟踪,而本文的方法仅在单帧上运行。在 MPII Cooking 上,本文的方法比 2D 方法有了很大的改进,证明了在多个视图中共同估计姿势的优势。

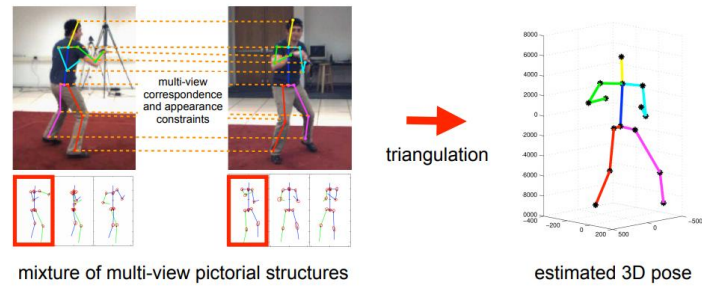


Figure 1: Our approach. (1) Projections of 3D pose in each view are jointly inferred using a mixture of multi-view pictorial structures models. The body layout priors of each mixture component are visualized below, activated components are highlighted in red. (2) 3D pose is recovered via triangulation.

## 1. Single-view model 单视图模型

### 1.1 Pictorial structures model

在下文中，作者描述了依赖于图形结构模型的 2D 姿态估计方法。作者介绍了一种更灵活的零件配置和多模式成对，颜色特征和图形结构的混合。

图片结构模型将人体表示为  $N$  个刚性零件的配置  $L = \{l_1, \dots, l_N\}$  和成对零件关系  $E$  的集合。每个部分的位置由  $l_i = (x_i, y_i, \theta_i)$  给出，其中  $(x_i, y_i)$  是该部分的图像位置，而  $\theta_i$  是绝对方向，分解为一元和成对项的乘积：

$$p(L|I) = \frac{1}{Z} \prod_{n=1}^N f_n(l_n; I) \cdot \prod_{(i,j) \in E} f_{ij}(l_i, l_j).$$

文中不是通过四肢的配置来编码身体姿态，而是通过身体关节的配置来编码。从肢体转换到关节的优点是，新模型可以更好地编码出平面旋转导致的身体部位的透视缩短。文中的新模型有 14 个部分，分别对应躯干、头部、左、右手腕、肘部、肩膀、脚踝、膝盖和臀部，MPII 数据集只使用了 10 个上半身部分。

### 1.2 Appearance representation

零件似然项用增强零件检测器表示，该零件检测器依赖于使用密集计算的形状上下文描述符网格对图像的编码。本文使用颜色特征来增强在增强部分检测器中使用的形状上下文特征，例如手或头部经常具有特征性的肤色。另外，某些颜色比身体部位更可能对应于相应的背景。为此，作者针对 RGB 颜色空间的每个维度使用 10 个面元的多维直方图对零件边界框的颜色进行编码，从而得出 1000 维度的特征向量，将形状上下文与颜色特征连接起来，并在此组合表示的基础上学习增强型零件检测器。

### 1.3 Spatial model

下列等式编码模型各部分之间的空间约束，并在两部分之间的关节的变换空间中以高斯分布进行建模：

$$f_{ij}(l_i, l_j) = \mathcal{N}(T_{ji}(l_i) - T_{ij}(l_j) | \mu^{ij}, \Sigma^{ij})$$

作者通过在这些成对的部分依赖层引入混合模型来扩展模型。为此，将上述公式中的单峰高斯项替换为  $K$  个模态的最大值项，并用高斯表示每个模态。新的多模态两两项为：

$$f_{ij}(l_i, l_j) = \max_{k=1}^K \mathcal{N}(T_{ji}^k(l_i) - T_{ij}^k(l_j) | \mu_{ij}^k, \Sigma_{ij}^k).$$

### 1.4 Mixtures of pictorial structures (Mixture PS)

作者将方法扩展到混合图形结构模型。作者通过将训练数据与  $k$  均值聚类并为每个聚类学习单独的模型来获得混合成分。组件通常对应于数据中的主要模式，例如人相对于相机的各种视点。组件的索引被视为潜在变量，在测试时进行推断。作者发现使用等式中的后验值对预测最佳混合成分不可靠，因此提出两种替代策略。

a. 组件分类器：作者训练了一个整体分类器，该分类器根据人员边界框的内容来区分混合成分。为此，作者使用的方法可以共同解决对象检测和视点分类的任务，但依赖于结构化的预测公式，该公式既可以鼓励正确的定位又可以进行组件检测。

b. 最小方差：作者使用与姿势估计的质量直接相关的标准选择混合分量，选择了身体各部位后缘分布具有最小不确定性的最佳组件。

## 2. Multi-view model 多视图模型

描述 3D 姿态估计的方法包括两个步骤。第一步，作者共同估算每个视图中 3D 人体关节的 2D 投影。作为表示和推断人体姿势的基本工具，作者依赖于单视图模型中引入的 2D 模型。在第二步中，作者使用估计的 2D 投影并通过三角剖分恢复 3D 姿势。

为了清楚起见，作者首先针对两个视图的情况介绍多视图模型。类似于单视图模型中的



公式，有条件的身体后部结构在两个视图中分解为一元和成对项的乘积，它们为每个视图独立定义零件之间的外观和空间约束。另外，作者在每个视图的每对对应部分之间引入成对因子。两种视图中的关节后部结构由下式给出：

$$p(L_1, L_2 | I_1, I_2) = \frac{1}{Z} f(L_1; I_1) f(L_2; I_2) \prod_n f_n^{app}(\mathbf{l}_n^1, \mathbf{l}_n^2; I_1, I_2) f_n^{cor}(\mathbf{l}_n^1, \mathbf{l}_n^2)$$

### 2.1 Multi-view appearance

$f_n^{app}$  因子编码从多个角度看到的身体部位的颜色和形状。作者通过连接来自多个视图的特征来定义关节外观特征向量，并使用此表示来训练增强部分检测器。外观因素取决于每个视图中零件的位置。与单视图增强部分检测器相比，多视图检测器在训练期间可以访问所有视图中的特征，并且可以利用视图中的特征的同时出现，来学习更具判别性的检测器。

$f_n^{cor}$  因子对每个视图中的零件位置应在同一 3D 位置上的一致性进行约束。给定一对对应的零件位置  $\mathbf{l}_n^1$  和  $\mathbf{l}_n^2$ ，作者首先使用线性三角测量在 3D 中重建零件的对应位置，多视图对应因子为

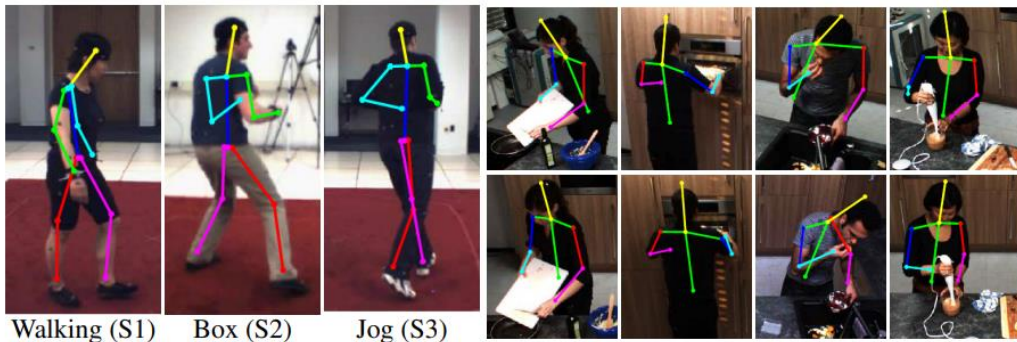
$$f_n^{cor}(\mathbf{l}_n^1, \mathbf{l}_n^2) = \exp(-(\|\mathbf{l}_n^1 - \hat{\mathbf{l}}_n^1\|^2 + \|\mathbf{l}_n^2 - \hat{\mathbf{l}}_n^2\|^2))$$

### 2.2 3D mixture model

多视图模型也采用图形结构的混合来表示每个视图的 2D 身体配置。但是，在多视图情况下，混合分量对应于 3D 中类似的姿势组。为了获得此类 3D 混合分量，作者首先将 3D 训练姿势与 k 均值聚类。然后，投影每个 3D 群集的训练数据，并从投影数据中学习 2D 模型。对于组件检测器，作者在所有视图添加相应组件的分数。对于基于不确定性的标准，作者将在单视图模型中的空间模型中为所有视图中的每个对应组件添加不确定性得分。

### 2.3 Inference

在简化的假设的前提下，图形结构方法可以进行有效而精确的推断。但是，这些假设限制了模型的表达能力。例如，等式中的成对因子以及等式中的多视图因子不是高斯模型，并在模型结构中创建循环依赖。为了用非高斯因素和循环模型进行推理，作者使用近似两阶段推理程序。在第一阶段，此过程依赖于具有高斯成对因子以及简单的形状和颜色外观术语的简化树结构模型，以便生成有关身体部位位置的建议。这个阶段可以看作是减少搜索空间的步骤，这是应用更复杂的模型所必需的，第一阶段的推论是通过求和积置信传播来执行的，允许计算每个身体部位  $p(l_i | \mathbf{I})$  的边缘分布。该推论是精确而有效的，因为该模型是树形结构的，并且可以使用高斯卷积来计算消息。在第二阶段中，作者从  $p(l_i | \mathbf{I})$  中采样了足够大的位置集，并在所有模型中使用所有因素在采样部分位置的缩减状态空间中进行了推断，使用最大乘积置信度传播，因为它允许获得整个身体配置的一致估计。最后，给定多视图图片结构模型估计的 2D 投影，使用三角剖分重建 3D 姿势。



(a) HumanEva-I

(b) MPII Cooking, top: cam 1, bottom: cam 2.

实验结果

---

结论:

传统上会使用 3D 人体模型解决 3D 人体姿势估计问题。在这项工作中, 作者采用了另一种方法, 并将问题重新表述为对每个摄像机视图中 3D 姿势的 2D 投影集的推断。这种替代的公式建立在最先进的图片结构模型的基础上, 可以从 2D 人体姿势估计的最新进展中受益。通过扩展之前提出的原始模型, 该模型具有灵活的部分, 颜色特征, 多模式成对术语以及图形结构的混合, 作者的 2D 姿态估计方法显著提高了用于评估的两个数据集的性能。为了利用多视图信息, 作者使用跨视图的外观和空间对应约束来扩充模型。

### 判别算法

判别算法避免了反复调整身体模型参数以适应图像的过程, 因此它们也通常被称为无模型算法。与生成方法相比, 它们通常会有更快的处理时间, 改进的健壮性和减少对初始猜测的依赖。然而, 它们可以降低精确度, 并且它们需要一个非常大的样本数据数据库(甚至比生成算法所使用的构建统计体型模型所需要的数据还要多), 从中它们可以学习如何推断结果。判别算法主要有两种方法, 一种方法是直接发现从图像特征到位姿描述的映射, 例如使用基于机器学习的回归, 从而就有可能“教”计算机如何仅使用图像数据来确定一个简单骨架模型的姿态。另一个方法是利用深度学习来训练一个系统, 这个系统可以识别多个人的身体部位, 然后快速解析这些部位来确定骨骼。或者, 也可以创建一个位姿示例数据库, 然后搜索当前图像中已知的最相似的位姿。

### “判别算法”相关文章解析

#### Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields

摘要:

作者提出一种有效检测图像中多个人的 2D 姿势的方法。该方法使用非参数表示(作者称为 PAF)来学习将身体部位与图像中的个体相关联。该体系结构对全局上下文进行编码, 从而允许自下而上的解析步骤, 无论图像中有多少人, 都可以在保持高精度的同时实现实时性能。该体系结构旨在通过同一顺序预测过程的两个分支共同学习零件位置及其关联。作者的方法在首次 COCO 2016 关键点挑战中排在首位, 在性能和效率上都大大超过了 MPII MultiPerson 基准测试之前的最新结果。

正文:

人体 2D 姿态估计-定位解剖学关键点或部位的问题, 主要集中在寻找个体的身体部位。推断图像中的多个人的姿势, 尤其是从事社交活动的人, 会带来一系列独特的挑战。首先, 每个图像可能包含未知数量的人, 这些人可以在任何位置或任何比例出现。其次, 由于接触, 咬合和肢体关节活动, 人与人之间的互动会引起复杂的空间干扰, 从而使零件的关联变得困难。第三, 运行时复杂度往往随着映像中的人数而增加, 这使实时性能成为挑战。

在本文中, 作者提出了一种有效的方法, 可以在多个公共基准上以最先进的精度进行多人姿势估计。作者通过 PAF 展示了关联评分的第一个自下而上的表示形式, 这是一组二维矢量场, 它们在图像域上编码肢体的位置和方向。作者证明, 同时推断出这些检测和关联的自下而上的表示方式, 可以很好地编码全局上下文, 以使贪婪的解析能够以少量的计算成本来获得高质量的结果。

Simultaneous Detection and Association:

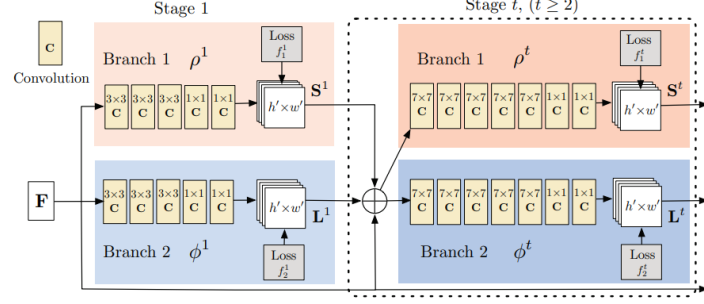


Figure 3. Architecture of the two-branch multi-stage CNN. Each stage in the first branch predicts confidence maps  $S^t$ , and each stage in the second branch predicts PAFs  $L^t$ . After each stage, the predictions from the two branches, along with the image features, are concatenated for next stage.

作者的架构，如上图所示，同时预测检测置信度映射和编码部分到部分关联的亲域。这个网络被分成两个分支，上面的分支(用米色表示)预测可信度图，下面的分支(用蓝色表示)预测亲和力和域。每个分支都是一个迭代架构，它细化了连续阶段的预测。首先通过卷积网络(由 VGG-19 和微调的前 10 层初始化)对图像进行分析，生成一组特征图  $F$ ，输入到每个分支的第一阶段。在第一阶段，网络生成一组检测置信度图和一组部分亲和场，其中，lawren1 和 czain1 是第一阶段进行推理的 CNNs。在随后的每一个阶段中，前一阶段的两个分支的预测，连同原始图像特征  $F$ ，被连接起来并用于产生精确的预测，并进行了跨阶段的置信度图和亲和度字段的细化。为了指导网络迭代预测第一个支路的身体部分和第二个支路的 PAF 的置信度图，作者在每个阶段末分别应用两个损失函数和一个损失函数。作者在估计预测和真实地面图和场之间使用 L2 损失。因此，作者对损失函数进行空间加权来解决一个实际问题，从而使一些数据集不能完全的标记所有人。具体而言， $t$  阶段两个分行的损失函数为：

$$f_S^t = \sum_{j=1}^J \sum_{\mathbf{p}} \mathbf{W}(\mathbf{p}) \cdot \|S_j^t(\mathbf{p}) - S_j^*(\mathbf{p})\|_2^2,$$

$$f_L^t = \sum_{c=1}^C \sum_{\mathbf{p}} \mathbf{W}(\mathbf{p}) \cdot \|L_c^t(\mathbf{p}) - L_c^*(\mathbf{p})\|_2^2,$$

每个阶段的中间监督通过定期补充梯度来解决梯度消失的问题，总目标是：

$$f = \sum_{t=1}^T (f_S^t + f_L^t).$$

#### Confidence Maps for Part Detection

在训练期间，作者从带注释的 2D 关键点生成接地的真实度置信图。每个置信度图是特定身体部位出现在每个像素位置的信念的 2D 表示。理想情况下，如果图像中只有一个人，则只要可见部分可见，则每个置信度图中应存在一个峰值；如果出现多个人，则应该为每个人  $k$  对应于每个可见部分  $j$  的峰值。作者先为每一个人生成置信图，网络中要预测的接地真实度置信度图是通过最大算子对单个置信度图进行的汇总。

#### Part Affinity Fields for Part Association

给定一组检测到的身体部位，如何组装它们以形成未知人数的全身姿势？作者是对每对身体部位检测的关联性进行置信度量，即它们属于同一个人。测量关联的一种可能方法是检测肢体上每对零件之间的附加中点，并检查候选零件检测之间其发生率，但是，当人们挤在一起时，这些中点很可能支持虚假的关联。这种假联想是由于表示方面的两个限制而引起的：

(1)它仅编码每个肢体的位置，而不是编码方向；(2)将肢体的支撑区域减少到单个点。

为了解决这些限制，作者提出了一种新颖的特征表示，称为零件亲和力量段，该字段保留了肢体支撑区域中的位置和方向信息。零件相似性是每个肢体的 2D 矢量场：对于属于特定肢体的区域中的每个像素，2D 向量编码从肢体的一部分指向另一部分的方向。每种肢体都有一个对应的亲和力量场，将其两个相关的身体部位连接在一起。

在测试期间，作者通过计算对应 PAF 上沿着连接候选零件位置的线段的线积分来测量候选零件检测之间的关联，即作者测量预测的 PAF 与候选肢体的对齐方式，该肢体将通过连接检测到的身体部位而形成。具体来说，对于两个候选零件位置  $d_{j1}$  和  $d_{j2}$ ，作者沿着线段对预测零件字段进行采样，以测量对其关联的置信度：

$$E = \int_{u=0}^{u=1} \mathbf{L}_c(\mathbf{p}(u)) \cdot \frac{\mathbf{d}_{j2} - \mathbf{d}_{j1}}{\|\mathbf{d}_{j2} - \mathbf{d}_{j1}\|_2} du$$

#### Multi-Person Parsing using PAFs

作者对检测置信度图执行非极大值抑制，以获得零件候选位置的离散集合。对于每个部分，由于图像中有多个人或误判断，作者可能有多个候选对象。这些候选零件定义了大量可能的肢体，作者使用定义的 PAF 上的线积分计算为每个候选肢体评分。找到最佳解析的问题对应于一个称为 NP-Hard 的 K 维匹配问题。在本文中，作者提出了一个贪婪的松弛，该松弛持续产生高质量的匹配项。作者推测原因是由于 PAF 网络的接收域很大，成对关联分数隐式编码全局上下文。

当要找到多个人的全身姿势时，确定 Z 是 K 维匹配问题。在这项工作中，作者为优化添加了两个放宽部分。首先，选择了最少数量的边缘来获得人体姿势的生成树骨架，而不是使用完整的图形。其次，进一步将匹配问题分解为一组二分匹配子问题，并独立确定相邻树节点中的匹配。本文中显示了详细的比较结果，该结果表明，最小的贪婪推理以很小的计算成本就能很好地逼近全局解。原因是，相邻树节点之间的关系是通过 PAF 显式建模的，但是内部，非相邻树节点之间的关系是由 CNN 隐式建模的。之所以会出现此属性，是因为 CNN 是在大接收域下训练的，并且来自非相邻树节点的 PAF 也影响预测的 PAF。

通过这两个松弛，优化可以简单地分解为：

$$\max_{\mathbf{Z}} E = \sum_{c=1}^C \max_{\mathbf{Z}_c} E_c.$$

使用相应等式分别获得每种肢体类型的肢体连接候选者。使用所有肢体连接候选者，可以将共享相同零件检测候选者的连接组装成多个人的全身姿势，在树结构上的优化方案比在全连接图上的优化快几个数量级。

实验结果：

Method	Hea	Sho	Elb	Wri	Hip	Kne	Ank	mAP	s/image
Subset of 288 images as in [22]									
Deepcut [22]	73.4	71.8	57.9	39.9	56.7	44.0	32.0	54.1	57995
Iqbal et al. [12]	70.0	65.2	56.4	46.1	52.7	47.9	44.5	54.7	10
DeeperCut [11]	87.9	84.0	71.9	63.9	68.8	63.8	58.1	71.2	230
Ours	<b>93.7</b>	<b>91.4</b>	<b>81.4</b>	<b>72.5</b>	<b>77.7</b>	<b>73.0</b>	<b>68.1</b>	<b>79.7</b>	<b>0.005</b>
Full testing set									
DeeperCut [11]	78.4	72.5	60.2	51.0	57.2	52.0	45.4	59.5	485
Iqbal et al. [12]	58.4	53.9	44.5	35.0	42.2	36.7	31.1	43.1	10
Ours (one scale)	89.0	84.9	74.9	64.2	71.0	65.6	58.1	72.5	0.005
Ours	<b>91.2</b>	<b>87.6</b>	<b>77.7</b>	<b>66.8</b>	<b>75.4</b>	<b>68.9</b>	<b>61.7</b>	<b>75.6</b>	<b>0.005</b>

Table 1. Results on the MPII dataset. Top: Comparison result on the testing subset. Middle: Comparison results on the whole testing set. Testing without scale search is denoted as “(one scale)”.

Team	AP	AP <sup>50</sup>	AP <sup>75</sup>	AP <sup>M</sup>	AP <sup>L</sup>
Test-challenge					
Ours	<b>60.5</b>	<b>83.4</b>	<b>66.4</b>	55.1	<b>68.1</b>
G-RMI [19]	59.8	81.0	65.1	<b>56.7</b>	66.7
DL-61	53.3	75.1	48.5	55.5	54.8
R4D	49.7	74.3	54.5	45.6	55.6
Test-dev					
Ours	<b>61.8</b>	<b>84.9</b>	<b>67.5</b>	57.1	<b>68.2</b>
G-RMI [19]	60.5	82.2	66.2	<b>57.6</b>	66.6
DL-61	54.4	75.3	50.9	58.3	54.3
R4D	51.4	75.0	55.9	47.4	56.7

Table 3. Results on the COCO 2016 keypoint challenge. Top: results on test-challenge. Bottom: results on test-dev (top methods only). AP<sup>50</sup> is for OKS = 0.5, AP<sup>L</sup> is for large scale persons.

在本文中，作者考虑了这种感知的关键组成部分：实时算法来检测图像中多个人的 2D 姿势。提出了关键点关联的显式非参数表示形式，它编码了人类四肢的位置和方向。其次，作者设计了一种用于共同学习零件检测和零件关联的架构。第三，作者证明了文中的解析算法足以产生高质量的人体姿势解析，即使图像中的人数增加，该算法也能保持效率。

发布链接: <https://mp.weixin.qq.com/s/mrDvYV3j7whGf2aqWkLXyg>

#### Reference:

1. A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System
2. Multi-view Pictorial Structures for 3D Human Pose Estimation
3. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields