Video Understanding

University of BRISTOL

Dima Damen
BinEgo-360 Workshop @ICCV2025

Definitions…

***Ego**… a person's sense of self-esteem or self-importance*

University of BRISTOL

Definitions…

***Ego**… a person's sense of self-esteem or self-importance*

***Egocentric vision**… the wearer serves as the central reference point in the study of interesting entities: objects, actions, interactions and intentions*

360 vs Egocentric…

Ego

360

360 vs Egocentric…

In today's talk…

Motivation and Datasets in Egocentric Video Understanding

Video Understanding Out of the Frame

- Point Tracking
- Object Tracking
- Gaze Priming
- Hand Tracking

Teaser: The Wizard of Oz & Genie 3

Outlook into the Future of Egocentric Vision

Conclusion

In today's talk…

Motivation and Datasets in Egocentric Video Understanding

Teaser: The Wizard of Oz & Genie 3

Video Understanding Out of the Frame

- Point Tracking
- Object Tracking
- Gaze Priming
- Hand Tracking

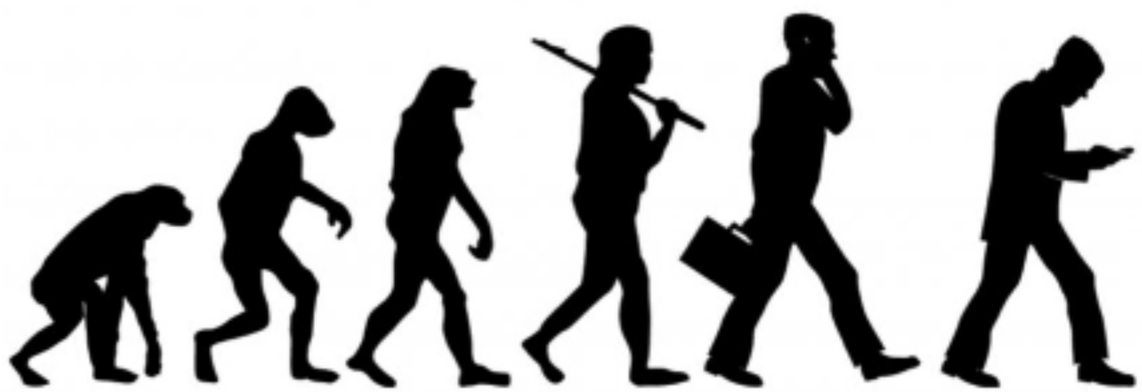
Outlook into the Future of Egocentric Vision

Conclusion

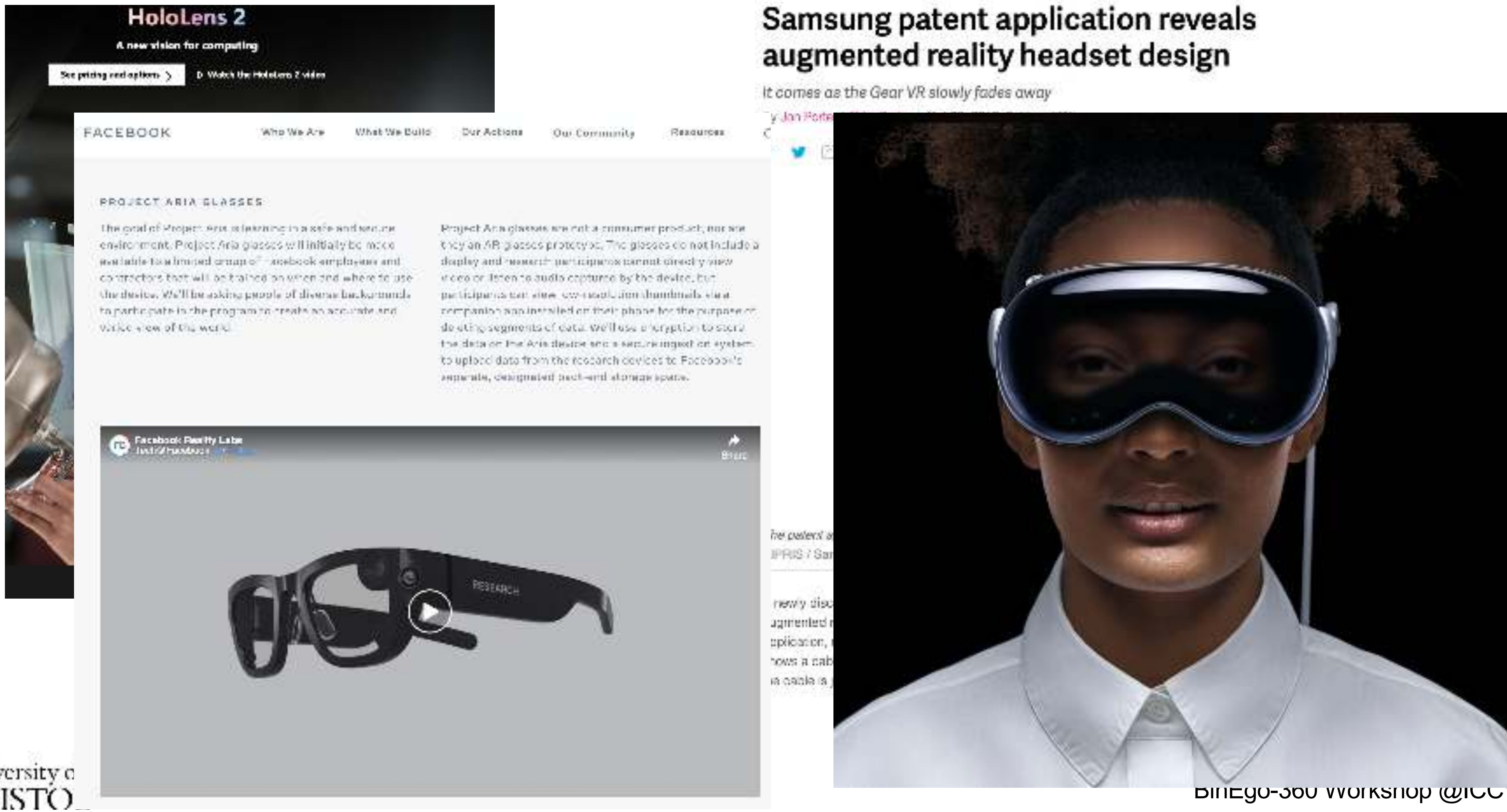
University of BRISTOL

The present…

Photo *Illustration* by Pelle Cass



The future…



University of BRISTOL

amen
BinEgo-360 Workshop @ICCV2025

The future is here…

The future can be imagined…

Egocentric Videos?

In today's talk…

Motivation and Datasets in Egocentric Video Understanding

Video Understanding Out of the Frame

- Point Tracking
- Object Tracking
- Gaze Priming
- Hand Tracking

Teaser: The Wizard of Oz & Genie 3

Outlook into the Future of Egocentric Vision

Conclusion

EPIC-KITCHENS

EPIC Fields

with: V Tschernezki*, A Darkhalil*, Z Zhu*, D Fouhey, I Laina, D Larlus, A Vedaldi

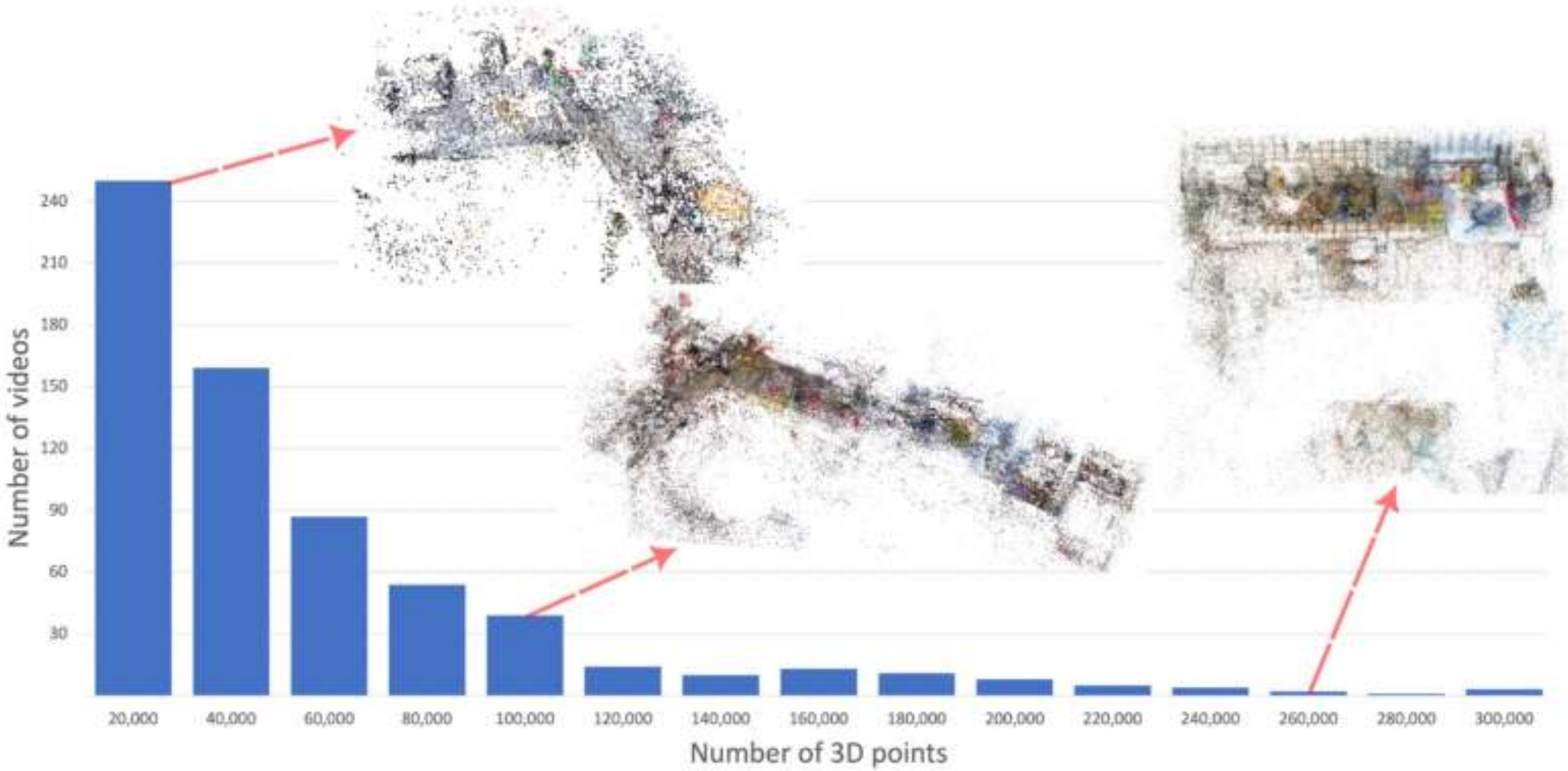


Figure 4: **Number of 3D points histogram.** The majority of our reconstructions generate less than 40,000 points that are enough to represent the kitchen. However, some reconstructions have more than 100,000, we include the point clouds for each points range showing the fine details covered by having more points

with: V Tschernezki*, A Darkhalil*, Z Zhu*,
D Fouhey, I Laina, D Larlus, A Vedaldi

Table 1: Comparison of datasets commonly used in dynamic new-view synthesis.

Dataset	#Scenes	Seq. Length	Monocular	Semantics
Nerfies [37]	4	8–15 sec	-	-
D-NeRF [41]	8	1–3 sec	-	-
Plenoptic Video [22]	6	10-60 sec	-	-
NVIDIA Dynamic Scene Dataset [65]	12	1–5 sec	4 / 12	-
HyperNeRF [38]	16	8–15 sec	13 / 16	-
iPhone [13]	14	8–15 sec	7 / 14	-
SAFF [25]	8	1–5sec	-	✓
EPIC Fields (ours)	50	6–37 min (Avg 22)	50 / 50	✓

A Darkhalil et al (2022). EPIC-KITCHENS VISOR Benchmark: VIdeo Segmentations and Object Relations. *NeurIPS*

In today's talk…

Motivation and Datasets in Egocentric Video Understanding

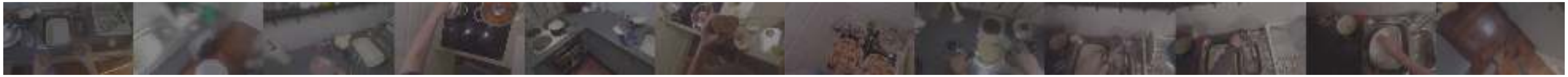
Video Understanding Out of the Frame

- Point Tracking
- Object Tracking
- Gaze Priming
- Hand Tracking

Teaser: The Wizard of Oz & Genie 3

Outlook into the Future of Egocentric Vision

Conclusion



EgoPoints: Advancing Point Tracking for Egocentric Videos

Ahmad Darkhalil[1] Rhodri Guerrier[1] Adam W. Harley[2] Dima Damen[1]

[1]University of Bristol [2]Stanford University

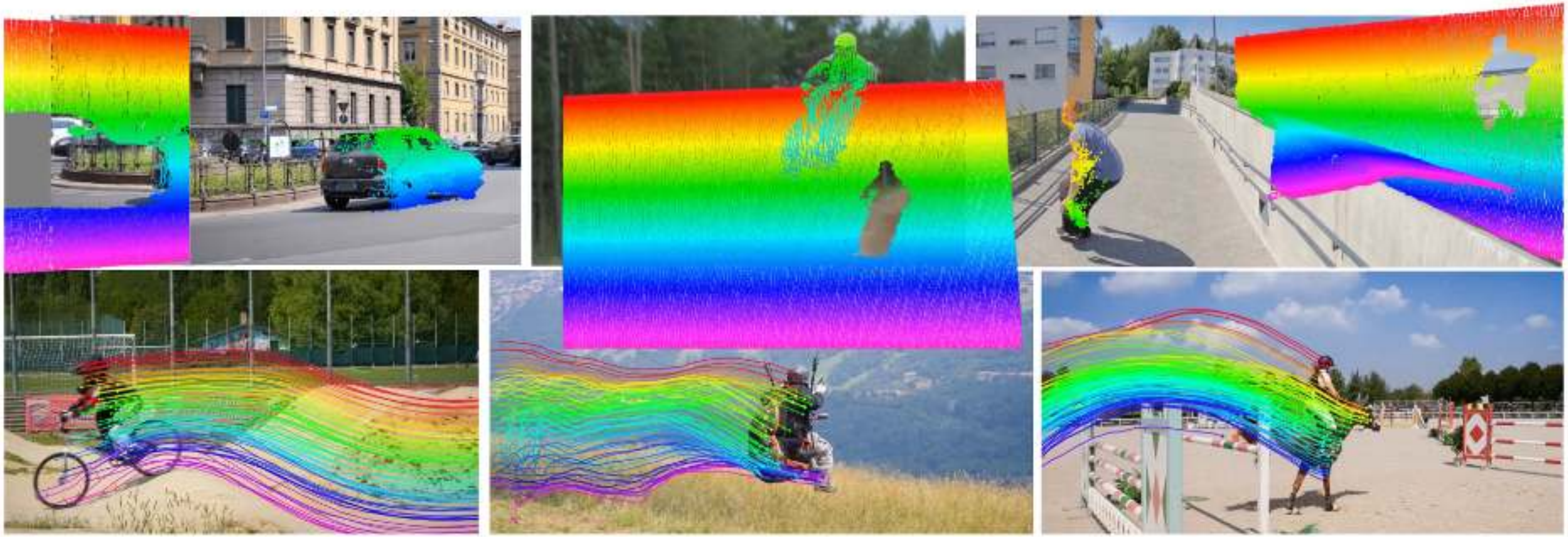
WACV 2025

TUCSON, ARIZONA • FEB 28 - MAR 4

What is point tracking?

with: Ahmad Darkhalil
Rhodri Guerrier
Adam W Harley

- Given: Query points in one frame
- Track these points throughout the video

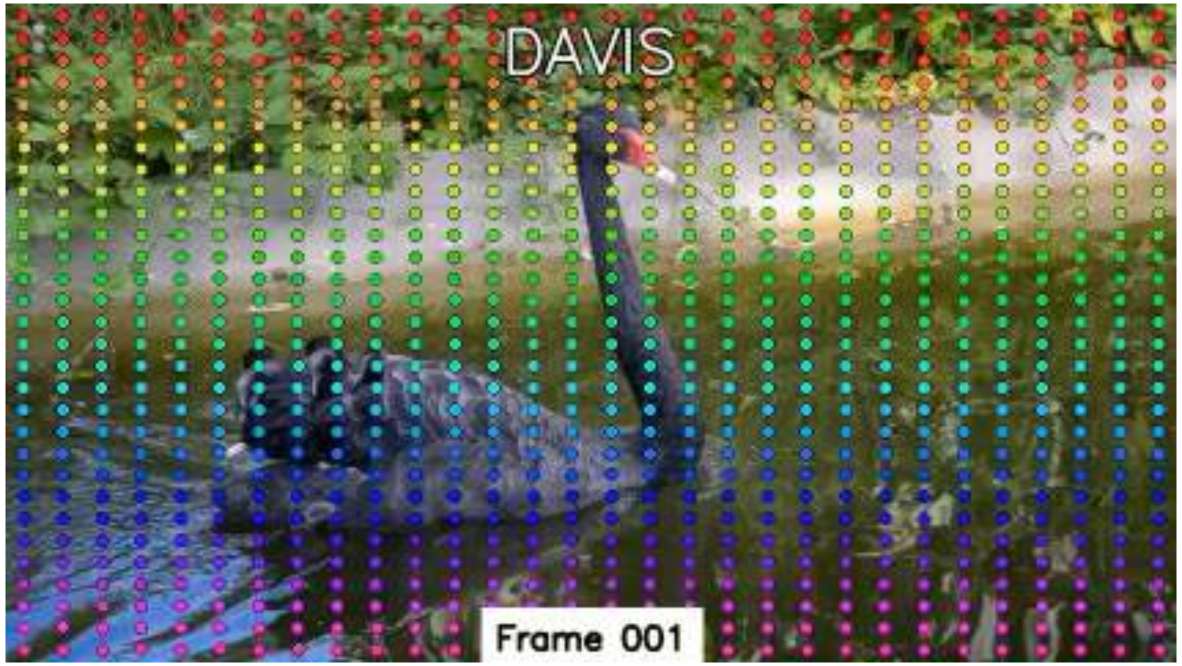


Darkhalil et al (2025). EgoPoints: Advancing Point Tracking for Egocentric Videos. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)

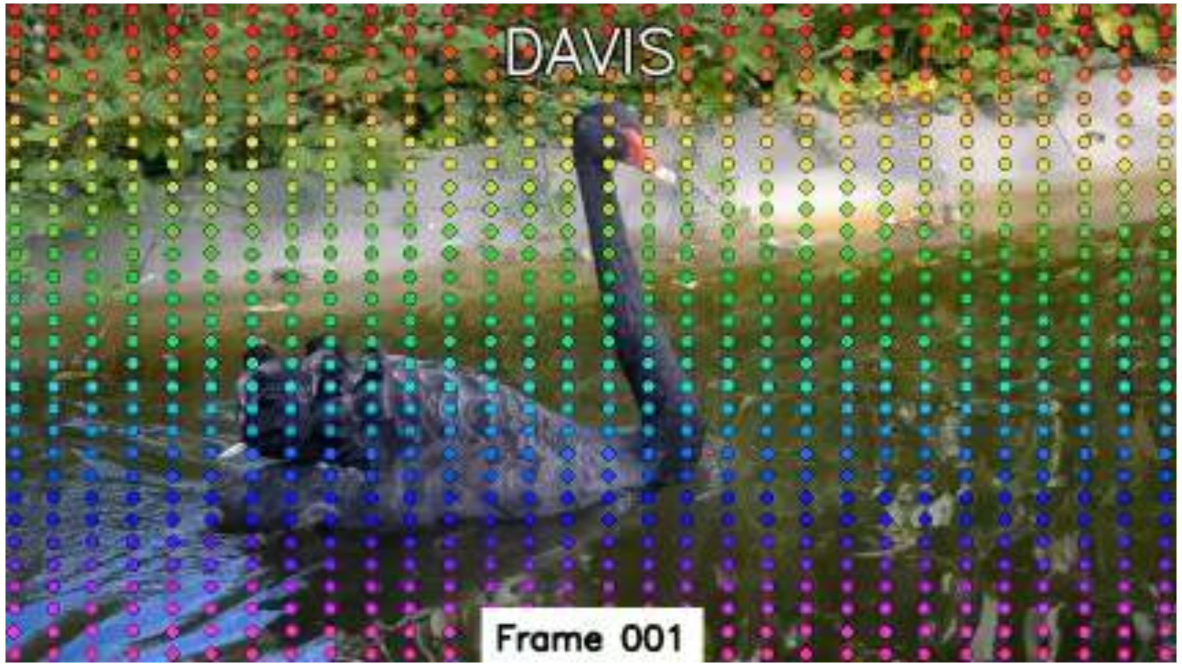
SOTA on current benchmarks

with: Ahmad Darkhalil
Rhodri Guerrier
Adam W Harley

LocoTrack



CoTracker3



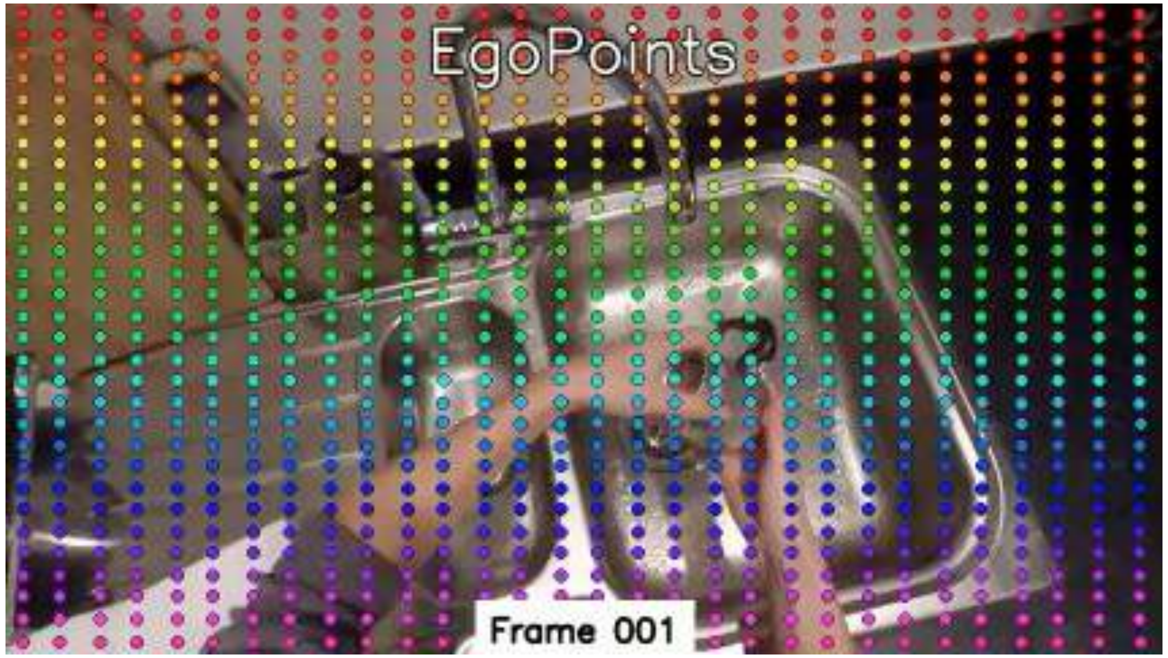
Darkhalil et al (2025). EgoPoints: Advancing Point Tracking for Egocentric Videos. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)

Current Models Struggle with Egocentric Videos

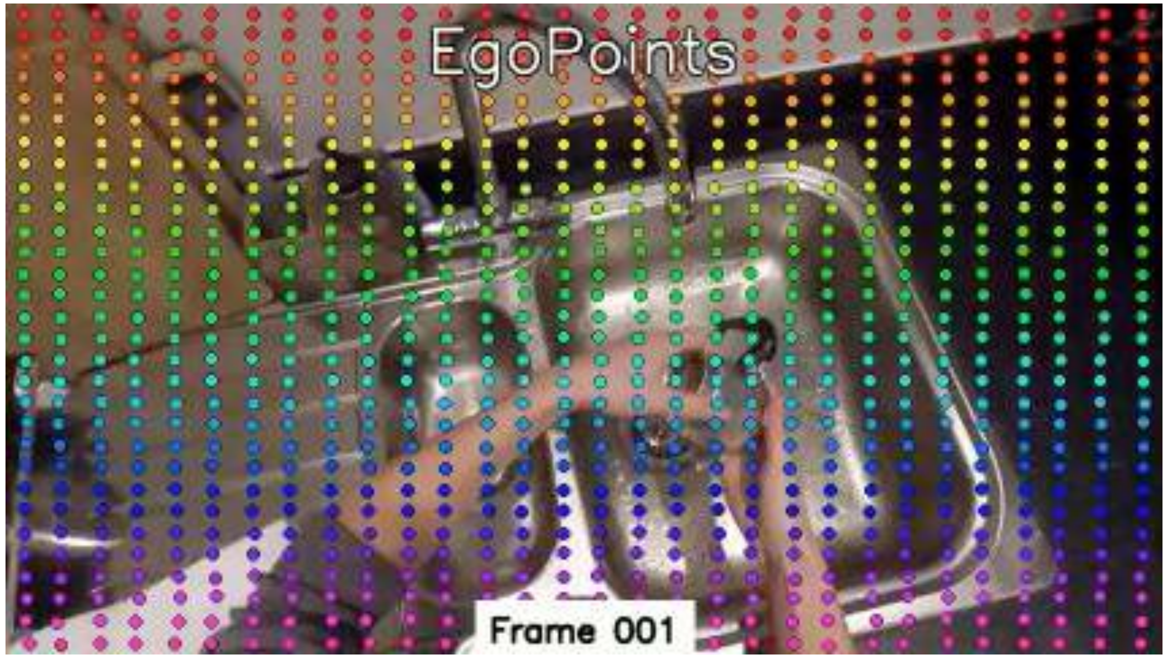
with: Ahmad Darkhalil
Rhodri Guerrier
Adam W Harley

- Head motion and motion blur
- Frequent re-identification

LocoTrack



CoTracker3



Darkhalil et al (2025). EgoPoints: Advancing Point Tracking for Egocentric Videos. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)

Main Contributions

with: Ahmad Darkhalil
Rhodri Guerrier
Adam W Harley

- Identify challenges that point trackers face in egocentric videos.
- Propose a new benchmark (EgoPoints) and new metrics to showcase these challenges
- Propose K-EPIC, a pipeline to generate semi-real training data

Darkhalil et al (2025). EgoPoints: Advancing Point Tracking for Egocentric Videos. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)

EgoPoints Annotation interface

with: Ahmad Darkhalil
Rhodri Guerrier
Adam W Harley

Darkhalil et al (2025). EgoPoints: Advancing Point Tracking for Egocentric Videos. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)

EgoPoints Benchmark

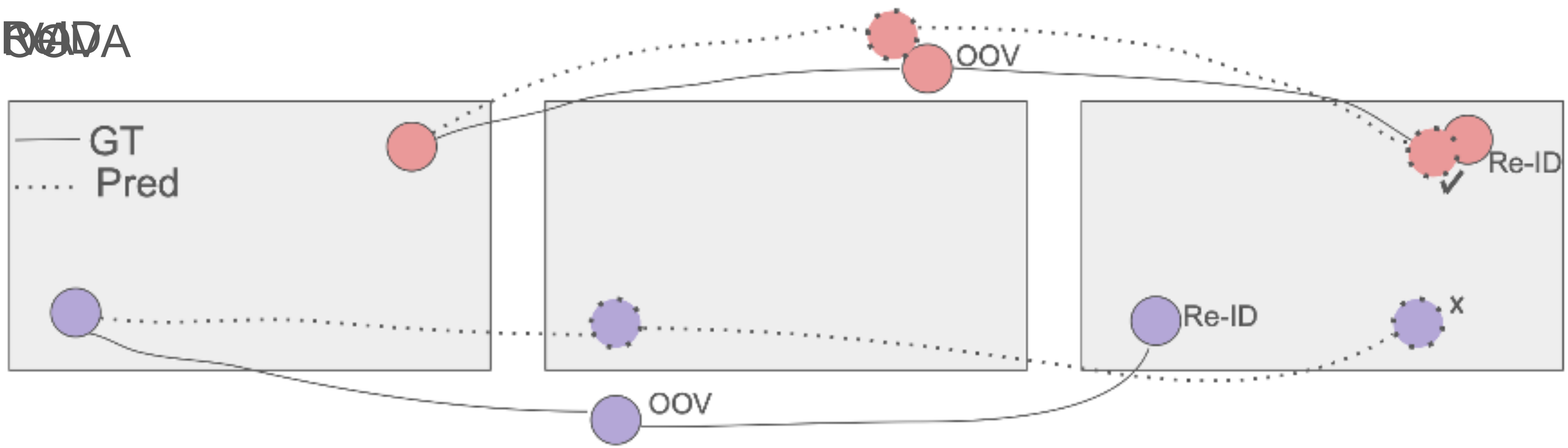
with: Ahmad Darkhalil
Rhodri Guerrier
Adam W Harley

Darkhalil et al (2025). EgoPoints: Advancing Point Tracking for Egocentric Videos. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)

Proposed Metrics

with: Ahmad Darkhalil
Rhodri Guerrier
Adam W Harley

ReID OOV



Darkhalil et al (2025). EgoPoints: Advancing Point Tracking for Egocentric Videos. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)

EgoPoints Benchmark

with: Ahmad Darkhalil
Rhodri Guerrier
Adam W Harley

Dataset	Total Tracks	OOV Tracks	ReID Tracks	Avg. Video Length	Avg. Points/Frame
TAP-Vid-DAVIS	650	94	10	66.6	**21.7**
EgoPoints	**4703**	**875**	**593**	**511.0**	8.5

Comparisons of our annotated sequences, EgoPoints, and the commonly used TAP-Vid-DAVIS [7] point tracking benchmarks

Darkhalil et al (2025). EgoPoints: Advancing Point Tracking for Egocentric Videos. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)

SOTA Models Struggle on EgoPoints

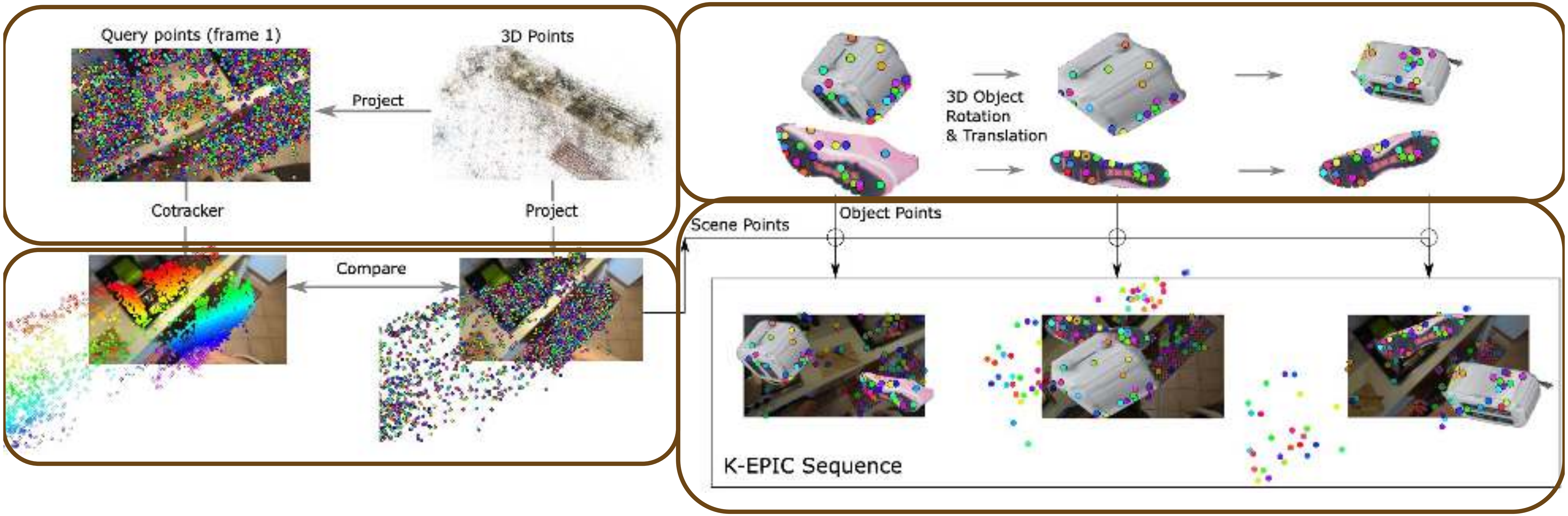
with: Ahmad Darkhalil
Rhodri Guerrier
Adam W Harley

	TAP-Vid-DAVIS	EgoPoints			
Model	$\delta_{avg}\uparrow$	$\delta_{avg}\uparrow$	ReID$\delta_{avg}\uparrow$	OOVA↑	IVA↑
PIPs++ [42]	64.0	36.9	14.6	50.4	89.2
CoTracker [22]	74.7	38.5	4.8	**81.4**	73.4
BootsTAPIR Online [8]	65.2	39.6	0.0	0.0	**100.0**
LocoTrack [4]	75.3	**59.4**	0.1	0.2	99.9
CoTracker v3 [21]	**77.2**	50.0	**15.0**	31.8	99.3

Darkhalil et al (2025). EgoPoints: Advancing Point Tracking for Egocentric Videos. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)

Pipeline of K-EPIC

with: Ahmad Darkhalil
Rhodri Guerrier
Adam W Harley



University of BRISTOL

Darkhalil et al (2025). EgoPoints: Advancing Point Tracking for Egocentric Videos. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)

Examples from K-EPIC

with: Ahmad Darkhalil
Rhodri Guerrier
Adam W Harley

Darkhalil et al (2025). EgoPoints: Advancing Point Tracking for Egocentric Videos. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)

Improvements After Fine-Tuning on K-EPIC

with: Ahmad Darkhalil
Rhodri Guerrier
Adam W Harley

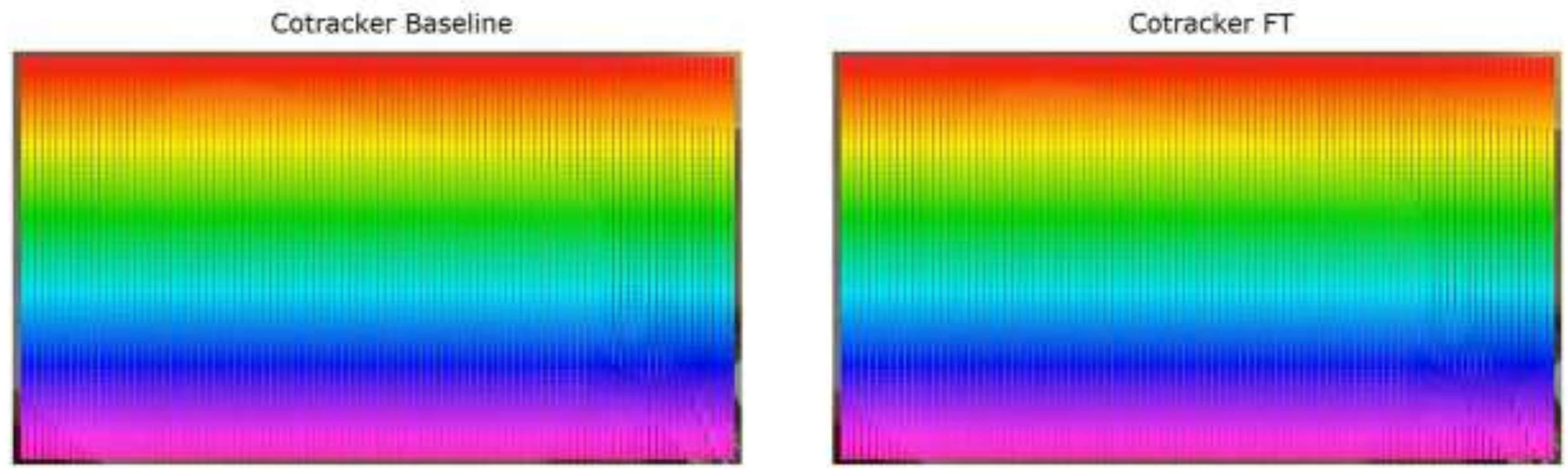
	δ Metrics			Accuracy Metrics			Error
Model	$\delta_{avg}\uparrow$	$\delta^*_{avg}\uparrow$	ReID$\delta_{avg}\uparrow$	IVA↑	OOVA↑	OA↑	MTE↓
PIPs++ [42]	**36.9**	57.8	14.0	89.2	50.4	–	22.9
PIPs++ w. K-EPIC FT (scene points only)	36.3	57.8	13.0	**90.1**	**53.0**	–	22.9
PIPs++ w. K-EPIC FT (scene and object points)	36.6	**58.1**	**16.8**	89.9	52.0	–	**22.2**
CoTracker [22]	38.5	54.8	4.8	73.4	81.4	81.0	52.1
CoTracker w. K-EPIC FT (scene points only)	38.9	56.0	6.3	74.8	**85.4**	80.7	51.3
CoTracker w. K-EPIC FT (scene and object points)	**39.6**	**57.5**	**7.2**	**78.1**	82.0	**81.8**	**40.5**

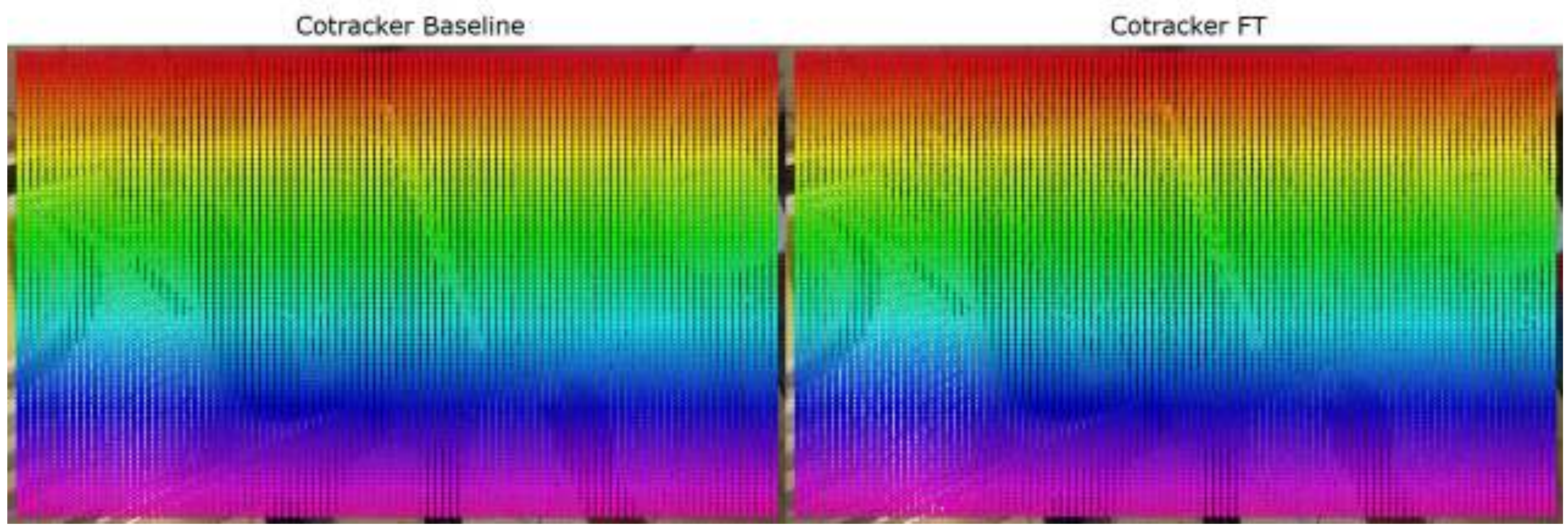
University of BRISTOL

Darkhalil et al (2025). EgoPoints: Advancing Point Tracking for Egocentric Videos. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)

Qualitative Examples of CoTracker

with: Ahmad Darkhalil
Rhodri Guerrier
Adam W Harley





Darkhalil et al (2025). EgoPoints: Advancing Point Tracking for Egocentric Videos. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)

AllTracker – newest work

Table 1. Comparison against recent point trackers and optical flow models, across nine datasets. We evaluate δ_{avg} (higher is better), using an input resolution of 384 × 512. The benchmarks are BADJA [3], CroHD [46], TAPVid-DAVIS [11], DriveTrack [1], EgoPoints [10], Horse10 [33], TAPVid-Kinetics [11], RGB-Stacking [28], and RoboTAP [52].

Method	Params.	Training	Bad.	Cro.	Dav.	Dri.	Ego.	Hor.	Kin.	Rgb.	Rob.	Avg.
RAFT [47]	5.26	Flow mix	23.7	29.3	48.5	44.8	41.0	27.8	64.3	82.8	72.2	48.3
SEA-RAFT [54]	19.66	Flow mix	23.9	21.9	48.7	49.4	44.0	33.1	64.3	85.7	67.6	48.7
AccFlow [55]	11.76	Flow mix	10.3	22.2	23.5	26.4	4.0	12.1	38.8	63.2	57.9	28.7
PIPs++ [59]	17.57	PointOdyssey	34.1	27.5	62.5	51.3	38.5	21.4	64.2	70.4	73.4	49.3
LocoTrack [6]	11.52	Kubric	41.4	43.1	68.0	66.5	58.4	48.9	70.0	80.3	76.9	61.5
BootsTAPIR [13]	54.70	Kubric+15M	42.7	34.9	67.9	66.9	56.8	48.8	70.6	81.0	78.2	60.9
DELTA [37]	59.17	Kubric	44.6	42.9	75.3	67.8	40.3	41.8	66.5	83.0	74.8	59.7
CoTracker2 [23]	45.43	Kubric	40.0	31.7	70.9	67.8	43.2	33.9	65.8	73.4	73	55.5
CoTracker3-Kub [25]	25.39	Kubric	47.5	**48.9**	**77.4**	**69.8**	58.0	47.5	70.6	83.4	77.2	64.5
CoTracker3 [25]	25.39	Kubric+15k	48.3	44.5	77.1	**69.8**	60.4	47.1	71.8	84.2	81.6	65.0
AllTracker-Tiny-Kub	6.29	Kubric	45.4	39.6	73.7	65.1	55.9	45.2	70.6	86.1	79.3	62.3
AllTracker-Tiny	6.29	Kubric+mix	47.5	39.8	74.3	63.9	58.3	45.5	71.5	88.1	80.7	63.3
AllTracker-Kub	16.48	Kubric	46.4	42.3	75.2	66.1	60.3	**49.0**	71.3	**90.1**	82.2	64.8
AllTracker	16.48	Kubric+mix	**51.5**	44.0	76.3	65.8	**62.5**	**49.0**	**72.3**	90.0	**83.4**	**66.1**

Harley et al (2025). AllTracker: Efficient Dense Point Tracking at High Resolution. ICCV

In today's talk…

Motivation and Datasets in Egocentric Video Understanding

Video Understanding Out of the Frame

- Point Tracking
- Object Tracking
- Gaze Priming
- Hand Tracking

Teaser: The Wizard of Oz & Genie 3

Outlook into the Future of Egocentric Vision

Conclusion

Spatial Cognition from Egocentric Video:

Out of Sight, Not Out of Mind

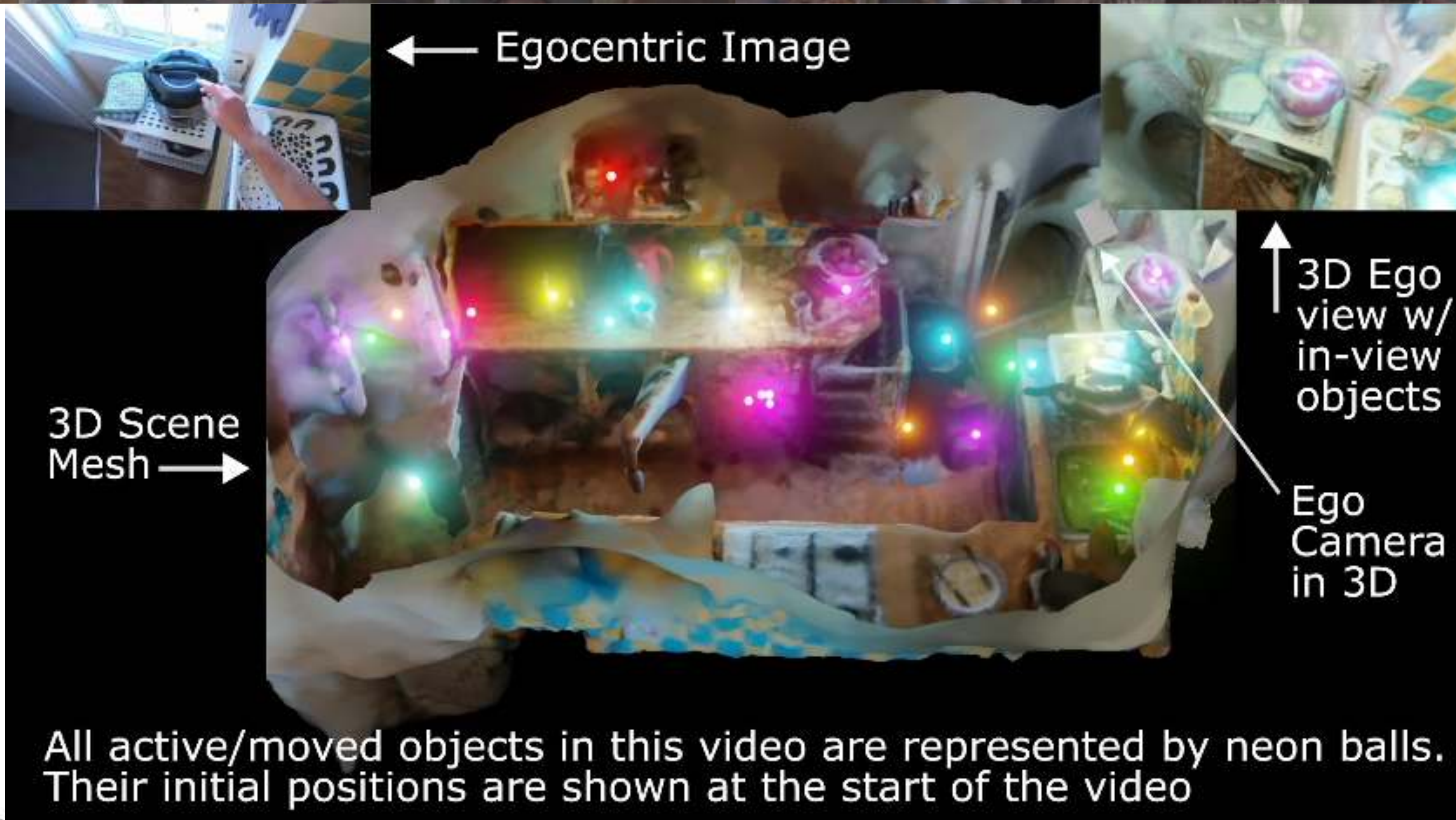
Chiara Plizzari Shubham Goel Toby Perrett Jacob Chalk
Angjoo Kanazawa Dima Damen

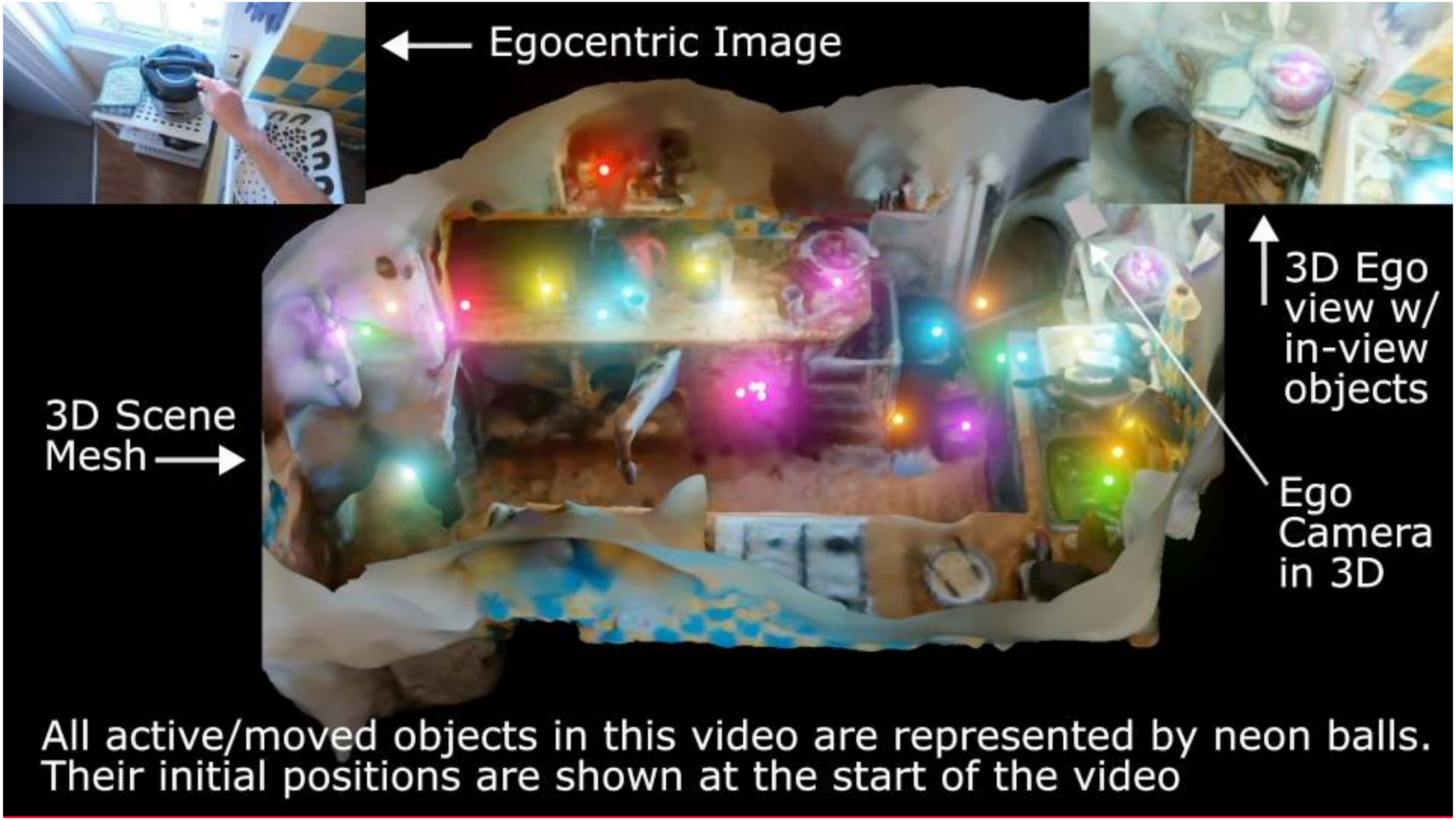
http://dimadamen.github.io/OSNOM

Politecnico di Torino
Berkeley UNIVERSITY OF CALIFORNIA
University of BRISTOL

Out of Sight, not Out of Mind

with: Chiara Plizzari
Toby Perrett
Shubham Goel
Angjoo Kanazawa



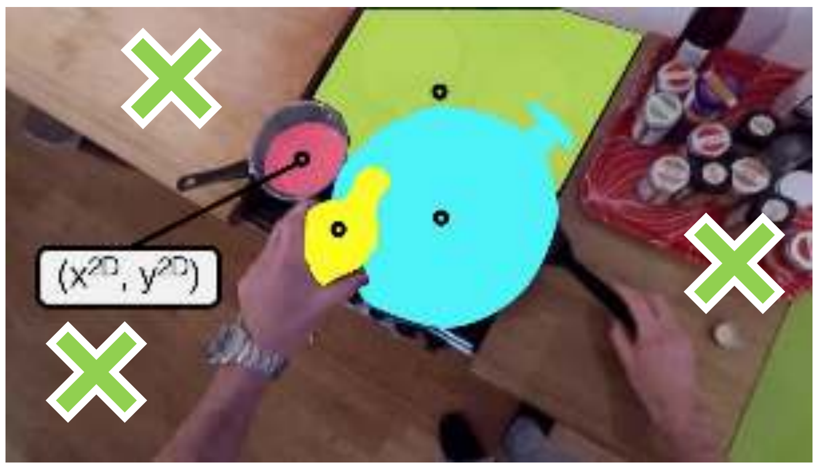


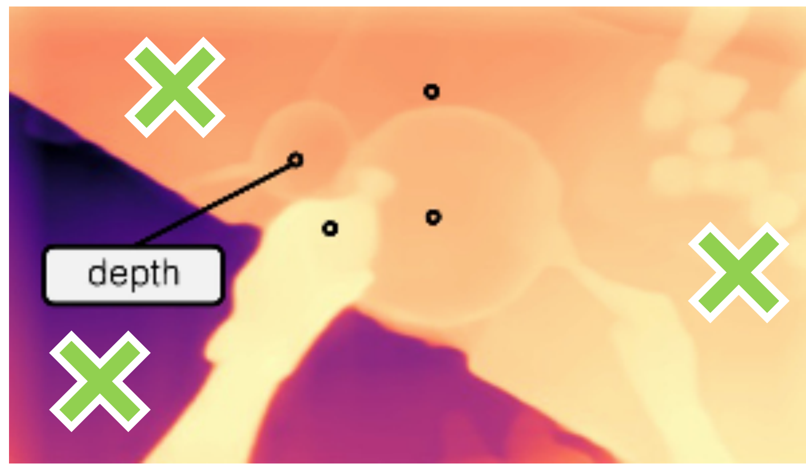
Egocentric Image
3D Ego view w/ in-view objects
3D Scene Mesh
Ego Camera in 3D
All active/moved objects in this video are represented by neon balls.
Their initial positions are shown at the start of the video

Out of Sight, not Out of Mind

with: Chiara Plizzari Toby Perrett Shubham Goel Angjoo Kanazawa

Lift → Match → Keep



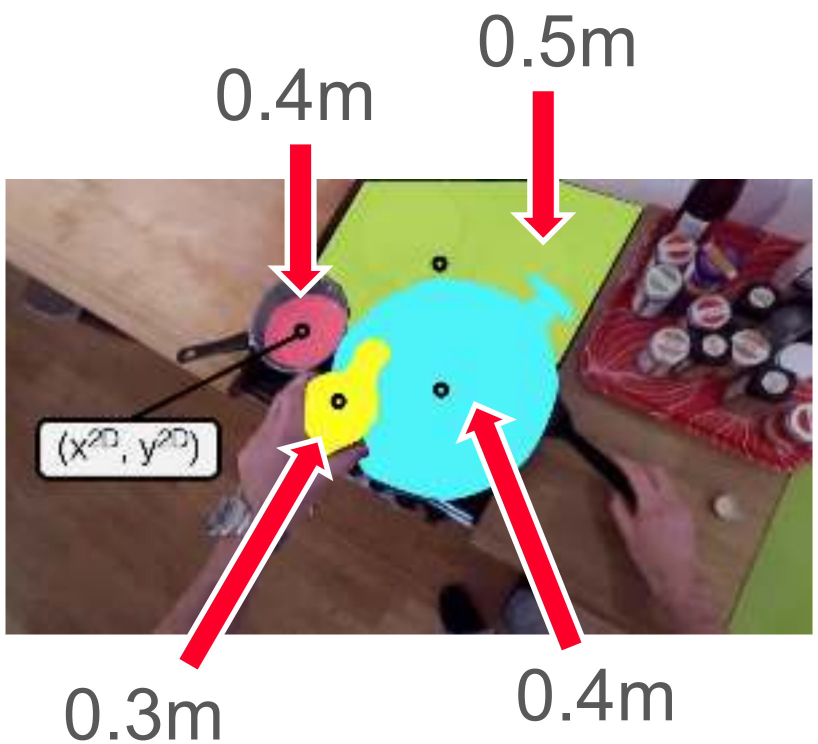


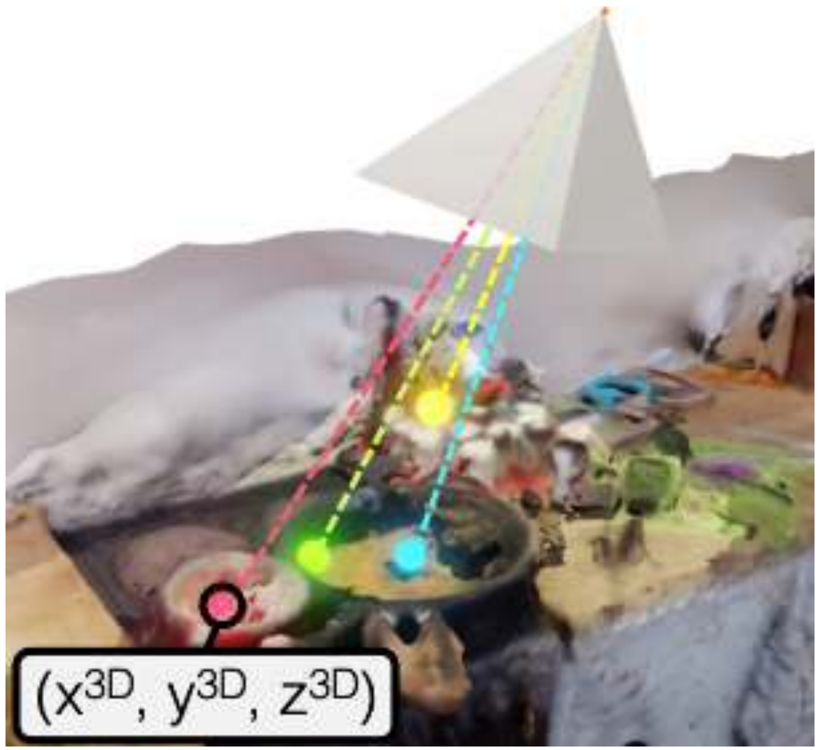
0.0 … 1.0

0.3m … 1.8m

with: Chiara Plizzari Toby Perrett Shubham Goel Angjoo Kanazawa

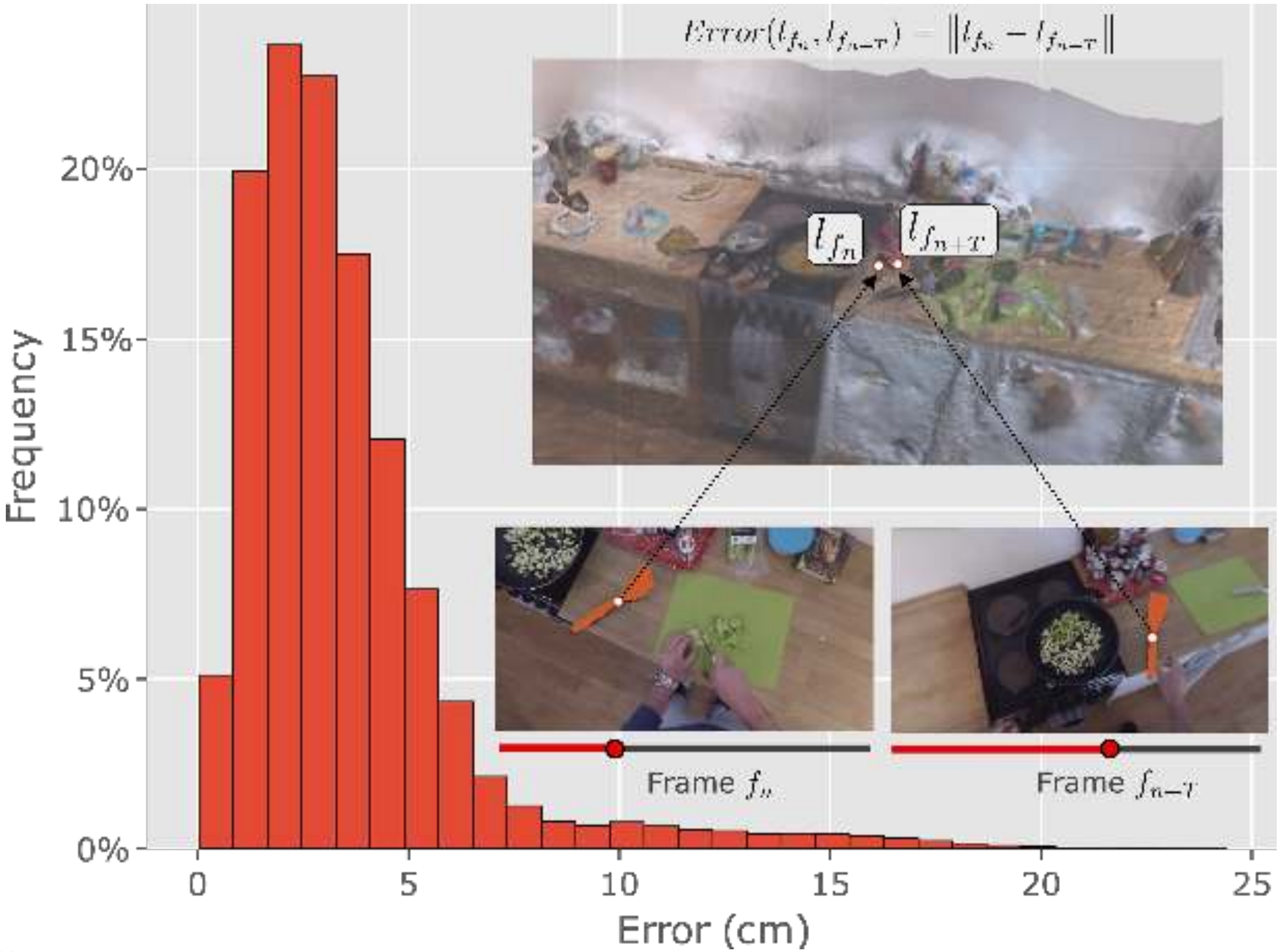
Lift → Match → Keep





Out of Sight, not Out of Mind

with: Chiara Plizzari, Toby Perrett, Shubham Goel, Angjoo Kanazawa



Plizzari et al (2025). Spatial Cognition from Egocentric Video: Out of Sight, Not Out of Mind. 3DV

with: Chiara Plizzari
Toby Perrett
Shubham Goel
Angjoo Kanazawa

Lift → Match → Keep

Instead of tracking in 2D, we track in 3D, using combination of appearance and location distances

Out of Sight, not Out of Mind

with: Chiara Plizzari Toby Perrett Shubham Goel Angjoo Kanazawa

After we Lift, Match and Keep (LMK), we can reason about an object's visibility and position

- In-View vs Out-of-View
- In-Sight vs Out-of-Sight (Occluded)
- Within-Reach vs Out-of-Reach (defining the camera wearer's near space)

Plizzari et al (2025). Spatial Cognition from Egocentric Video: Out of Sight, Not Out of Mind. 3DV

Out of Sight, not Out of Mind

with: Chiara Plizzari
Toby Perrett
Shubham Goel
Angjoo Kanazawa

After we Lift, Match and Keep (LMK), we can reason about an object's visibility and position

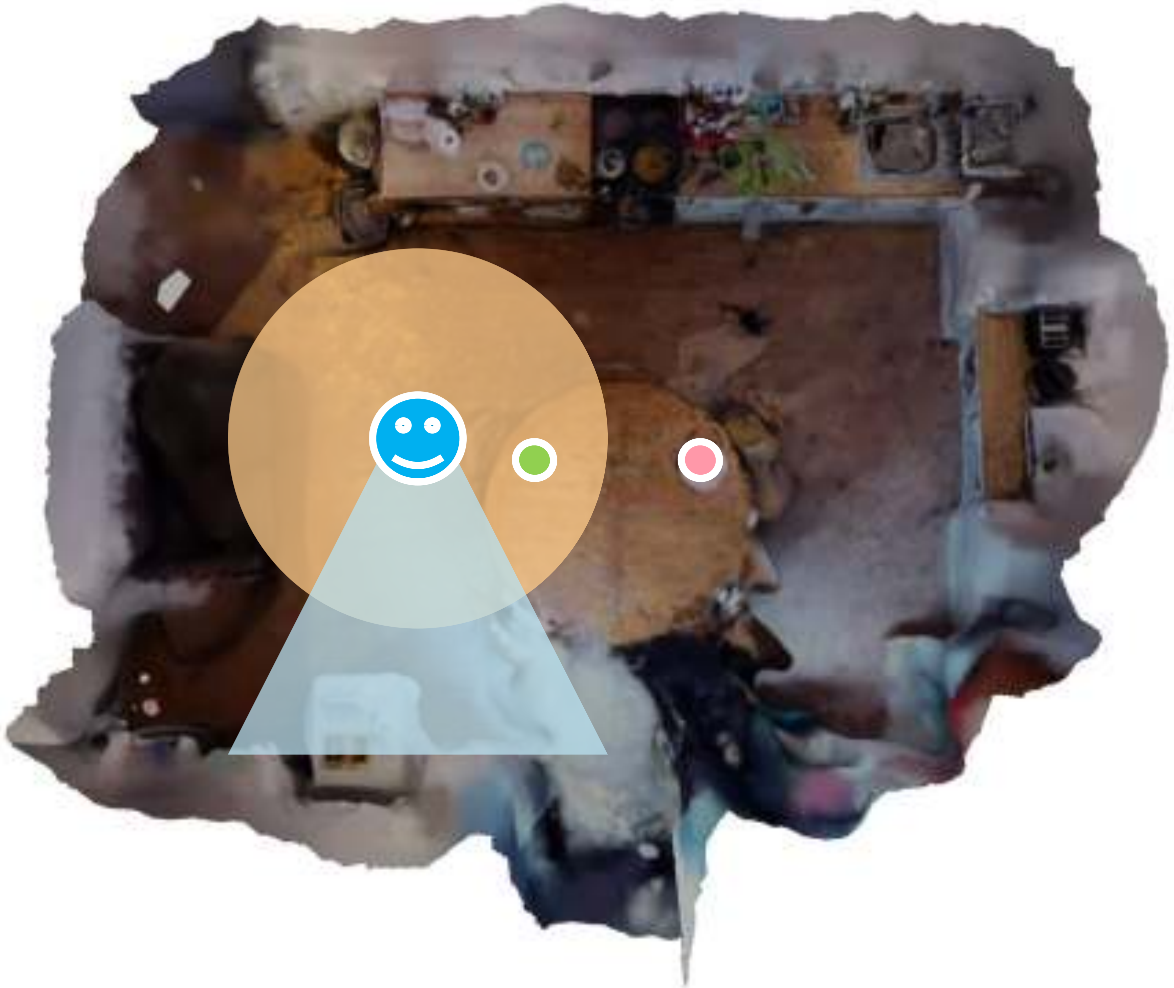
- In-View vs Out-of-View
- In-Sight vs Out-of-Sight (Occluded)
- Within-Reach vs Out-of-Reach (defining the camera wearer's near space)

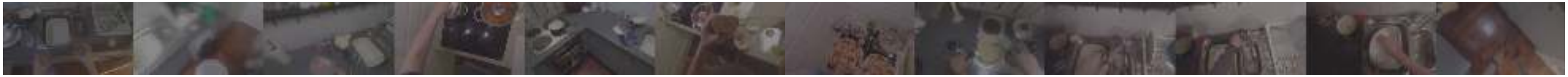
Out of Sight, not Out of Mind

with: Chiara Plizzari, Toby Perrett, Shubham Goel, Angjoo Kanazawa

After we Lift, Match and Keep (LMK), we can reason about an object's visibility and position

- In-View vs Out-of-View
- In-Sight vs Out-of-Sight (Occluded)
- Within-Reach vs Out-of-Reach (defining the camera wearer's near space)





Spatial Cognition from Egocentric Video:

Out of Sight, Not Out of Mind

Ground-Truth??

Chiara Plizzari Shubham oby Perrett Jacob Chalk

Ang a Dima Damen

ttp://dimadamen.github.io/OSNOM

CVPR Nashville JUNE 11-15, 2025

HD-EPIC: A Highly-Detailed Egocentric Video Dataset

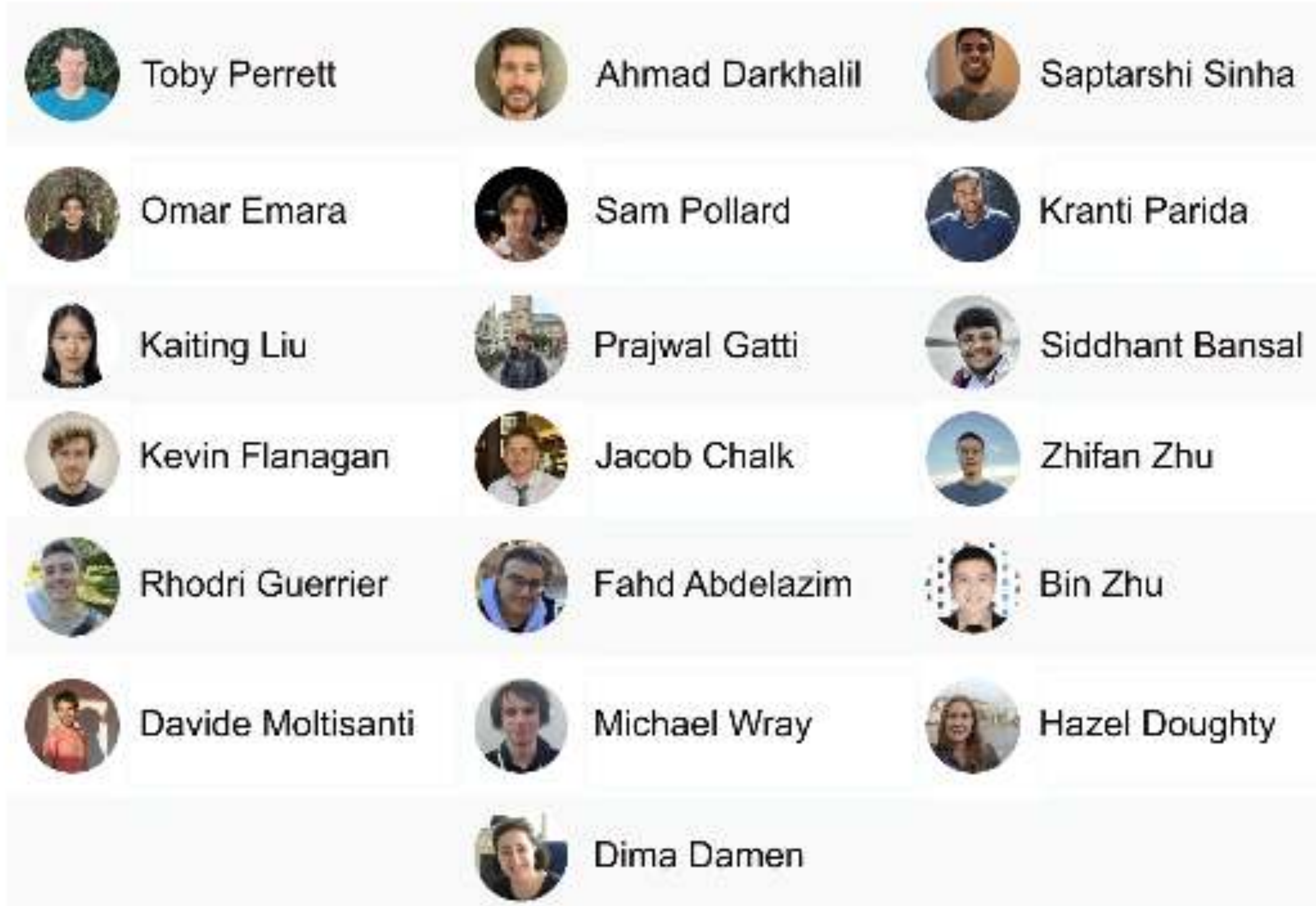
HD-EPIC

University of Bristol

Universiteit Leiden

Singapore Management University

University of Bath

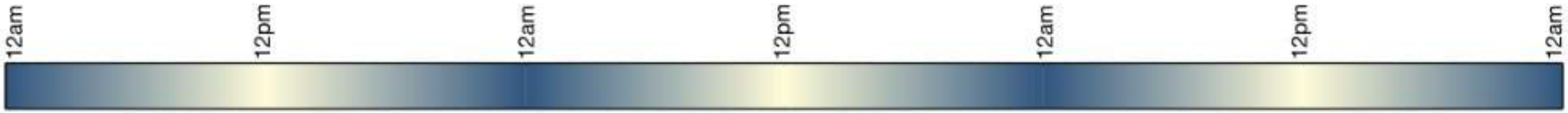


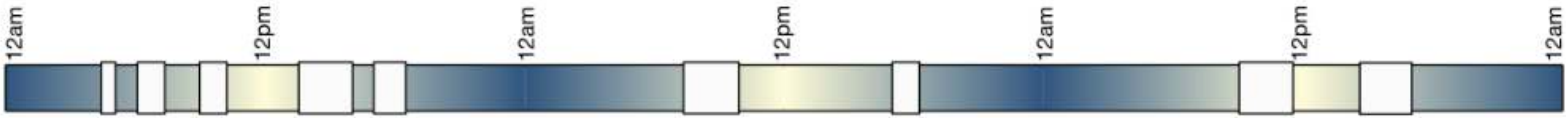
University of Bristol

Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR

HD-EPIC

HD-EPIC





Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

Dima Damen
BinEgo-360 Workshop @ICCV2025

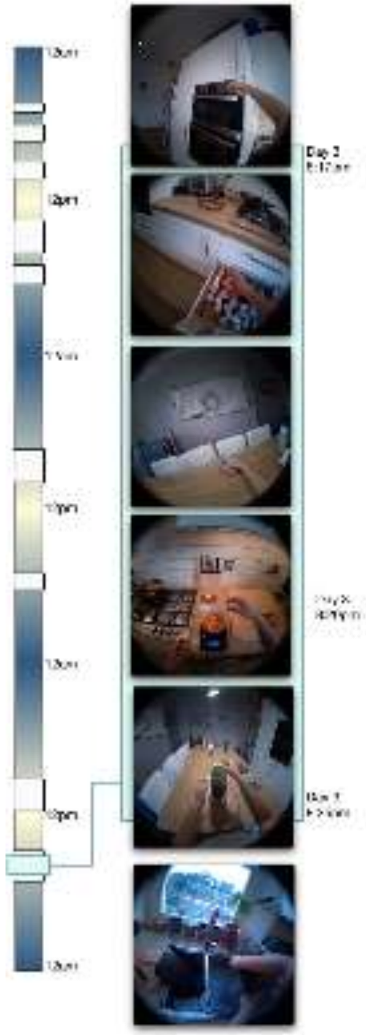
HD-EPIC

HD-EPIC

Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

Dima Damen
BinEgo-360 Workshop @ICCV2025

HD-EPIC



Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

HD-EPIC

ay 3
:17pm

Recipe: Southwestern Salad

1: Preheat the oven to 400F

2: Wash and peel the sweet potatoes and chop into bite-sized pieces. Put the sweet potatoes in a bowl and add the olive oil, cumin, and chili powder. Pour onto tray and roast for 10 mins.

3: Pulse all the dressing ingredients in a food processor until mostly smooth.

Recipe and nutrition

HD-EPIC

Cacio e Pepe (modified)

Ingredients:

200 g ~~400g~~ of pasta of your choice penne (we recommend bucatini)

1 ~~2~~ tablespoon of black peppercorn

30 g ~~200g~~ of freshly grated pecorino cheese parmigiano

+25g of slightly salted butter

Steps:

1. Toast the peppercorns until fragrant in a dry frying pan over medium heat, about 2 minutes. Keep them moving to prevent them from burning. ~~Once toasted, roughly crush.~~ → step 2

2. Cook your choice of pasta in a large pot of generously salted boiling water ~~for around 4-6 minutes~~, or until al dente. → step 1

3. While the pasta cooks, add freshly grated cheese and crushed black peppercorns to a large serving bowl on very low heat. Gradually add a cup of the boiling cooking water constantly mixing to obtain a silky, smooth sauce that's able to completely coat the pasta. → step 3

University of BRISTOL

Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

Dima Damen
BinEgo-360 Workshop @ICCV2025

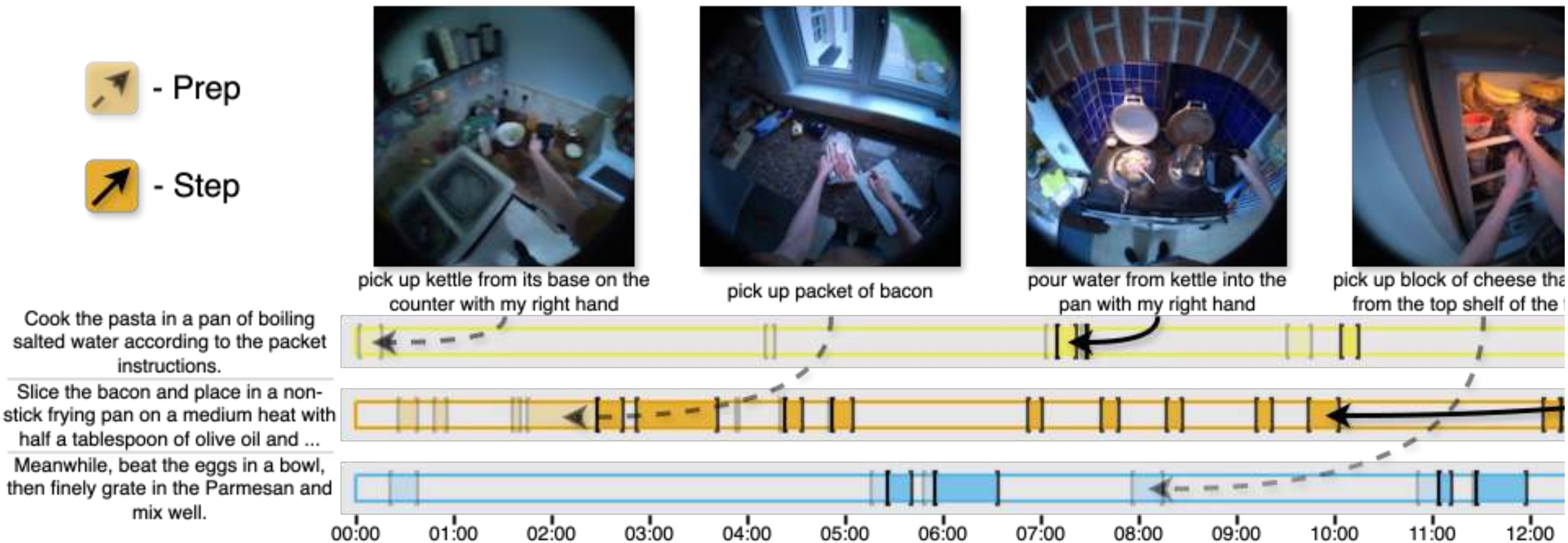
HD-EPIC

- The **prep** of a corresponding **step** is defined as
all essential actions the participant takes to get ready to execute a given step.

- For example, the **step** ‘chop tomato’:
  - **Prep:** retrieve tomato from storage, wash tomato, retrieve a knife and chopping board.

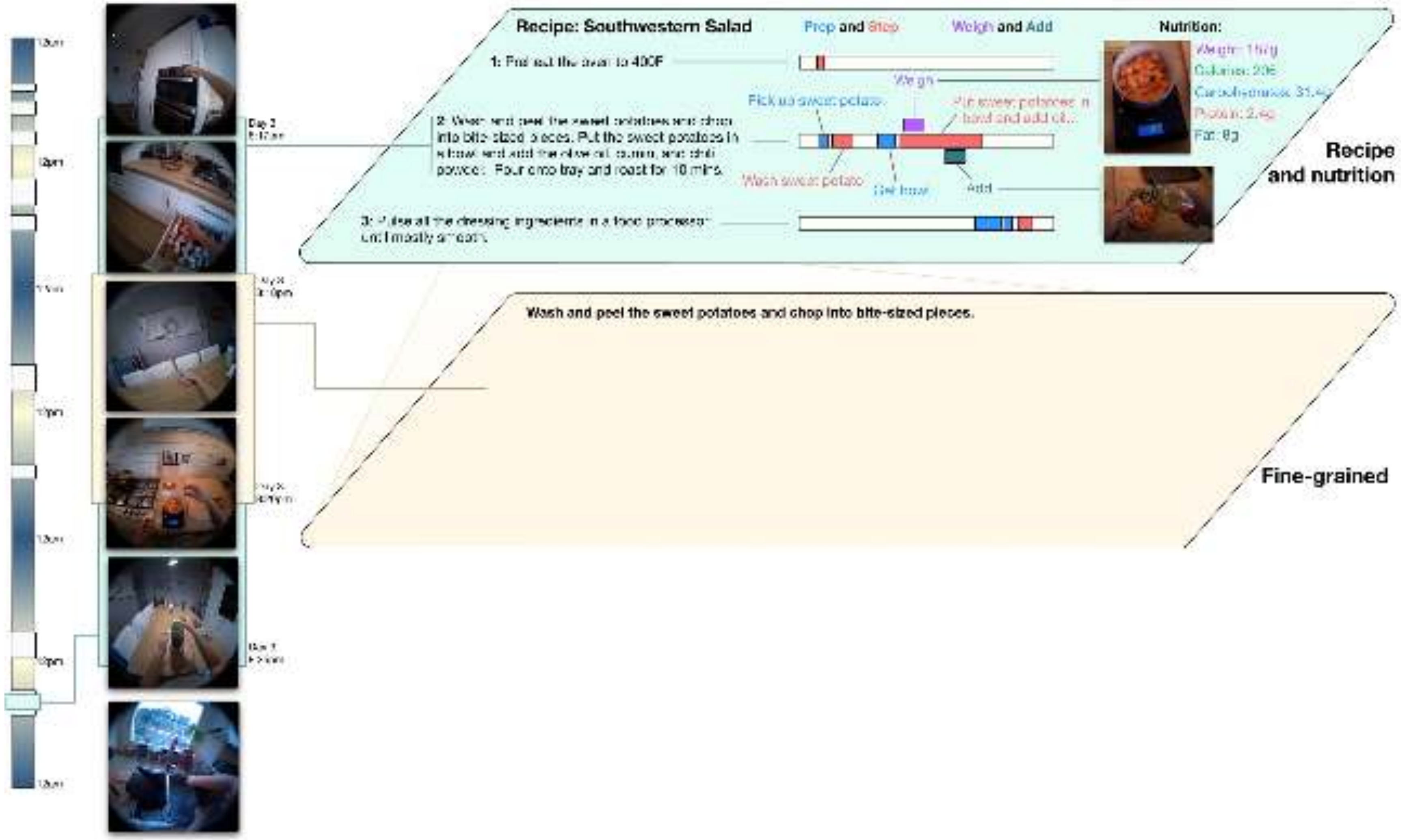
- the **step** ‘add chopped onions and stir’:
  - **Prep:** retrieve onion from storage, retrieve a knife and chopping board, **and chop the onions.**

HD-EPIC



- Prep
- Step
pick up kettle from its base on the counter with my right hand
pick up packet of bacon
pour water from kettle into the pan with my right hand
pick up block of cheese tha
from the top shelf of the
Cook the pasta in a pan of boiling salted water according to the packet instructions.
Slice the bacon and place in a non-stick frying pan on a medium heat with half a tablespoon of olive oil and ...
Meanwhile, beat the eggs in a bowl, then finely grate in the Parmesan and mix well.
00:00
01:00
02:00
03:00
04:00
05:00
06:00
07:00
08:00
09:00
10:00
11:00
12:00

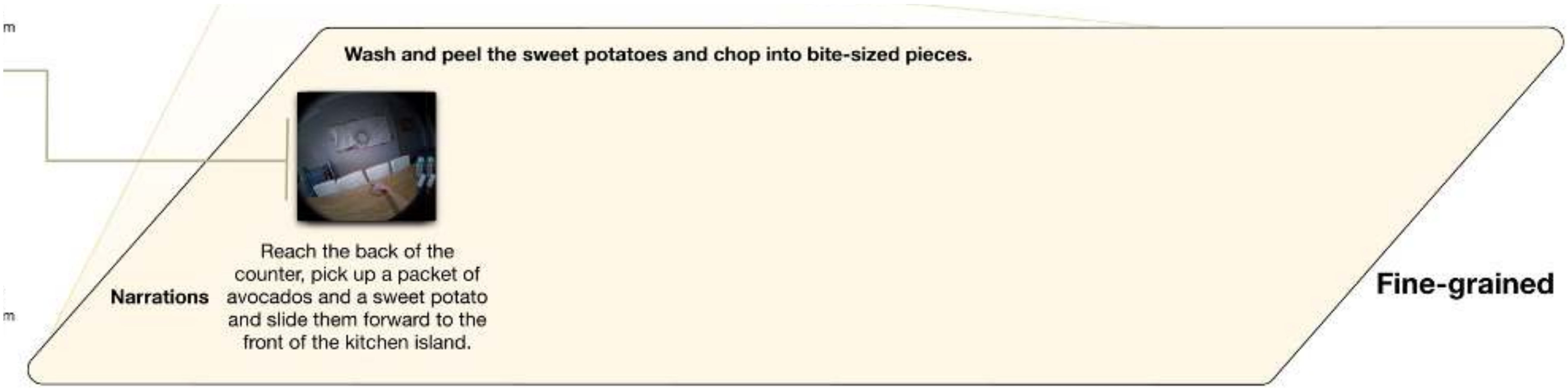
HD-EPIC



Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

Dima Damen
BinEgo-360 Workshop @ICCV2025

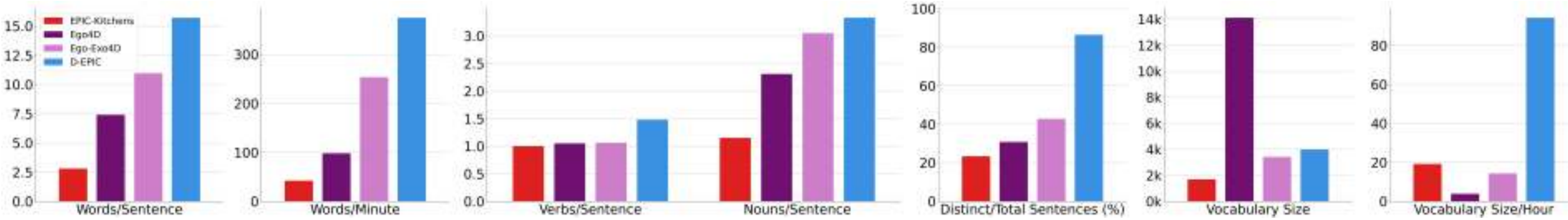
HD-EPIC


m
Wash and peel the sweet potatoes and chop into bite-sized pieces.
Narrations
Reach the back of the counter, pick up a packet of avocados and a sweet potato and slide them forward to the front of the kitchen island.
m
Fine-grained

Highly-Detailed Narrations

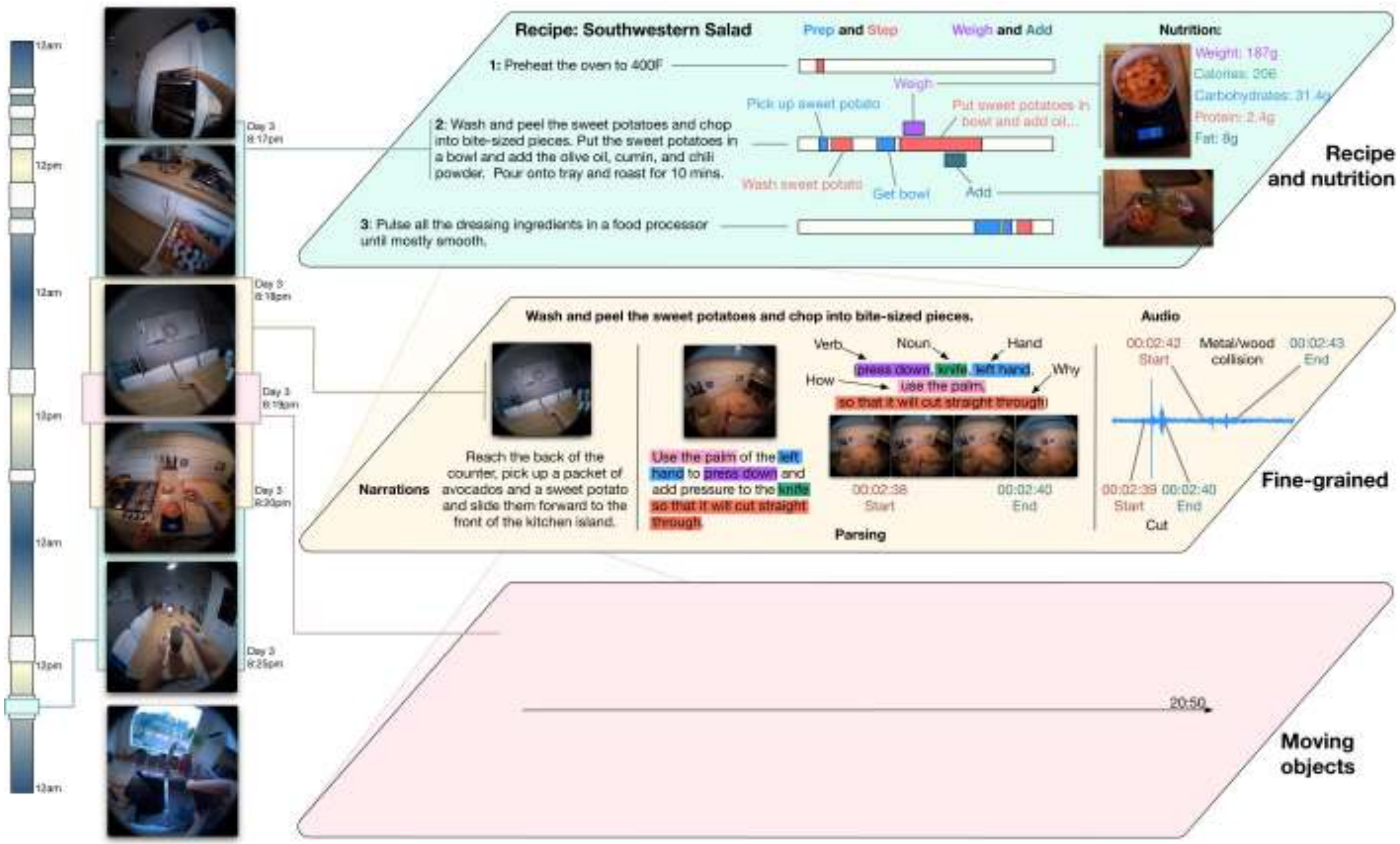
HD-EPIC

- 59,454 fine-grained actions, with a mean duration of 2.0s (±3.4s).



Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

HD-EPIC

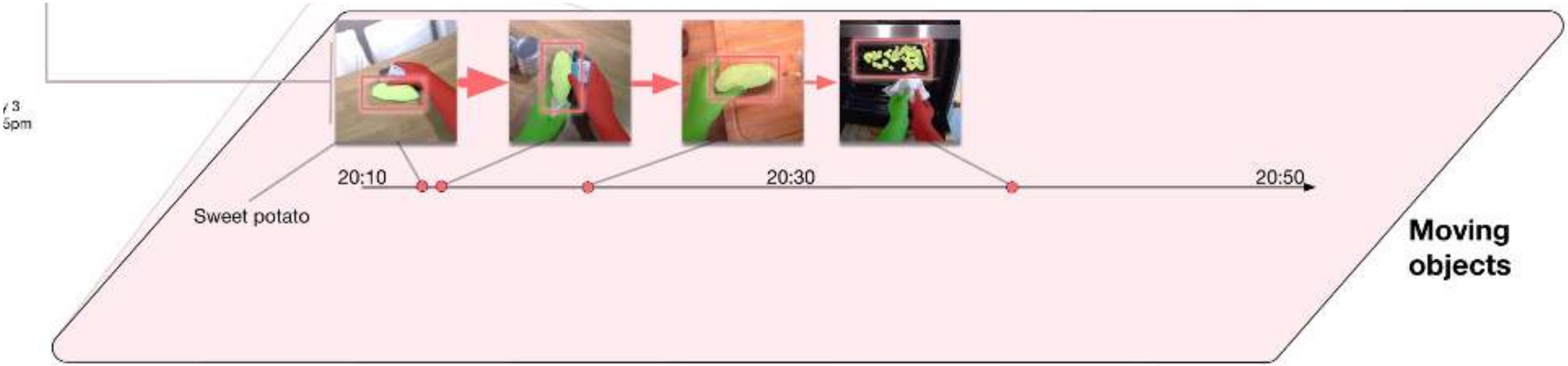


University of BRISTOL

Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

Dima Damen
BinEgo-360 Workshop @ICCV2025

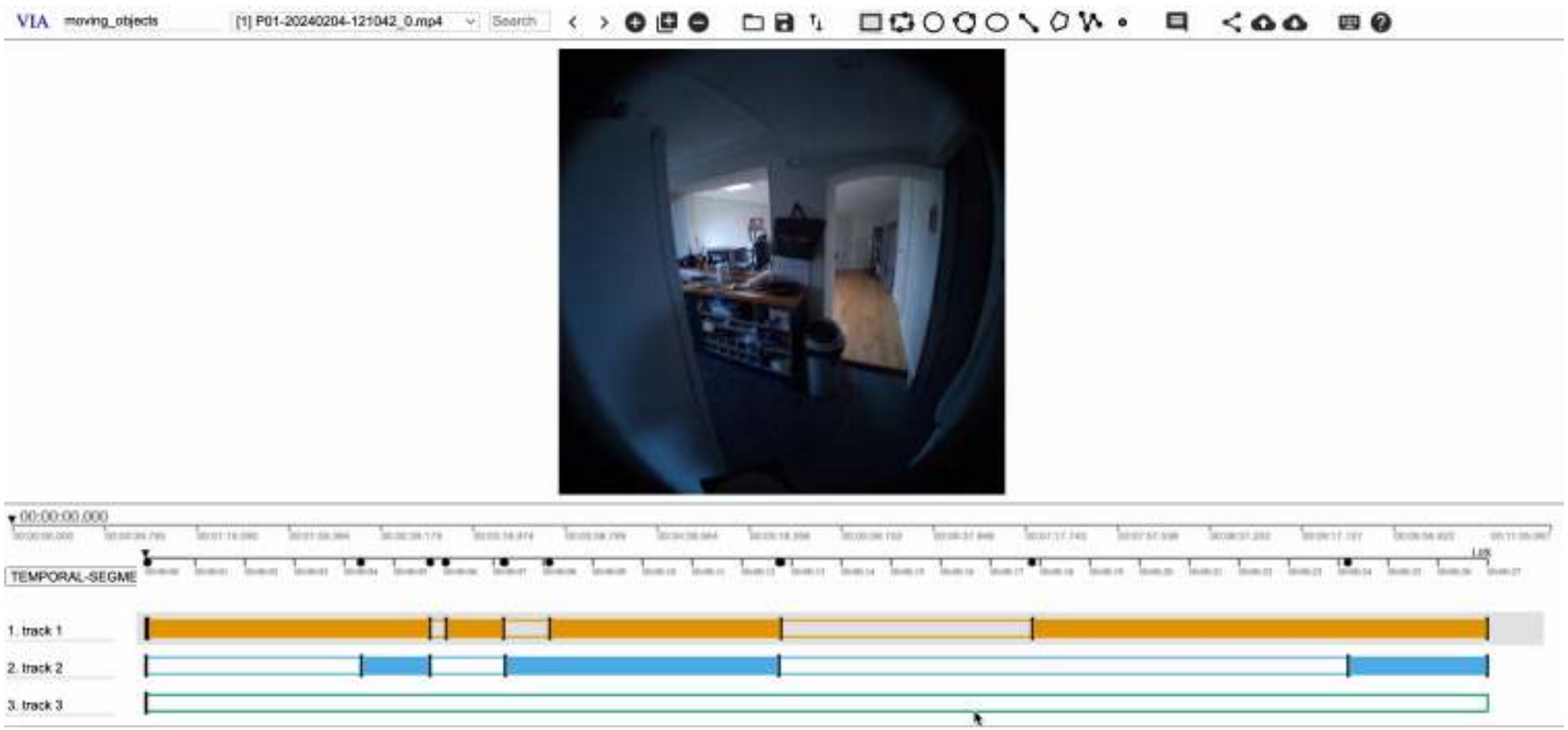
HD-EPIC



HD-EPIC

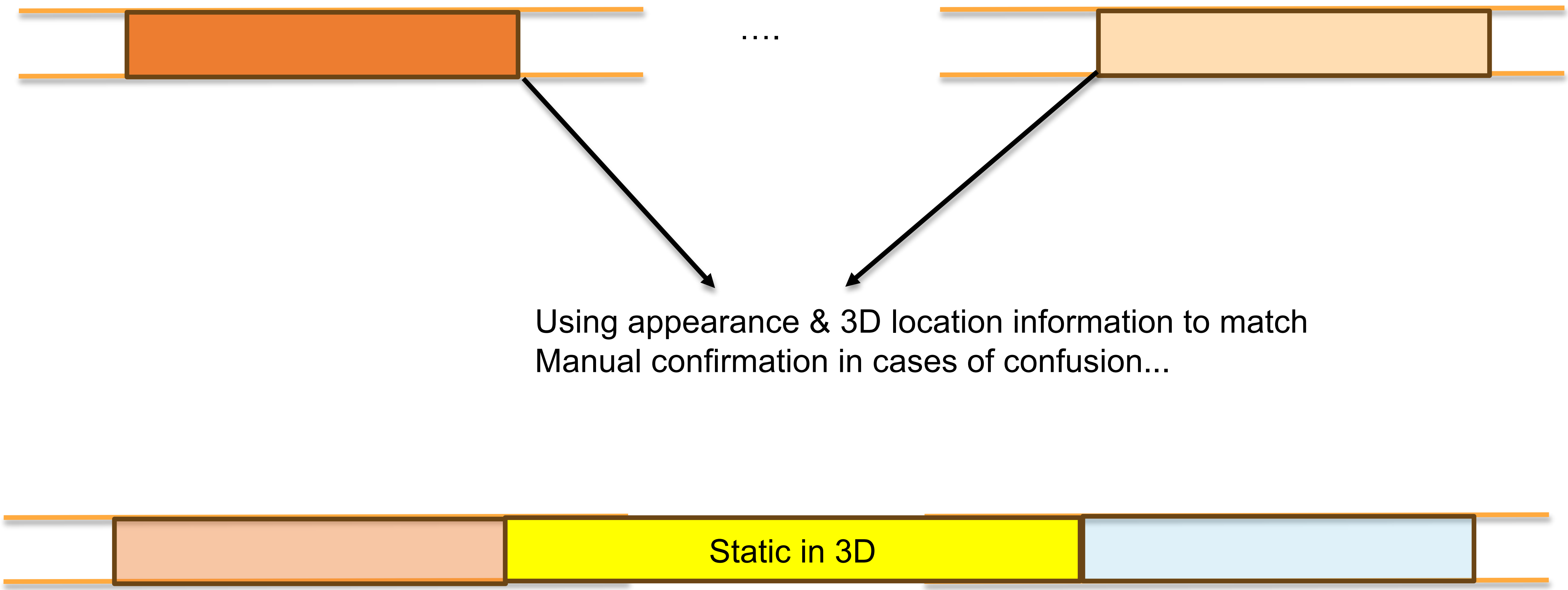
HD-EPIC

- How to minimize the annotations for tracking objects…



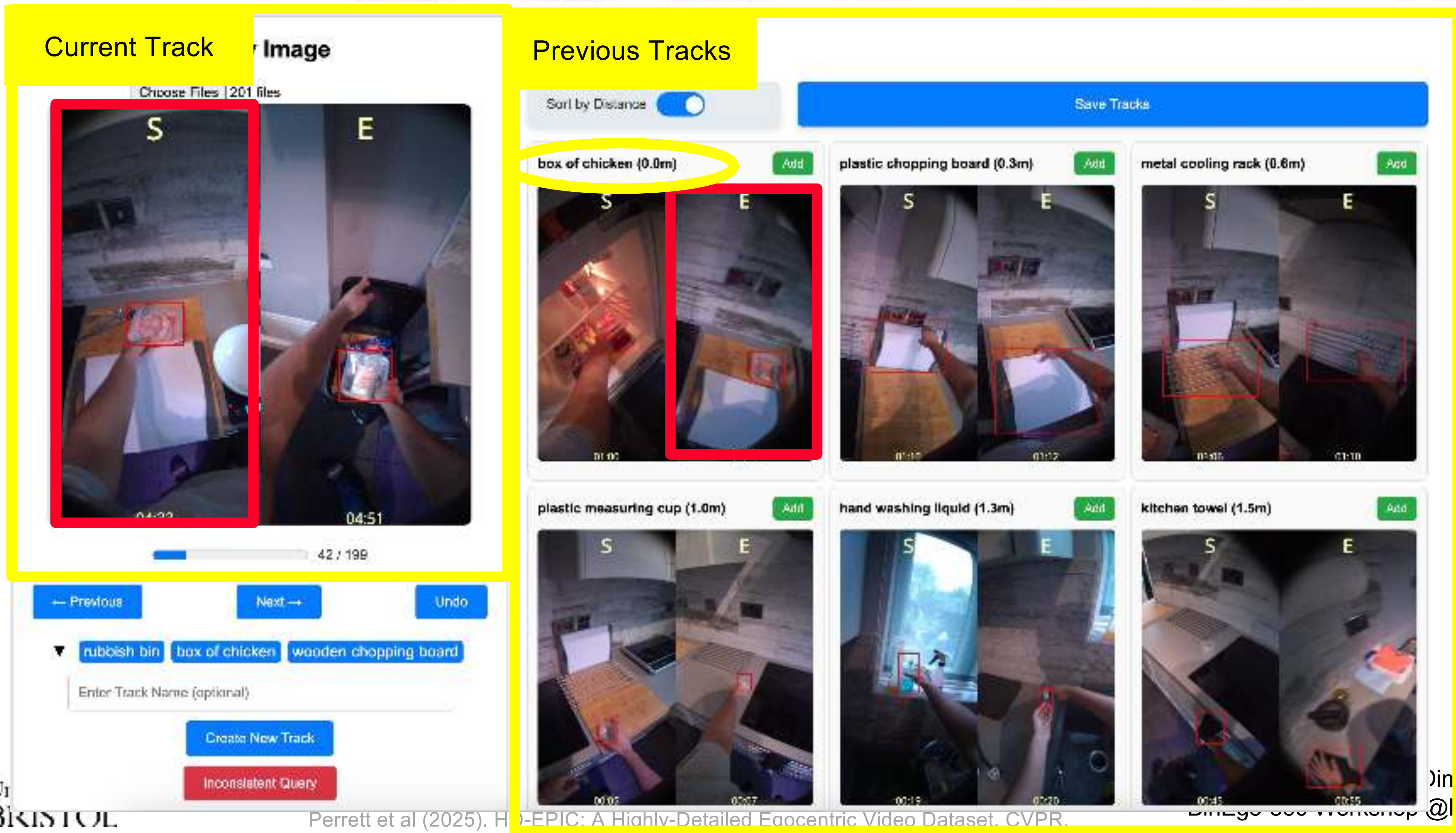
HD-EPIC

- How to minimize the annotations for tracking objects…



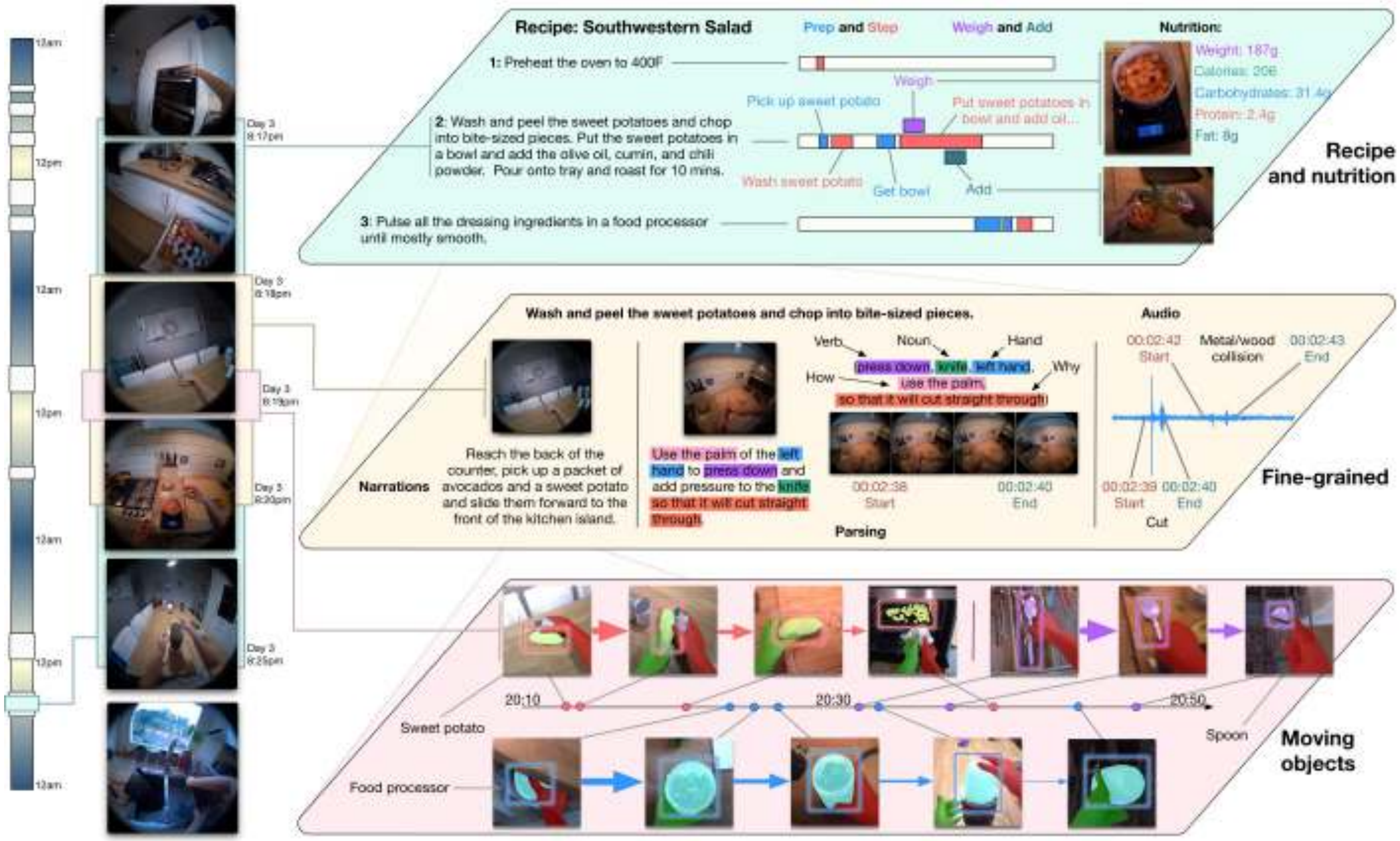
Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

HD-EPIC



Current Track
Image
Choose Files | 201 files
S
E
04:51
42 / 199
← Previous
Next →
Undo
rubbish bin
box of chicken
wooden chopping board
Enter Track Name (optional)
Create New Track
Inconsistent Query
Previous Tracks
Sort by Distance
Save Tracks
box of chicken (0.0m)
Add
plastic chopping board (0.3m)
metal cooling rack (0.6m)
plastic measuring cup (1.0m)
hand washing liquid (1.3m)
kitchen towel (1.5m)

HD-EPIC

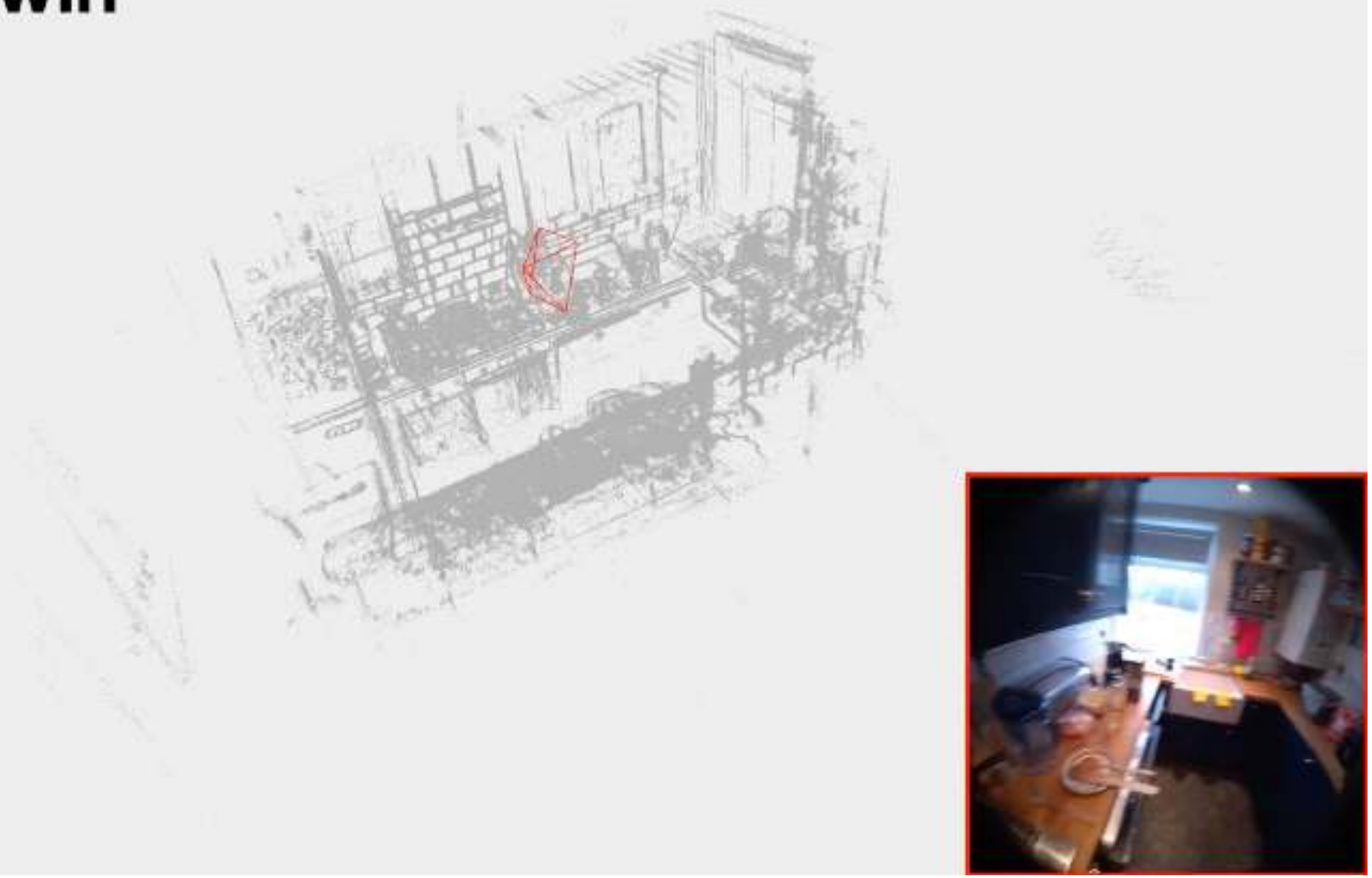


University of BRISTOL

Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

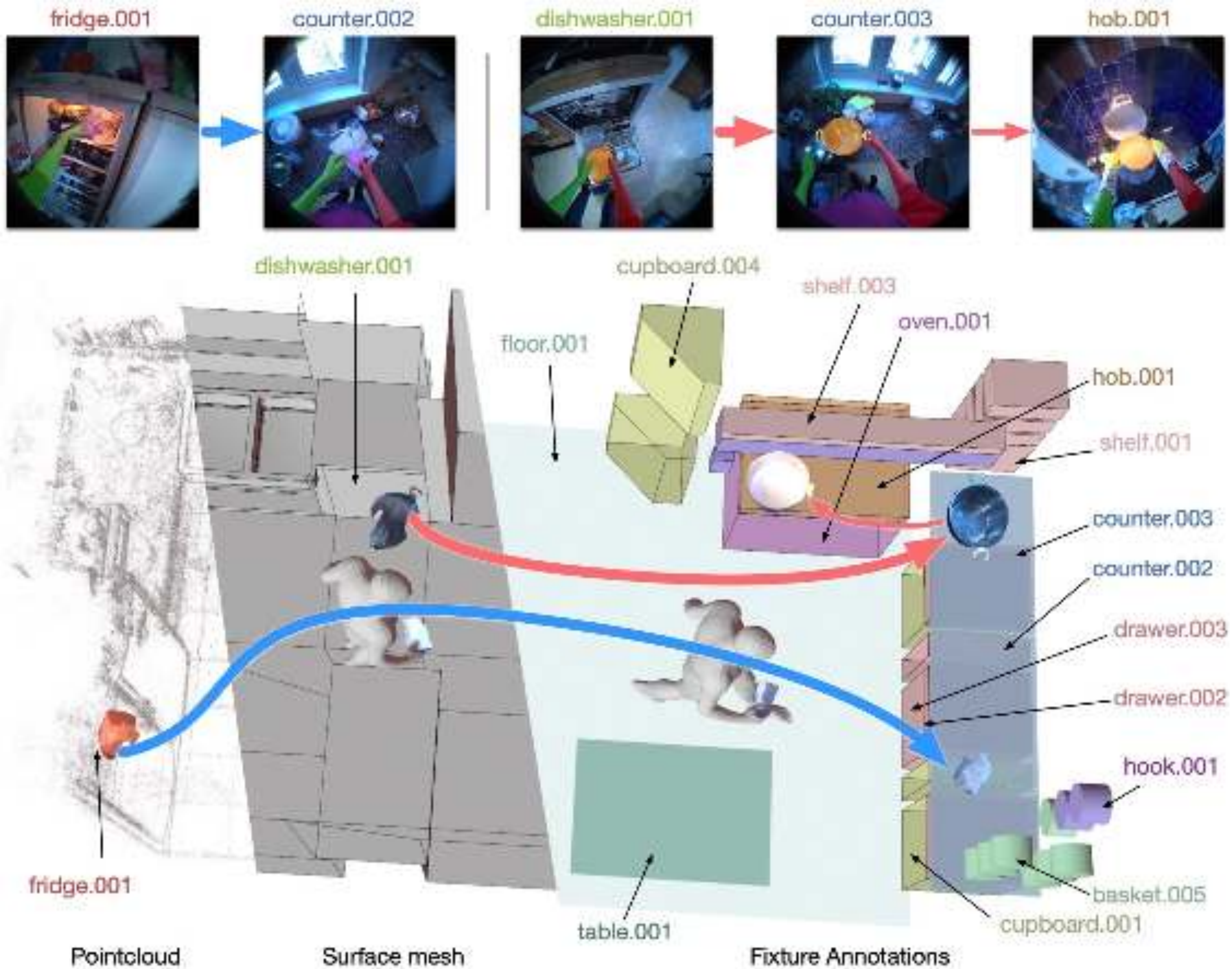
HD-EPIC

Digital Twin



Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

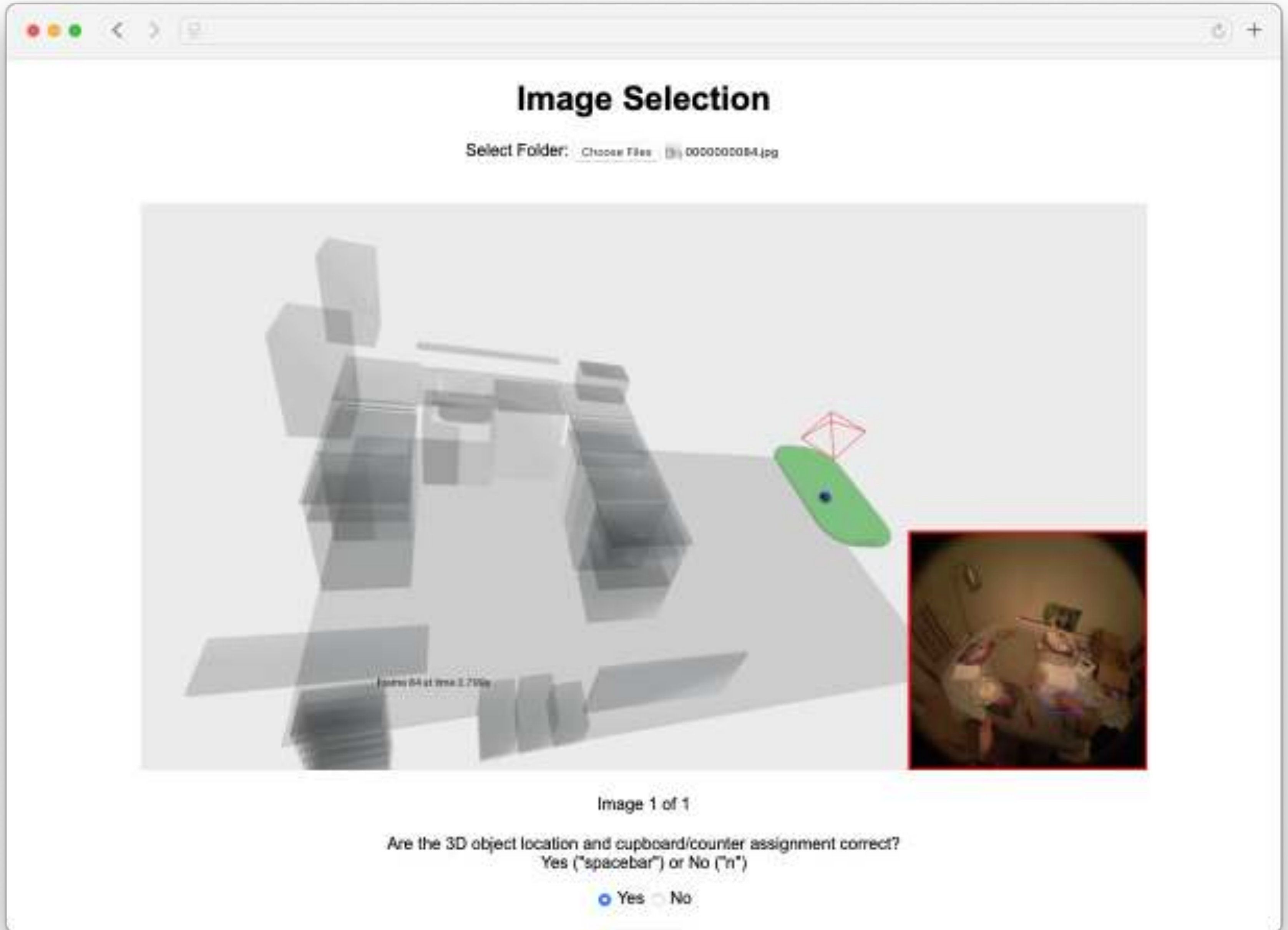
HD-EPIC



Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

Dima Damen
BinEgo-360 Workshop @ICCV2025

HD-EPIC



Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

Dima Damen
BinEgo-360 Workshop @ICCV2025

[redacted]

In today's talk…

Motivation and Datasets in Egocentric Video Understanding

Video Understanding Out of the Frame

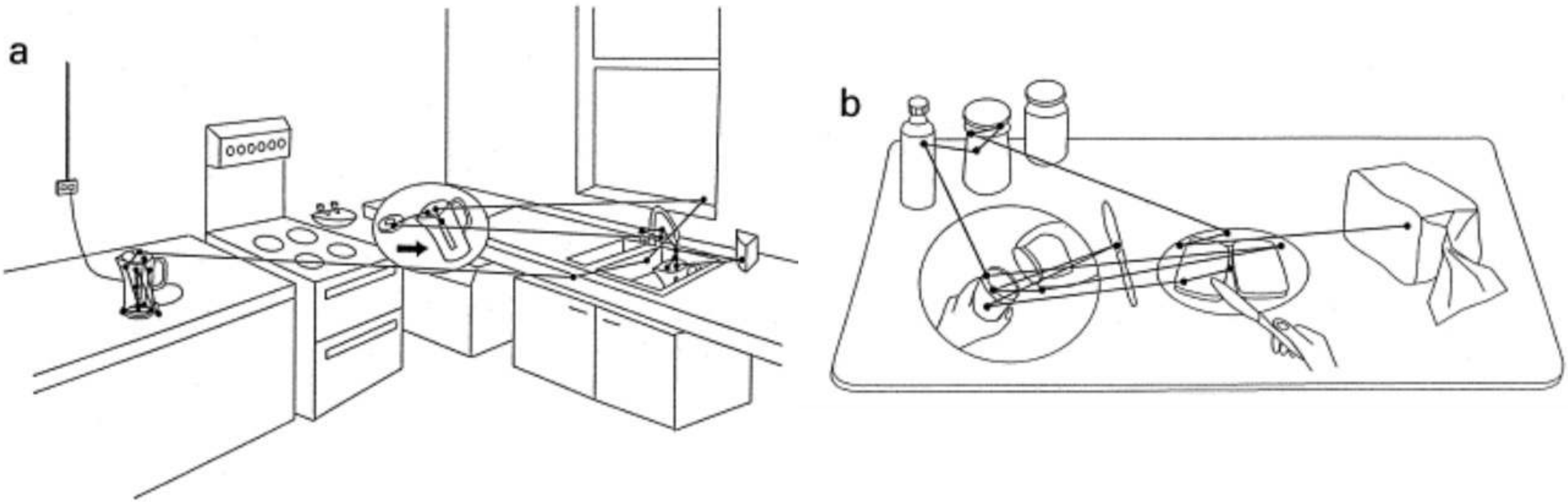
- Point Tracking
- Object Tracking
- Gaze Priming
- Hand Tracking

Teaser: The Wizard of Oz & Genie 3

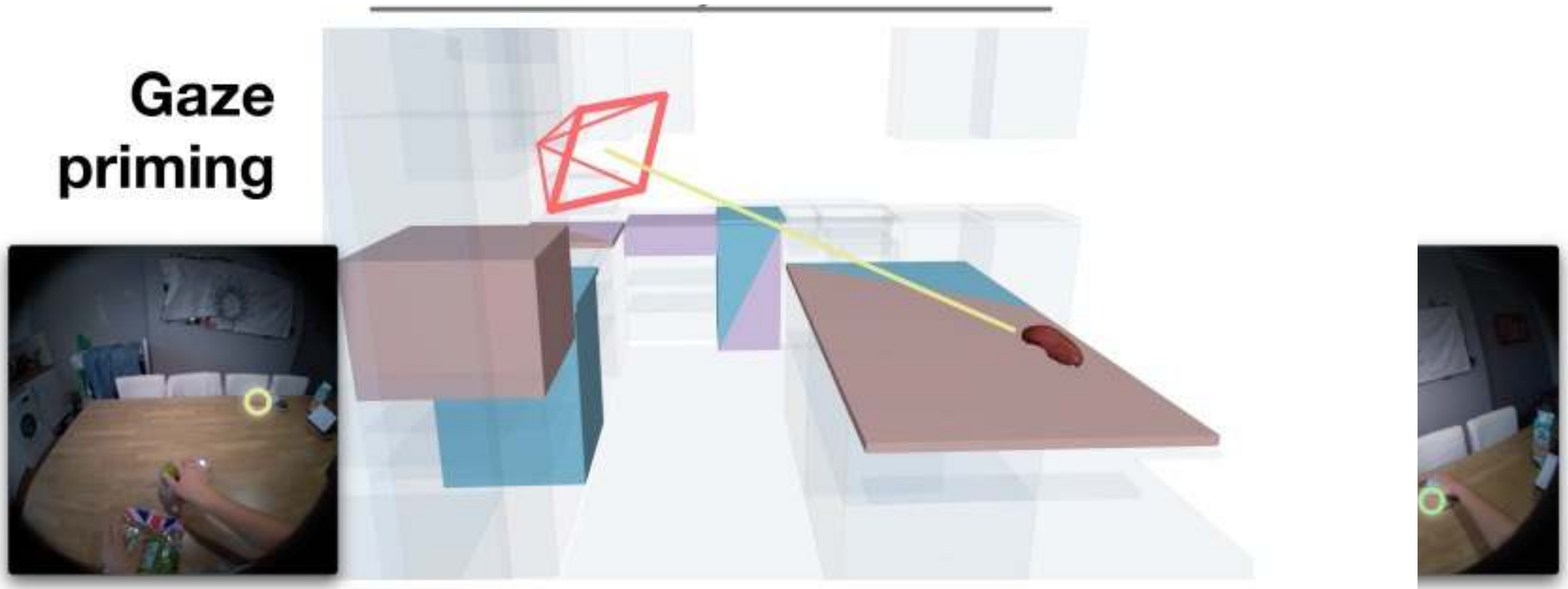
Outlook into the Future of Egocentric Vision

Conclusion

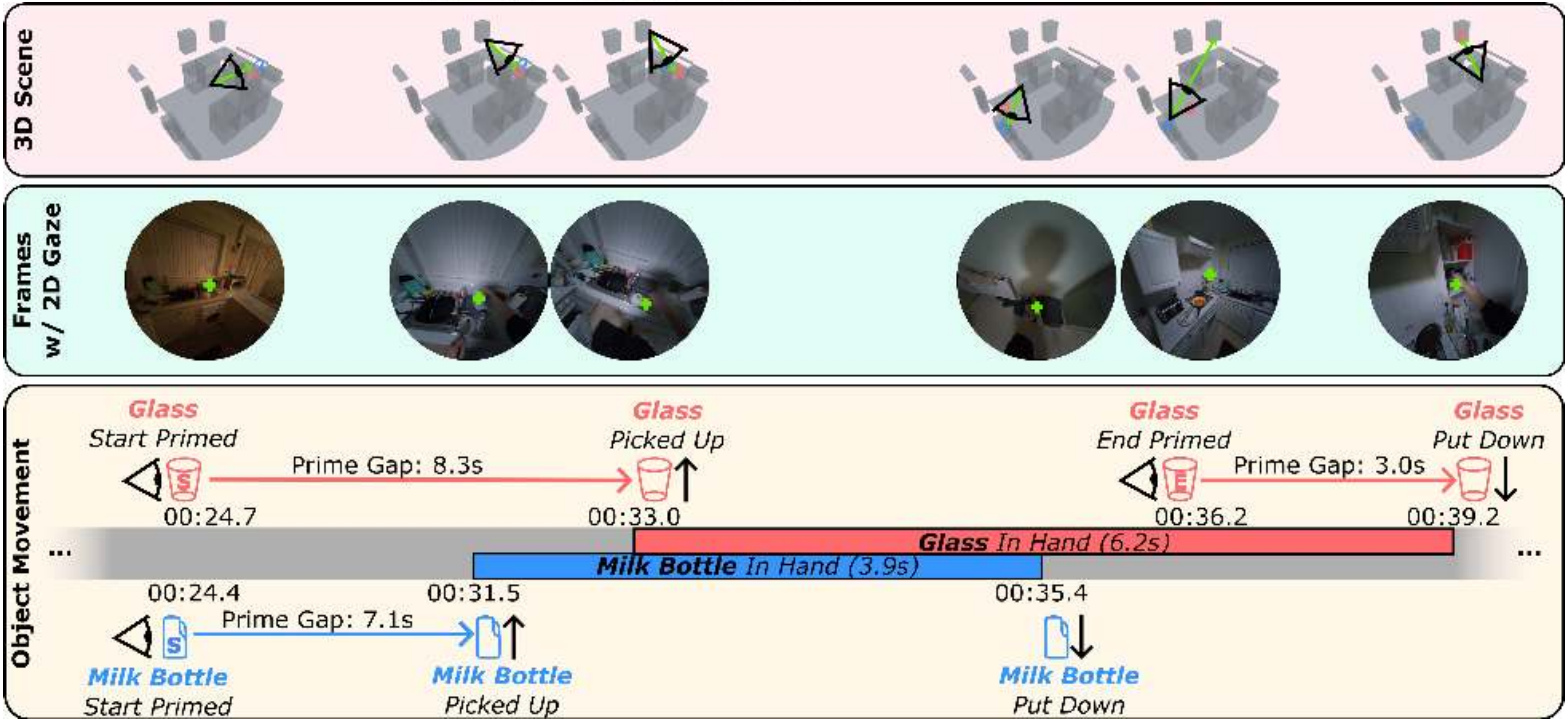
Gaze and Fixations



Gaze priming



HD-EPIC

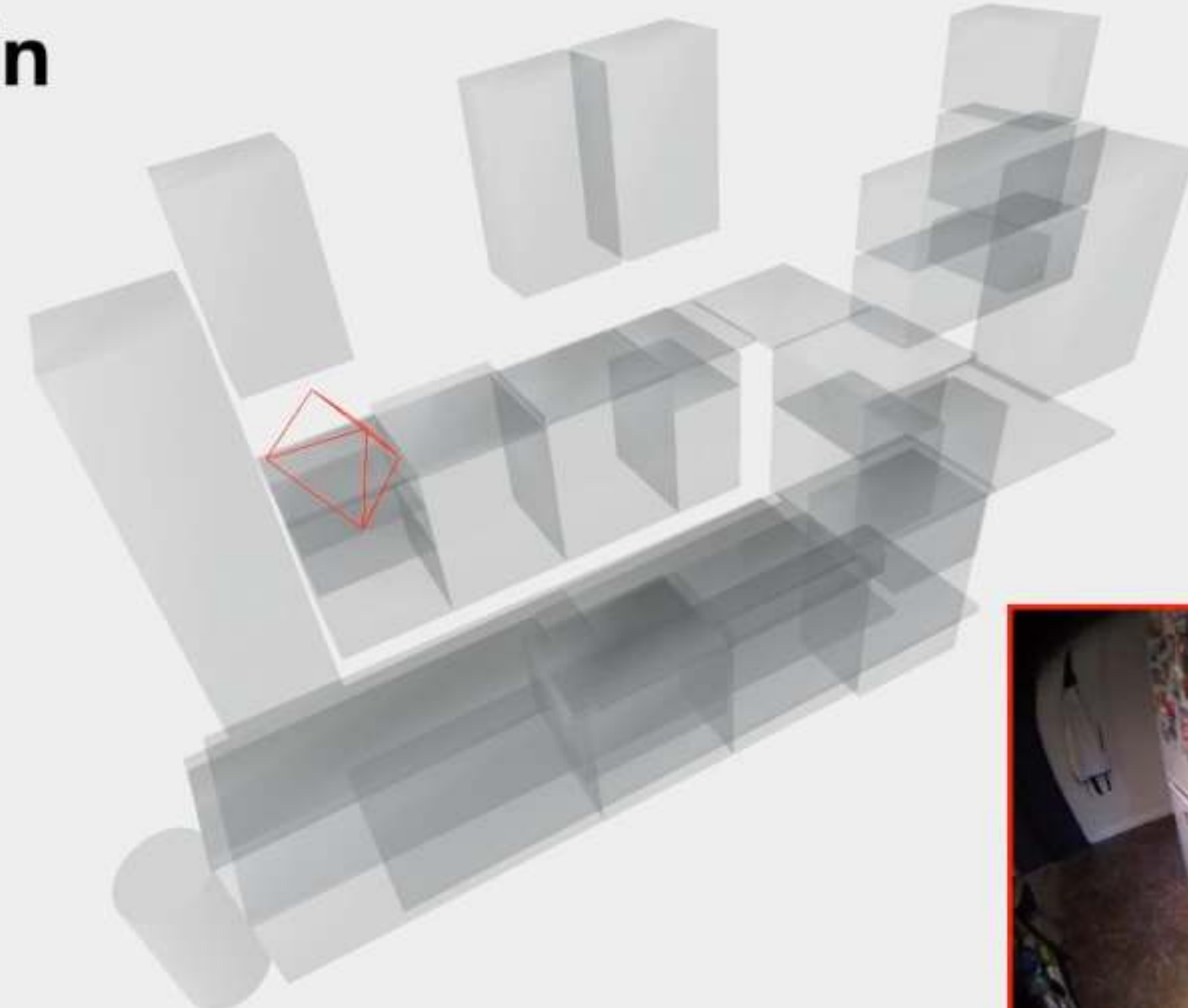


Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

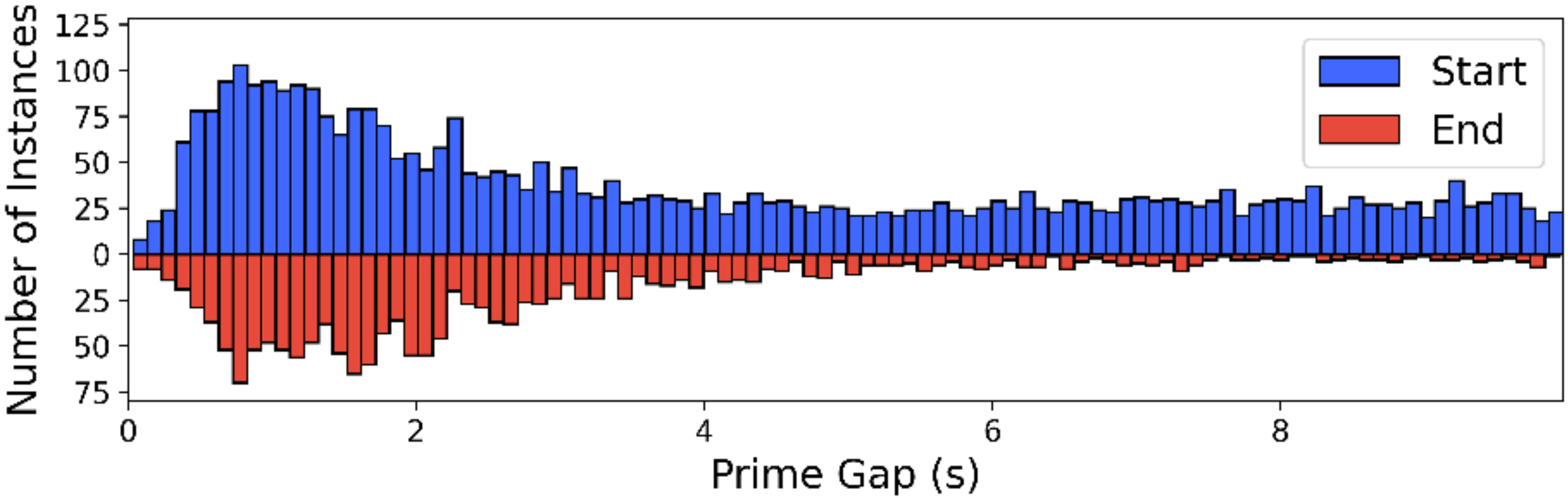
Digital Twin

Fixtures

Open drawer



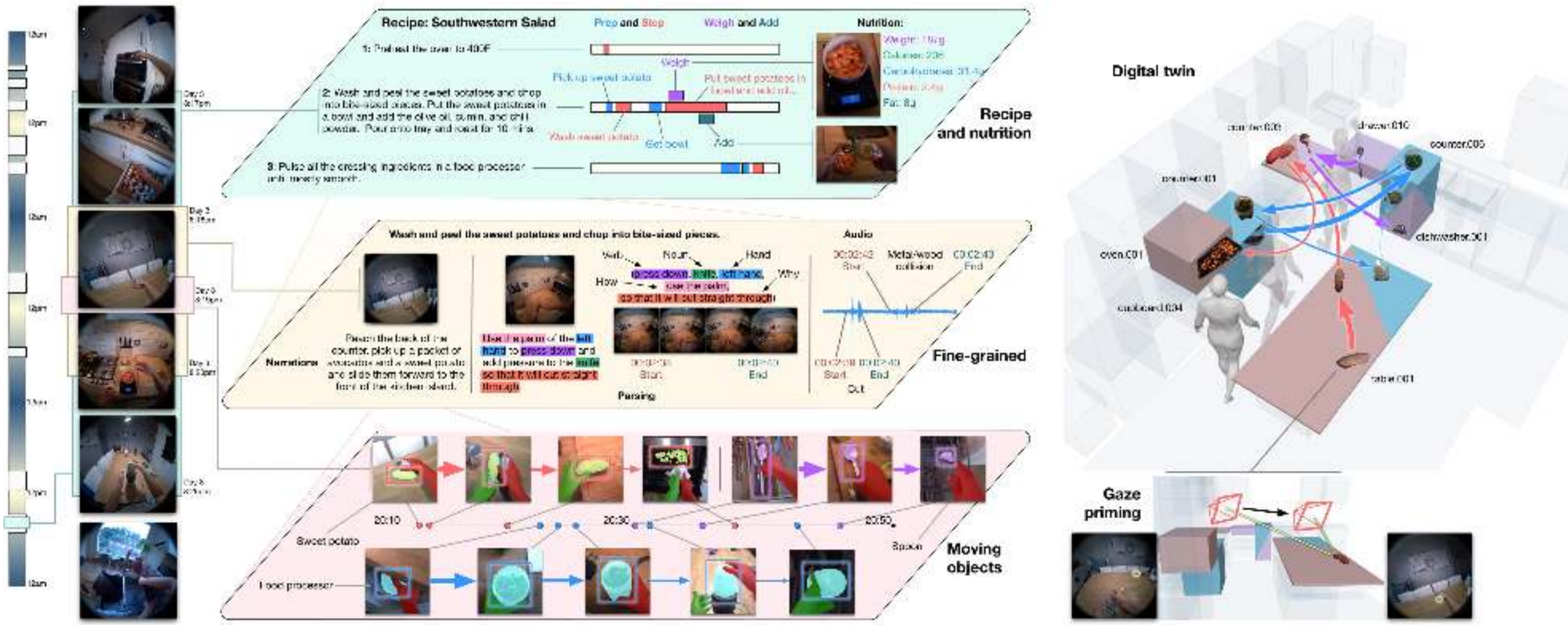
HD-EPIC



Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

HD-EPIC

HD-EPIC



Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

University of Bristol

Dima Damen
BinEgo-360 Workshop @ICCV2025

HD-EPIC

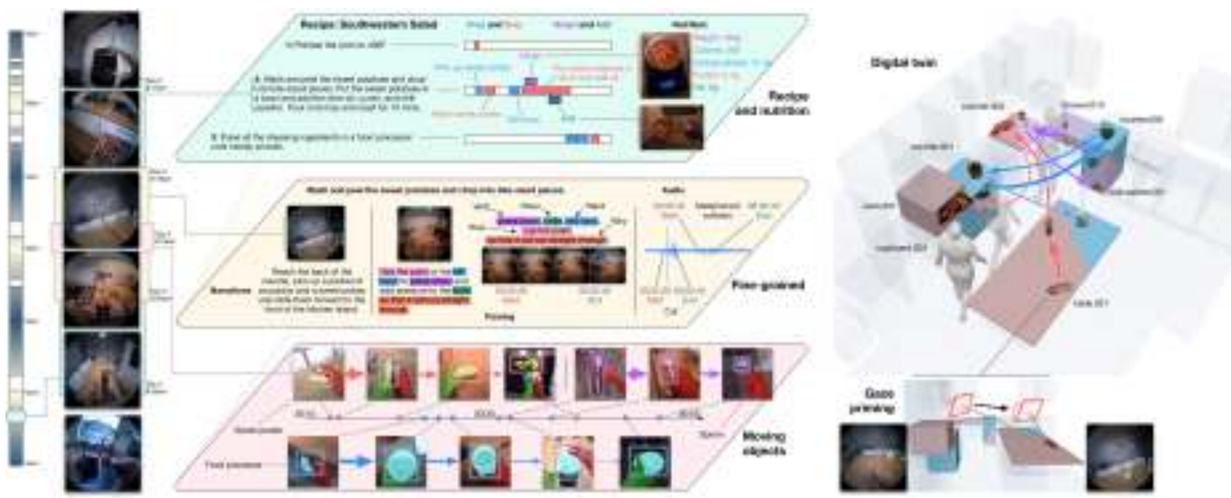
Annotation Type	Total annotations	Annotations/min
Narrations	59,454	24.0
Parsing (Verbs + Nouns + Hands + How + Why)	303,968	122.7
Recipes (Preps + Steps)	4,052	1.6
Sound	50,968	20.6
Action boundaries	59,454	24.0
Object Motion (Pick up + Put down + Fixtures + Bboxes + Masks)	153,480	62.0
Object Itinerary	4,881	2.0
Object Priming (Starts + Ends)	18,264	7.4
Total		263.2

Table A3. HD-EPIC annotations per minute

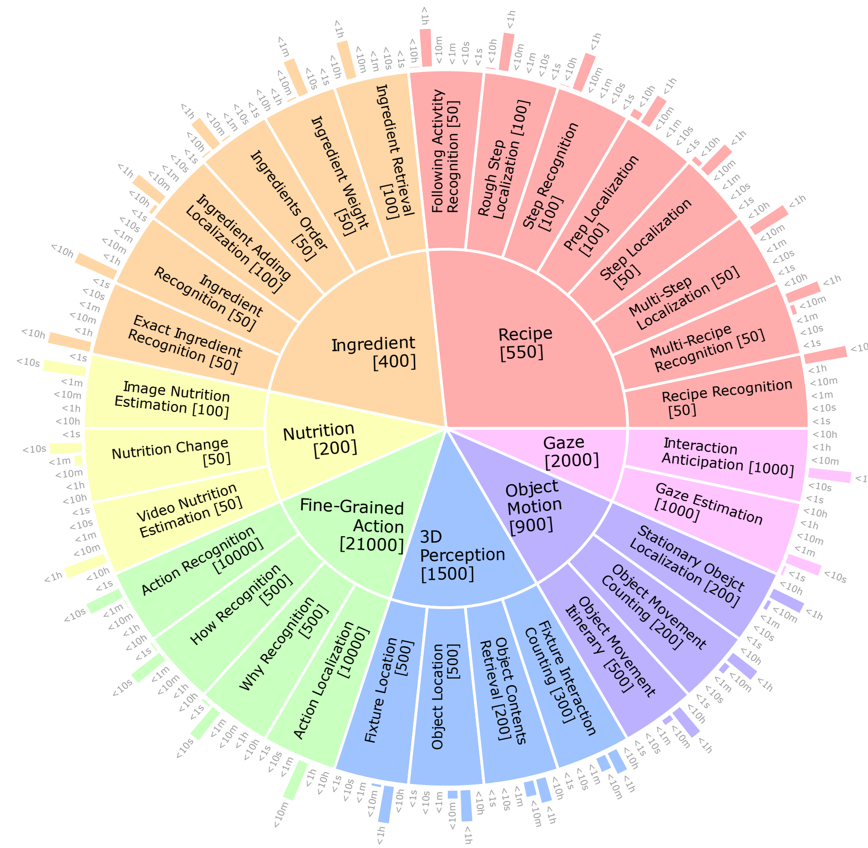
University of BRISTOL

Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

HD-EPIC



Sec 1: Highly-Detailed Dataset



Sec 2: HD-EPIC VQA Benchmark

Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

Dima Damen
BinEgo-360 Workshop @ICCV2025

HD-EPIC

Try it Yourself

Use Wise to Search through HD-EPIC

SCAN ME

https://meru.robots.ox.ac.uk/HD-EPIC/

University of BRISTOL

Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

In today's talk…

Motivation and Datasets in Egocentric Video Understanding

Teaser: The Wizard of Oz & Genie 3

Video Understanding Out of the Frame

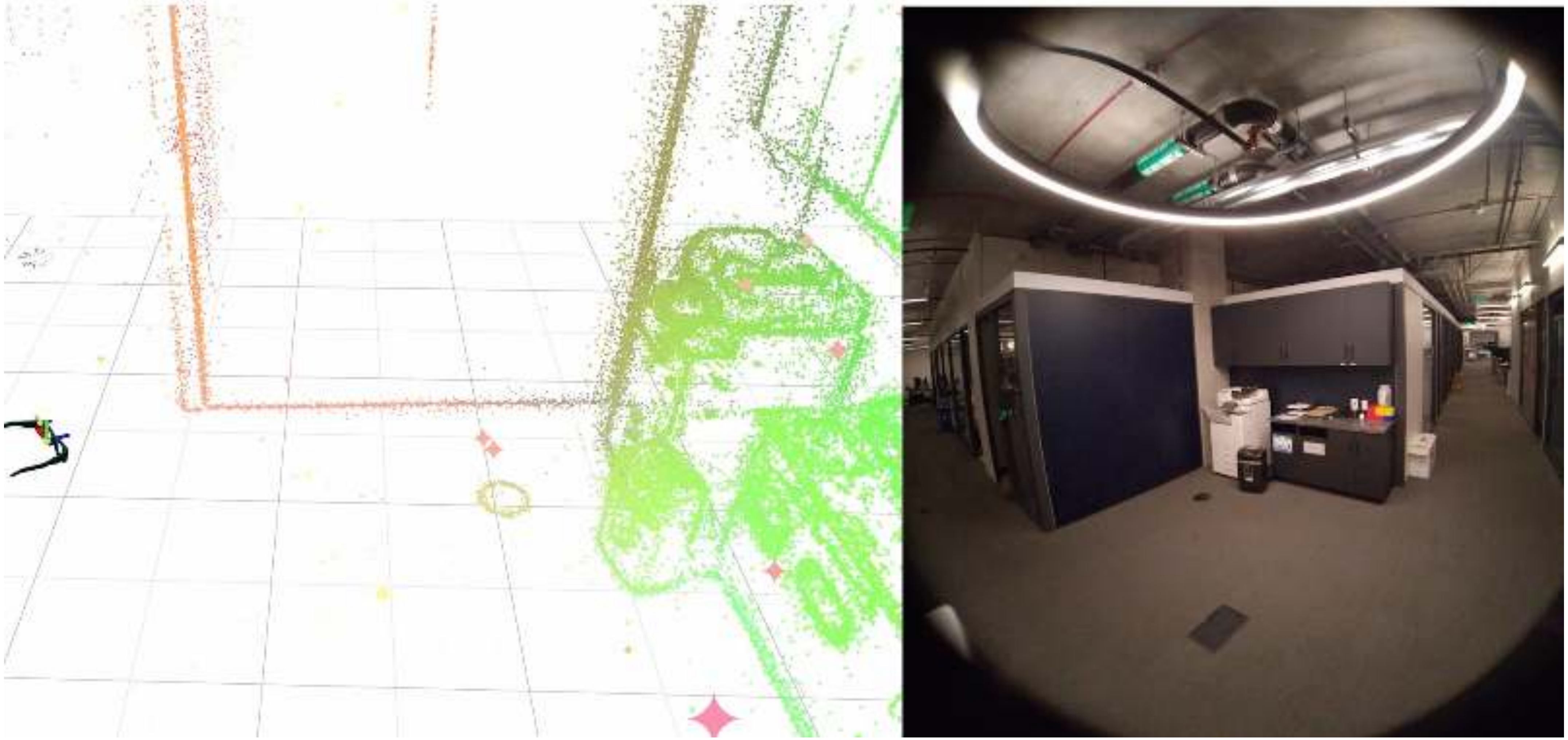
- Point Tracking
- Object Tracking
- Gaze Priming
- Hand Tracking

Outlook into the Future of Egocentric Vision

Conclusion

EgoBody

EgoAllo uses egocentric (👓) SLAM poses and images



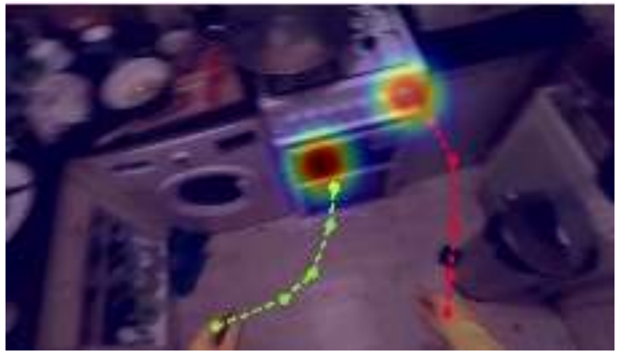
Yi et al (2025). Estimating Body and Hand Motion in an Ego-sensed World. CVPR

with: Masashi Hatano
Zhifan Zhu
Hideo Saito

2D Hand Forecasting

Given an egocentric video,
forecast 2D hand positions of both hands

→ Limited in 2D image plane



OCT [CVPR’22]

Diff-IP2D [IROS’25]

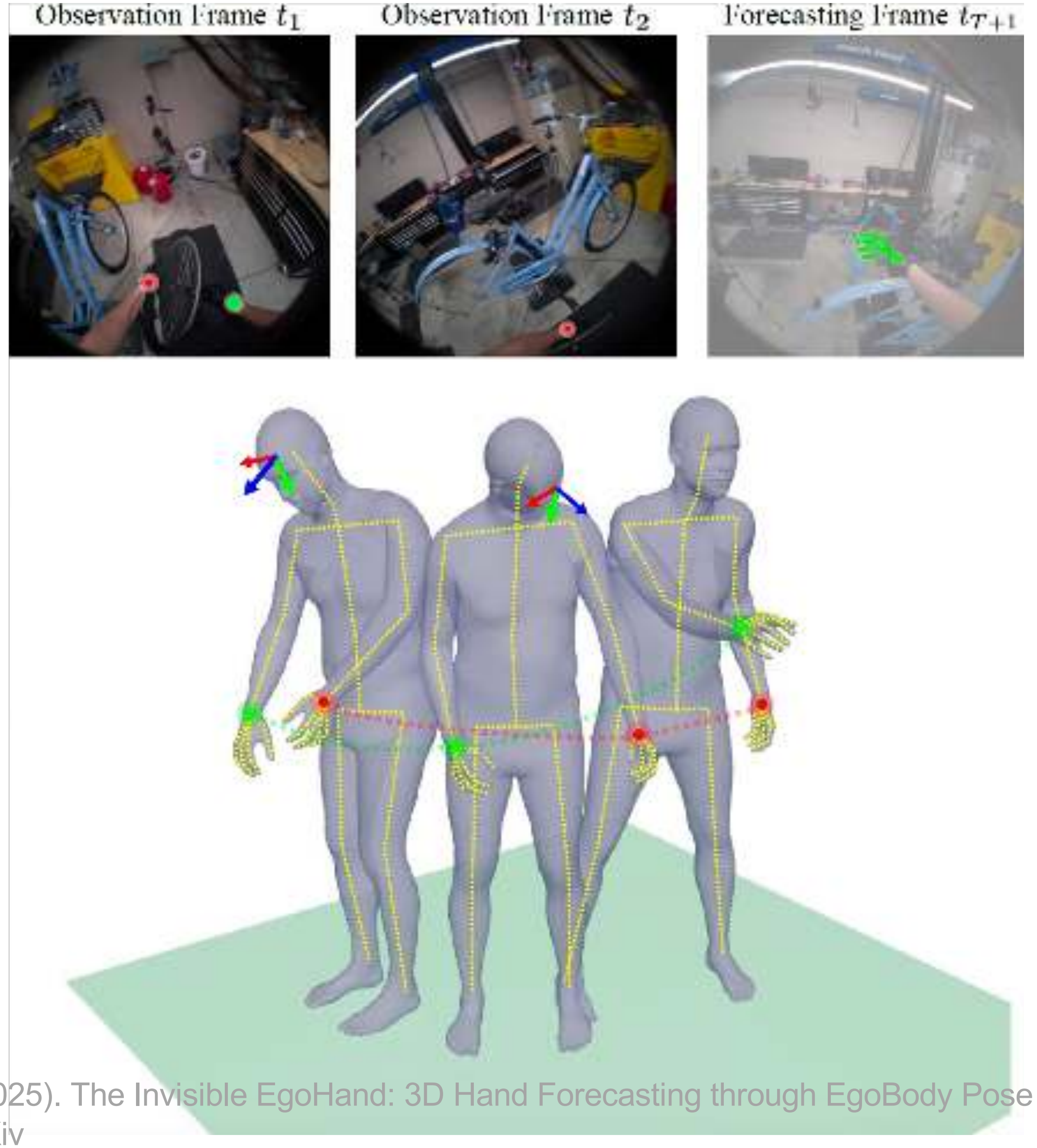
3D Hand Forecasting

Given an egocentric video & 3D hand trajectory,
forecast 3D hand positions of one hand

USST [ICCV’23]

The Invisible EgoHand

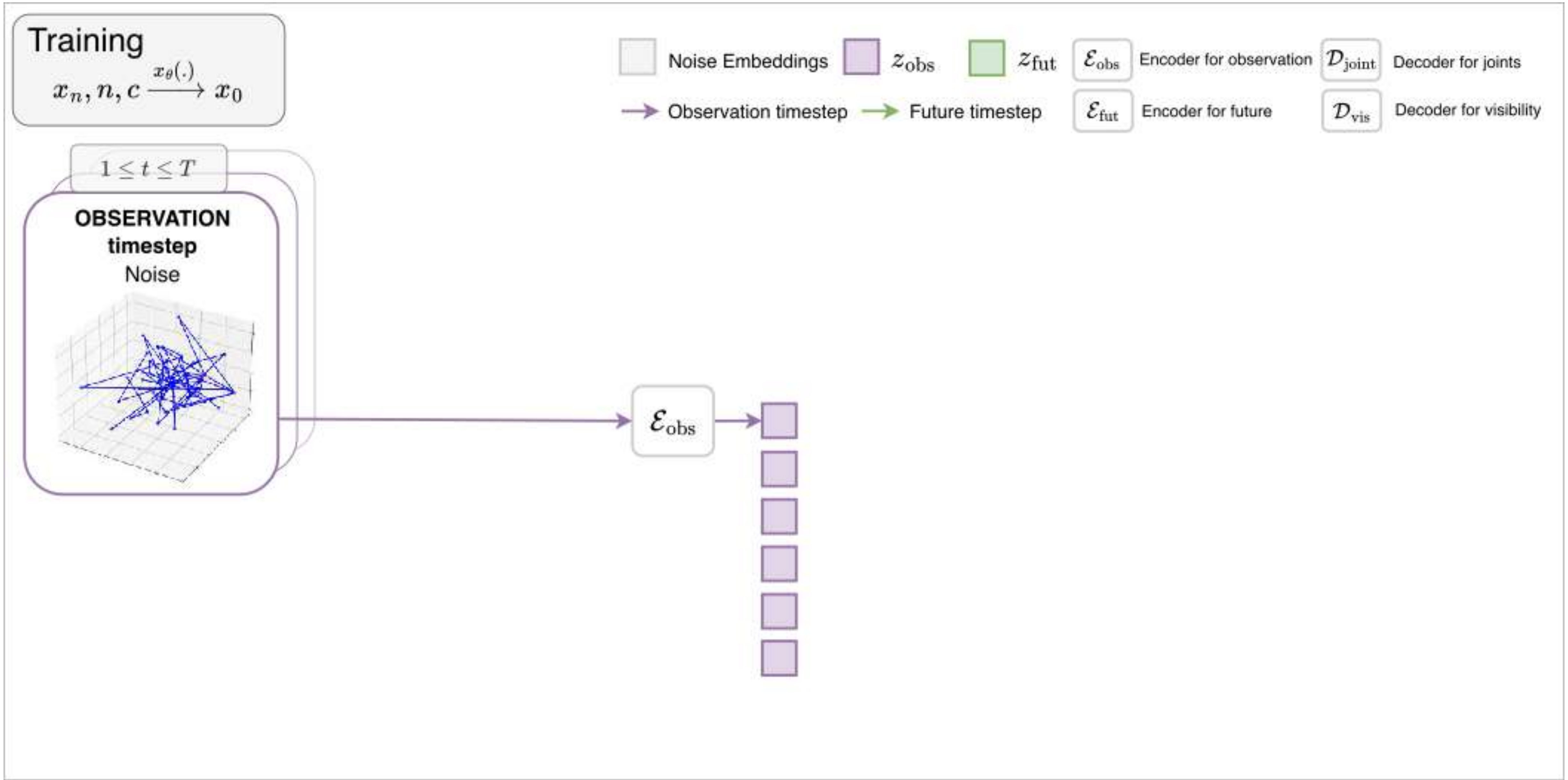
with: Masashi Hatano
Zhifan Zhu
Hideo Saito



Hatano et al (2025). The Invisible EgoHand: 3D Hand Forecasting through EgoBody Pose Estimation. ArXiv

The Invisible EgoHand

with: Masashi Hatano
Zhifan Zhu
Hideo Saito



Hatano et al (2025). The Invisible EgoHand: 3D Hand Forecasting through EgoBody Pose Estimation. ArXiv

The Invisible EgoHand

with: Masashi Hatano
Zhifan Zhu
Hideo Saito

Method	Hand Trajectory Forecasting		Hand Pose Forecasting	
	All		All	
	ADE	FDE	MPJPE	MPJPE-F
Static	0.335	0.405	0.166	0.179
CVM [61]	0.346	0.467	0.166	0.183
EgoEgoForecast	0.295	0.352	0.166	0.177
USST [3]	0.562	0.581	-	-
Ours	**0.261**	**0.324**	**0.115**	**0.143**

Hatano et al (2025). The Invisible EgoHand: 3D Hand Forecasting through EgoBody Pose Estimation. ArXiv

with: Masashi Hatano
Zhifan Zhu
Hideo Saito

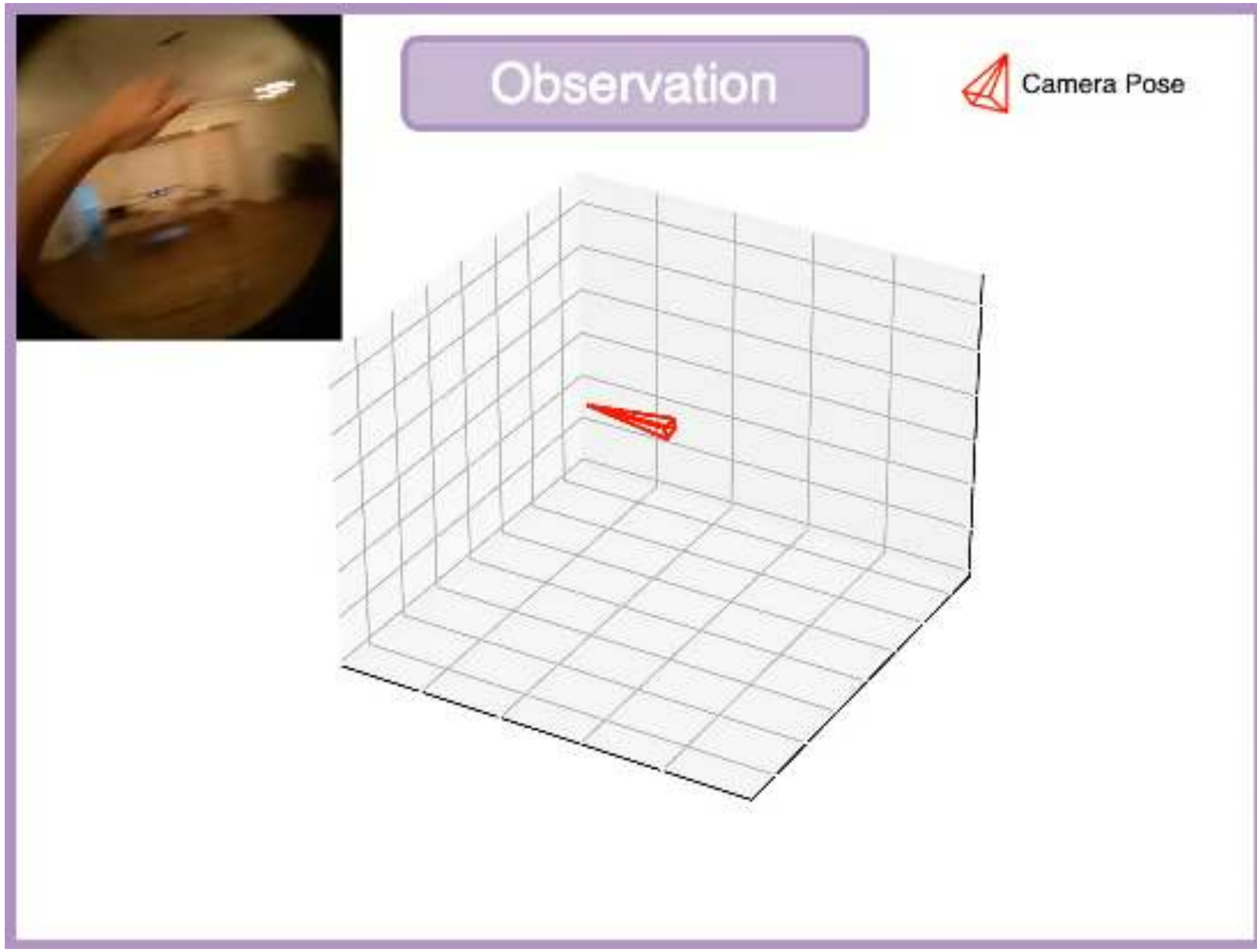
Method	Hand Trajectory Forecasting			Hand Pose Forecasting		
	In-view	Out-of-view	All	In-view	Out-of-view	All
EgoEgoForecast	0.171	0.385	0.295	0.162	0.299	0.166
Ours w/o. 2D joint	0.151	0.377	0.282	0.139	0.269	0.142
Ours w/o. image	**0.116**	0.367	**0.261**	0.117	**0.234**	0.120
Ours w/o. $\mathcal{L}_{\text{reproj}}$	0.132	0.368	0.269	0.125	0.250	0.128
Ours w/o. $\mathcal{L}_{\text{vis}}$	0.127	0.377	0.272	0.121	0.240	0.124
Ours w/o. $\mathcal{L}_{\text{body}}$	0.129	0.385	0.277	0.120	0.258	0.123
Ours w/o. $\mathcal{L}_{\text{obs}}$	0.149	0.390	0.289	0.139	0.250	0.142
Ours	**0.116**	**0.366**	**0.261**	**0.112**	0.240	**0.115**

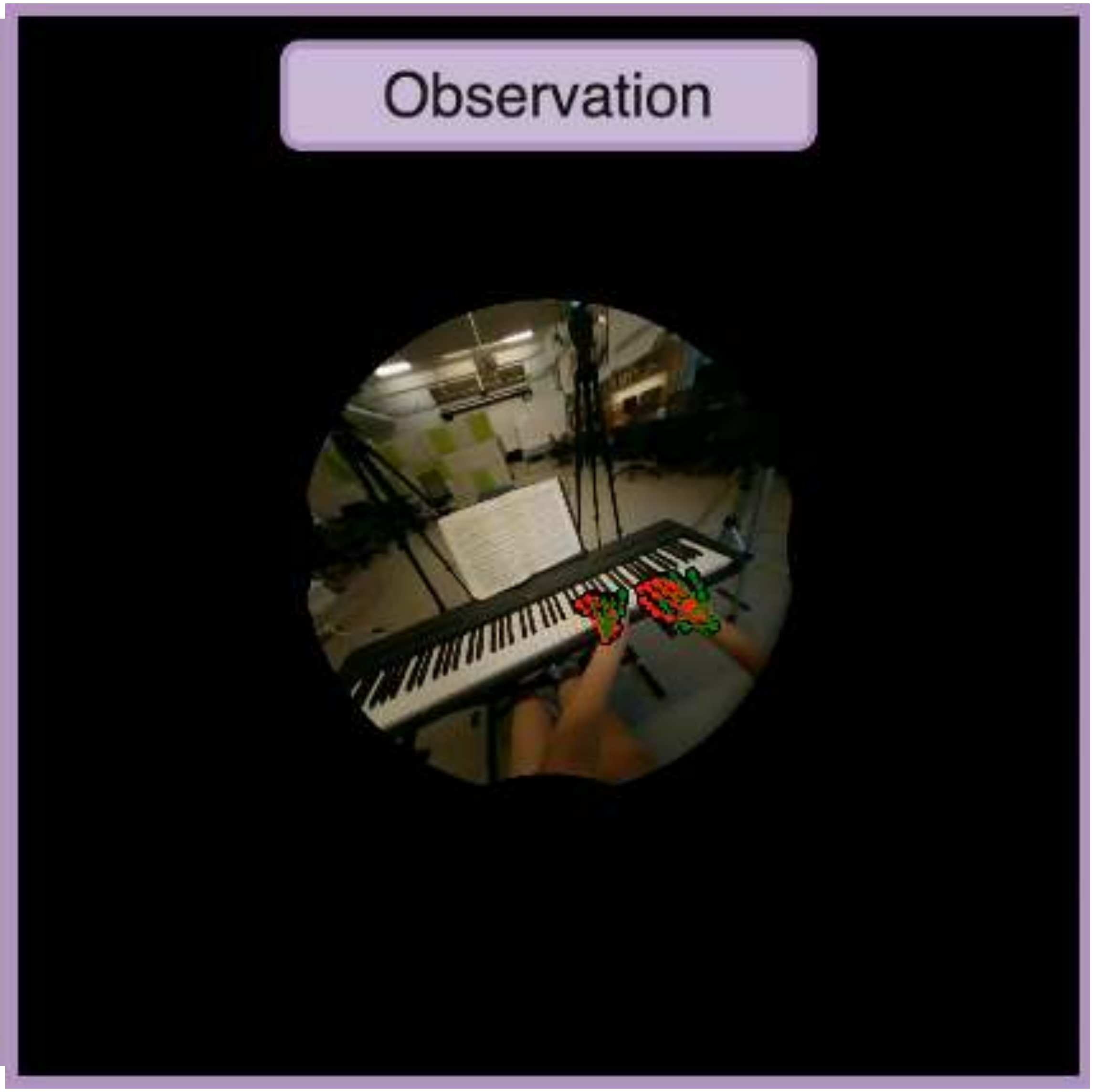
- Without visible 2D joints, significant performance drops can be seen
- 2D reprojection loss serves as effective regularization
- Visibility loss & Body joints loss contribute for out-of-view scenario

Hatano et al (2025). The Invisible EgoHand: 3D Hand Forecasting through EgoBody Pose Estimation. ArXiv

The Invisible EgoHand

with: Masashi Hatano
Zhifan Zhu
Hideo Saito





In today's talk…

Motivation and Datasets in Egocentric Video Understanding

Teaser: The Wizard of Oz & Genie 3

Video Understanding Out of the Frame

- Point Tracking
- Object Tracking
- Gaze Priming
- Hand Tracking

Outlook into the Future of Egocentric Vision

Conclusion

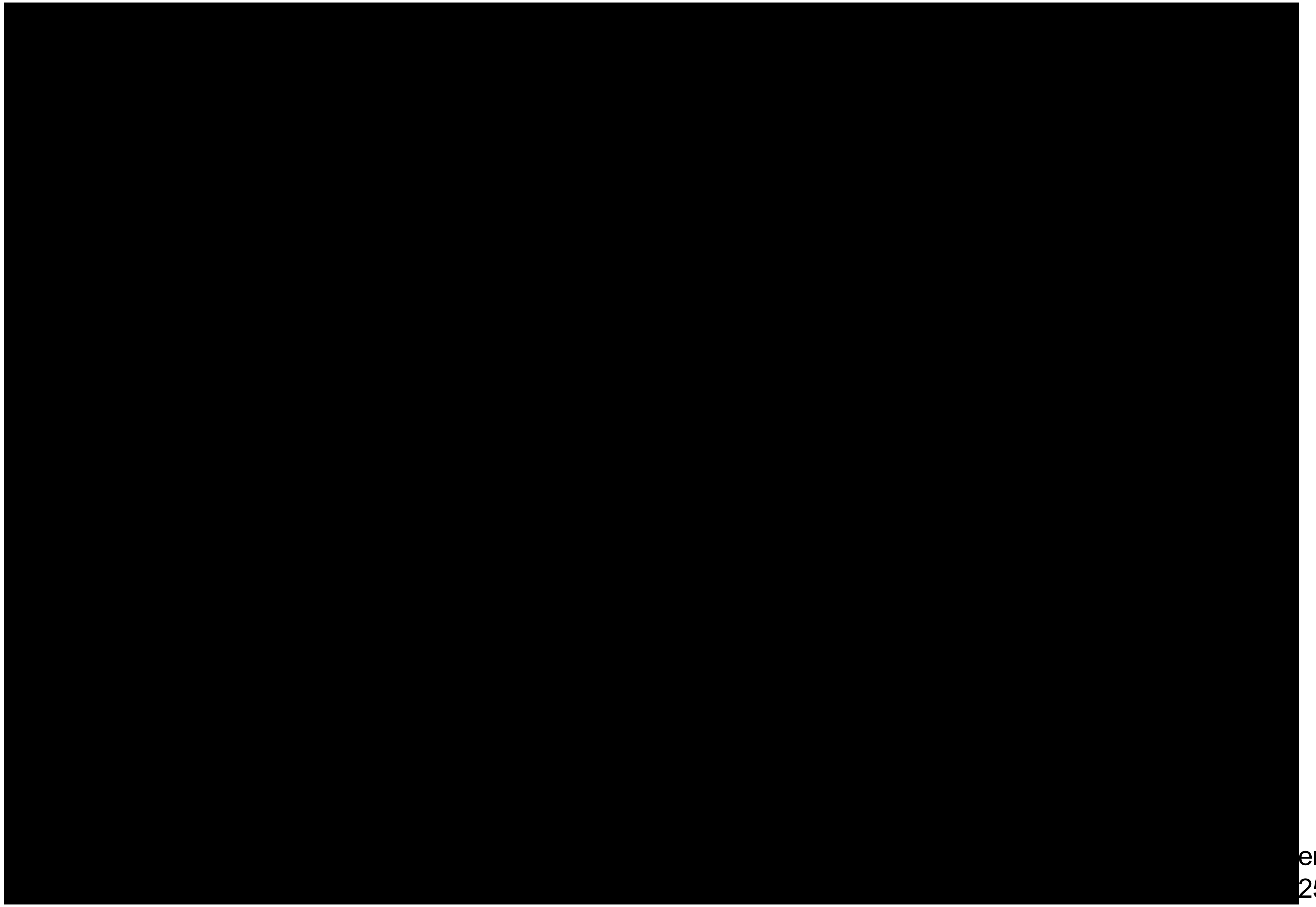
The Wizard of Oz at the Sphere

Sold out tickets – August 2025

https://behindthecurtain.withgoogle.com

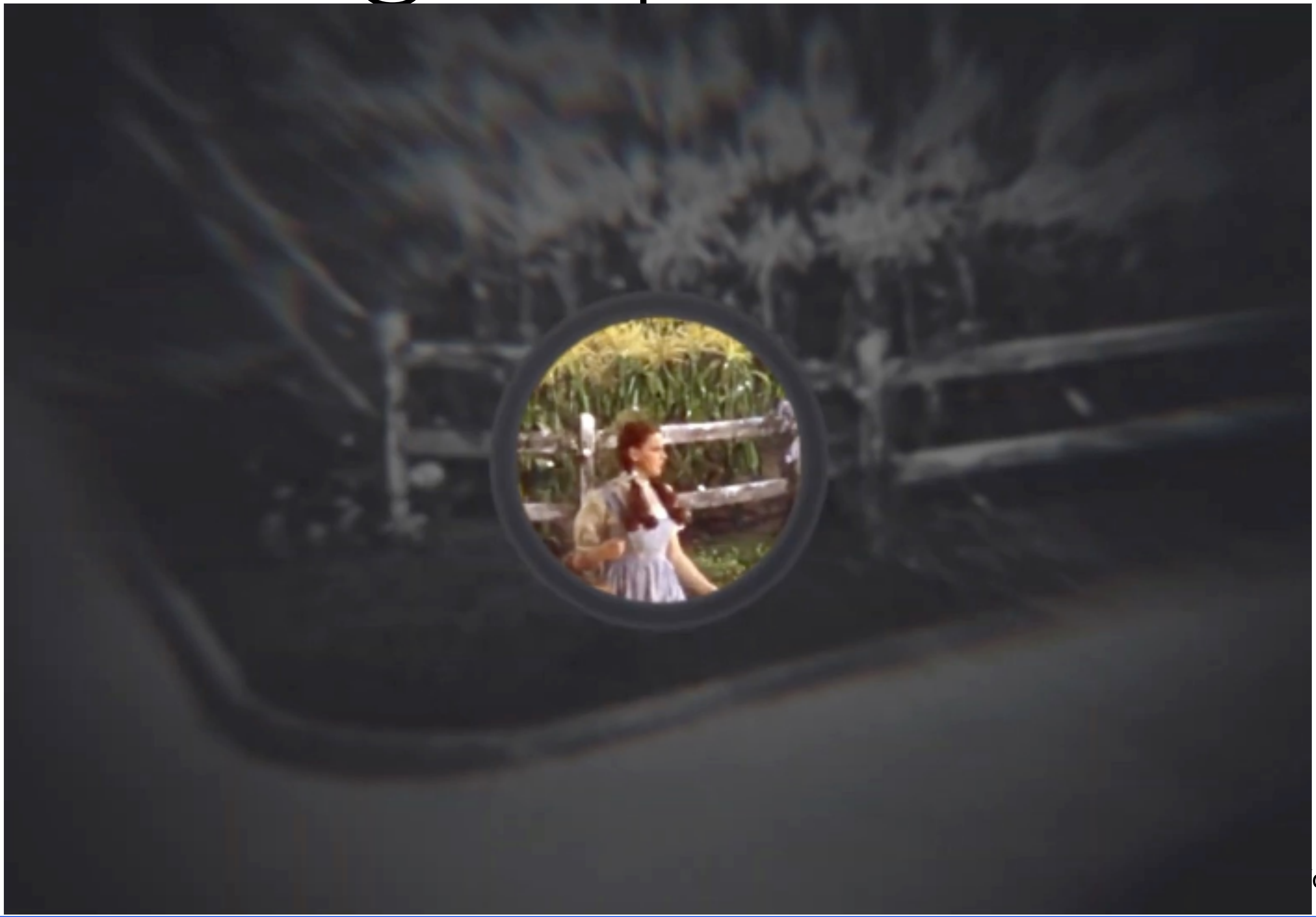
The Wizard of Oz @ The Sphere

- The Movie (1939)
- Technocolour pioneer
- Iconic characters



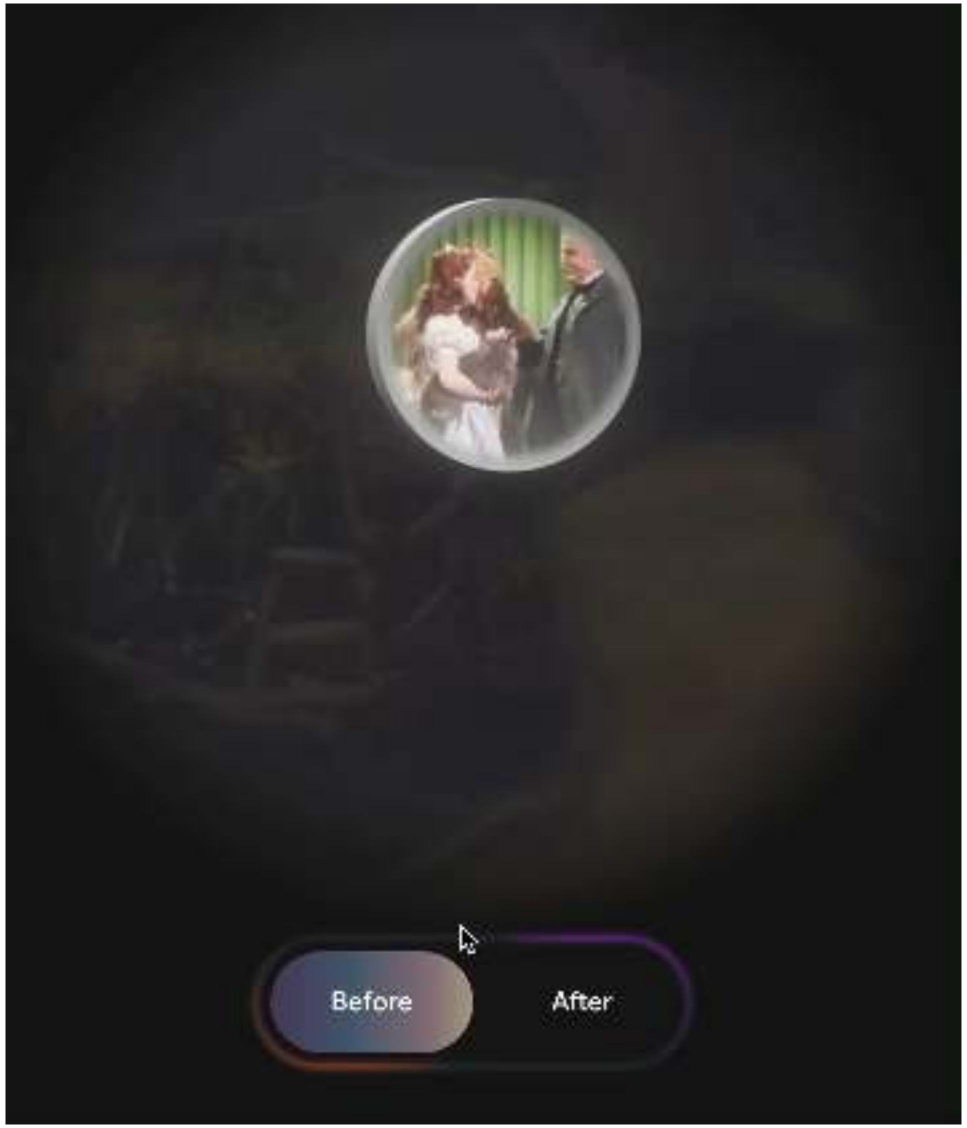
The Wizard of Oz @ The Sphere

The Wizard of Oz @ The Sphere



The Wizard of Oz @ The Sphere

- Super-resolution,
- Outpainting…



https://behindthecurtain.withgoogle.com

The Wizard of Oz @ The Sphere

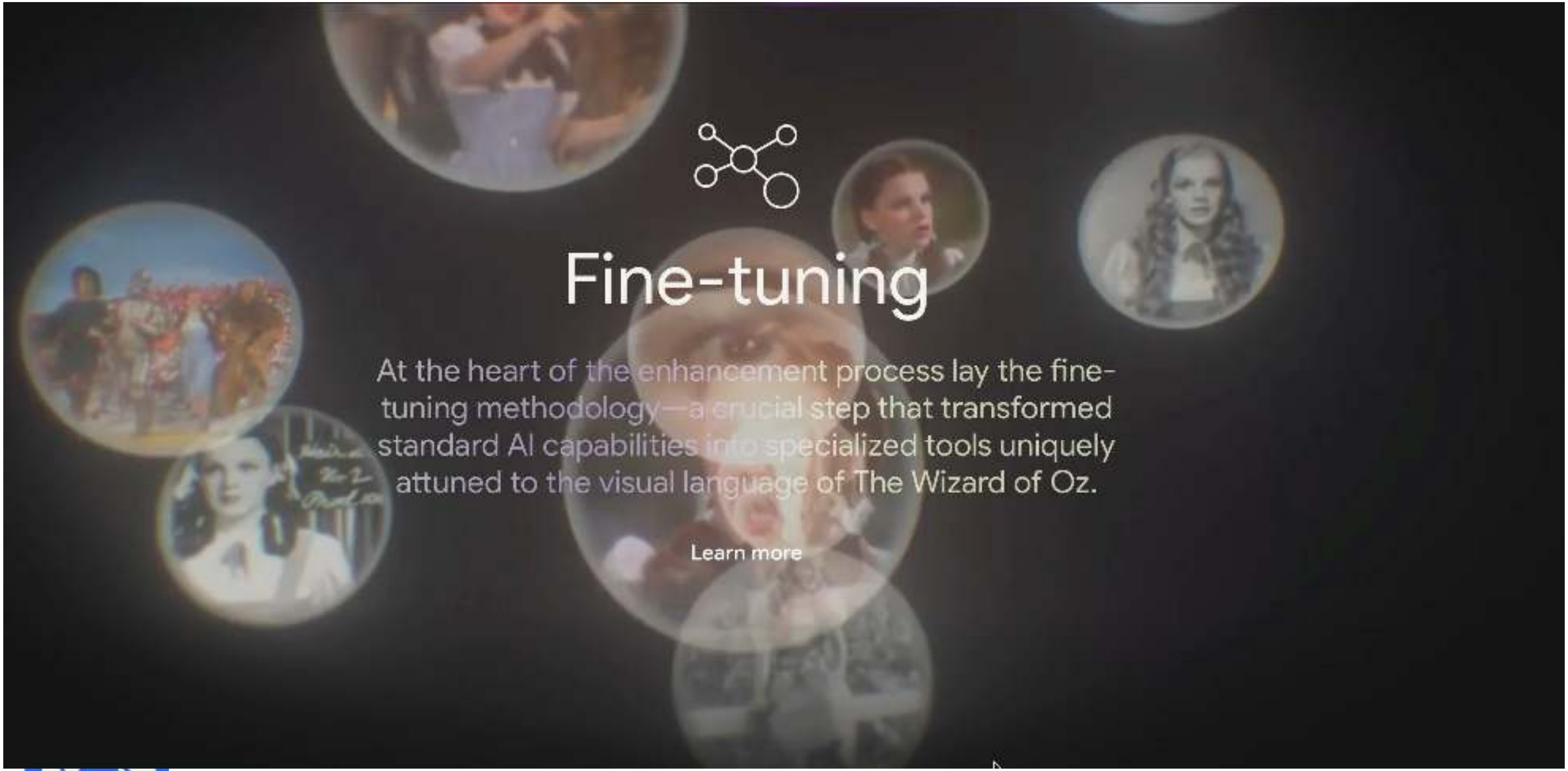
- Performance Interpolation,

The Wizard of Oz @ The Sphere

- Auto-Director

https://behindthecurtain.withgoogle.com

[redacted]

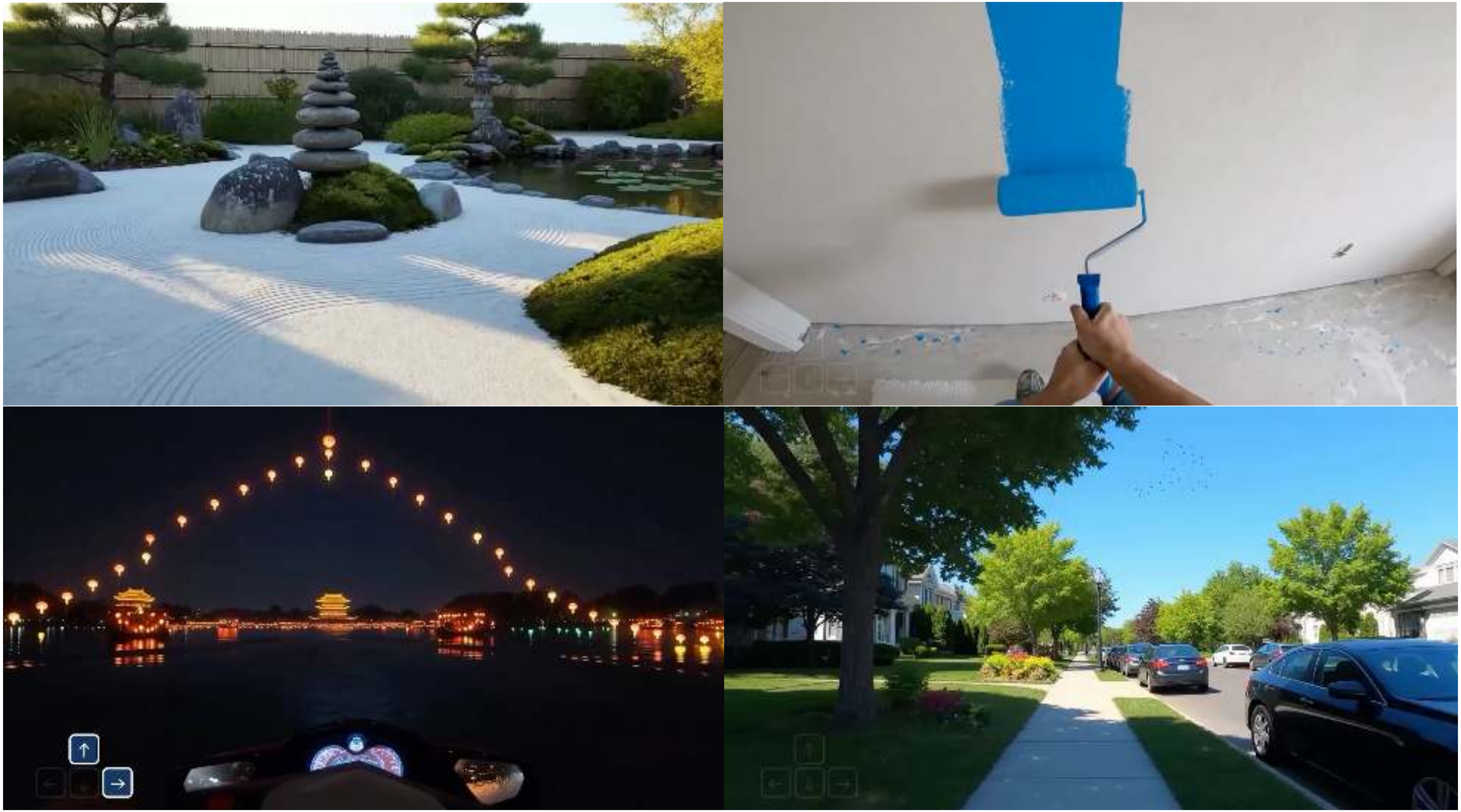


https://behindthecurtain.withgoogle.com

Genie 3

Released Aug 2025

https://deepmind.google/discover/blog/genie-3-a-new-frontier-for-world-models/



Genie 3

- A new frontier of world models
- From a text prompt or a starting image/video
- Generate interactive worlds at 24fps and 720px
- Remains consistent for several minutes
- Genie 3’s consistency is an emergent capability

In today's talk…

Motivation and Datasets in Egocentric Video Understanding

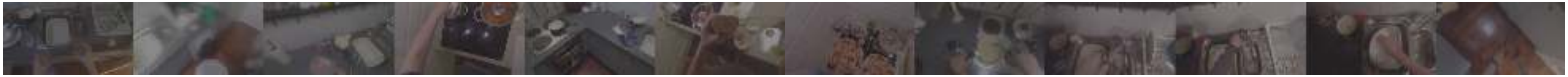
Video Understanding Out of the Frame

- Point Tracking
- Object Tracking
- Gaze Priming
- Hand Tracking

Teaser: The Wizard of Oz & Genie 3

Outlook into the Future of Egocentric Vision

Conclusion



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[†] 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

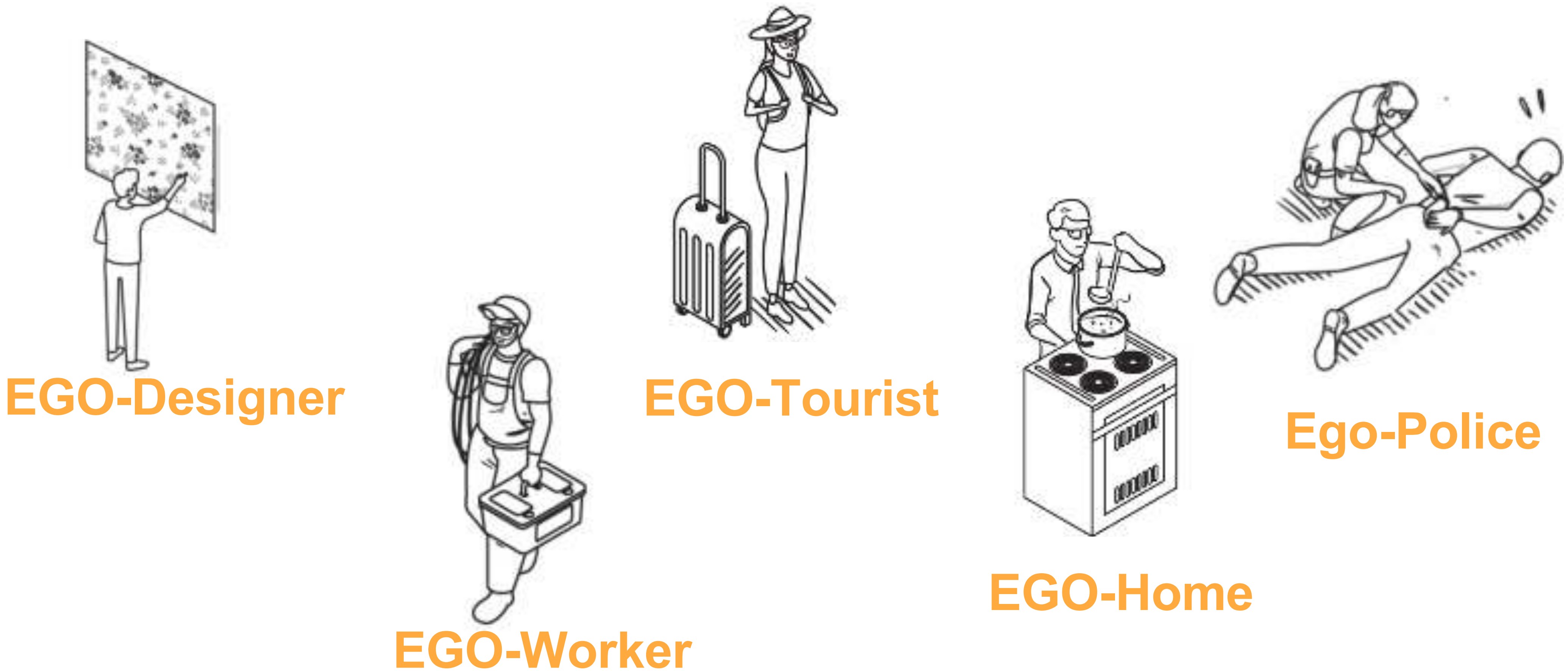
How did we do this?

Dima Damen
BinEgo-360 Workshop @ICCV2025

C Plizzari et al (2025). An Outlook into the Future of Egocentric Vision. IJCV

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

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

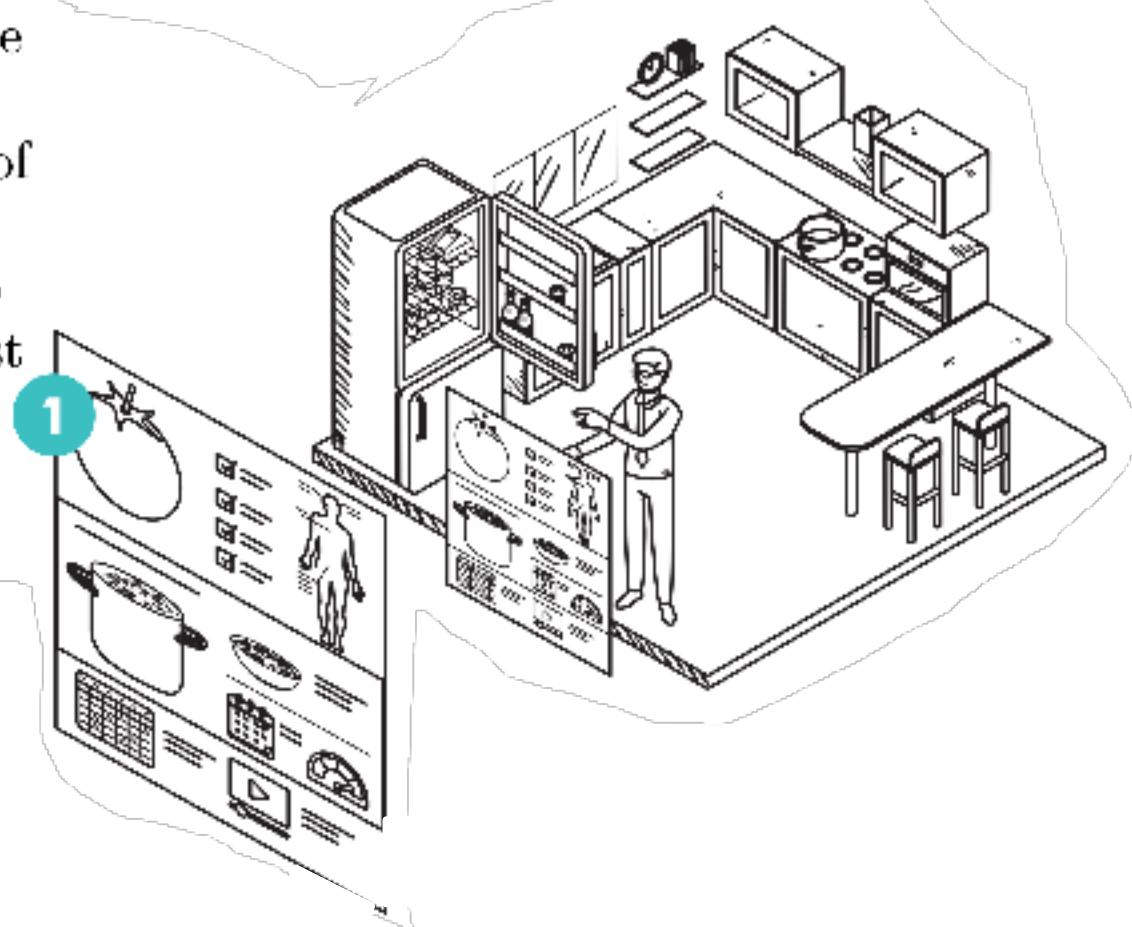


C Plizzari et al (2025). An Outlook into the Future of Egocentric Vision. IJCV

EGO-Home

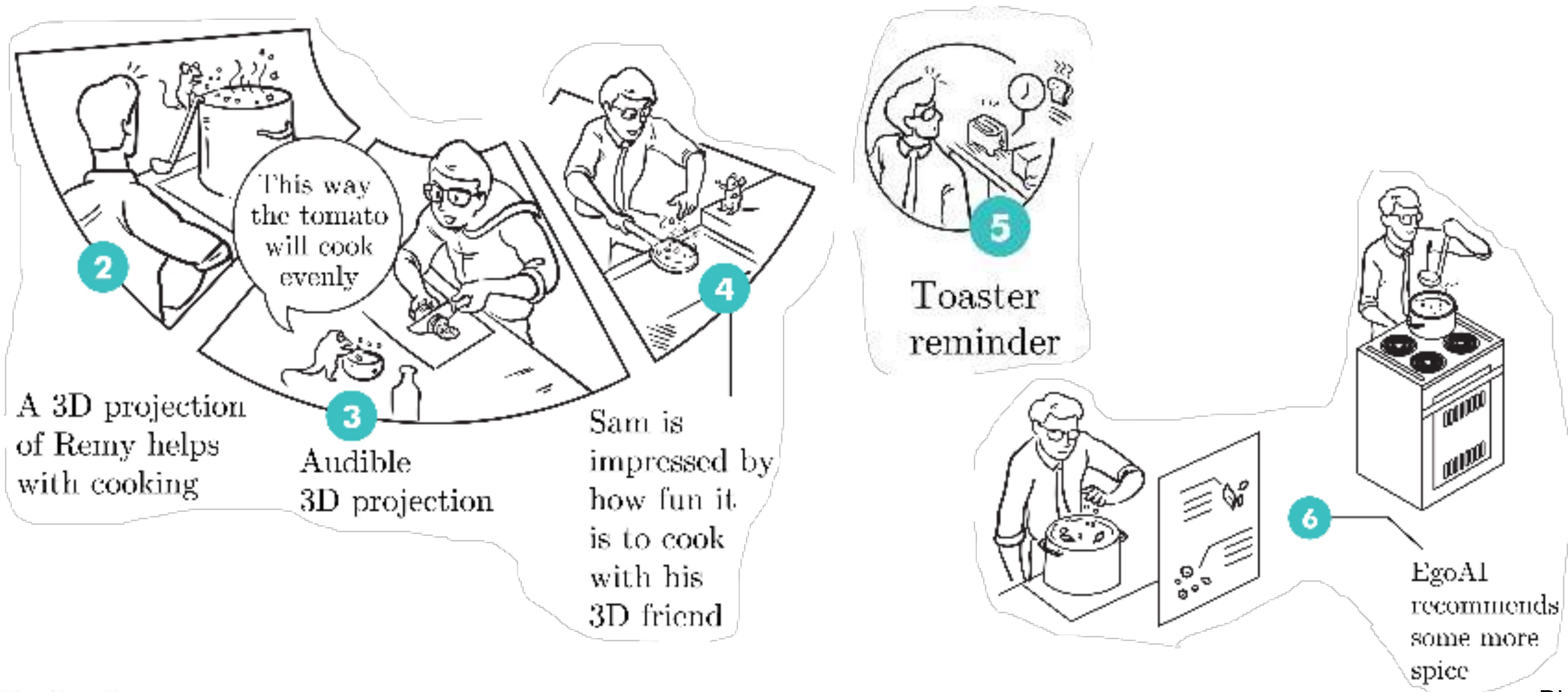
with: Chiara Plizzari, Gabriele Goletto, Antonio Furnari, Siddhant Bansal, Francesco Ragusa, Giovanni Maria Farinella, Tatiana Tommasi

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



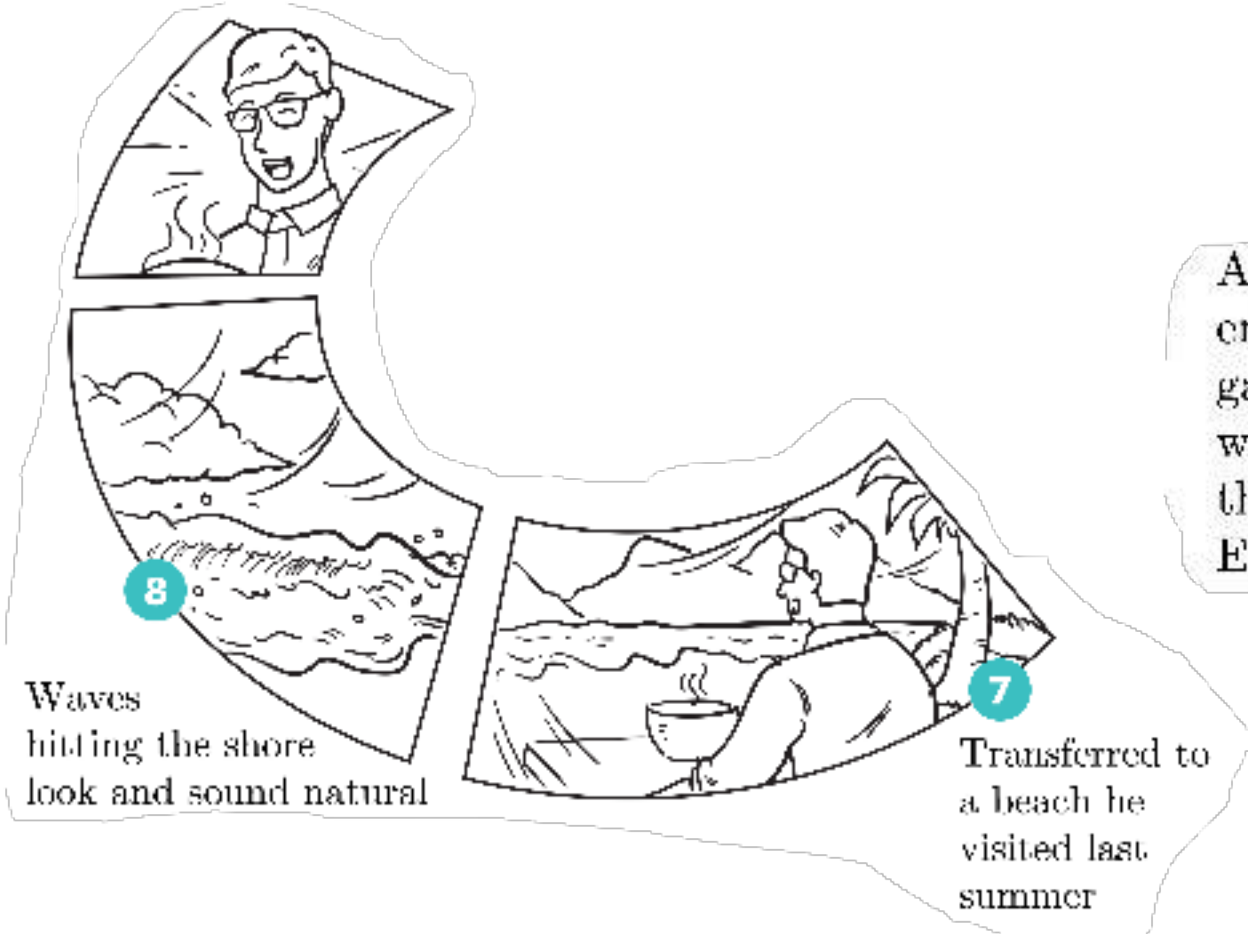
EGO-Home

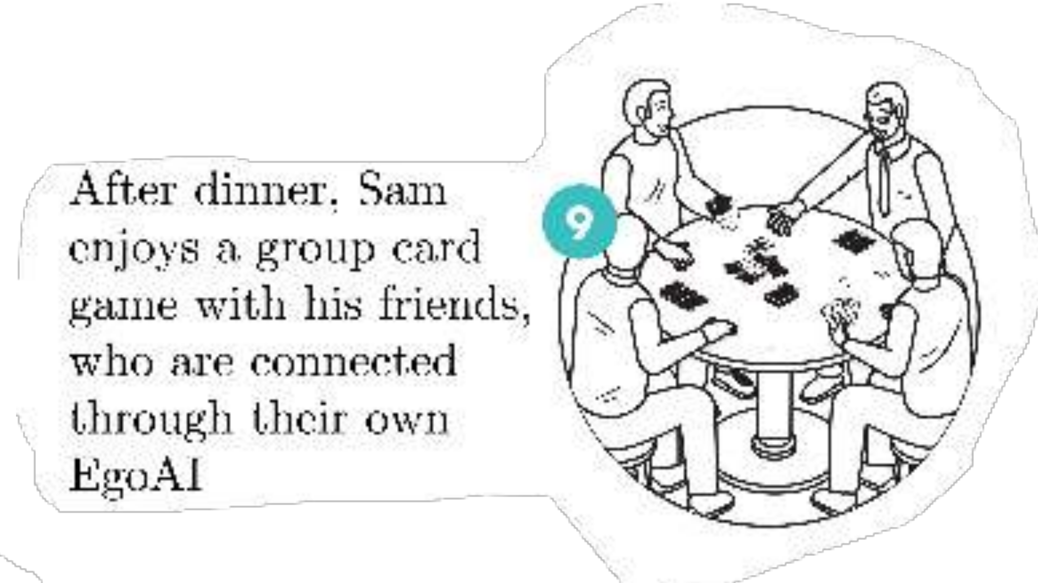
with: Chiara Plizzari, Gabriele Goletto, Antonio Furnari, Siddhant Bansal, Francesco Ragusa, Giovanni Maria Farinella, Tatiana Tommasi



EGO-Home

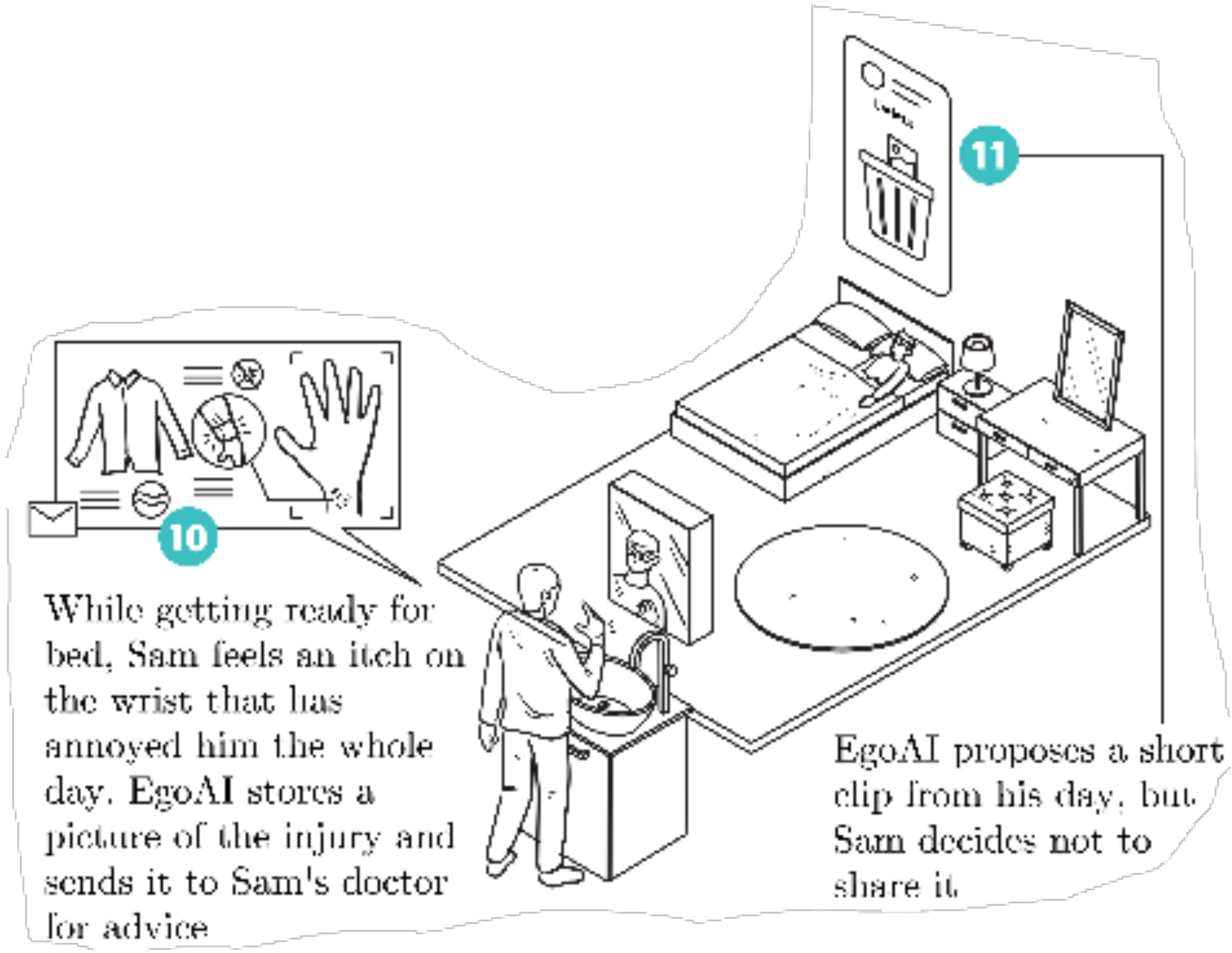
with: Chiara Plizzari, Gabriele Goletto, Antonio Furnari, Siddhant Bansal, Francesco Ragusa, Giovanni Maria Farinella, Tatiana Tommasi





EGO-Home

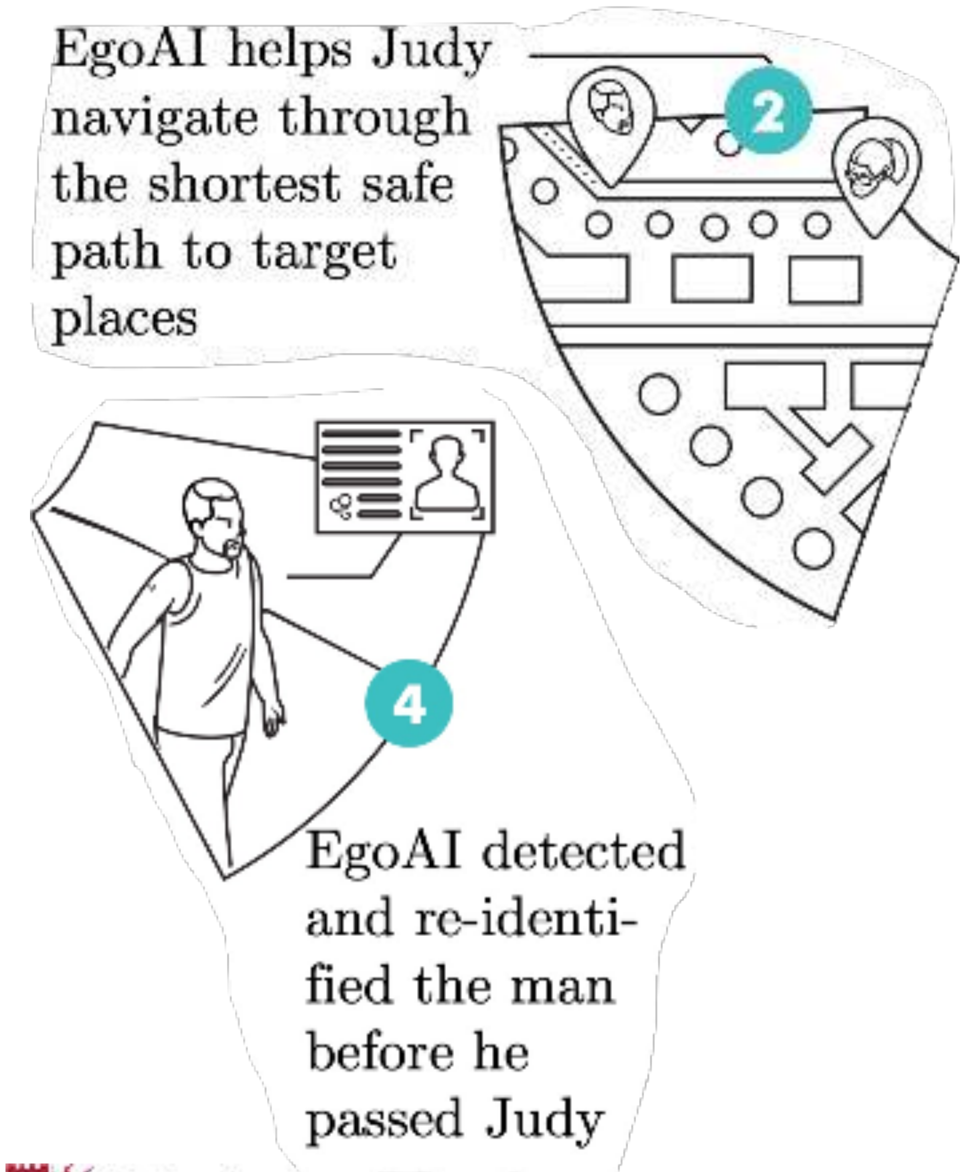
with: Chiara Plizzari, Gabriele Goletto, Antonio Furnari, Siddhant Bansal, Francesco Ragusa, Giovanni Maria Farinella, Tatiana Tommasi

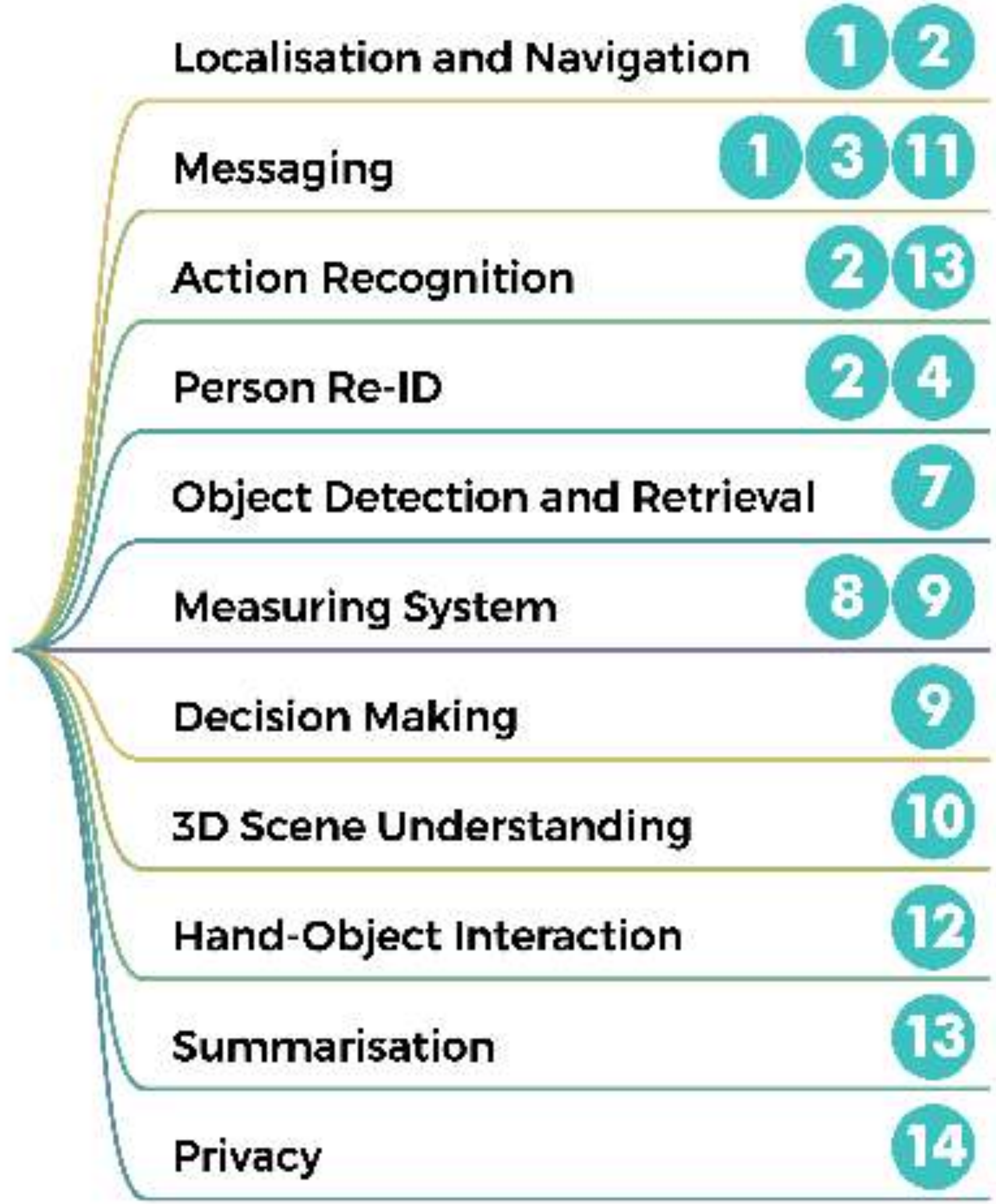


C Plizzari et al (2025). An Outlook into the Future of Egocentric Vision. IJCV

From Stories to Tasks

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



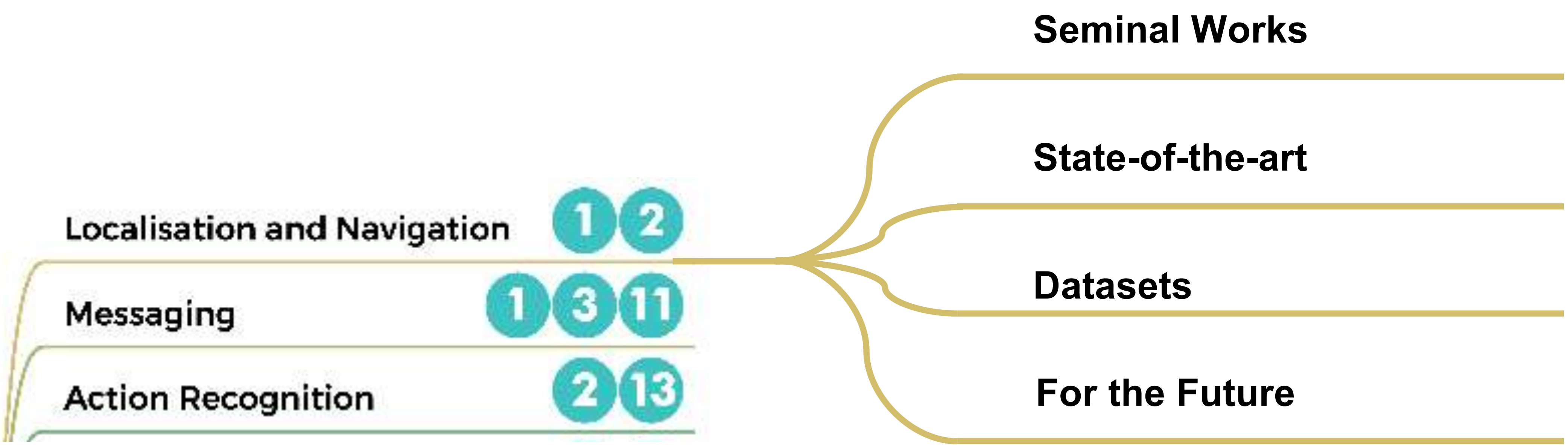


University of Bristol

C Plizzari et al (2025). An Outlook into the Future of Egocentric Vision. IJCV

The Survey Part

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



In today's talk…

Motivation and Datasets in Egocentric Video Understanding

Video Understanding Out of the Frame

- Point Tracking
- Object Tracking
- Gaze Priming
- Hand Tracking

Teaser: The Wizard of Oz & Genie 3

Outlook into the Future of Egocentric Vision

Conclusion

My research team…

grateful

Thank you

For further info, datasets, code, publications…

http://dimadamen.github.io

@dimadamen

@dimadamen.bsky.social

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

Q&A