Deeper Action

ECCV TEL AVIV 2022

DeeperAction Workshop @ ECCV2022
FineAction Track on Temporal Action Localization

Yi Liu[1] Limin Wang[2] Yali Wang[1] Xiao Ma[1] Yu Qiao[1]

[1] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[2] State Key Laboratory for Novel Software Technology, Nanjing University

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

南京大学
NANJING UNIVERSITY

Deeper Action

Outline

ECCV TEL AVIV 2022

Part 1: Dataset Introduction

Part 2: FineAction Competition

Deeper Action

Part 1

ECCV TEL AVIV 2022

Dataset Introduction

Deeper Action

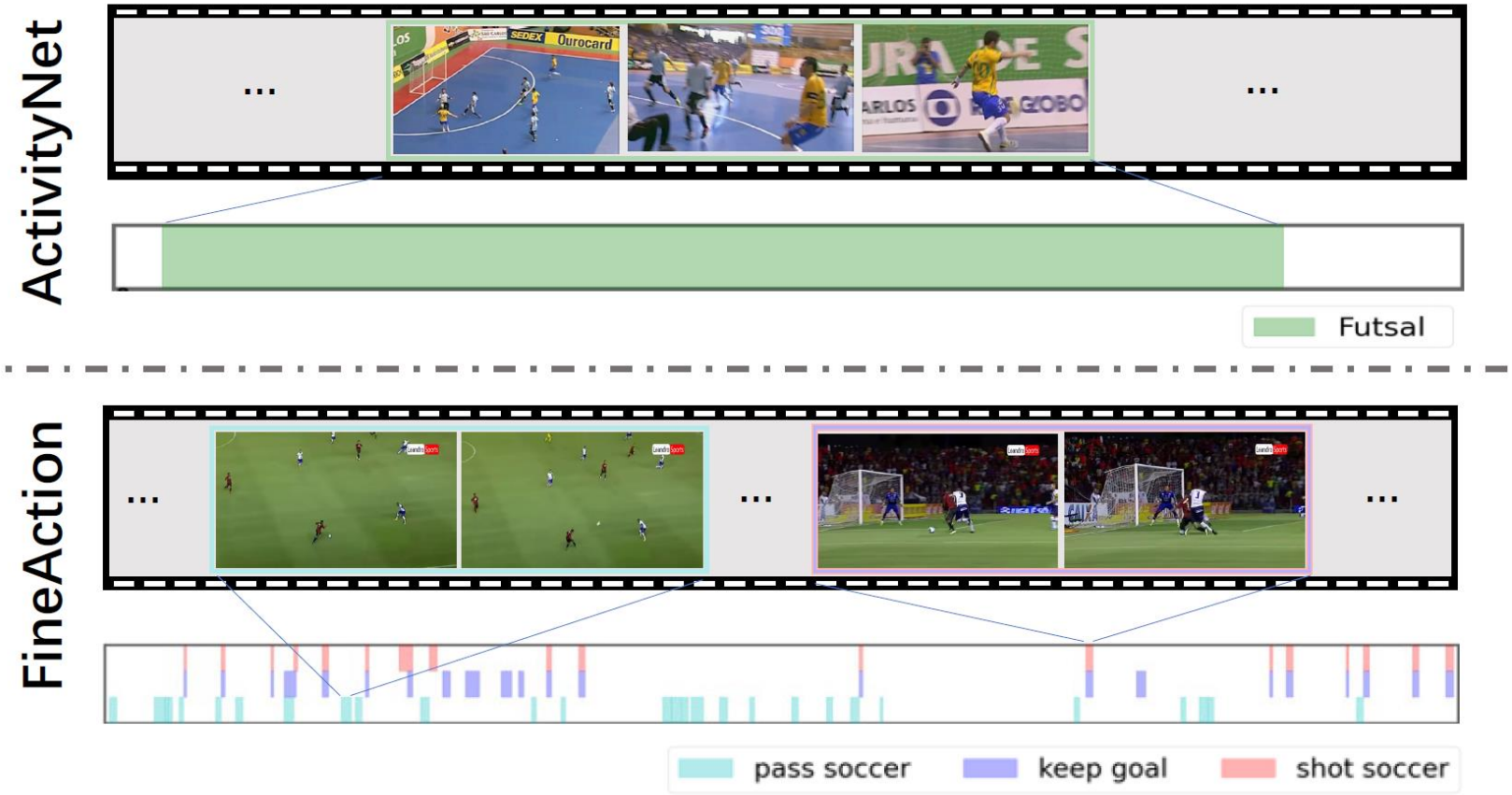
Why to do?

ECCV TEL AVIV 2022

ActivityNet

- ❑ *Sparse annotations*
- ❑ *Coarse-level action*

- ❑ **Dense annotations**
- ❑ **Fine-Grained action**



Deeper Action

Why to do?

ECCV TEL AVIV 2022

THUMOS-14

- ***Small scale***
- ***Specific domain***

- **Large scale**
- **Diversity domain**

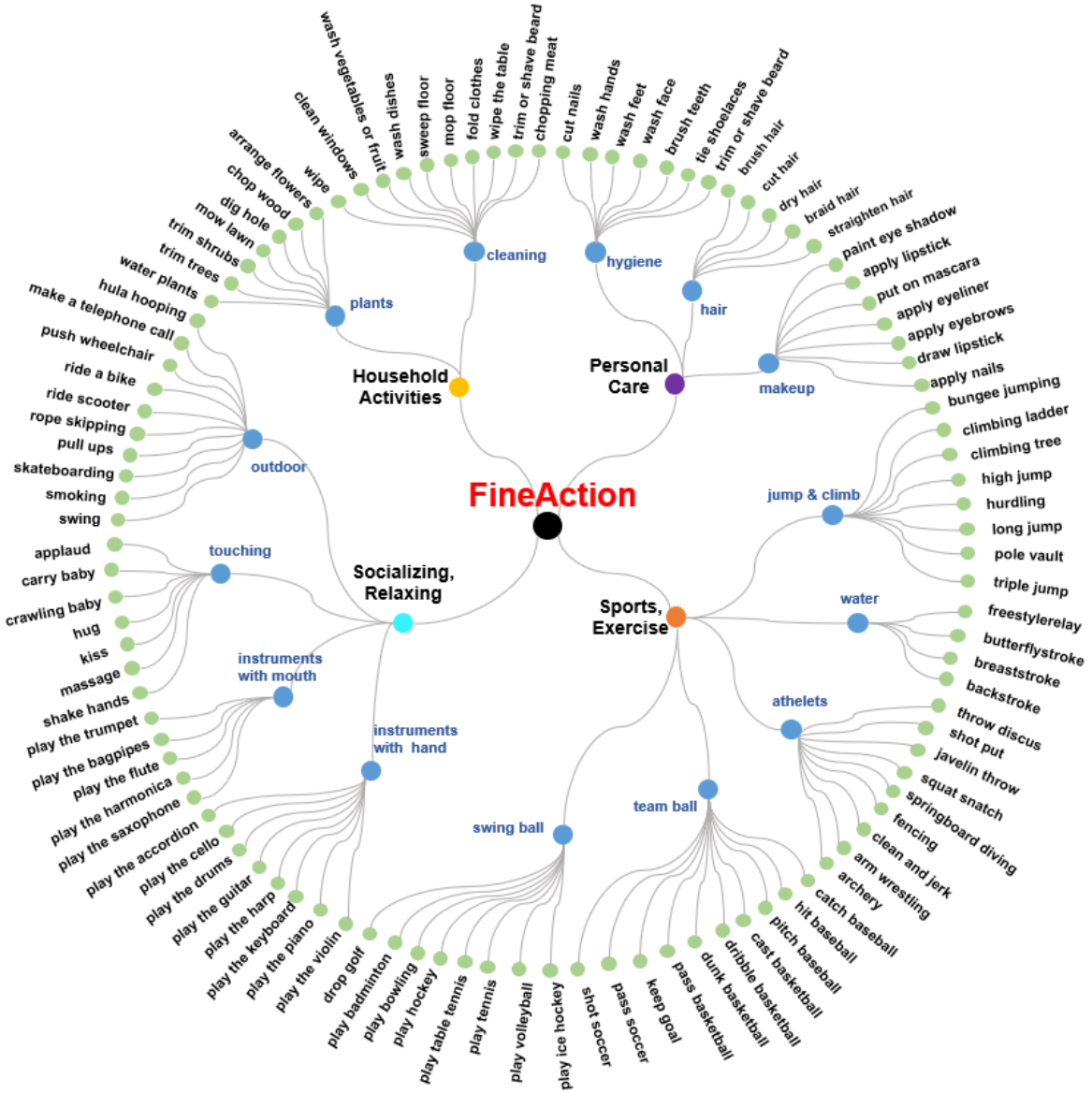
Deeper Action

FineAction: New Taxonomy

ECCV TEL AVIV 2022

FineAction

- ❑ 4 top-level categories
- ❑ 14 middle-level categories
- ❑ 106 bottom-level categories



Deeper Action

FineAction: Video Collection

ECCV TEL AVIV 2022

Video Selection

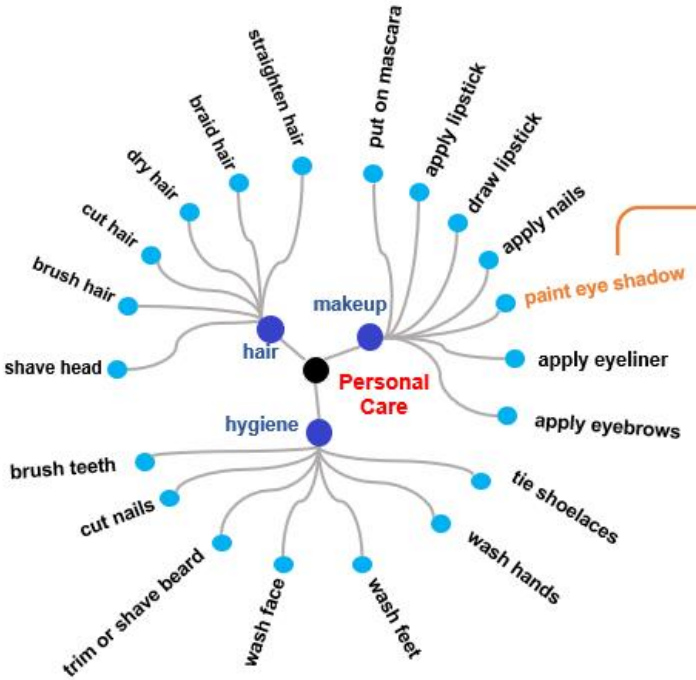
- ❑ **Exisiting video datasets**
  - *YouTube 8M*
  - *Kinetics 400*
  - *FCVID*
- ❑ **Crawl from YouTube videos**

Deeper Action

FineAction: Annotation

ECCV TEL AVIV 2022

Framework of Annotation Process

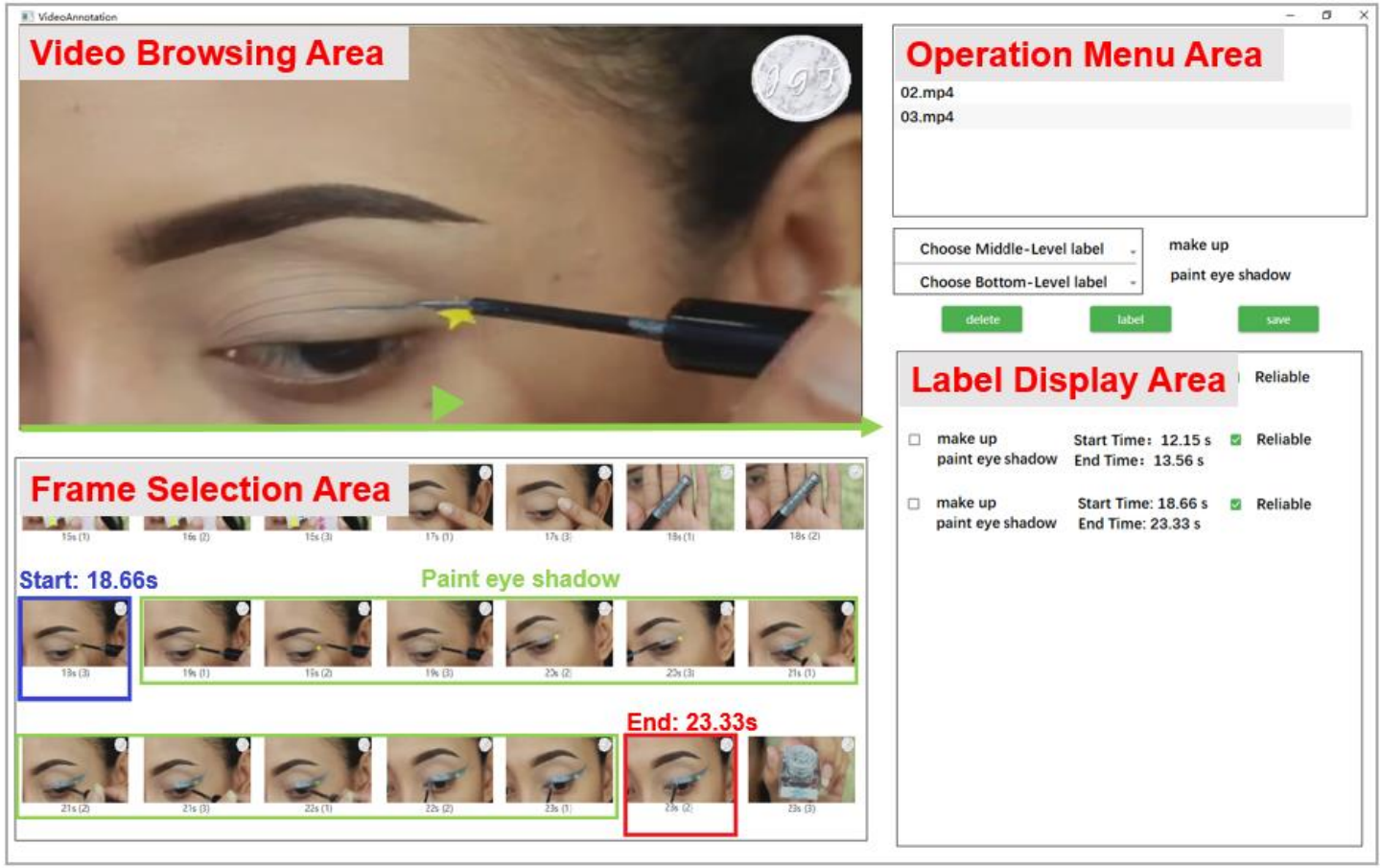


(a) Category Taxonomy

Start Definition:
Draw eye shadow with tools such as brushes.

End Definition:
Take off the tool.

(b) Annotation Guidance



(c) Annotation Tool

Deeper Action

FineAction: Statistics

ECCV TEL AVIV 2022

❑ Large-Scale & Fine-Grained

Database	Category	M-L	Video	Instance	Overlap	Duration	Action type	Main task
MPII Cooking [24]	65	✓	45	5,609	0.1%	11.1 m	kitchens	Action Classification
EPIC-Kitchens [26]	4,025	✓	700	89,979	28.1%	3.1 s	kitchens	Action Classification
FineGym V1.0 [23]	530	✓	303	32,697	0.0%	1.7 s	sports	Action Classification
THUMOS14 [14]	20	×	413	6,316	17.5%	4.3 s	sports	Temporal Action Localization
ActivityNet [15]	200	×	19,994	23,064	0.0%	49.2 s	daily events	Temporal Action Localization
HACS Segment [16]	200	×	49,485	122,304	0.0%	33.2 s	daily events	Temporal Action Localization
FineAction	106	✓	16,732	103,324	11.5%	7.1 s	daily events	Temporal Action Localization

❑ Instance duration

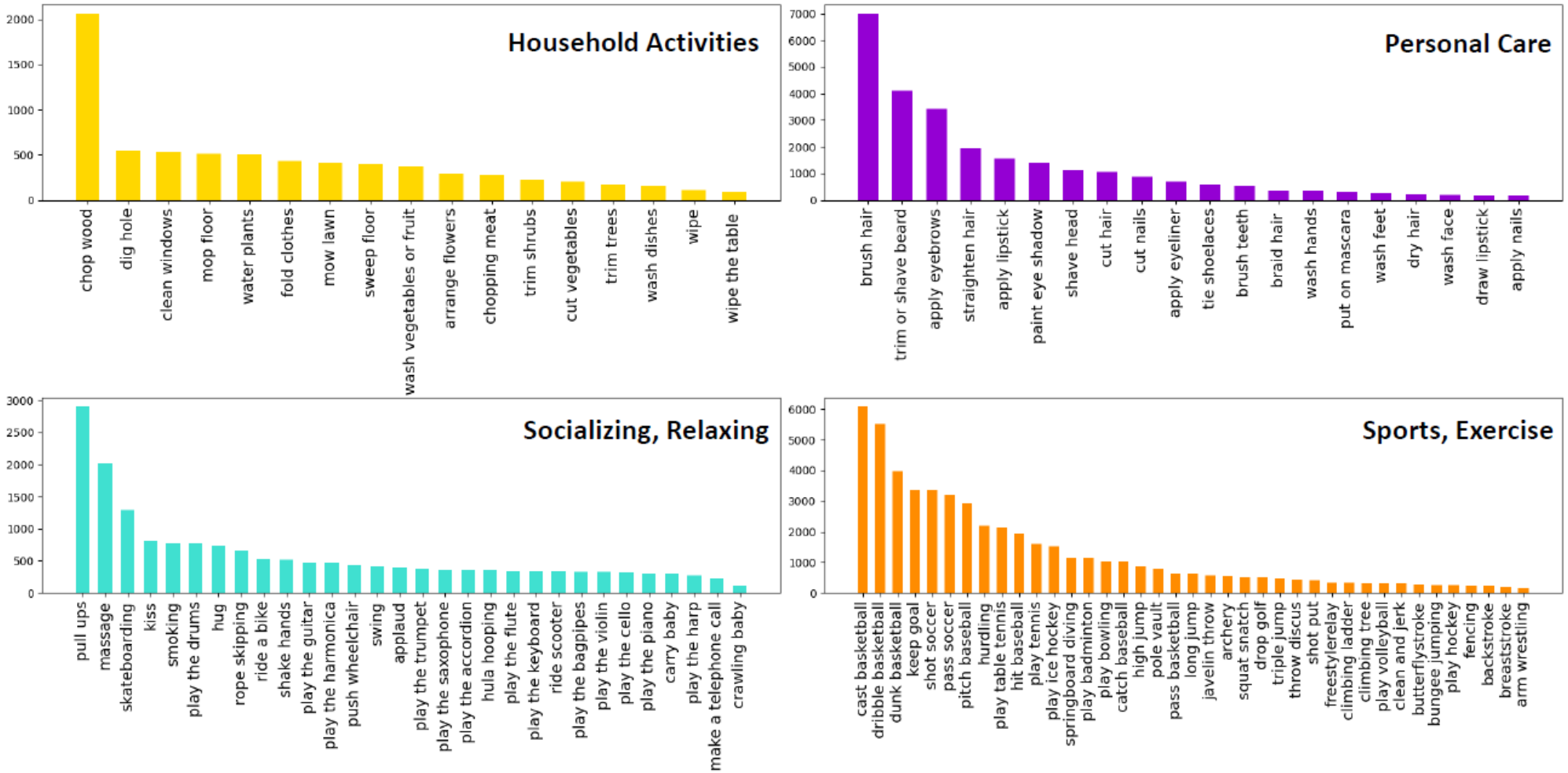
Database	0-2 s	2-6 s	6-15 s	>15 s	Ins / Vid
THUMOS14 [14]	2,029	2,753	1,437	99	15.29
ActivityNet [15]	900	3,253	4,426	14,485	1.15
HACS Segment [16]	8,874	29,644	31,982	51,804	2.47
FineAction	**66,890**	15,253	10,523	10,586	**6.17**

Deeper Action

FineAction: Statistics

ECCV TEL AVIV 2022

❑ Instance distribution



Deeper Action

FineAction: Experiment

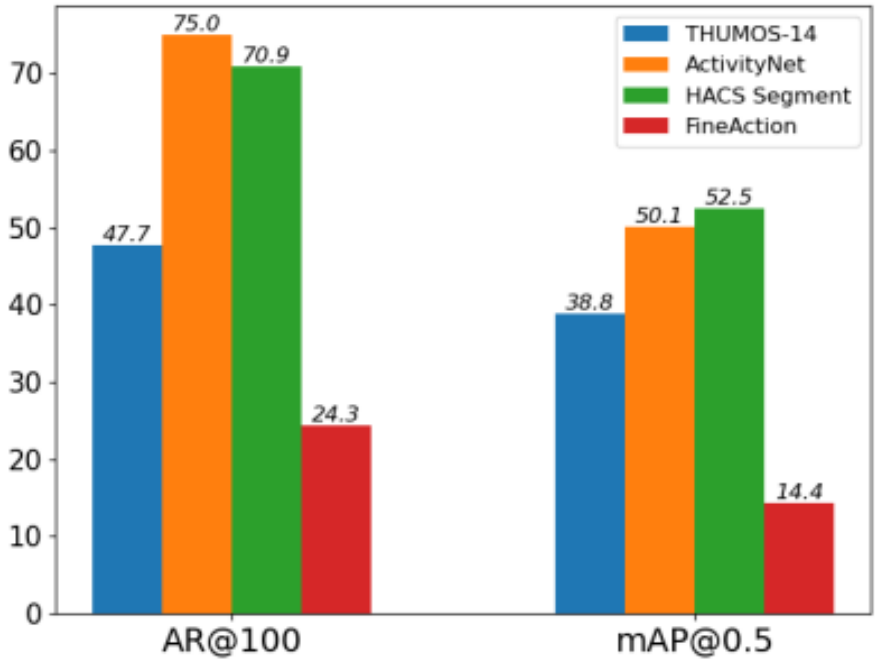
ECCV TEL AVIV 2022

❑ SOAT methods on FineAction

Method	Modality	Action Proposal Generation			
		AR@5	AR@10	AR@100	AUC
BMN [11]	RGB	8.62	11.20	22.74	17.49
	Flow	9.85	12.72	24.18	18.94
	RGB+Flow	9.99	12.84	24.34	19.19
DBG [13]	RGB	6.82	9.01	21.26	15.48
	Flow	8.27	10.90	23.37	17.70
	RGB+Flow	7.82	10.45	23.07	17.24
G-TAD [12]	RGB	7.96	10.45	20.86	16.06
	Flow	8.87	11.60	22.01	17.09
	RGB+Flow	9.02	11.83	23.17	17.65

Method	Modality	Temporal Action Localization			
		mAP@0.50	mAP@0.75	mAP@0.95	Avg.mAP
BMN [11]	RGB	12.56	7.49	2.62	7.86
	Flow	14.49	8.92	3.19	9.23
	RGB+Flow	14.44	8.92	3.12	9.25
DBG [13]	RGB	8.57	5.01	1.93	5.31
	Flow	11.03	6.95	2.70	7.20
	RGB+Flow	10.65	6.43	2.50	6.75
G-TAD [12]	RGB	10.88	6.52	2.19	6.87
	Flow	12.58	8.18	2.56	8.26
	RGB+Flow	13.74	8.83	3.06	9.06

❑ Benchmark comparison



Deeper Action

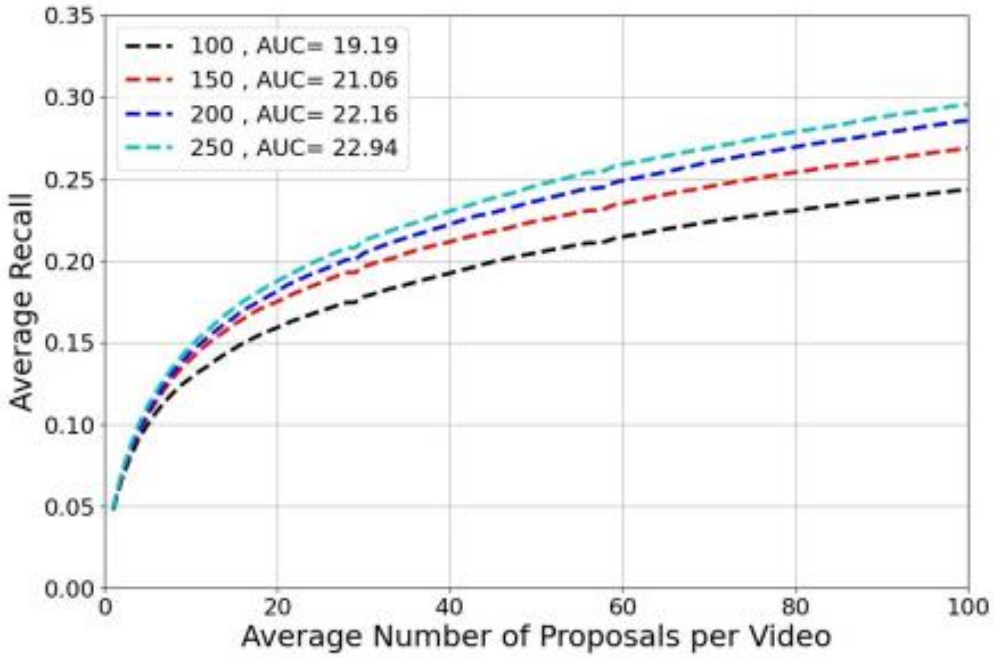
FineAction: Experiment

ECCV TEL AVIV 2022

❑ Cross-dataset evaluation

Dataset	Parameters Initial	AUC(%)
ActivityNet [15]	Scrath	67.29
FineAction	Scrath	19.19
ActivityNet [15]	FineAction	63.17 (↓ 4.12)
FineAction	ActivityNet [15]	18.42 (↓ 0.77)

❑ Different sequence length

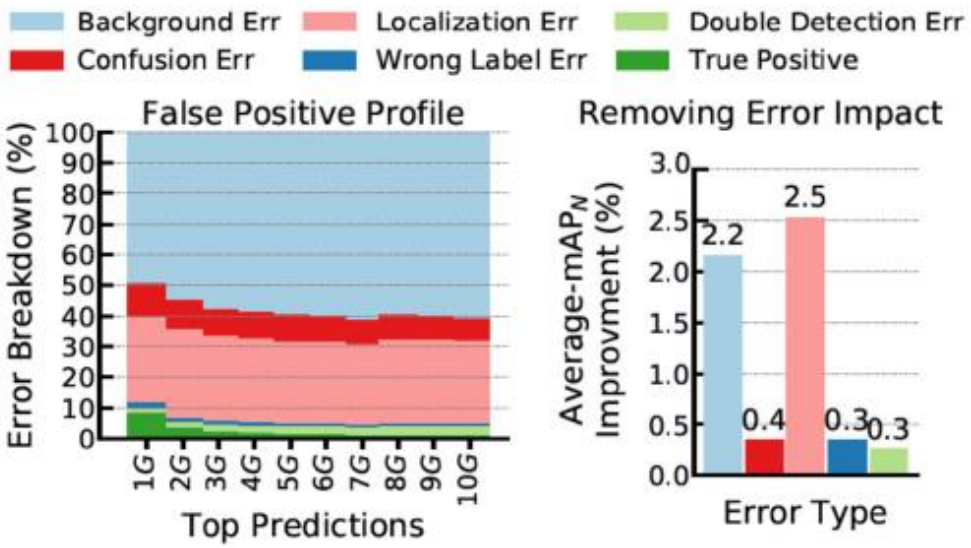


Deeper Action

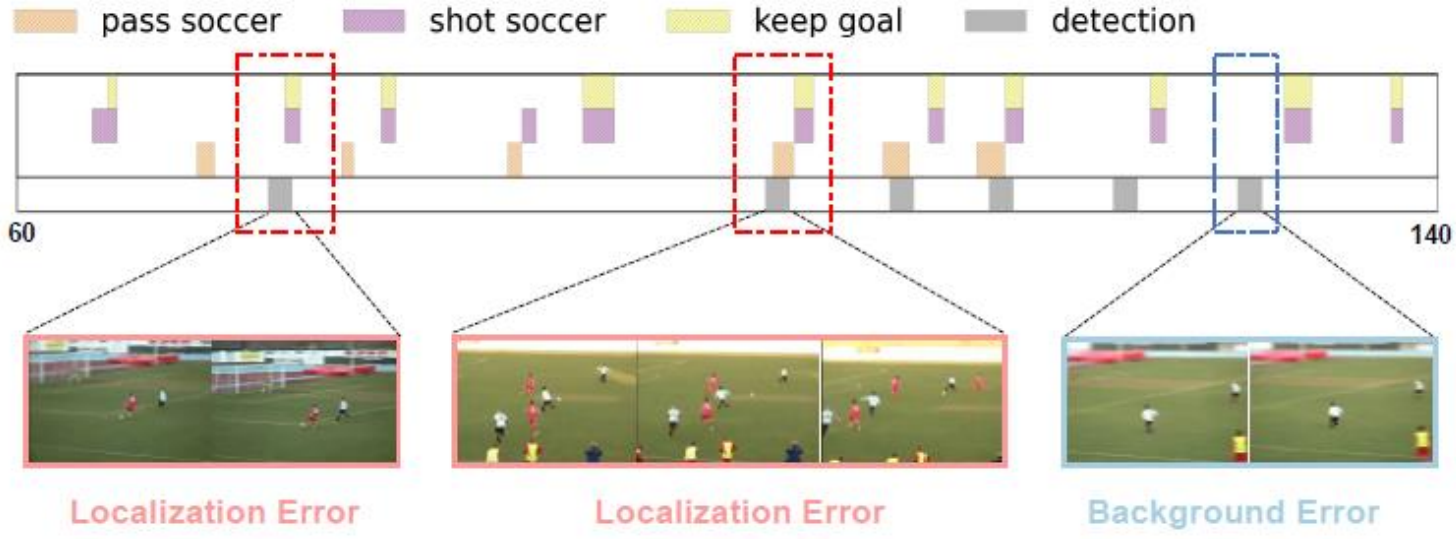
FineAction: Experiment

ECCV TEL AVIV 2022

❑ Error analysis & visualization



(a) Error Analysis.



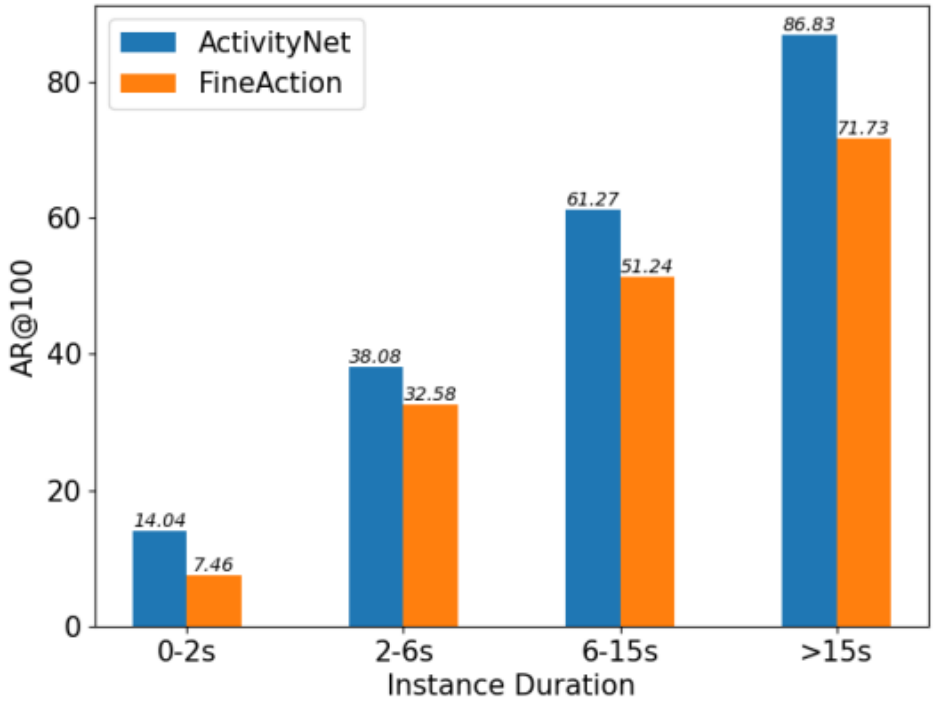
(b) Visualization of Typical Errors.

Deeper Action

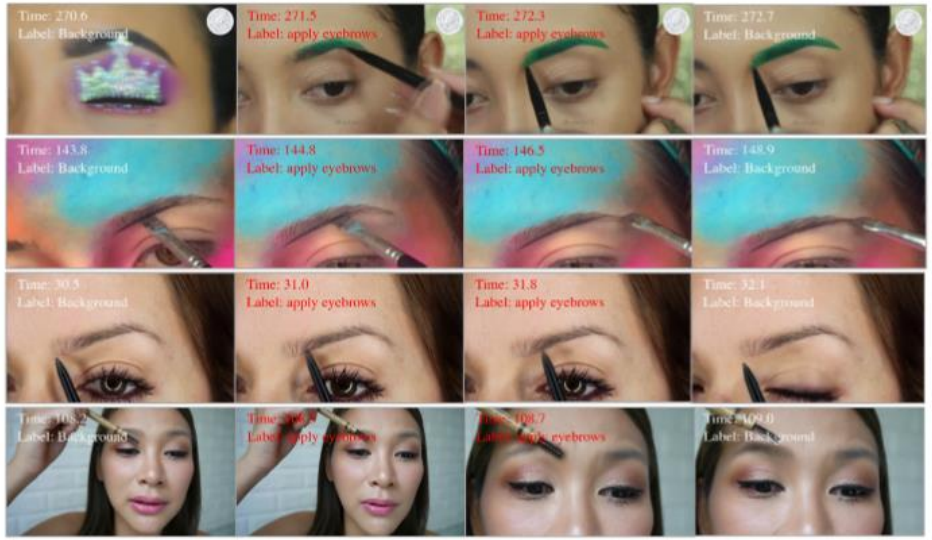
FineAction: Experiment

ECCV TEL AVIV 2022

❑ Challenging categories



FineAction	AR@100	Avg. Duration	Num
straighten hair	1.6	1.17 s	1934
apply eyebrows	1.7	1.13 s	3428
dig hole	2.2	0.83 s	544
freestyle relay	88.4	53.12 s	345
breaststroke	94.5	53.59 s	211
play the harp	96.2	69.21 s	268



Deeper Action

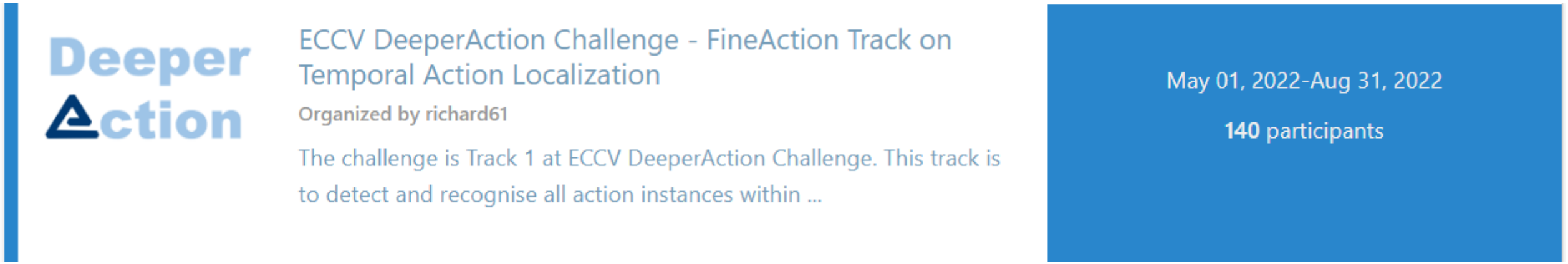
Part 2

ECCV TEL AVIV 2022

FineAction Competition

Competition: FineAction Track

https://codalab.lisn.upsaclay.fr/competitions/4386



- **Validation phase:** 2022.05.01-2022.08.15
- **Testing phase:** 2022.08.16-2022.08.31

Deeper Action

FineAction Track: Evaluation

ECCV TEL AVIV 2022

❑ Goal

The challenge requires detecting and recognizing all-action instances within an untrimmed video.

❑ Metric

Mean Average Precision (mAP) is a conventional evaluation metric, where Average Precision (AP) is calculated for each action category with tIoU thresholds [0.5:0.05:0.95].

Deeper Action

FineAction Track: Statistics

ECCV TEL AVIV 2022

- Valid Participants : 137
- Valid Teams: 8+9 (Company, University, Institute)

北京交通大學 BEIJING JIAOTONG UNIVERSITY

中国科学院 自动化研究所 INSTITUTE OF AUTOMATION CHINESE ACADEMY OF SCIENCES

中山大學 SUN YAT-SEN UNIVERSITY

oppo

中国科学技术大学 University of Science and Technology of China

華中科技大學

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

南京理工大学 NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

Agency for Science, Technology and Research SINGAPORE

Deeper Action

FineAction Track: Results

ECCV TEL AVIV 2022

- Valid Submission: 58 (Val Phase) + 84 (Test Phase)

Test Set (Mean Average Precision - mAP)													
#	User	Entries	Date of Last Entry	mAP@0.50 ▲	mAP@0.55 ▲	mAP@0.60 ▲	mAP@0.65 ▲	mAP@0.70 ▲	mAP@0.75 ▲	mAP@0.80 ▲	mAP@0.85 ▲	mAP@0.90 ▲	mAP@0.95 ▲	Avg.mAP ▲
1	yptang	24	08/31/22	35.95 (1)	33.48 (1)	30.64 (1)	27.89 (1)	24.92 (1)	21.78 (1)	18.25 (1)	14.25 (1)	9.00 (1)	3.52 (3)	21.97 (1)
2	linxt	12	08/25/22	28.55 (2)	26.14 (2)	23.78 (2)	21.35 (2)	18.67 (2)	15.91 (2)	12.85 (3)	9.54 (4)	5.93 (4)	2.16 (4)	16.49 (2)
3	Strangelove	14	08/31/22	25.50 (3)	23.58 (3)	21.66 (3)	19.46 (3)	17.59 (3)	15.70 (3)	13.74 (2)	11.35 (2)	8.36 (2)	4.25 (2)	16.12 (3)

Deeper Action

1st Place Winner

ECCV TEL AVIV 2022

A One-Stage Method for FineAction Localization from Multiple Views

Yepeng Tang[1,2†], Weining Wang[3], Chunjie Zhang[1,2*], Jie Jiang[3,4], Weitao Yuan[3,4], Sihan Chen[3,4], Jing Liu[3,4], Yao Zhao[1,2]

[1] Institute of Information Science, Beijing Jiaotong University
[2] Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing Jiaotong University
[3] Institute of Automation, Chinese Academy of Sciences
[4] School of Artificial Intelligence, University of Chinese Academy of Sciences

北京交通大學
BEIJING JIAOTONG UNIVERSITY

中国科学院
自动化研究所
INSTITUTE OF AUTOMATION
CHINESE ACADEMY OF SCIENCES

Deeper Action

2nd Place Winner

ECCV TEL AVIV 2022

Technical Report for FineAction 2022 - Temporal Action Localization

Yaokun Zhong , Xiaotong Lin
Sun Yat-sen University, Guangzhou, China

Deeper Action

3rd Place Winner

ECCV TEL AVIV 2022

Technical Report for FineAction 2022 - Temporal Action Localization

Shimin Chen[1] Yijia Duan[1,2] Wei Li[1] Changlong Li[1,2] Chen Chen[1]
[1]OPPO Research Institute. [2]Shanghai Jiao Tong University.
{chenshimin1, liwei19, chenchen}@oppo.com
{yjduan, hurray-long}@sjtu.edu.cn

oppo

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Deeper Action

ECCV TEL AVIV 2022

Over & Thank you !

Homepage: https://deeperaction.github.io/datasets/fineaction

GitHub: https://github.com/Richard-61/FineAction

IEEE TRANSACTIONS ON IMAGE PROCESSING

FineAction: A Fine-Grained Video Dataset for Temporal Action Localization

Yi Liu, Limin Wang *Member, IEEE*, Yali Wang, Xiao Ma, Yu Qiao *Senior Member, IEEE*