

Kinetics-TPS Challenge on Part-level Action Parsing

Kunchang Li¹ Xiao Ma¹ Ding Xia¹ Dongliang Wang² Yali Wang¹ Yi Liu¹

Weihaio Gan² Jing Shao² Wei Wu² Junjie Yan² 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 INSTITUTES OF ADVANCED TECHNOLOGY
CHINESE ACADEMY OF SCIENCES



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

Kinetics-TPS Track Organizers



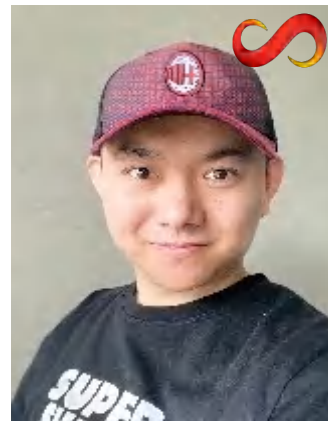
Kunchang Li



Xiao Ma



Ding Xia



Dongliang Wang



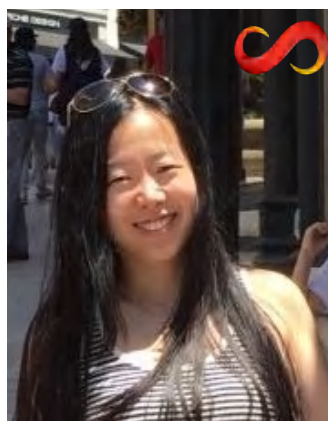
Yali Wang



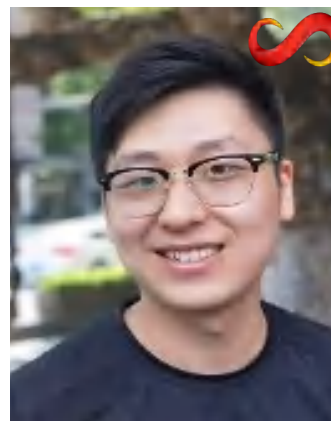
Yi Liu



Weihao Gan



Jing Shao



Wei Wu



Junjie Yan



Yu Qiao

Motivation

Dataset Introduction

Kinetics-TPS Competition

Why to do?

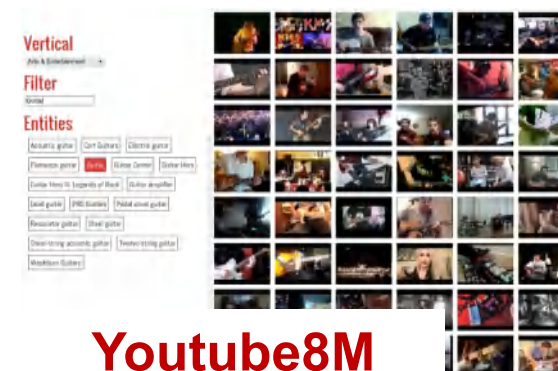
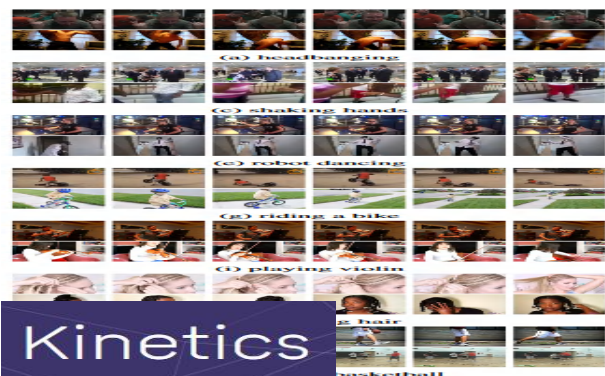
- Action recognition is treated as a high-level video classification



Deep Learning Classifiers

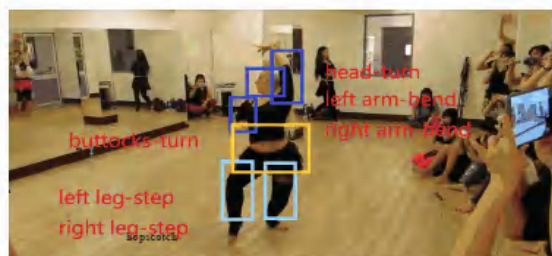
Pull_Ups

Large-Scale Video Benchmarks with Only Action Label



Why to do?

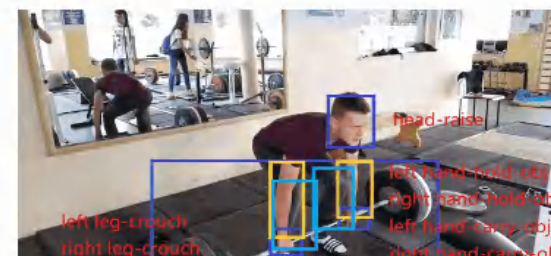
- Human action is spatio-temporal composition of body part state



belly dancing



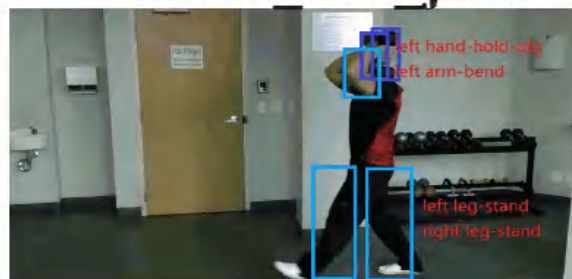
clean and jerk



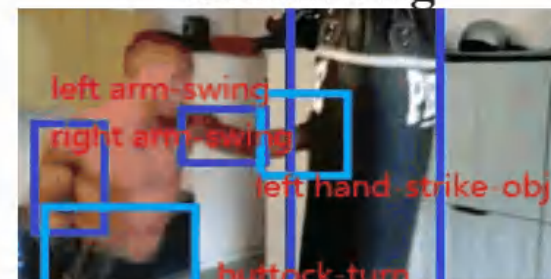
deadlifting



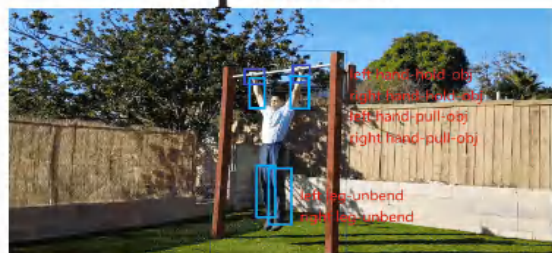
hopscotch



lunge



punching_bag



pull_ups



push_up



throwing_discus

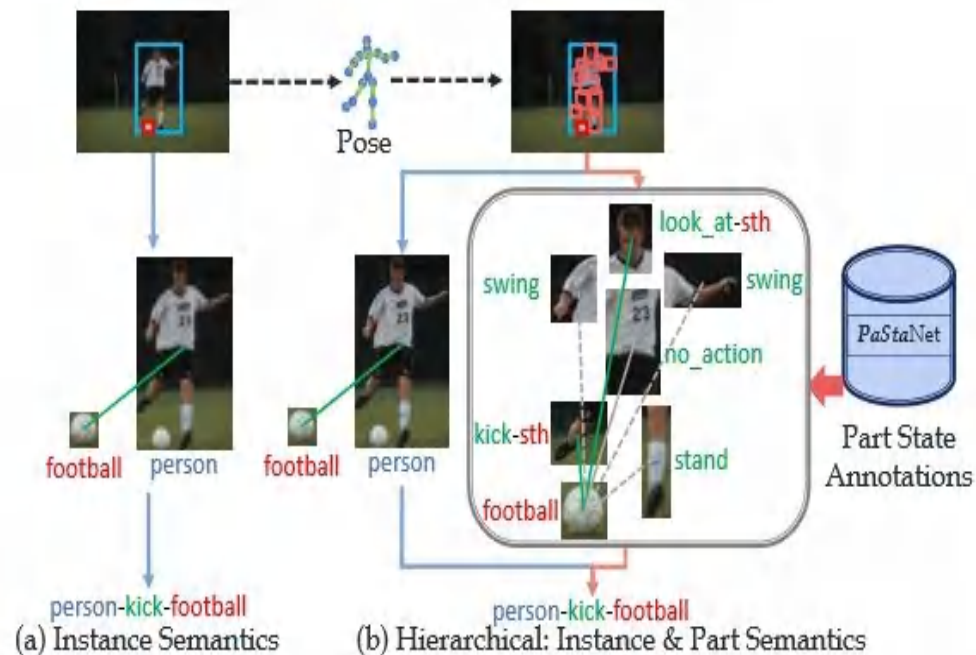
Why to do?

Action Genome (Without Body Part State)



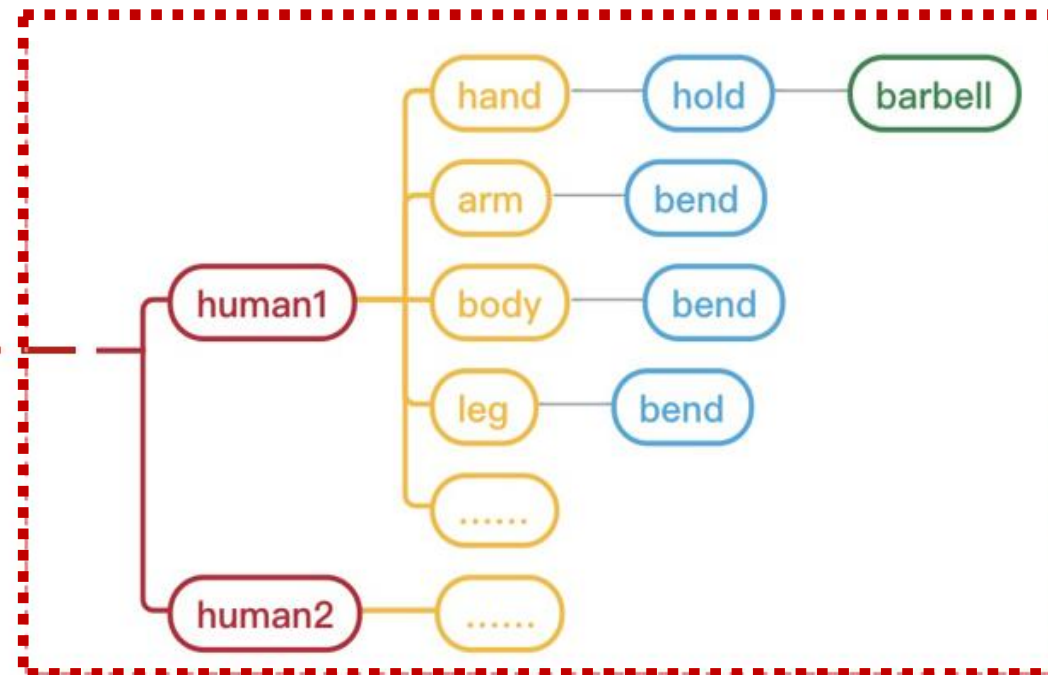
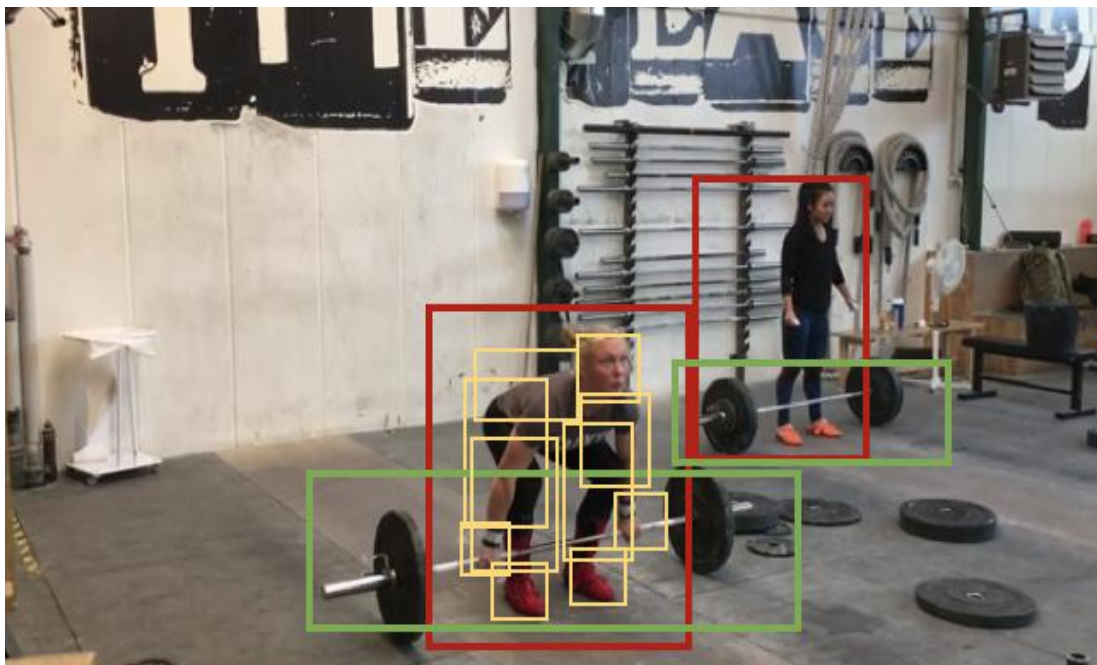
Ji et al., Action Genome: Actions as Composition of Spatio-temporal Scene Graphs, CVPR2020

HAKE (Image-based HOI)



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

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



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

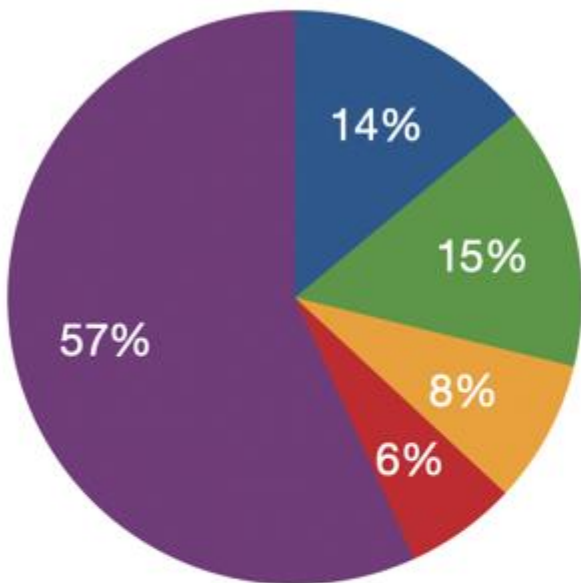
Body Part Annotation

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

Key Data Statistics

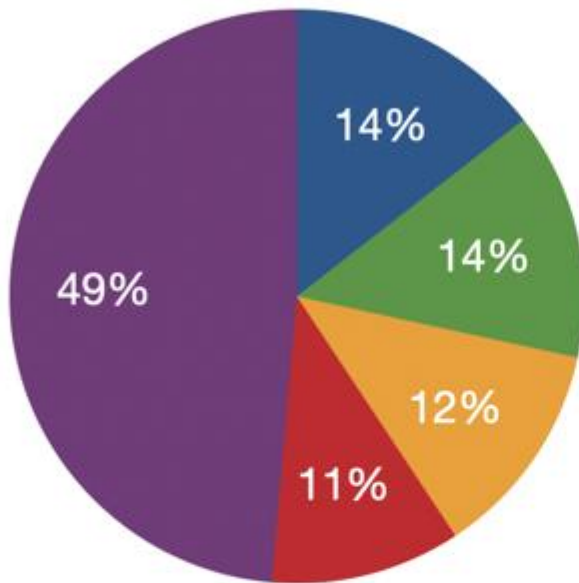
Rich diversity of (body part, part state) for various human actions

belly_dancing



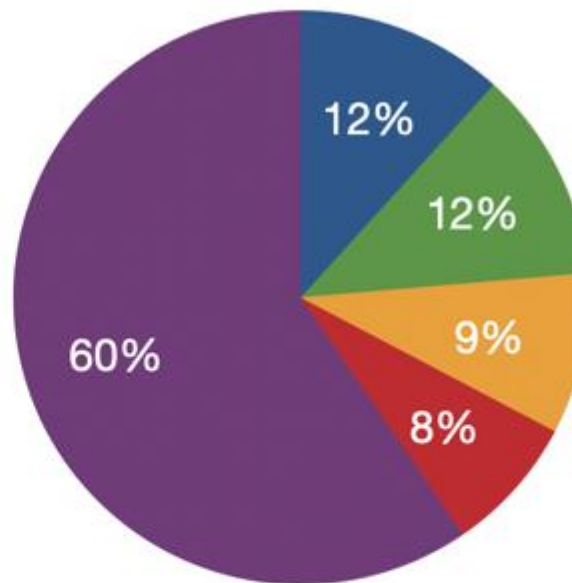
- hip, turn
- arm, bend
- arm, unbend
- leg, step
- other pairs

deadlifting



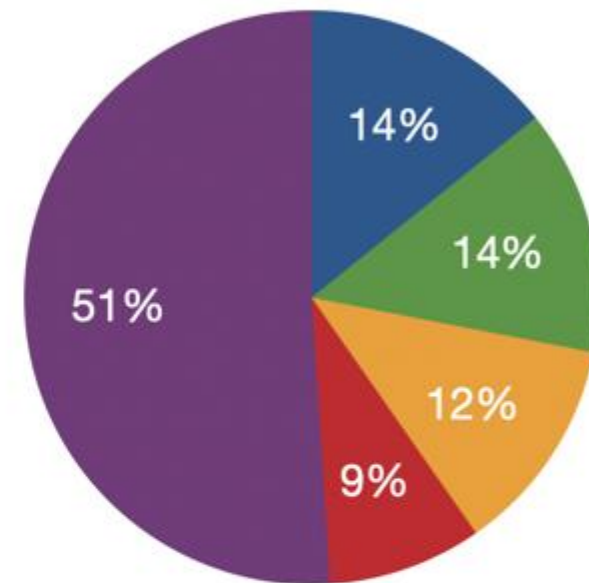
- hand, hold
- foot, step_on
- arm, carry
- leg, stand
- other pairs

hitting_baseball



- leg, step_on
- foot, stand
- arm, swing
- hand, hold
- other pairs

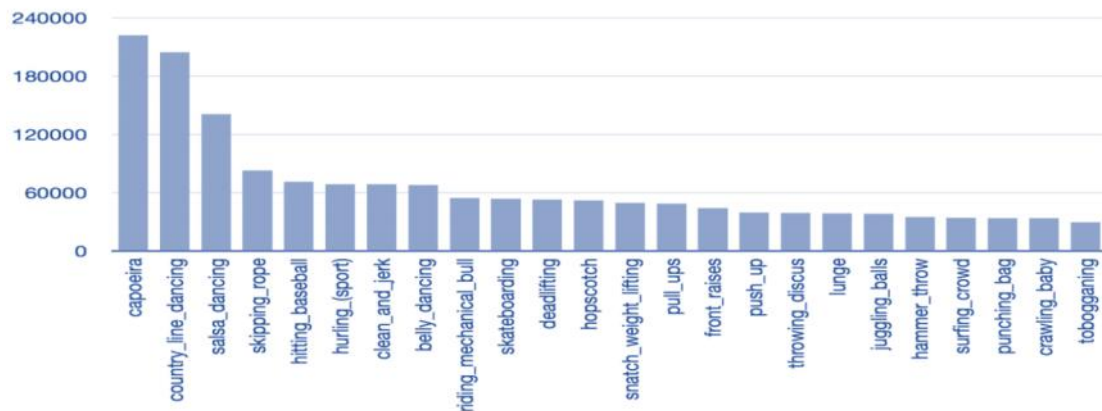
punching_bag



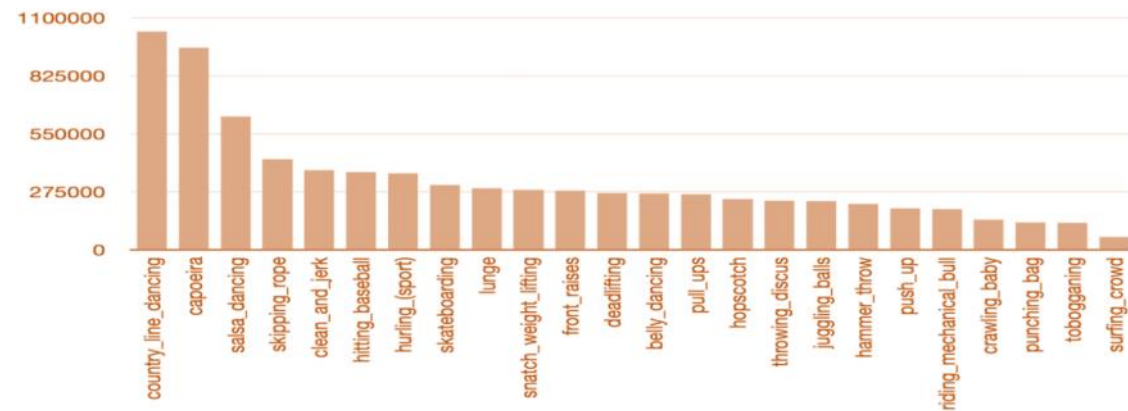
- arm, swing
- hand, strike
- leg, stand
- foot, step_on
- other pairs

Long-tailed distribution over all the levels of annotations

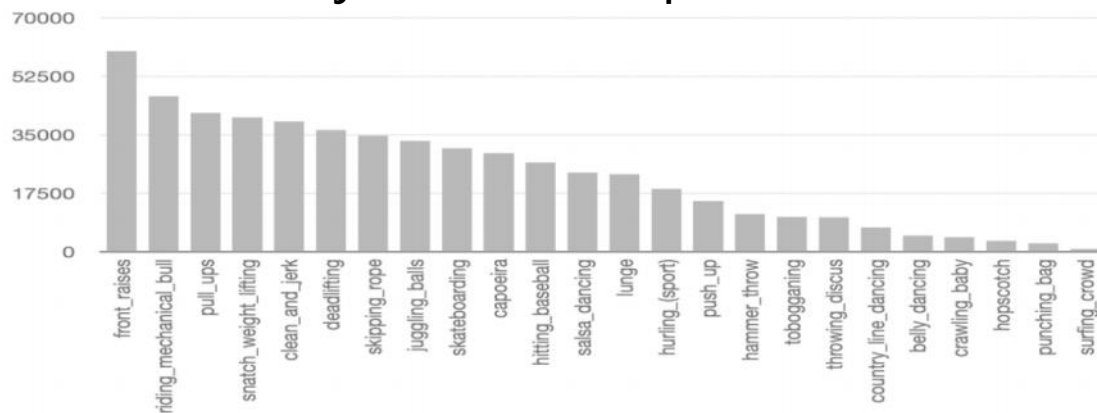
No. of human instances per action class



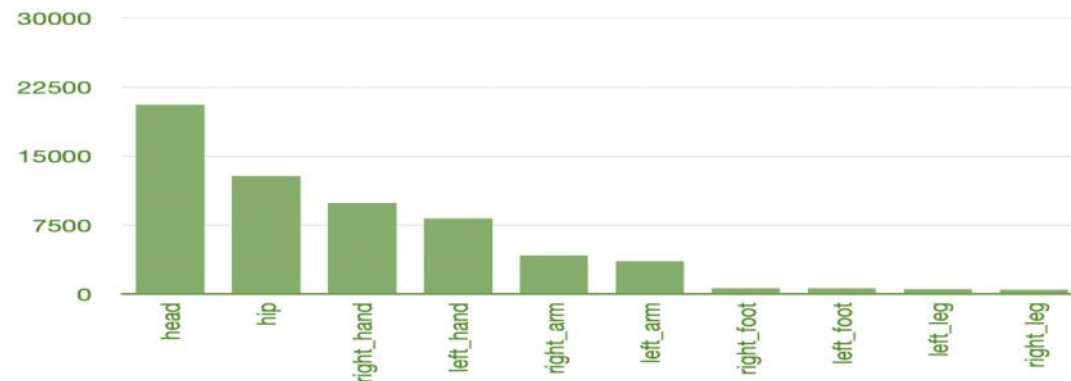
No. of part instances per action class



No. of object instances per action class

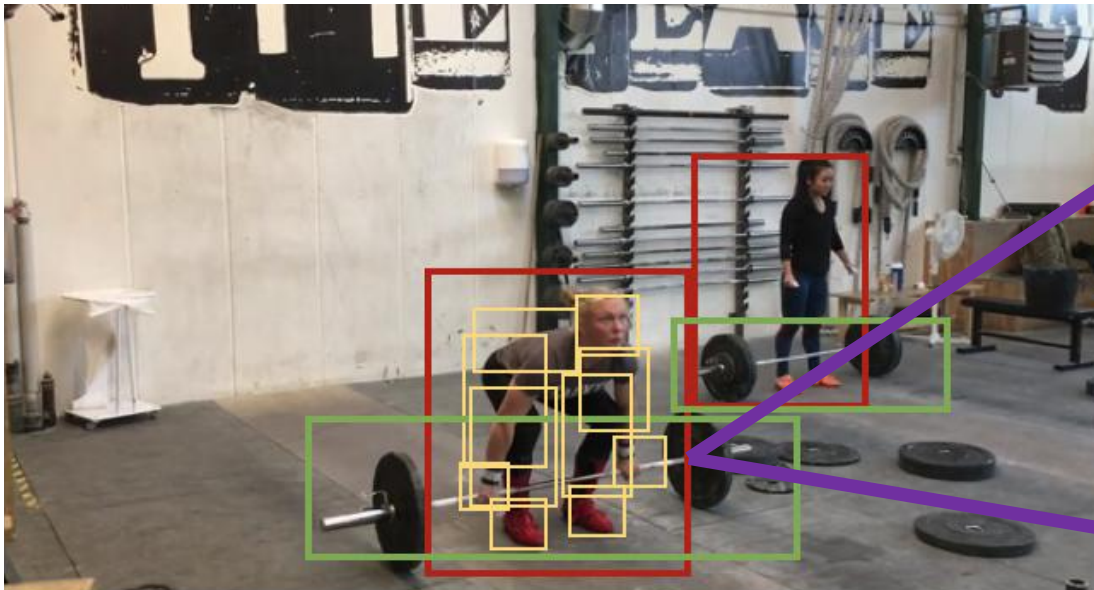


No. of part state annotations per body part

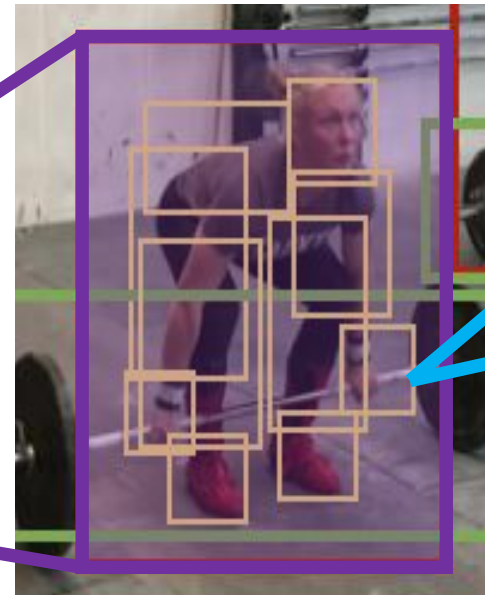


1) Part State Parsing

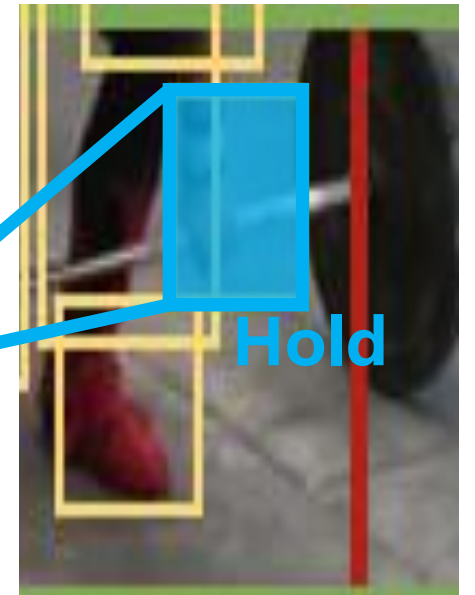
- Predicted boxes of human instances
- Predicted boxes of body parts & Predicted part state of each box



each sampled frame in a test video



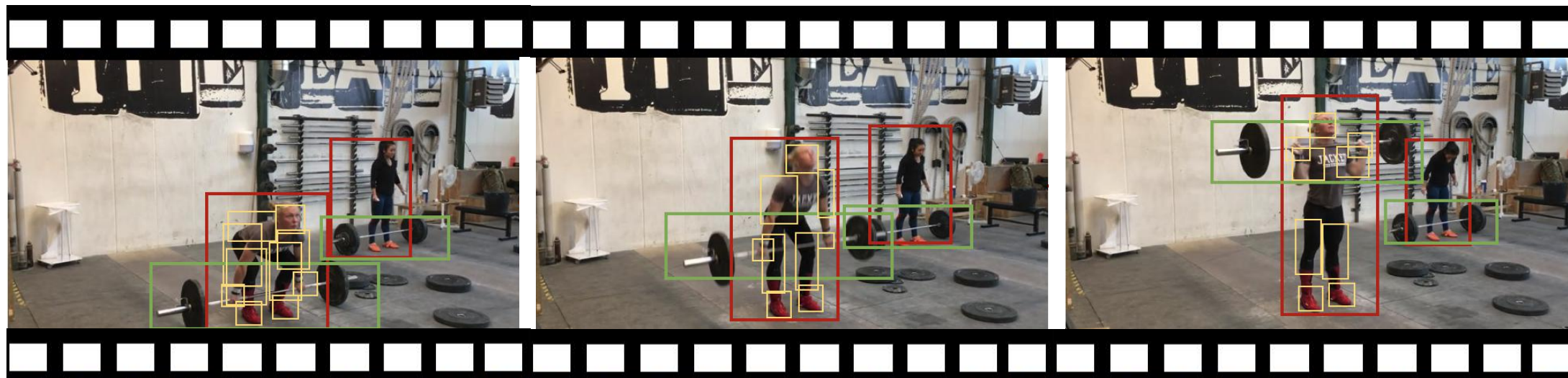
Human Box



Part Box & State

2) Action Recognition (for each test video)

- The predicted action label



deadlifting

Goal

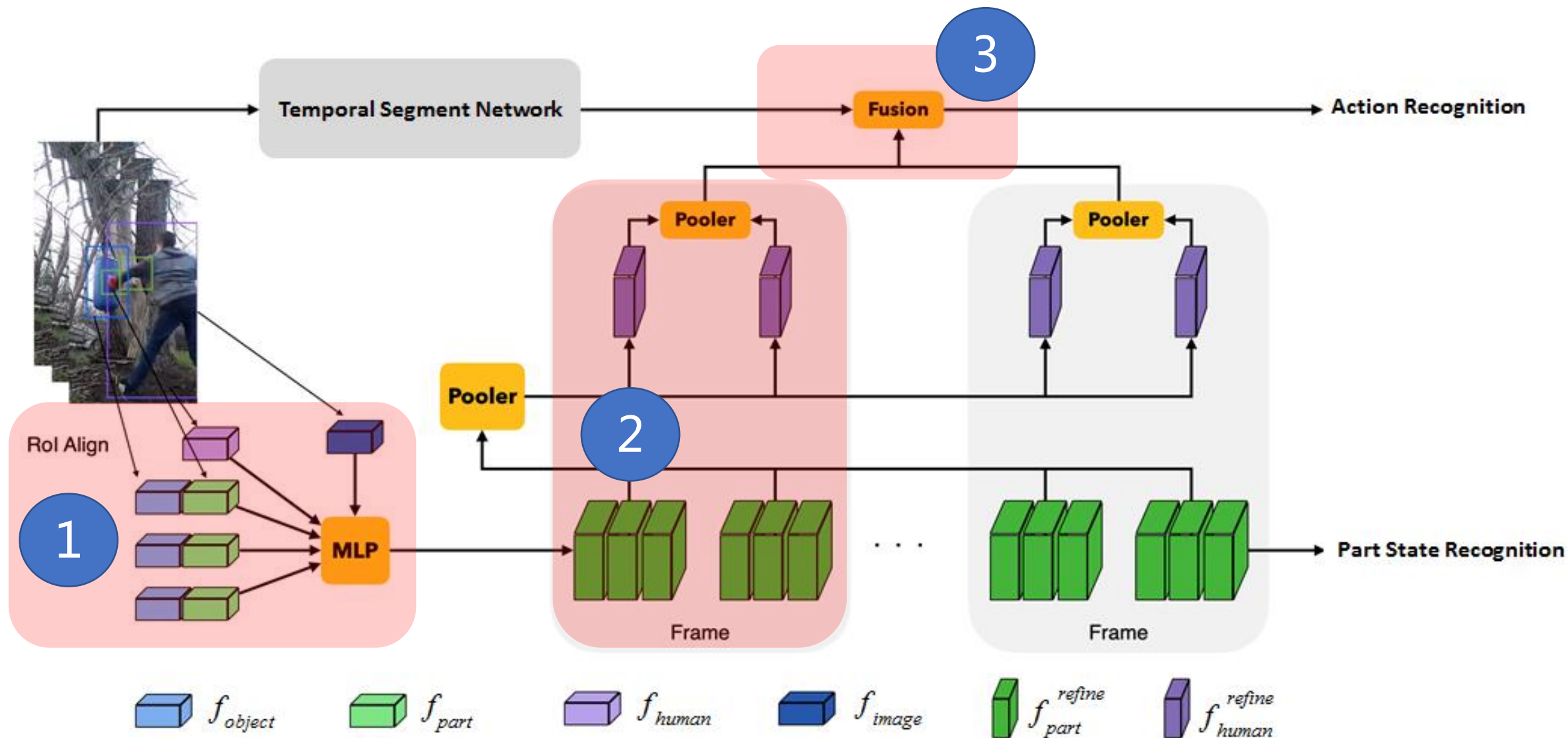
- Leveraging part state parsing for action recognition

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)

Kinetics-TPS Track: Baseline



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	OverWhelmingFlt	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)

1st Place Winner

Jiawei Dong¹, Yuliang Chen¹, Shuo Wang¹
¹Shanghai Paidao Intelligent Technology Co., Ltd.



Xiaojia Chen¹, Xuanhan Wang¹, Yan Dai¹, Jingkuan Song¹²

¹Center for Future Media, University of Electronic Science and
Technology of China, Chengdu, China

²Pengcheng Lab, Shenzhen, China



3rd Place Winner

Lianli Gao¹, Ji Zhang¹, Beitao Chen¹, Pengpeng Zeng¹

¹Center for Future Media, University of Electronic Science and
Technology of China, Chengdu, China

