1. Introduction

Problem: Multi-person pose estimation in videos

Challenges:
- Existing methods for multi-person pose estimation cannot be applied directly
- In addition to pose estimation, person association over time is also required
- Large pose and scale variation
- Varying number of persons or body parts due to occlusion or truncation

2. Contributions

- A novel method that jointly models multi-person pose estimation and tracking in a single formulation
- The method can implicitly handle occlusion and truncation of persons
- A novel and challenging dataset to quantitatively evaluate the performance
- A completely unconstrained evaluation protocol
- No assumptions about the scale, size and location or the number of persons

3. Pose Tracking

Joint Detection Candidates
- Generate a set of body joint detection hypotheses $D = \{D_f\}_{f \in F}$
- Every detection $d \in D$ at location $x_d \in \mathbb{R}^2$ in frame $f$ belong to joint type $j$
- We aim to identify joint hypotheses that belong to an individual person in the entire video

Spatial Edges
- Connect all joint candidates in frame $f$
  $E_s = \bigcup_{f \in F} E_s^f$ and $E_s' = \{(d, d') : d \neq d' \land d, d' \in D_f\}$
- Joint candidates are connected independent of the associated joint type $j$

Temporal Edges
- Connect joint hypotheses of the same joint type across frames
  $E_t = \{(d, d') : j = j' \land d \in D_f \land d' \in D_{f'} \land 1 \leq |f - f'| \leq 5 \land f, f' \in F\}$
- TEMPoral relations up to $|f - f'| = 5$ frames are considered to handle short-term occlusion and missing detections

Spatio-Temporal Edges
- The spatial and temporal edges are connected to each other

Graph Partitioning (Integer Linear Programming)

- $G = (V, E)$, $B = E_s \cup E_t$ and $D = (D_f)_{f \in F}$
- Partition the graph by removing edges and nodes such that each partition corresponds to a tracked pose of an individual person

Potentials

- **Unaries**: confidence of detected joint candidates $p_d$ obtained using DeeperCut [1]

- **Spatio binaries**: $\psi_s(d_j, d'_j)$
  - Case 1: same joint type
  - Case 2: different joint type

- **Temporal binaries**: $\psi_t(d_j, d'_j)$

Constraints

- **Consistency**
  - Spatial
  - Temporal

- **Transitivity**
  - Spatial
  - Temporal

- **Spatio-temporal**
  - Spatial
  - Temporal