Deeper Action

ICCV 2021 Workshop

Kinetics-TPS Challenge on Part-level Action Parsing

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

中国科学院深圳先进技术研究院
SHENZHEN INSTITUTE OF ADVANCED TECHNOLOGY
CHINESE ACADEMY OF SCIENCES

商汤 sensetime

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

Kinetics-TPS Track Organizers

Xiao Ma*

Ding Xia*

Dongliang Wang

Yali Wang

Yi Liu

Weihao Gan

Jing Shao

Wei Wu

Junjie Yan

Yu Qiao

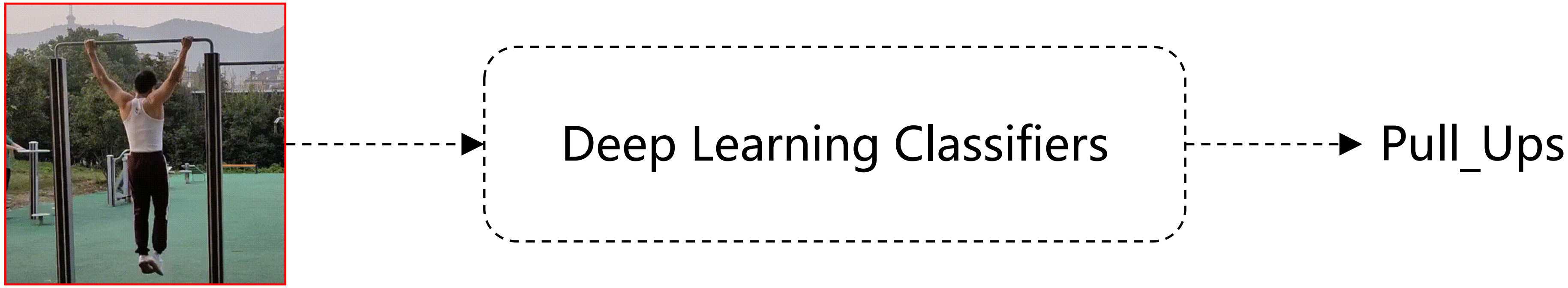
* Contributed Equally

Outline

- Motivation
- Dataset Introduction
- Kinetics-TPS Competition

Why to do?

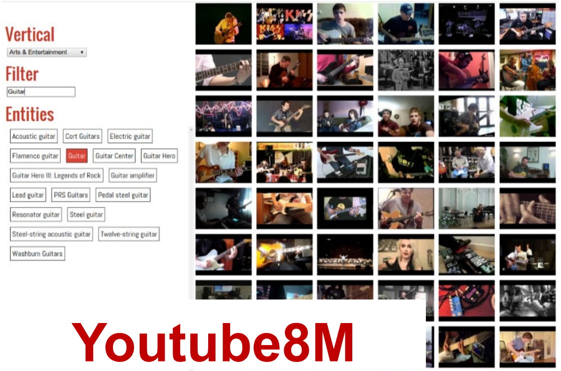
- ❑ Action recognition is treated as a high-level video classification



Deep Learning Classifiers

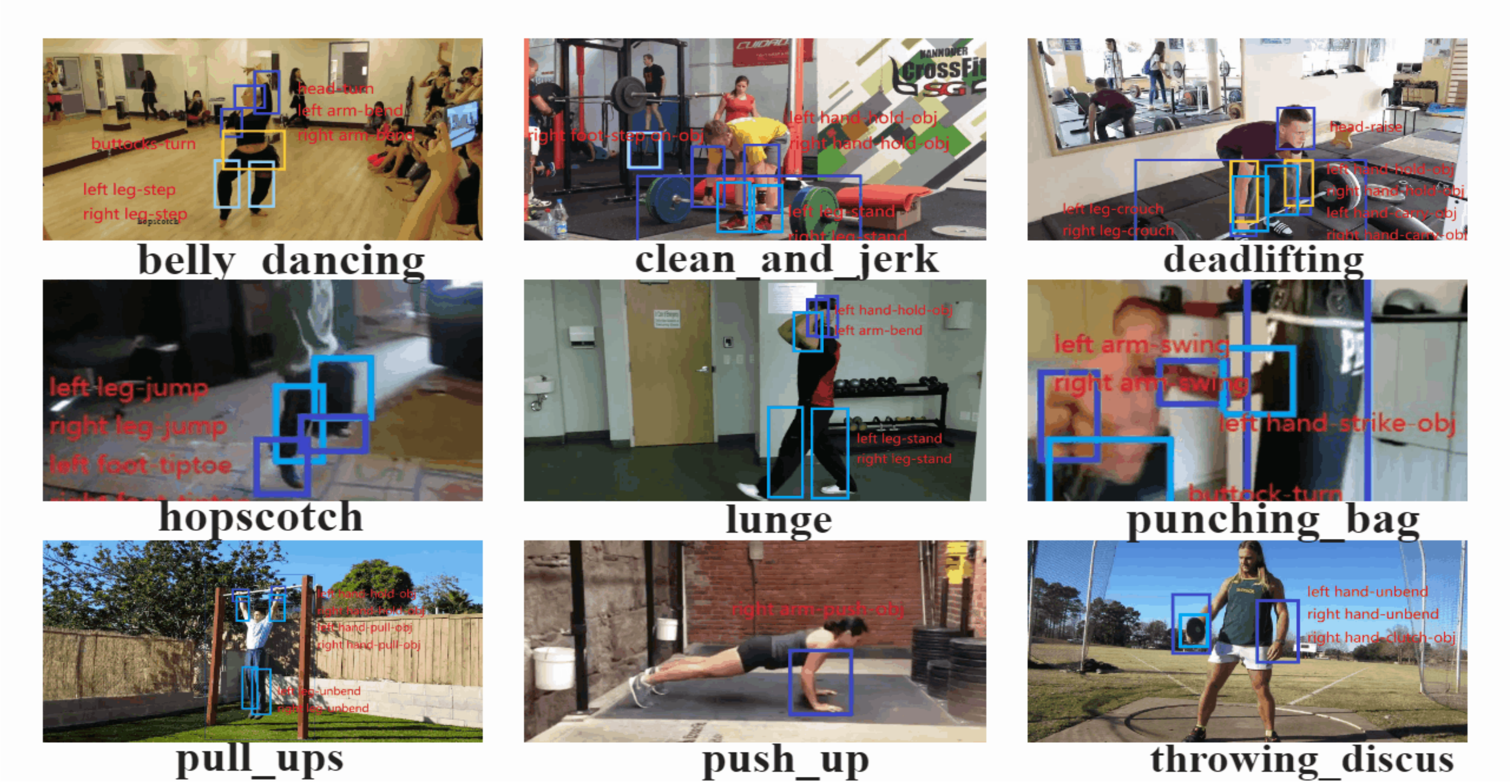
Pull_Ups

Large-Scale Video Benchmarks with Only Action Label



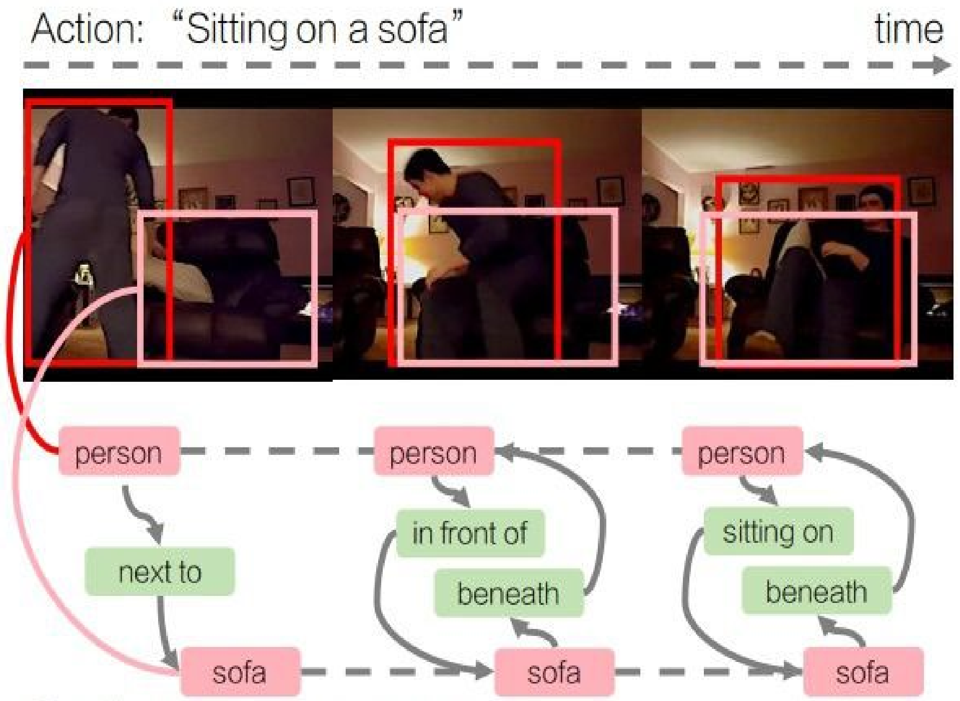
Why to do?

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



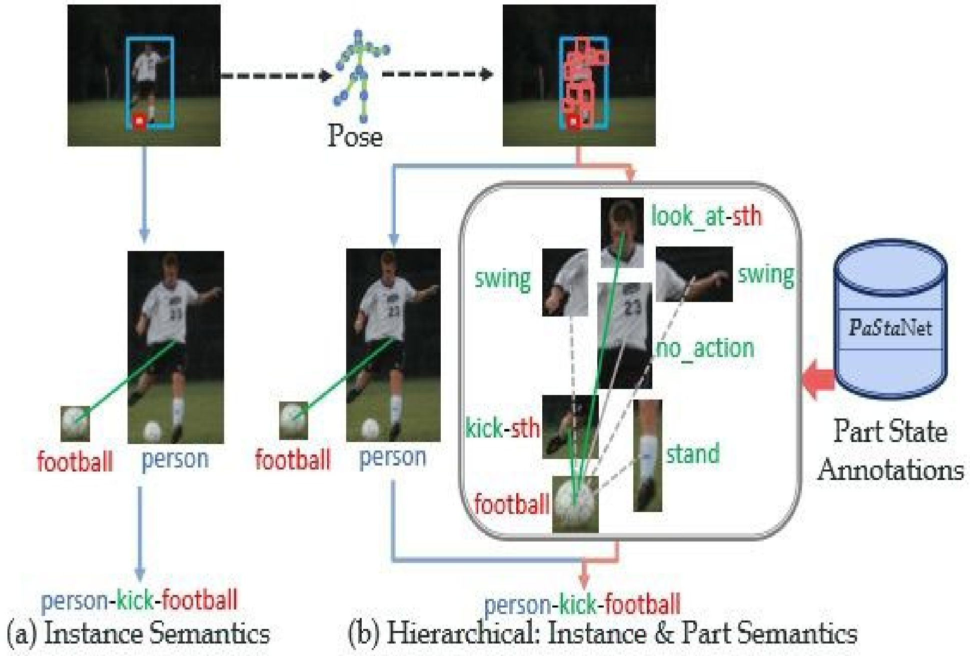
Why to do?

Action Genome (Without Body Part State)



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

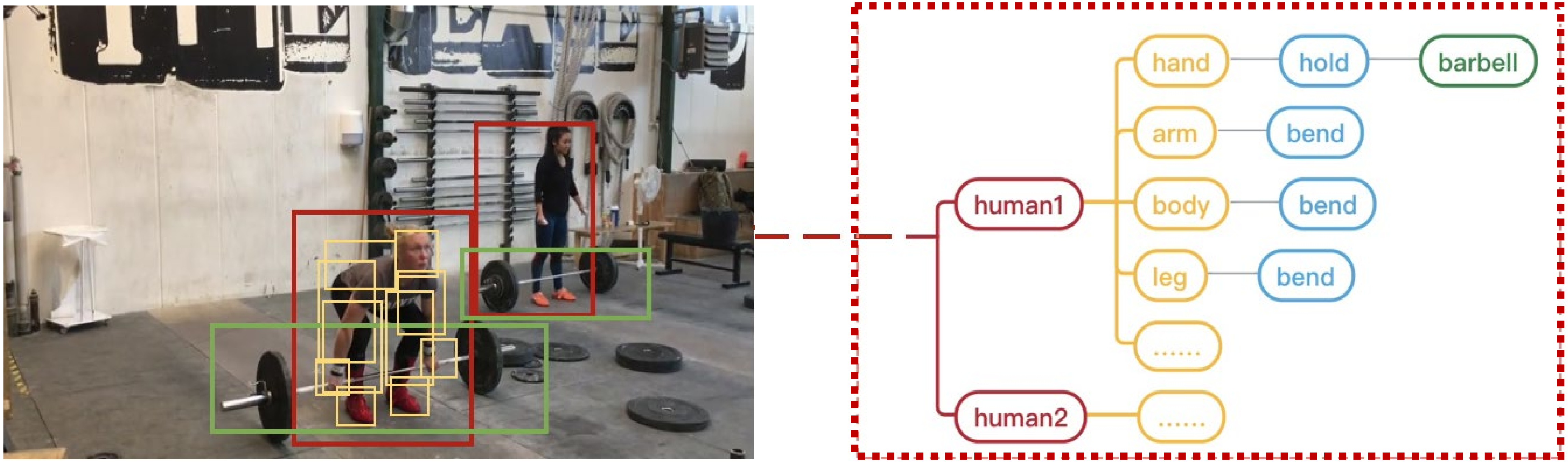
HAKE (Image-based HOI)



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

Kinetics-TPS Dataset

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



Key Data Statistics

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

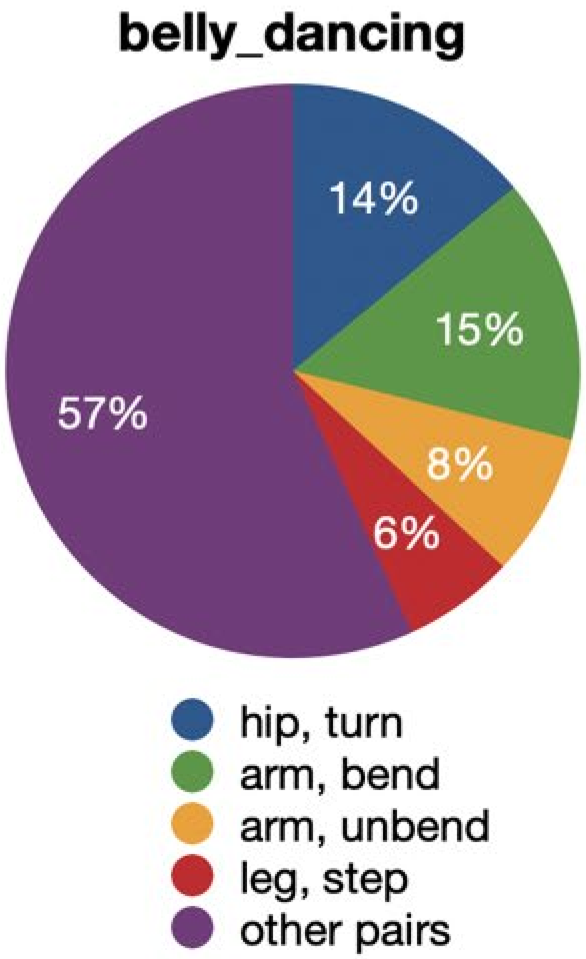
- 0.5M bboxes of objects
- 0.5 M object tags over 75 classes

Body Part Annotation

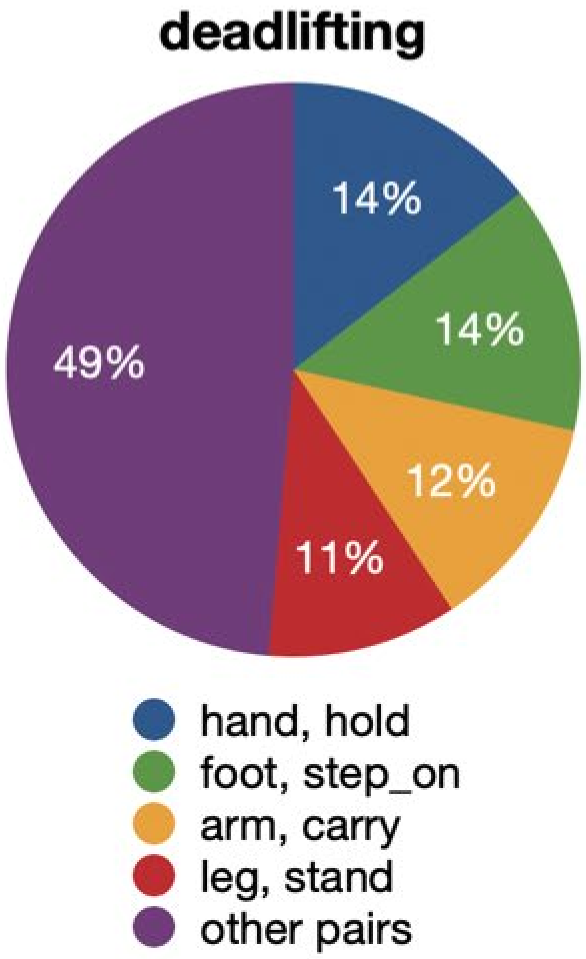
- 7.9M bboxes of body parts
- 7.9M part state tags over 74 classes

Key Data Statistics

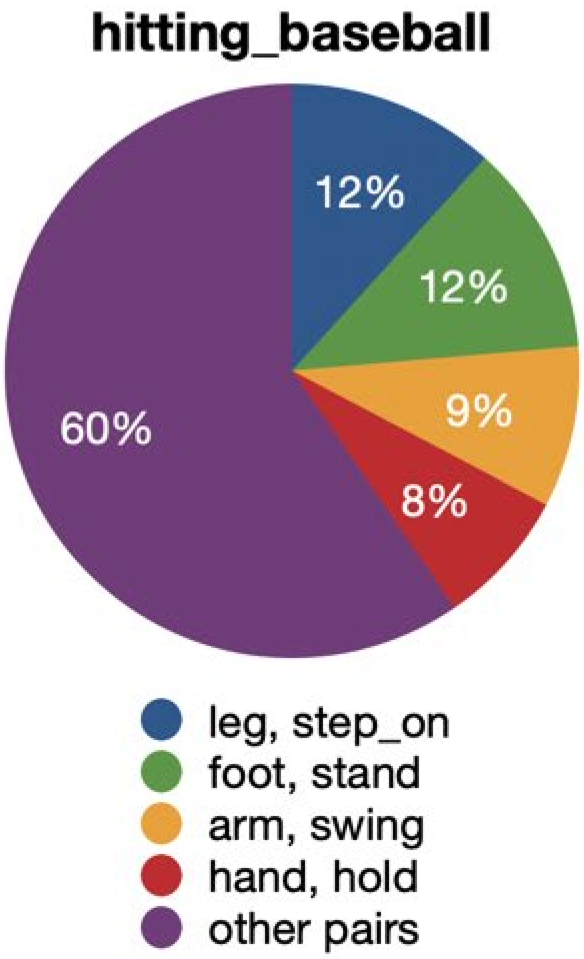
- ❑ **Rich diversity** of (body part, part state) for various human actions



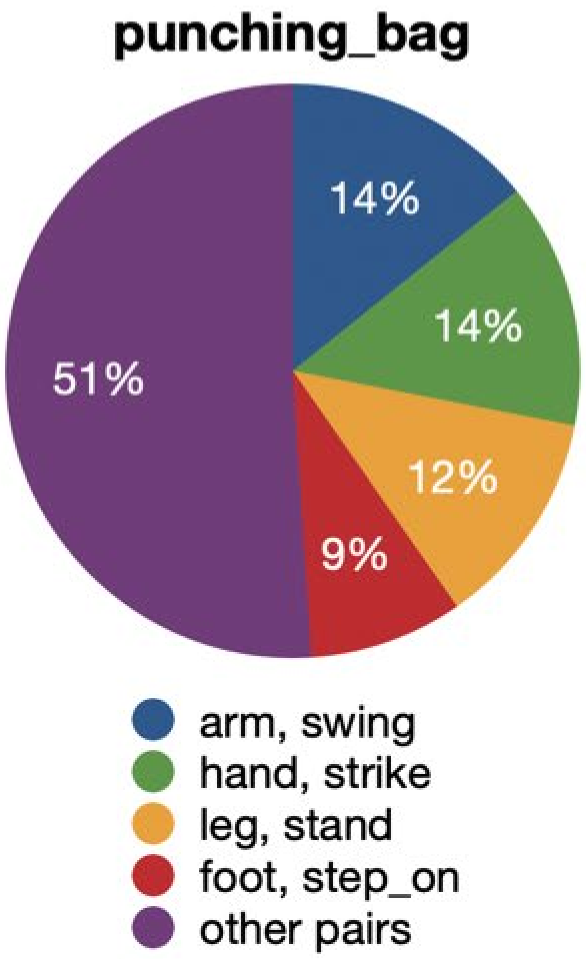
belly_dancing



deadlifting



hitting_baseball

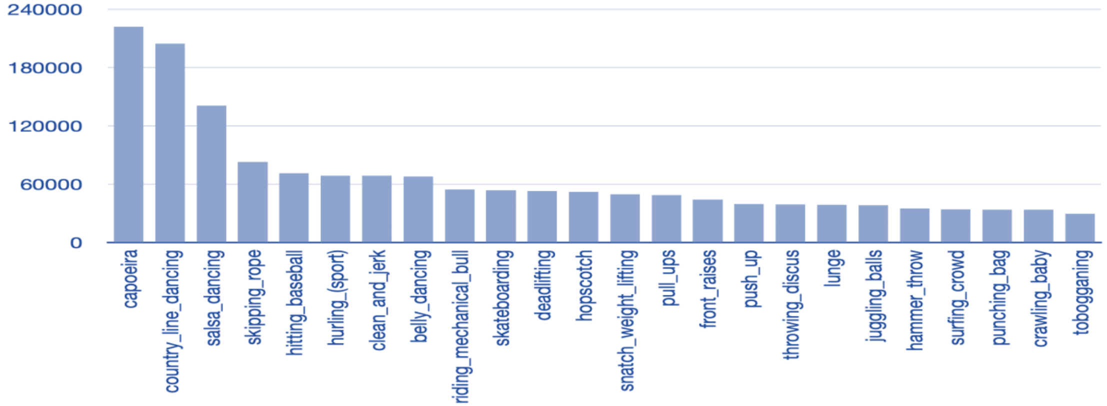


punching_bag

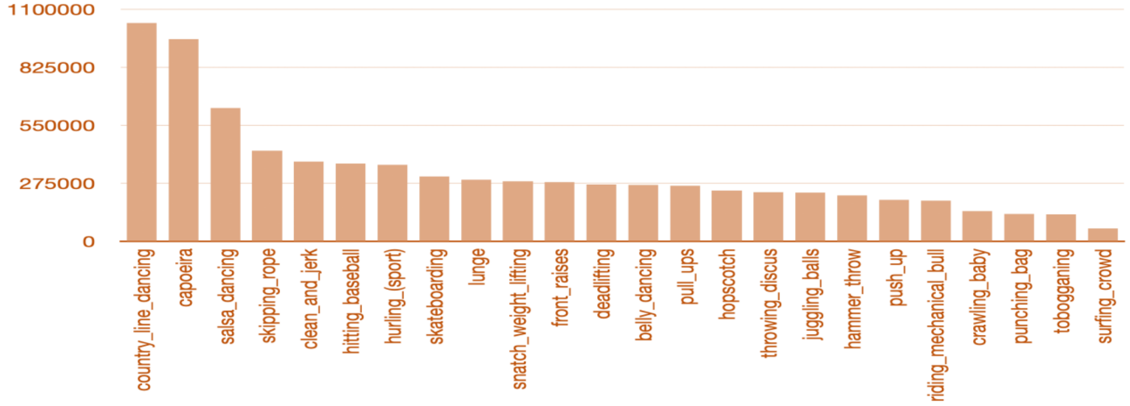
Key Data Statistics

- **Long-tailed distribution** over all the levels of annotations

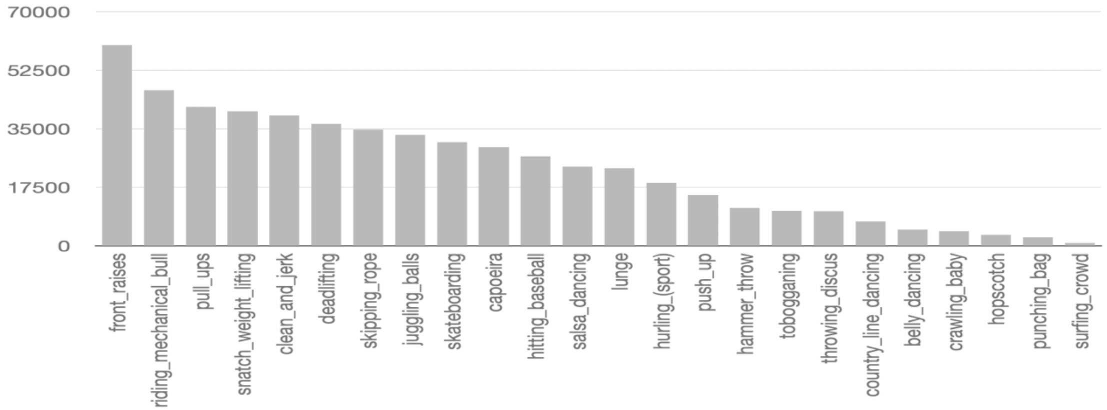
No. of human instances per action class



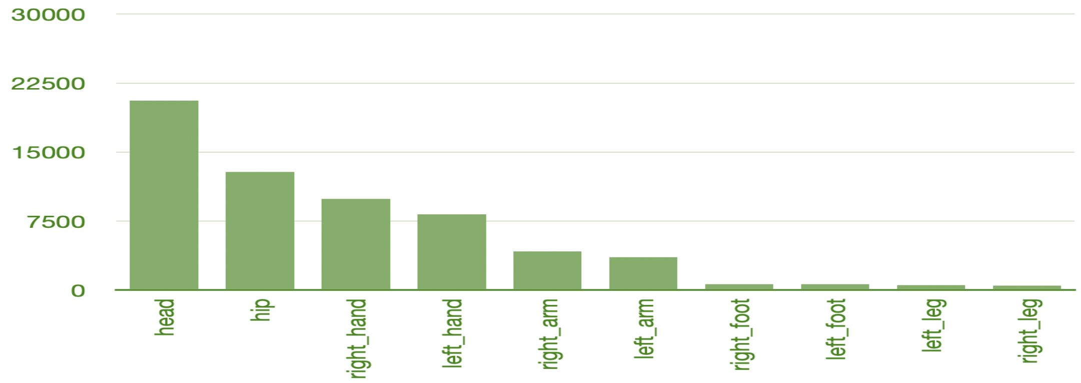
No. of part instances per action class



No. of object instances per action class



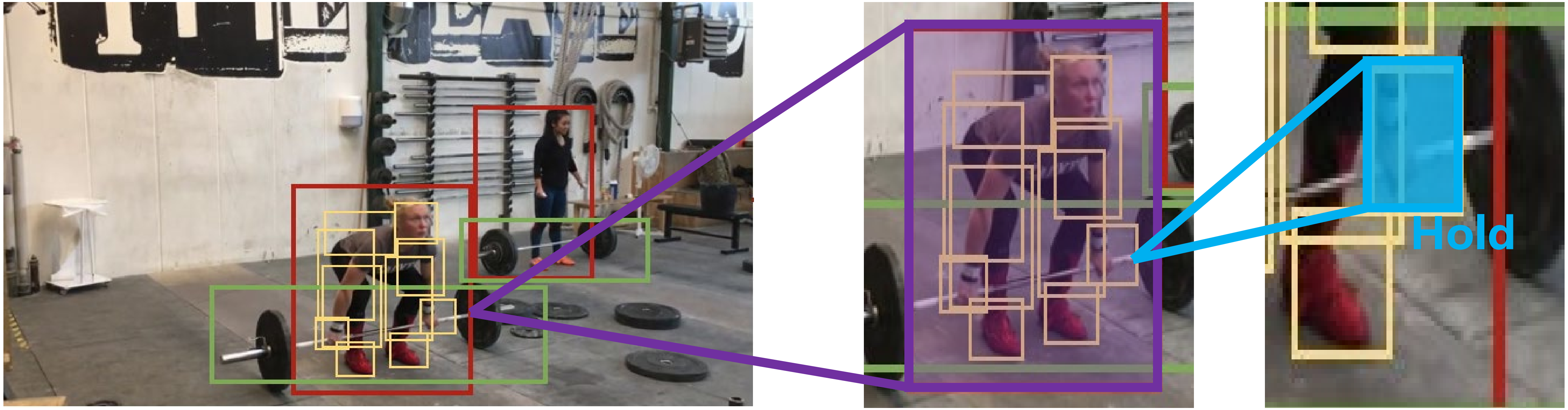
No. of part state annotations per body part



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

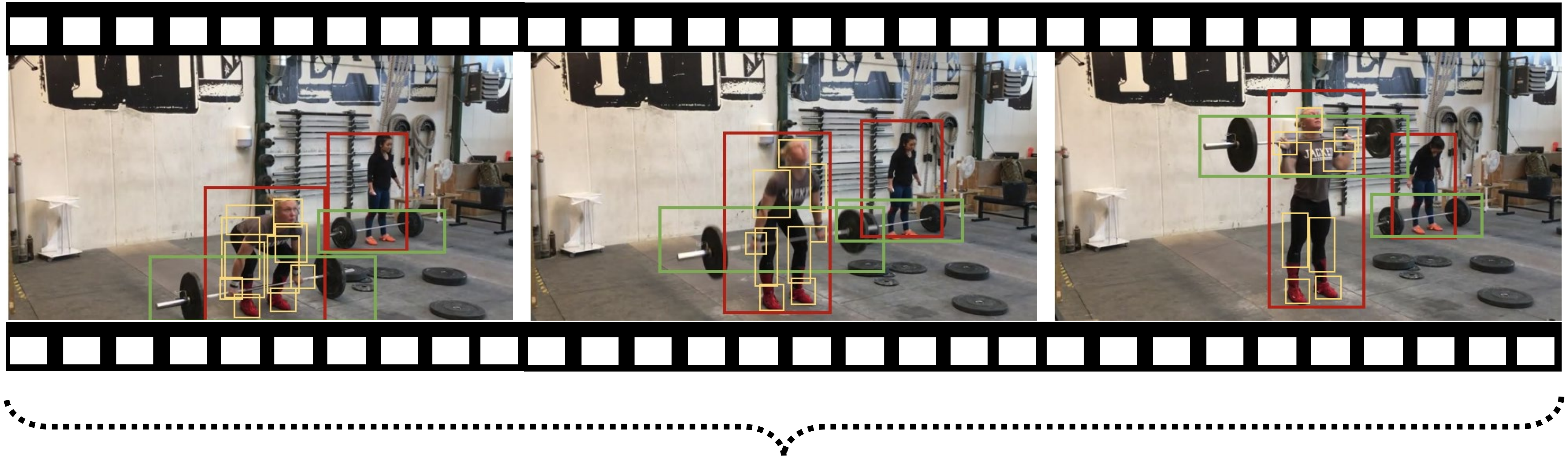
Deeper Action

Kinetics-TPS Track: Task

ICCV 2021 Workshop

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**

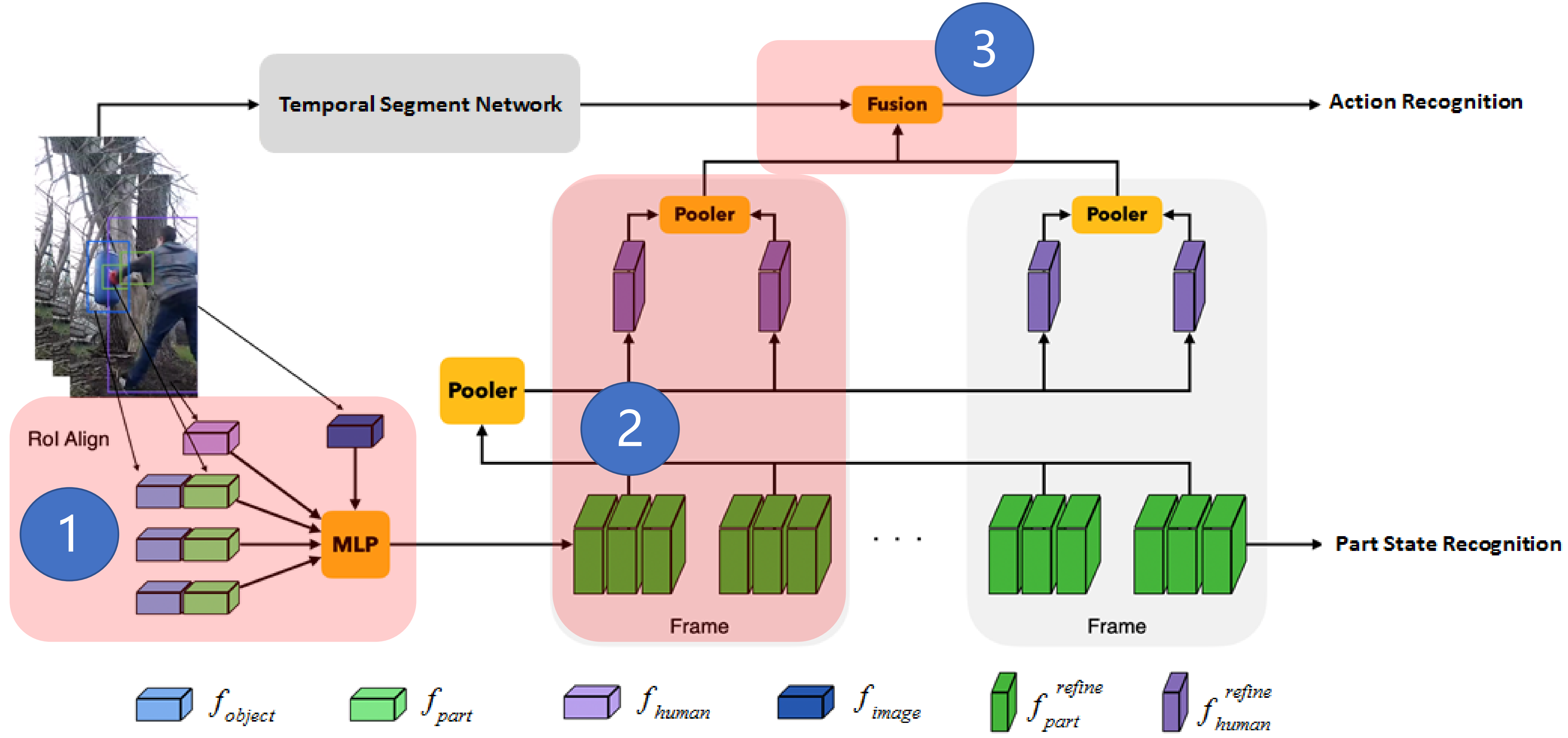
- ➢ Action recognition accuracy (ACC) conditioned on part state correctness (PSC)
- ➢ The area under PSC-ACC curve as our final evaluation metric

(https://competitions.codalab.org/competitions/32360#learn_the_details-evaluation)

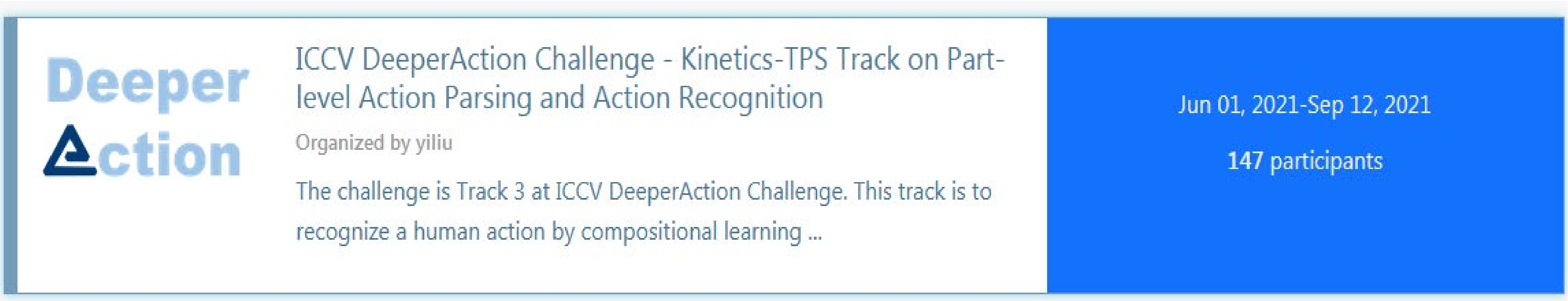
Deeper Action

Kinetics-TPS Track: Baseline

ICCV 2021 Workshop



Kinetics-TPS Track: Results



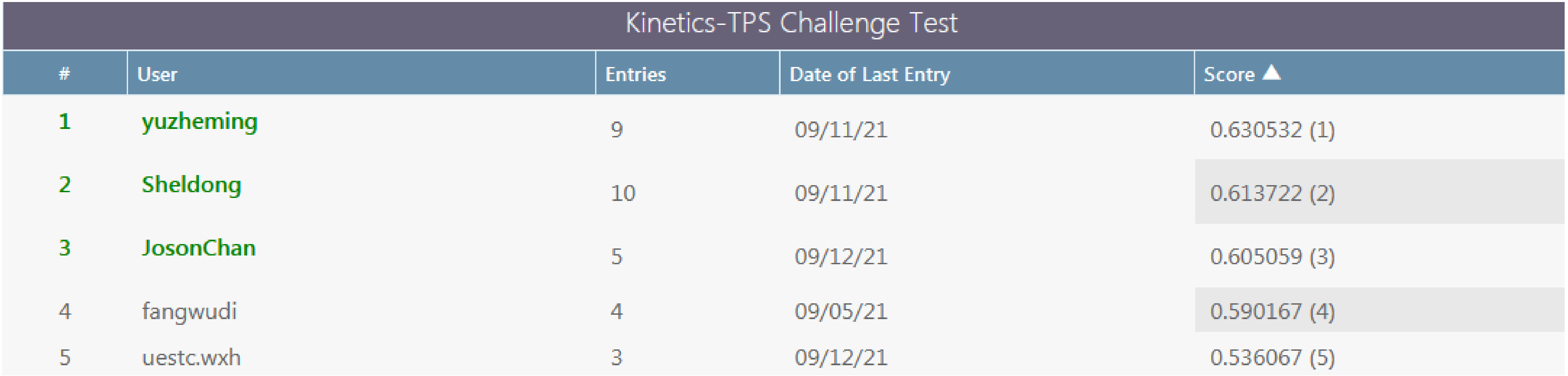
ICCV DeeperAction Challenge - Kinetics-TPS Track on Part-level Action Parsing and Action Recognition

Organized by yiliu

The challenge is Track 3 at ICCV DeeperAction Challenge. This track is to recognize a human action by compositional learning ...

Jun 01, 2021-Sep 12, 2021

147 participants



Kinetics-TPS Challenge Test

#	User	Entries	Date of Last Entry	Score ▲
1	yuzheming	9	09/11/21	0.630532 (1)
2	Sheldong	10	09/11/21	0.613722 (2)
3	JosonChan	5	09/12/21	0.605059 (3)
4	fangwudi	4	09/05/21	0.590167 (4)
5	uestc.wxh	3	09/12/21	0.536067 (5)

1st Place Winner

Zheming Yu, Lin Li
Hikvision Research Institute

HIKVISION
海康威视

2nd Place Winner

Xiaodong Chen[1*] Xinchen Liu[2] Kun Liu[2] Wu Liu[2] Tao Mei[2]

[1]University of Science and Technology of China, Hefei, China

[2]JD AI Research, Beijing, China

中国科学技术大学

University of Science and Technology of China

JDAI Research

3rd Place Winner

Xuanhan Wang Xiaojia Chen Lianli Gao Lechao Cheng Jingkuan Song

Center for Future Media, University of Electronic Science and Technology of China, Chengdu, China
Zhejiang Lab, Hangzhou, China