

状态

- [已开源](#)
- [文章链接](#)

简介

主要思想是如何利用一些时间序列信息来提高多人姿态追踪的效果，工程的东西比较多，也比较复杂。

才用的top-down模式，提出多个帧的人物姿态关系的pose flow (PF-Builder) 和基于pose flow的NMS方法。

目前主流的pose tracking还是基于top down的方法，因为bottom-up的很容易导致肢体分配出现错误，如下图所示：



即时采用top-down方法，也很难避免以下问题：

1. 视频帧模糊
2. 遮挡
3. 部分肢体部位未出现在视频中

b)



c)



d)



Groundtruth

False cases

最后效果是13mAP 25MOTA和6mAP 3MOTA，速度10fps。

方法简介

首先介绍pose flow这个概念，其含义是在多个帧当中代表某个特定人物实例的pose序列。大体步骤分两步，第一步是在时间序列上找到置信度最高的一段序列，第二步是删除多余的pose flow并且将间断的pose flow重新链接起来。

Preliminary

- Intra-Frame Pose Distance: 衡量某一帧当中两个pose P1和P2的相似性。soft matching function:

$$K_{Sim}(P_1, P_2 | \sigma_1) = \begin{cases} \sum_n \tanh \frac{c_1^n}{\sigma_1} \cdot \tanh \frac{c_2^n}{\sigma_1} & \text{if } p_2^n \text{ is within } B(p_1^n) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

其中P的上标n代表第n个关键点， $B(p_1^n)$ 是中心在p1的bbox，c代表对应的score。

关键点间的空间相似性如下：

$$H_{Sim}(P_1, P_2 | \sigma_2) = \sum_n \exp\left[-\frac{(p_1^n - p_2^n)^2}{\sigma_2}\right] \quad (2)$$

最后的相似性：

$$\begin{aligned} d_f(P_1, P_2 | \Lambda) \\ = K_{Sim}(P_1, P_2 | \sigma_1)^{-1} + \lambda H_{Sim}(P_1, P_2 | \sigma_2)^{-1} \end{aligned} \quad (3)$$

where $\Lambda = \{\sigma_1, \sigma_2, \lambda\}$. These parameters can be determined in a data-driven manner.

- Inter-frame Pose DistanceL 衡量某一帧的pose P1和下一帧的poses P2之间的相关性，这个指标需要考虑时间因素，让B代表对应的bounding box，其中t代表从B1当中提取出来的DeepMatch feature points：

$$d_c(P_1, P_2) = \sum_n \frac{f_2^n}{f_1^n} \quad (4)$$

其他tricks

- 数据增强： 利用一种叫做iDPG的improved deep proposal generator来产生出多个crop之后的human proposal可以使得结果更加鲁棒
- Motion-Guided Box Propagation: 由于运动模糊和遮挡等各种原因，容易发生检测丢失，会导致人物id不一致的问题。作者采用的方法很简单，将当前帧的proposal通过匹配策略(文中用deepmatching)传递到上一帧和下一帧当中。

Pose Flow Building

这个模块是为了实现在多个帧当中找到同一个人的poses。具体是通过在时间轴上滑窗的方式，从pose proposal当中迭代的估计出pose flow。

之后采用的方法并非greedy match，利用 $d_c(P_1, P_2)$ 找到下一帧当中距离最近的pose。而是在所有潜在的flow当中找到置信度最高的，这种优化方法有利于稳定pose flow并且可以更好的处理间断性人物丢失的问题。

假定 P_i^j 代表第i个pose在第j帧，首先可以求其候选的其它帧同一人物Pose:

$$\begin{aligned} \mathcal{T}(P_i^j) = \{P | d_c(P, P_i^j) \leq \varepsilon\}, \\ s.t. P \in \Omega_{j+1} \end{aligned} \quad (5)$$

其中 Ω_{j+1} 代表第j+1帧的所有pose集合。

$$\begin{aligned} F(t, T) &= \max_{Q_t, \dots, Q_{t+T}} \sum_{i=t}^{t+T} s(Q_i), \\ \text{s.t. } Q_0 &= P_i^t, \\ \text{s.t. } Q_i &\in \mathcal{T}(Q_{i-1}) \end{aligned} \quad (6)$$

where $s(Q_i)$ is a function that outputs confidence score of Q_i , which is defined as

$$s(Q_i) = s_{box}(Q_i) + \text{mean}(s_{pose}(Q_i)) + \max(s_{pose}(Q_i)) \quad (7)$$

where $s_{box}(P)$, $\text{mean}(s_{pose}(P))$ and $\max(s_{pose}(P))$ are score of human box, mean score and max score of all keypoints within this human proposal, respectively. The optimum $\{Q_t, \dots, Q_{t+T}\}$ is our pose flow for P_i^t from t^{th} to $(t+T)^{th}$ frame.

不难看出优化的目标是一个连加的形式，这样就算当中某一帧分数比较低，也可以通过Pose flow最大化的形式补偿。

可以利用动态规划进行求解，逐帧求解直到达到u:

$$F(t, u+r) - F(t, u) < \gamma$$

以为着在接下来的r帧之内分数之和很小，所以pose flow就此停止。文章通过交叉验证r=3。当poses flow停下来后会用平均的置信度代替每一帧的置信度，因为作者认为pose flow是一个基本单元，应该用单一的置信度来表示。这个步骤称为confidence unification。

PF-NMS

以pose flow为基本单元进行NMS，在时间域当中进行。关键在于定义同一人物不同pose flow的距离。

- Pose Flow Distance: 首先可以把两个pose flow在时间域内重叠的部分提取出来，然后计算每一帧两个pose的距离（相似性）：

$$d_{PF}(\mathcal{Y}_a, \mathcal{Y}_b) = \text{median}[\{d_f(P_a^1, P_b^1), \dots, d_f(P_a^N, P_b^N)\}] \quad (8)$$

- Pose Flow Merging: 接下来的过程和NMS差不多，首先选一个分数最高的(after confidence unification)pose flow，然后计算别的flow和它的距离，首先距离近的flow可以先进行group，group之后可以merge成更长更robust的flow，新的pose flow称为representative pose flow。其坐标和置信度为：

$$\hat{\mathbf{x}}_{t,i} = \frac{\sum_j s_{t,i}^j \mathbf{x}_{t,i}^j}{\sum_j s_{t,i}^j} \quad \text{and} \quad \hat{s}_{t,i} = \frac{\sum_j s_{t,i}^j}{\sum \mathbb{1}(s_{t,i}^j)} \quad (9)$$

这一过程是通过滑窗的方式进行得(L=20)，所以也是在线的。

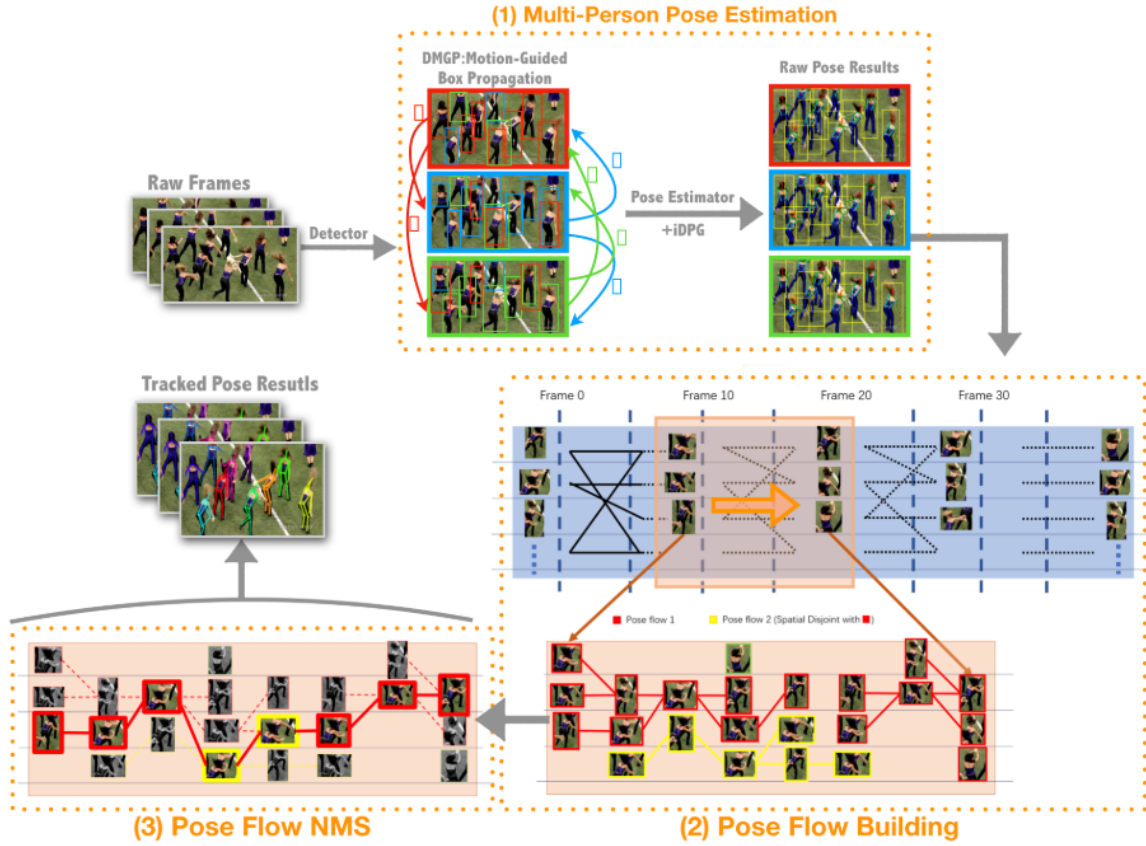


Figure 2: Overall Pipeline: 1) Pose Estimator. 2) Pose Flow Builder. 3) Pose Flow NMS. First, we estimate multi-person poses. Second, we build pose flows by maximizing overall confidence and purify them by Pose Flow NMS. Finally, reasonable multi-pose trajectories can be obtained.

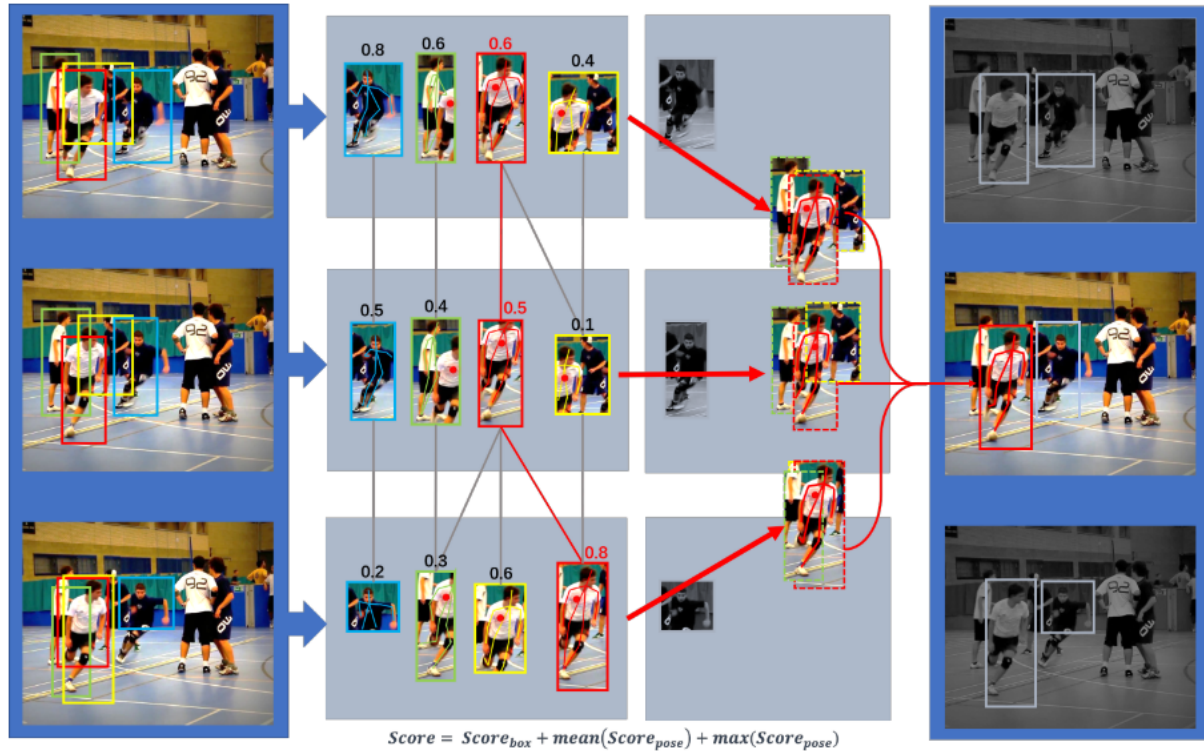


Figure 3: Pose Flow Merging

结果分析

最后实验证明虽然map没有太多上升但是MOTA类的指标的上升还是非常明显的