文献分享

20200813

人体姿态估计

- 单人姿态估计 (Single-Person Skeleton Estimation)
  常见的数据集有MPII, LSP, FLIC, LIP。

- 多人姿态估计 (Multi-person Pose Estimation)
  一般有两种做法，分别是top-down以及bottom-up的方法。

- 人体姿态跟踪 (Video Pose Tracking)
  主要是针对视频场景中的每一个行人，进行人体以及每个关键点的跟踪。

- 3D人体姿态估计 (3D Skeleton Estimation)
  输入RGB图像，输出3D的人体关键点。

应用

- 人体的动作行为估计
- 人体交互，美体等
- 其他算法的辅助环节

姿态跟踪

- **姿态跟踪：** 在视频中估计多人姿态并为帧中的每个关键点分配唯一实例ID的任务。
- **数据集：**

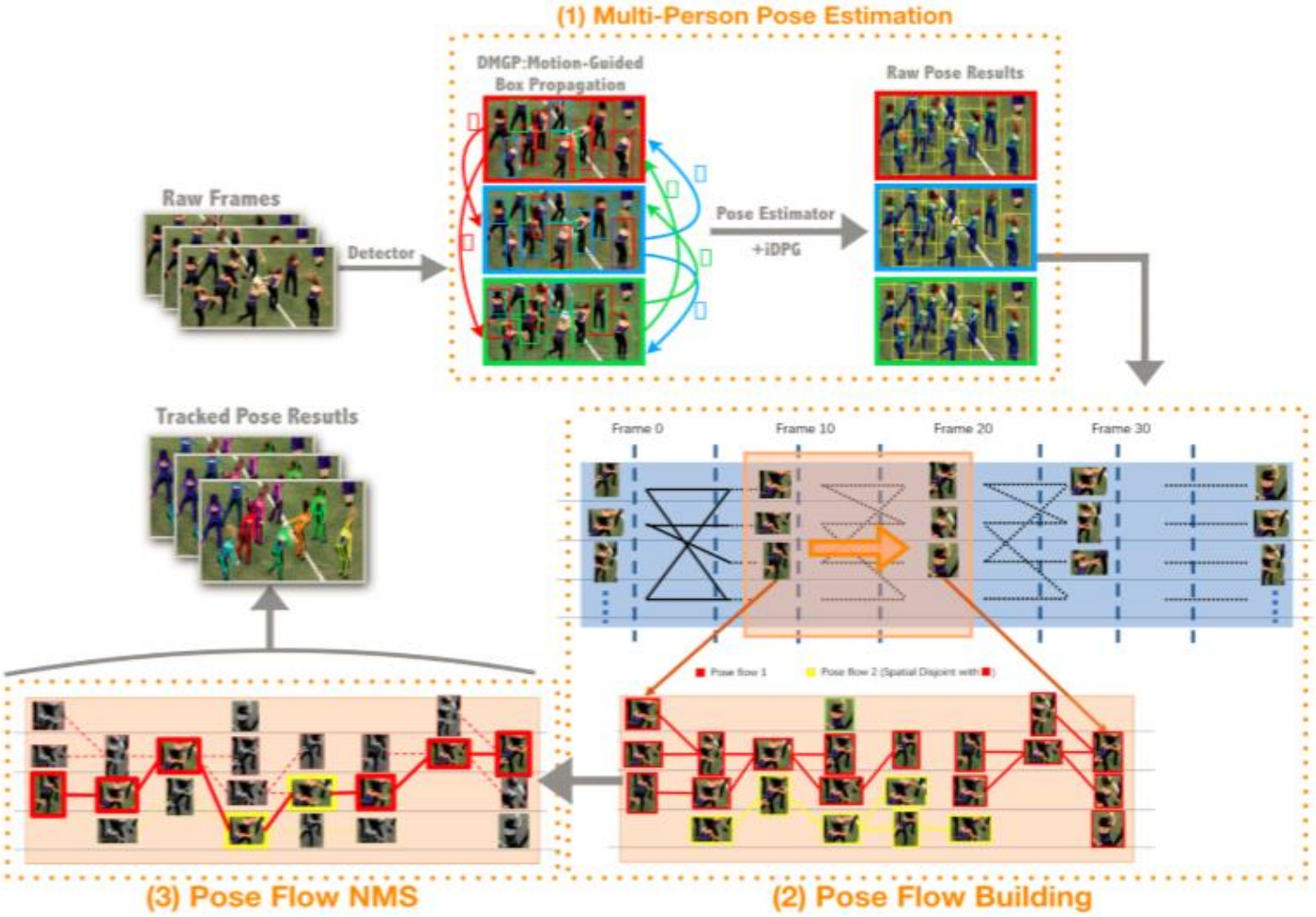
  PoseTrack数据集

  MPII视频姿态数据集

Efficient Online Pose Tracking

- 目前自上而下的PT方法仍然有以下主要问题：
- 1）如何利用相邻帧的信息过滤掉多余的检测bbox；
- 2）如何利用时间信息产生鲁棒的PF；
- 3）如何把同一个目标的检测框关联到一起，同时消除尺度的变化。

Approach


(1) Multi-Person Pose Estimation
DMGP:Motion-Guided Box Propagation
Raw Pose Results
Raw Frames
Detector
Pose Estimator +iDPG
Tracked Pose Resutls
Frame 0
Frame 10
Frame 20
Frame 30
(3) Pose Flow NMS
(2) Pose Flow Building

Preliminary

将P_1和 P_2中的第 n^{th}个关键点分别表示为 p_1^n 和 p_2^n ，$B(p_1^n)$是一个中心位于p_1^n的box。

- Intra-Frame Pose Distance:

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

- Inter-frame Pose Distance:

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

Pose Flow Building

- 选择不同帧中 $d_c(P_1, P_2)$ 最为接近的Pose。

$$\mathcal{T}(P_i^j) = \{P | d_c(P, P_i^j) \leq \epsilon\}, s.t.\ P \in \Omega_{j+1}$$

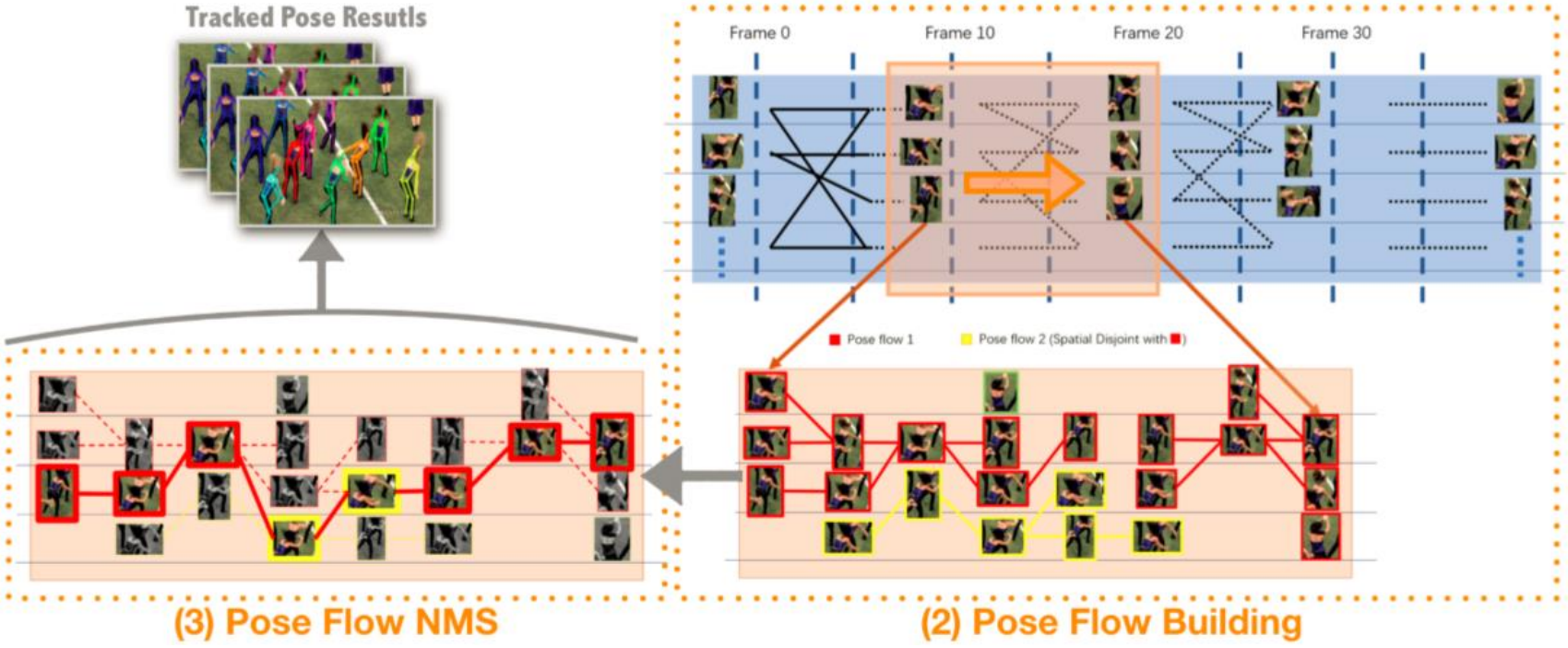
$$F(t, T) = \max_{Q_t, .., Q_{t+T}} \sum_{i=t}^{t+T} s(Q_i),$$

$$s.t.\ Q_0 = P_i^t,$$

$$s.t.\ Q_i \in \mathcal{T}(Q_{i-1})$$

- 通过求解最优化问题来获得我们最优的t到t+T帧的PF。

Pose Flow NMS



Pose Flow NMS

- Pose Flow Distance: 给定两个PF：ya和yb，提取它们的时域重叠子流

$$\{P_a^1, \ldots, P_a^N\}, \{P_b^1, \ldots, P_b^N\}$$

$$d_{PF}(y_a, y_b) = \text{median}[\{d_f(P_a^1, P_b^1), \ldots, d_f(P_a^N, P_b^N)\}]$$

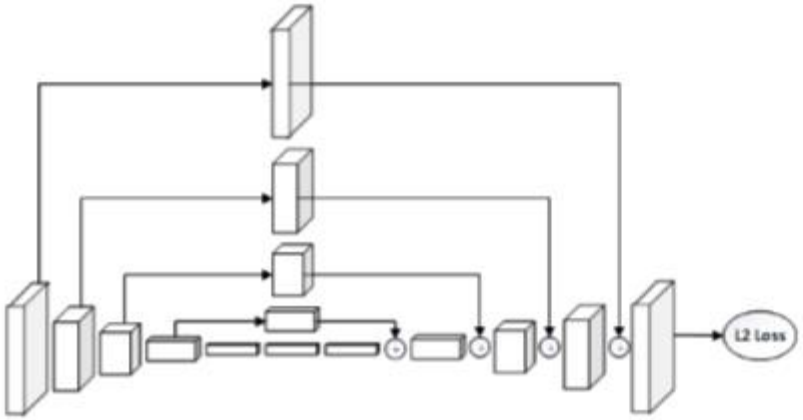
- Pose Flow Merging: 给定距离函数 ，我们可以用传统方法进行NMS。有最大置信度的PF会作为reference PF。

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

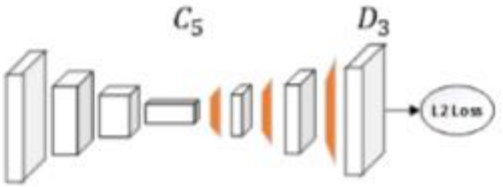
result

Method	Dataset	MOTA Head	MOTA Shou	MOTA Elb	MOTA Wri	MOTA Hip	MOTA Knee	MOTA Ankl	MOTA Total	MOTP Total	Prcn	Rcll
Girdhar *et al.* [8]	validation	61.7	65.5	57.3	45.7	54.3	53.1	45.7	55.2	61.5	88.1	66.5
Ours		59.8	67.0	59.8	51.6	60.0	58.4	50.5	**58.3**	67.8	87.0	70.3
Girdhar *et al.* [8]	*(Mini)Test v1.0 testset	55.9	59.0	51.9	43.9	47.2	46.3	40.1	49.6	34.1	81.9	67.4
Ours		52.0	57.4	52.8	46.6	51.0	51.2	45.3	**51.0**	16.9	78.9	71.2

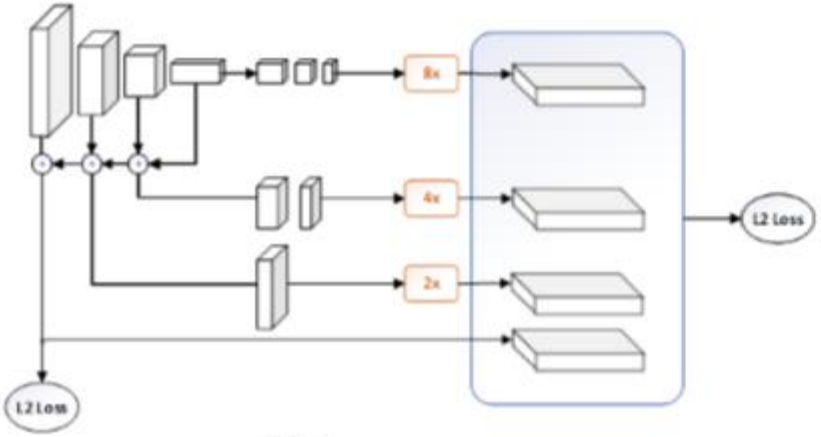
Simple Baselines for Human Pose Estimation and Tracking



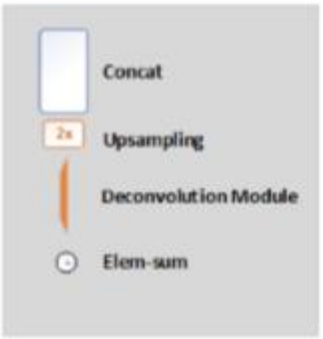
(a) Hourglass



(c) Our Network



(b) CPN



Approach

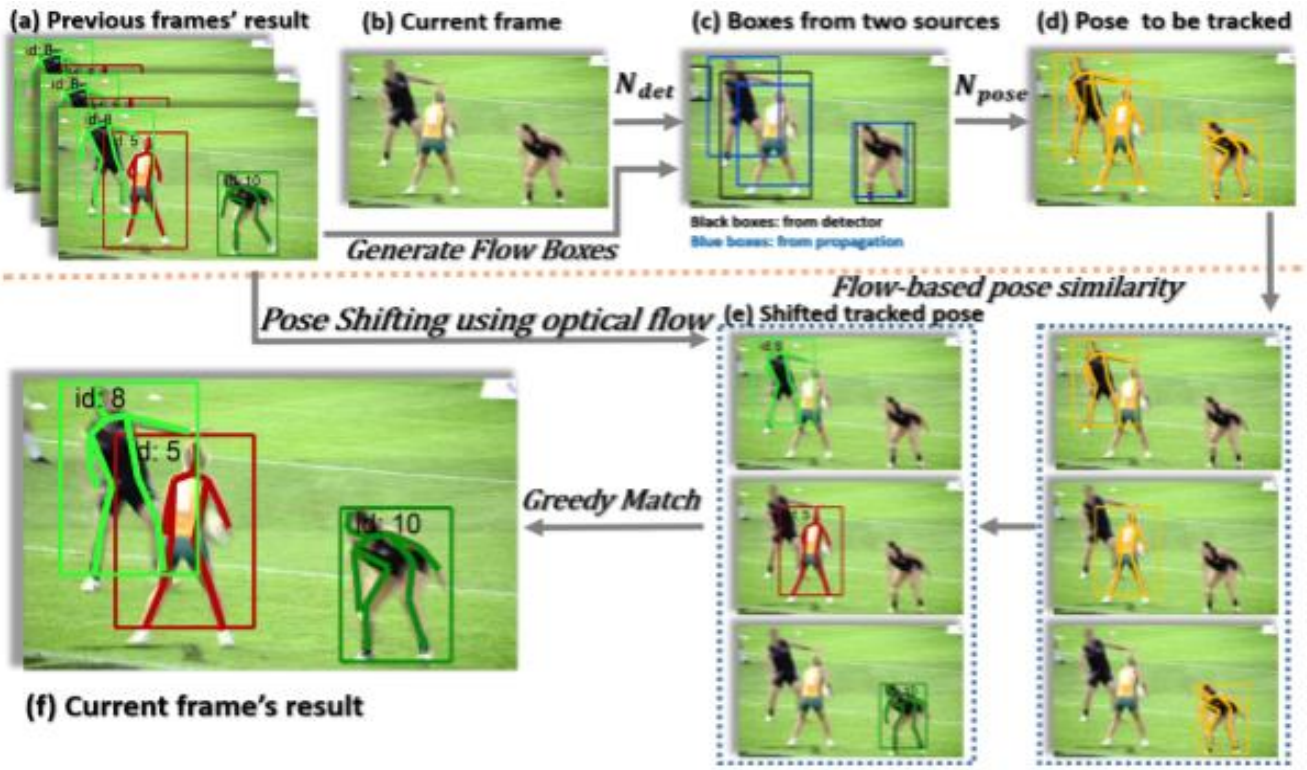
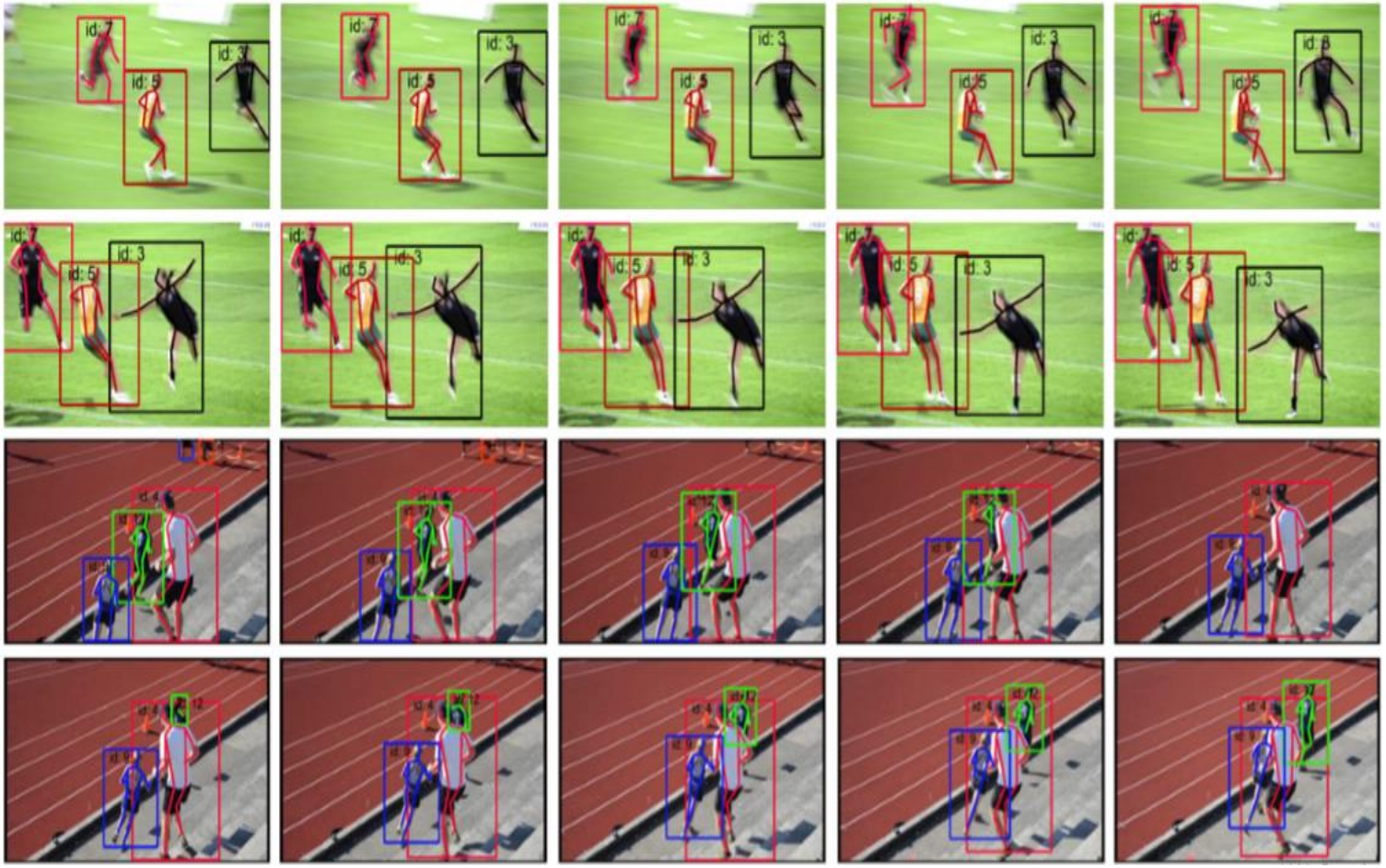


Fig. 2. The proposed flow-based pose tracking framework.

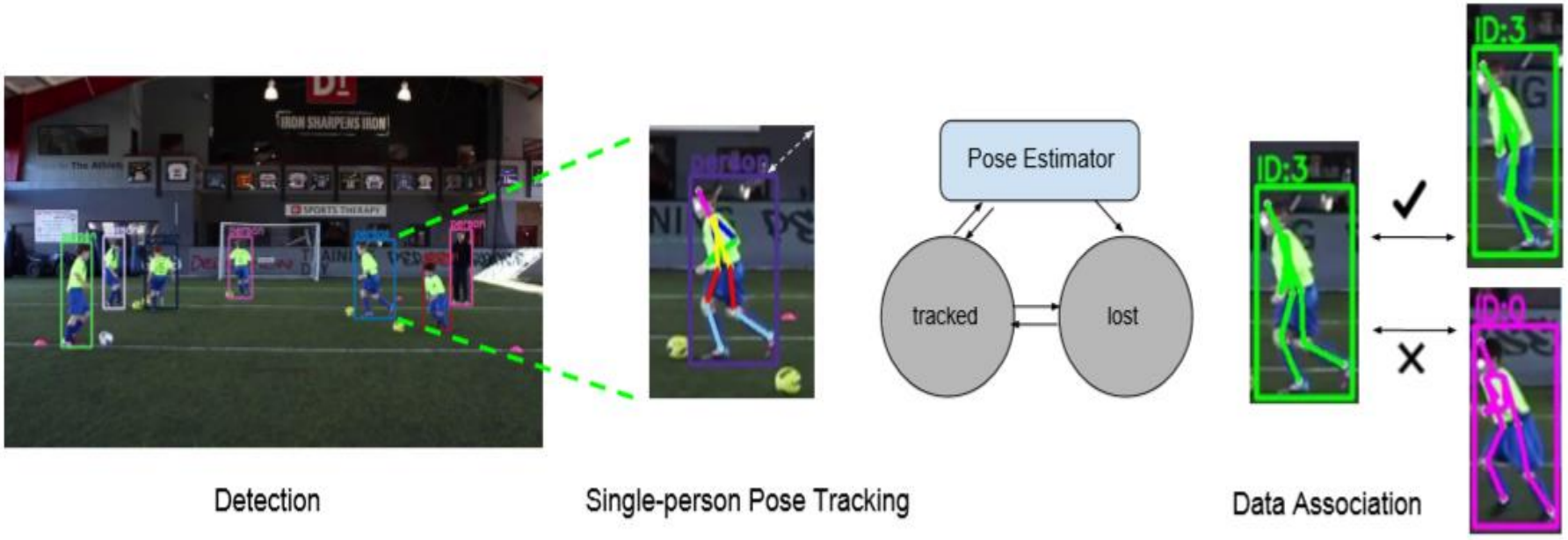
Object Keypoint Similarity (OKS)

$$OKS = \sum_i [\exp(-d_i^2 / 2s^2 k_i^2) \delta(v_i > 0)] \; / \; \sum_i [\delta(v_i > 0)]$$



LightTrack: A Generic Framework for Online Top-Down Human Pose Tracking

- （1）我们提出了一种通用的在线姿态跟踪框架，适用于人体姿态估计的自上而下的方法。
- （2）我们在姿态跟踪系统中提出了一种用于人体姿态匹配的孪生图卷积网络（SGCN）作为Re-ID模块。



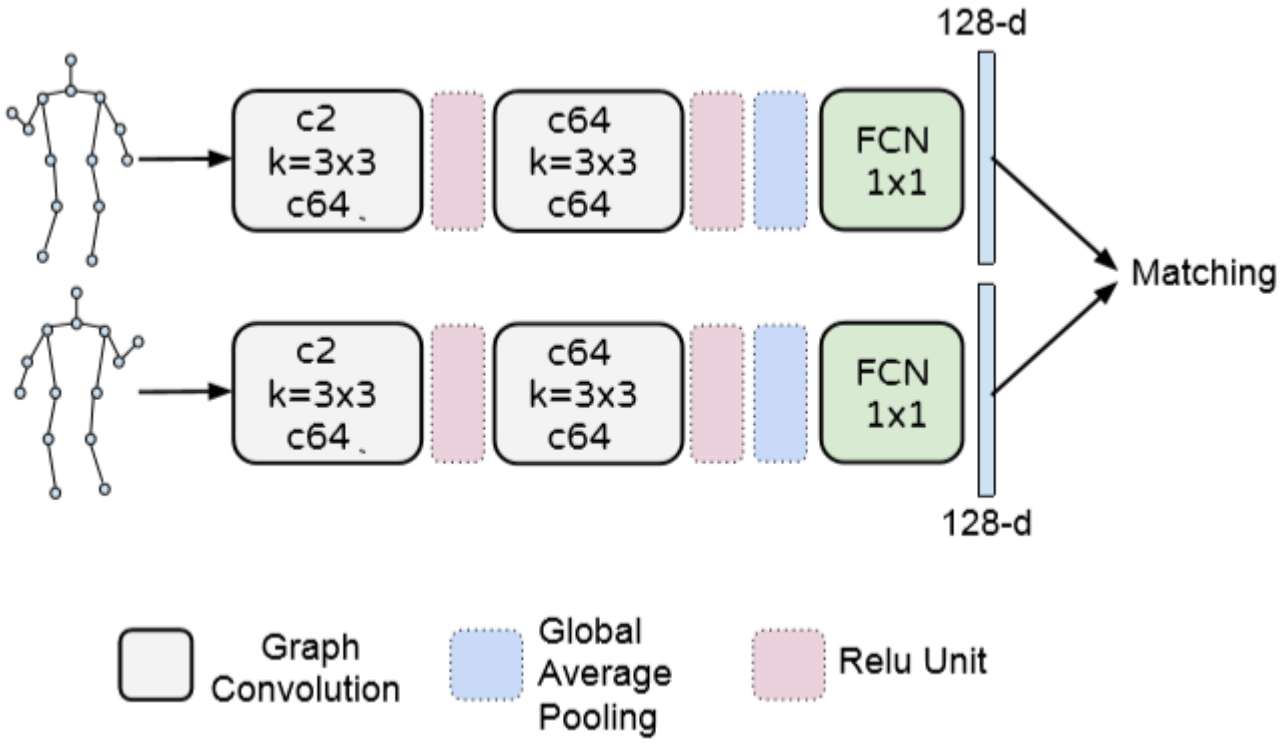
如果目标丢失，我们有两种模式：

- （1）固定关键帧间隔（FKI）模式。 忽略该目标直到预定的下一个关键帧，其中检测模块重新生成候选者，然后将他们的ID与跟踪历史相关联。
- （2）自适应关键帧间隔（AKI）模式。 通过候选检测和身份关联立即恢复缺失。

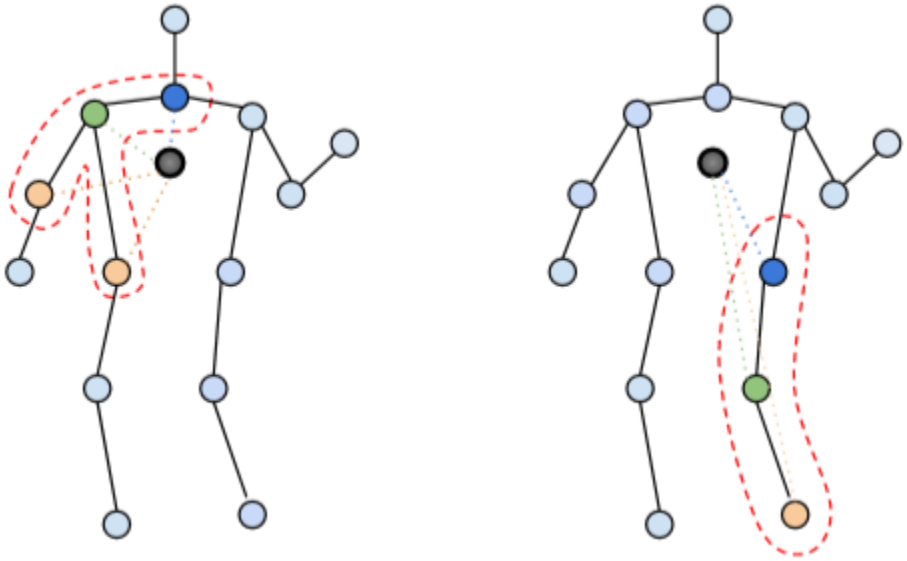
$$m(t_k, d_k) = \begin{cases} 1, & \text{if } o(t_k, \mathcal{D}_{i,k}) > \tau_o \\ 0, & \text{otherwise.} \end{cases}$$

Siamese Network

- 给定2D坐标形式的身体关节序列，我们构建一个空间图，其中关节为图形节点，人体结构中的连接为图边。我们的图卷积网络的输入是图节点上的联合坐标向量。



骨架的图形卷积



$$l_i(v_j) = \begin{cases} 0 & \text{if } r_j = r_i \\ 1 & \text{if } r_j < r_i \\ 2 & \text{if } r_j > r_i \end{cases}$$

result

	Method	Wrist-AP	Ankles-AP	mAP	MOTA	fps
	Posetrack 2017 Test Set					
Offline	PoseTrack, CVPR'18 [3]	54.3	49.2	59.4	48.4	-
Offline	BUTD, ICCV'17 [19]	52.9	42.6	59.1	50.6	-
Offline	Detect-and-track, CVPR'18 [12]	-	-	59.6	51.8	-
Offline	Flowtrack-152, ECCV'18 [36]	71.5	65.7	74.6	57.8	-
Offline	HRNet, CVPR'19[33]	72.0	67.0	74.9	57.9	-
Offline	**Ours-CPN101 (offline)**	**68.0 / 59.7**	**62.6 / 56.3**	**70.7 / 63.9**	**55.1**	-
Offline	**Ours-MSRA152 (offline)**	**68.9 / 61.8**	**63.2 / 58.4**	**71.5 / 65.7**	**57.0**	-
Offline	**Ours-manifold (offline)**	**- / 64.6**	**- / 58.4**	**- / 66.7**	**58.0**	-