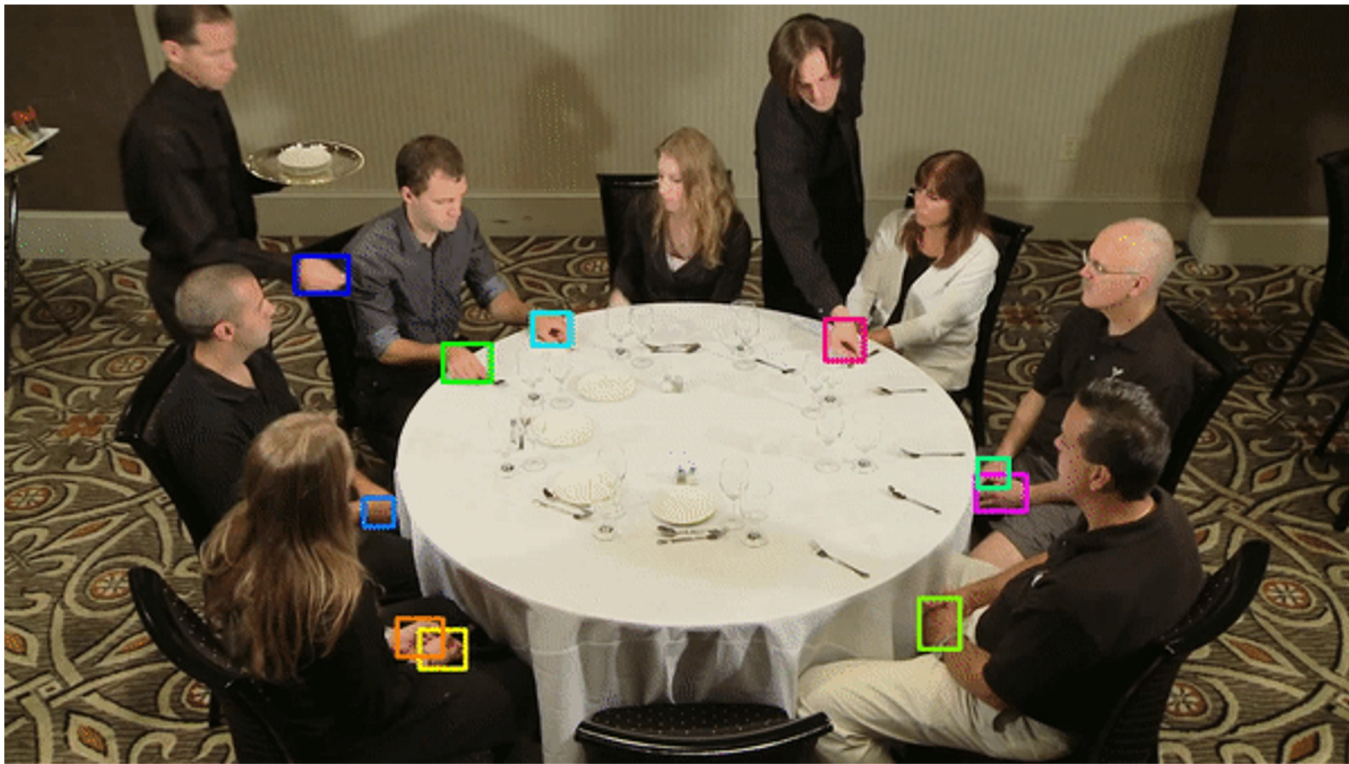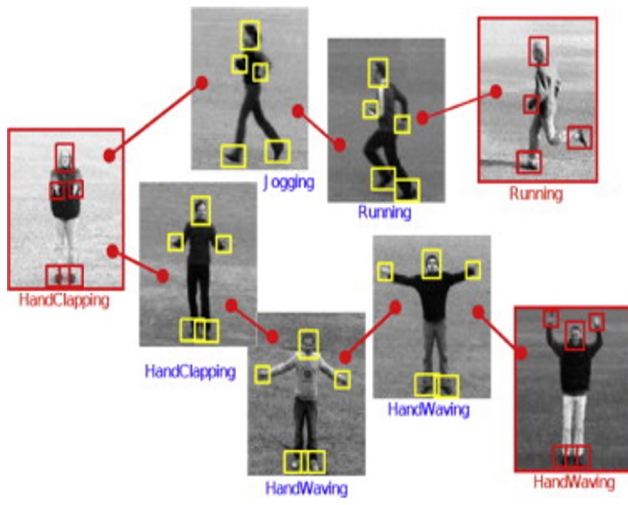# Forward Propagation, Backward Regression, and Pose Association for Hand Tracking in the Wild

Mingzhen Huang, Supreeth Narasimhaswamy, Saif Vazir,
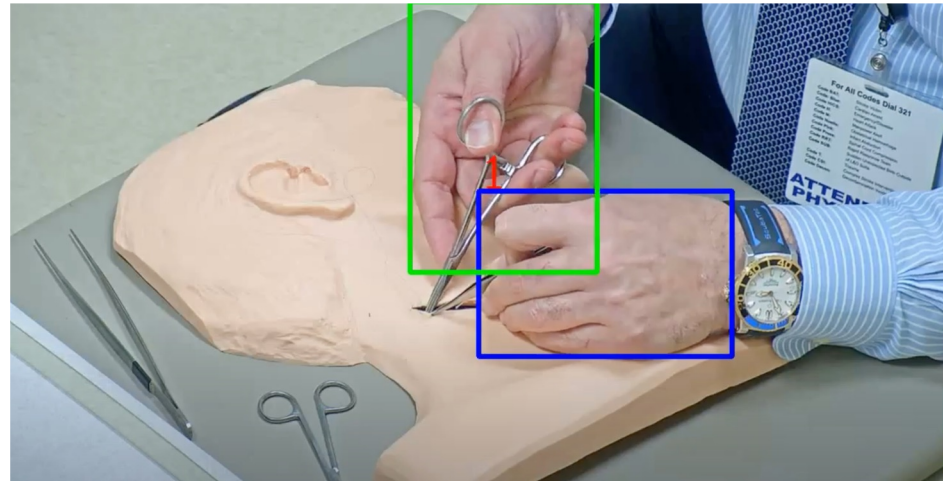Haibin Ling, Minh Hoai

# Goal: Tracking Hands

# Why is this task important?



Activity recognition



Skill evaluation

# Challenges for Hand Tracking



Fast motion, easy to be occluded

# Challenges for Hand Tracking



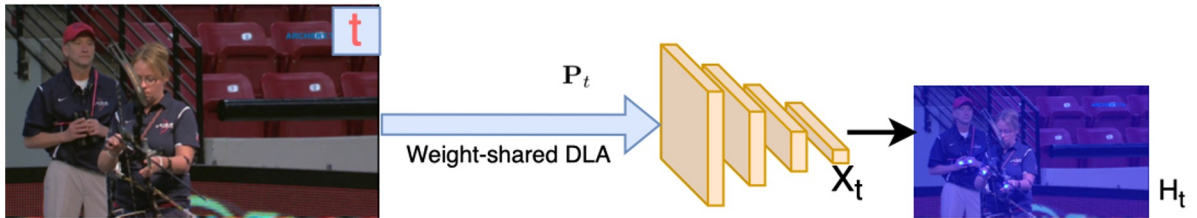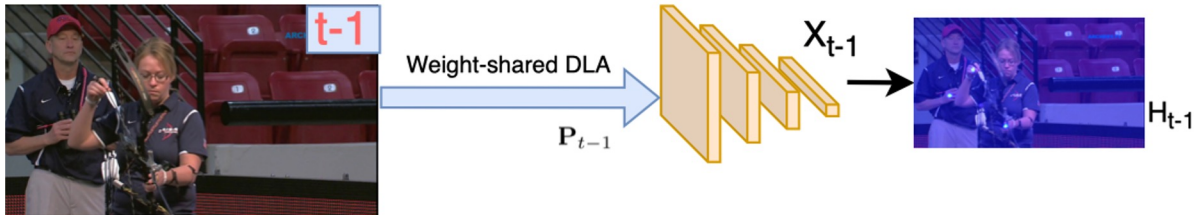Fast motion, easy to be occluded
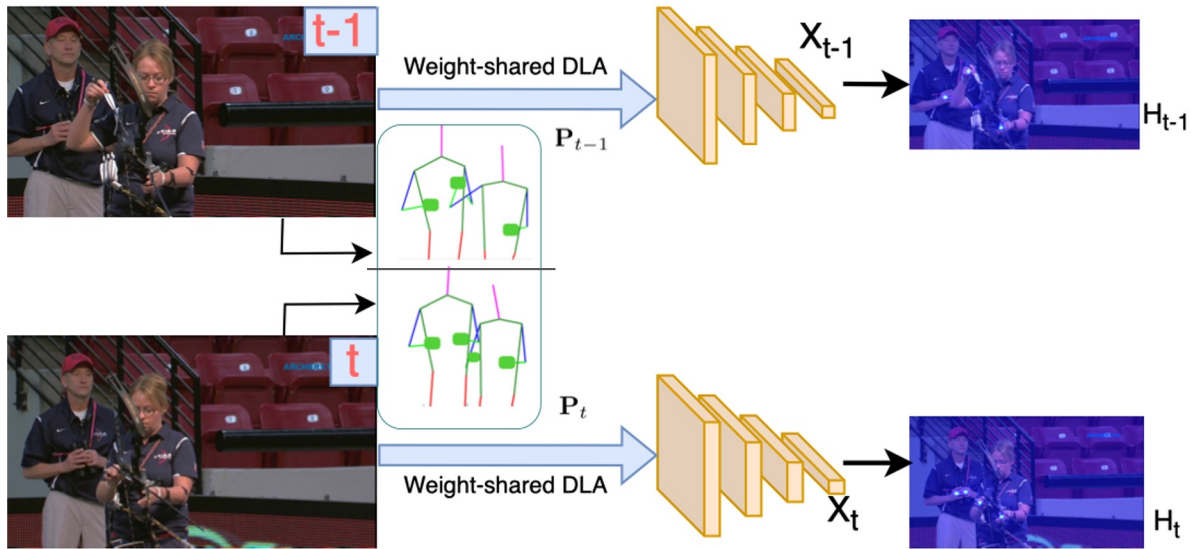
Motion blur and appearance change

# Contributions

- An online method that
    - detects and tracks hand identities across frames
    - Links hands to body pose for long-term tracking
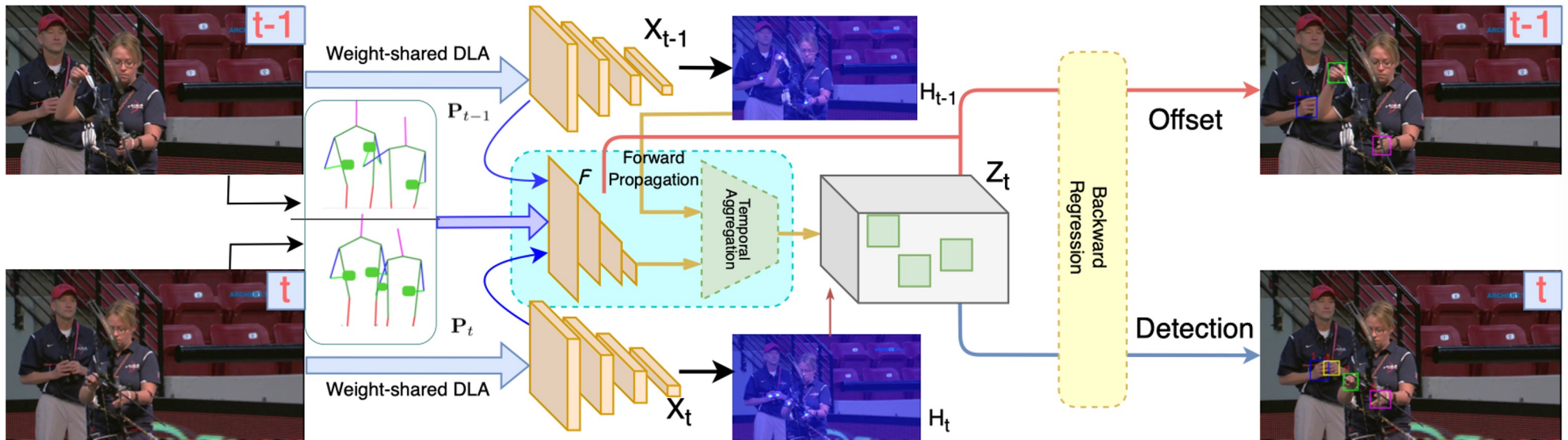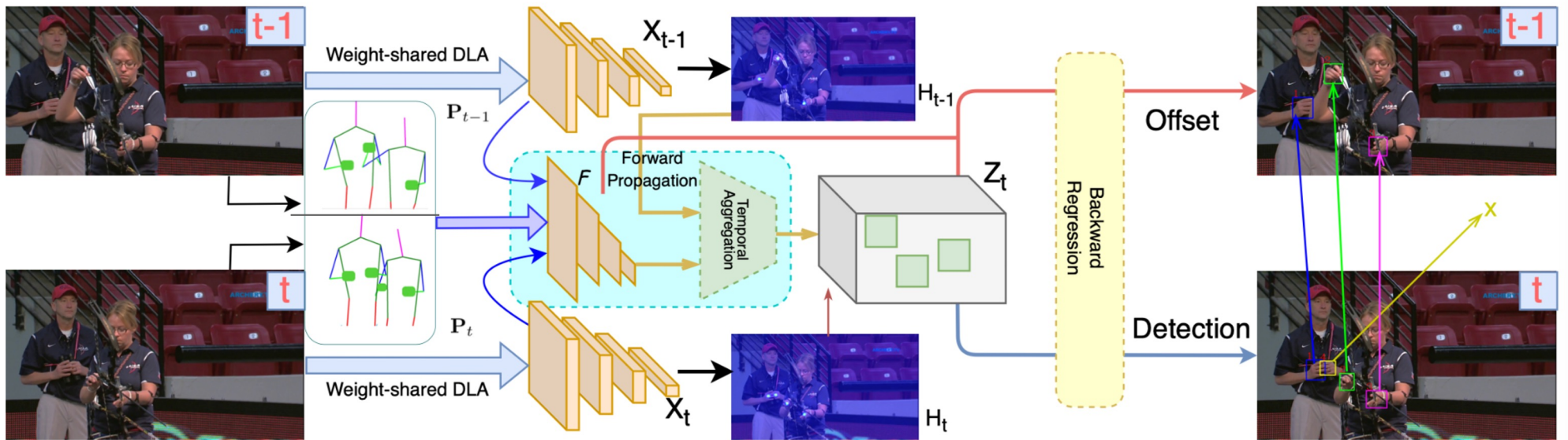- A large-scale hand tracking dataset

# Methods

# Methods

# Methods
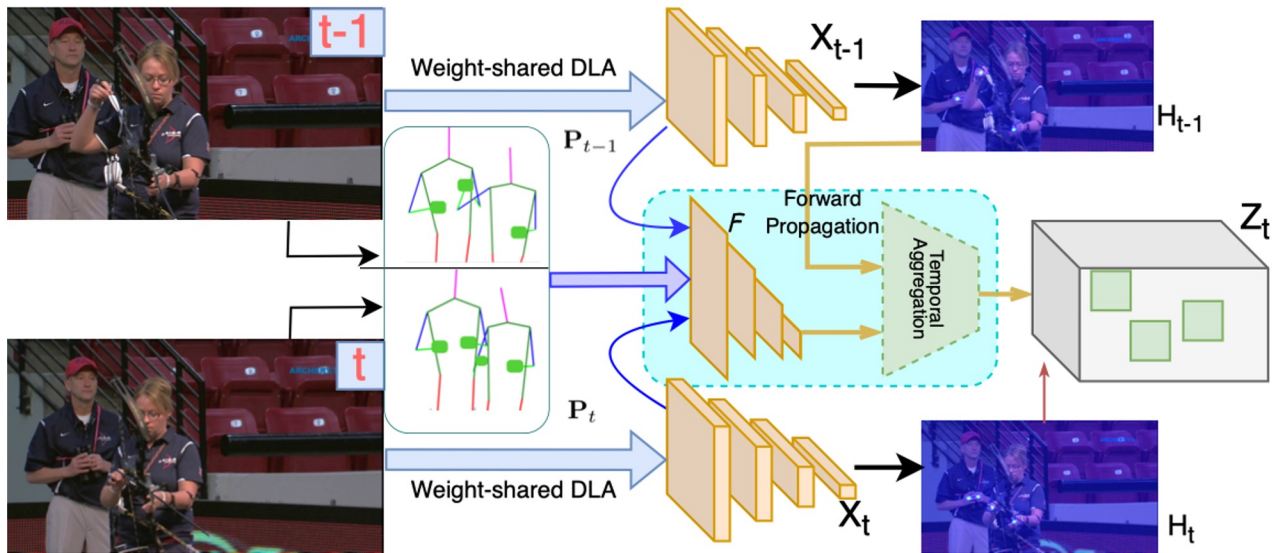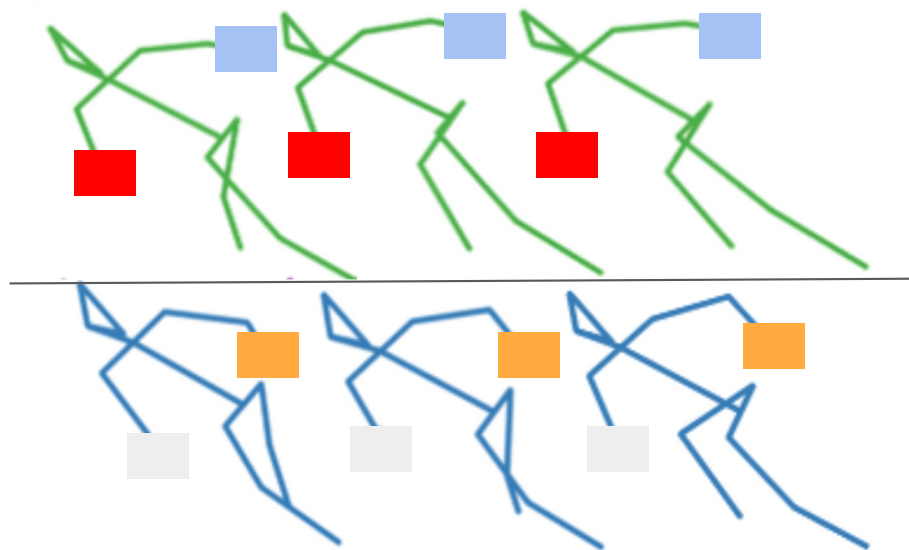
# Methods

# Methods

# Methods



Hand bounding box

# Youtube-Hand Dataset
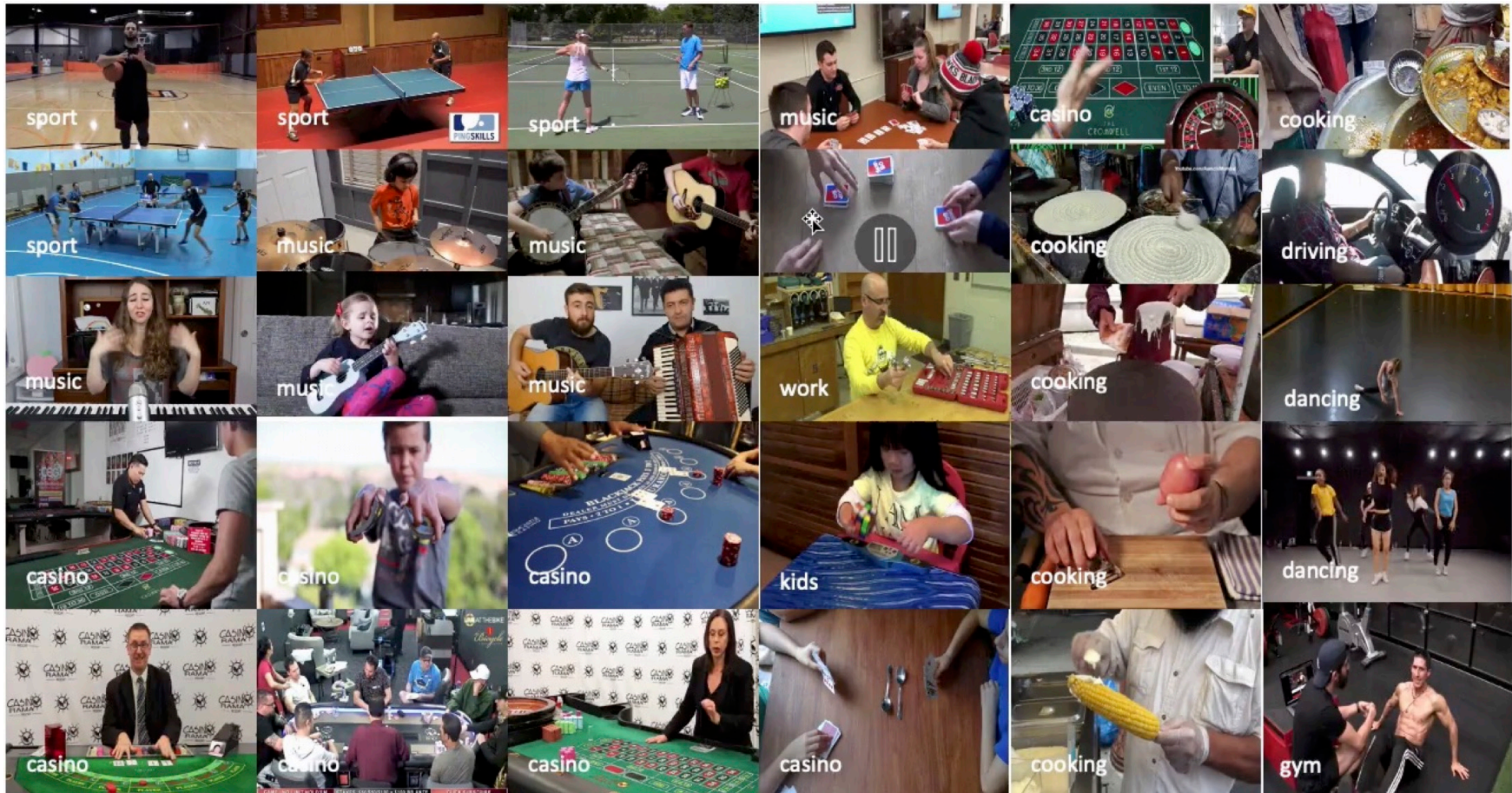
# Youtube-Hand Dataset

| | Total |
|---|---|
| #Videos | 240 |
| #Frames | 232K |
| #Anno. hands | 60K |
| #Trajectories | 864 |

# Youtube-Hand Dataset

| | Total |
|---|---|
| #Videos | 240 |
| #Frames | 232K |
| #Anno. hands | 60K |
| #Trajectories | 864 |
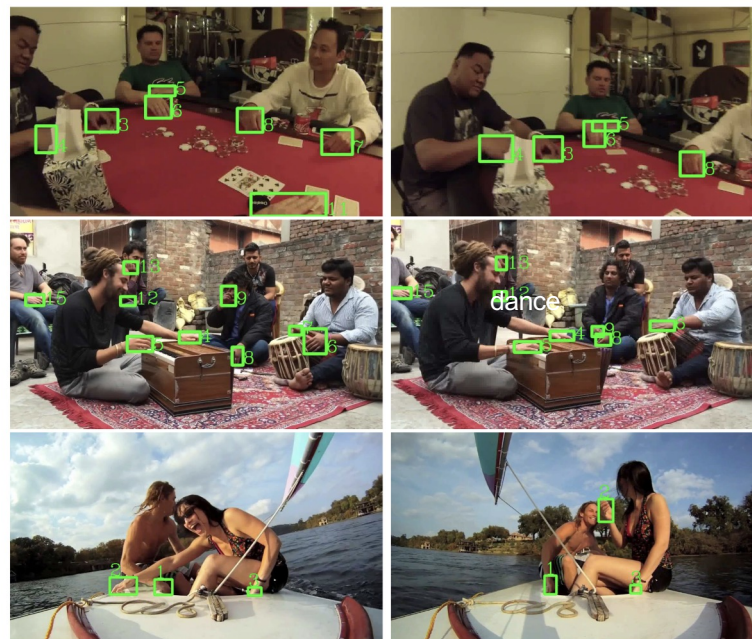
# Youtube-Hand Dataset

| Dataset | Scene/camera Constraints | Has Video | #Hand Trajs. | Maximum #trajs/video |
|---|---|:---:|:---:|:---:|
| EgoHands [1] | Google glasses | | 0 | n/a |
| Handseg [18] | Color gloves | | 0 | n/a |
| NYUHands [42] | Hands keypoints | | 0 | n/a |
| ColorHandPose [59] | 3D hands keypoints | | 0 | n/a |
| HandNet [45] | Fingertips | | 0 | n/a |
| GANeratedHands [22] | Synthetic | | 0 | n/a |
| Oxford-Hand [20] | Unconstrained | | 0 | n/a |
| TV-Hand [24] | Unconstrained | | 0 | n/a |
| COCO-Hand [24] | Unconstrained | | 0 | n/a |
| Contact-Hand [25] | Unconstrained | | 0 | n/a |
| 100DOH [36] | Unconstrained | ✓ | 0 | n/a |
| GTEA [15] | Ego-centric | ✓ | 0 | n/a |
| WorkingHands [39] | Down-facing cam. | ✓ | 0 | n/a |
| BSL [28] | TV show, segmentation | ✓ | 2 | 2 |
| SynthHands [21] | Ego-centric | ✓ | 1 | 1 |
| ICP-PSO [31] | Hand keypoints | ✓ | 6 | 1 |
| EpicKitchen [6] | Ego-centric, auto-label | ✓ | 1400 | 2 |
| VIVA [33] | Vehicle-mounted | ✓ | 45 | 4 |
| YouTube-Hand | Unconstrained | ✓ | 864 | 15 |

| | Total |
|---|:---:|
| #Videos | 240 |
| #Frames | 232K |
| #Anno. hands | 60K |
| #Trajectories | 864 |
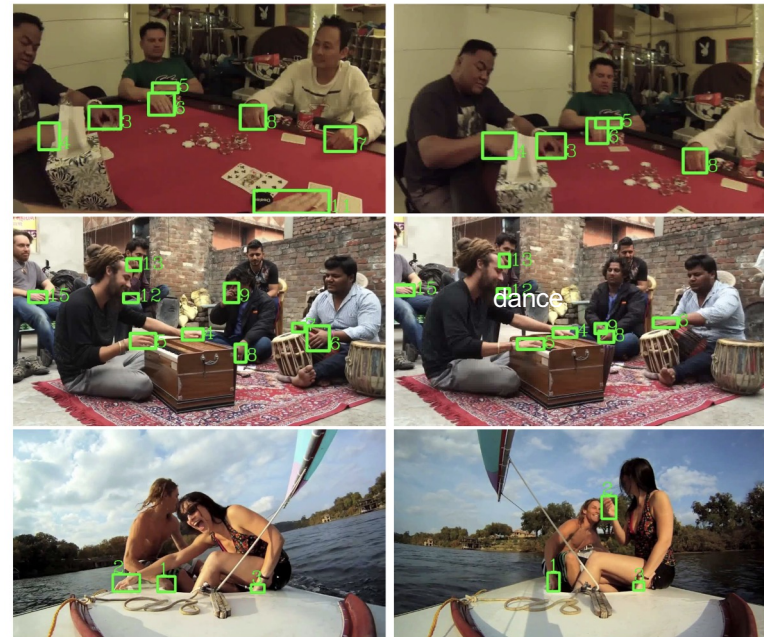
# Quantitative Results

| Methods | MOTA↑ | LTR↑ |
|---|---|---|
| LightTrack [29] (Pose) | 30.8 | 48.4 |
| FairMot [59] | 39.9 | 31.3 |
| MPNTrack [5] (Offline) | 40.0 | 44.3 |
| CenterTrack[62] | 40.7 | 27.3 |
| CenterTrack*[62] | 50.0 | 37.5 |
| SORT [4] | 44.9 | 47.6 |
| TraDeS [51] | 52.7 | 44.4 |
| HandLer (proposed) | **70.0** | **64.3** |

LTR = avg(longest tracklet length / total length)

# Quantitative Results

| Methods | MOTA↑ | LTR↑ | IDF1↑ | MT↑ | ML↓ | FP↓ | FN↓ | IDs↓ | MOTP↑ |
|---|---|---|---|---|---|---|---|---|---|
| LightTrack [29] (Pose) | 30.8 | 48.4 | 53.4 | 101 | 70 | 6240 | 12816 | 1955 | 74.5 |
| FairMot [59] | 39.9 | 31.3 | 41.4 | 96 | 57 | 2065 | 12753 | 3448 | 76.8 |
| MPNTrack [5] (Offline) | 40.0 | 44.3 | 49.0 | 156 | 66 | 5918 | 11263 | 1039 | 77.0 |
| CenterTrack[62] | 40.7 | 27.3 | 37.2 | 113 | 62 | 2279 | 12379 | 3362 | 76.5 |
| CenterTrack*[62] | 50.0 | 37.5 | 57.8 | 137 | 43 | 3208 | 10317 | 1647 | 79.0 |
| SORT [4] | 44.9 | 47.6 | 48.3 | 101 | 72 | 2295 | 12960 | 1475 | 76.7 |
| TraDeS [51] | 52.7 | 44.4 | 53.6 | 168 | 43 | 3271 | 9102 | 1982 | 76.4 |
| HandLer (proposed) | 70.0 | 64.3 | 70.9 | 218 | 23 | 2412 | 5986 | 712 | 79.9 |

# Qualitative Result - 1
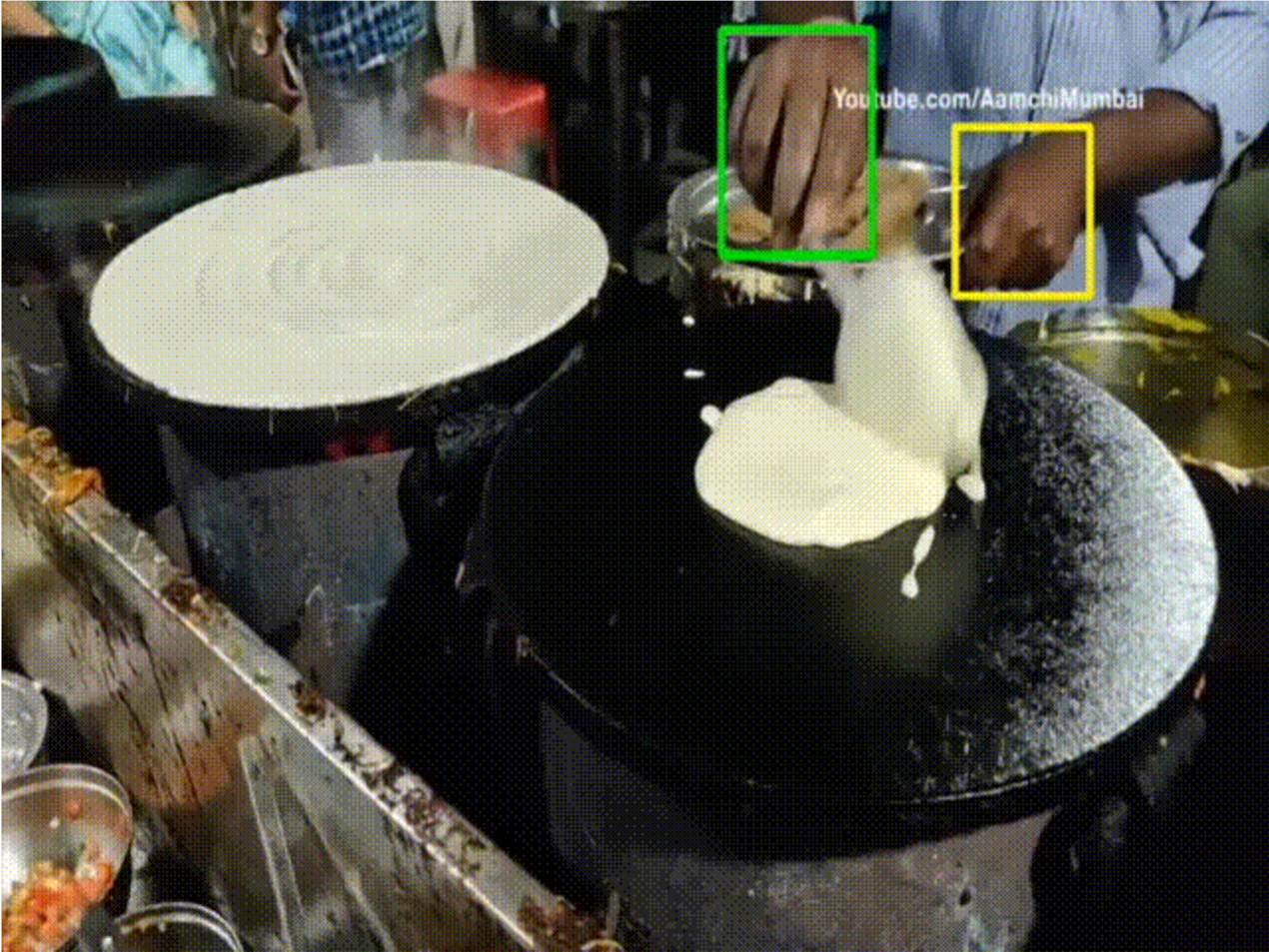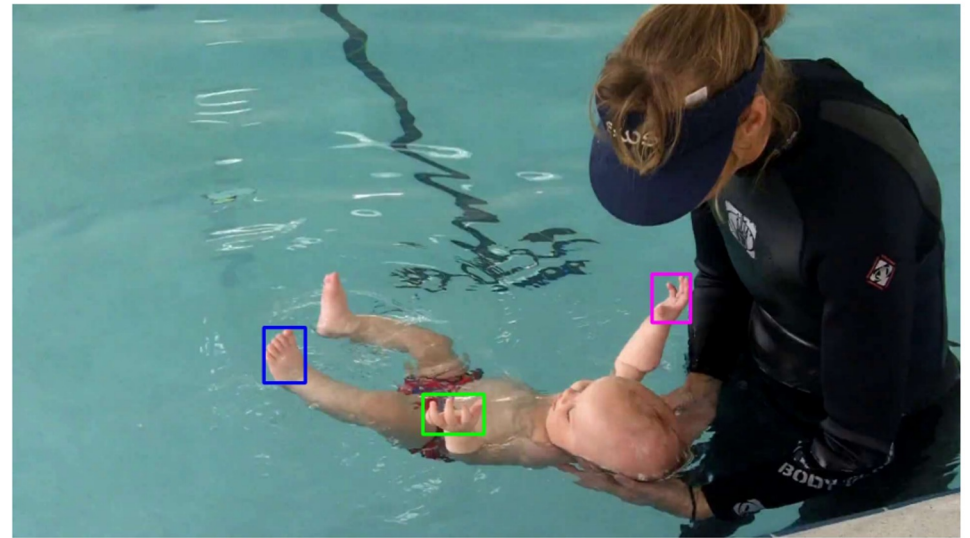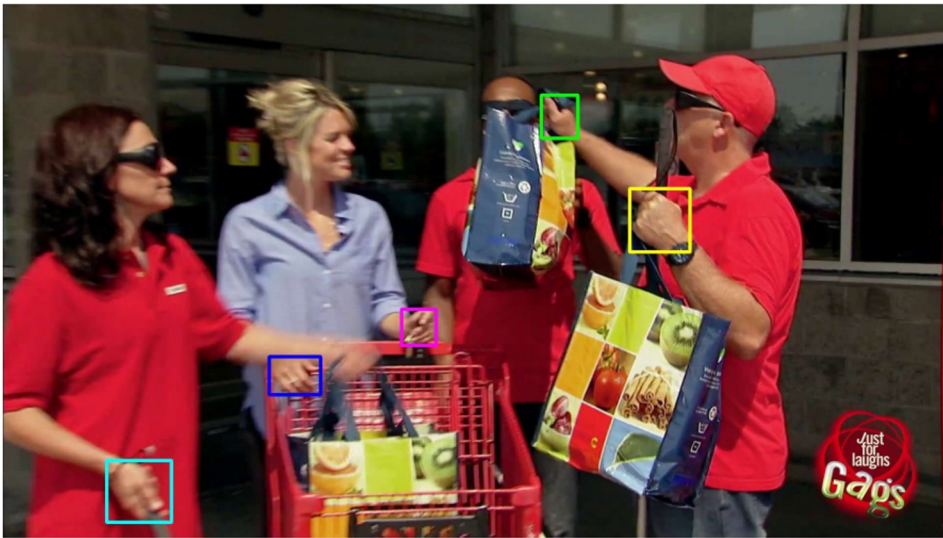
# Qualitative Result - 2
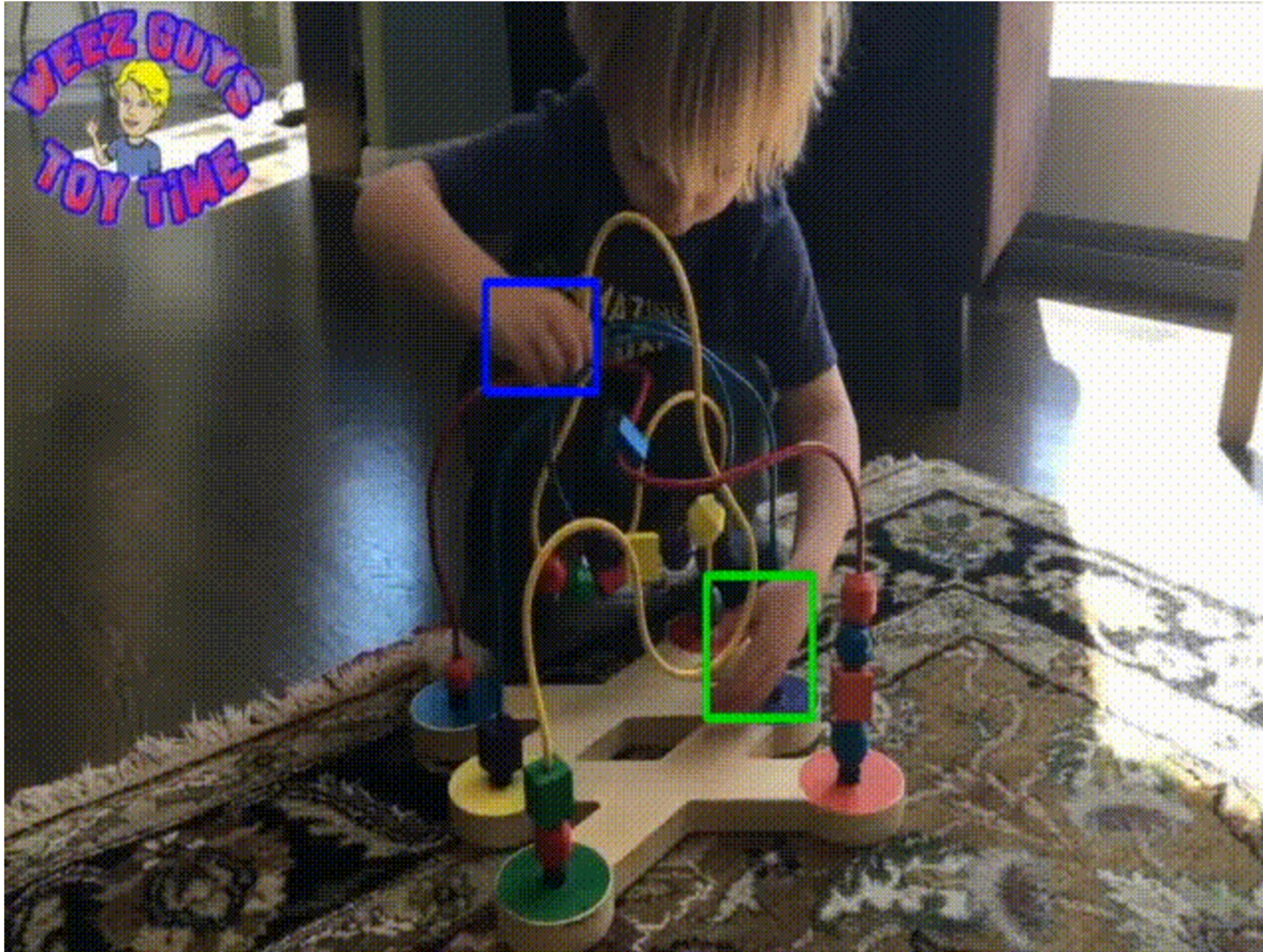
# Qualitative Result - 3
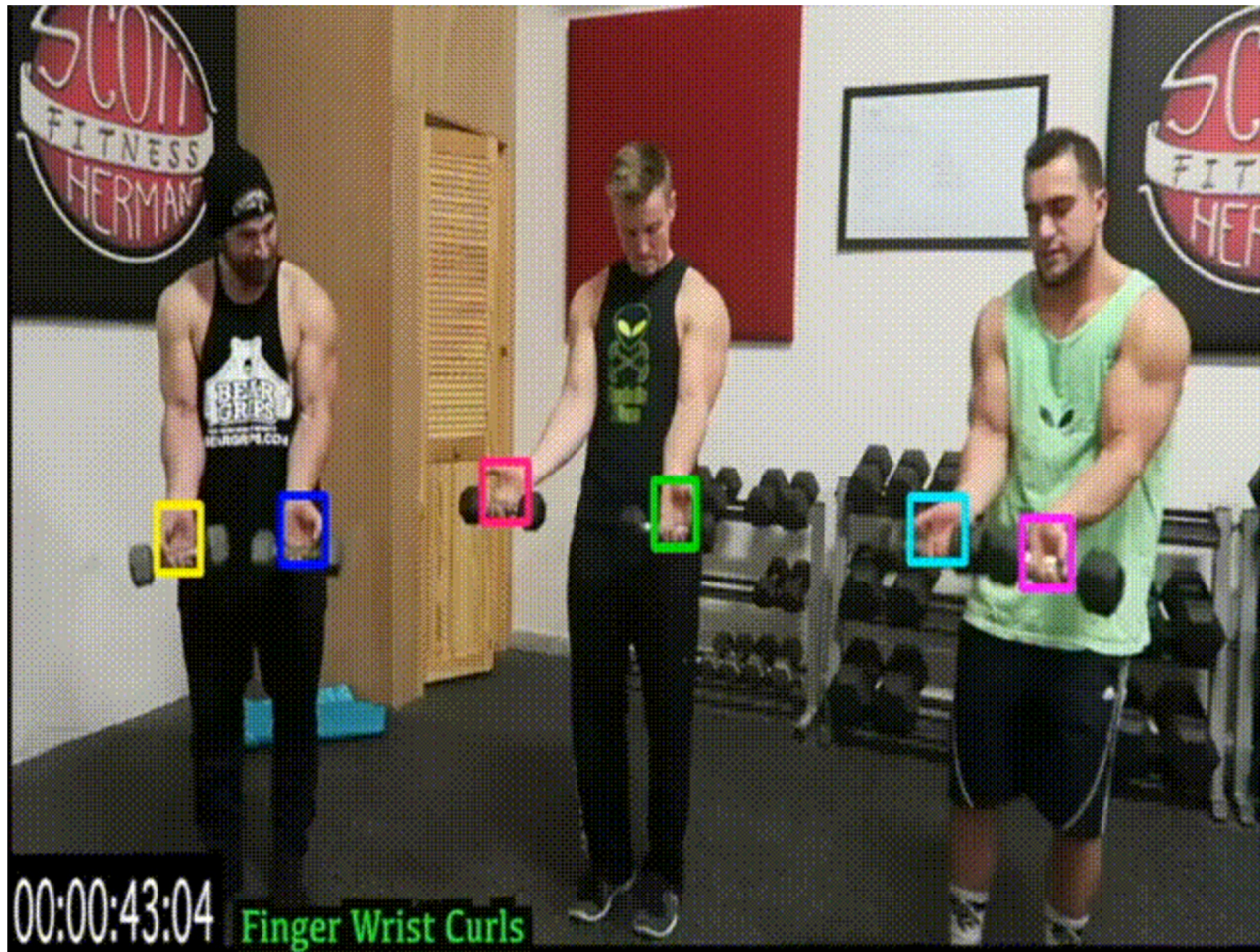
# Failure Case

# Tracking Occluded Hands



HandCNN                                    HandLer
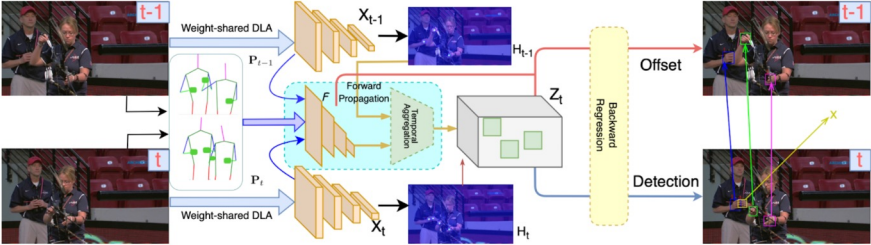
|                | Dataset        |                   |
|----------------|----------------|-------------------|
| Method         | YouTube-Hand   | Blur&Occ Split    |
| HandCNN [26]   | 72.4           | 62.8(13.1% ↓)     |
| HandLer        | **84.1**       | **76.7**(8.8% ↓)  |

Detecting heavy occluded hands

# Takeaways



HandLer for hand tracking

Youtube-Hand
Dataset



Project Page