Deeper Action

ECCV TEL AVIV 2022

Kinetics-TPS Challenge on Part-level Action Parsing

Kunchang Li[1] Xiao Ma[1] Ding Xia[1] Dongliang Wang[2] Yali Wang[1] Yi Liu[1]

Weihao Gan[2] Jing Shao[2] Wei Wu[2] Junjie Yan[2] Yu Qiao[1,3]

[1] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

[2] SenseTime [3] Shanghai Artificial Intelligence Laboratory

Deeper Action

Kinetics-TPS Track Organizers

ECCV TEL AVIV 2022

Kunchang Li

Xiao Ma

Ding Xia

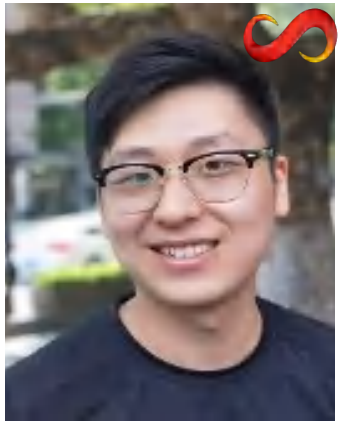
Dongliang Wang

Yali Wang

Yi Liu

Weihao Gan

Jing Shao



Wei Wu

Junjie Yan

Yu Qiao

Deeper Action

Outline

ECCV TEL AVIV 2022

- Motivation
- Dataset Introduction
- Kinetics-TPS Competition

Deeper Action

Why to do?

ECCV TEL AVIV 2022

- Action recognition is treated as a high-level video classification

Deep Learning Classifiers → Pull_Ups

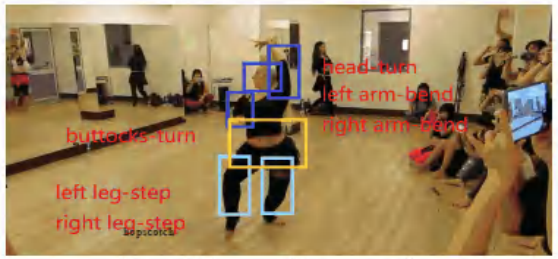
Large-Scale Video Benchmarks with Only Action Label

Deeper Action

Why to do?

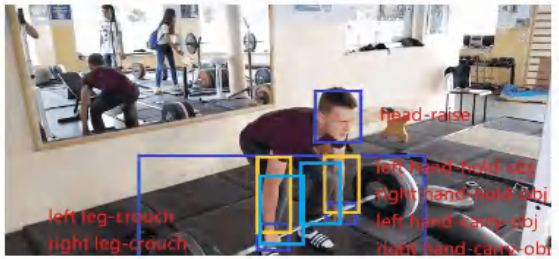
ECCV TEL AVIV 2022

- Human action is spatio-temporal composition of body part state



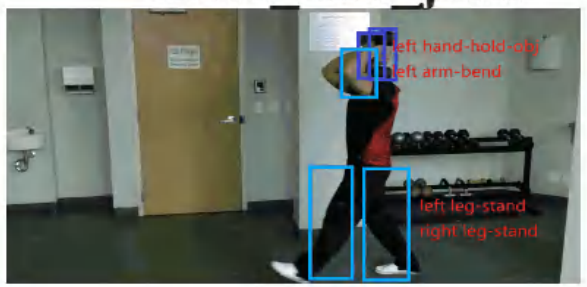
belly_dancing

clean_and_jerk



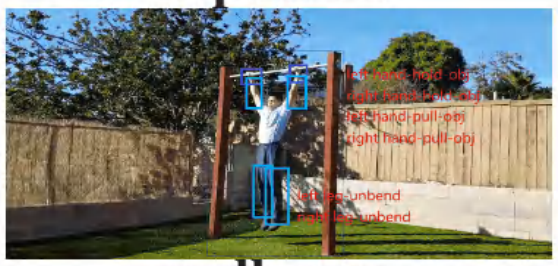
deadlifting

hopscotch

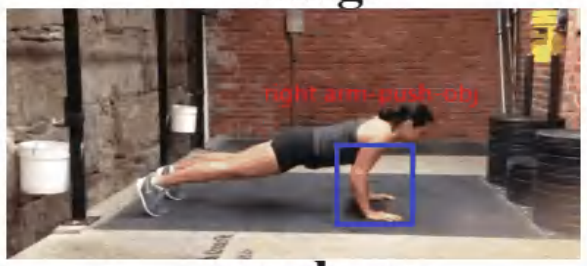


lunge

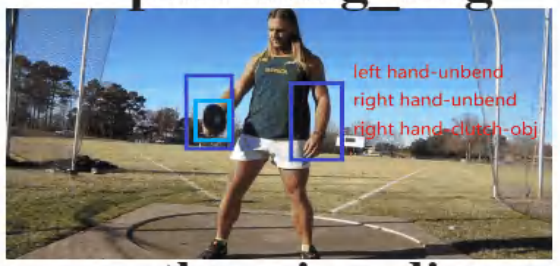
punching_bag



pull_ups



push_up



throwing_discus

Deeper Action

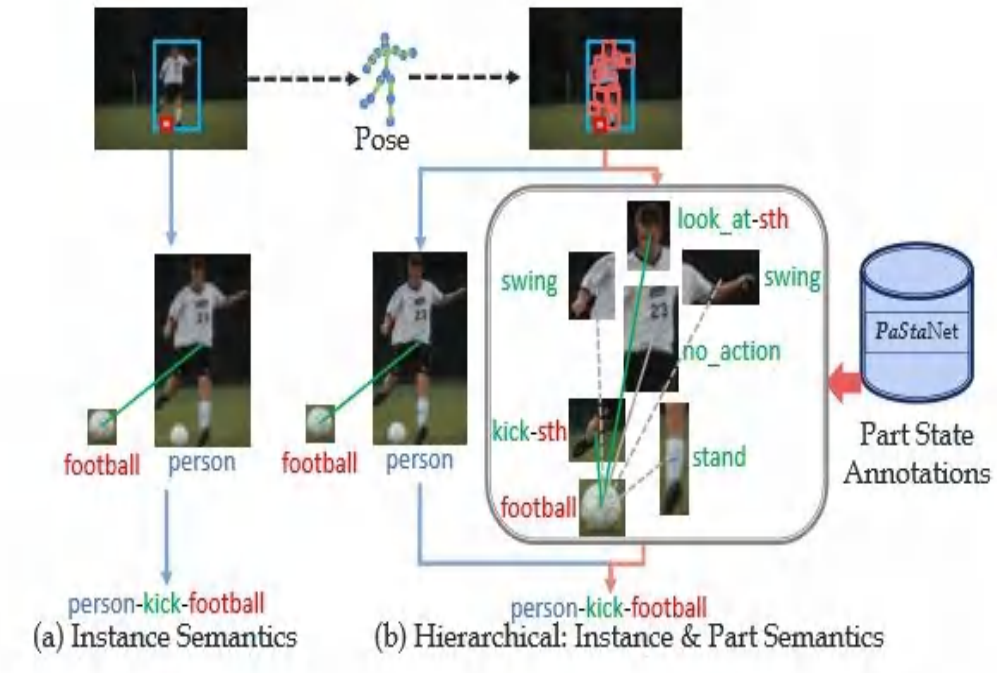
Why to do?

ECCV TEL AVIV 2022

Action Genome
(Without Body Part State)

Ji et al., Action Genome: Actions as Composition of Spatio-temporal Scene Graphs, CVPR2020

HAKE
(Image-based HOI)



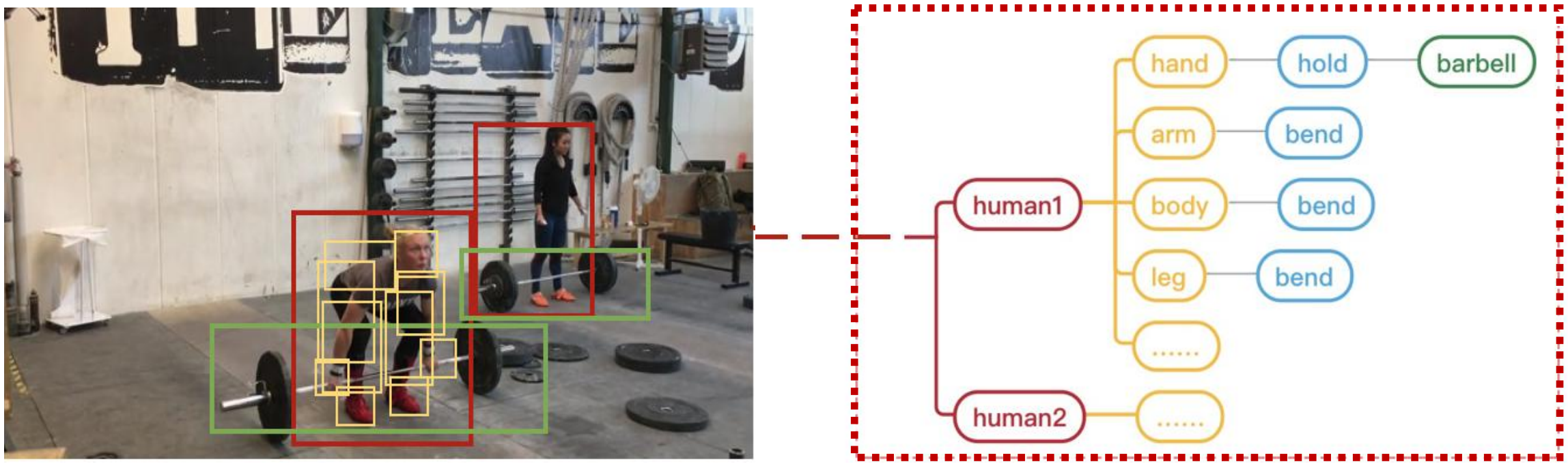
Li et al., PaStaNet: Toward Human Activity Knowledge Engine, CVPR2020

Deeper Action

Kinetics-TPS Dataset

ECCV TEL AVIV 2022

A large-scale video dataset for **Part-level Action Parsing**



Deeper Action

Key Data Statistics

ECCV TEL AVIV 2022

10 Million detailed annotations for understanding human actions

Videos Collection

- 24 action classes from Kinetics-700
- 4741 videos (3809/932 for Train/Test)

Human Annotation

- 1.6 M bboxes of human instances

Object Annotation

- 0.5M bboxes of objects
- 0.5 M object tags over 75 classes

Body Part Annotation

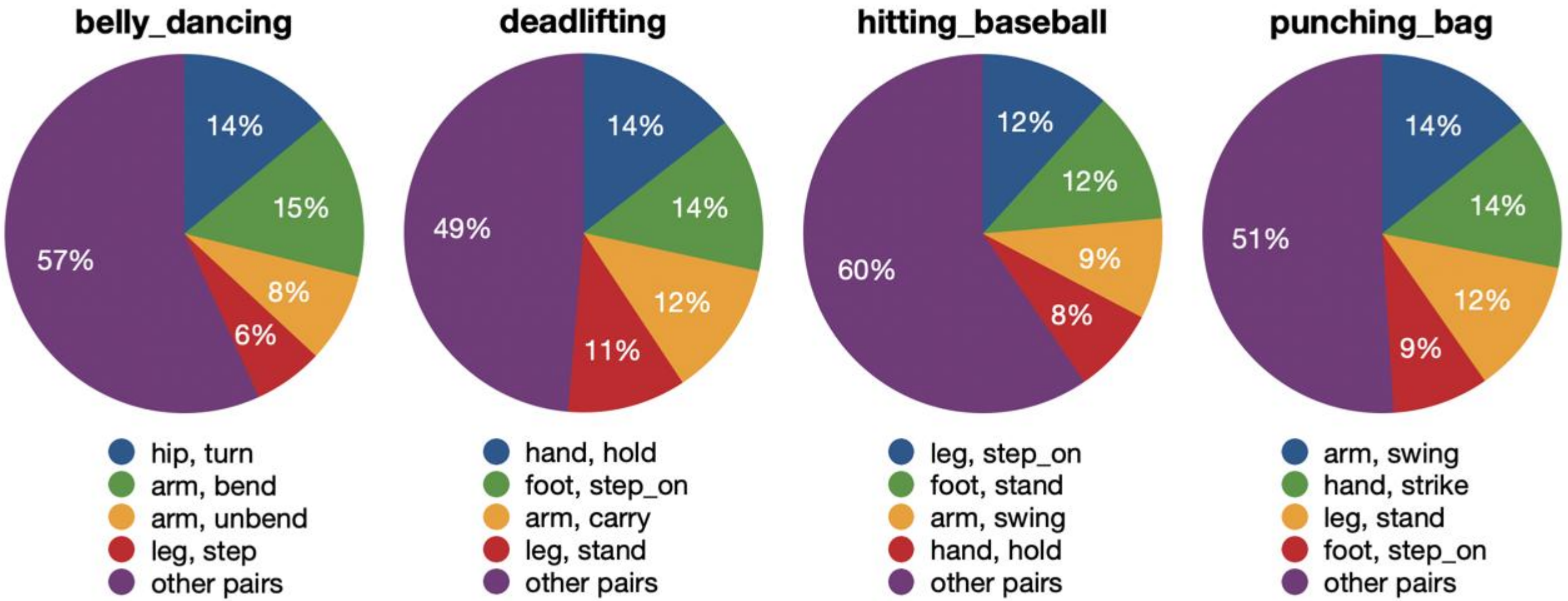
- 7.9M bboxes of body parts
- 7.9M part state tags over 74 classes

Deeper Action

Key Data Statistics

ECCV TEL AVIV 2022

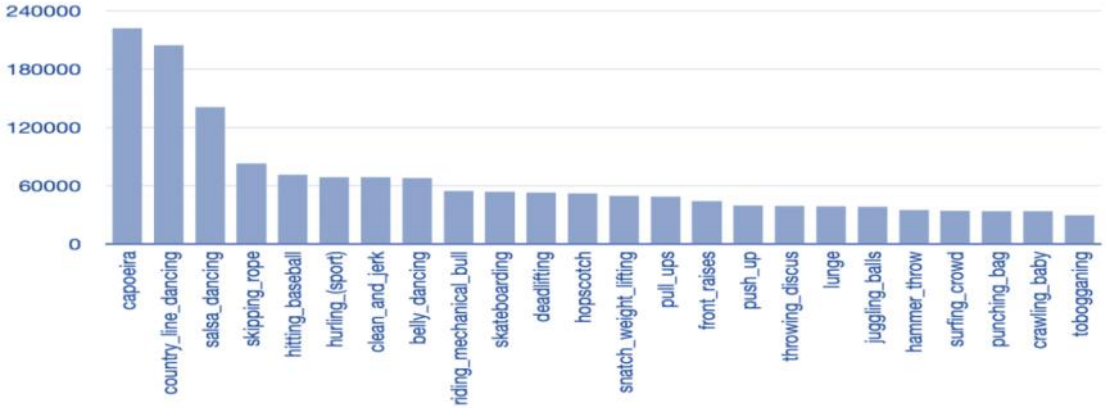
Rich diversity of (body part, part state) for various human actions



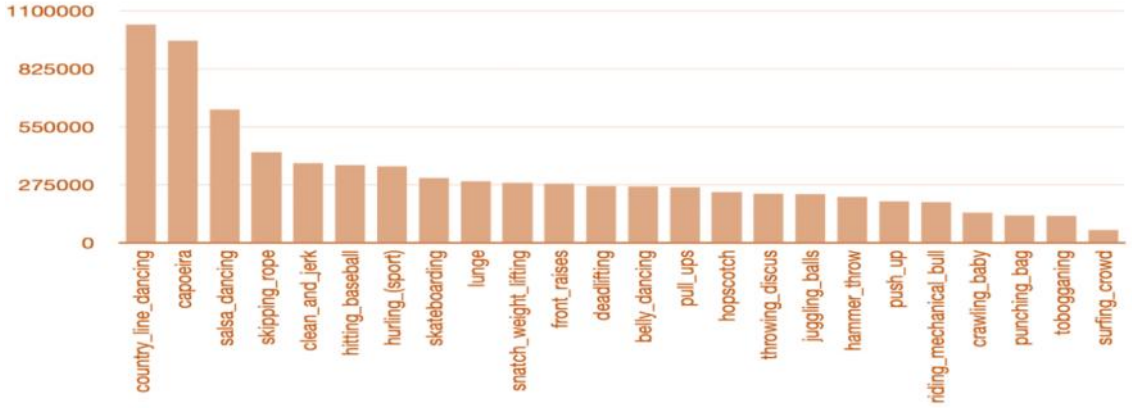
Key Data Statistics

Long-tailed distribution over all the levels of annotations

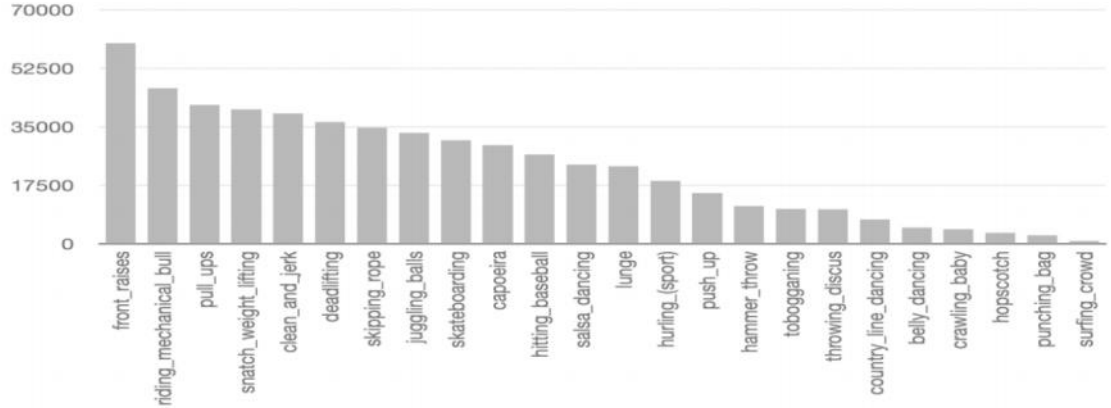
No. of human instances per action class



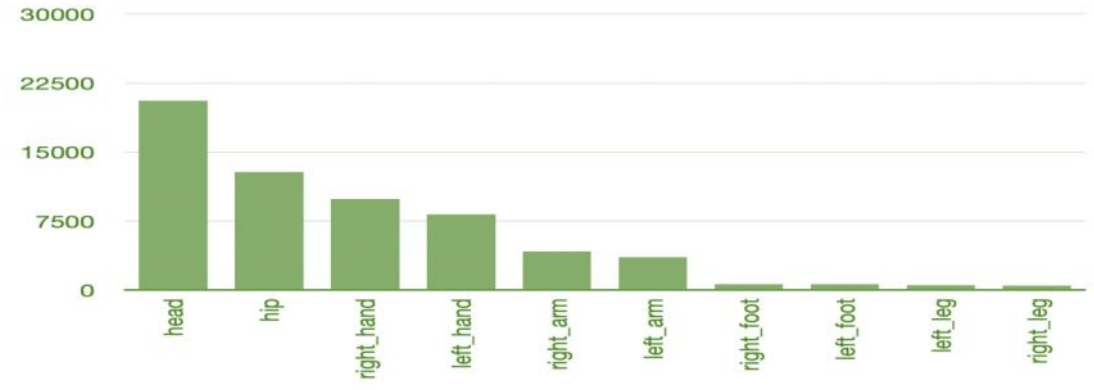
No. of part instances per action class



No. of object instances per action class



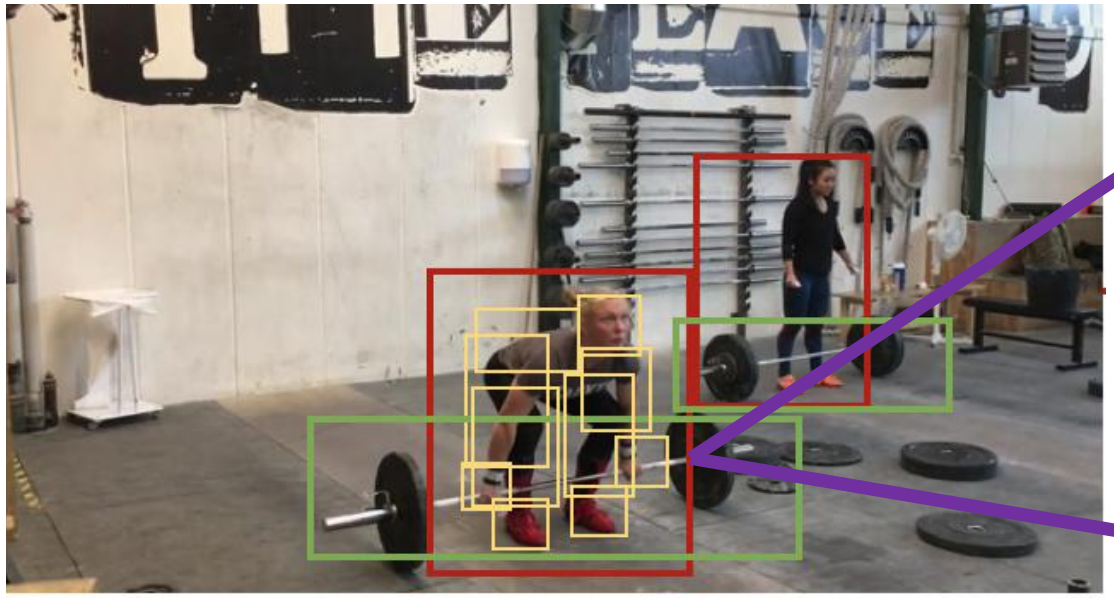
No. of part state annotations per body part



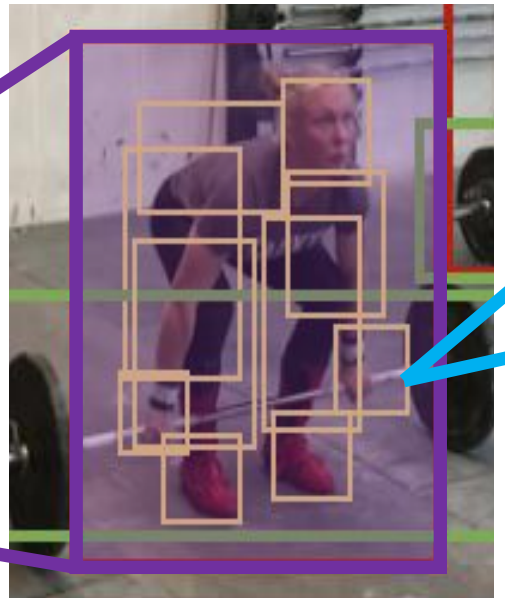
Kinetics-TPS Track: Task

1) Part State Parsing

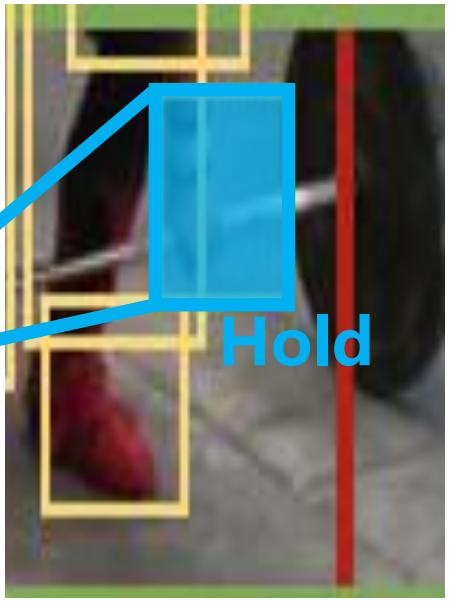
- Predicted boxes of human instances
- Predicted boxes of body parts & Predicted part state of each box



each sampled frame in a test video



Human Box

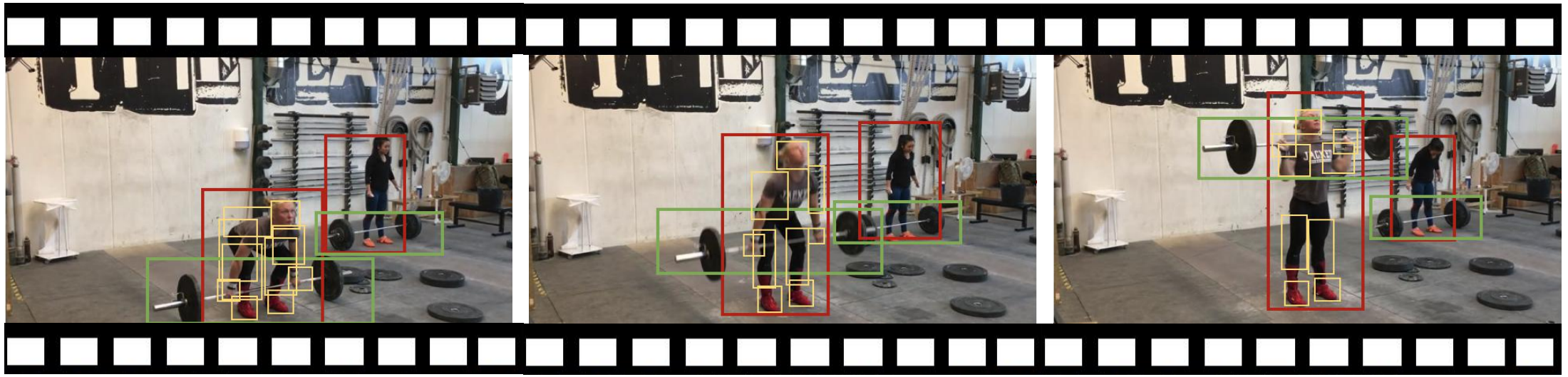


Part Box & State

Kinetics-TPS Track: Task

2) Action Recognition (for each test video)

- The predicted action label



deadlifting

Kinetics-TPS Track: Evaluation

Goal

- Leveraging part state parsing for action recognition

Metric

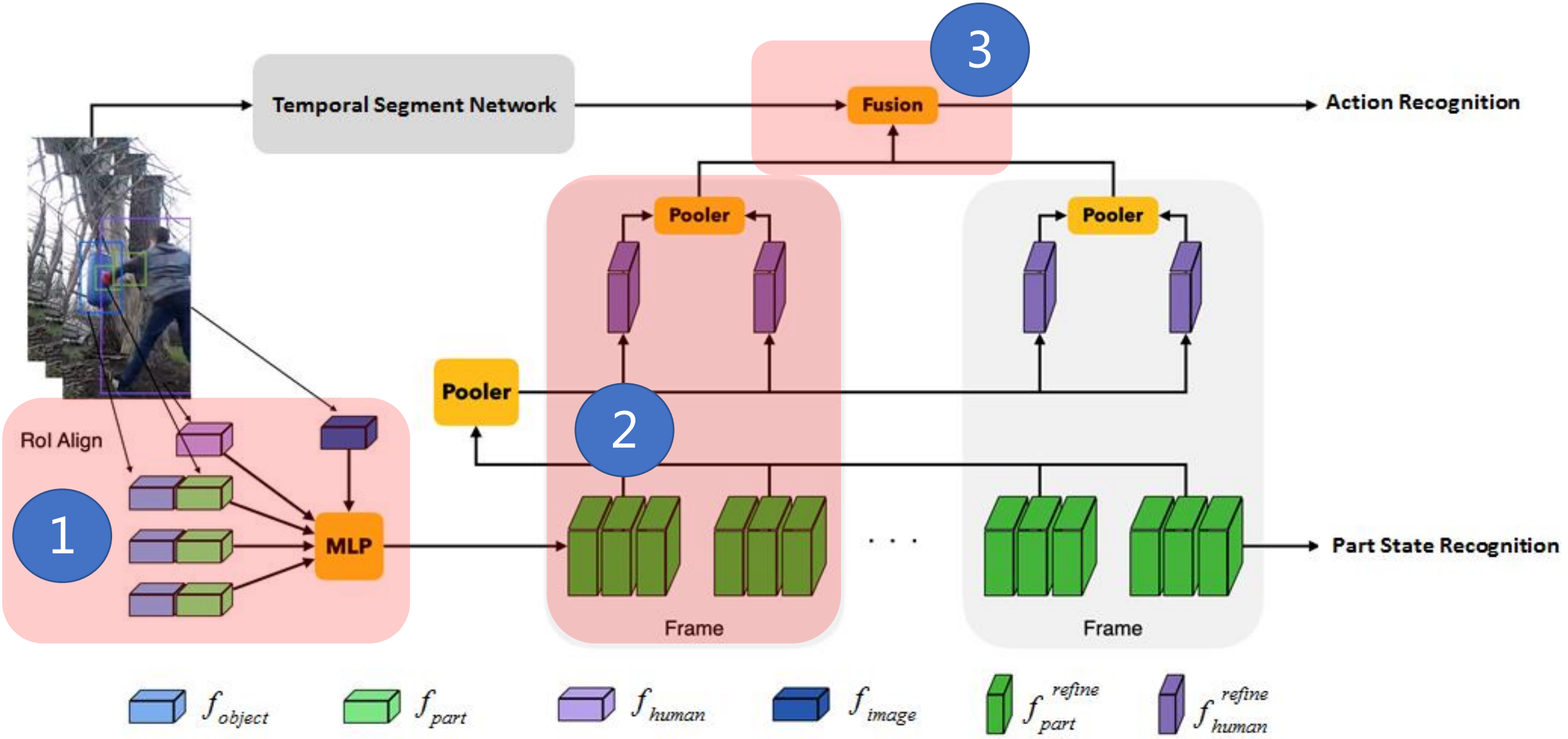
- Action recognition accuracy (ACC) conditioned on part state correctness (PSC)
- The area under PSC-ACC curve as our final evaluation metric

(https://codalab.lisn.upsaclay.fr/competitions/4392#learn_the_details-evaluation)

Deeper Action

Kinetics-TPS Track: Baseline

ECCV TEL AVIV 2022



Kinetics-TPS Track: Results

ECCV DeeperAction Challenge - Kinetics-TPS Track on Part-level Action Parsing and Action Recognition

Organized by SakuraD

The challenge is Track 4 at ECCV DeeperAction Challenge. This track is to recognize a human action by compositional learning ...

May 01, 2022-Aug 31, 2022

68 participants

Kinetics-TPS Challenge Test				
#	User	Entries	Date of Last Entry	Score ▲
1	OverWhelmingFIt	30	08/31/22	0.74 (1)
2	JosonChan	17	08/31/22	0.69 (2)
3	sota-wxh	3	08/31/22	0.66 (3)
4	Tamakoko	1	08/26/22	0.25 (4)

Deeper Action

1st Place Winner

ECCV TEL AVIV 2022

Jiawei Dong[1], Yuliang Chen[1], Shuo Wang[1]
[1]Shanghai Paidao Intelligent Technology Co., Ltd.

Deeper Action

2nd Place Winner

ECCV TEL AVIV 2022

Xiaojia Chen[1], Xuanhan Wang[1], Yan Dai[1], Jingkuan Song[12]
[1]Center for Future Media, University of Electronic Science and Technology of China, Chengdu, China
[2]Pengcheng Lab, Shenzhen, China

Deeper Action

3rd Place Winner

ECCV TEL AVIV 2022

Lianli Gao[1], Ji Zhang[1], Beitao Chen[1], Pengpeng Zeng[1]
[1]Center for Future Media, University of Electronic Science and Technology of China, Chengdu, China