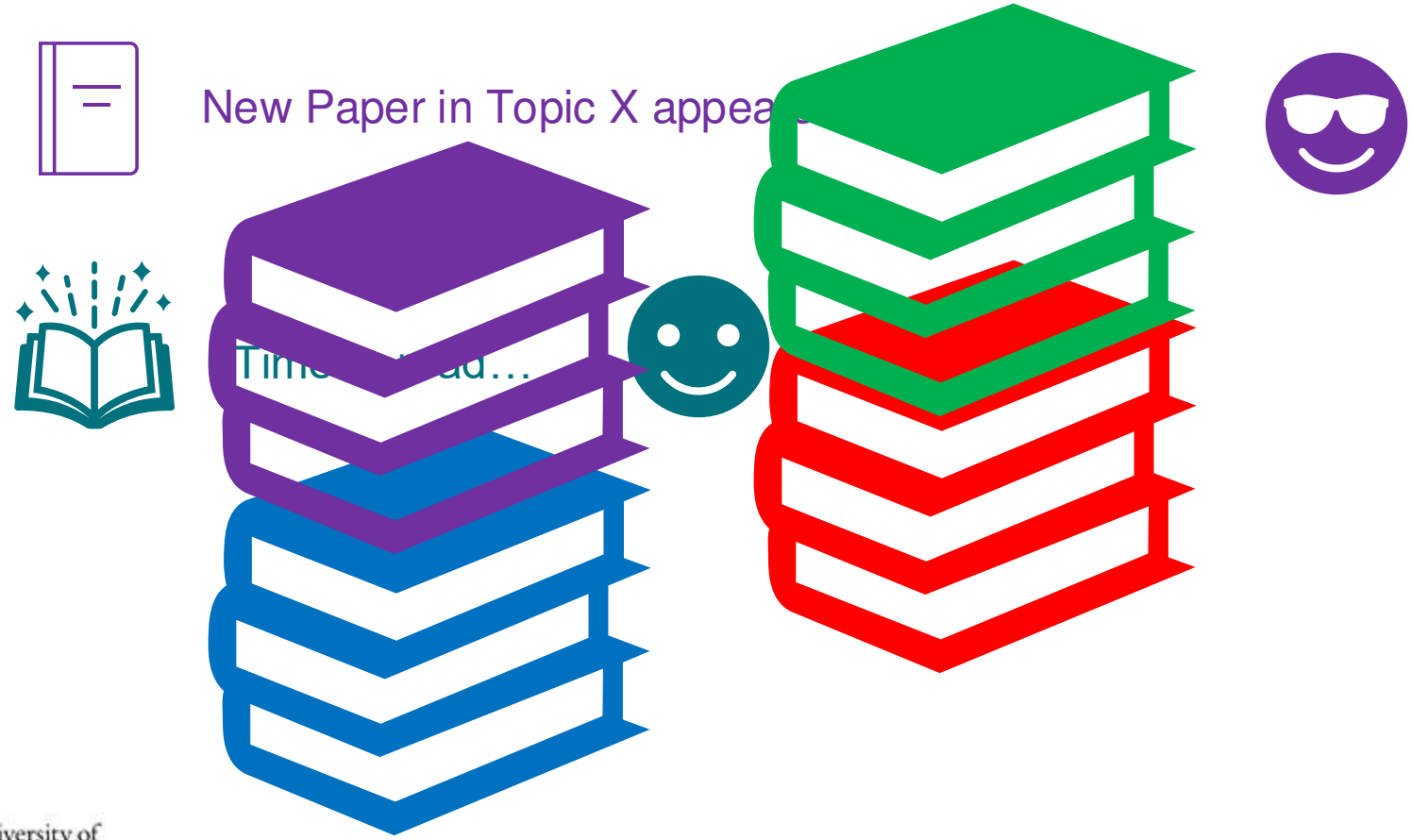
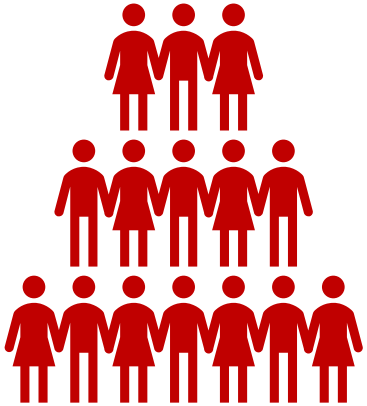
More Exploration, Less Exploitation…

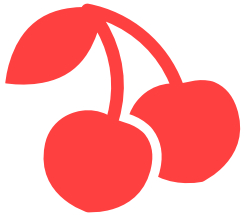
Current Status of Computer Vision



Why is this happening??



Community is very large

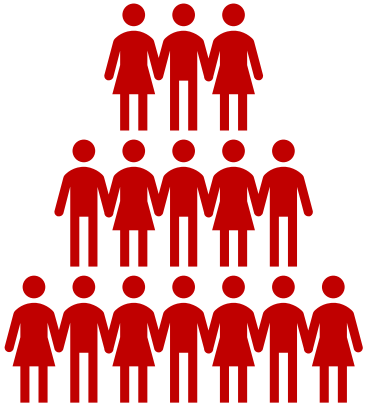


Low-Hanging Fruit

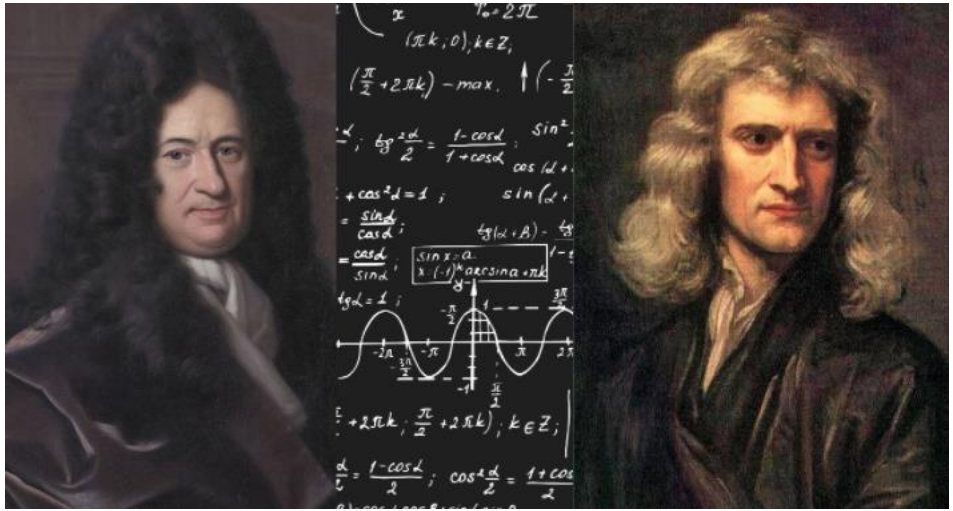
Current Wave

FOMO

Why is this happening??



Community is very large

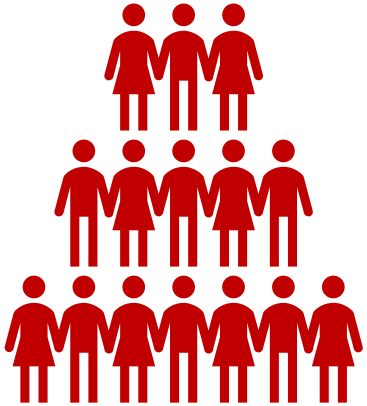


Leibniz
(1684)

Newton

I have not yet seen the book published against me being at Vienna which is in the furthest part of Germany where such books come very slowly

Why is this happening??

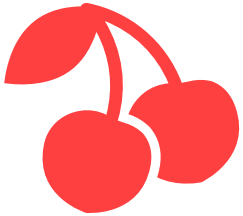


Community is very large

Solutions:

- Keep/Showcase Evidence of Work Progress
- Acknowledge others -- parallel/recent work
- Accept -- some will stay bitter
- Collaborate…

Why is this happening??



Low-Hanging Fruit

Solution:

- This is a low-hanging fruit (obvious) → Reject

Why is this happening??

Current Wave

A jack of all trades is a master of none

My research interests include a wide range of topics within computer vision, Machine Learning and Data Science, including but not limited to ….

I am interested in motion. What does motion tell us about the structure of the world and how can we compute this from video? How do humans and animals move? What goals drive behavior?

Why is this happening??

Current Wave

Solution:

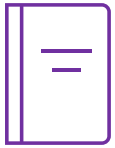
“A smooth sea never made a skilled sailor”

Why is this happening??

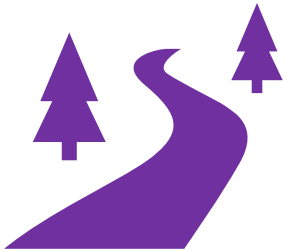
FOMO

Seek Support

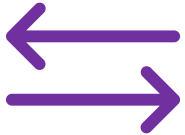
What should you do?



New Paper in Topic X appears on ArXiv



In your path



Paradigm Shift

If it really matters,
will be mentioned again

What should you do?

Stay Focused….

Impactful Research Takes Years…

Credit is rarely given to the original individuals

Ex… Wearable Cameras

1907

1970s

..

1987

2004
GoPro

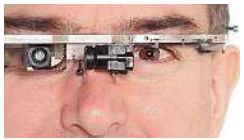
Ex… Wearable Cameras

1907

1970s

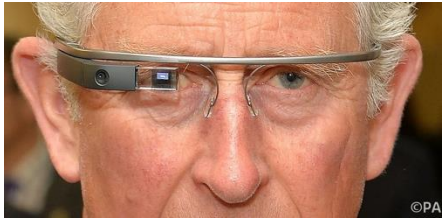
1987

1995



1999
Eyetrap

2004
GoPro



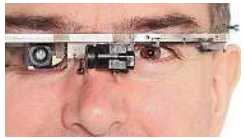
2013
Google
Glass

2024

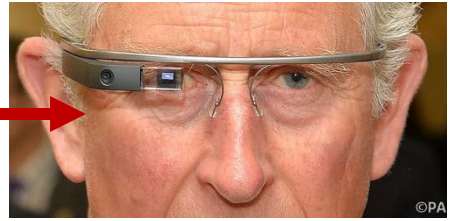
Ex… Wearable Cameras

https://www.youtube.com/watch?v=fCco6FMCRmk

Ex… Wearable Cameras



14 years



1907

70s

1999
Eyetrap

2013
Google
Glass

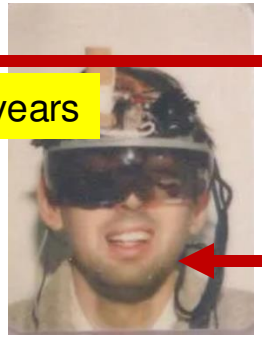
1987

1995

2004
GoPro

2024

17 years



29 years

How to Explore?

Data-Driven Research

Keep a file of your ideas

Grow a Team

Imagine the future

How to Explore?

Data-Driven Research

Data Collection

EPIC KITCHENS

2017 - now

100 hours
45 kitchens
4 countries
Long-term recording
Kitchen-based activities

EGO 4D

2020 - now

6730 hours
923 participants
74 locations
9 countries
Short-term recording
All daily activities

Data Collection Exercises

EGO-EXO4D

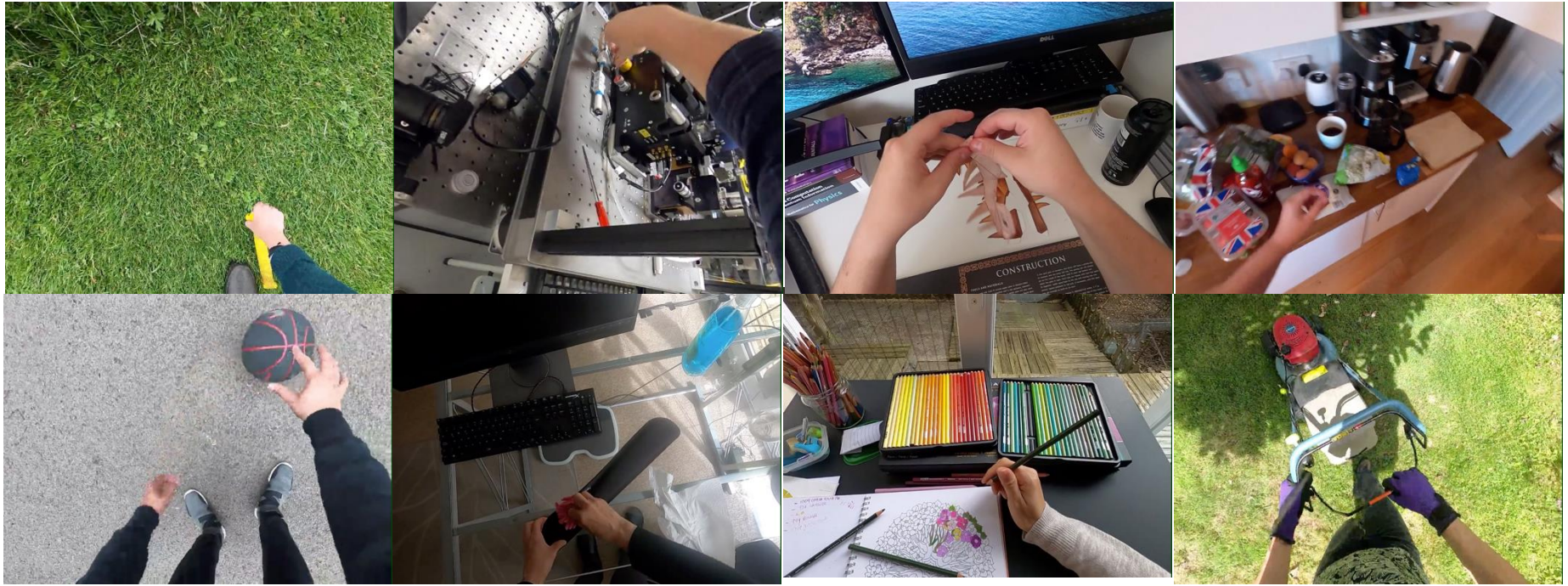
2022 - now

Released Dec 2023
1422 hours
8 skilled activities
839 camera wearers
Ego-Exo recordings

2024 – [coming]

[new recordings]

Egocentric Videos



Egocentric Videos

How to Explore?

Keep a file of your ideas

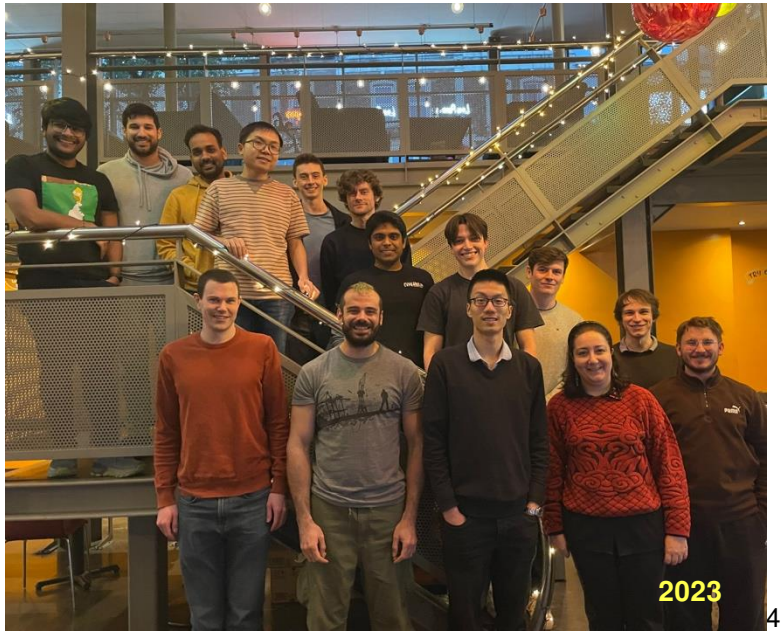
Sometimes you need to wait for the right time…

How to Explore?

Your most valuable output is your PhD students / postdocs….
NOT your papers!

Grow a Team

The Team



How to Explore?

Imagine the future

Share it when ready…

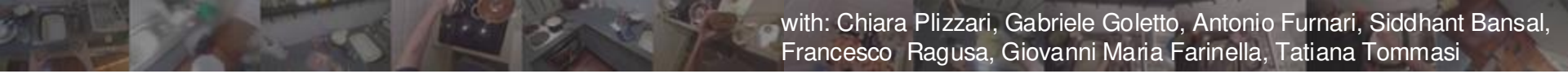
An Outlook into the Future of Egocentric Vision

Chiara Plizzari*, Gabriele Goletto*, Antonino Furnari*, Siddhant Bansal*, Francesco Ragusa*, Giovanni Maria Farinella†, Dima Damen†, Tatiana Tommasi†

Politecnico di Torino

University of BRISTOL

UNIVERSITÀ degli STUDI di CATANIA



with: Chiara Plizzari, Gabriele Goletto, Antonio Furnari, Siddhant Bansal, Francesco Ragusa, Giovanni Maria Farinella, Tatiana Tommasi

Envisioning an Ambitious Future and **Analysing** the Current Status of Egocentric Vision

How did we do this?

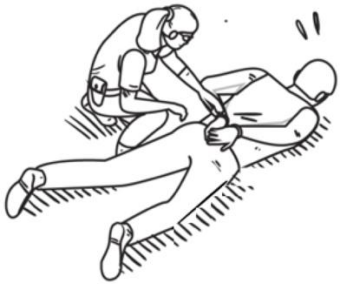
We imagined a device – *EgoAI* and envisioned its utility in multiple scenarios

EGO-Designer

EGO-Worker

EGO-Tourist

EGO-Home

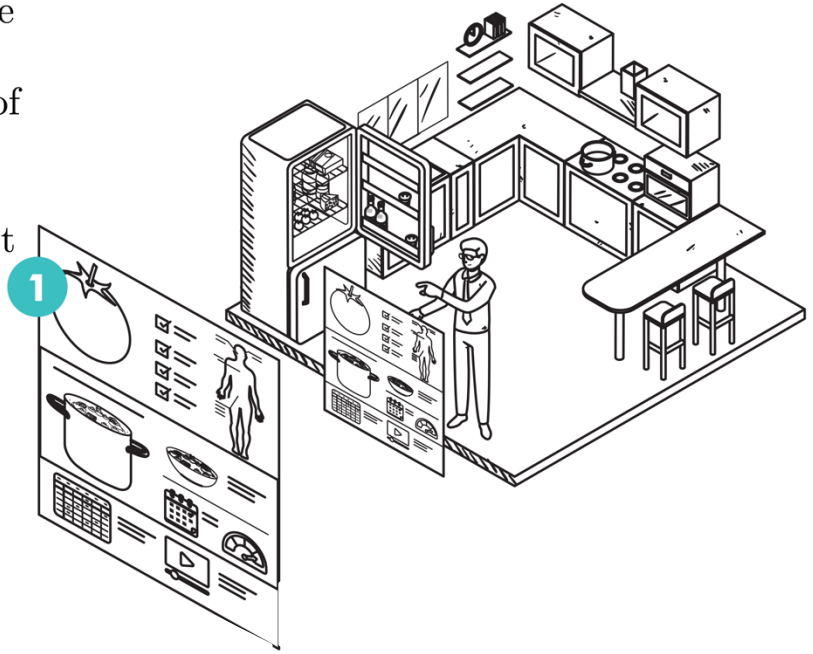


Ego-Police

EGO-Home

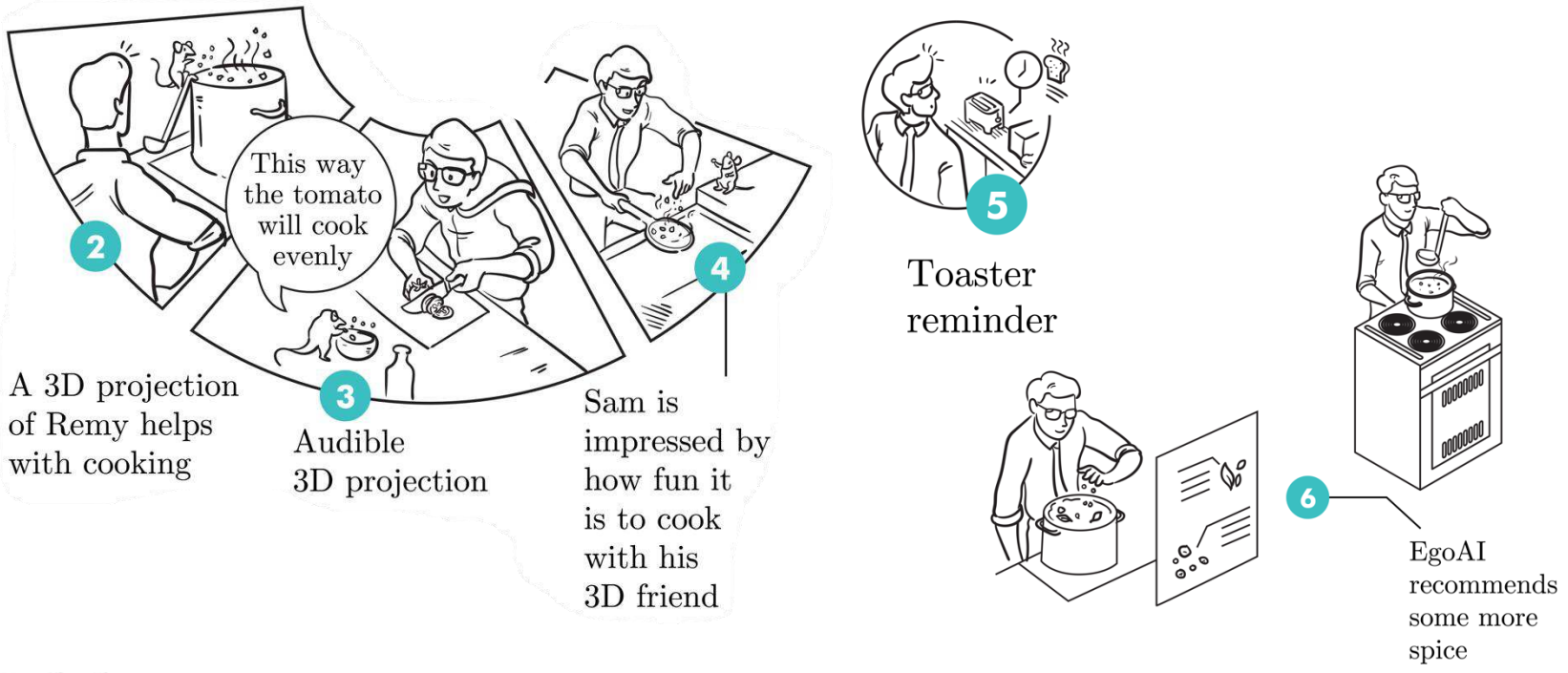
with: Chiara Plizzari, Gabriele Goletto, Antonio Furnari, Siddhant Bansal, Francesco Ragusa, Giovanni Maria Farinella, Tatiana Tommasi

Sam is finally home after a long day. EgoAI kept track of Sam's food intake and a tomato soup sounds like the best complementary nutrition



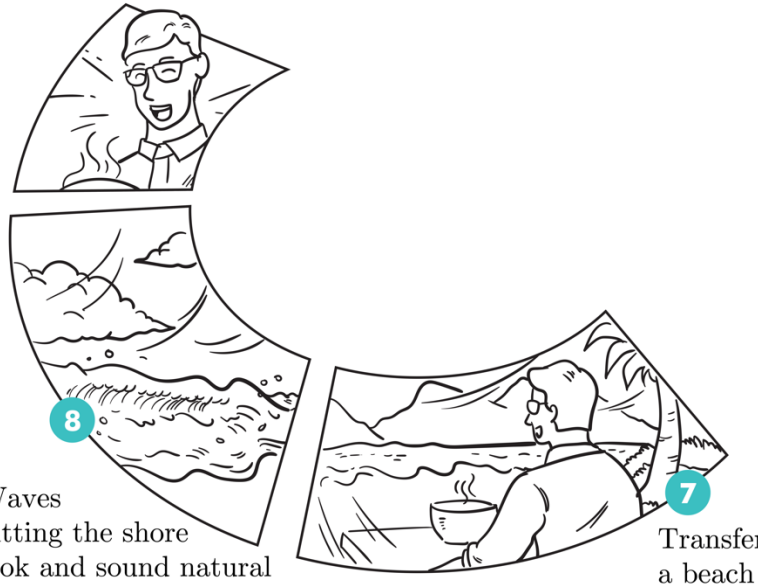
EGO-Home

with: Chiara Plizzari, Gabriele Goletto, Antonio Furnari, Siddhant Bansal, Francesco Ragusa, Giovanni Maria Farinella, Tatiana Tommasi



EGO-Home

with: Chiara Plizzari, Gabriele Goletto, Antonio Furnari, Siddhant Bansal, Francesco Ragusa, Giovanni Maria Farinella, Tatiana Tommasi



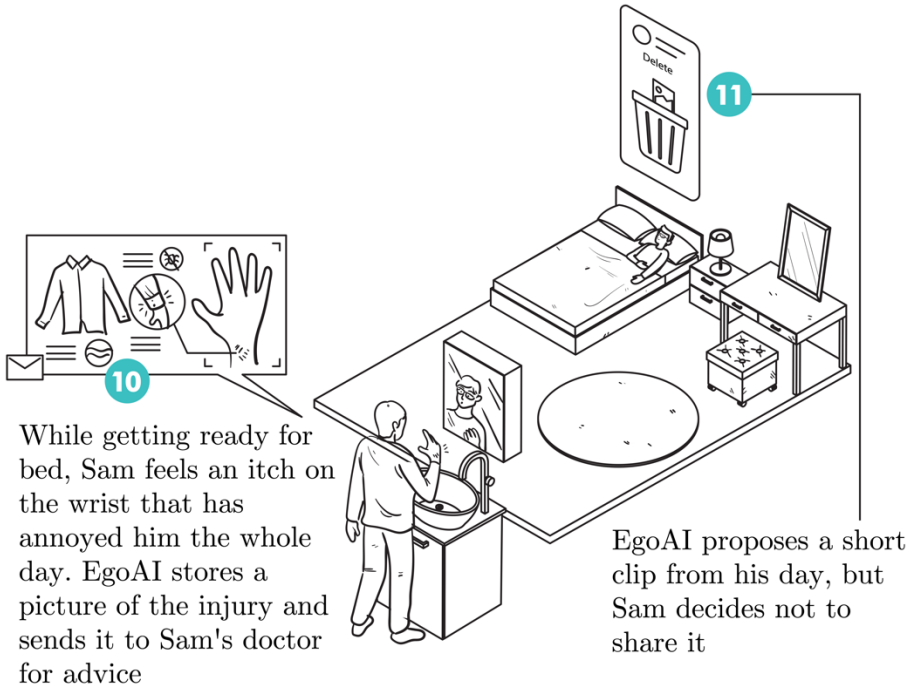
Waves hitting the shore look and sound natural

Transferred to a beach he visited last summer

After dinner, Sam enjoys a group card game with his friends, who are connected through their own EgoAI

EGO-Home

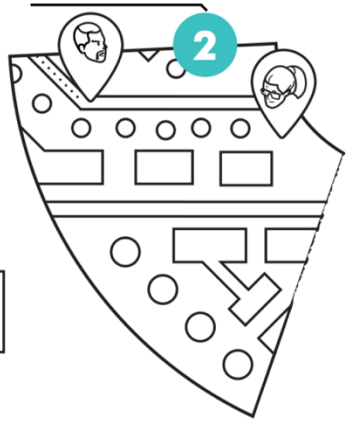
with: Chiara Plizzari, Gabriele Goletto, Antonio Furnari, Siddhant Bansal, Francesco Ragusa, Giovanni Maria Farinella, Tatiana Tommasi

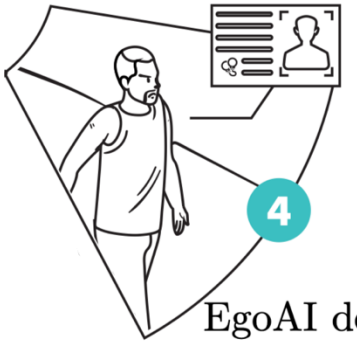


From Stories to Tasks

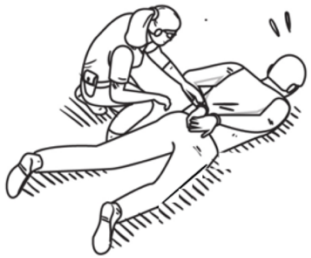
with: Chiara Plizzari, Gabriele Goletto, Antonio Furnari, Siddhant Bansal, Francesco Ragusa, Giovanni Maria Farinella, Tatiana Tommasi

EgoAI helps Judy navigate through the shortest safe path to target places





EgoAI detected and re-identified the man before he passed Judy

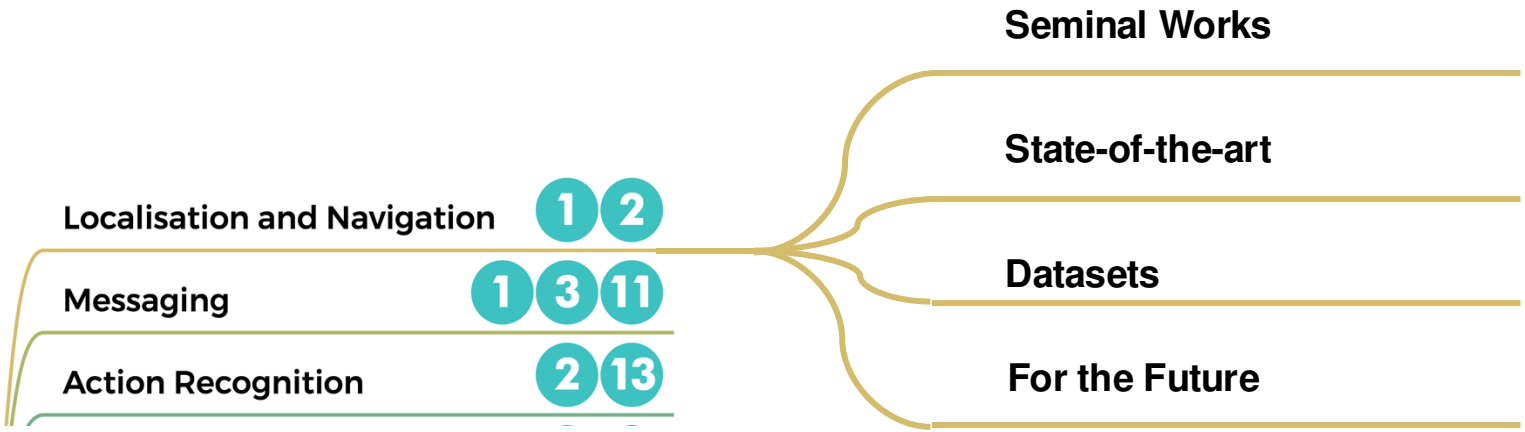


EGO-Police

- Localisation and Navigation 1 2
- Messaging 1 3 11
- Action Recognition 2 13
- Person Re-ID 2 4
- Object Detection and Retrieval 7
- Measuring System 8 9
- Decision Making 9
- 3D Scene Understanding 10
- Hand-Object Interaction 12
- Summarisation 13
- Privacy 14

The Survey Part

with: Chiara Plizzari, Gabriele Goletto, Antonio Furnari, Siddhant Bansal, Francesco Ragusa, Giovanni Maria Farinella, Tatiana Tommasi



Examples where things still fail…

Every Shot Counts

Get A Grip

Every Shot Counts:
Using Exemplars for Repetition Counting in Videos

Saptarshi Sinha, Alexandros Stergiou, Dima Damen

Every Shot Counts

with: Saptarshi Sinha
Alexandros Stergiou

RepCount

GT:6

Pred:6

Countix

GT:9

Pred:9

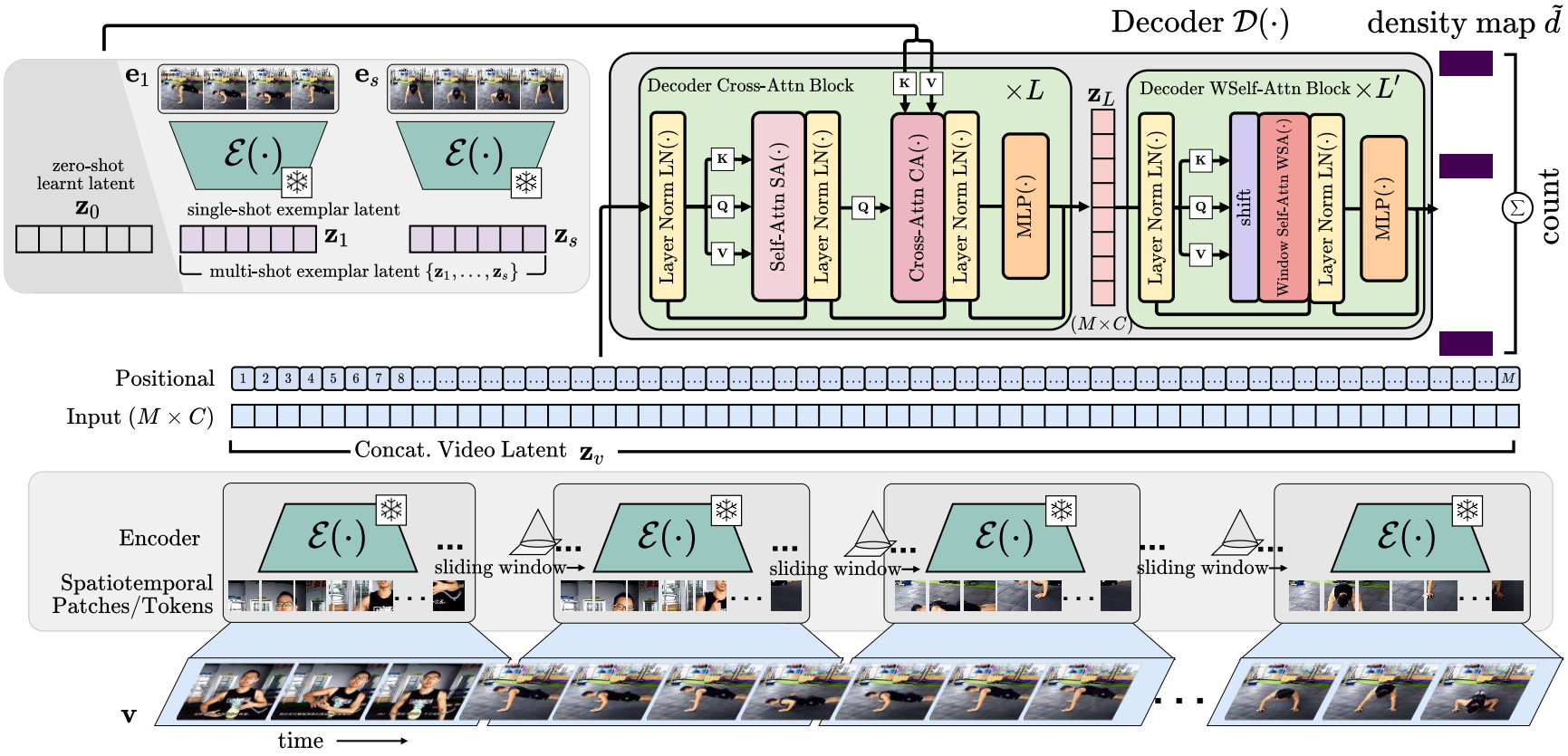
RepCount

GT:32

Pred:32

Every Shot Counts

with: Saptarshi Sinha
Alexandros Stergiou



Every Shot Counts

with: Saptarshi Sinha
Alexandros Stergiou

(a) RepCount

Method	Encoder	RMSE↓	MAE↓	OBZ↑	OBO↑
RepNet [15]	R2D50	-	0.995	-	0.013
TransRAC [18]	VSwinT	9.130*	0.443	0.085*	0.291
MFL [27]†	VSwinT	-	0.384	-	0.386
ESCounts	VSwinT	6.905	0.298	0.183	0.403
ESCounts	VMAE	**4.455**	**0.213**	**0.245**	**0.563**

(c) UCFRep

Method	Encoder	RMSE↓	MAE↓	OBZ↑	OBO↑
Levy & Wolf [25]	RX3D101	-	0.286	-	0.680
RepNet [15]	R2D50	-	0.998	-	0.009
Context (F) [62]	RX3D101	5.761*	0.653*	0.143*	0.372*
TransRAC [18]	VSwinT	-	0.640	-	0.324
MFL [27]†	RX3D101	-	0.388	-	0.510
ESCounts	RX3D101	2.004	0.247	0.343	0.731
ESCounts	VMAE	**1.972**	0.216	0.381	0.704

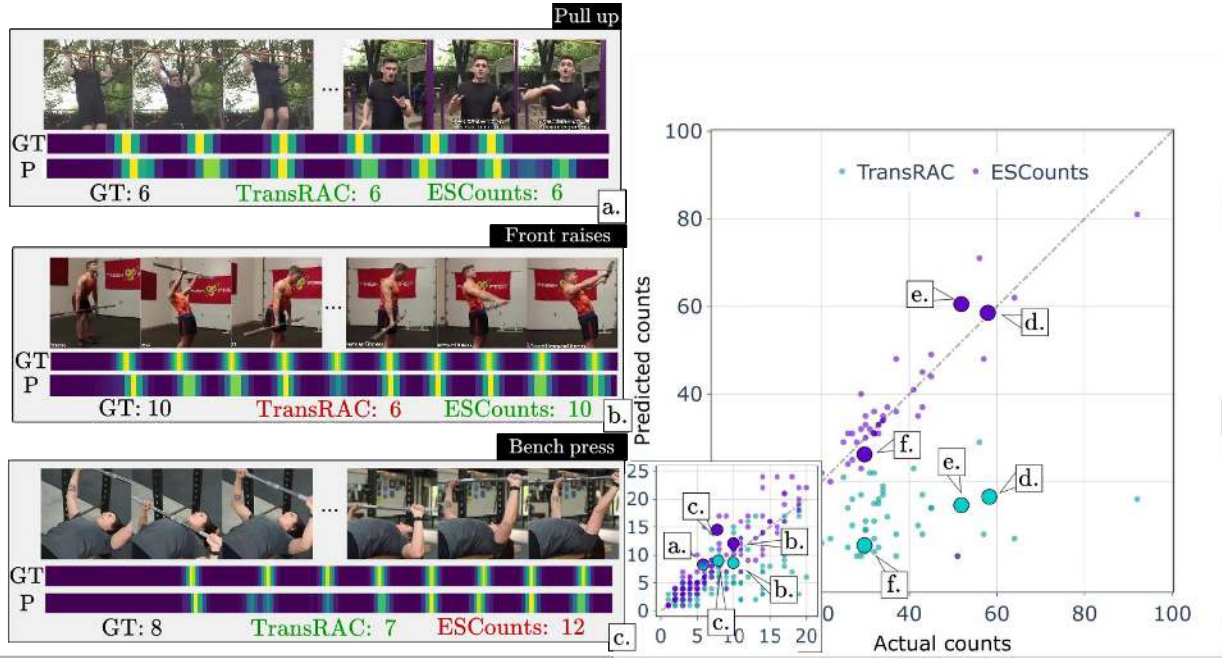
(b) Countix

Method	Encoder	RMSE↓	MAE↓	OBZ↑	OBO↑
RepNet [15]	R2D50	-	0.364	-	0.697
Sight & Sound [64]†	R(2+1)D18	-	0.307	-	0.511
ESCounts	R(2+1)D18	3.536	0.293	0.286	**0.701**
ESCounts	VMAE	**3.029**	**0.276**	**0.319**	0.673

Every Shot Counts

with: Saptarshi Sinha
Alexandros Stergiou

RepCount



Every Shot Counts - Generalisation

with: Saptarshi Sinha
Alexandros Stergiou

Table 2: Cross-dataset generalisation scores. Arrows $X \rightarrow Y$ denote train dataset X and test dataset Y. Results obtained using provided checkpoints are denoted with *.

	RepCount → UCFRep				RepCount → Countix			
	RMSE↓	MAE↓	OBZ↑	OBO↑	RMSE↓	MAE↓	OBZ↑	OBO↑
RN [16]	-	0.998	-	0.009	-	-	-	-
TRAC [20]	6.701*	0.640	0.087*	0.324	6.867*	0.593*	0.132*	0.364*
MFL [30]	-	0.523	-	0.350	-	-	-	-
ESCounts	**3.536**	**0.317**	**0.219**	**0.571**	**4.429**	**0.374**	**0.185**	**0.521**

Table X2. **Close and open-set setting results on RepCount.**

Task	Method	benchmark		open-set	
		MAE↓	OBO↑	MAE↓	OBO↑
TAL	Huang *et al.*	0.527	0.159	1.000	0.000
VRC	TRAC	0.443	0.291	0.625	0.204
	ESCounts	**0.213**	**0.563**	**0.436**	**0.519**

Every Shot Counts - Generalisation

with: Saptarshi Sinha
Alexandros Stergiou

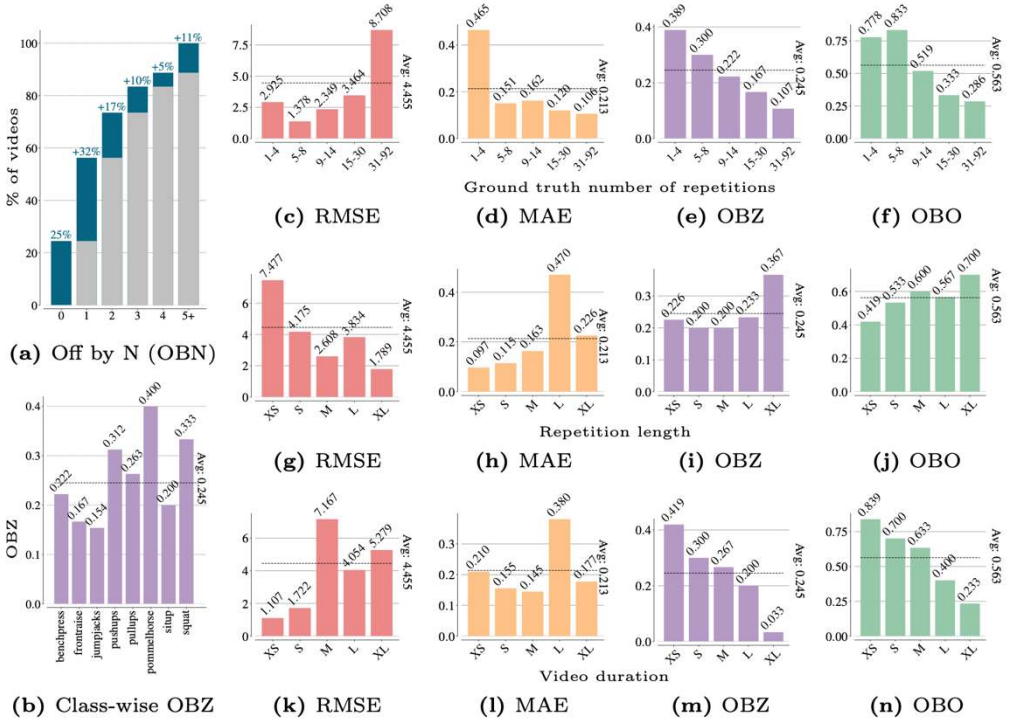


Fig. 6: Grouped VRC scores over different number of repetitions and lengths. (a) overviews the Off by N accuracy for increasing Ns. (b) shows OBZ by action class. The first row (c–f) reports results over different counts. (g–j) reports scores over groups by repetition durations. (k–n) reports metrics grouped by video duration.

Rejected @ECCV 2024

Accepted @ACCV 2024

ot Counts:

Us petition Counting in Videos

Is it still a significant task that needs to be explored?

no guarantee that ESCount can generalize to non-human activities

Saptarshi Sinha, Alexandros Stergiou, Dima Damen

Get a Grip

Reconstructing Hand-Object Stable Grasps in Egocentric Videos

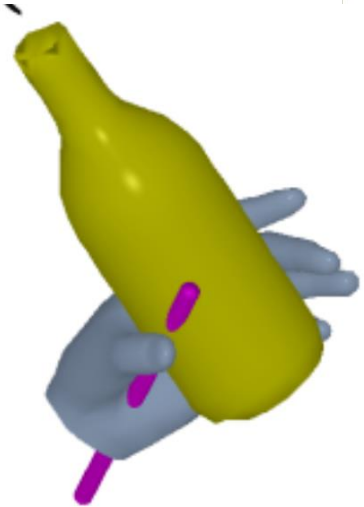
Zhifan Zhu and Dima Damen

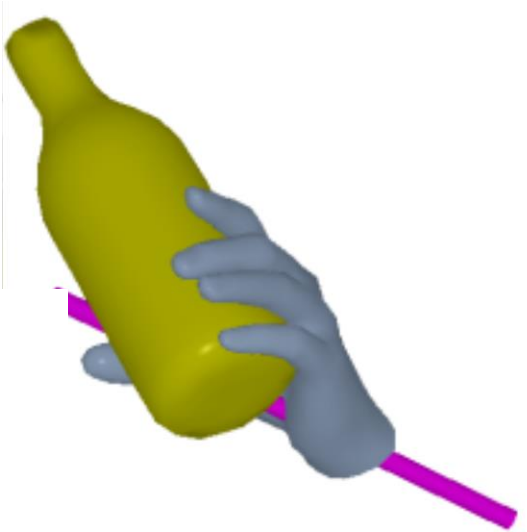
Get a Grip

with: Zhifan Zhu

left hand

bottle





Get a Grip

with: Zhifan Zhu

Non-Ego Views

Ego Views

Z Zhu and D Damen (2024). Get a Grip: Reconstructing Hand-Object Stable Grasps in Egocentric Videos. *ArXiv*

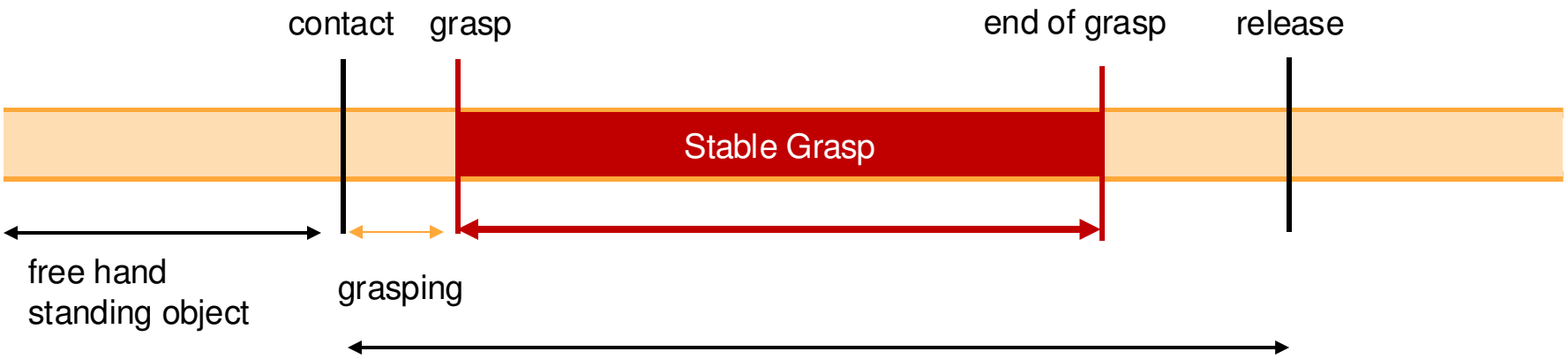
Get a Grip

with: Zhifan Zhu

Z Zhu and D Damen (2024). Get a Grip: Reconstructing Hand-Object Stable Grasps in Egocentric Videos. *ArXiv*

Get a Grip

with: Zhifan Zhu



Z Zhu and D Damen (2024). Get a Grip: Reconstructing Hand-Object Stable Grasps in Egocentric Videos. *ArXiv*

Get a Grip

with: Zhifan Zhu

ARCTIC (CVPR 2023)

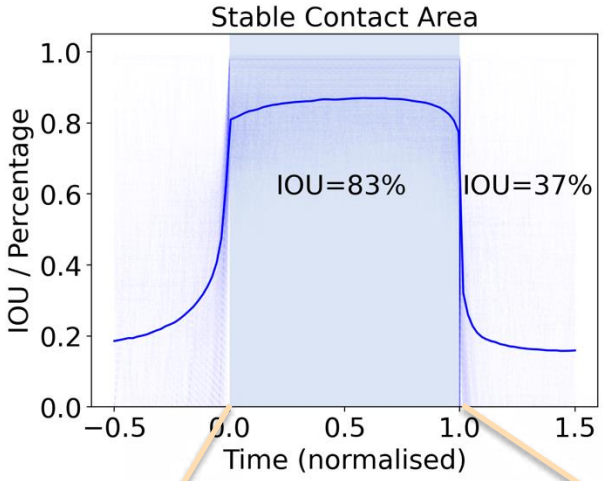
Z Fan, et al.
ARCTIC: A dataset for dexterous bimanual hand- object manipulation. **CVPR 2023**

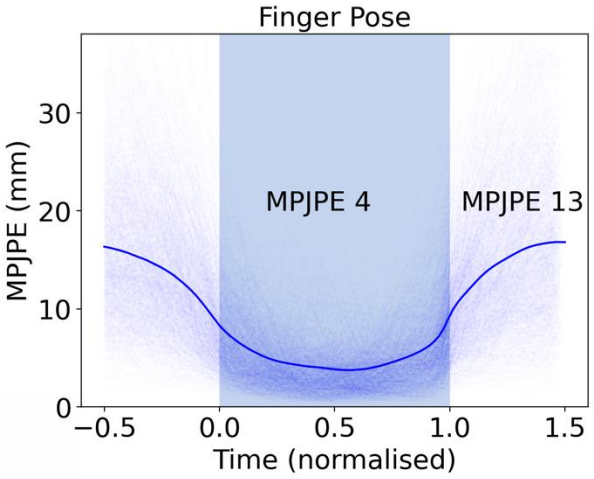
HOI4D (CVPR 2022)

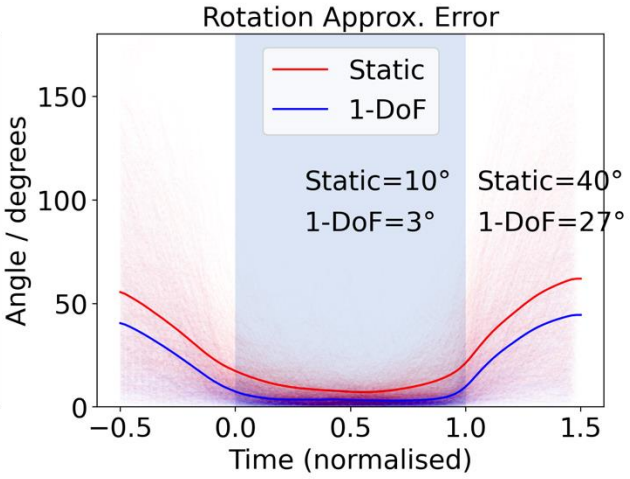
Yunze Liu, et al.
HOI4D: A 4D Egocentric Dataset for Category-Level Human-Object Interaction. **CVPR 2022**

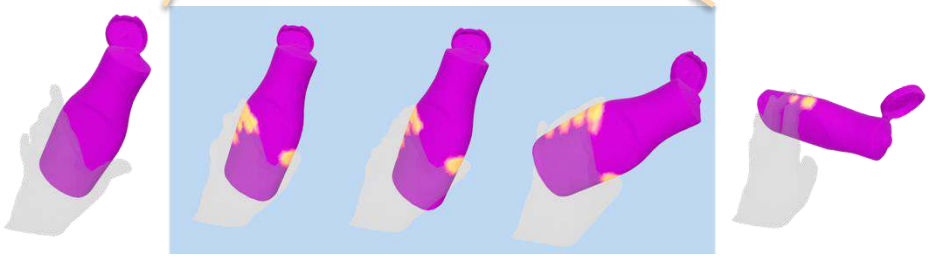
Get a Grip

with: Zhifan Zhu



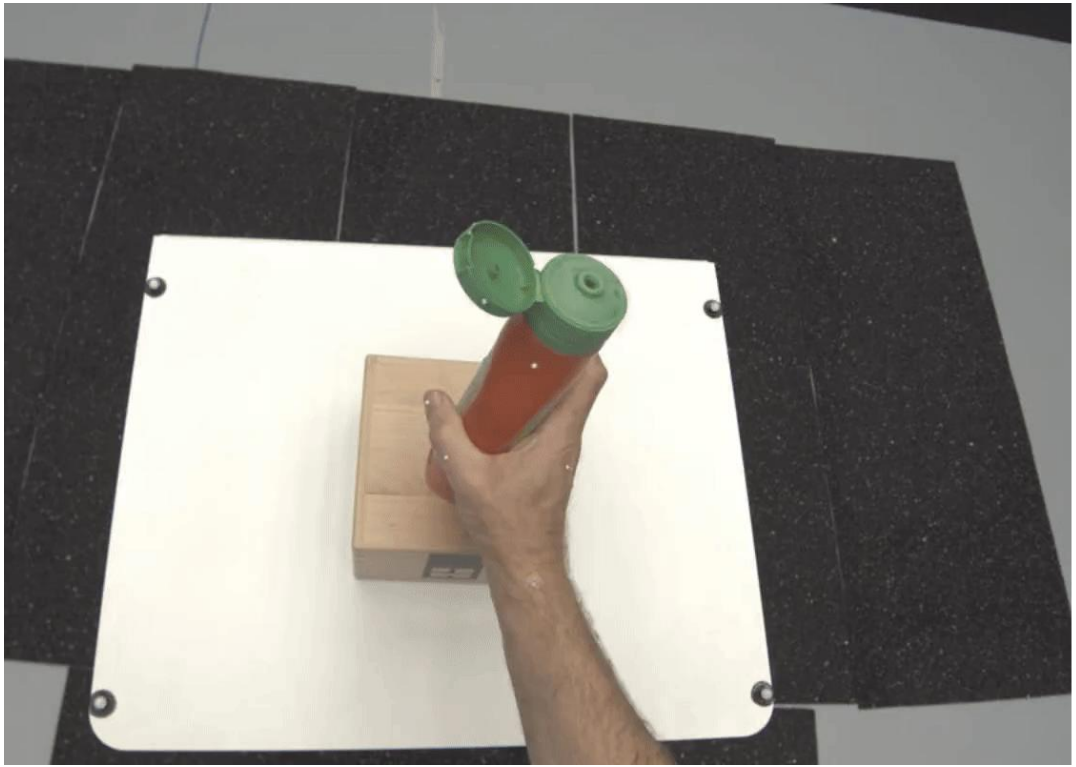






Get a Grip

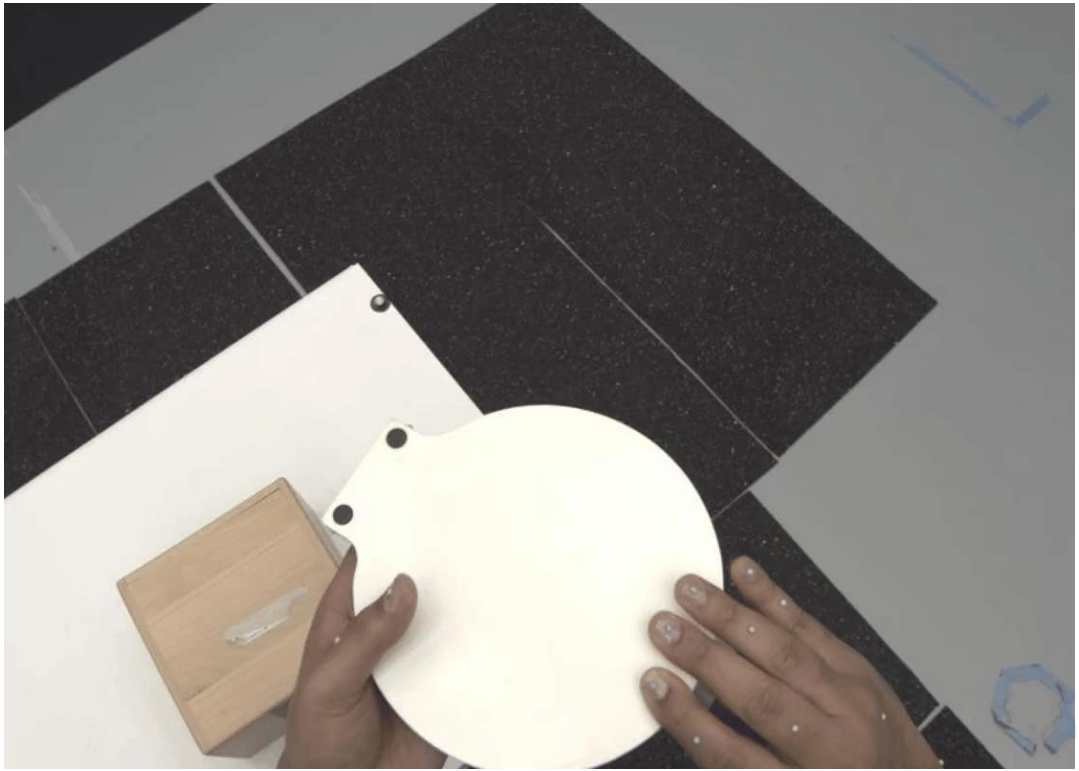
with: Zhifan Zhu

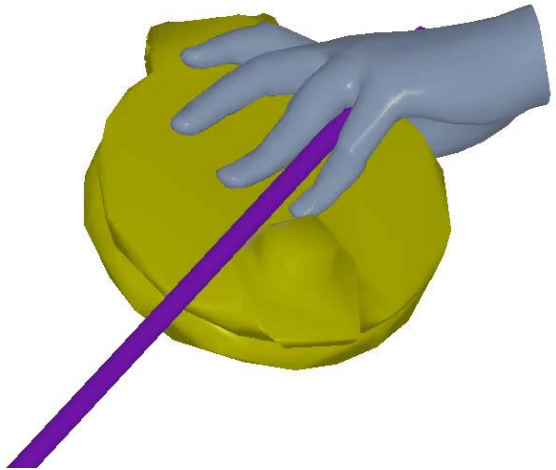


Z Zhu and D Damen (2024). Get a Grip: Reconstructing Hand-Object Stable Grasps in Egocentric Videos. *ArXiv*

Get a Grip

with: Zhifan Zhu





Z Zhu and D Damen (2024). Get a Grip: Reconstructing Hand-Object Stable Grasps in Egocentric Videos. *ArXiv*

Get a Grip

with: Zhifan Zhu

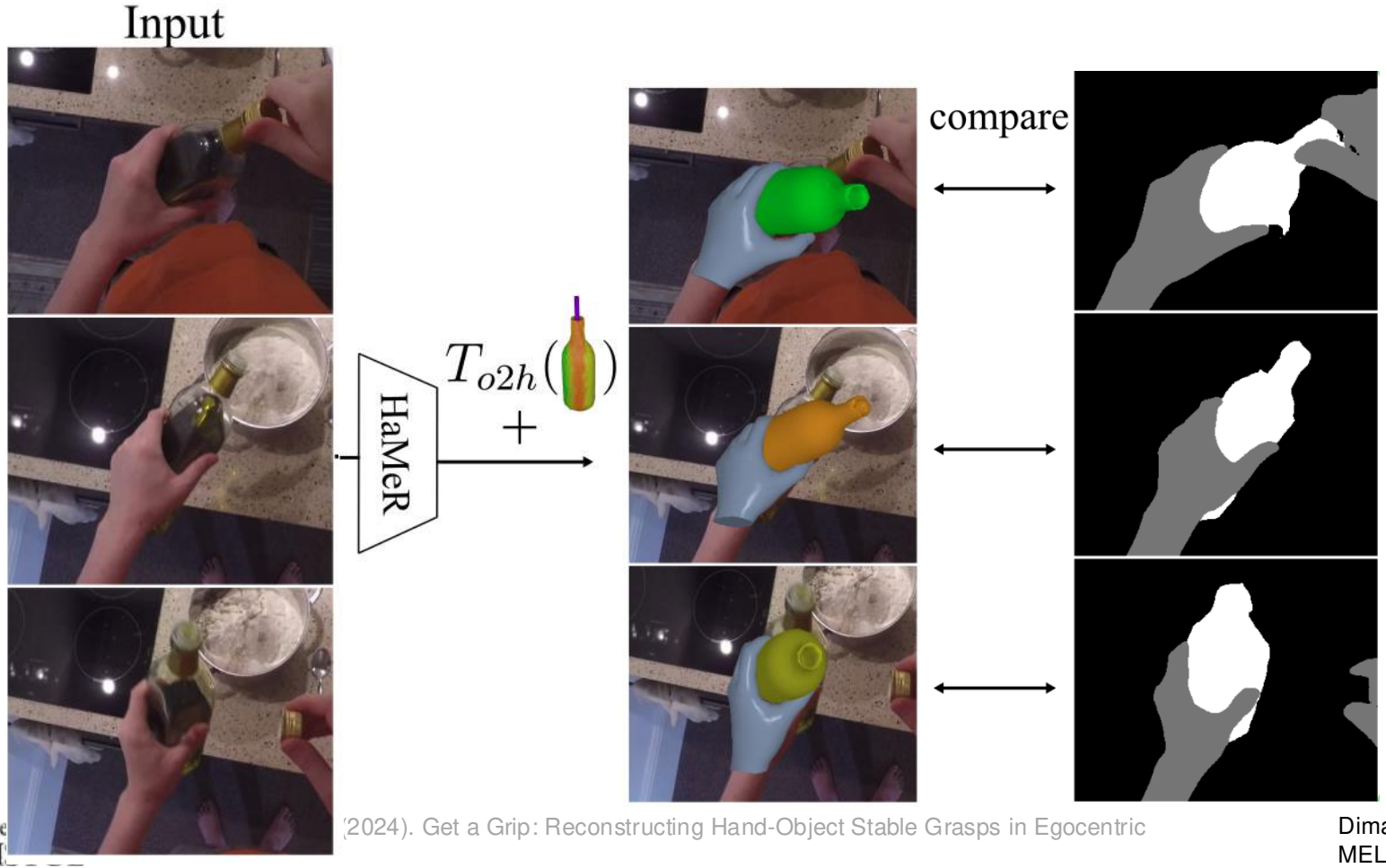
Sequences	Instances	Categories	Subjects
2431	~390	9	31

1446 left hands  , 985 right hands 

Z Zhu and D Damen (2024). Get a Grip: Reconstructing Hand-Object Stable Grasps in Egocentric Videos. *ArXiv*

Get a Grip

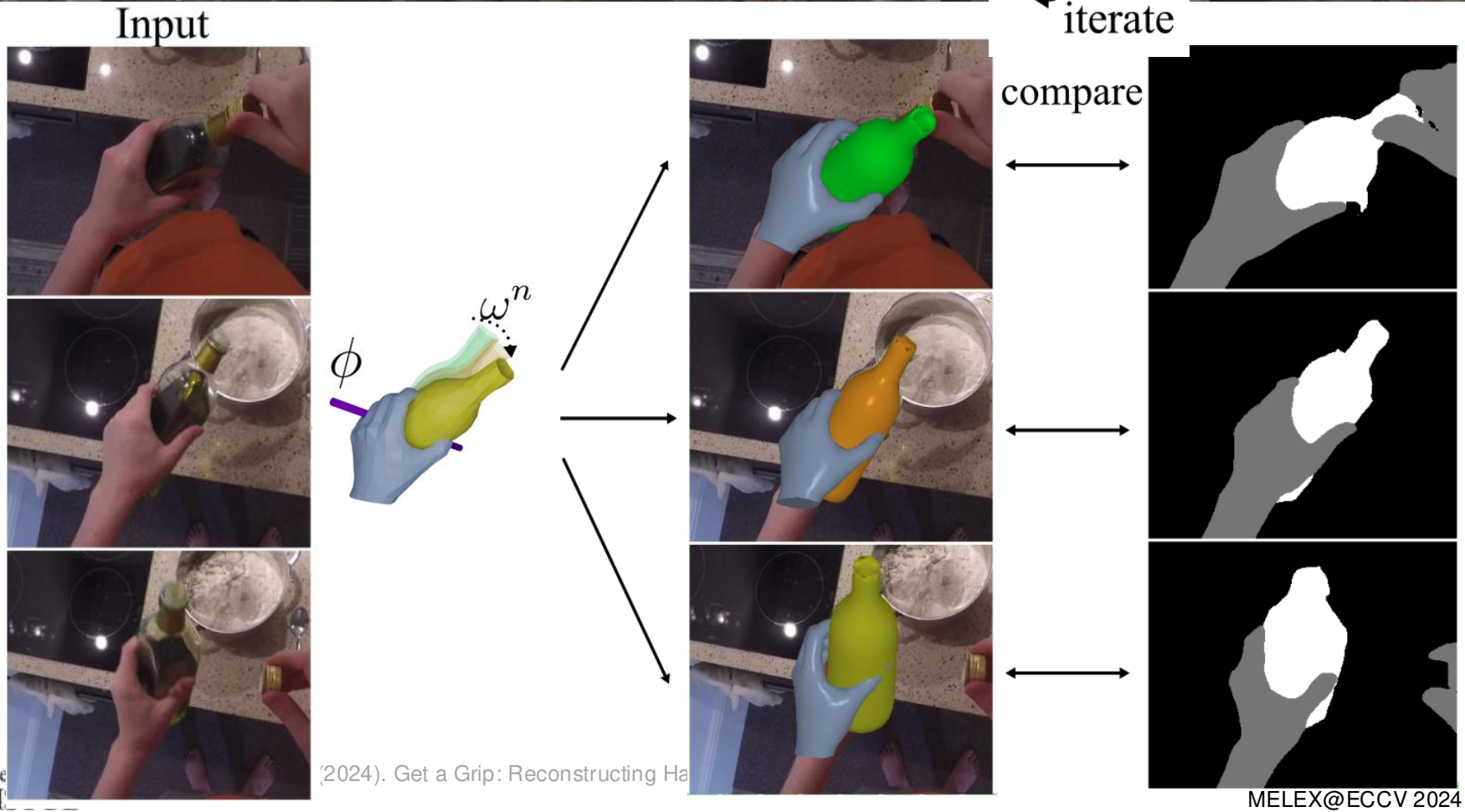
with: Zhifan Zhu



(2024). Get a Grip: Reconstructing Hand-Object Stable Grasps in Egocentric

Get a Grip

with: Zhifan Zhu



(2024). Get a Grip: Reconstructing Ha

Get a Grip

with: Zhifan Zhu

Z Zhu and D Damen (2024). Get a Grip: Reconstructing Hand-Object Stable Grasps in Egocentric Videos. *ArXiv*

Rejected @ECCV 2024

Get a Grip

Stable Grasps in
os

This method only targets a very specific scenario, the stable hand-object grasp period

It is currently unclear to me the key motivations of reconstruction under stable grasps

Zhifan Zhu and Dima Damen

Exploration vs Exploitation

Exploration	Exploitation
Harder	
Riskier	Rarely worth the time
Slower	

Thank you

For further info, datasets, code, publications…

http://dimadamen.github.io


@dimadamen


http://www.linkedin.com/in/dimadamen

Q&A