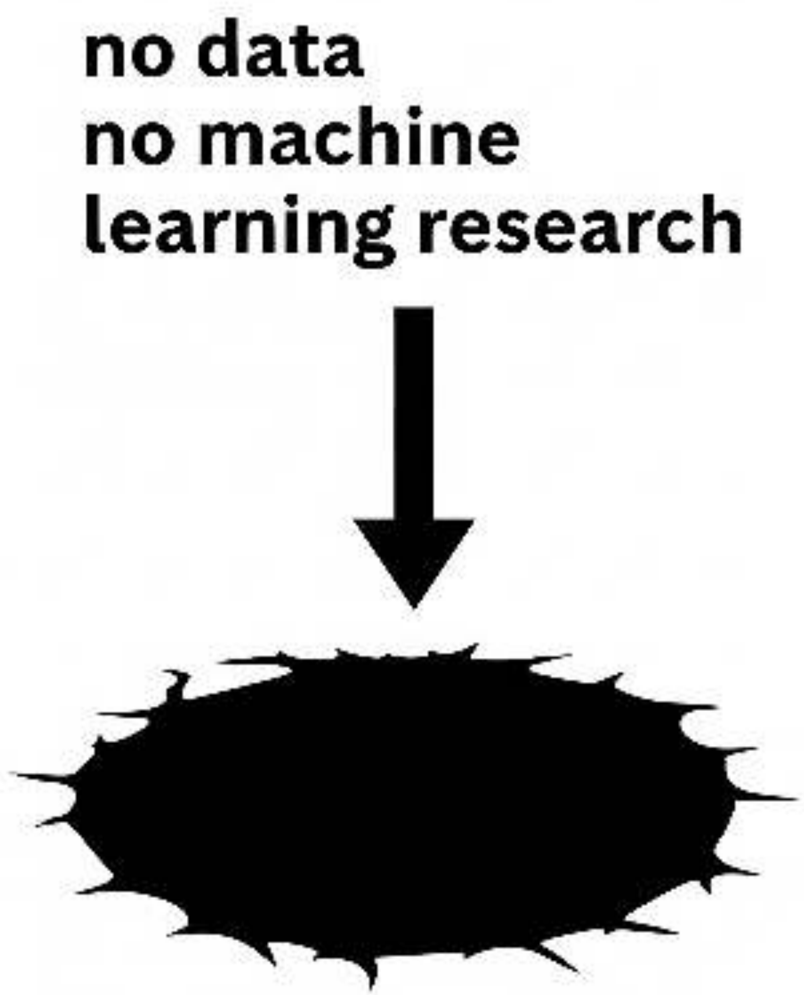
Collecting a Dataset for Computer Vision in 2025…

an Egocentric Perspective

University of BRISTOL

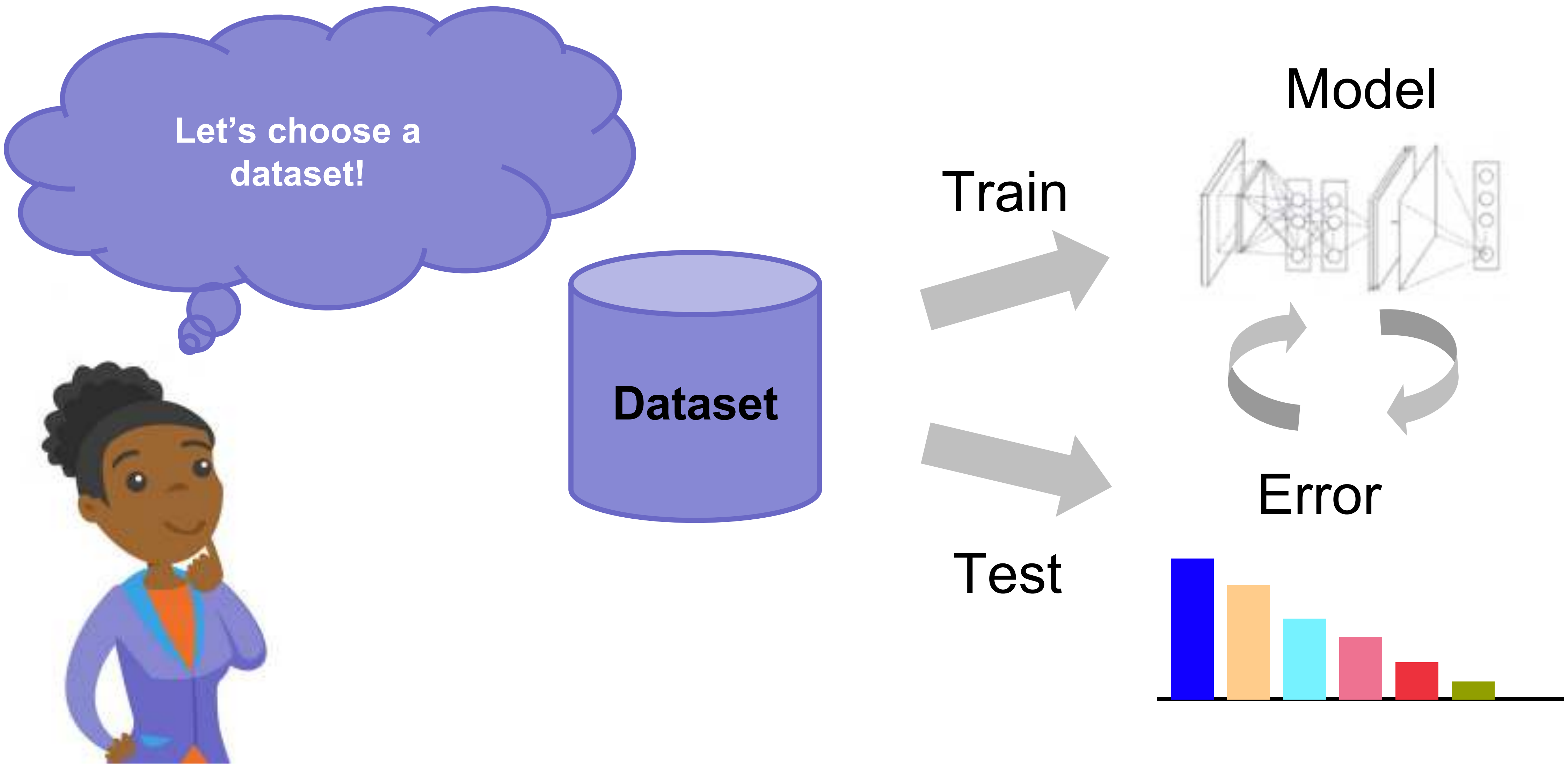
Dima Damen
DataCV W – ICCV2025

Machine Learning in Practice

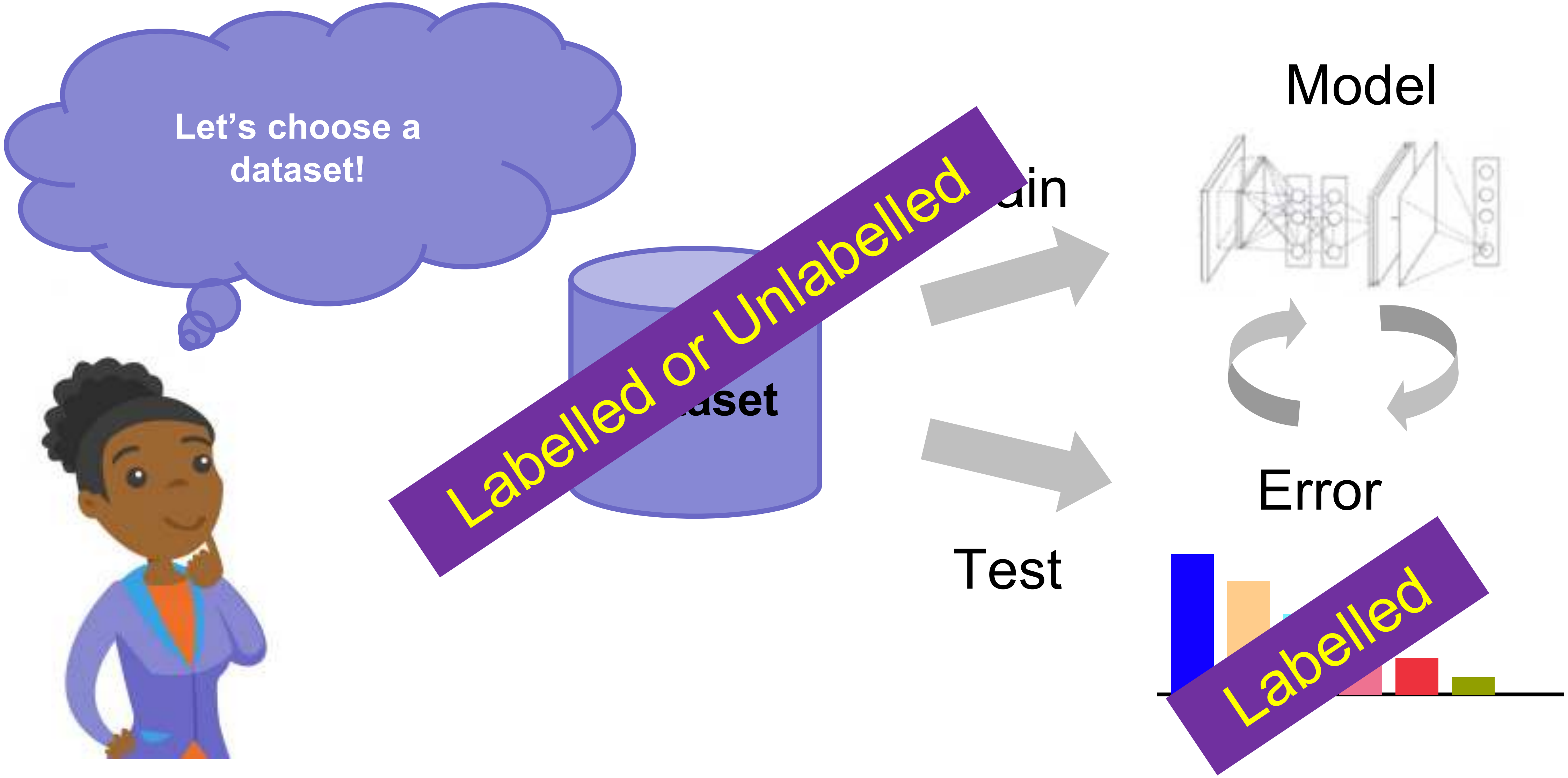


The current paradigm of Computer Vision Research

The current paradigm of Computer Vision Research



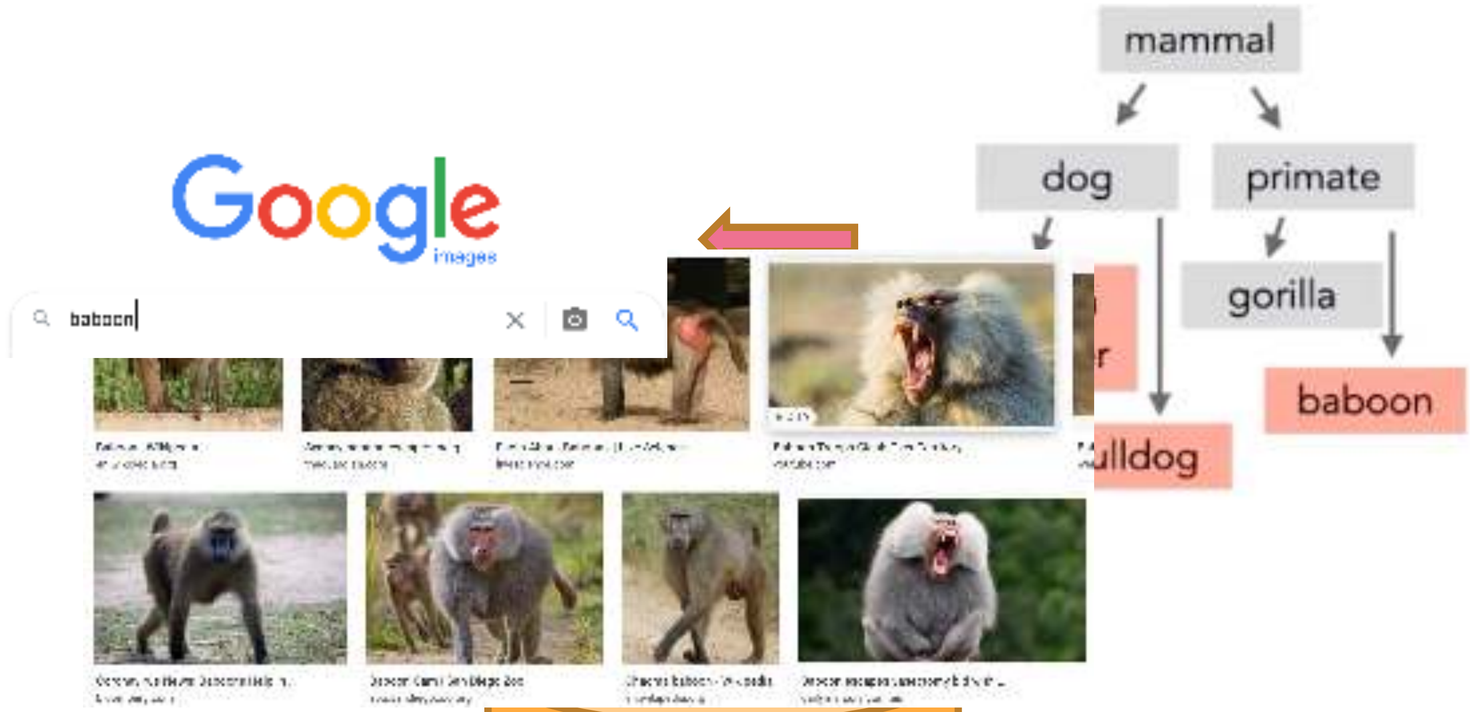
The current paradigm of Computer Vision Research



University of BRISTOL

Ideas from: Aleksander Madry, CVPR 2020 workshop on Adversarial machine learning

Machine Learning in Practice



Dataset

Machine Learning in Practice

- Applies to Most ML research at the moment
    - Object Recognition (Pascal, ImageNet, Places, …)
    - Action Recognition (Kinetics-400, -600, -700, AVA, SS, …)
    - …
- Datasets:
    - Methods overfit to the dataset
    - useful for **one** task
    - unnaturally balanced (or nearly balanced) – unrelated to priors outside the dataset itself

One Exception

Machine Learning in Practice

- Autonomous Driving…

Welcome to the KITTI Vision Benchmark Suite!

We take advantage of our autonomous driving platform Annieway to develop novel challenging real-world computer vision benchmarks. Our tasks of interest are: stereo, optical flow, visual odometry, 3D object detection and 3D tracking. For this purpose, we equipped a standard station wagon with two high-resolution color and grayscale video cameras. Accurate ground truth is provided by a Velodyne laser scanner and a GPS localization system. Our datsets are captured by driving around the mid-size city of Karlsruhe, in rural areas and on highways. Up to 15 cars and 30 pedestrians are visible per image. Besides providing all data in raw format, we extract benchmarks for each task. For each of our benchmarks, we also provide an evaluation metric and this evaluation website. Preliminary experiments show that methods ranking high on established benchmarks such as Middlebury perform below average when being moved outside the laboratory to the real world. Our goal is to reduce this bias and complement existing benchmarks by providing real-world benchmarks with novel difficulties to the community.

Share

To get started, grab a cup of your favorite beverage and watch our video trailer (5 minutes):

stereo flow sceneflow depth odometry object tracking road semantics raw data

Machine Learning in Practice

~~Object Recognition~~

Let’s collect Data!

EPIC KITCHENS-100

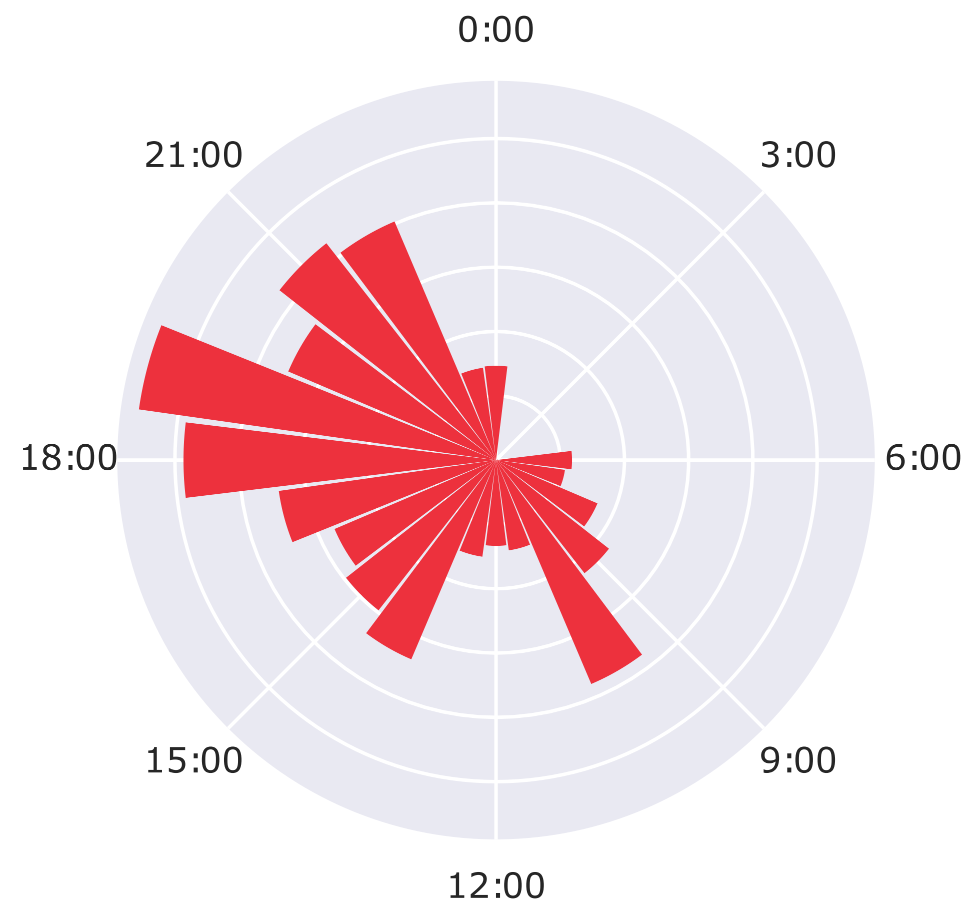

GoPro

Scaling and Rescaling Egocentric Vision

- Head-Mounted Go-Pro, adjustable mounting
- Recording starts immediately before entering the kitchen
- Only stopped before leaving the kitchen

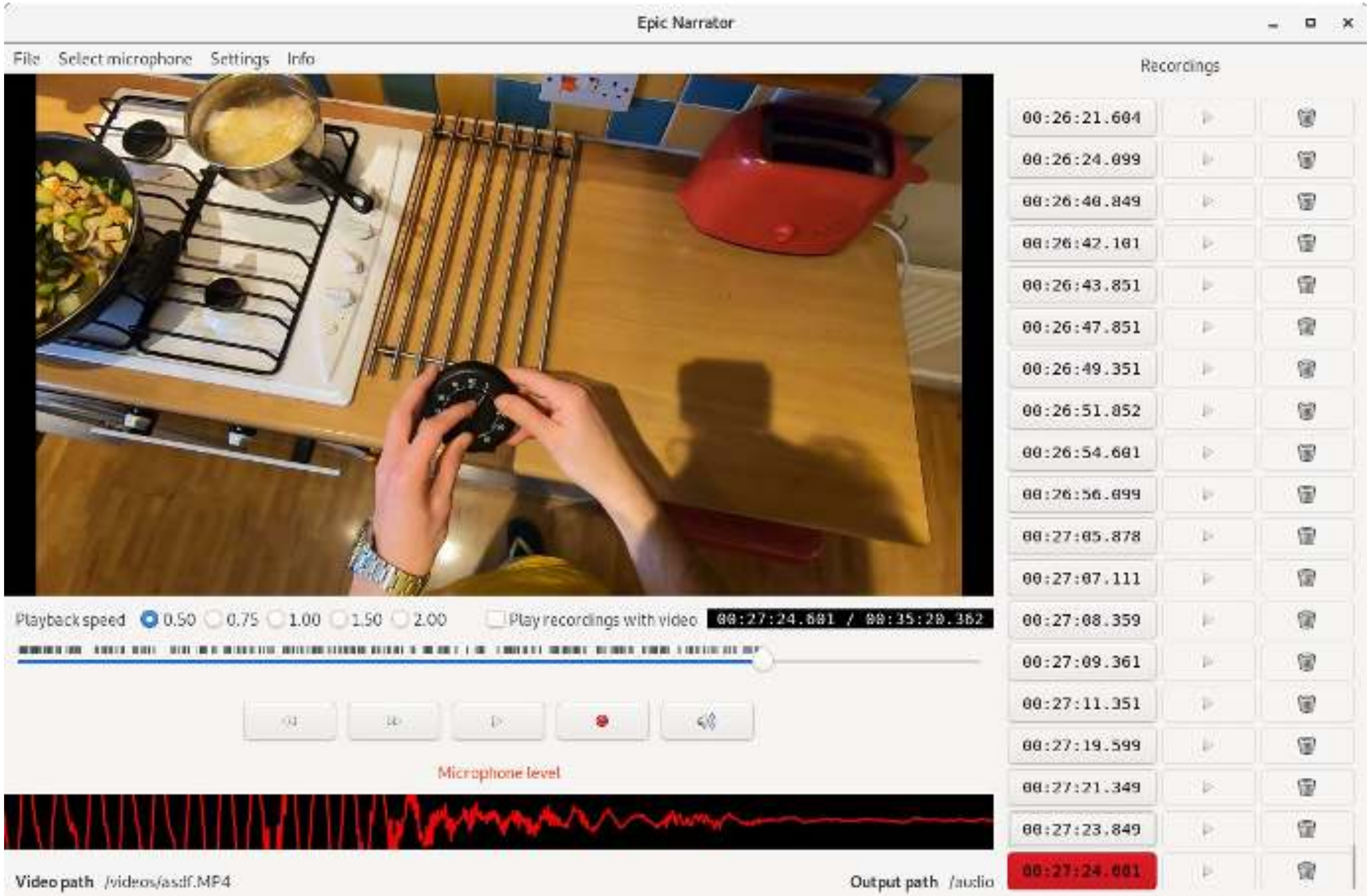
Scaling and Rescaling Egocentric Vision

- 45 kitchens
- Single-person environments
- 4 cities
- May – Nov 2017 – 55 hours
- May – Dec 2019 – 45 hours
- 10 nationalities
- 3 days - all kitchen activities

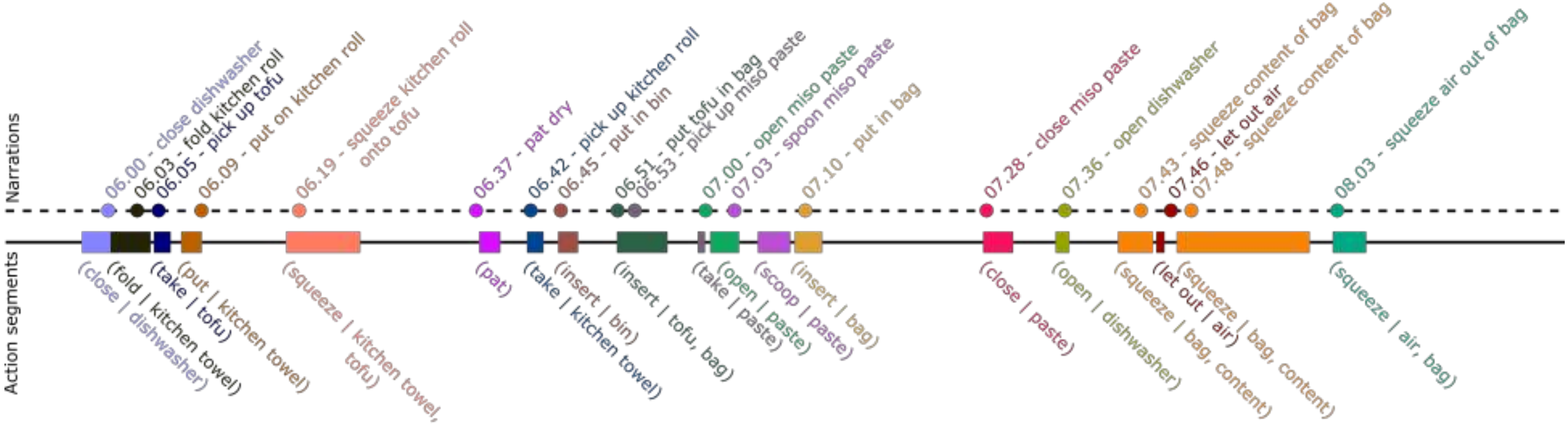


EPIC-KITCHENS

EPIC-KITCHENS



EPIC-KITCHENS



EPIC-KITCHENS and Ego4D

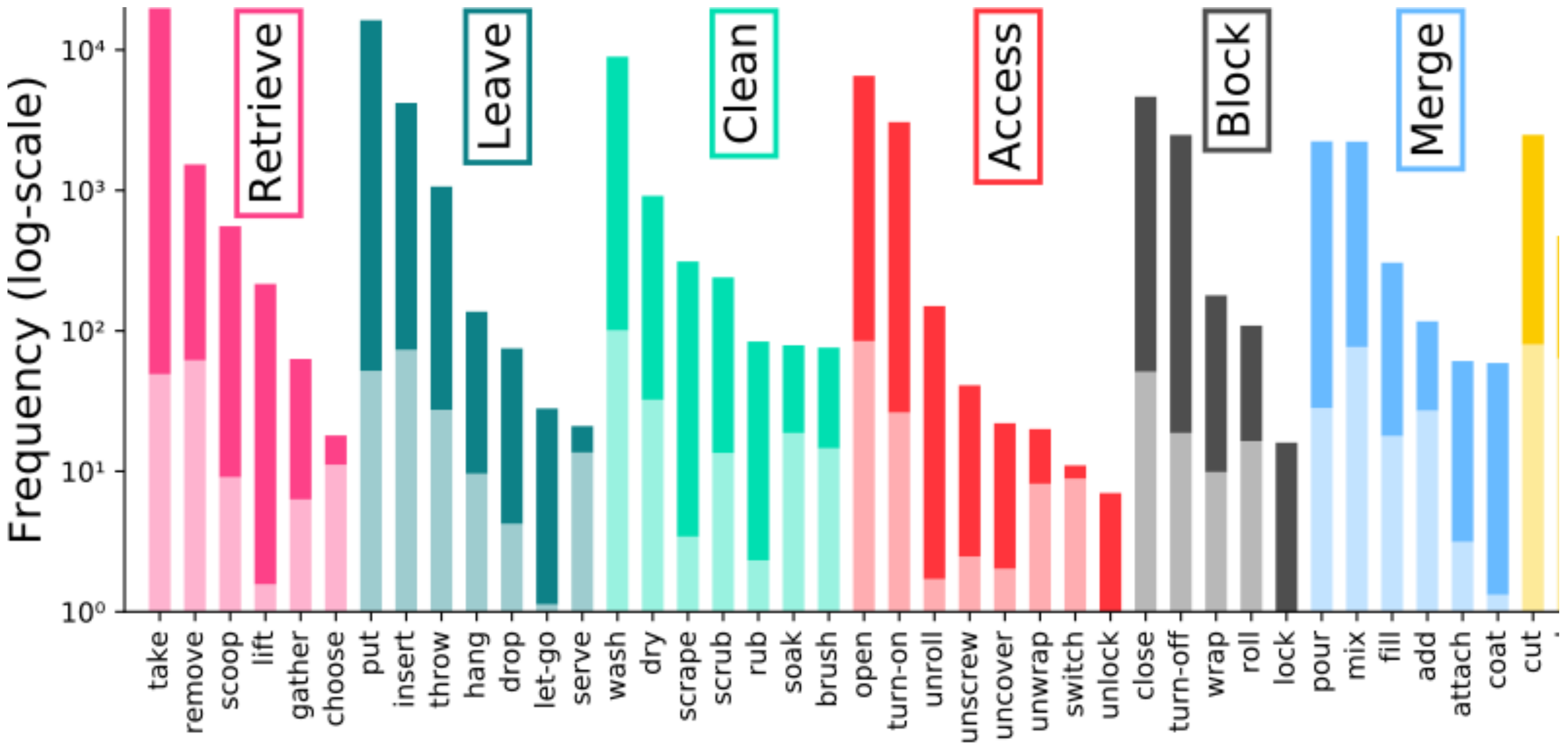
Store
Lay-down **PUT**
Place Rest-on

stove burner
hob
stovetop gas

open vocab	lay-down	stovetop
closed vocab	put	hob
category	leave	appliance


open oven
put spoon on counter
put on glove
pick up fork

EPIC-KITCHENS-100 Statistics



University of BRISTOL

The Data….

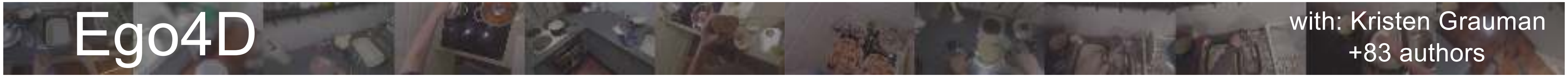
EPIC KITCHENS

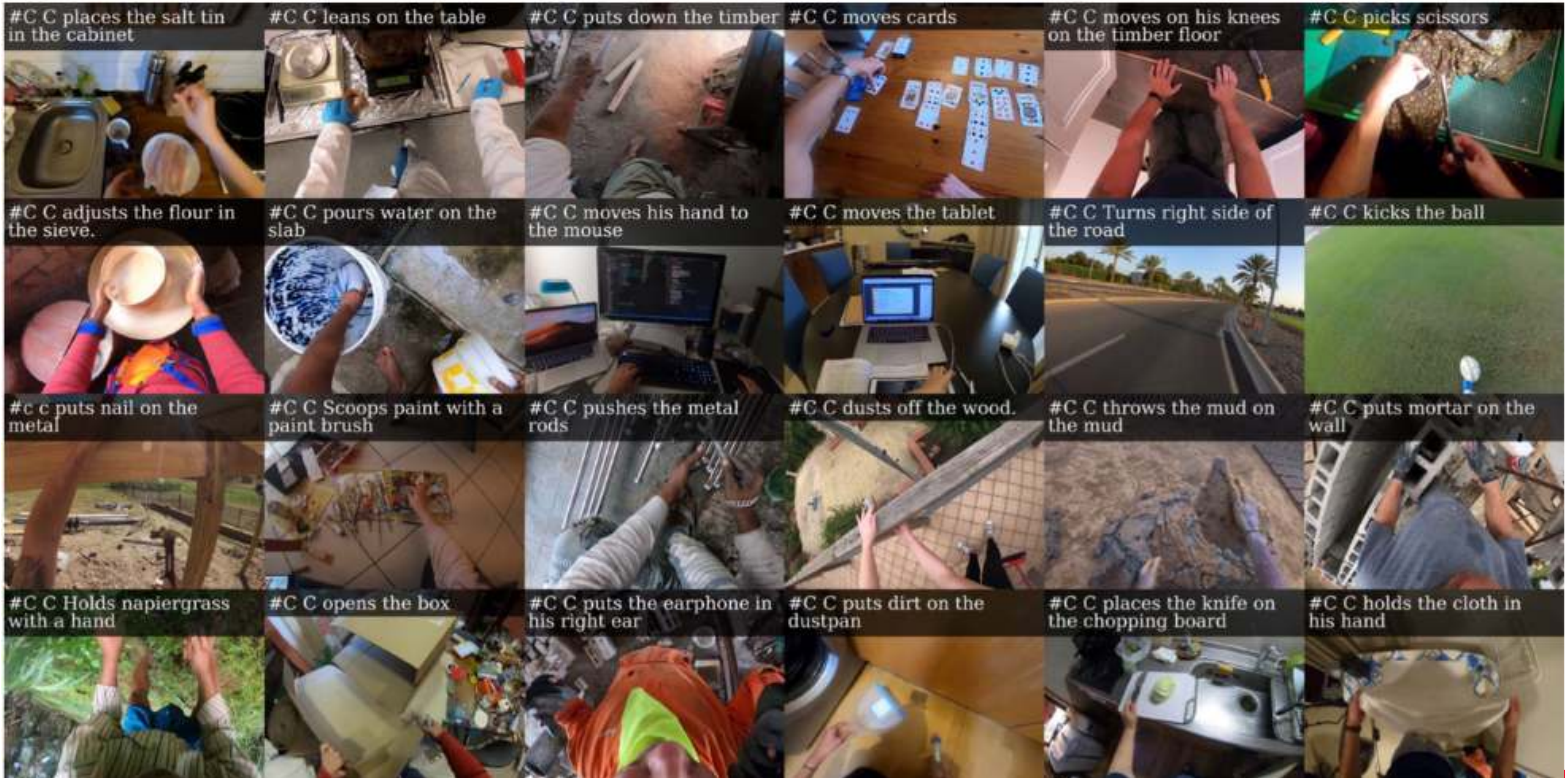
EGO 4D

2017 - now	2020 - now
100 hours 45 kitchens 4 countries Long-term recording Kitchen-based activities	6730 hours 923 participants 74 cities 9 countries Short-term recording All daily activities

Ego4D

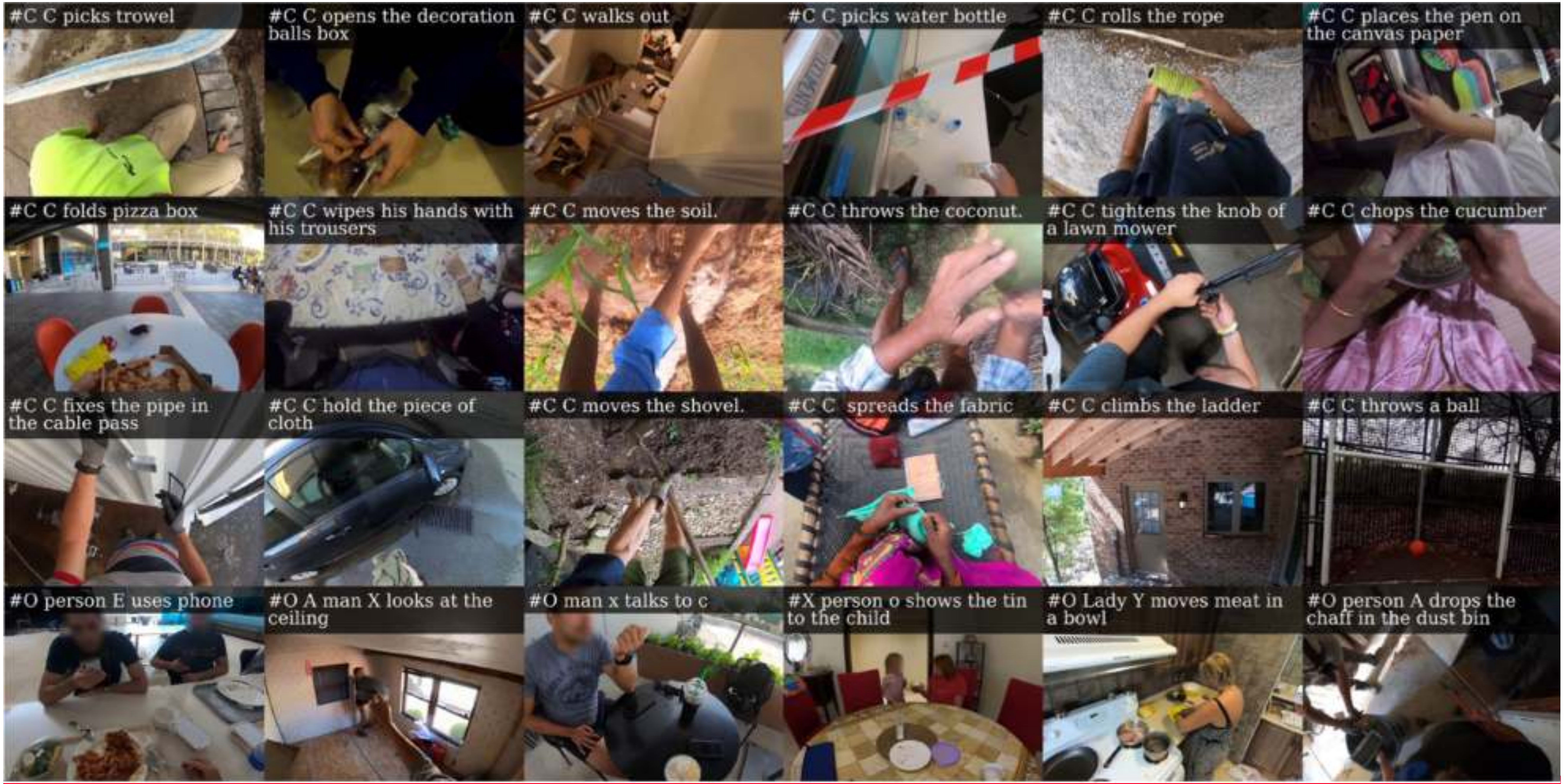
with: Kristen Grauman
+83 authors


Ego4D
with: Kristen Grauman
+83 authors


#C C places the salt tin in the cabinet
#C C leans on the table
#C C puts down the timber
#C C moves cards
#C C moves on his knees on the timber floor
#C C picks scissors
#C C adjusts the flour in the sieve.
#C C pours water on the slab
#C C moves his hand to the mouse
#C C moves the tablet
#C C Turns right side of the road
#C C kicks the ball
#c c puts nail on the metal
#C C Scoops paint with a paint brush
#C C pushes the metal rods
#C C dusts off the wood.
#C C throws the mud on the mud
#C C puts mortar on the wall
#C C Holds napiergrass with a hand
#C C opens the box
#C C puts the earphone in his right ear
#C C puts dirt on the dustpan
#C C places the knife on the chopping board
#C C holds the cloth in his hand

Ego4D

with: Kristen Grauman
+83 authors



Data Collection Exercise

EGO-EXO4D

2024 - now

1286 hours
740 camera wearers
Skilled activities

HD-EPIC

2025 - now

Validation dataset
41 hours
9 participants
Highly-Detailed
Digital Twin

Data Collection Exercise

Pascal VOC
ImageNet
Kinetics
Something-Something

EPIC-KITCHENS
Ego4D
Ego-Exo4D
HD-EPIC
…
KITTI

The chicken or the egg...

Data

Naturally unbalanced

Harder to label (exposes ambiguity)

Closer to application

Multiple tasks

Labels

Unnaturally balanced (or nearly)

Easier to label (hides ambiguity)

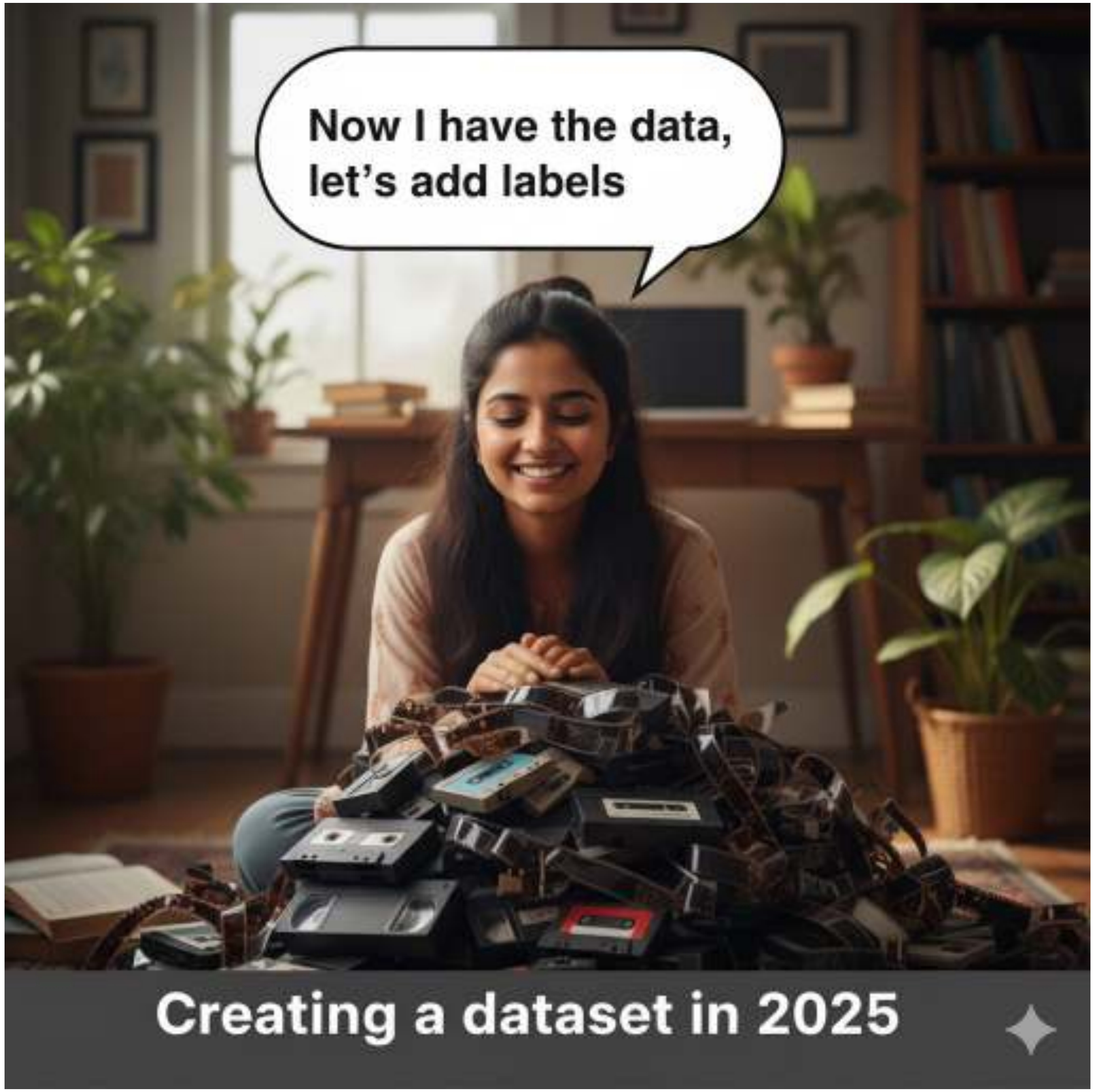
Can be expanded

Single task

The chicken or the egg...

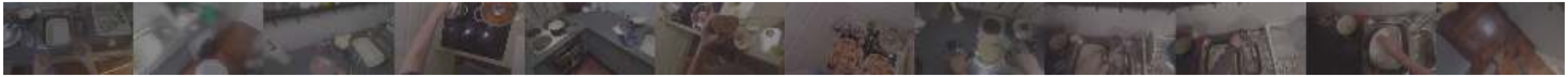
University of BRISTOL

And then?



University of BRISTOL

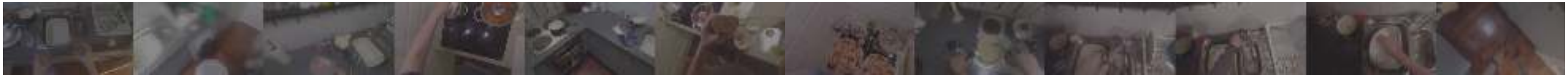
Dima Damen
DataCV W – ICCV2025




Do
It
Yourself

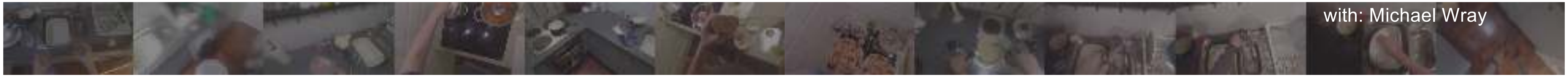
Verb?

Noun?



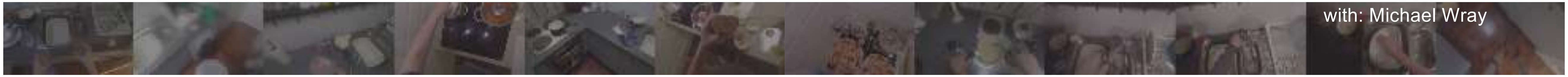
sprinkle salt

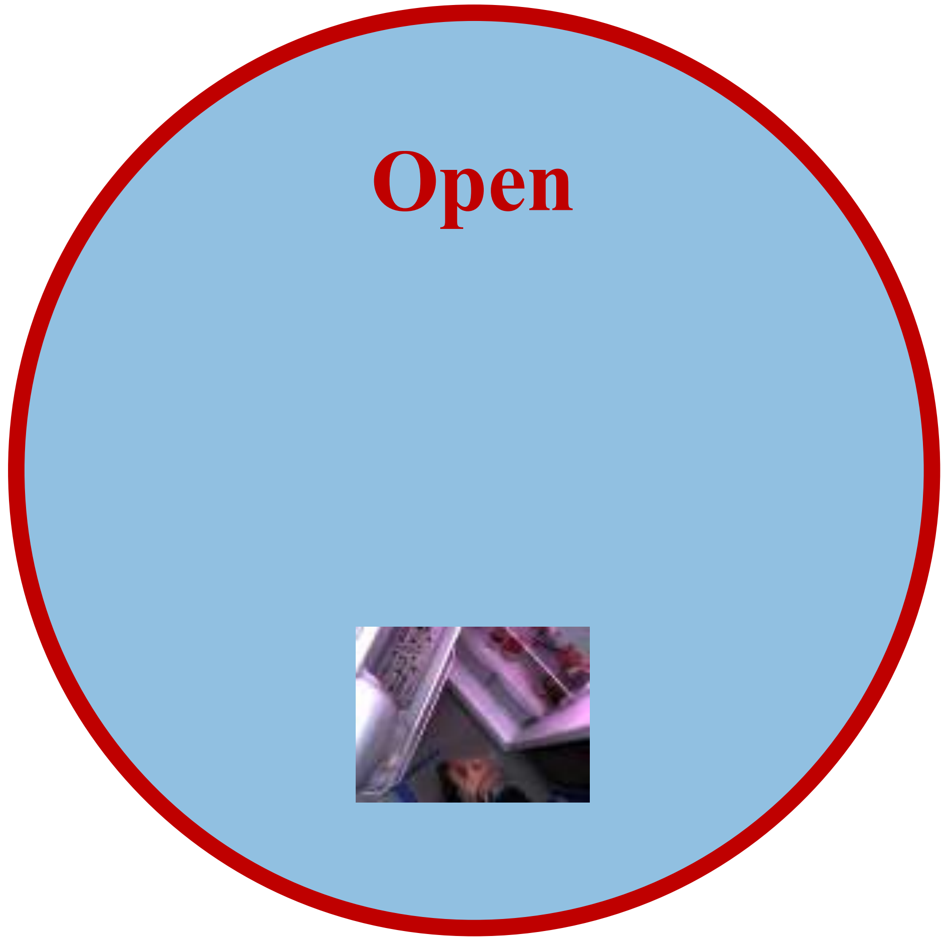
season meat



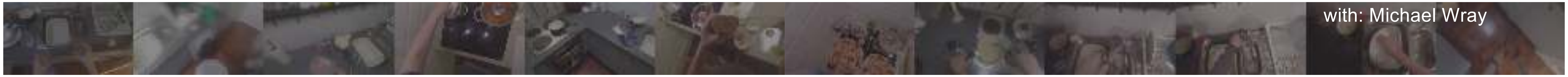
with: Michael Wray

M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. British Machine Vision Conference (BMVC).



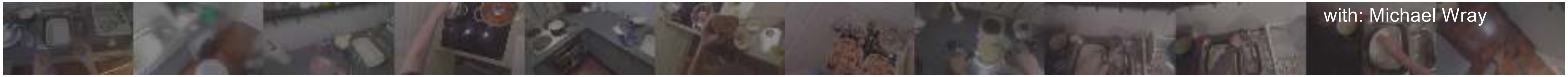


M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. British Machine Vision Conference (BMVC).



with: Michael Wray

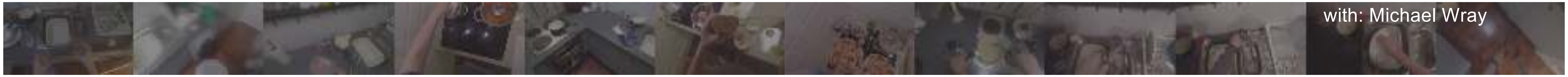
M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. British Machine Vision Conference (BMVC).

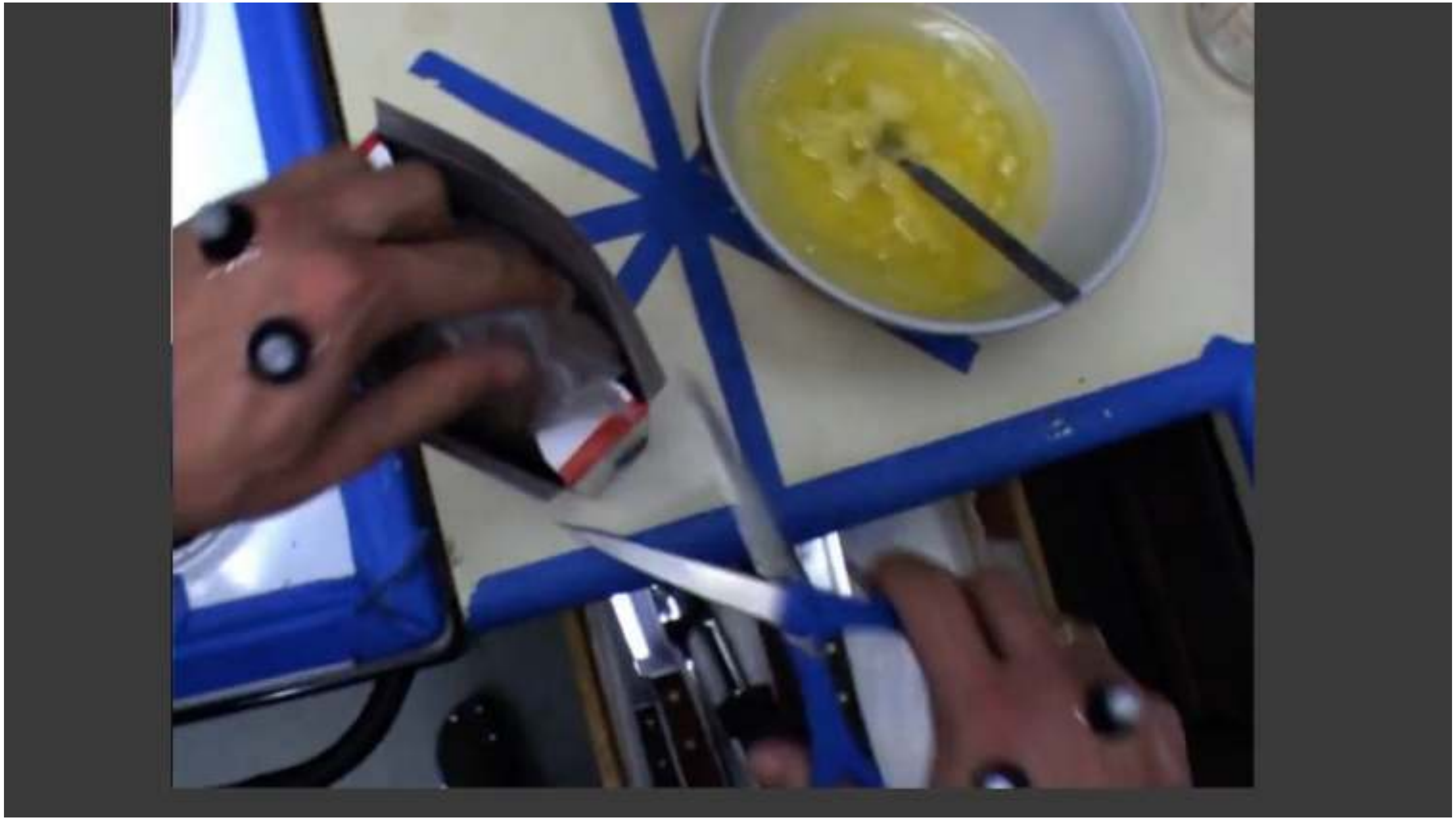


Open

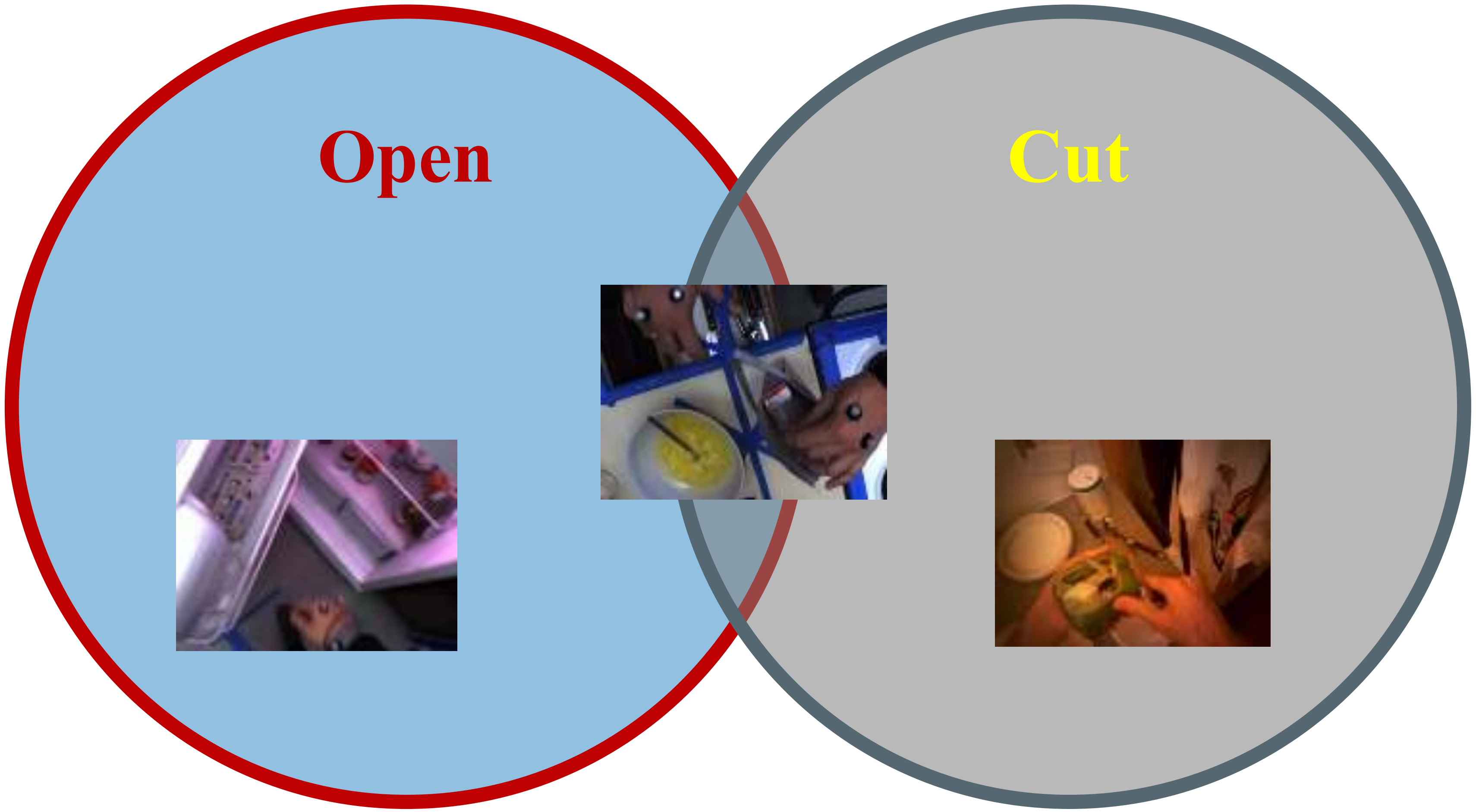
Cut

M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. British Machine Vision Conference (BMVC).





M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. British Machine Vision Conference (BMVC).



M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. British Machine Vision Conference (BMVC).

EPIC-KITCHENS VISOR

with: Ahmad Darkhalil, Dandan Shan, Bin Zhu, Jian Ma,
Amlan Kar, Richard Higgins, David Fouhey, Sanja Fidler

EPIC-KITCHENS VISOR

with: Ahmad Darkhalil, Dandan Shan, Bin Zhu, Jian Ma,
Amlan Kar, Richard Higgins, David Fouhey, Sanja Fidler

University of BRISTOL

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

EPIC-KITCHENS VISOR

with: Ahmad Darkhalil, Dandan Shan, Bin Zhu, Jian Ma,
Amlan Kar, Richard Higgins, David Fouhey, Sanja Fidler

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

EPIC-KITCHENS VISOR

with: Ahmad Darkhalil, Dandan Shan, Bin Zhu, Jian Ma,
Amlan Kar, Richard Higgins, David Fouhey, Sanja Fidler

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

EPIC-KITCHENS VISOR

with: Ahmad Darkhalil, Dandan Shan, Bin Zhu, Jian Ma,
Amlan Kar, Richard Higgins, David Fouhey, Sanja Fidler

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

EPIC-KITCHENS VISOR

with: Ahmad Darkhalil, Dandan Shan, Bin Zhu, Jian Ma,
Amlan Kar, Richard Higgins, David Fouhey, Sanja Fidler

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

EPIC-KITCHENS VISOR

with: Ahmad Darkhalil, Dandan Shan, Bin Zhu, Jian Ma, Amlan Kar, Richard Higgins, David Fouhey, Sanja Fidler

University of BRISTOL

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

EPIC-KITCHENS VISOR

with: Ahmad Darkhalil, Dandan Shan, Bin Zhu, Jian Ma,
Amlan Kar, Richard Higgins, David Fouhey, Sanja Fidler

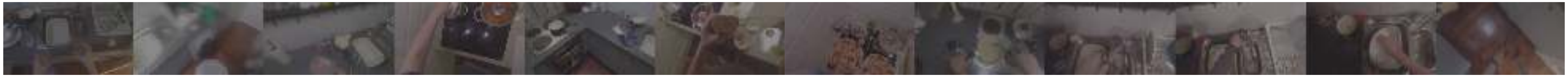
A Darkhalil et al (2022). EPIC-KITCHENS VISOR Benchmark: VIdeo Segmentations and Object Relations. *NeurIPS*

EPIC-KITCHENS VISOR

with: Ahmad Darkhalil, Dandan Shan, Bin Zhu, Jian Ma,
Amlan Kar, Richard Higgins, David Fouhey, Sanja Fidler

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

Semantics are harder than you think…
There are significant ambiguities



From Data to Video

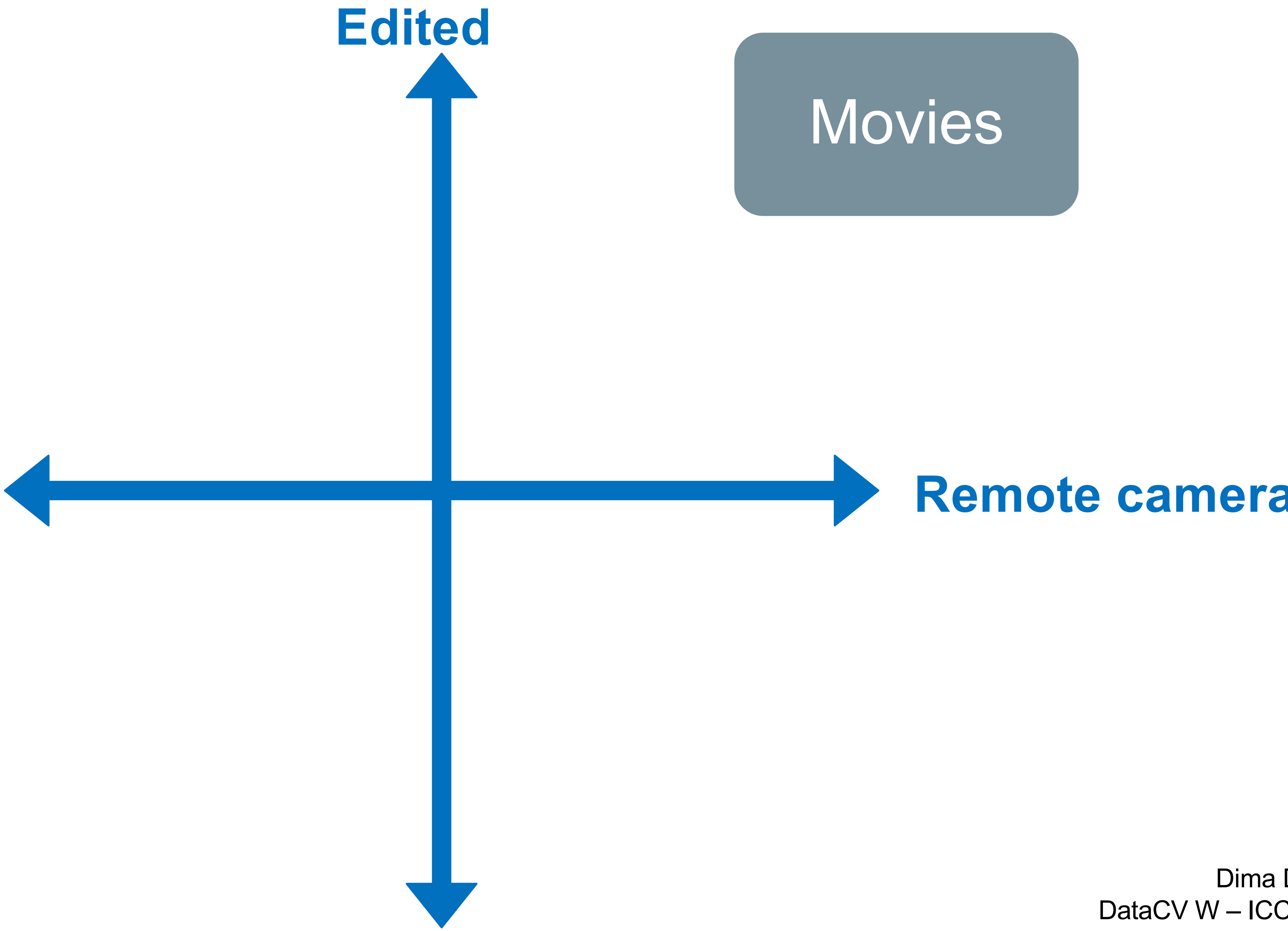
The history of VIDEO

Arrival of a Train at La Ciotat (The Lumière Brothers, 1895)

The history of VIDEO

University of BRISTOL

The history of VIDEO understanding



The history of VIDEO understanding

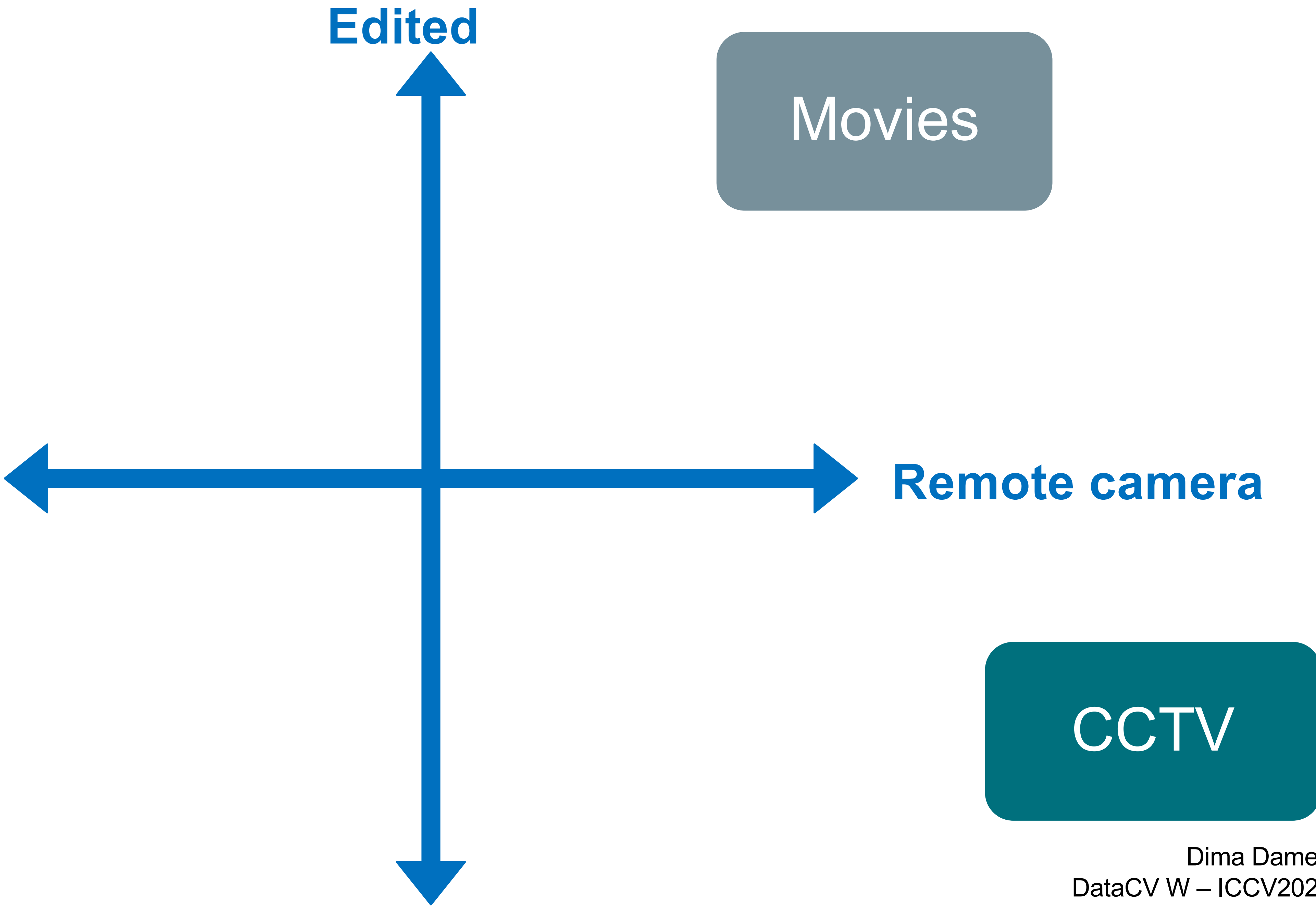
Figure 1. Examples of two action classes (drinking and smoking) from the movie "Coffee and Cigarettes". Note the high within-

Laptev and Perez (2007)

The history of VIDEO understanding

Laptev et al (2008). Learning realistic human actions from movies. CVPR

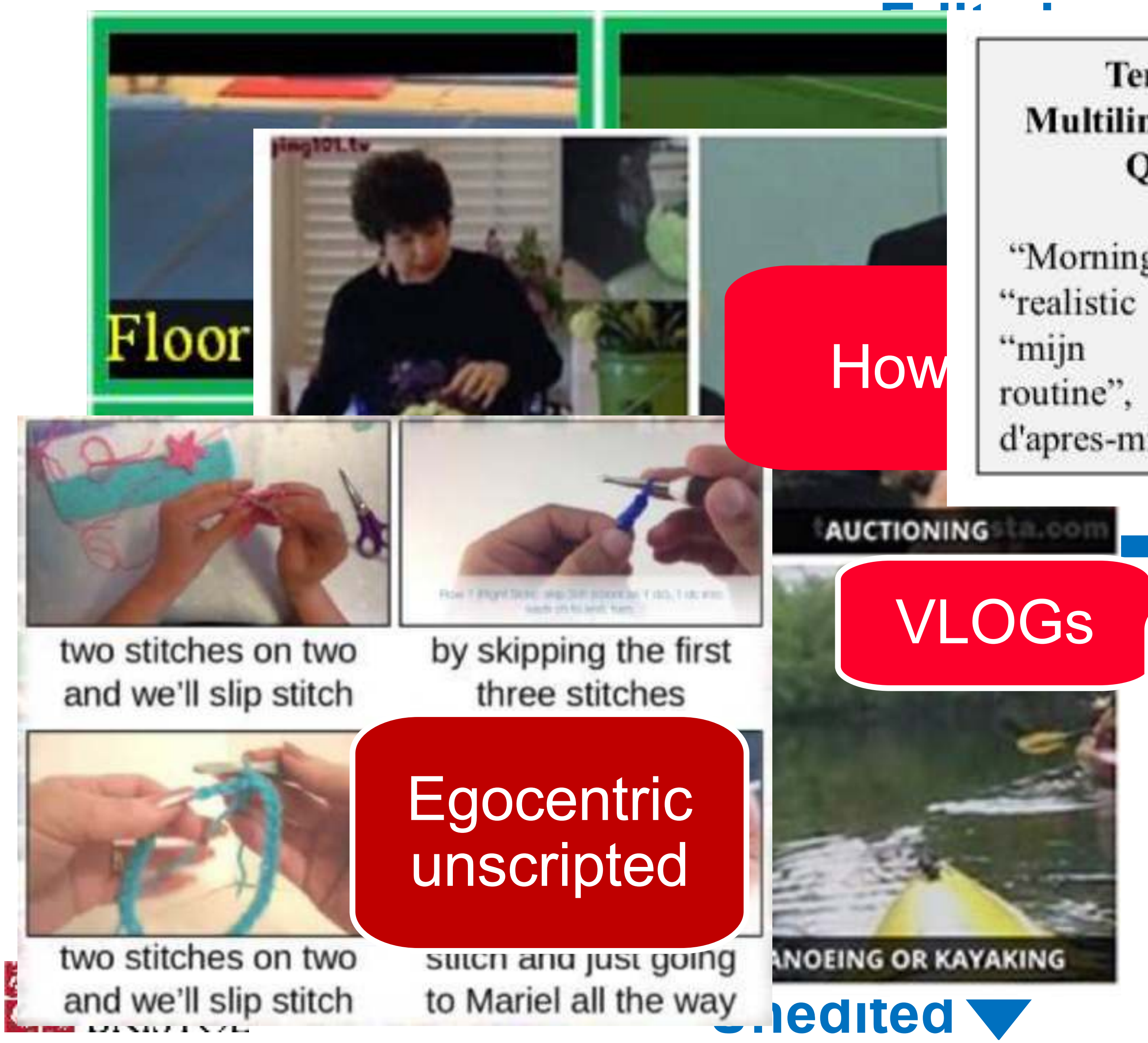
The history of VIDEO understanding

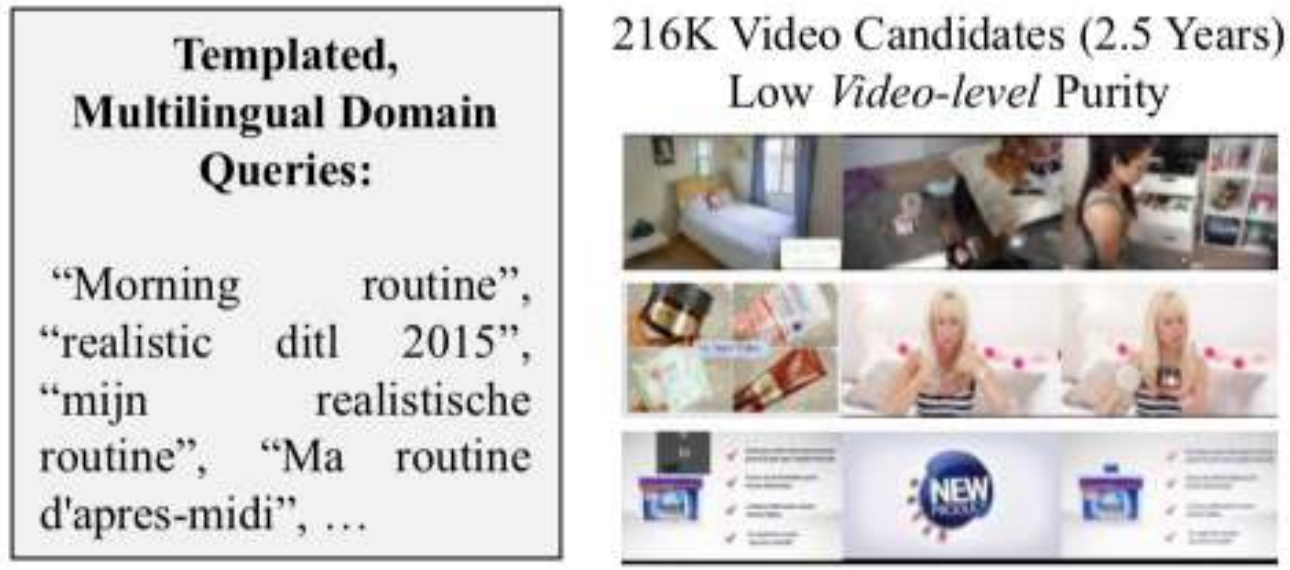


The history of VIDEO understanding

Damen and Hogg (2009). Recognizing linked events: Searching the space of feasible explanations. CVPR1

The history of VIDEO understanding





Remote camera

YouTube Videos

CCTV

Videos are not *one* thing


Videos are multi-modal

Multi-modal learning…

with: Vangelis Kazakos
Arsha Nagrani.
Andrew Zisserman
Jaesung Huh
Jacob Chalk

- The magic of audio-visual understanding…
- Object-Object interactions

Multi-modal learning…

with: Vangelis Kazakos
Arsha Nagrani.
Andrew Zisserman
Jaesung Huh
Jacob Chalk

- The magic of audio-visual understanding…
- Object-Object interactions
- Material sounds

Multi-modal learning…

with: Vangelis Kazakos
Arsha Nagrani.
Andrew Zisserman
Jaesung Huh
Jacob Chalk

- The magic of audio-visual understanding…
- Object-Object interactions
- Material sounds
- Sound-emitting objects

ICASSP 2023
4 - 10 JUNE, RHODES ISLAND, GREECE

EPIC-Sounds: A Large-scale Dataset of Actions That Sound

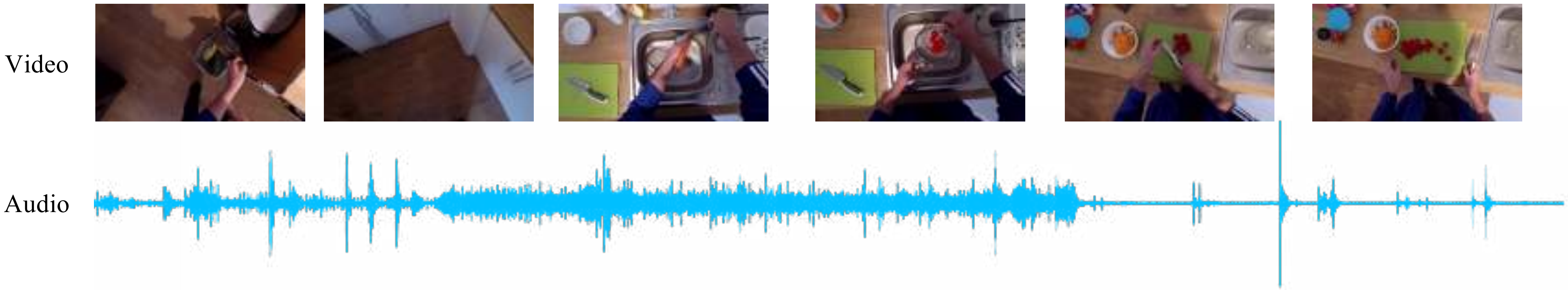
Jaesung Huh*, Jacob Chalk*, Evangelos Kazakos, Dima Damen, Andrew Zisserman
* : Equal contribution

VGG University of Oxford

University of Bristol

Motivation

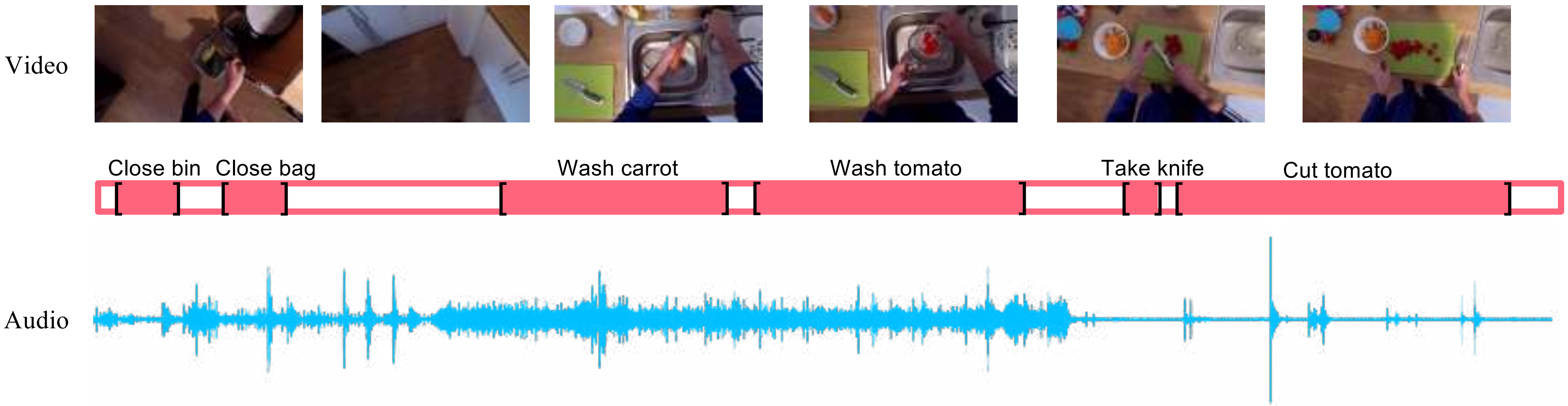
with: Jaesung Huh* & Jacob Chalk*
Vangelis Kazakos Andrew Zisserman



J Huh*, J Chalk* E Kazakos, D Damen, A Zisserman (2023). EPIC-Sounds: A Large-scale Dataset of Actions That Sound. ICASSP

Motivation

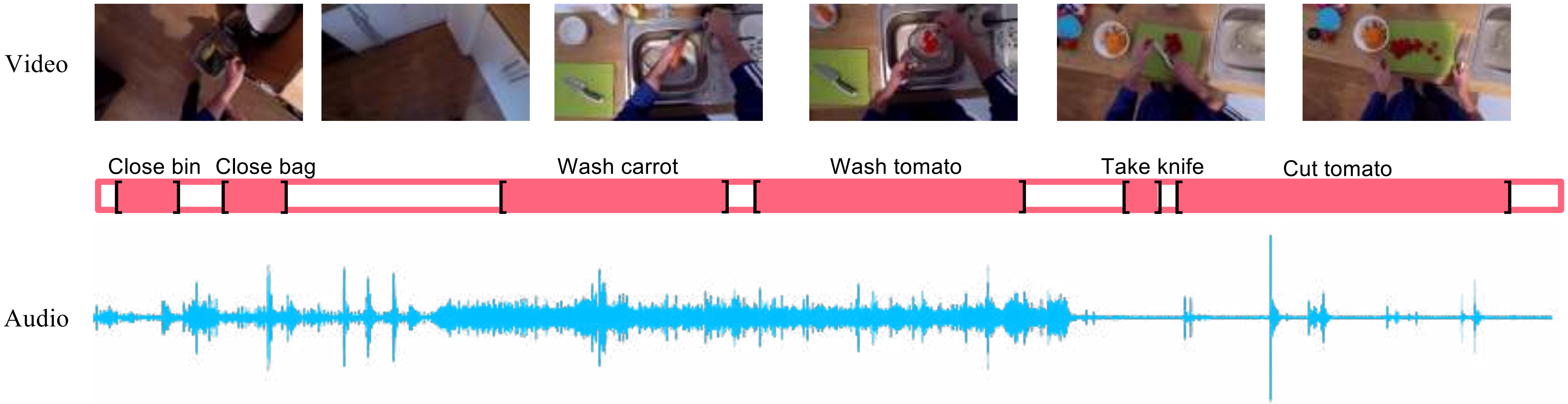
with: Jaesung Huh* & Jacob Chalk*
Vangelis Kazakos Andrew Zisserman



J Huh*, J Chalk* E Kazakos, D Damen, A Zisserman (2023). EPIC-Sounds: A Large-scale Dataset of Actions That Sound. ICASSP

Motivation

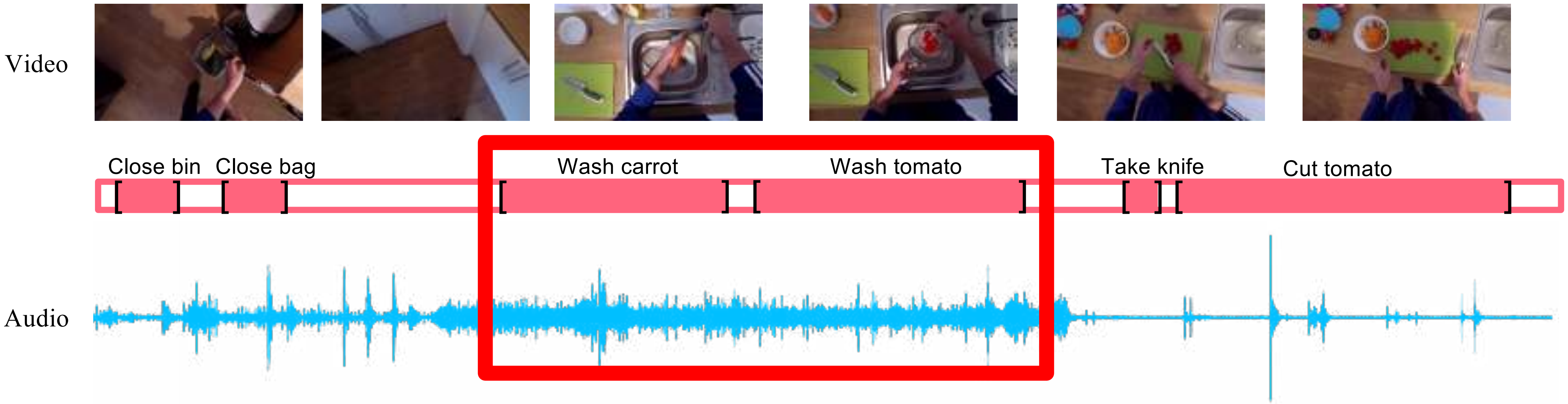
with: Jaesung Huh* & Jacob Chalk*
Vangelis Kazakos Andrew Zisserman



J Huh*, J Chalk* E Kazakos, D Damen, A Zisserman (2023). EPIC-Sounds: A Large-scale Dataset of Actions That Sound. ICASSP

Motivation

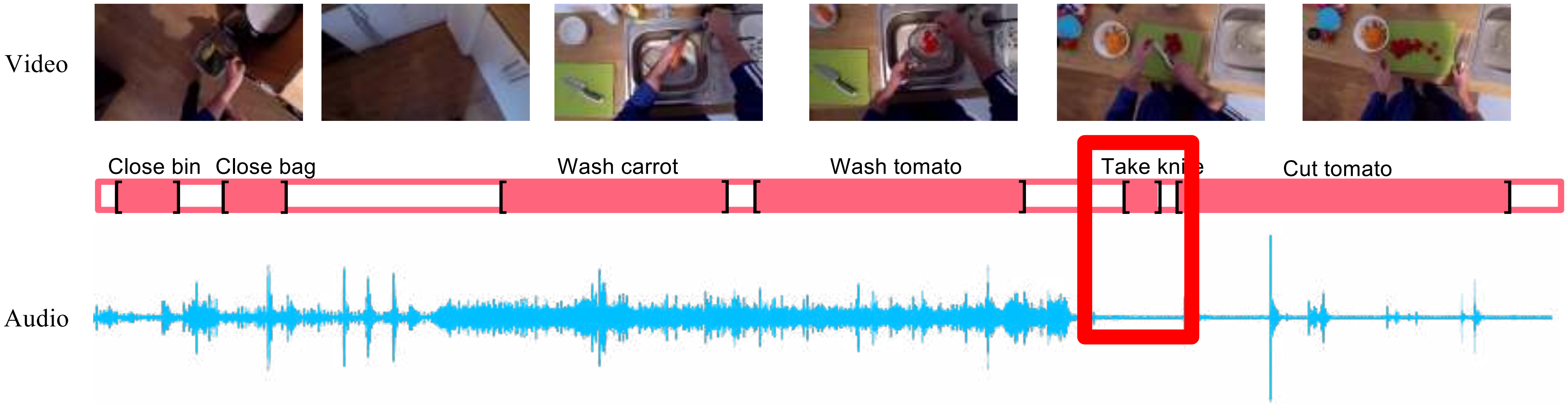
with: Jaesung Huh* & Jacob Chalk*
Vangelis Kazakos Andrew Zisserman



J Huh*, J Chalk* E Kazakos, D Damen, A Zisserman (2023). EPIC-Sounds: A Large-scale Dataset of Actions That Sound. ICASSP

Motivation

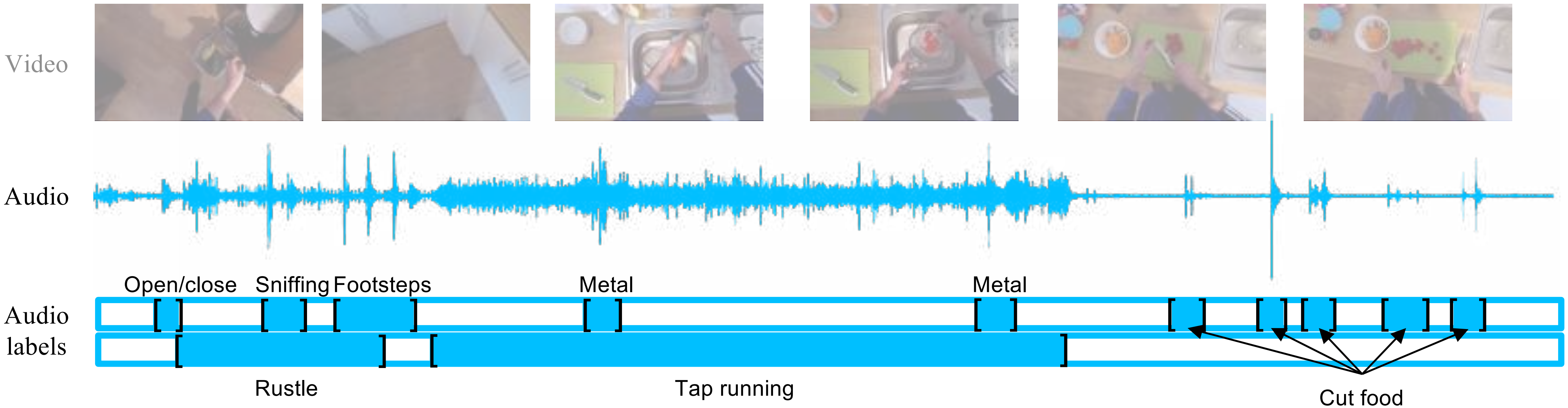
with: Jaesung Huh* & Jacob Chalk*
Vangelis Kazakos Andrew Zisserman



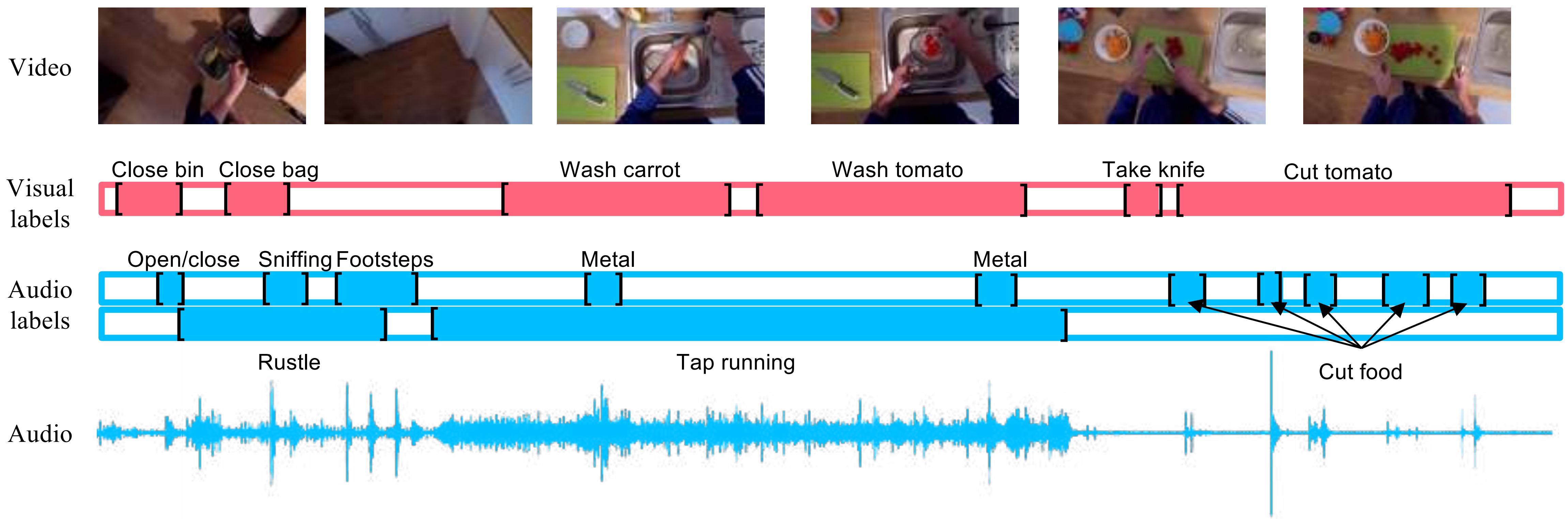
J Huh*, J Chalk* E Kazakos, D Damen, A Zisserman (2023). EPIC-Sounds: A Large-scale Dataset of Actions That Sound. ICASSP

Motivation

with: Jaesung Huh* & Jacob Chalk*
Vangelis Kazakos Andrew Zisserman

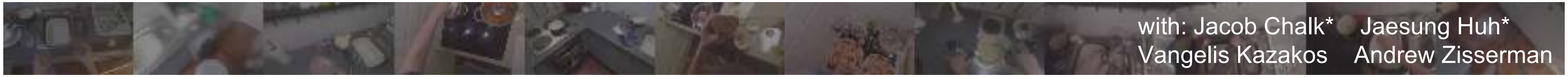


J Huh*, J Chalk* E Kazakos, D Damen, A Zisserman (2023). EPIC-Sounds: A Large-scale Dataset of Actions That Sound. ICASSP



Video
Visual labels
Close bin
Close bag
Wash carrot
Wash tomato
Take knife
Cut tomato
Audio labels
Open/close
Sniffing
Footsteps
Metal
Metal
Rustle
Tap running
Cut food
Audio


spray



CVPR JUNE 17-21, 2024 SEATTLE, WA

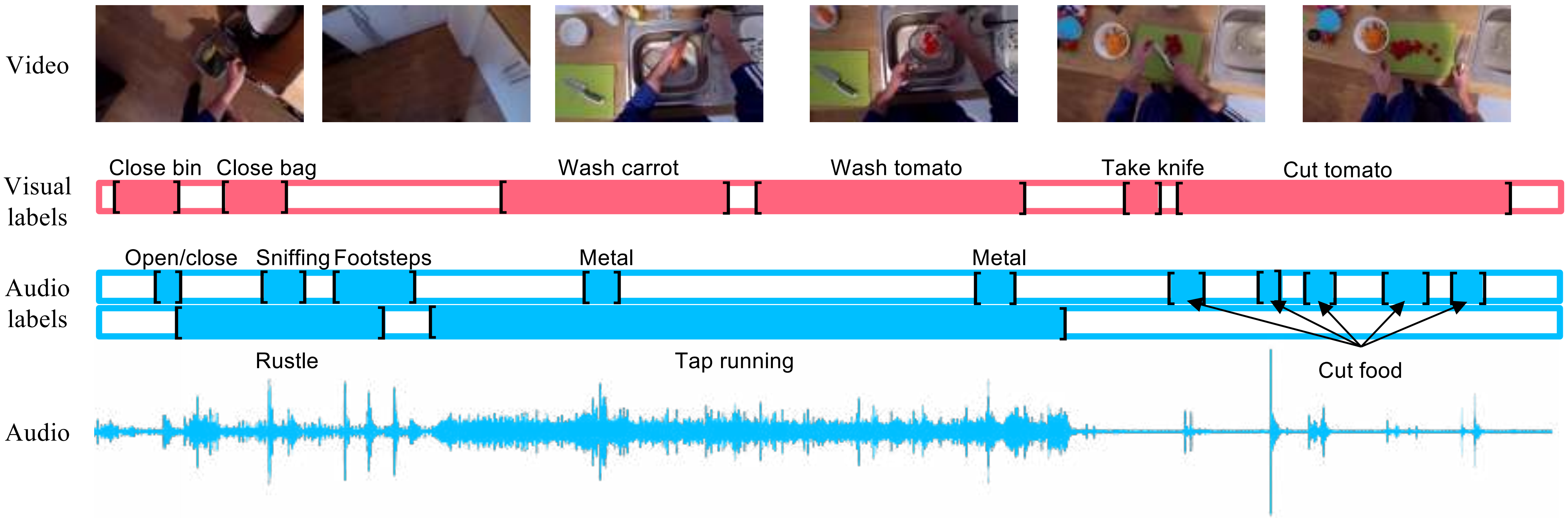
TIM: A Time Interval Machine for Audio-Visual Action Recognition

Jacob Chalk*, Jaesung Huh*, Evangelos Kazakos, Andrew Zisserman, Dima Damen
* : Equal contribution

University of Bristol
VGG University of Oxford
Czech Technical University in Prague

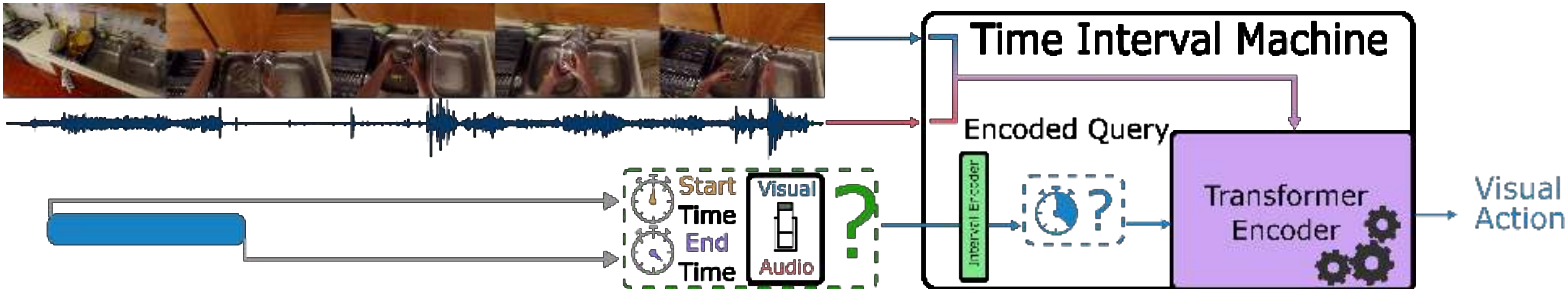
Multi-Modal Long-Form Dataset

with: Jacob Chalk* Jaesung Huh*
Vangelis Kazakos Andrew Zisserman



TIM: A Time-Interval Audio-Visual Machine

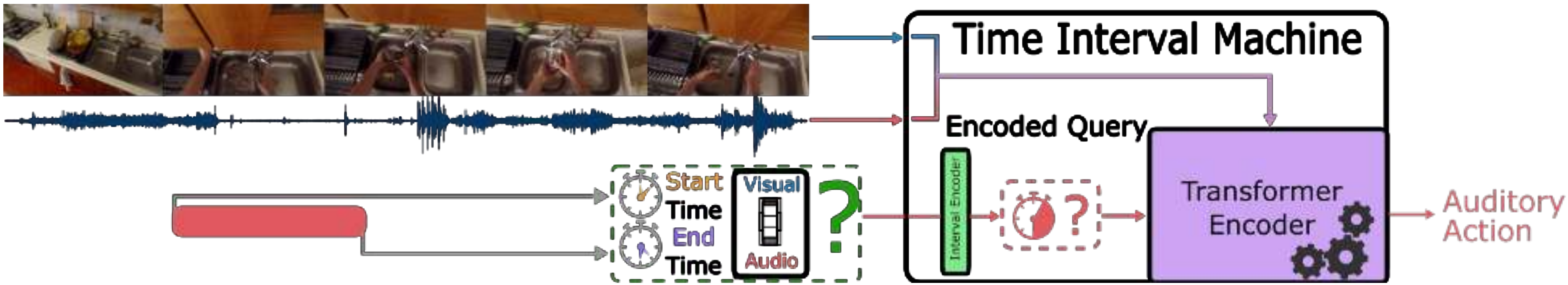
with: Jacob Chalk* Jaesung Huh*
Vangelis Kazakos Andrew Zisserman



J Chalk*, J Huh*, E Kazakos, A Zisserman, D Damen (2024). TIM: A Time Interval Machine for Audio-Visual Action Recognition. CVPR

TIM: A Time-Interval Audio-Visual Machine

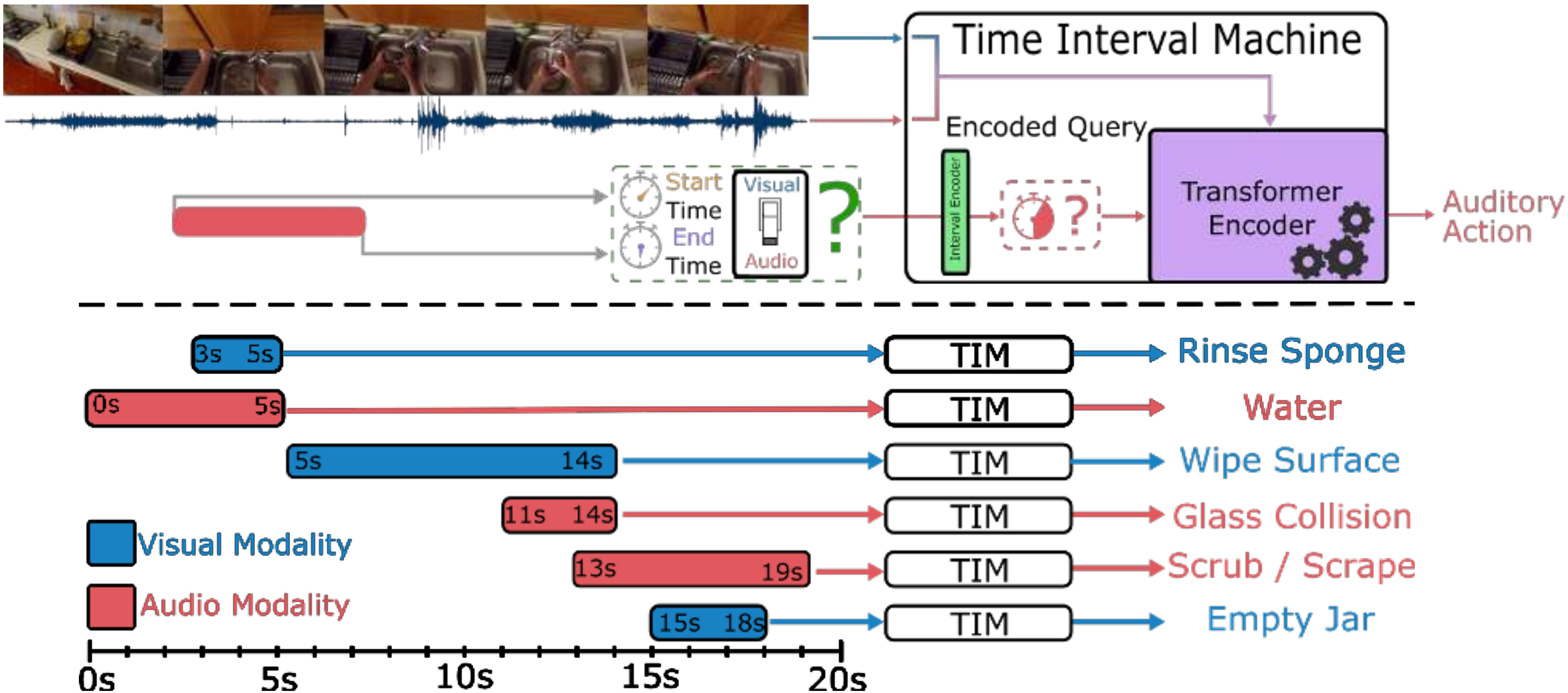
with: Jacob Chalk* Jaesung Huh*
Vangelis Kazakos Andrew Zisserman



J Chalk*, J Huh*, E Kazakos, A Zisserman, D Damen (2024). TIM: A Time Interval Machine for Audio-Visual Action Recognition. CVPR

TIM: A Time-Interval Audio-Visual Machine

with: Jacob Chalk* Jaesung Huh*
Vangelis Kazakos Andrew Zisserman

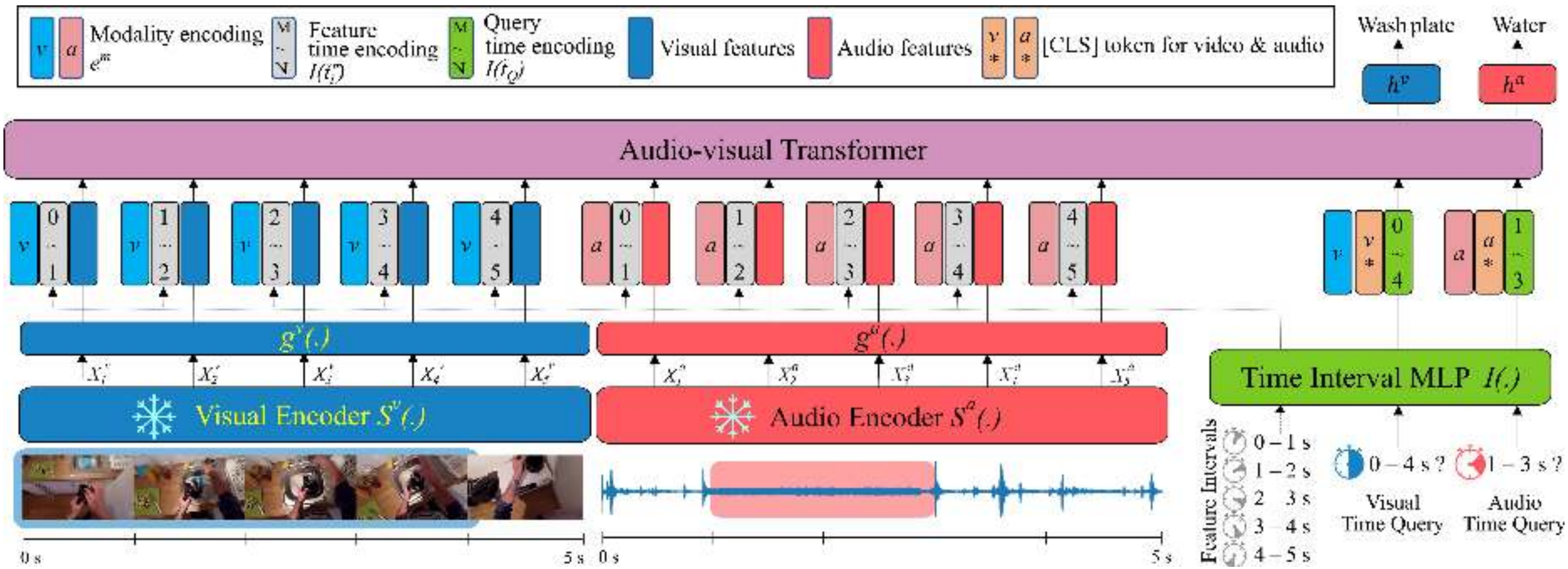


University of BRISTOL

J Chalk*, J Huh*, E Kazakos, A Zisserman, D Damen (2024). TIM: A Time Interval Machine for Audio-Visual Action Recognition. CVPR

TIM: A Time-Interval Audio-Visual Machine

with: Jacob Chalk* Jaesung Huh*
Vangelis Kazakos Andrew Zisserman



J Chalk*, J Huh*, E Kazakos, A Zisserman, D Damen (2024). TIM: A Time Interval Machine for Audio-Visual Action Recognition. CVPR

TIM: A Time-Interval Audio-Visual Machine

with: Jacob Chalk* Jaesung Huh*
Vangelis Kazakos Andrew Zisserman

Model	*xp*	LLM	Verb	Noun	Action
Visual-only models					
MFormer-HR [37]	336p	✗	67.0	58.5	44.5
MoViNet-A6 [27]	320p	✗	72.2	57.3	47.7
MeMViT [55]	224p	✗	71.4	60.3	48.4
Omnivore [14]	224p	✗	69.5	61.7	49.9
MTV [59]	280p	✗	69.9	63.9	50.5
LaViLa (TSF-L) [63]	224p	✓	72.0	62.9	51.0
AVION (ViT-L) [62]	224p	✓	73.0	65.4	54.4
TIM (ours)	224p	✗	**76.2**	**66.4**	**56.4**
Audio-visual models					
TBN [24]	224p	✗	66.0	47.2	36.7
MBT [34]	224p	✗	64.8	58.0	43.4
MTCN [25]	336p	✗	70.7	62.1	49.6
M&M [57]	420p	✗	72.0	66.3	53.6
TIM (ours)	224p	✗	**77.5**	**67.4**	**57.9**

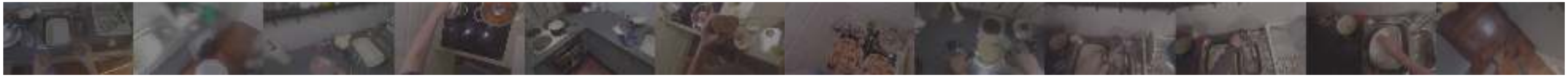
Perception Test Action				
Model	MLP (V)	MTCN [25](A+V)	**TIM (V)**	**TIM (A+V)**
Top-1 acc	43.7	51.2	56.1	**61.1**
Perception Test Sound				
Model	MLP (A)	MTCN [25](A+V)	**TIM (A)**	**TIM (A+V)**
Top-1 acc	50.6	52.9	54.8	**56.1**

Table 5. Comparisons to trained recognition baselines on the Perception Test validation split. We show both action and sound recognition and the benefit of including audio-visual in TIM for both challenges. **V** : visual and **A** : audio input features. MLP is the result by training an MLP classifier with the features directly.

University of BRISTOL

J Chalk*, J Huh*, E Kazakos, A Zisserman, D Damen (2024). TIM: A Time Interval Machine for Audio-Visual Action Recognition. CVPR

Labelling Multi-modalities
is under-explored




COLLECTING A DATASET
IN 2025

HD-EPIC: A Highly-Detailed Egocentric Video Dataset

HD-EPIC

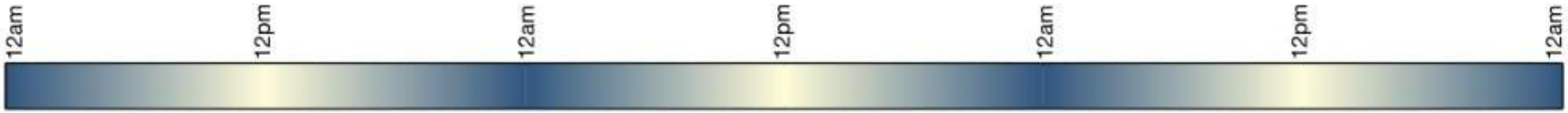
University of Bristol
Universiteit Leiden
Singapore Management University
University of Bath

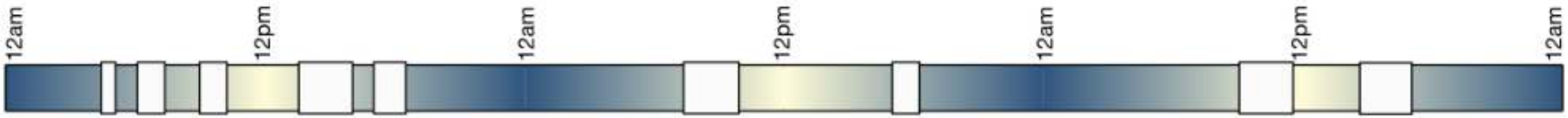
Toby Perrett, Ahmad Darkhalil, Saptarshi Sinha
Omar Emara, Sam Pollard, Kranti Parida
Kaiting Liu, Prajwal Gatti, Siddhant Bansal
Kevin Flanagan, Jacob Chalk, Zhifan Zhu
Rhodri Guerrier, Fahd Abdelazim, Bin Zhu
Davide Moltisanti, Michael Wray, Hazel Doughty
Dima Damen

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

HD-EPIC

HD-EPIC





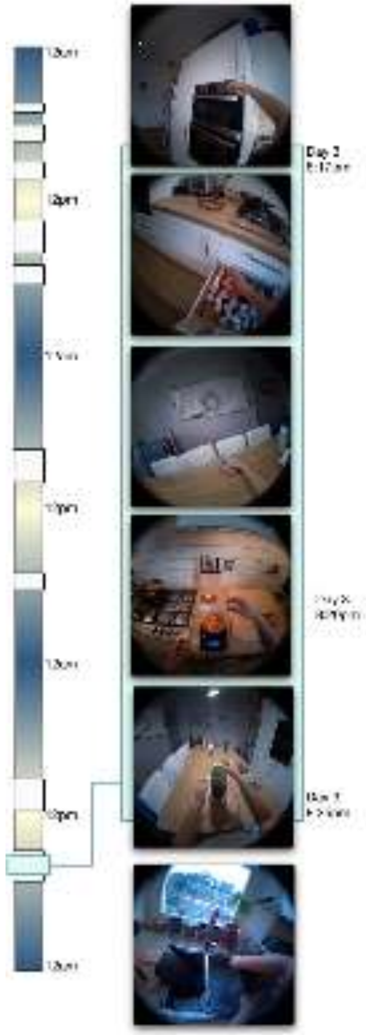
Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

HD-EPIC

HD-EPIC

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.

HD-EPIC

Recipe: Southwestern Salad

1: Preheat the oven to 400F

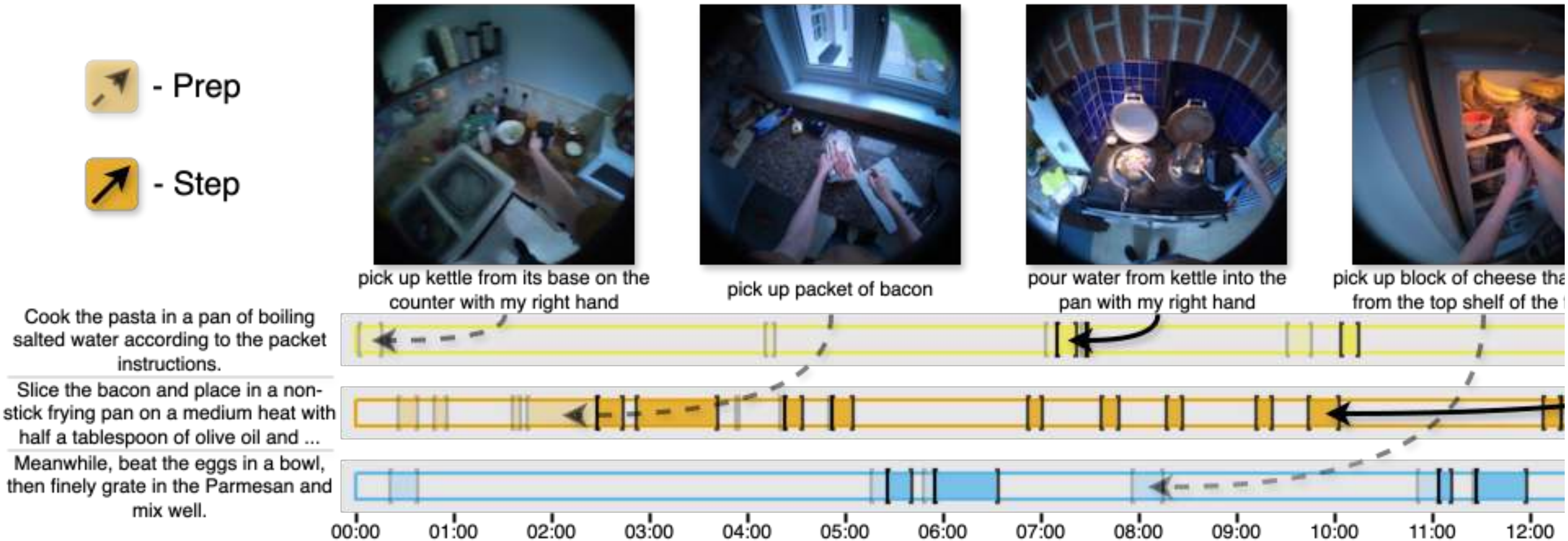
Day 3
:17pm

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

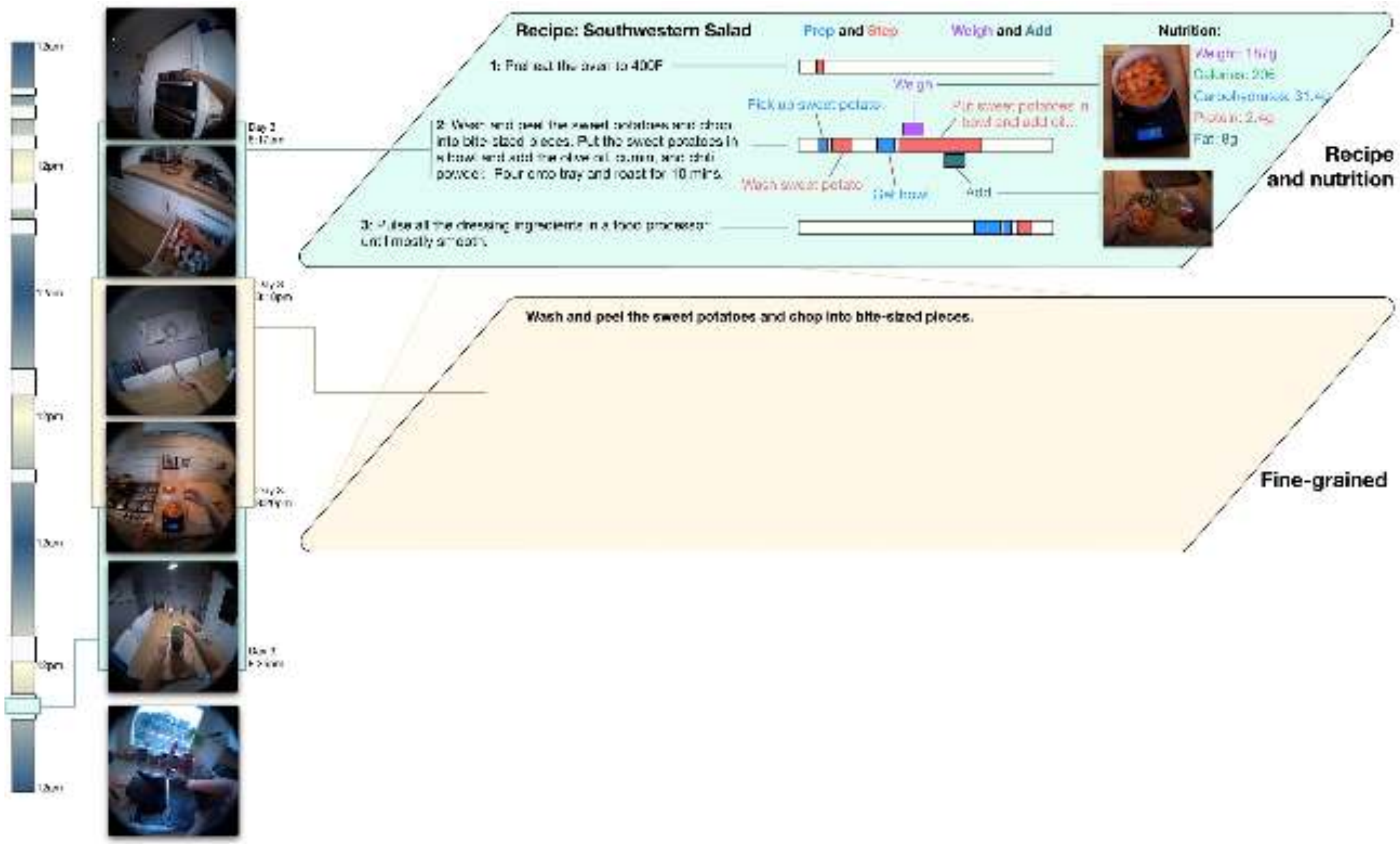
HD-EPIC



University of BRISTOL

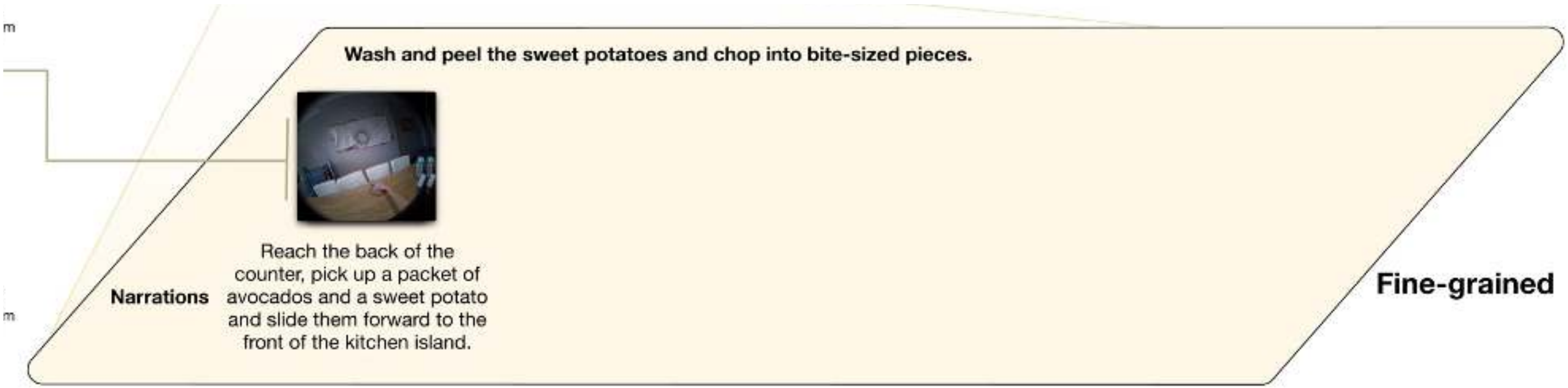
Dima Damen

HD-EPIC



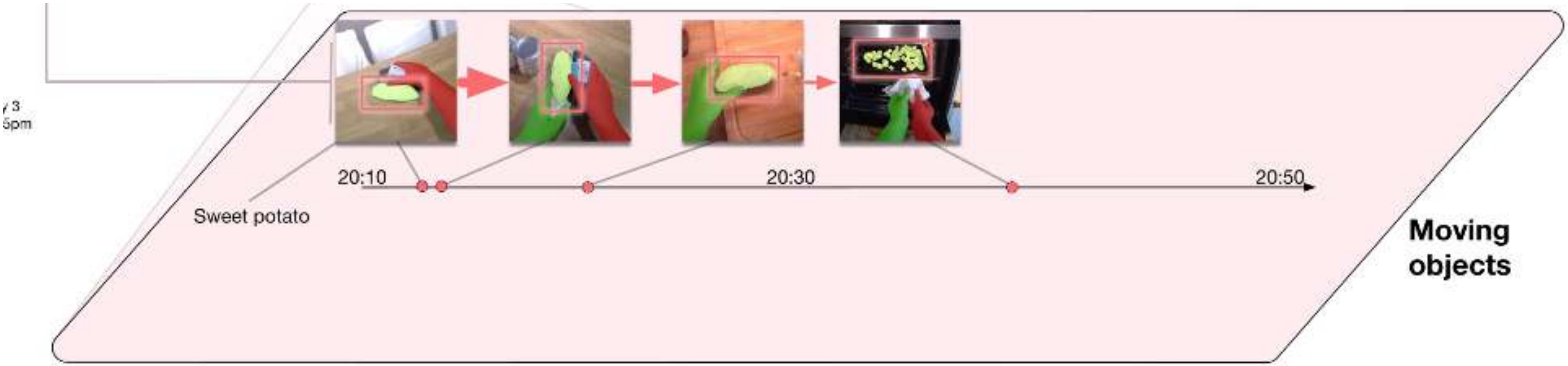
Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

HD-EPIC


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.
Fine-grained

Highly-Detailed Narrations

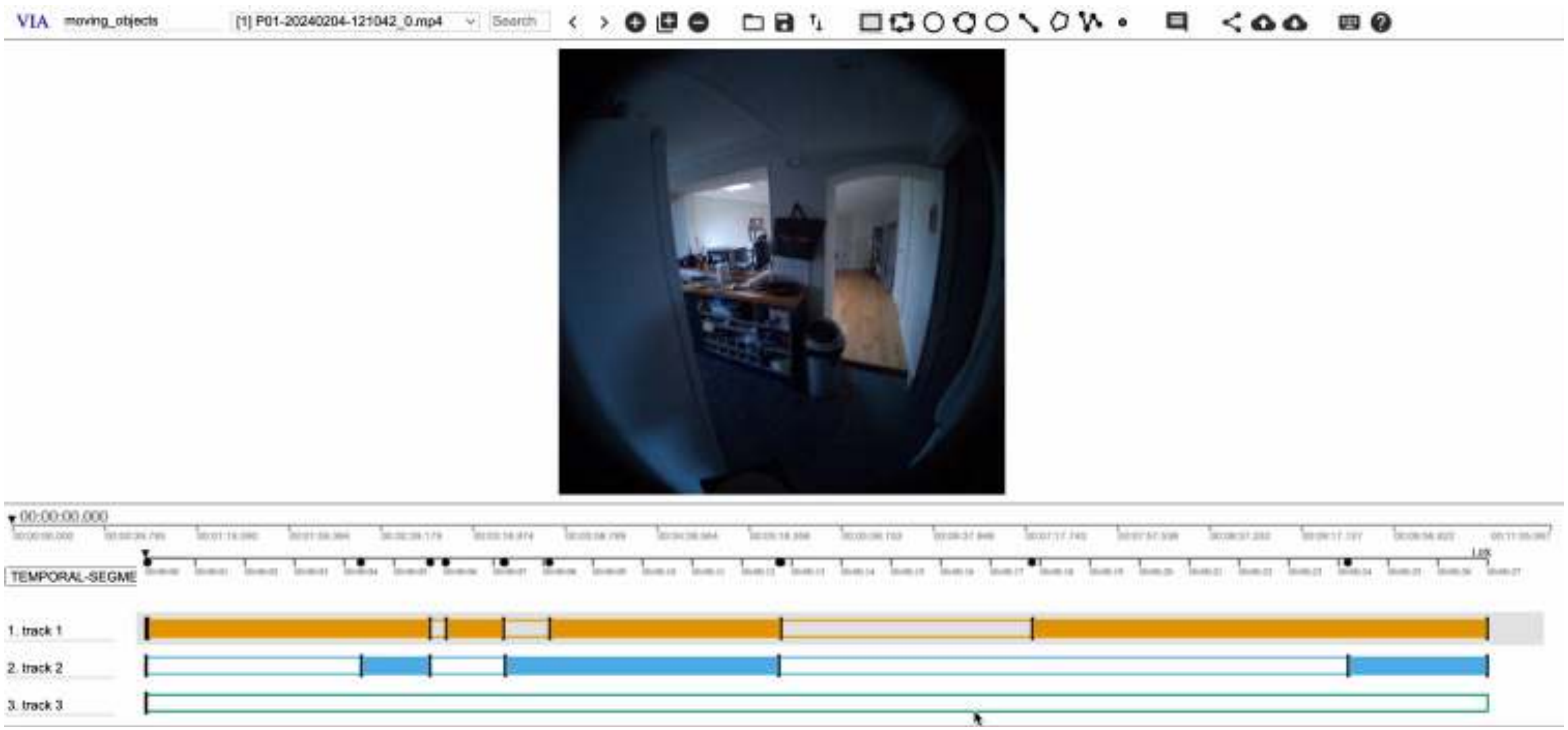
HD-EPIC



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

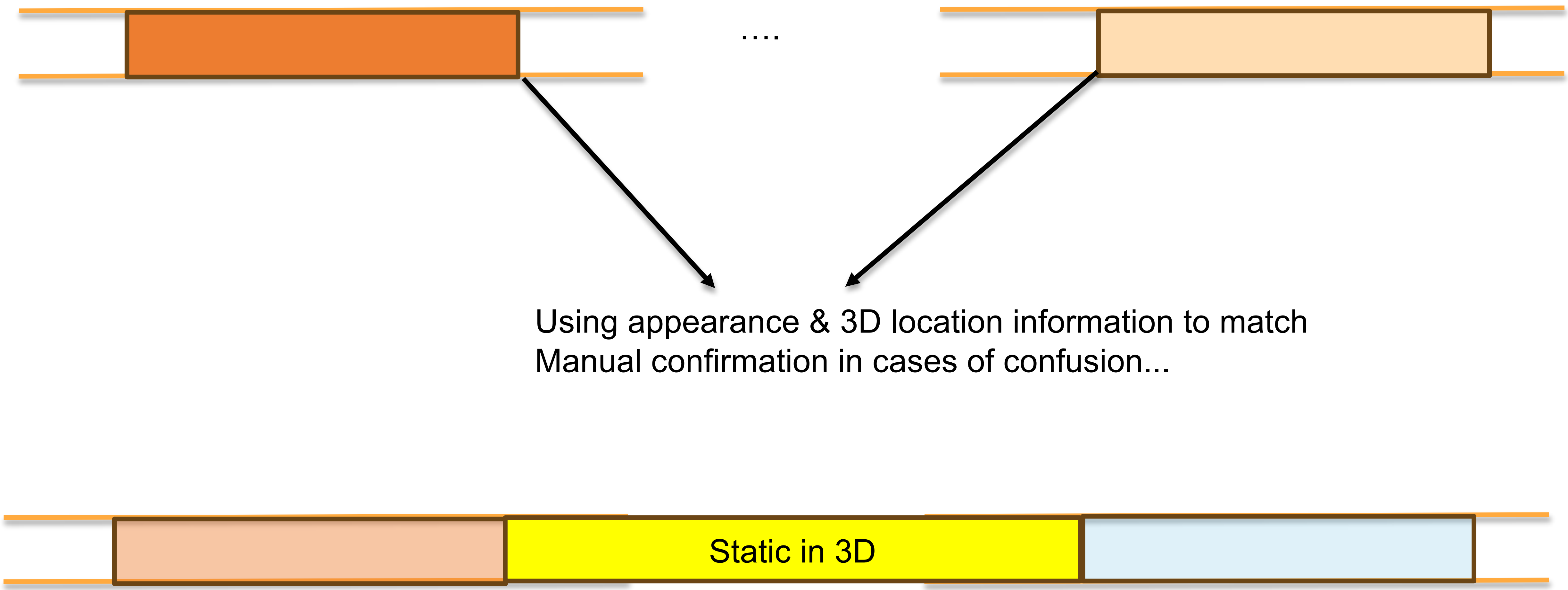
HD-EPIC

- How to minimize the annotations for tracking objects…

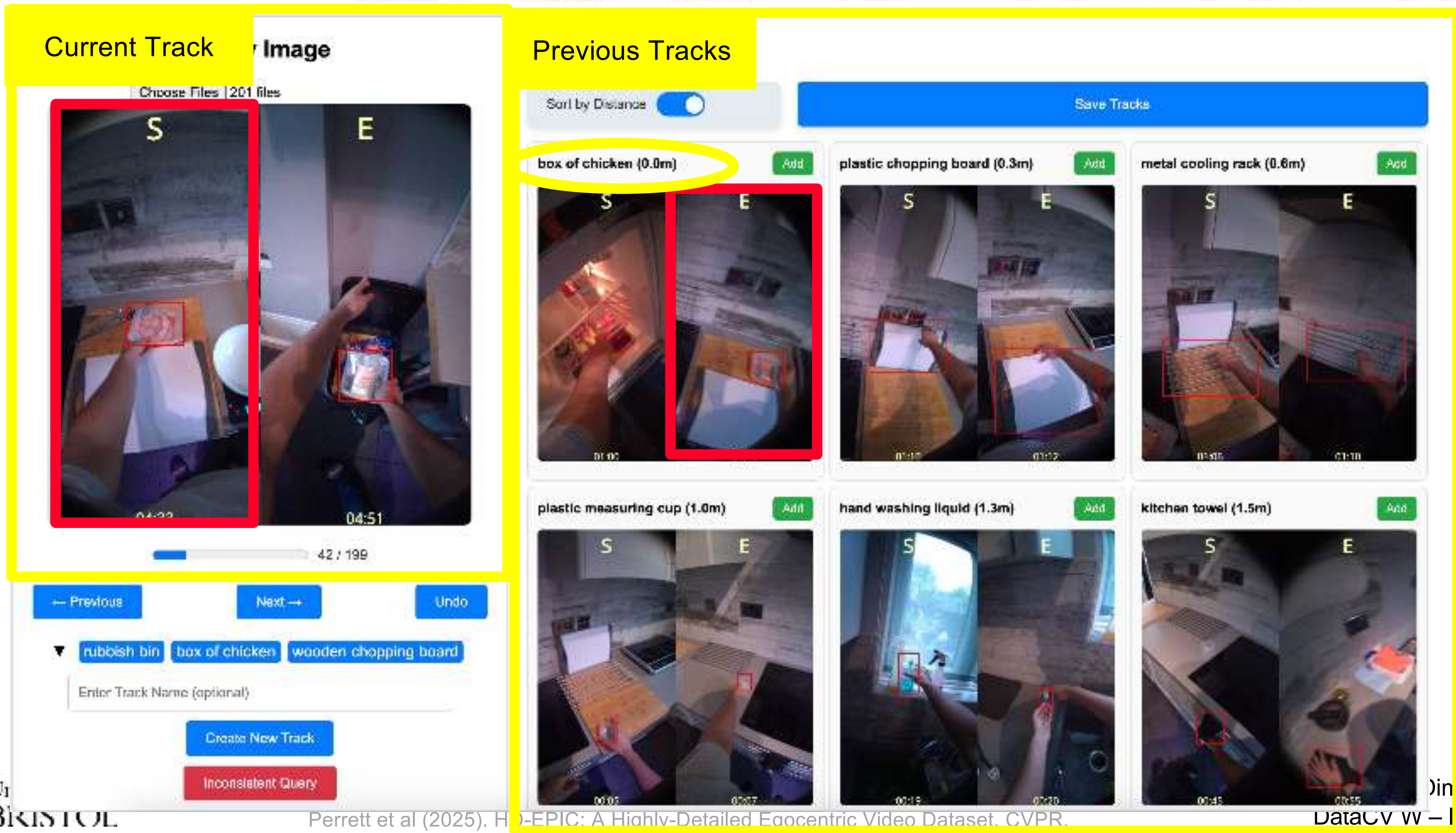


HD-EPIC

- How to minimize the annotations for tracking objects…

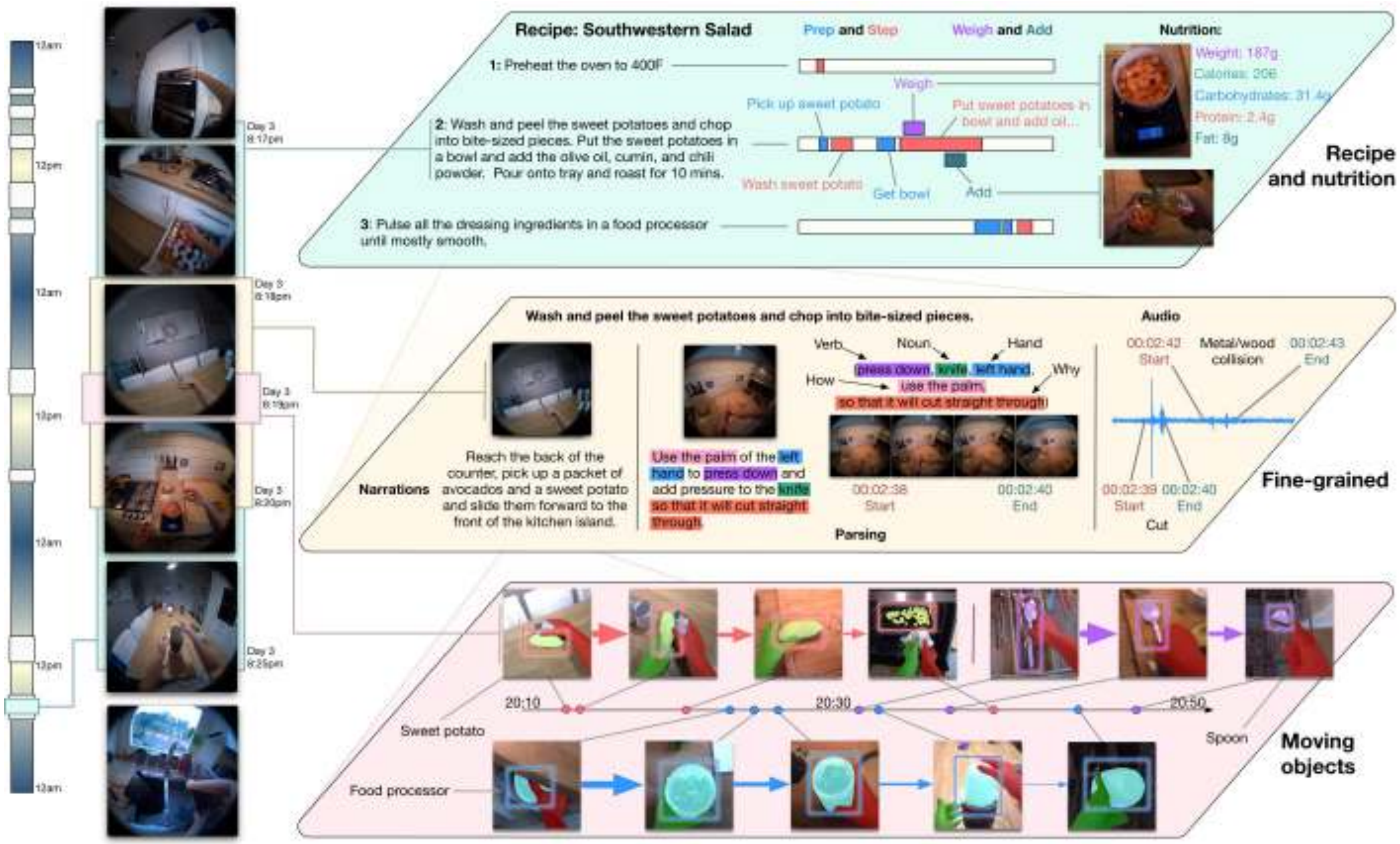


HD-EPIC



Current Track
Image
Choose Files | 201 files
S
E
04:51
42 / 199
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)
← Previous
Next →
Undo
rubbish bin
box of chicken
wooden chopping board
Enter Track Name (optional)
Create New Track
Inconsistent Query

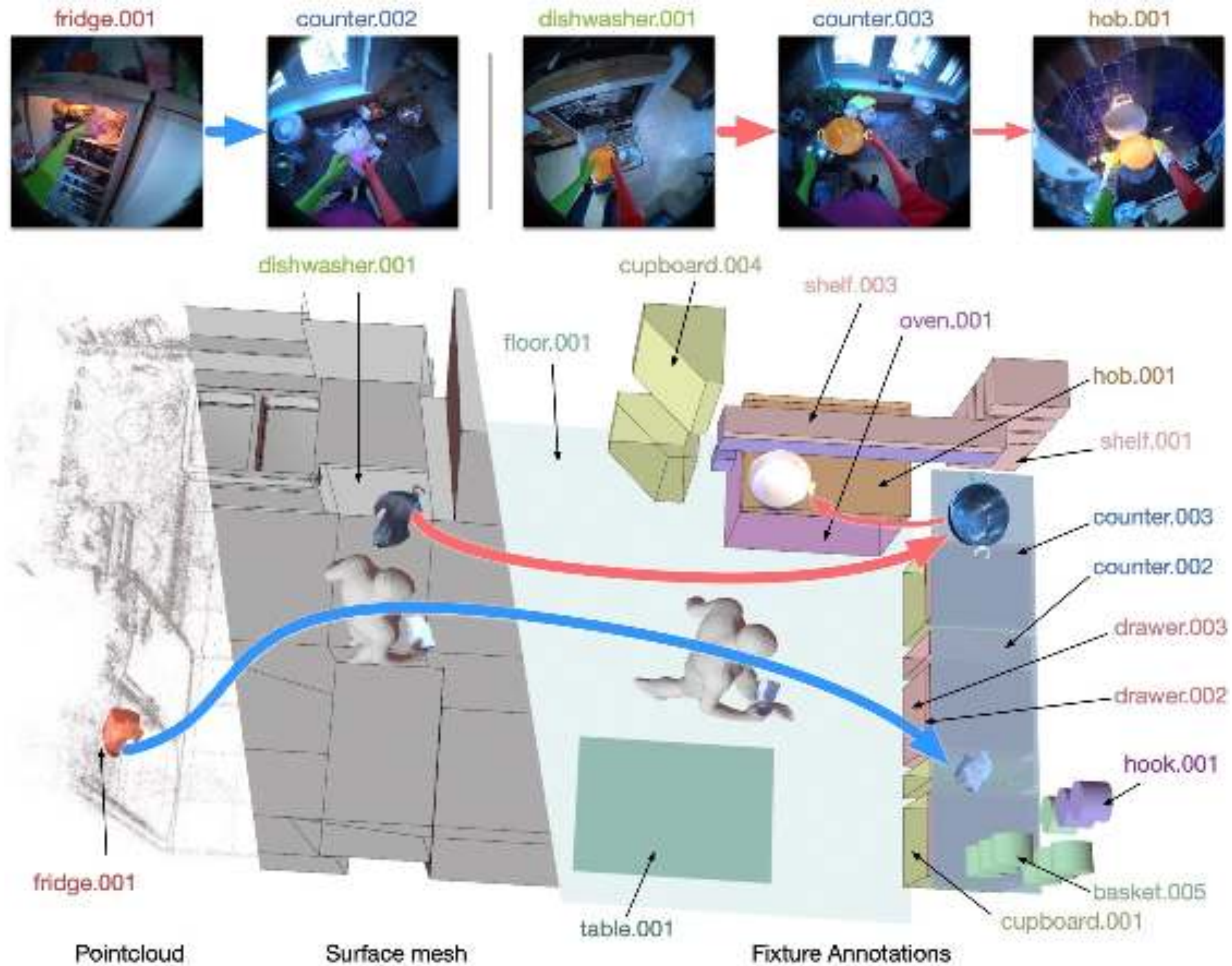
HD-EPIC



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

Digital Twin

HD-EPIC

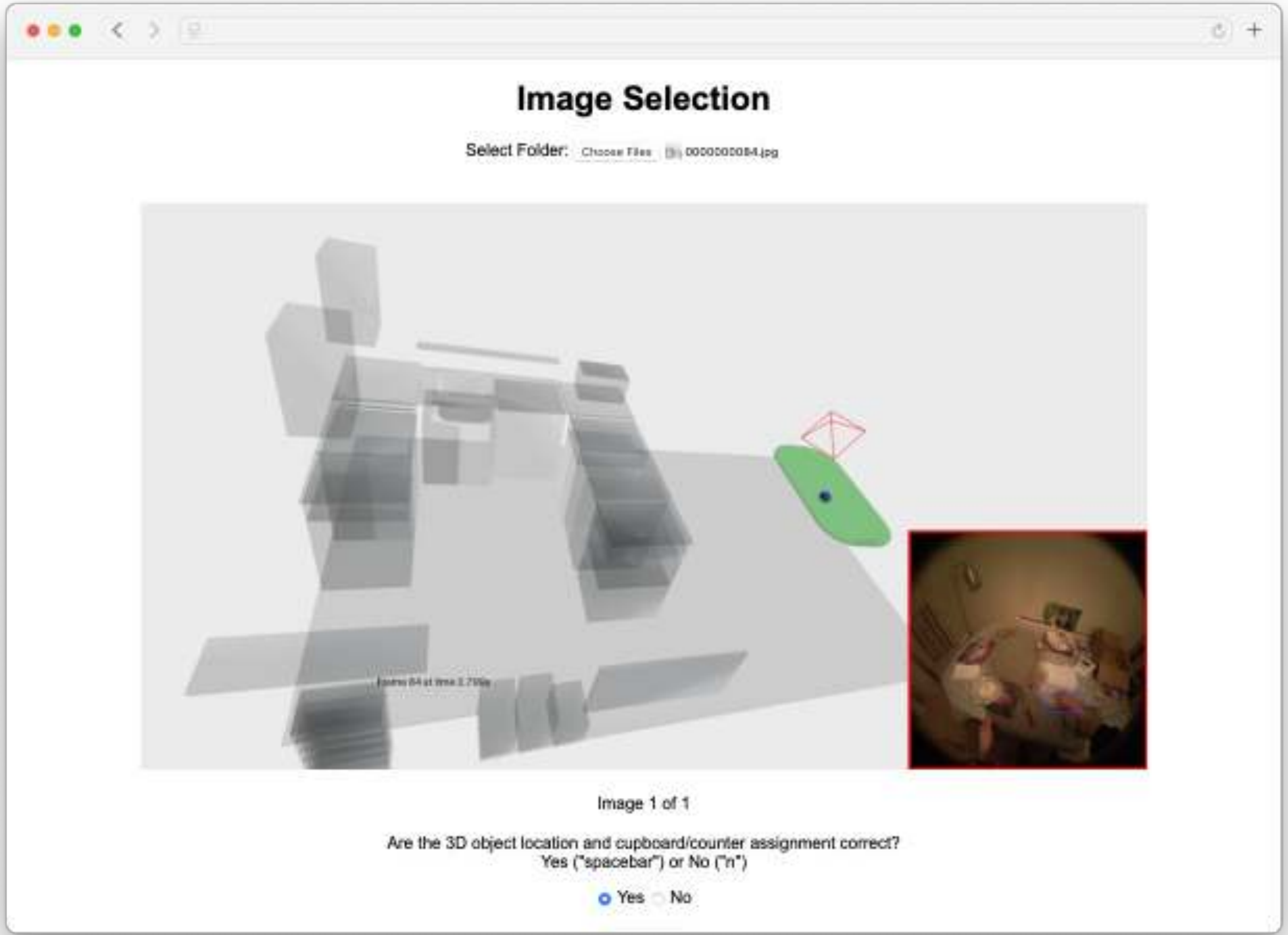


Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

University of BRISTOL

Dima Damen
DataCV W – ICCV2025

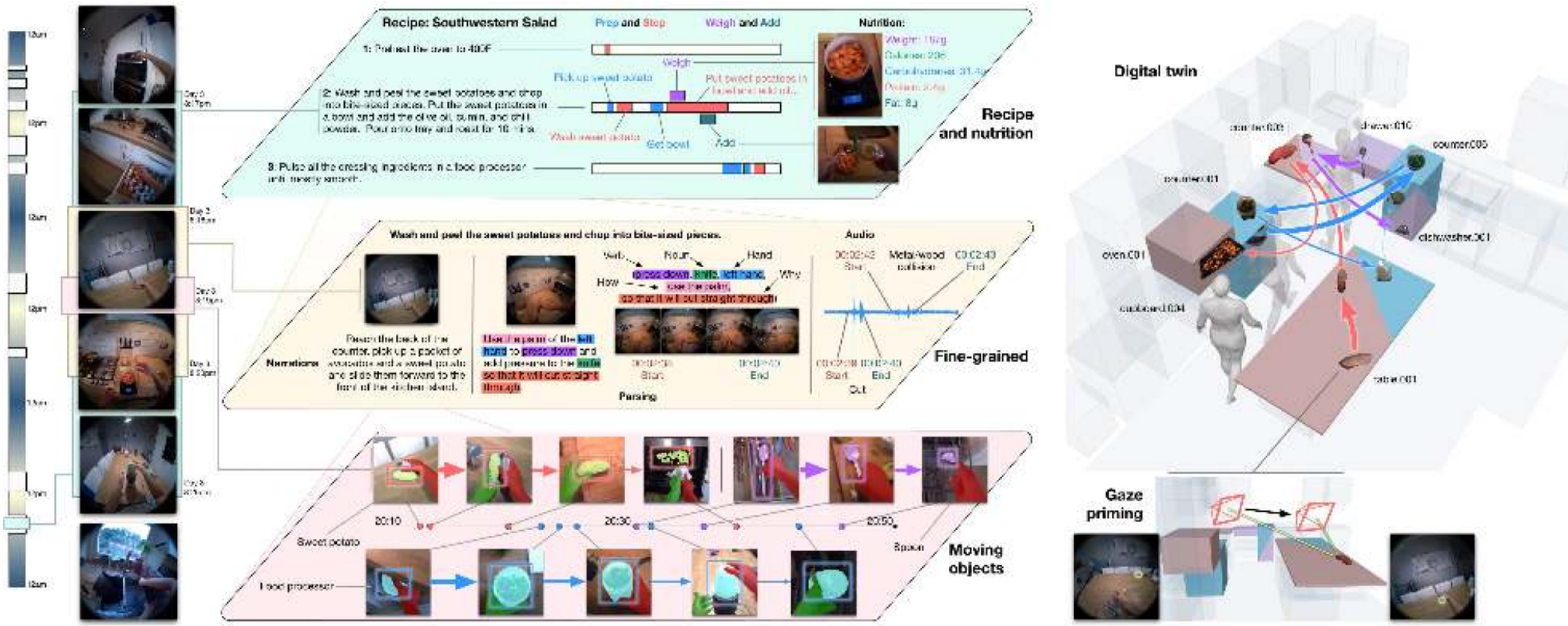
HD-EPIC



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

[redacted]

HD-EPIC



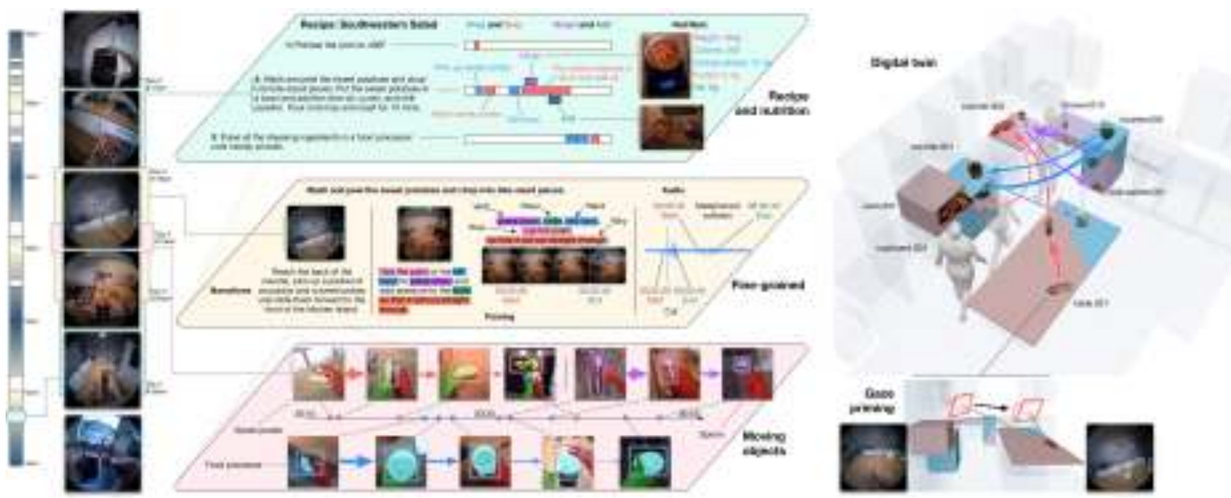
Recipe: Southwestern Salad
Prep and Stop
Weigh and Add
Nutrition:
1: Preheat the oven to 400F
3: Pulse all the dressing ingredients in a food processor until nearly smooth.
Weigh
Add
Get bowl
Recipe and nutrition
Wash and peel the sweet potatoes and chop into bite-sized pieces.
Audio
Verb
Noun
Hand
How
Why
Metal/wood collision
Narrations
Parsing
Start
End
Fine-grained
Sweet potato
Food processor
Spoon
Moving objects
Digital twin
drawer.010
counter.005
oven.001
dishwasher.001
cupboard.034
table.001
Gaze priming

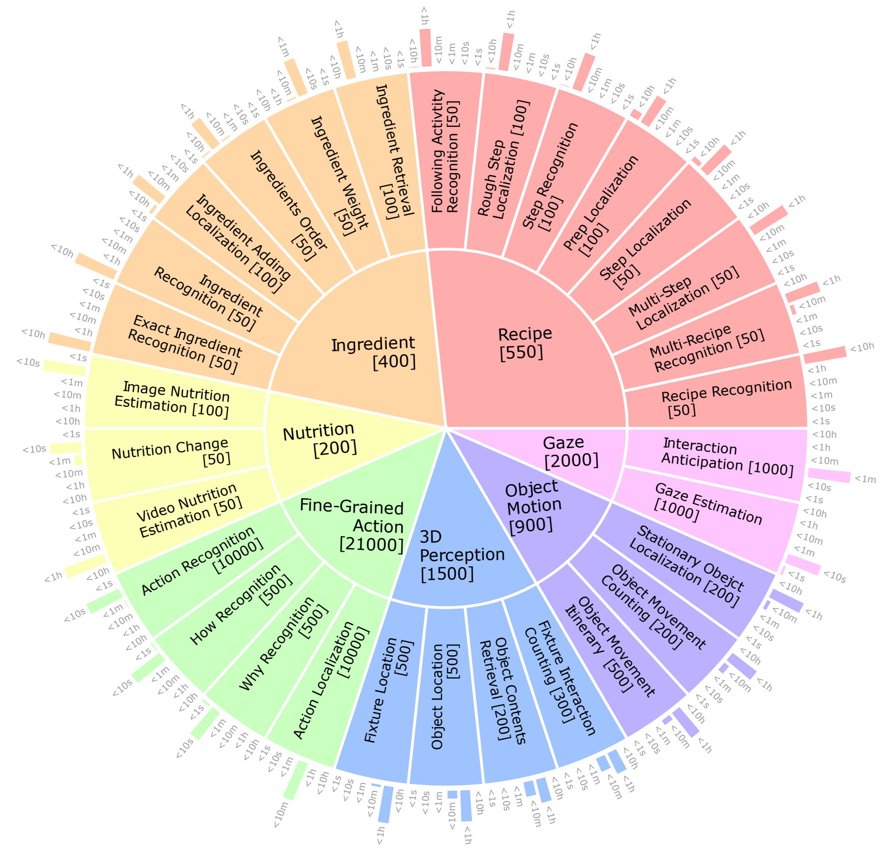
HD-EPIC

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

HD-EPIC





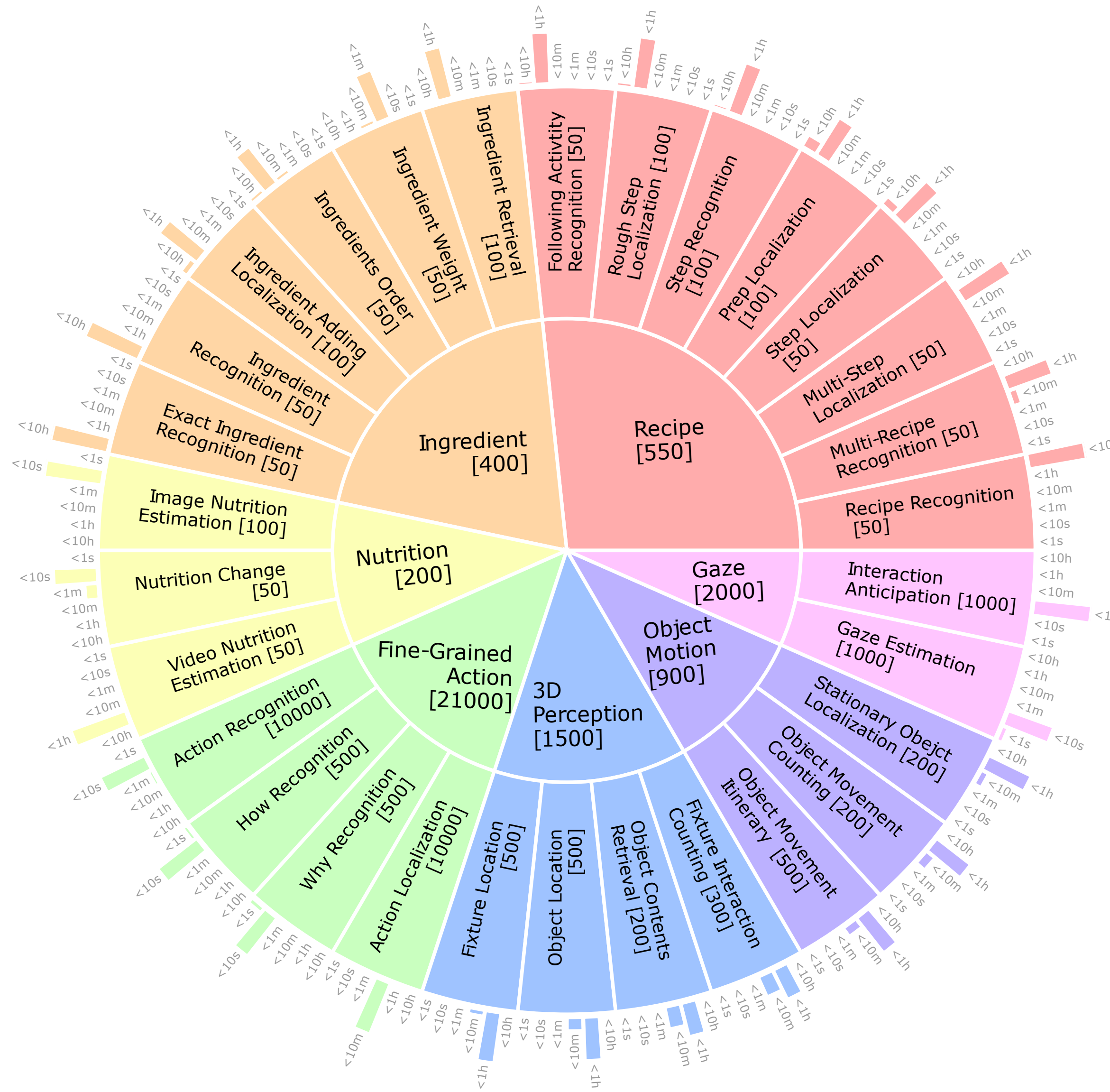
Sec 1: Highly-Detailed Dataset

Sec 2: HD-EPIC VQA Benchmark

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

HD-EPIC

1. Recipe. Questions on temporally localising, retrieving, or recognising recipes and their steps.
2. Ingredient. Questions on the ingredients used, their weight, their adding time and order.
3. Nutrition. Questions on nutrition of ingredients and nutritional changes as ingredients are added to recipes.
4. Fine-grained action. What, how, and why of actions and their temporal localisation.
5. 3D perception. Questions that require the understanding of relative positions of objects in the 3D scene.
6. Object motion. Questions on where, when and how many times objects are moved across long videos.
7. Gaze. Questions on estimating the fixation on large landmarks and anticipating future object interactions.



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

Dima Damen
DataCV W – ICCV2025

HD-EPIC

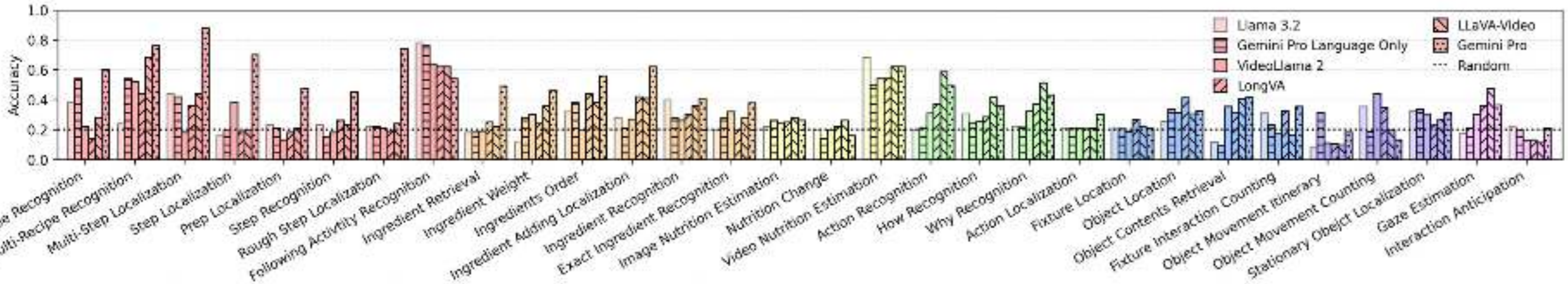
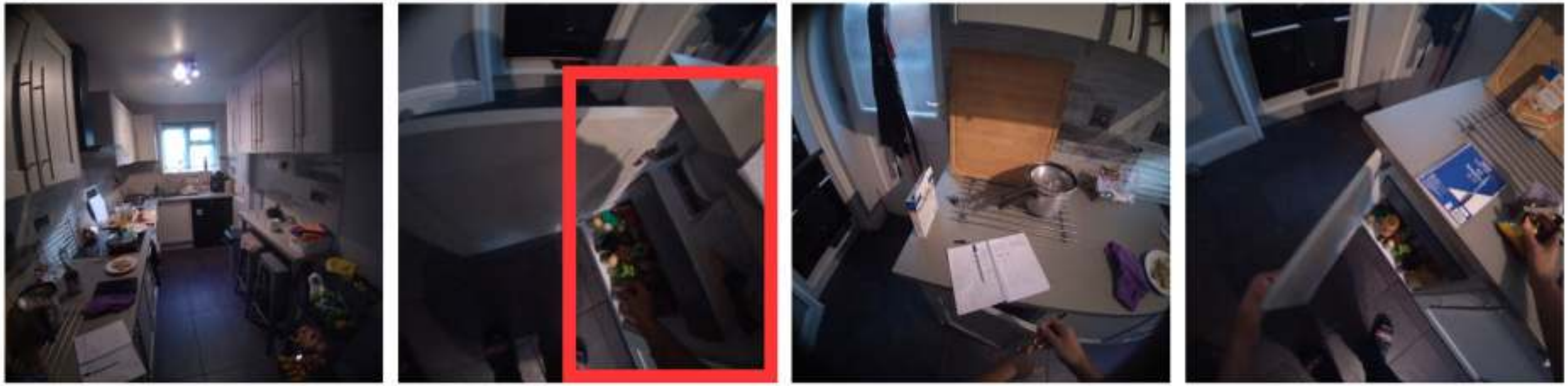


Figure 11. **VQA Results per Question Prototype.** Our benchmark contains many challenging questions for current models.

Model	Recipe	Ingredient	Nutrition	Action	3D	Motion	Gaze	Avg.
Blind - Language Only								
Llama 3.2	33.5	25.0	36.7	23.3	22.3	25.5	19.5	26.5
Gemini Pro	38.0	26.8	30.0	22.1	21.5	27.7	20.5	26.7
Video-Language								
VideoLlama 2	30.8	25.7	32.7	27.2	25.7	28.5	21.2	27.4
LongVA	29.6	30.8	33.7	30.7	32.9	22.7	24.5	29.3
LLaVA-Video	36.3	33.5	38.7	43.0	27.3	18.9	29.3	32.4
Gemini Pro	64.3	48.6	34.7	39.6	32.5	20.8	28.7	38.5
Sample Human Baseline	*96.7*	*96.7*	*85.0*	*92.5*	*93.8*	*92.7*	*75.0*	*90.3*

HD-EPIC



How many times did I open the item at bounding box (165, 452, 1408, 1408) in 00:00:57?

A. 3 **B.** 1 **C.** 4 **D.** 5 **E.** 2

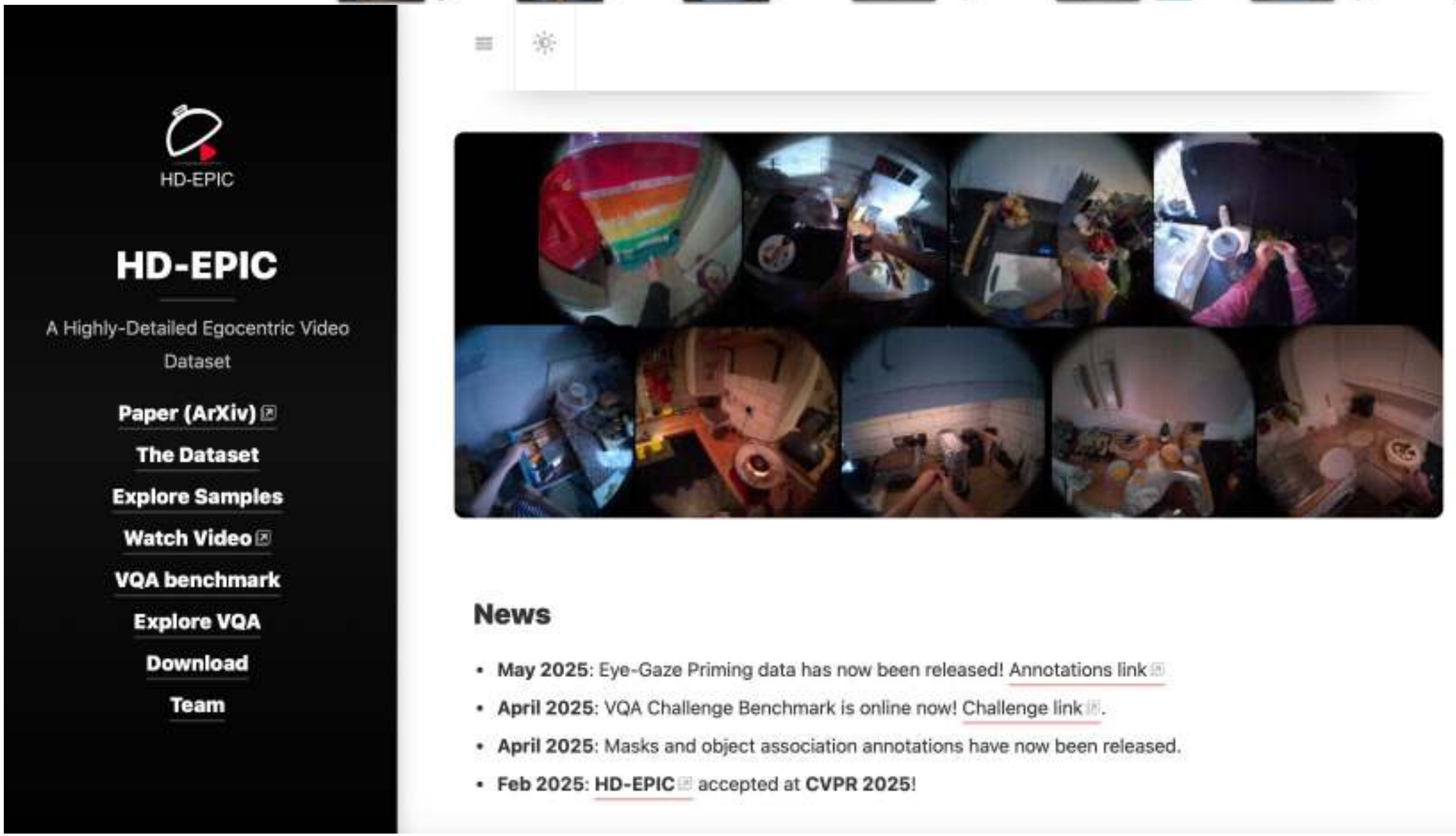
Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

HD-EPIC

University of BRISTOL

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

HD-EPIC



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

HD-EPIC

Try it Yourself

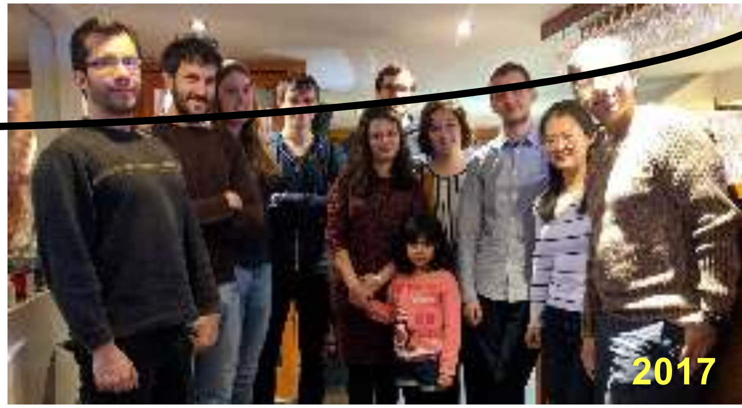
Use Wise to Search through HD-EPIC

SCAN ME

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

My research team…

grateful



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