



DeeperAction@ICCV 2021

Localized and Detailed Understanding of Human Actions in Videos

Organizers: Limin Wang, Yali Wang, Jing Shao, Wei Wu, Junjie Yan, Yu Qiao

11th October, 2021



Deeper Action Organizers

ICCV 2021
Workshop

Limin Wang

Nanjing University



Yali Wang

SIAT, CAS
Shanghai AI Lab



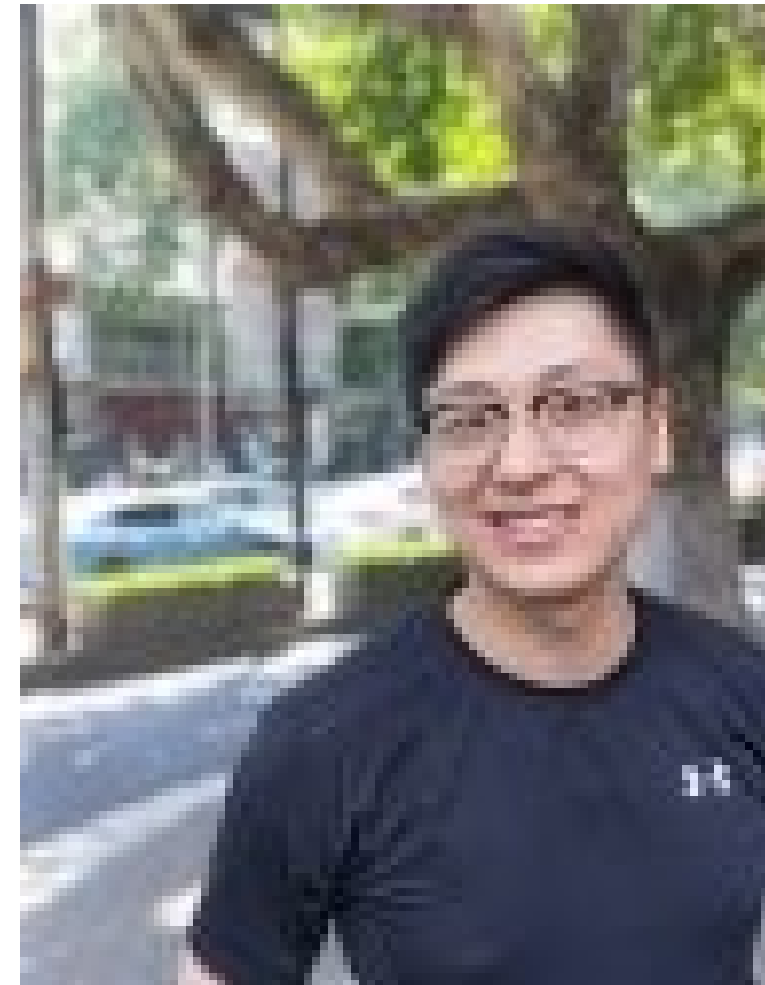
Jing Shao

SenseTime



Wei Wu

SenseTime



Junjie Yan

SenseTime



Yu Qiao

SIAT, CAS
Shanghai AI Lab



Rahul Sukthankar

Google Research



Luc Van Gool

CVL ETHZ, KU Leuven



Sponsors



AMAX Information Technologies (Suzhou) CO.,LTD.

14:00-14:10	Opening Remarks
14:10-14:40	Invited Talk 1 (Chen Sun): Detailed Human Action Understanding from Unlabeled Videos
15:40-15:00	Introduction to Track 1 - FineAction (Temporal Action Localization)
15:00-15:20	Winner Talk for Track 1 (Dahua Technology)
15:20-15:50	Invited Talk 2 (Gül Varol): Beyond Action Recognition: Detailed Video Modeling
15:50-16:10	Introduction to Track 2 - MultiSports (Spatio-Temporal Action Detection)
16:10-16:30	Winner Talk for Track (Huawei Noah's Ark Lab & USTC)
16:30-16:50	Introduction to Track 3 - Kinetics-TPS (Part-level Action Parsing)
16:50-17:10	Winner Talk for Track 3 (Hikvision Research Institute)
17:10-17:20	Closing and Awards

Previous Datasets - Action Recognition

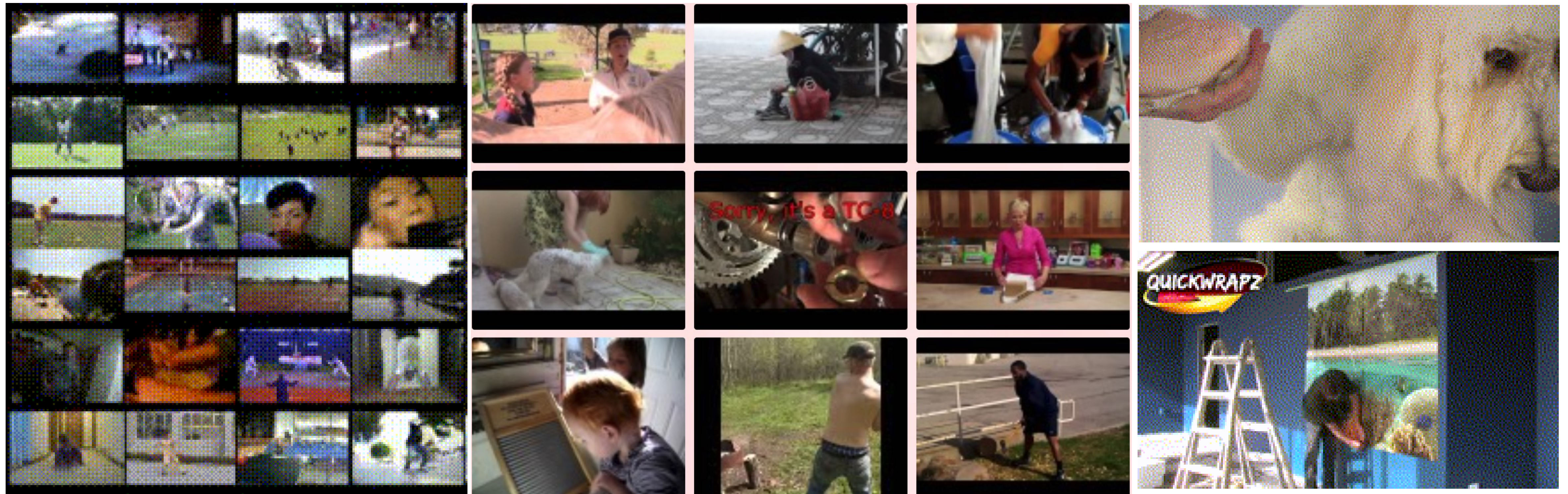
UCF101, HMDB51, Kinetics: rely on appearance, no fine-grained actions

Something-Something, Diving48: fine-grained, but no localization



Previous Datasets - Temporal Action Localization

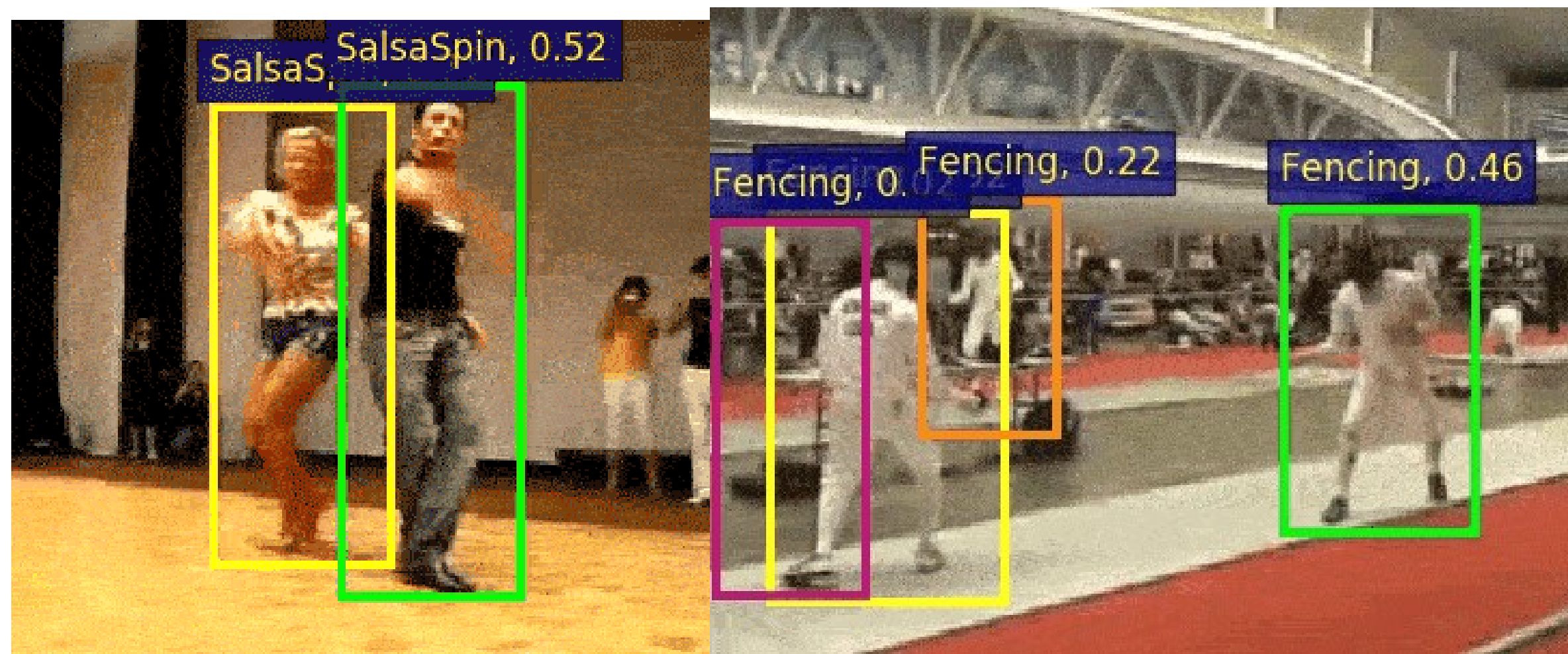
THUMOS, ActivityNet, HACs: no fine-grained actions, also rely on appearances



Previous Datasets - Spatio-Temporal Action Localization

UCF101-24, JHMDB: small scale, no fine-grained actions and only one action category per video

AVA: large scale but no accurate temporal localization (3s per annotation) and no fine-grained actions



Left: Sit, Talk to, Watch; Right: Crouch/Kneel, Listen to, Watch

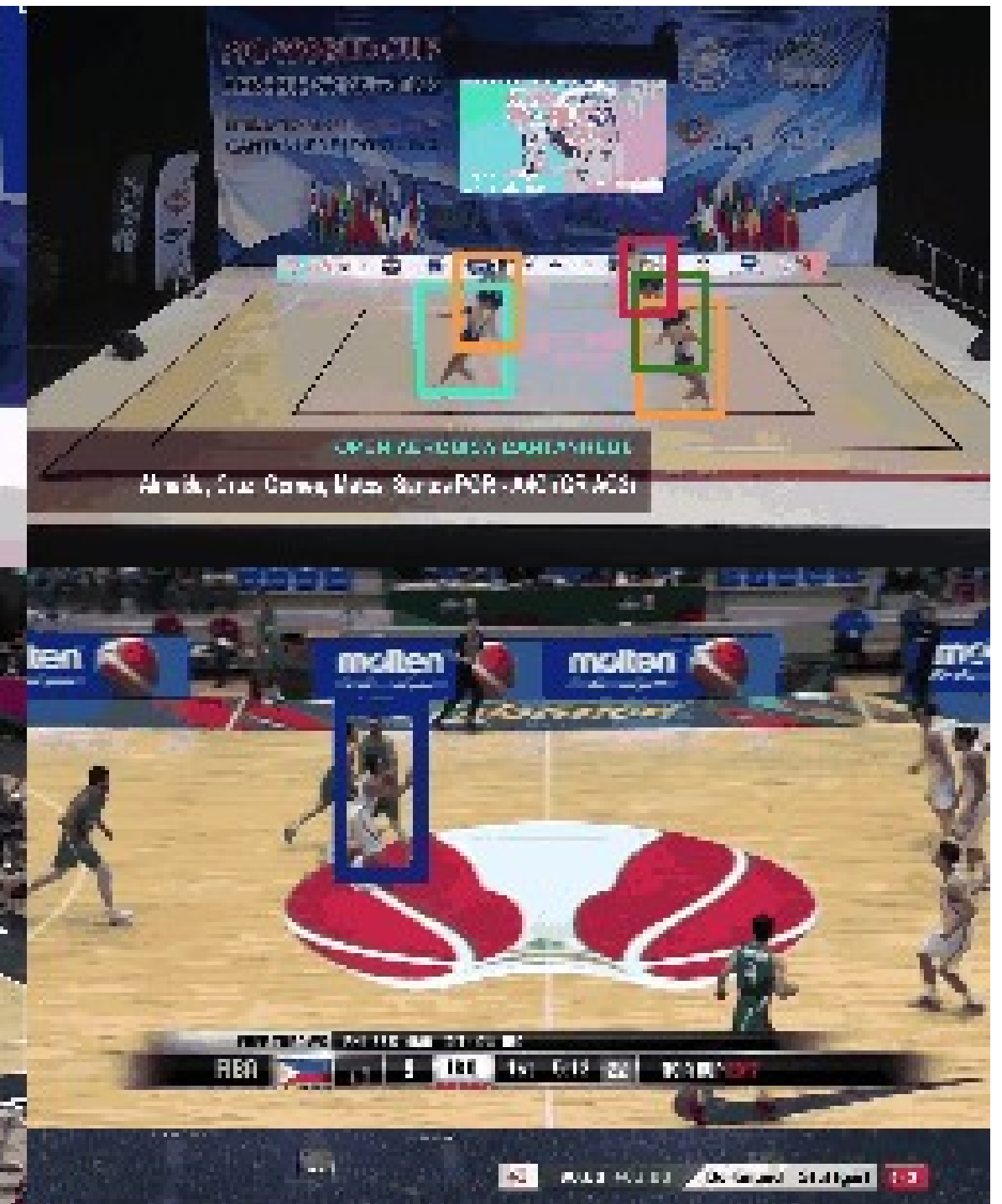
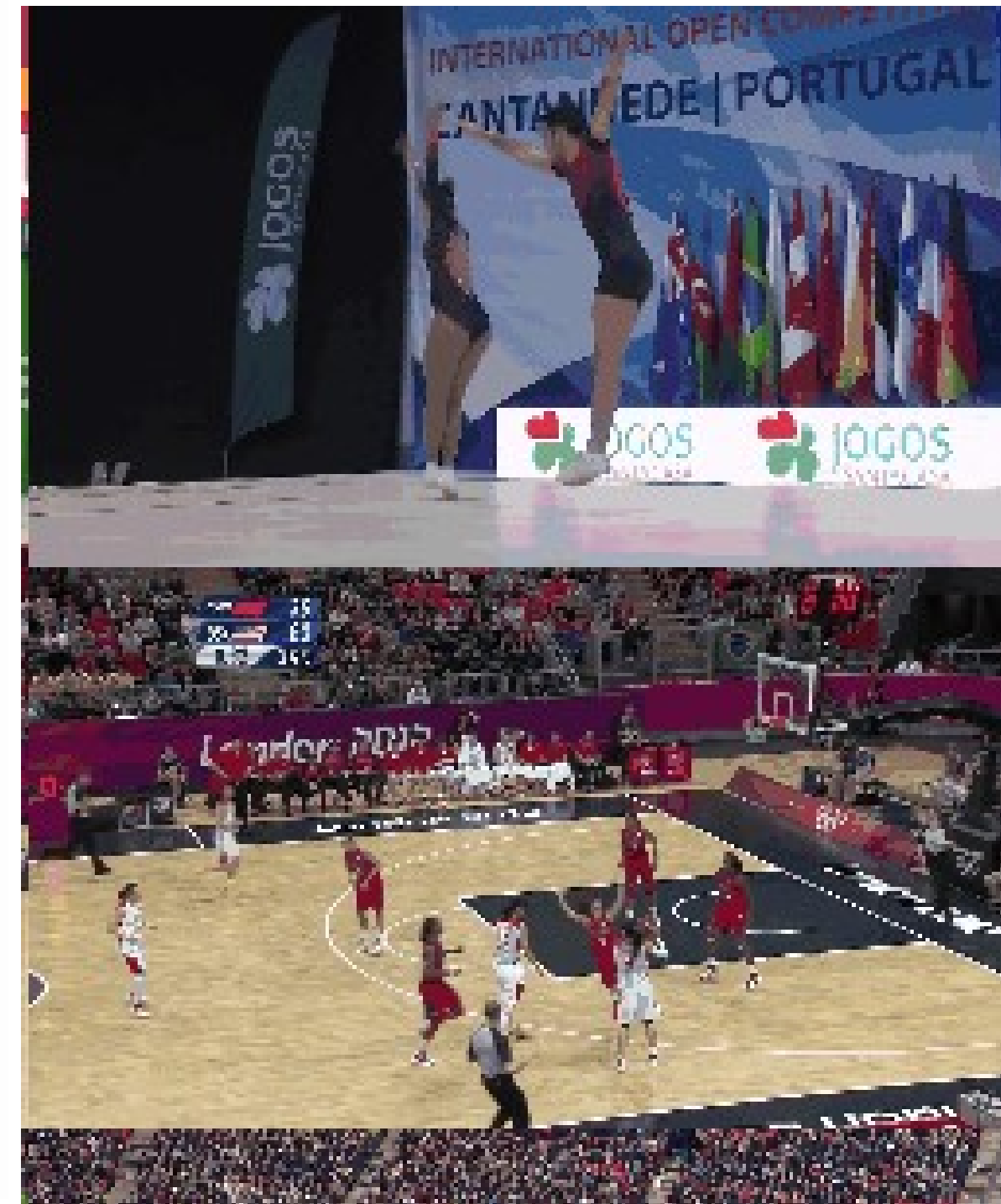
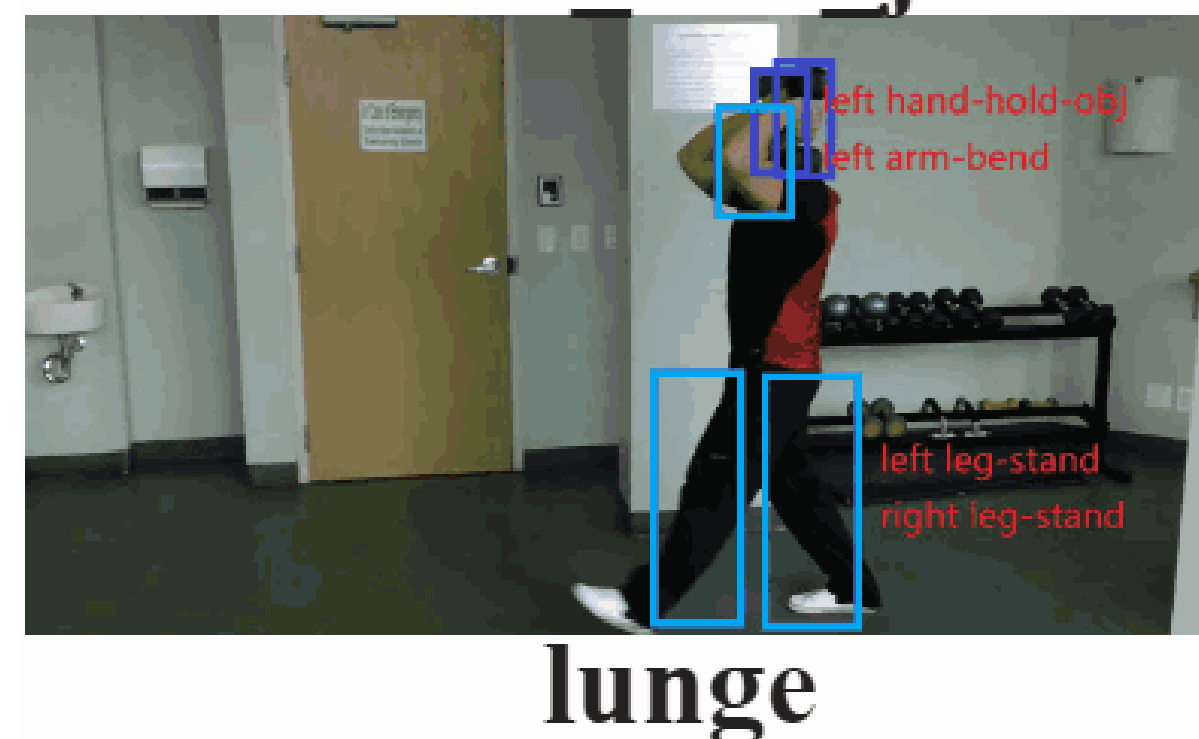
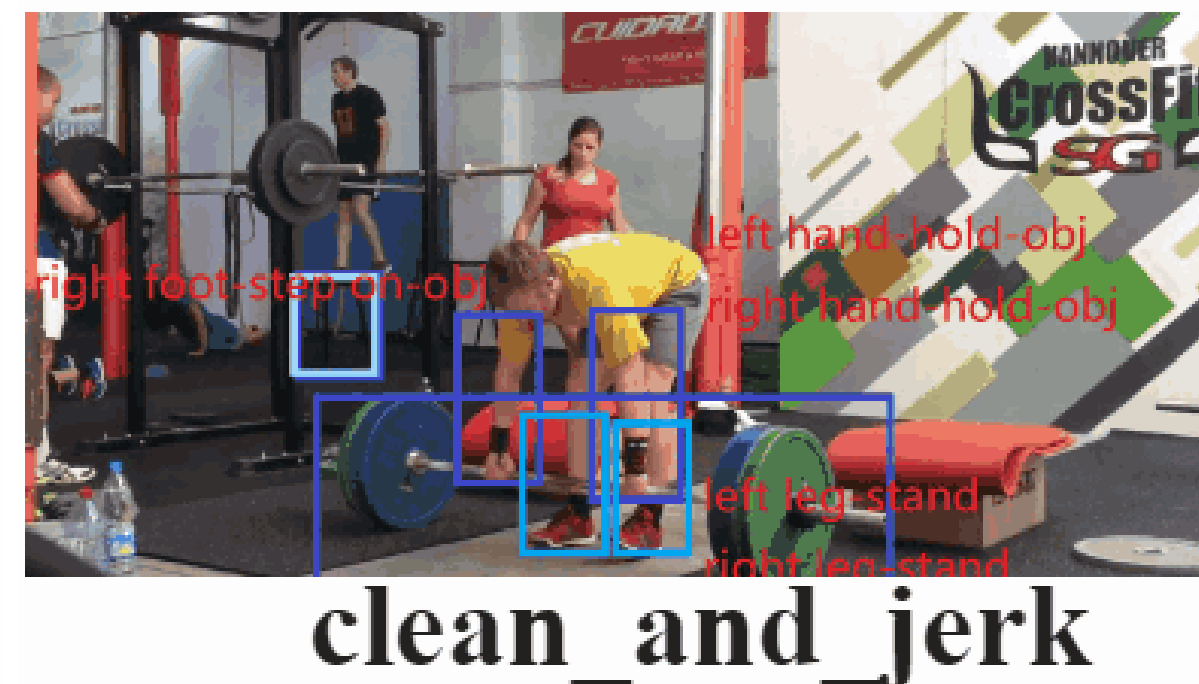
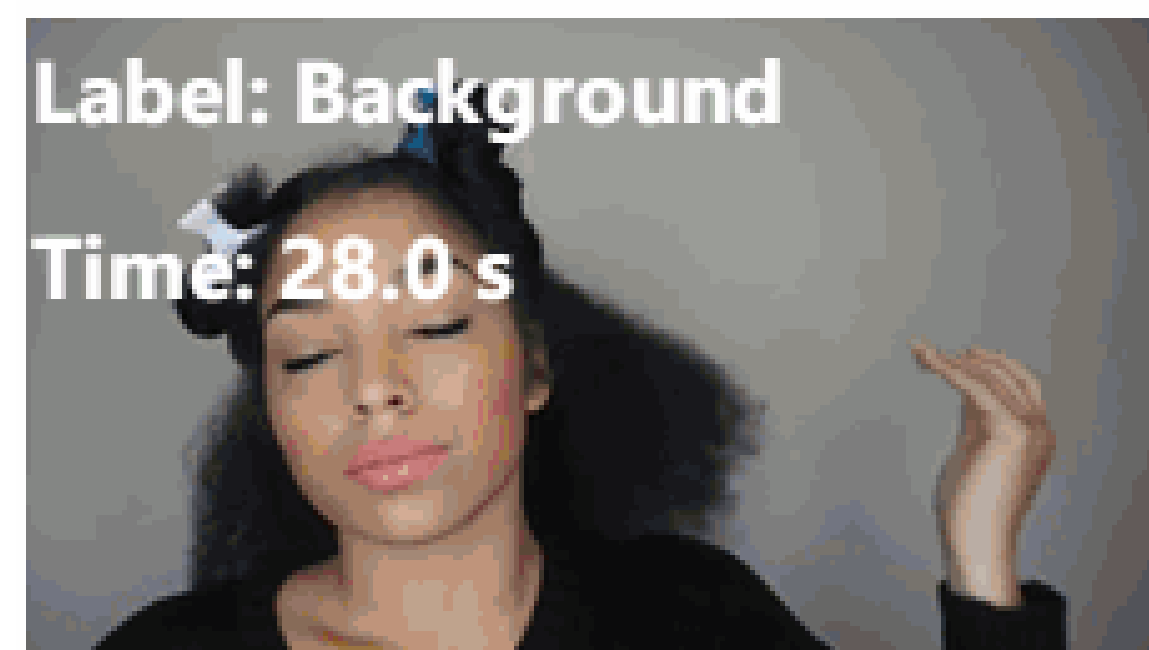
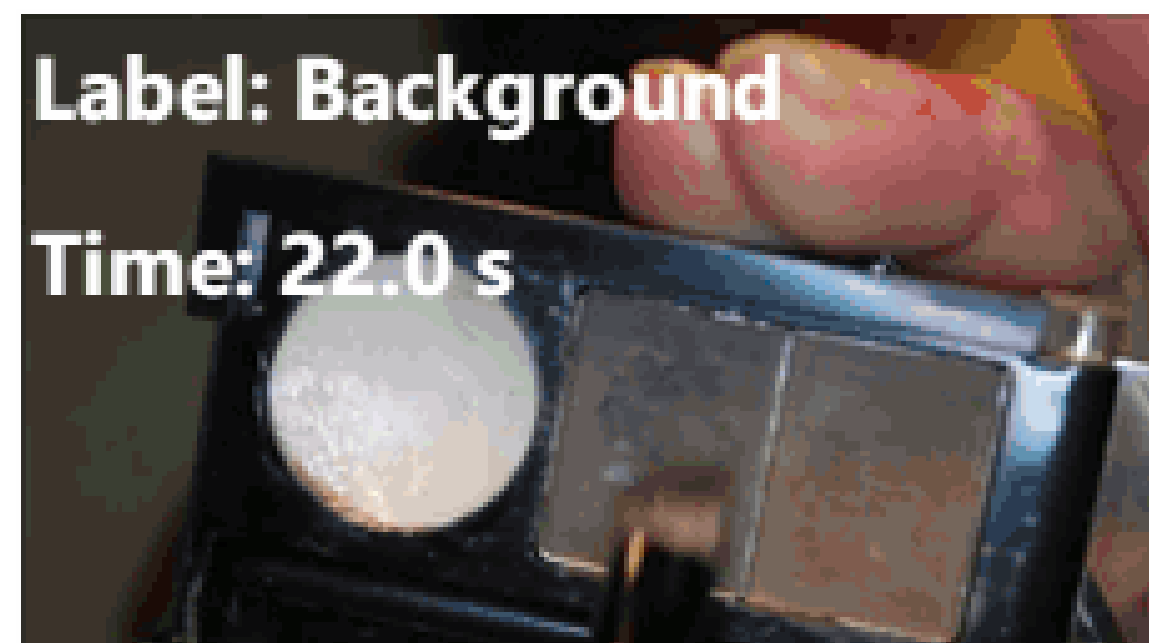


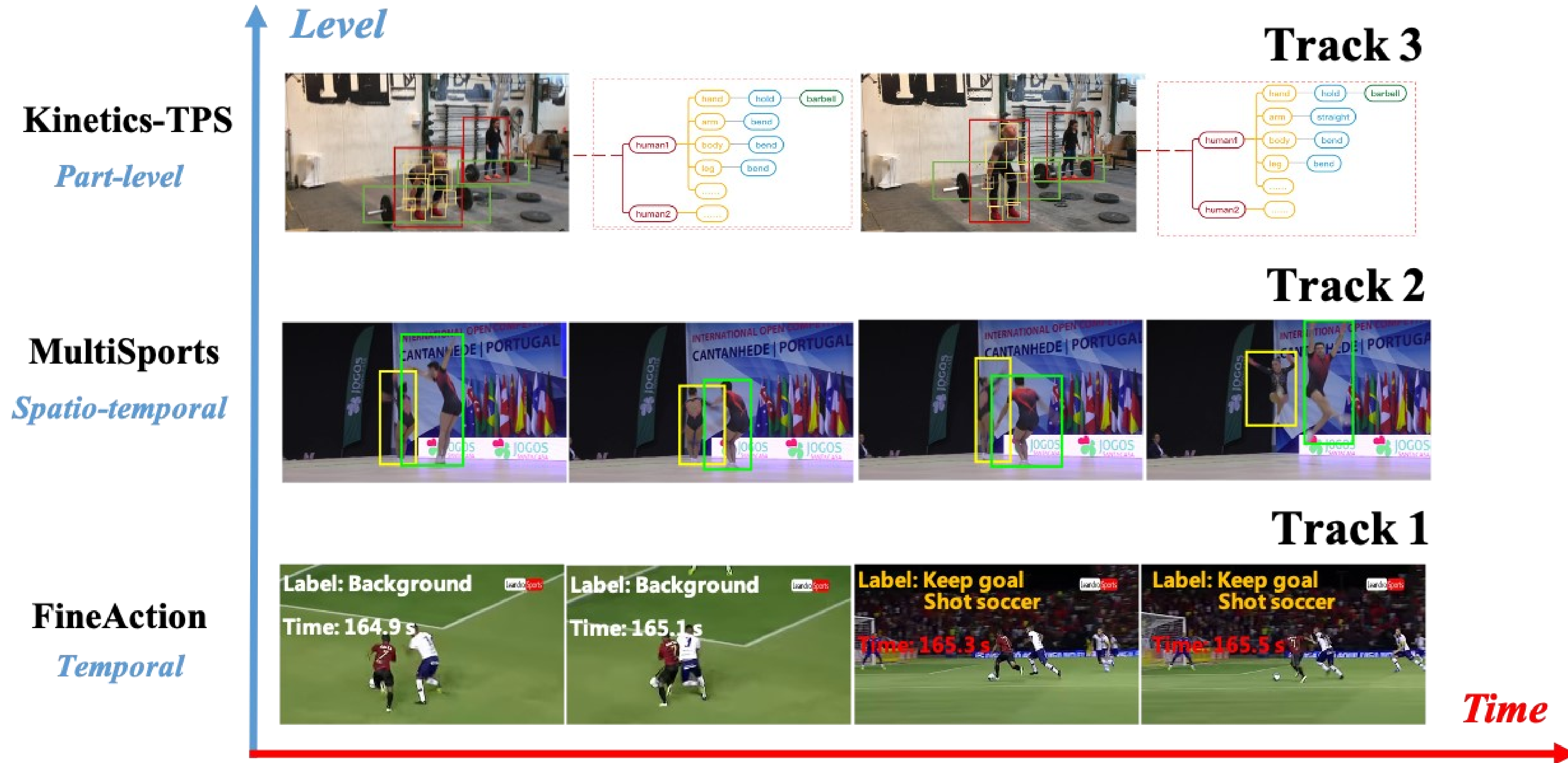
Left: Stand, Carry/Hold, Listen to; Middle: Stand, Carry/Hold, Talk to; Right: Sit, Write

Our Datasets - Localized and Detailed Understanding of Actions

Detailed: fine-grained classification of actions

Localized: temporal, spatio-temporal or part-level localization

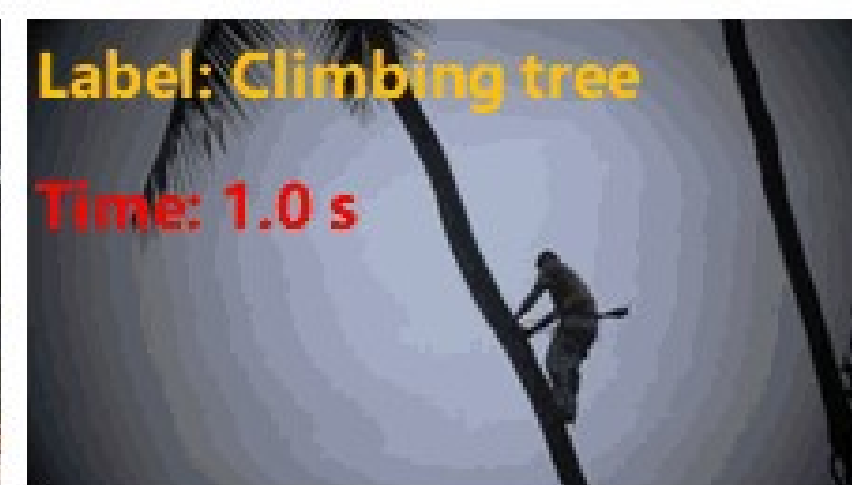
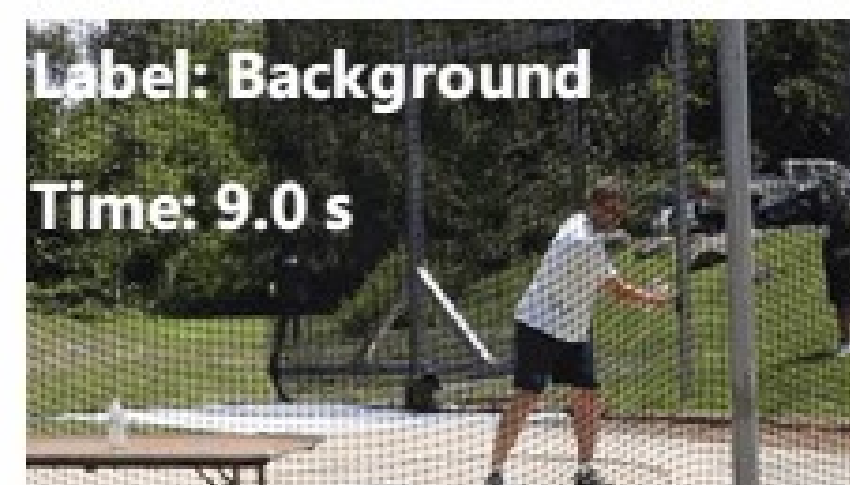
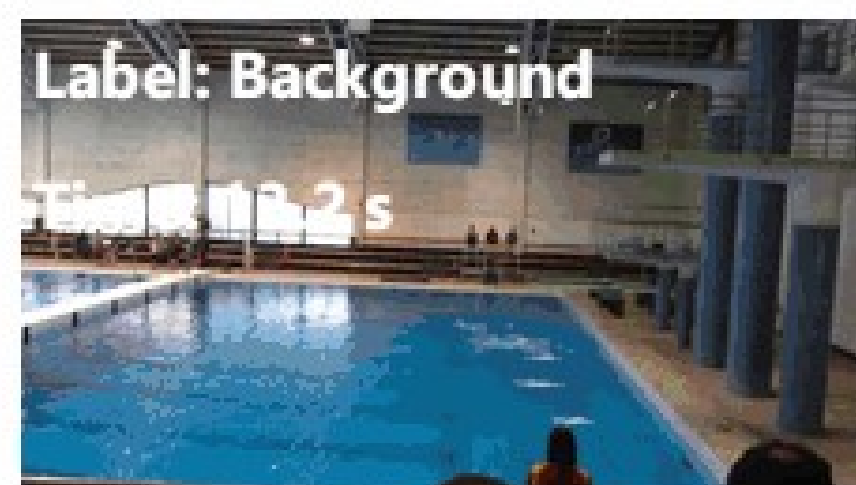
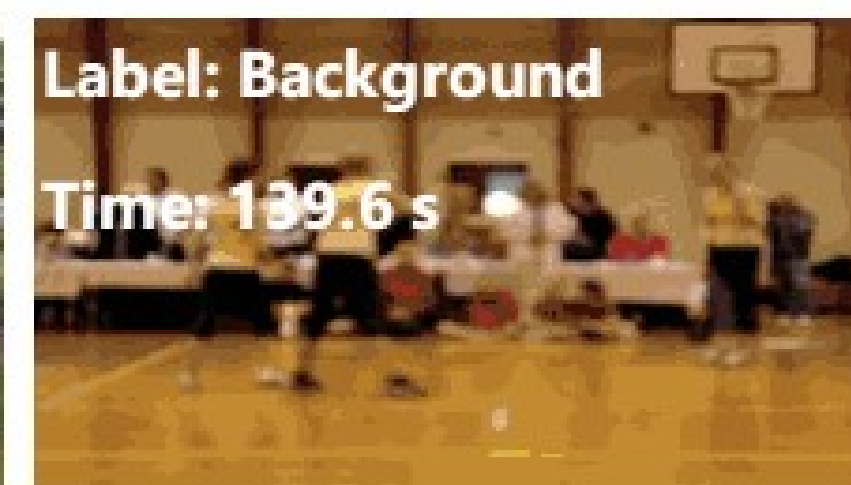
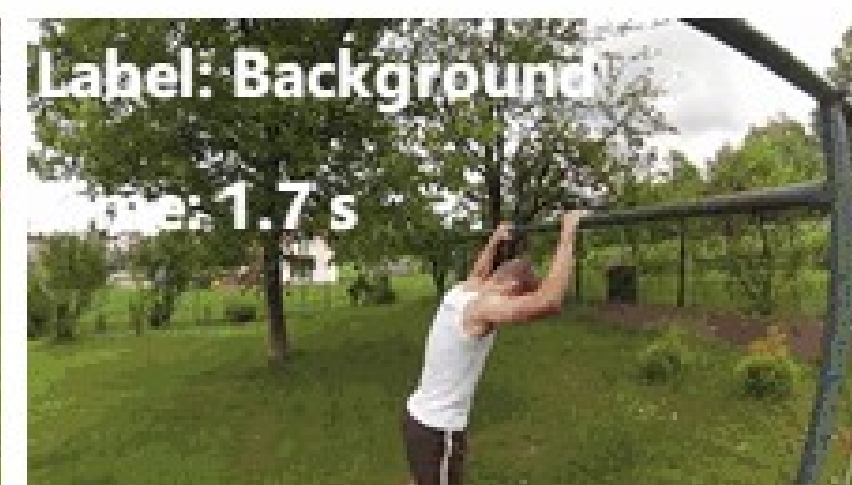
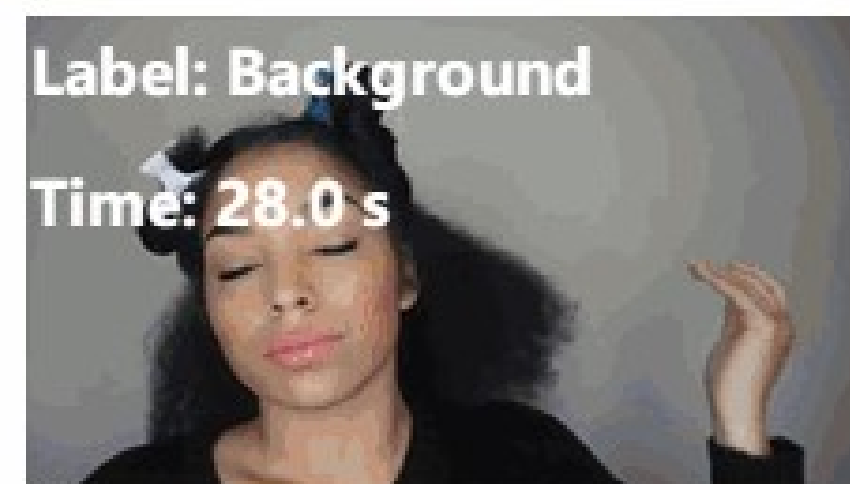
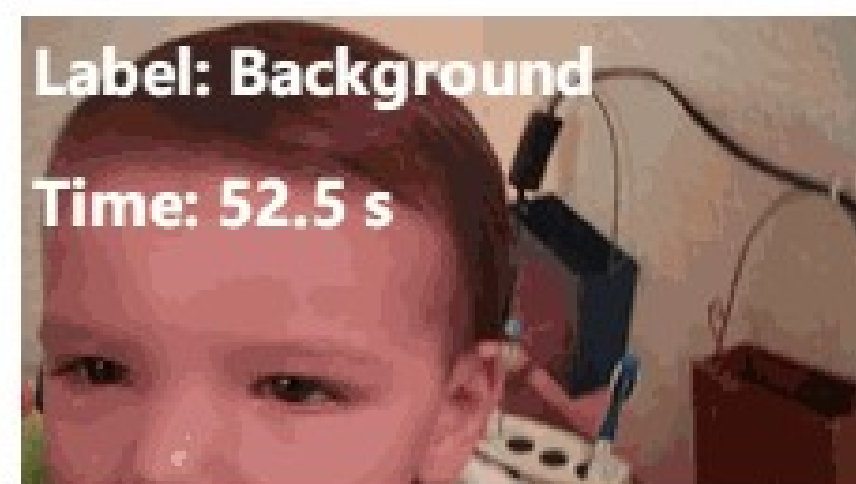
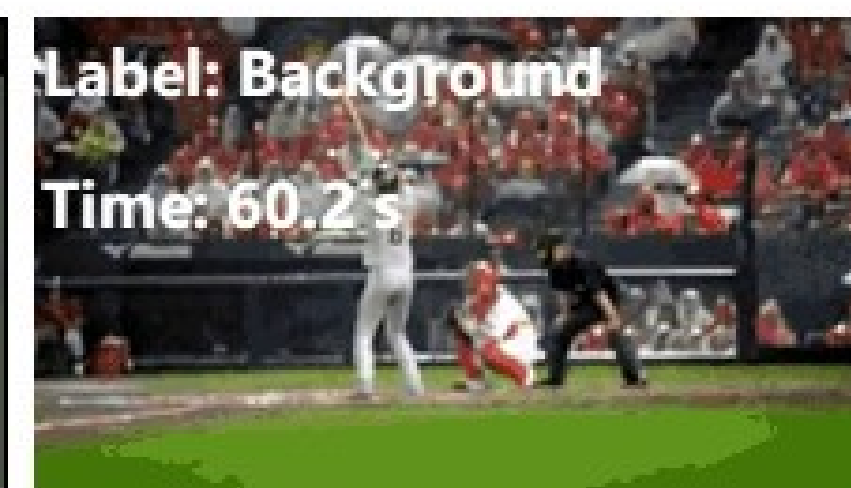
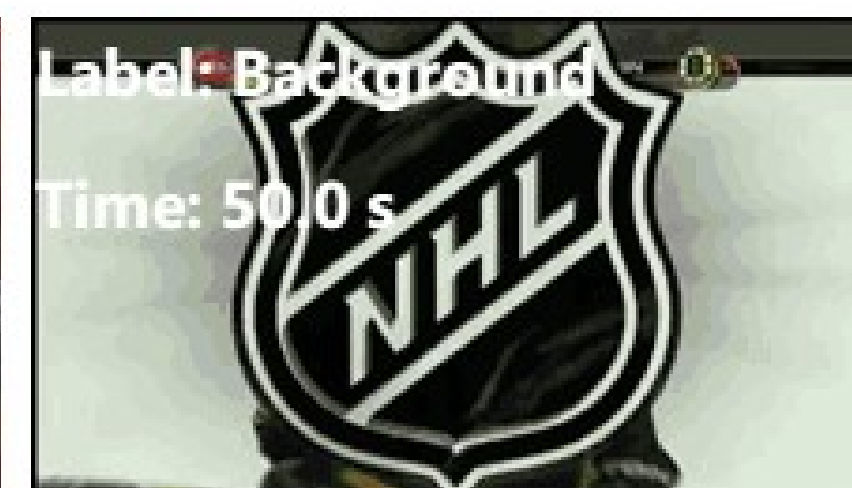
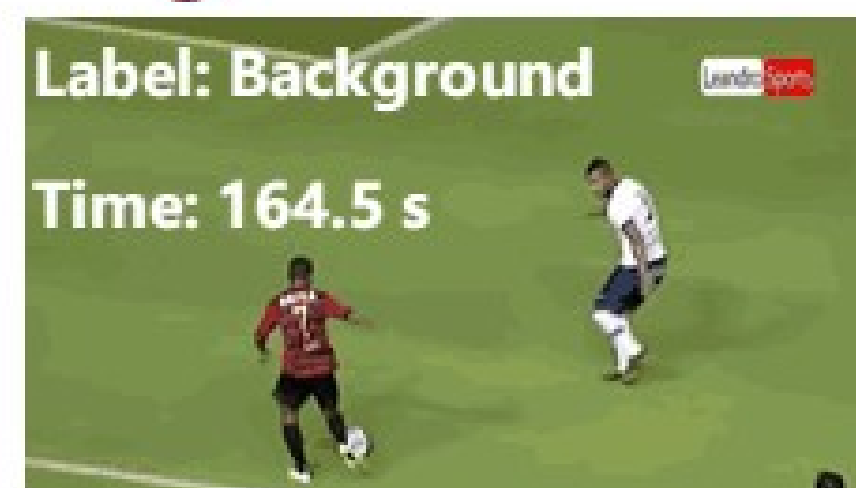




Time →

Track
Organizers:

Yi Liu , Xiao Ma



Track

Organizers:

Yixuan Li,

Lei Chen,

Runyu He,

Zhenzhi Wang



Track Organizers:

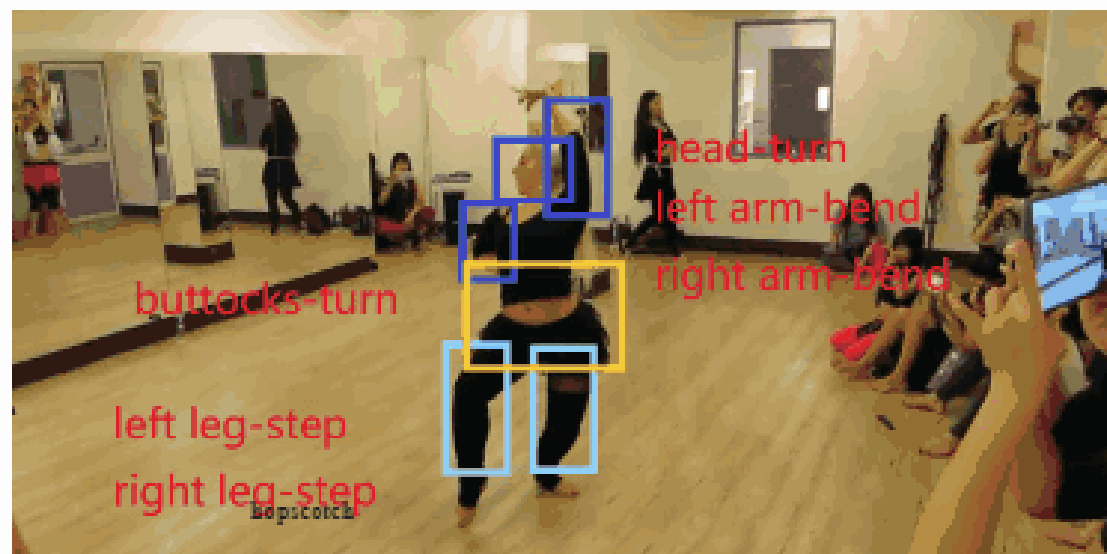
Xiao Ma,

Ding Xia,

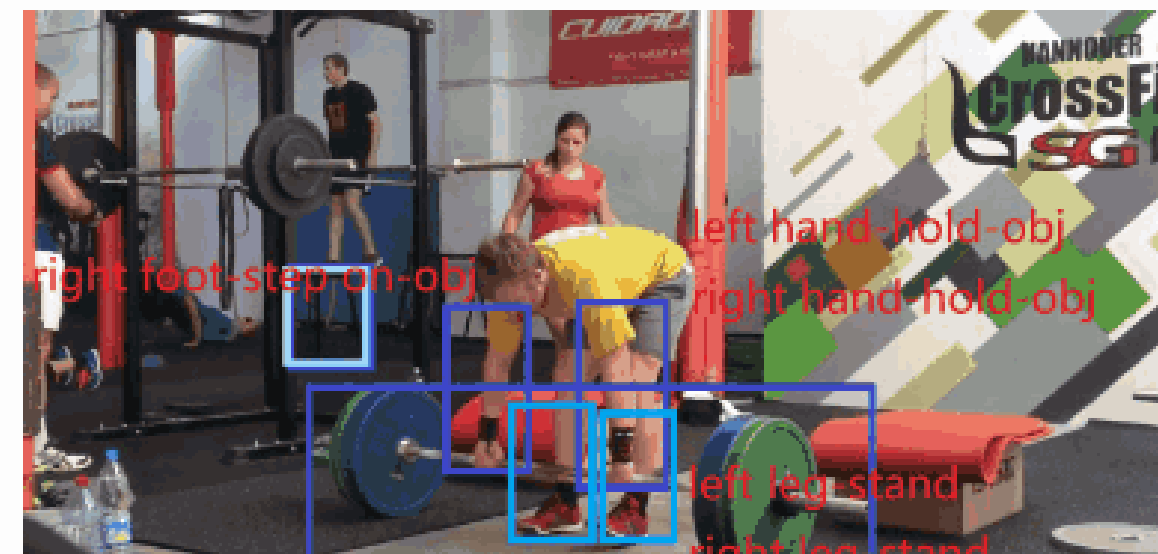
Dongliang Wang,

Weihaio Gan,

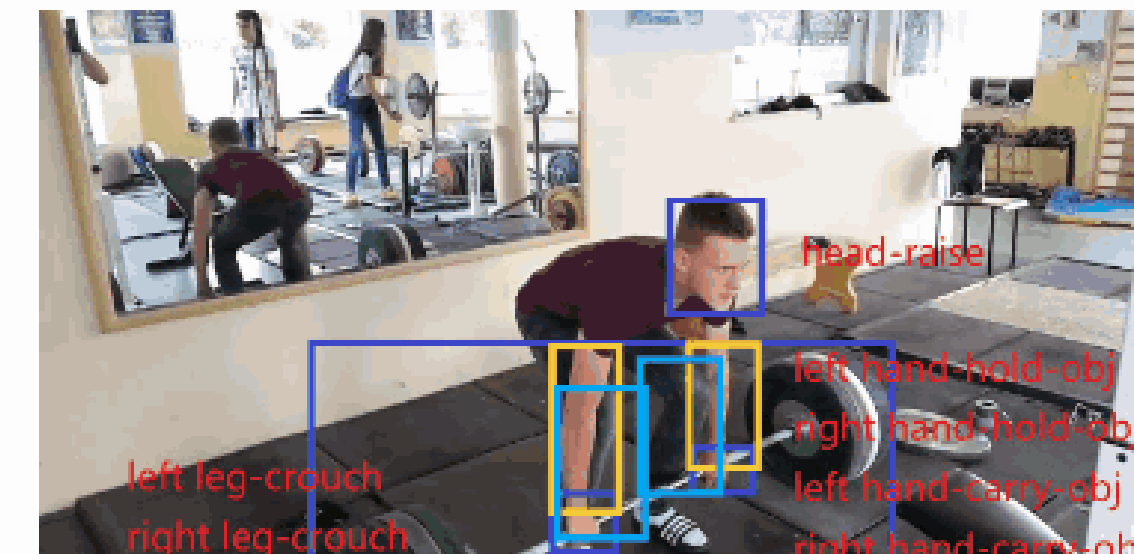
Junwei He



belly dancing



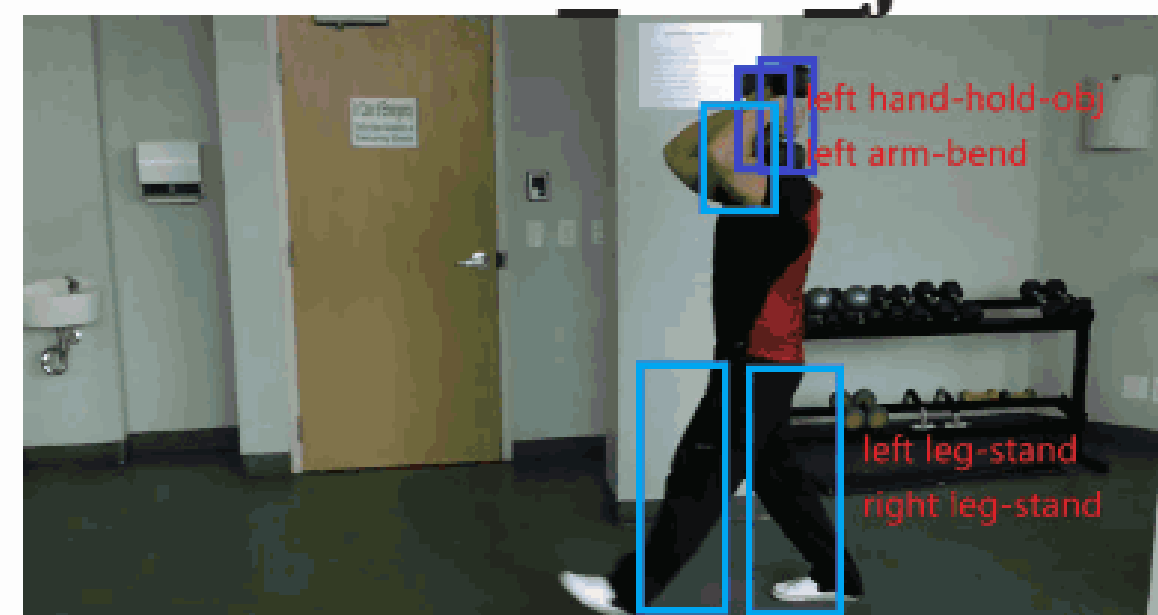
clean_and_jerk



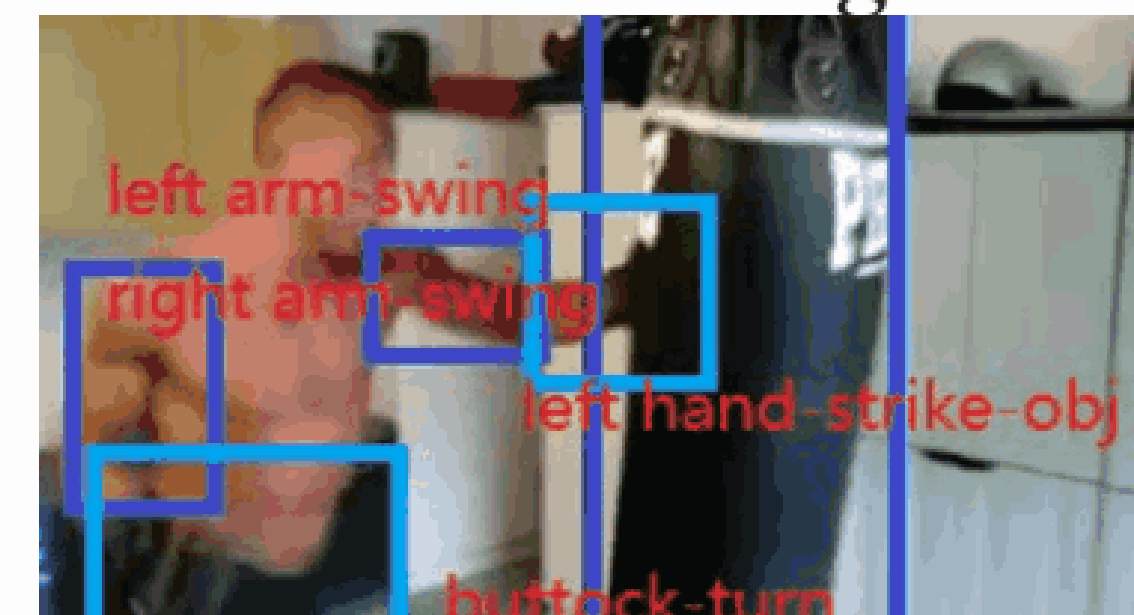
deadlifting



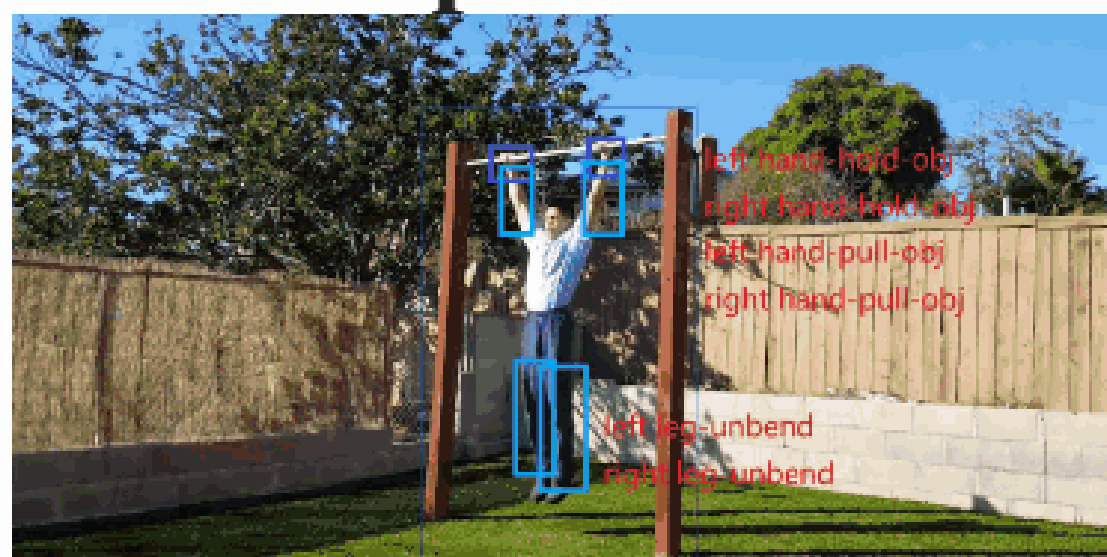
hopscotch



lunge



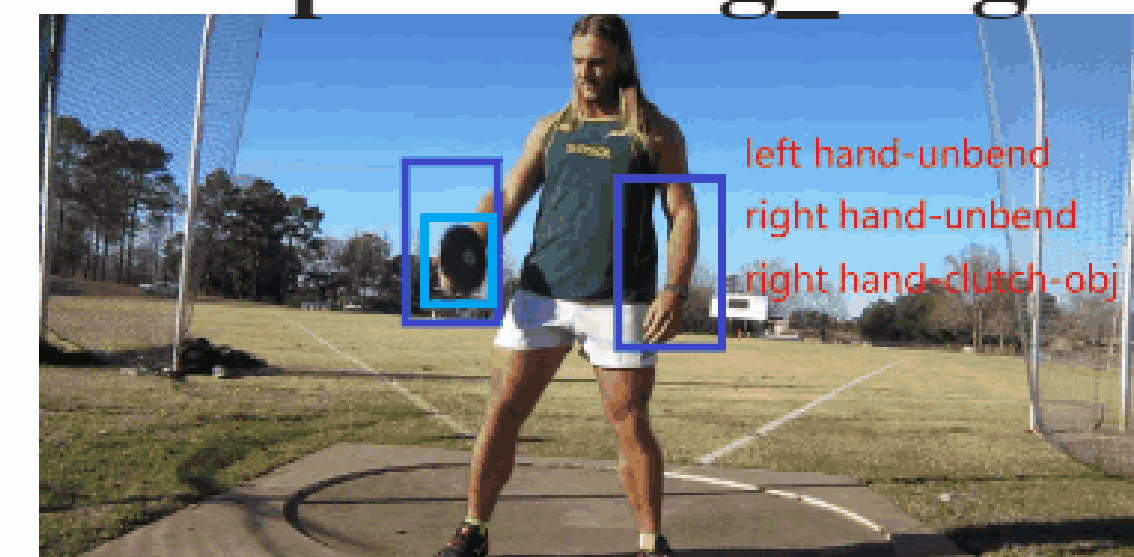
punching_bag



pull_ups



push_up



throwing_discus

Contribution of Our Workshop

FineAction Dataset

- 106 categories
- 16732 videos
- 103324 instances

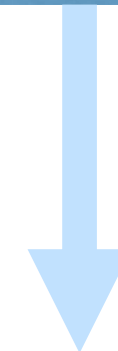


DeeperAction Challenge

- FineAction Track

MultiSports Dataset

- 66 categories
- 3200 clips
- 37701 instances
- 902k bboxes

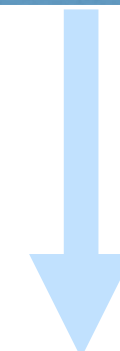


DeeperAction Challenge

- MultiSports Track

Kinetics-TPS Dataset

- 24 categories
- 4741 videos
- 1.6 M bboxes of human instances
- 0.5M bboxes of objects
- 7.9M bboxes of body parts



DeeperAction Challenge

- Kinetics-tps Track

Pushing the direction of deeper and finer understanding of human

14:00-14:10	Opening Remarks
14:10-14:40	Invited Talk 1 (Chen Sun): Detailed Human Action Understanding from Unlabeled Videos
15:40-15:00	Introduction to Track 1 - FineAction (Temporal Action Localization)
15:00-15:20	Winner Talk for Track 1 (Dahua Technology)
15:20-15:50	Invited Talk 2 (Gül Varol): Beyond Action Recognition: Detailed Video Modeling
15:50-16:10	Introduction to Track 2 - MultiSports (Spatio-Temporal Action Detection)
16:10-16:30	Winner Talk for Track (Huawei Noah's Ark Lab & USTC)
16:30-16:50	Introduction to Track 3 - Kinetics-TPS (Part-level Action Parsing)
16:50-17:10	Winner Talk for Track 3 (Hikvision Research Institute)
17:10-17:20	Closing and Awards

Chen Sun

Brown University and Google



14:00-14:10	Opening Remarks
14:10-14:40	Invited Talk 1 (Chen Sun): Detailed Human Action Understanding from Unlabeled Videos
15:40-15:00	Introduction to Track 1 - FineAction (Temporal Action Localization)
15:00-15:20	Winner Talk for Track 1 (Dahua Technology)
15:20-15:50	Invited Talk 2 (Gül Varol): Beyond Action Recognition: Detailed Video Modeling
15:50-16:10	Introduction to Track 2 - MultiSports (Spatio-Temporal Action Detection)
16:10-16:30	Winner Talk for Track (Huawei Noah's Ark Lab & USTC)
16:30-16:50	Introduction to Track 3 - Kinetics-TPS (Part-level Action Parsing)
16:50-17:10	Winner Talk for Track 3 (Hikvision Research Institute)
17:10-17:20	Closing and Awards

Gül Varol

École des Ponts ParisTech

