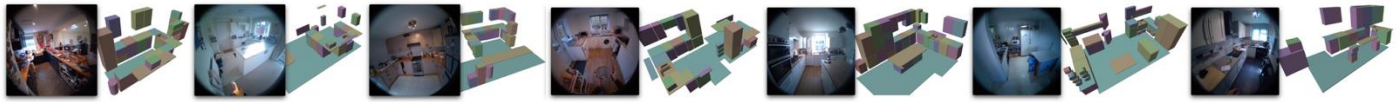
CVPR Nashville JUNE 11-15, 2025



HD-EPIC: A Highly-Detailed Egocentric Video Dataset

HD-EPIC

University of BRISTOL


Toby Perrett


Ahmad Darkhalil


Saptarshi Sinha

Universiteit Leiden


Omar Emara


Sam Pollard


Kranti Parida


Kaiting Liu


Prajwal Gatti


Siddhant Bansal

SINGAPORE MANAGEMENT UNIVERSITY


Kevin Flanagan


Jacob Chalk


Zhifan Zhu


Rhodri Guerrier


Fahd Abdelazim


Bin Zhu

UNIVERSITY OF BATH

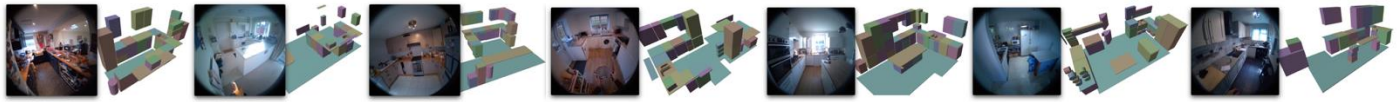

Davide Moltisanti


Michael Wray


Hazel Doughty

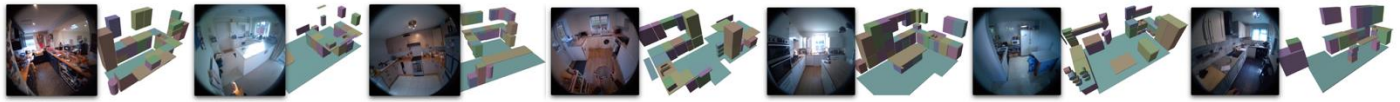

Dima Damen

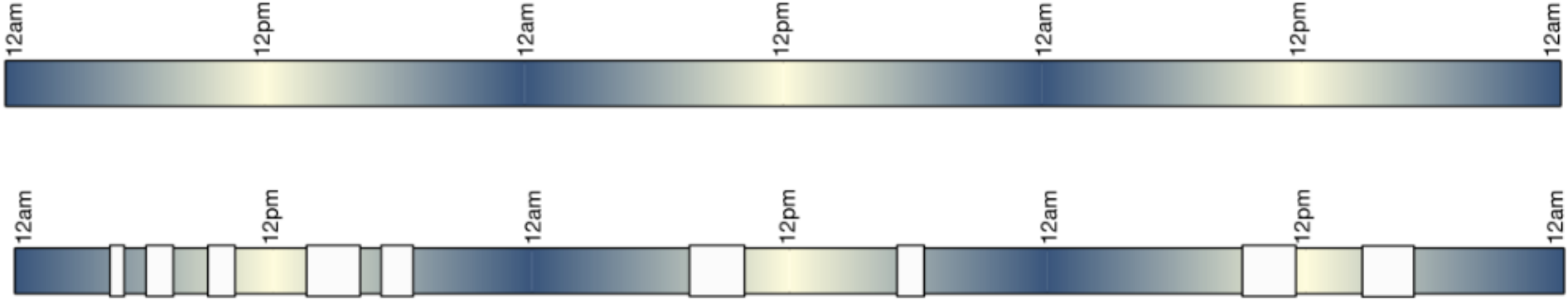
HD-EPIC



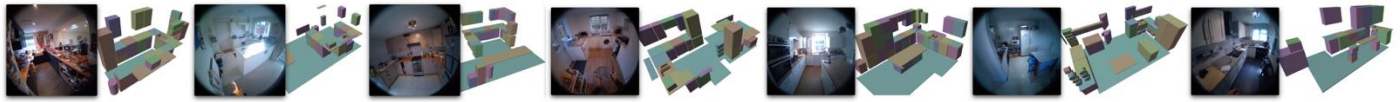
Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

HD-EPIC

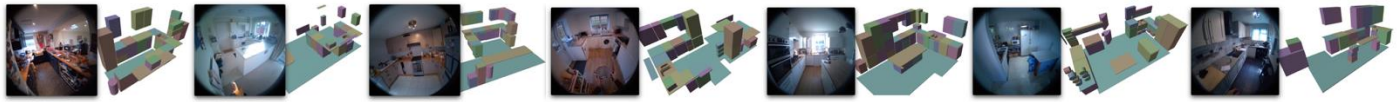



12am
12pm
12am
12pm
12am
12pm
12am
12am
12pm
12am
12pm
12am
12pm
12am

HD-EPIC

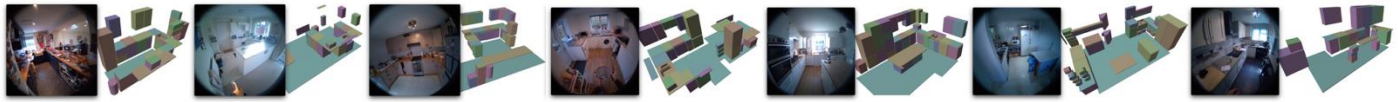


HD-EPIC

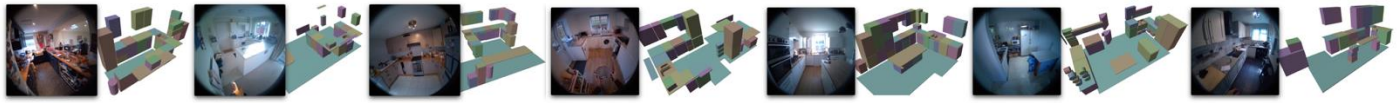


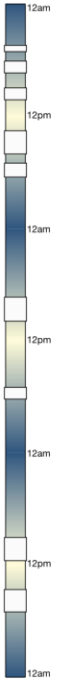
- 9 participants (6 male, 3 female)
- Participants recorded 3.5 to 7.2 hours (avg. 4.6 hours).
- We collected 41.3 hours in 156 videos.
- Long-time commitment (avg. 50 hours): training, data collection, reviewing footage, providing activities/recipe/nutrition and detailed narrations.
- Frames are manually anonymised in RGB as well as SLAM cameras (no easy way to directly map)

HD-EPIC



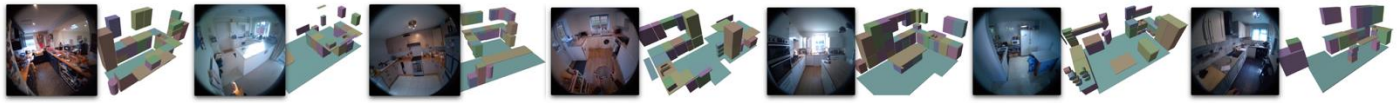
HD-EPIC

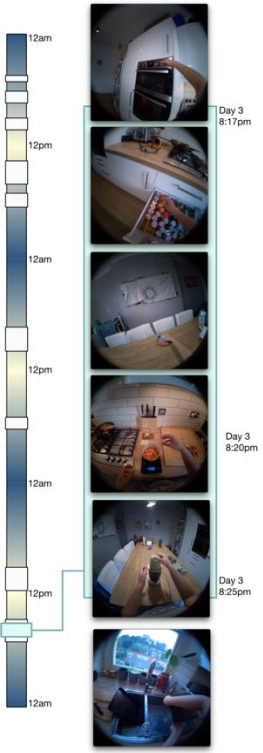




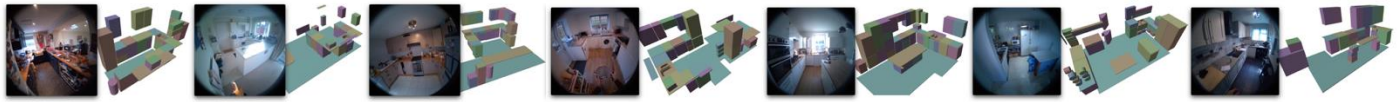
Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

HD-EPIC




12am
12pm
12am
12pm
12am
12pm
12am
Day 3
8:17pm
Day 3
8:20pm
Day 3
8:25pm

HD-EPIC



Recipe: Southwestern Salad

1: Preheat the oven to 400F

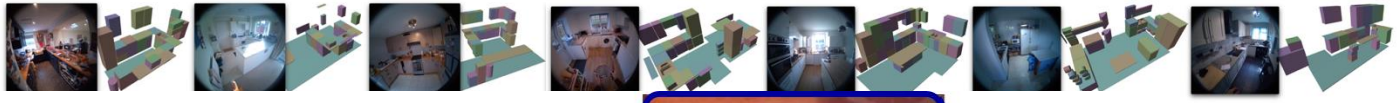
Day 3
:17pm

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



Cacio e Pepe (modified)

Ingredients:

~~200 g~~ penne

~~400g~~ of pasta of your choice

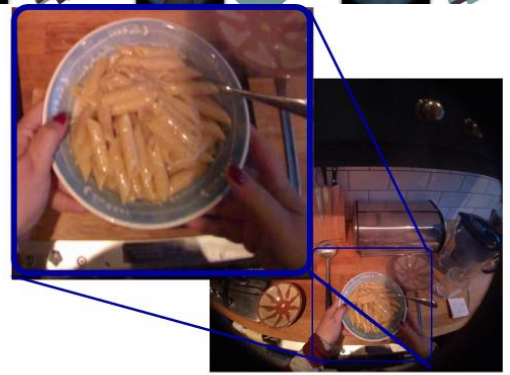
(we recommend bucatini)

~~2~~ 1 tablespoon of black peppercorn

~~30 g~~ parmigiano

200g of freshly grated pecorino cheese

+25g of slightly salted butter



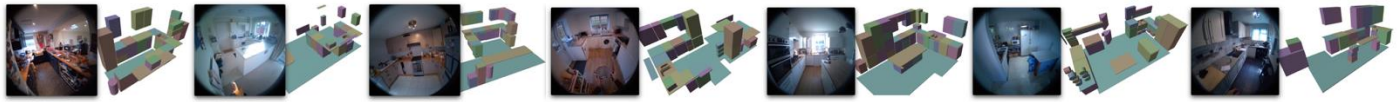
Steps:

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

HD-EPIC



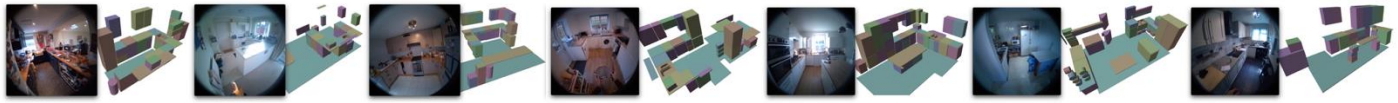
- The **prep** of a corresponding **step** is defined as

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.

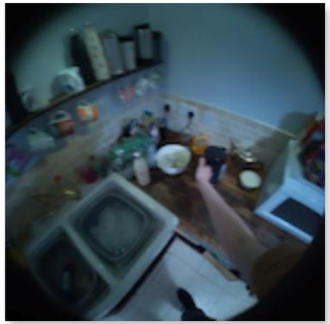
- the **step** ‘add chopped onions and stir’:
  - **Prep:** retrieve tomato from storage, wash tomato, retrieve a knife and chopping board, **and chop the onions.**

HD-EPIC



- Prep

- Step



pick up kettle from its base on the counter with my right hand

pick up packet of bacon

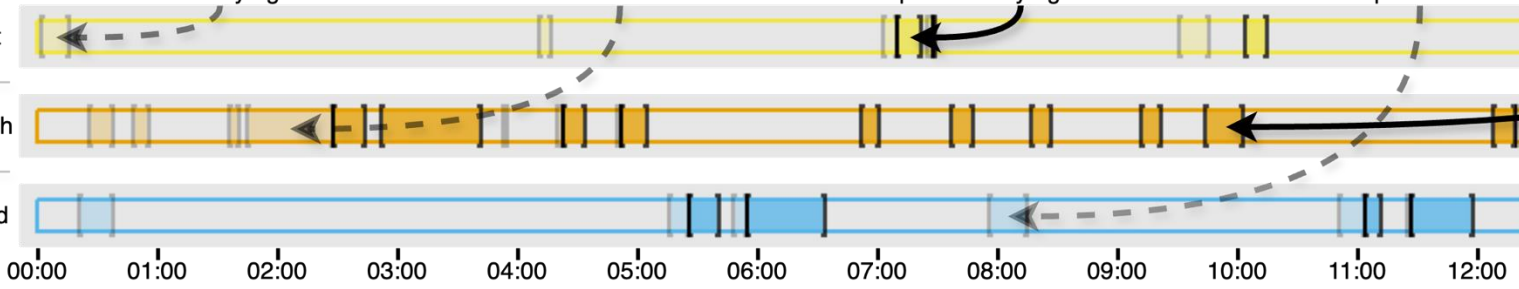
pour water from kettle into the pan with my right hand

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 ...

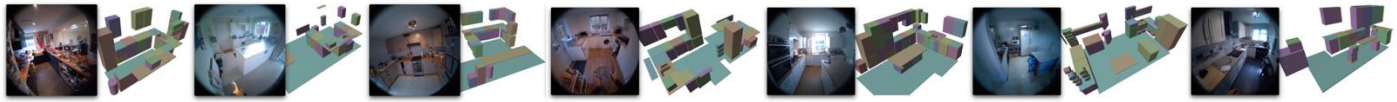
Meanwhile, beat the eggs in a bowl, then finely grate in the Parmesan and mix well.



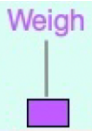
University of BRISTOL

Dima Damen
April 2025

HD-EPIC

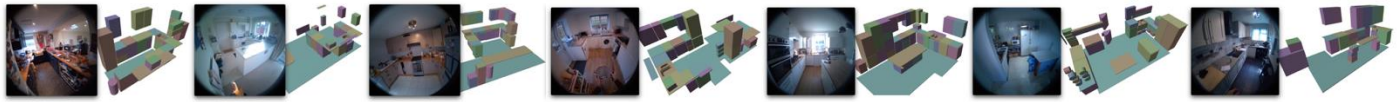



```
"P01_R03_I01": {
    "name": "penne pasta",
     "amount": 125,
     "amount_unit": "g",
     "calories": 445,
     "fat": 1.9,
     "carbs": 90,
     "protein": 15,
```

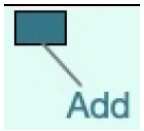



Dima Damen
April 2025

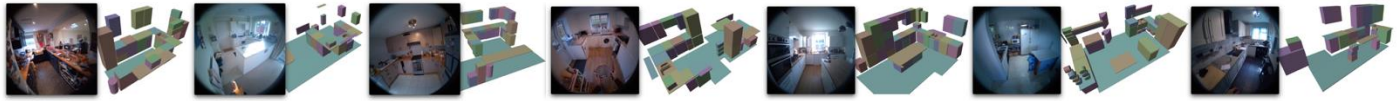
HD-EPIC

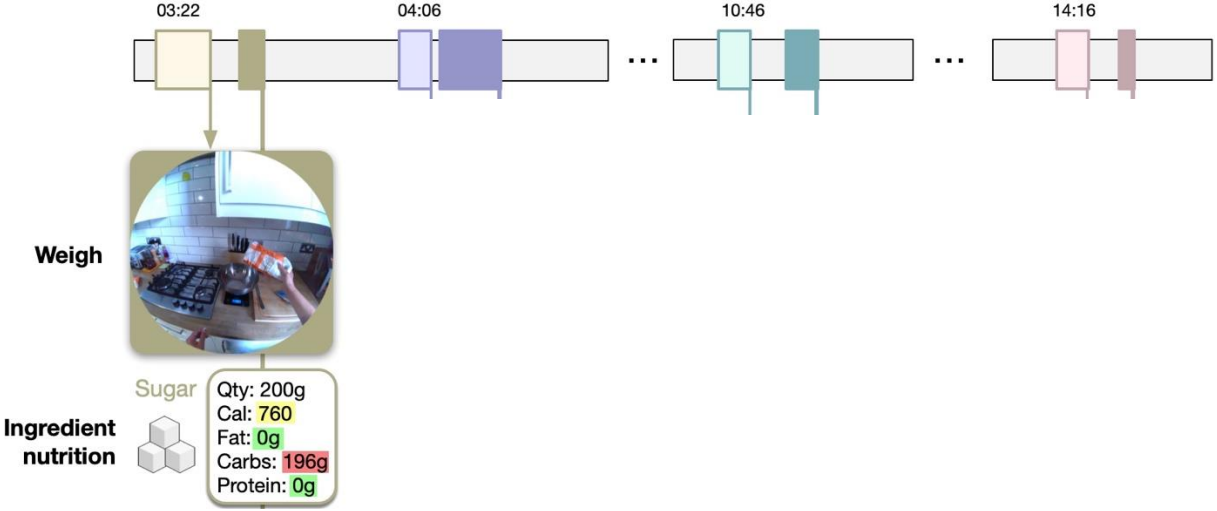



```
"P01_R03_I01": {
    "name": "penne pasta",
     "amount": 125,
     "amount_unit": "g",
     "calories": 445,
     "fat": 1.9,
     "carbs": 90,
     "protein": 15,
```

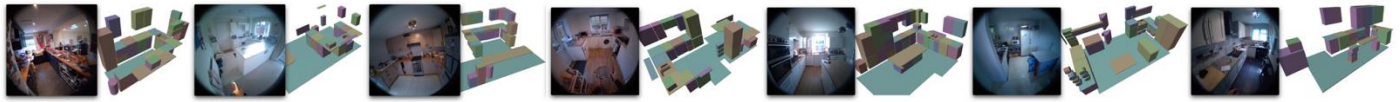



HD-EPIC




03:22
04:06
10:46
14:16
Weigh
Sugar
Ingredient nutrition
Qty: 200g
Cal: 760
Fat: 0g
Carbs: 196g
Protein: 0g

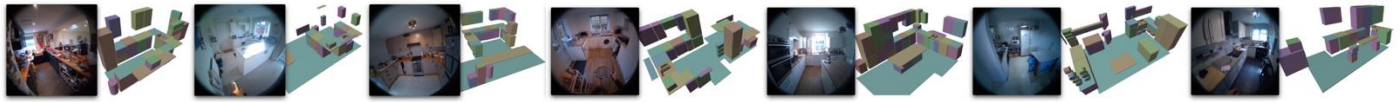
HD-EPIC



- 69 recipes. Avg: 6.6 steps, 8.1 ingredients, 4.8 hours, 2.1 videos.
- Our longest recipe took 2 days and 6 hours to complete.

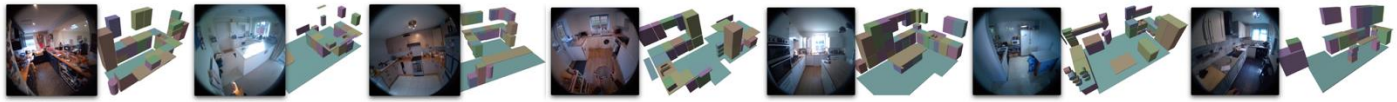
HD-EPIC Videos	Recipe Videos
Multiple recipes / video, Multiple videos / recipe	One video == one recipe
All the action including preparatory and cleaning/clearing	Only the steps (prep edited out)
Faster (no explanations)	Slower (with descriptions)
Ingredients weighed on camera	Ingredients pre-weighed
Usually average skill level	Usually above average skill level

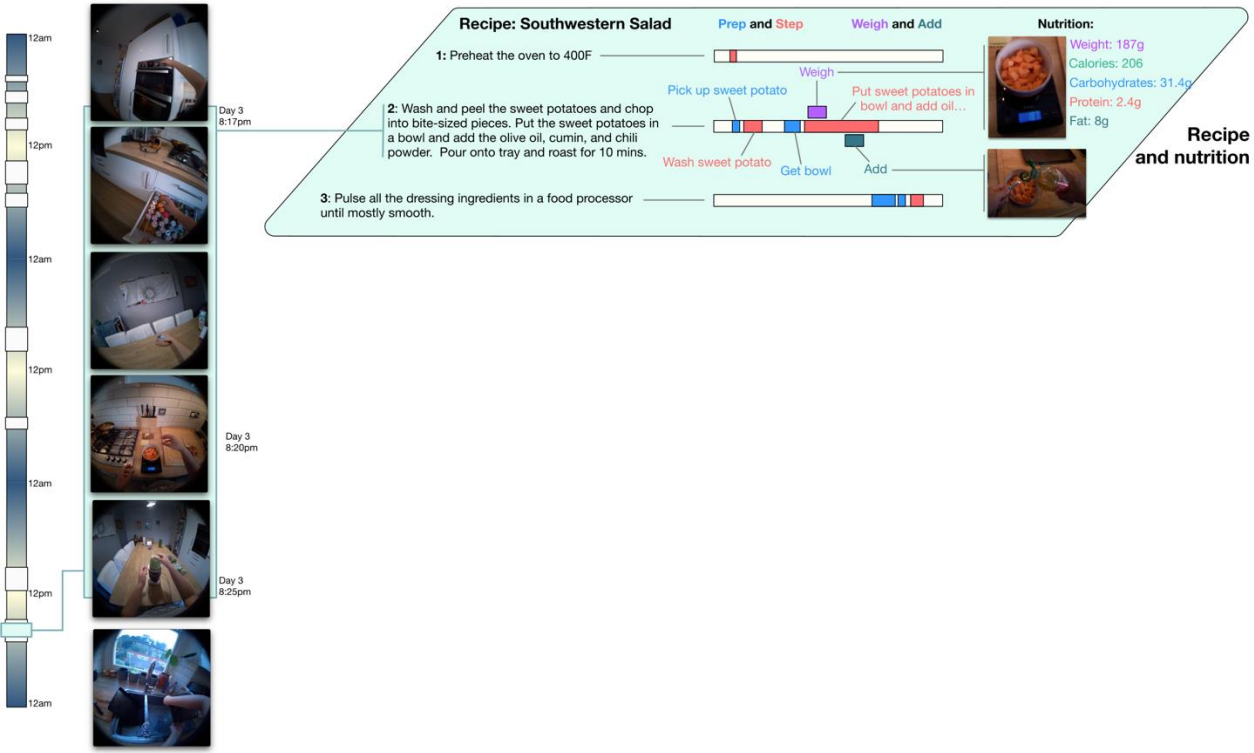
HD-EPIC



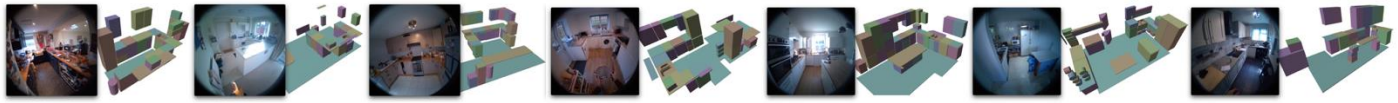

Recipe and Nutrition
Recipe: Lazy Cake
Step 1: Start by mixing the melted butter, sweetened condensed milk with the cocoa powder. Combine them really well together, but it's okay if there are a few small lumps.
Step 2: Taste the mixture and adjust it to your desired sweetness by either adding more condensed milk for more sweetness or more cacao powder. I prefer it with the amounts stated in the recipe.
Step 3: Next you need to crumble up the biscuit into the mixture.
Step 4: It's important to use small sized biscuit crumbs like in the picture, but also try and incorporate really small crumbs in there as well to help the mixture hold.
Step 5: The last step is shaping the mixture. Do this by spreading out some plastic wrap (or cling film) and spooning the mixture in the centre in the shape of a log. Slowly wrap the plastic around the mixture and tighten as you do this to shape into a cylindrical log.
Step 6: Next wrap the log with foil to help it maintain its shape and place it in the fridge for at least 5 hours (depending on your freezer).

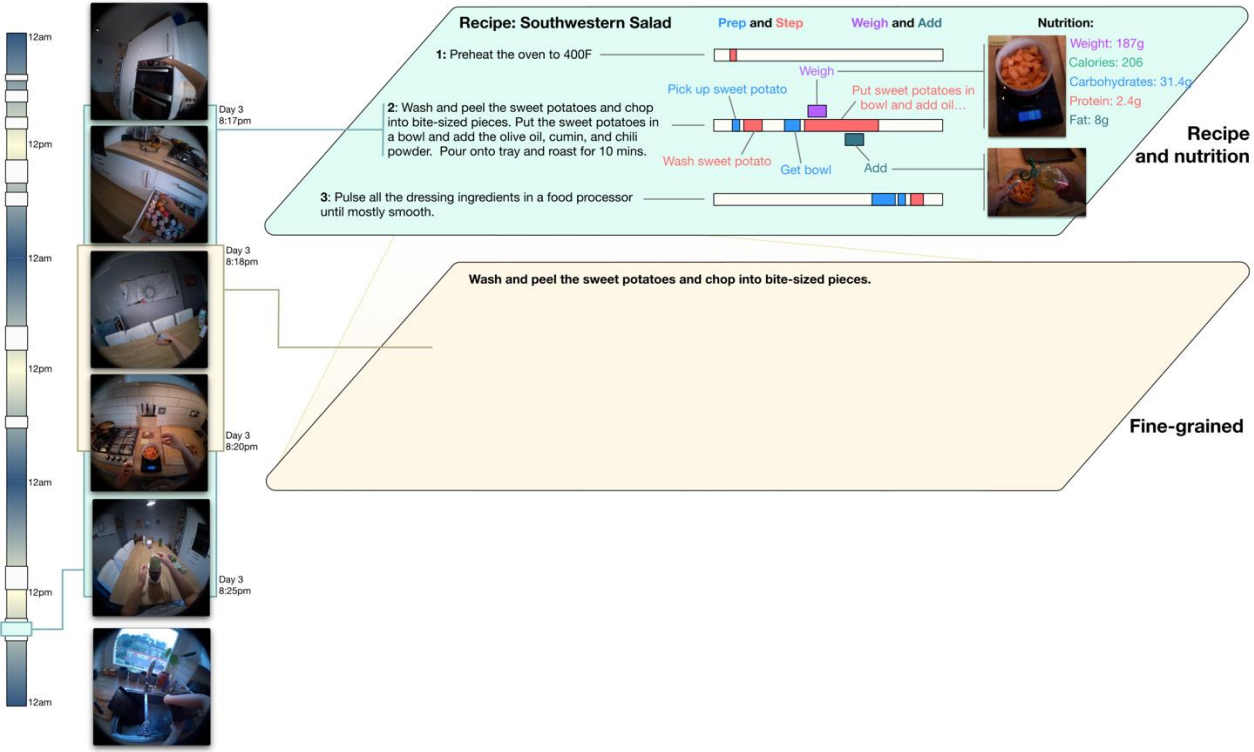
HD-EPIC



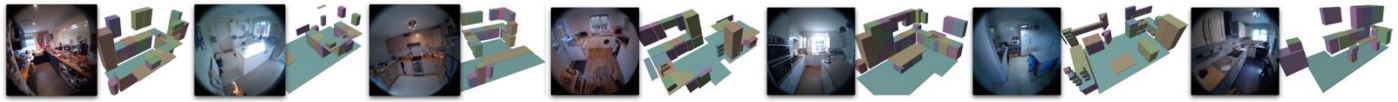

12am
12pm
12am
12pm
12am
12pm
12am
Day 3
8:17pm
Day 3
8:20pm
Day 3
8:25pm
Recipe: Southwestern Salad
Prep and Step
Weigh and Add
Nutrition:
1: Preheat the oven to 400F
2: Wash and peel the sweet potatoes and chop into bite-sized pieces. Put the sweet potatoes in a bowl and add the olive oil, cumin, and chili powder. Pour onto tray and roast for 10 mins.
3: Pulse all the dressing ingredients in a food processor until mostly smooth.
Weigh
Pick up sweet potato
Put sweet potatoes in bowl and add oil…
Wash sweet potato
Get bowl
Add
Weight: 187g
Calories: 206
Carbohydrates: 31.4g
Protein: 2.4g
Fat: 8g
Recipe and nutrition

HD-EPIC




12am
12pm
12am
12pm
12am
12pm
12am
Day 3 8:17pm
Day 3 8:18pm
Day 3 8:20pm
Day 3 8:25pm
Recipe: Southwestern Salad
Prep and Step
Weigh and Add
Nutrition:
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.
Weigh
Pick up sweet potato
Put sweet potatoes in bowl and add oil…
Wash sweet potato
Get bowl
Add
Weight: 187g
Calories: 206
Carbohydrates: 31.4g
Protein: 2.4g
Fat: 8g
Recipe and nutrition
Wash and peel the sweet potatoes and chop into bite-sized pieces.
Fine-grained

HD-EPIC



Wash and peel the sweet potatoes and chop into bite-sized pieces.

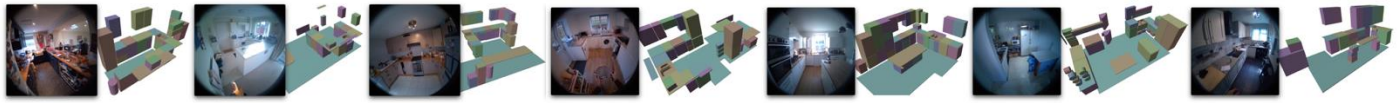
Narrations

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.

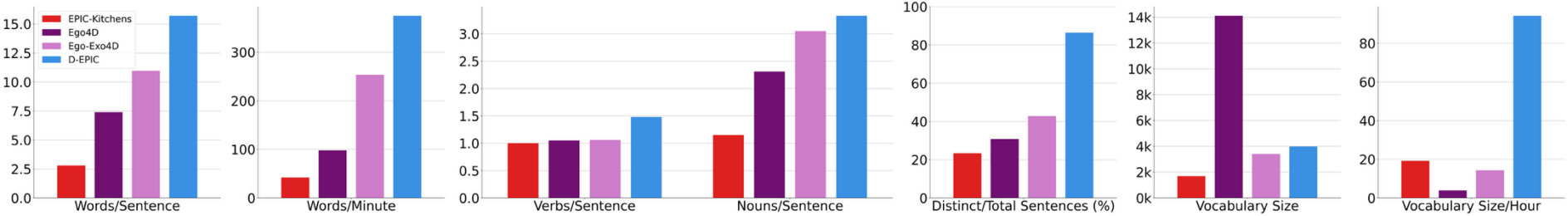
Fine-grained

Highly-Detailed Narrations

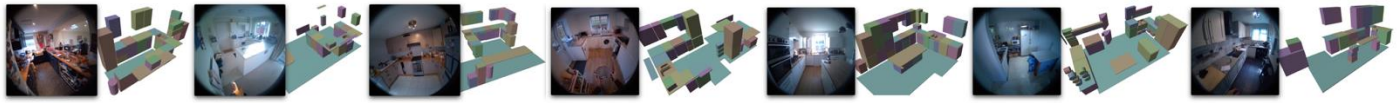
HD-EPIC



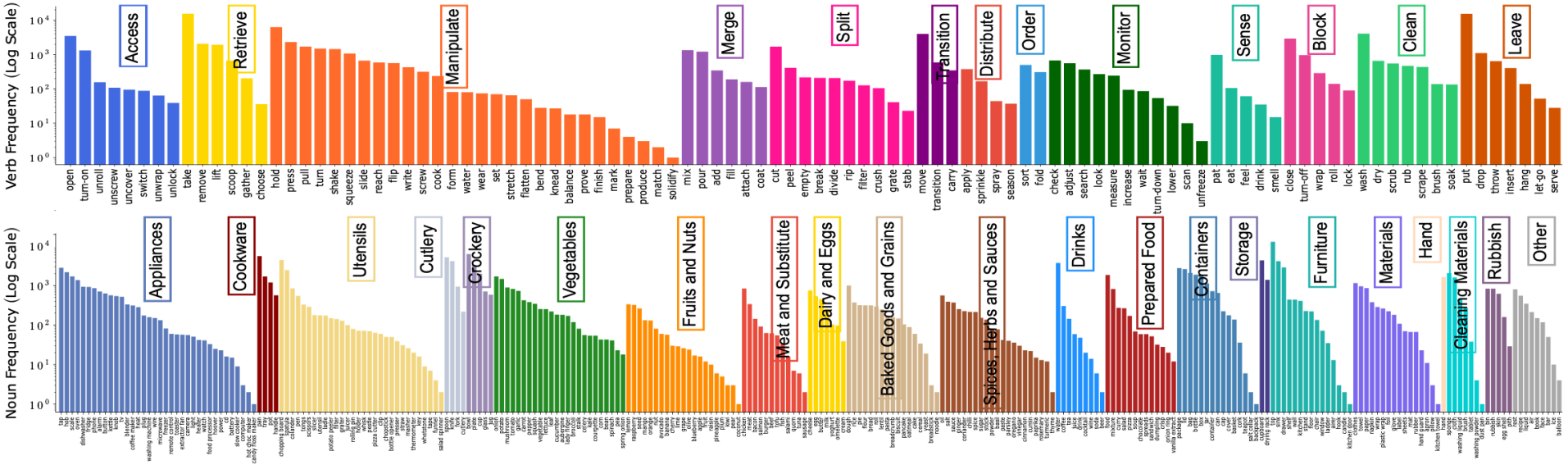
- 59,454 fine-grained actions, with a mean duration of 2.0s (±3.4s).



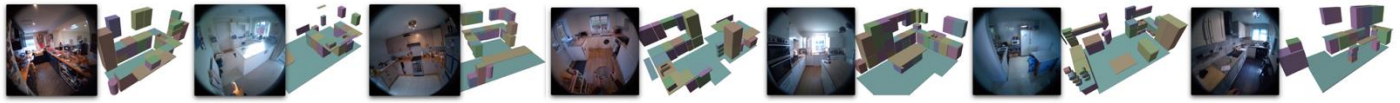
HD-EPIC

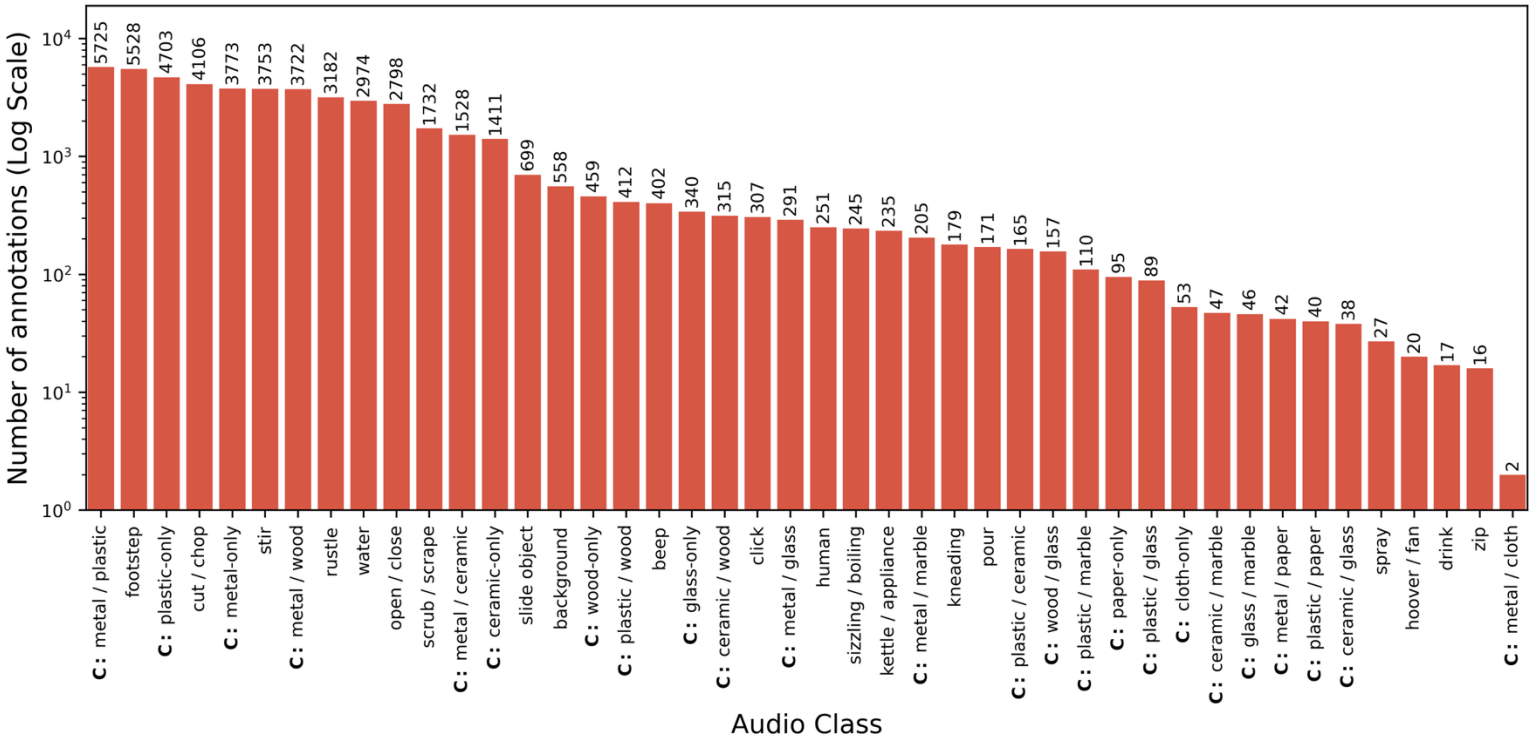


- 59,454 fine-grained actions, with a mean duration of 2.0s (±3.4s).

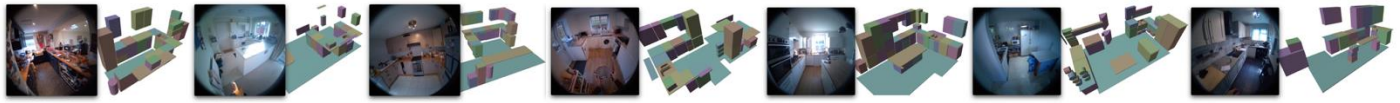


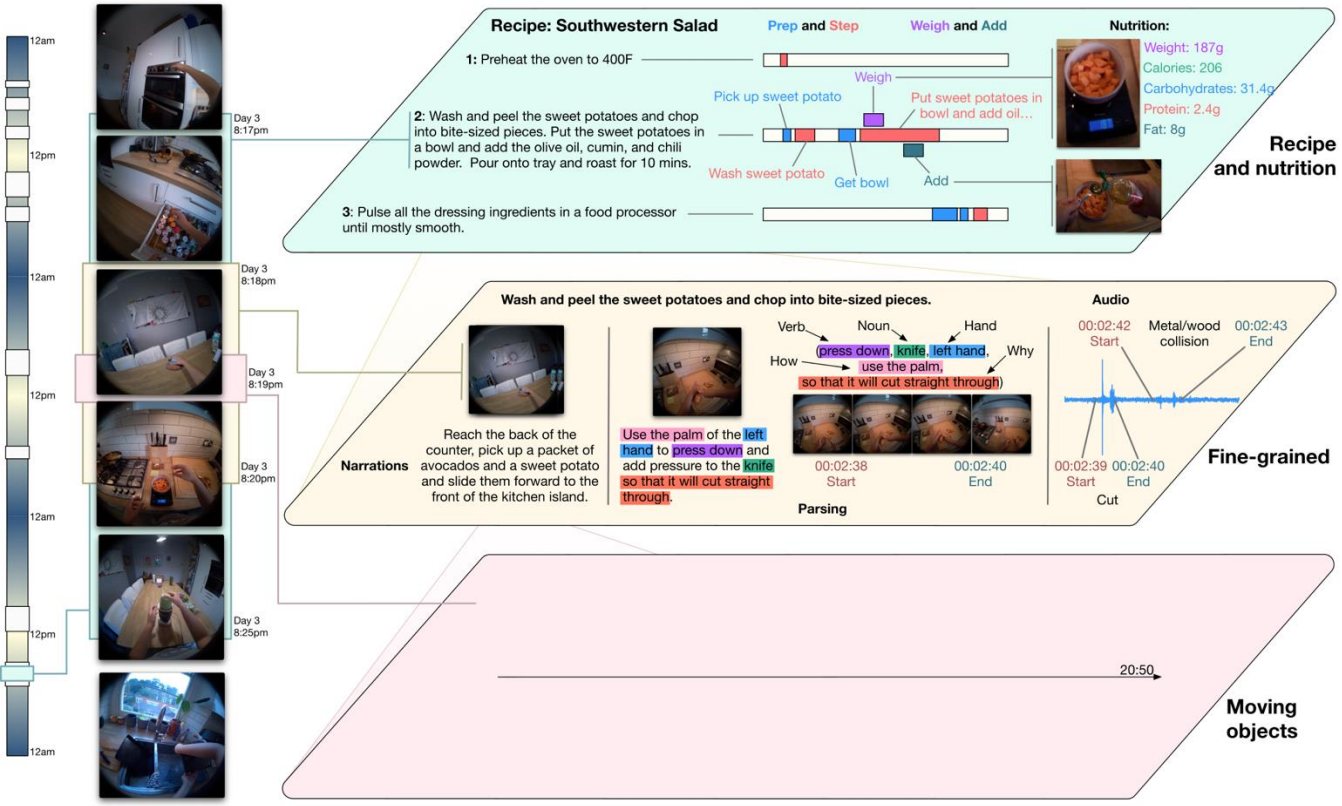
HD-EPIC



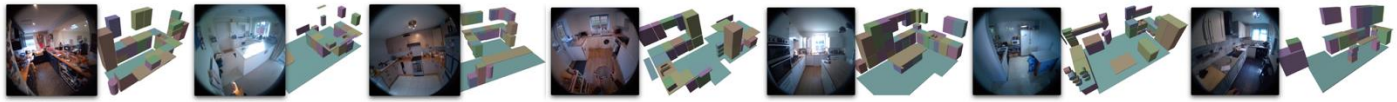

Number of annotations (Log Scale)
10^0
10^1
10^2
10^3
10^4
5725
5528
4703
4106
3773
3753
3722
3182
2974
2798
1732
1528
1411
699
558
459
412
402
340
315
307
291
251
245
235
205
179
171
165
157
110
95
89
53
47
46
42
40
38
27
20
17
16
2
C: metal / plastic
footstep
C: plastic-only
cut / chop
C: metal-only
stir
C: metal / wood
rustle
water
open / close
scrub / scrape
C: metal / ceramic
C: ceramic-only
slide object
background
C: wood-only
C: plastic / wood
beep
C: glass-only
C: ceramic / wood
click
C: metal / glass
human
sizzling / boiling
kettle / appliance
C: metal / marble
kneading
pour
C: plastic / ceramic
C: wood / glass
C: plastic / marble
C: paper-only
C: plastic / glass
C: cloth-only
C: ceramic / marble
C: glass / marble
C: metal / paper
C: plastic / paper
C: ceramic / glass
spray
hoover / fan
drink
zip
C: metal / cloth
Audio Class

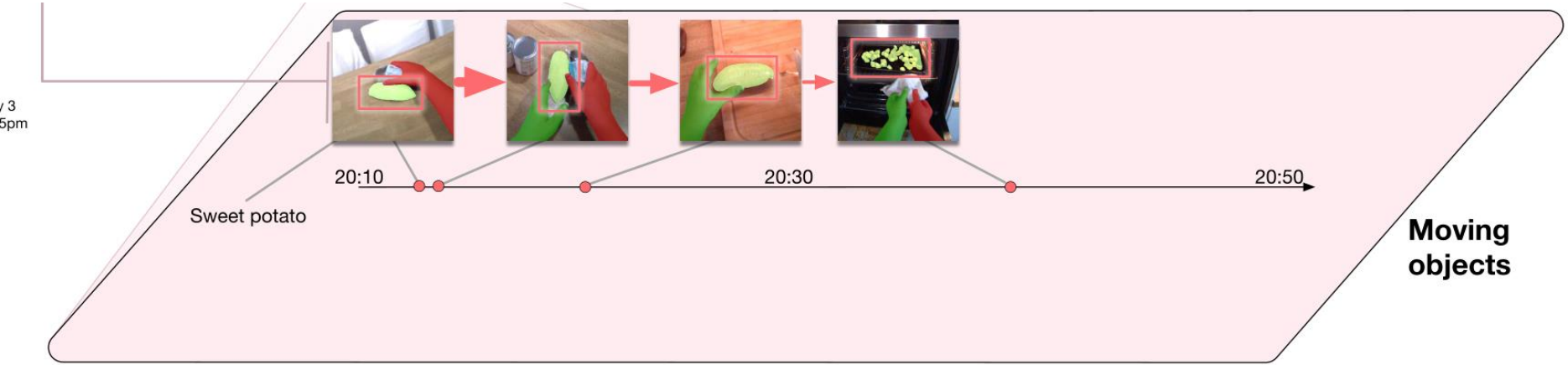
HD-EPIC



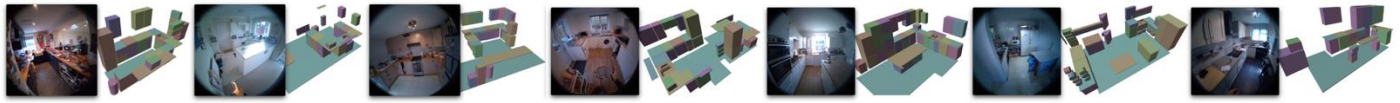

12am
12pm
12am
12pm
12am
12pm
12am
Day 3 8:17pm
Day 3 8:18pm
Day 3 8:19pm
Day 3 8:20pm
Day 3 8:25pm
Recipe: Southwestern Salad
Prep and Step
Weigh and Add
Nutrition:
1: Preheat the oven to 400F
2: Wash and peel the sweet potatoes and chop into bite-sized pieces. Put the sweet potatoes in a bowl and add 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.
Weigh
Pick up sweet potato
Put sweet potatoes in bowl and add oil…
Wash sweet potato
Get bowl
Add
Weight: 187g
Calories: 206
Carbohydrates: 31.4g
Protein: 2.4g
Fat: 8g
Recipe and nutrition
Wash and peel the sweet potatoes and chop into bite-sized pieces.
Audio
Verb
Noun
Hand
How
Why
(press down, knife, left hand, use the palm, so that it will cut straight through)
00:02:42 Start
Metal/wood collision
00:02:43 End
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.
Use the palm of the left hand to press down and add pressure to the knife so that it will cut straight through.
00:02:38 Start
00:02:40 End
Parsing
00:02:39 Start
00:02:40 End
Cut
Fine-grained
20:50
Moving objects

HD-EPIC

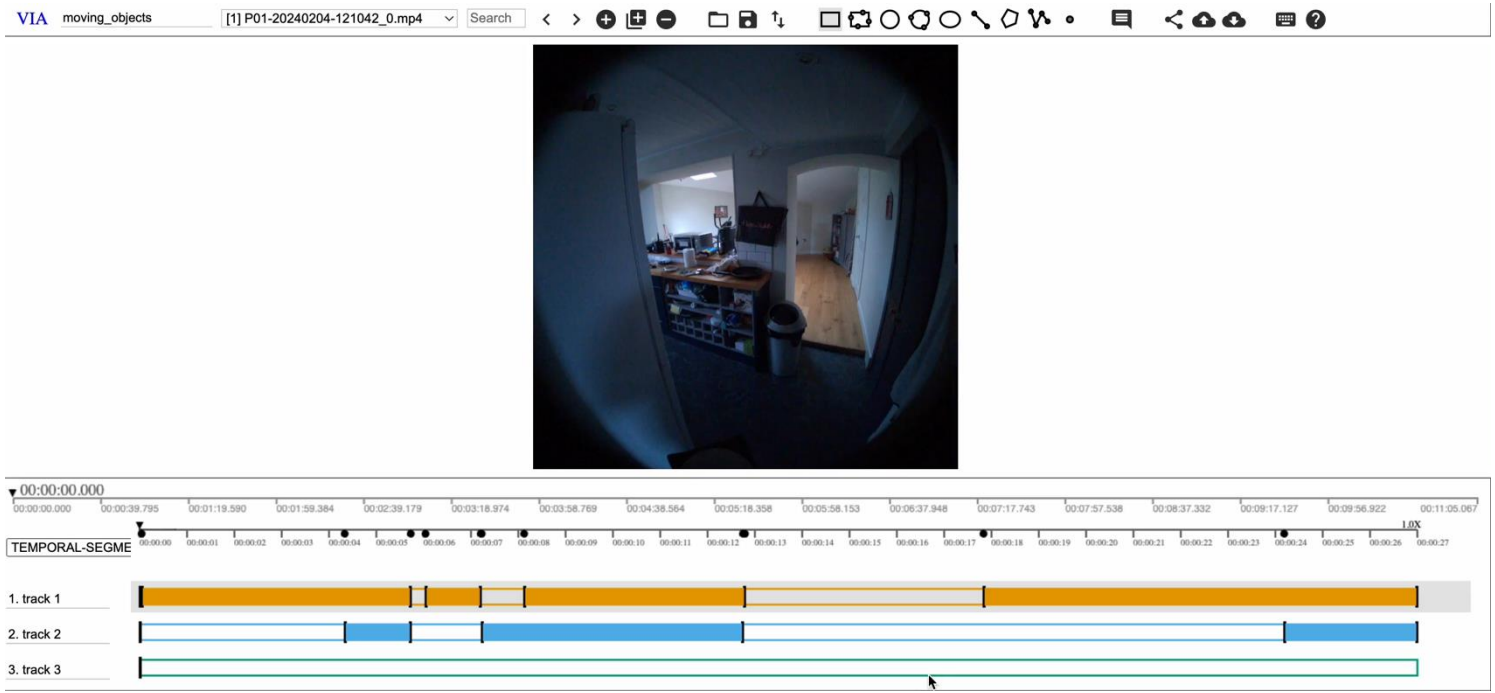




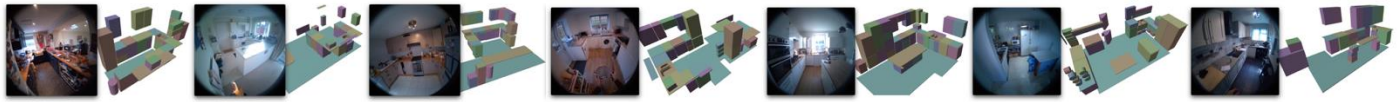
HD-EPIC



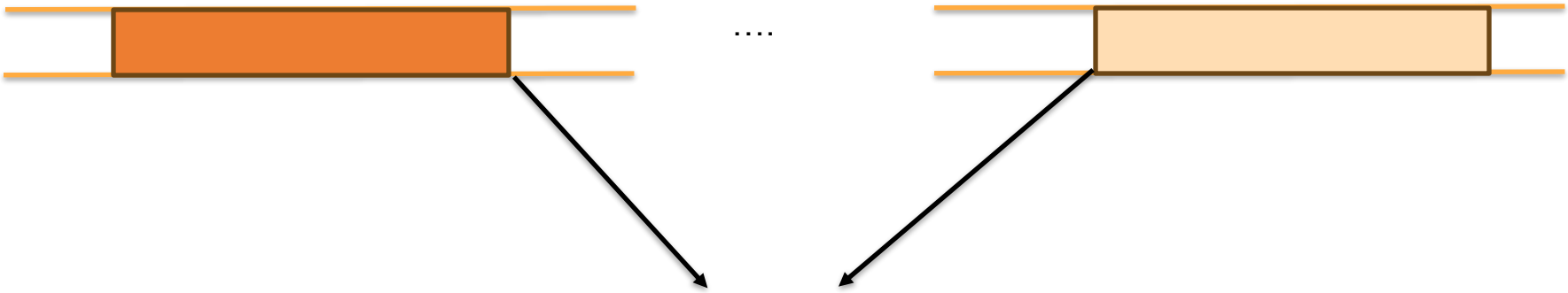
- How to minimize the annotations for tracking objects…



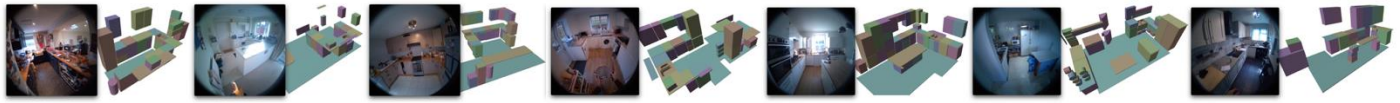
HD-EPIC



- How to minimize the annotations for tracking objects…

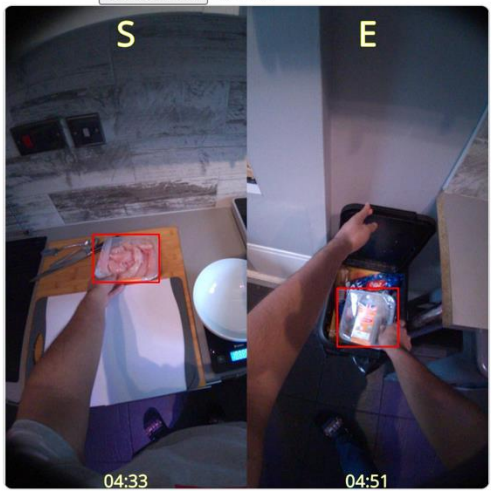


HD-EPIC



Query Image

Choose Files 201 files



42 / 199

← Previous | Next → | Undo

▼ rubbish bin | box of chicken | wooden chopping board

Enter Track Name (optional)

Create New Track

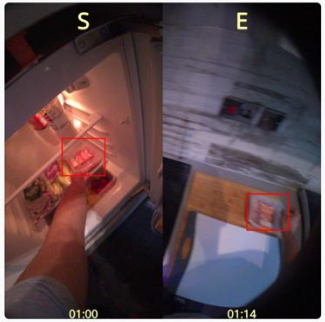
Inconsistent Query

Object Tracks

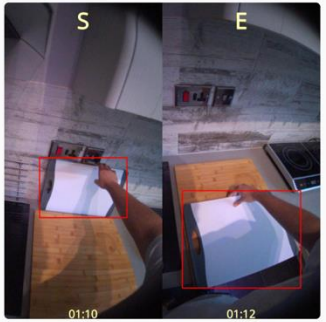
Sort by Distance

Save Tracks

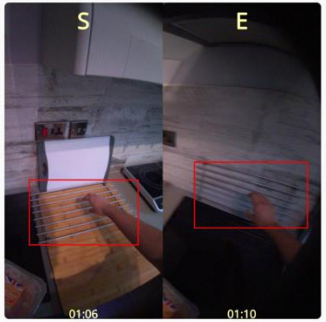
box of chicken (0.0m) Add



plastic chopping board (0.3m) Add

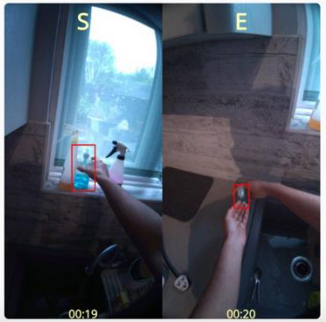


metal cooling rack (0.6m) Add

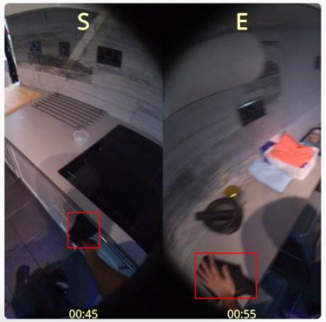


plastic measuring cup (1.0m) Add

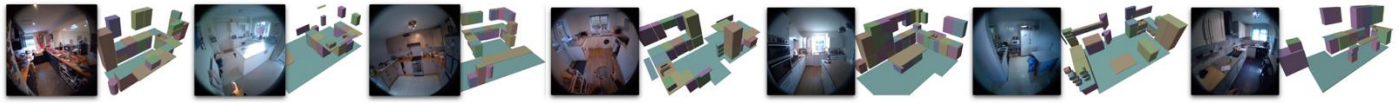
hand washing liquid (1.3m) Add

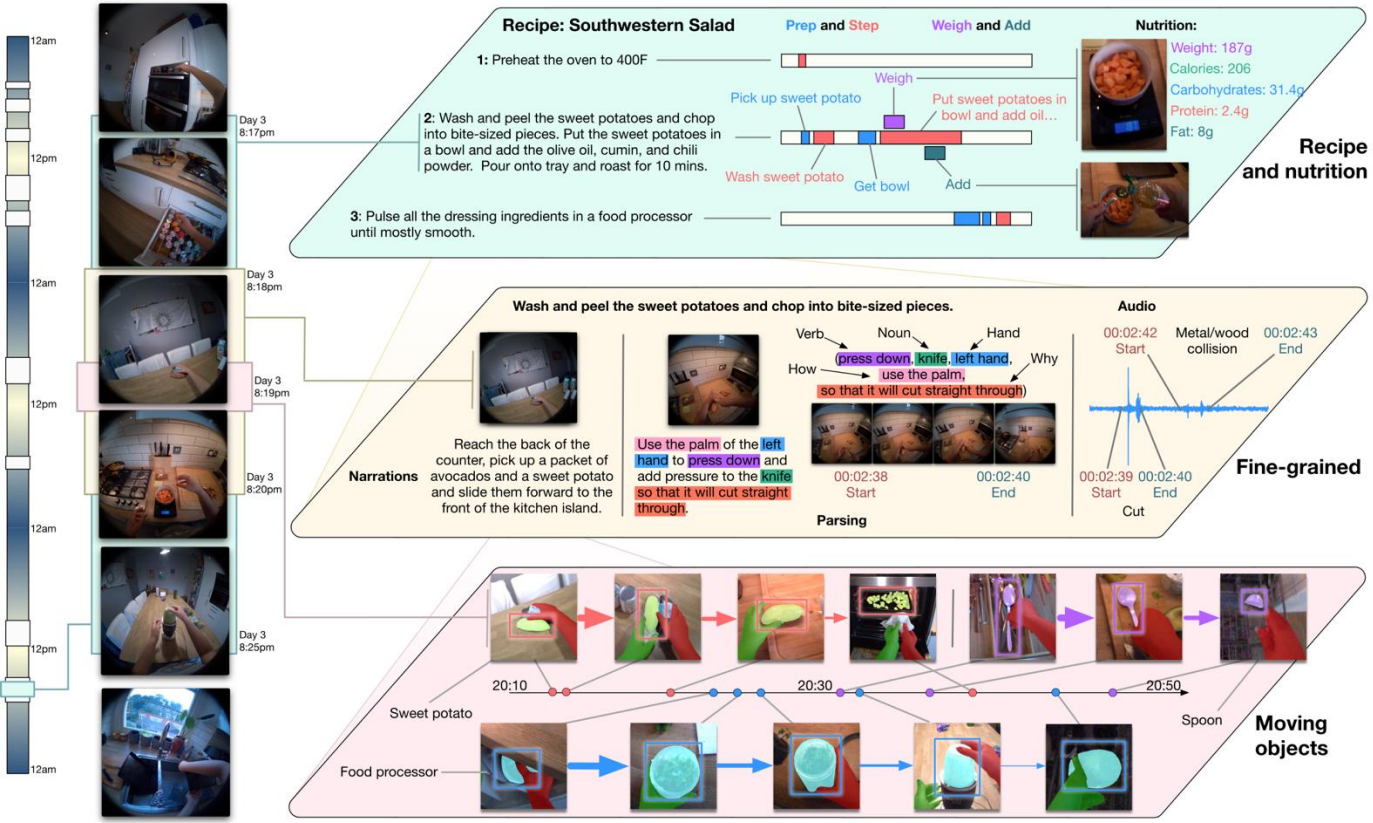


kitchen towel (1.5m) Add

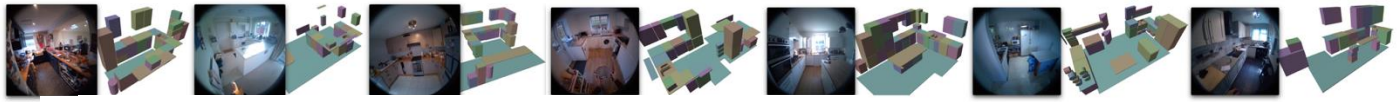


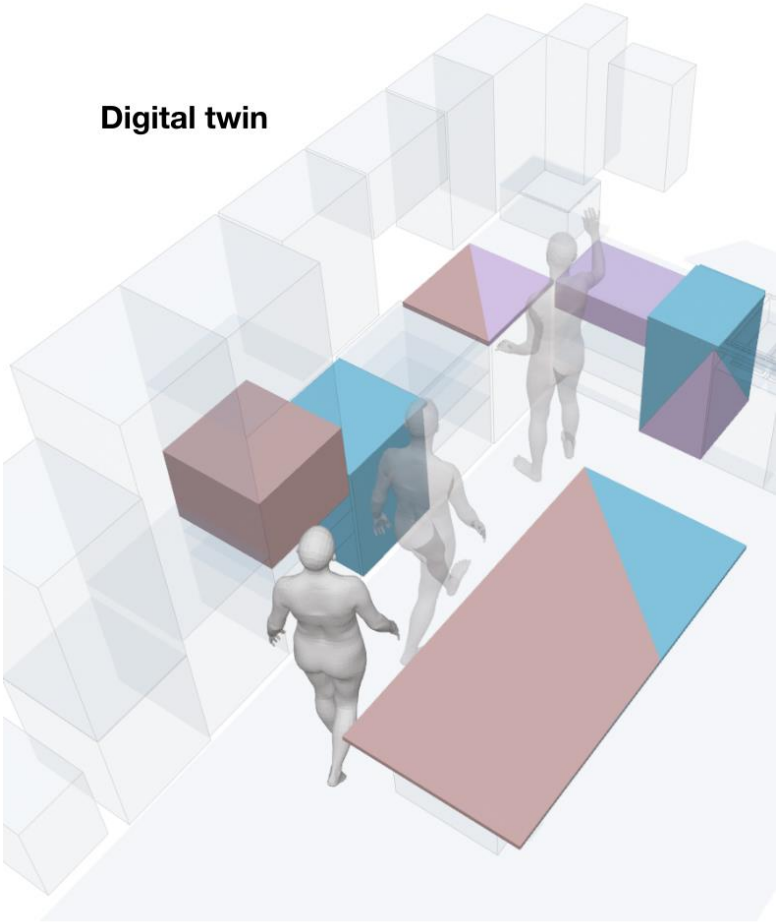
HD-EPIC



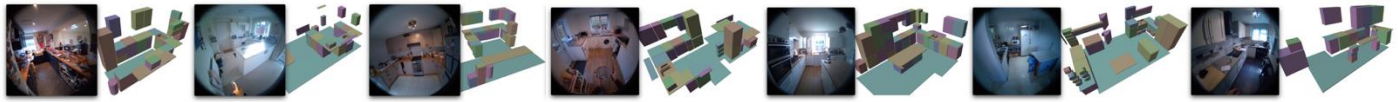

12am
12pm
12am
12pm
12am
12pm
12am
Day 3 8:17pm
Day 3 8:18pm
Day 3 8:19pm
Day 3 8:20pm
Day 3 8:25pm
Recipe: Southwestern Salad
Prep and Step
Weigh and Add
Nutrition:
1: Preheat the oven to 400F
2: Wash and peel the sweet potatoes and chop into bite-sized pieces. Put the sweet potatoes in a bowl and add 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.
Weigh
Pick up sweet potato
Put sweet potatoes in bowl and add oil…
Wash sweet potato
Get bowl
Add
Weight: 187g
Calories: 206
Carbohydrates: 31.4g
Protein: 2.4g
Fat: 8g
Recipe and nutrition
Wash and peel the sweet potatoes and chop into bite-sized pieces.
Audio
Verb
Noun
Hand
(press down, knife, left hand,
How
use the palm,
Why
so that it will cut straight through)
00:02:42 Start
Metal/wood collision
00:02:43 End
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.
Use the palm of the left hand to press down and add pressure to the knife so that it will cut straight through.
00:02:38 Start
00:02:40 End
Parsing
00:02:39 Start
00:02:40 End
Cut
Fine-grained
20:10
20:30
20:50
Sweet potato
Spoon
Food processor
Moving objects

HD-EPIC





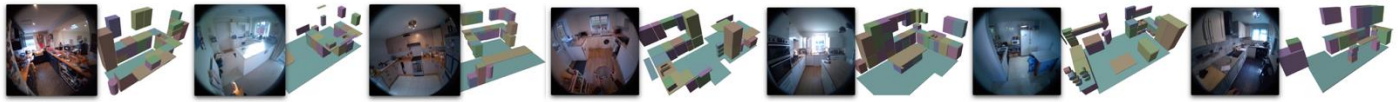
HD-EPIC

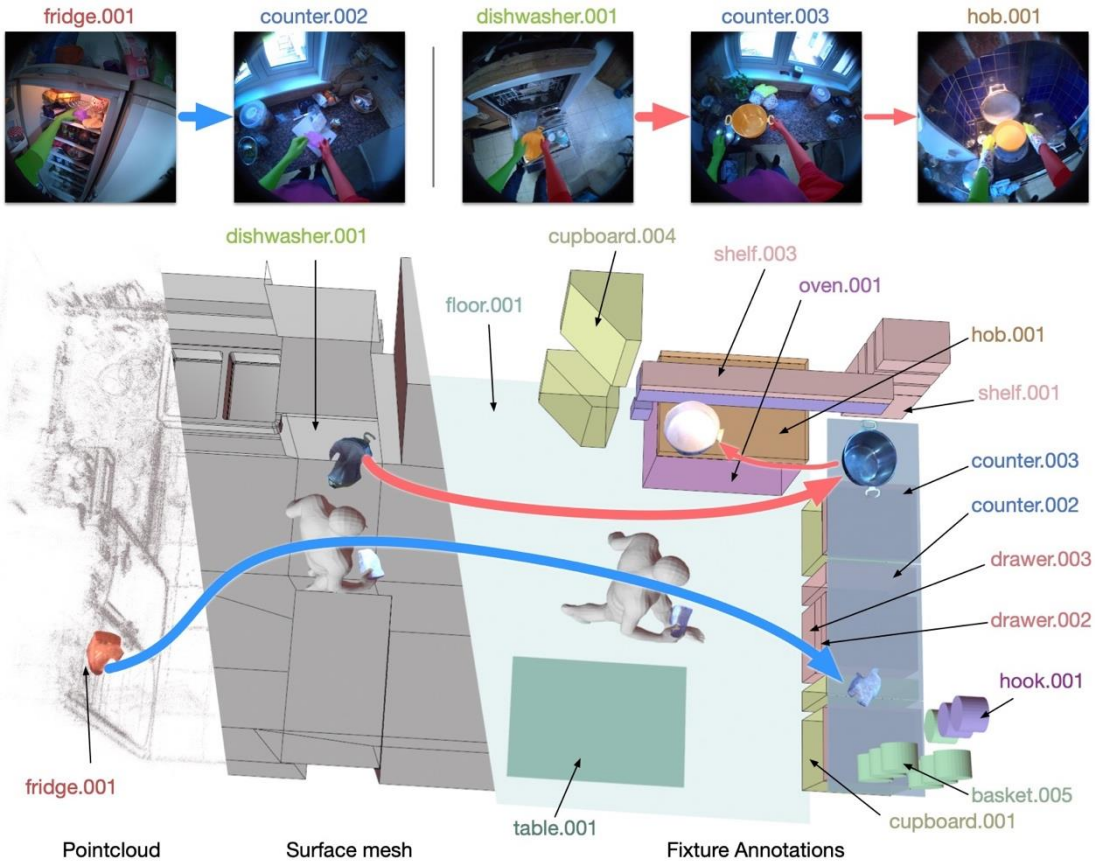


Digital Twin

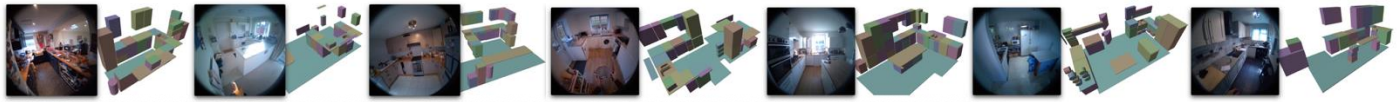
Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

HD-EPIC




fridge.001
counter.002
dishwasher.001
counter.003
hob.001
dishwasher.001
cupboard.004
shelf.003
oven.001
floor.001
hob.001
shelf.001
counter.003
counter.002
drawer.003
drawer.002
hook.001
basket.005
cupboard.001
fridge.001
table.001
Pointcloud
Surface mesh
Fixture Annotations

HD-EPIC



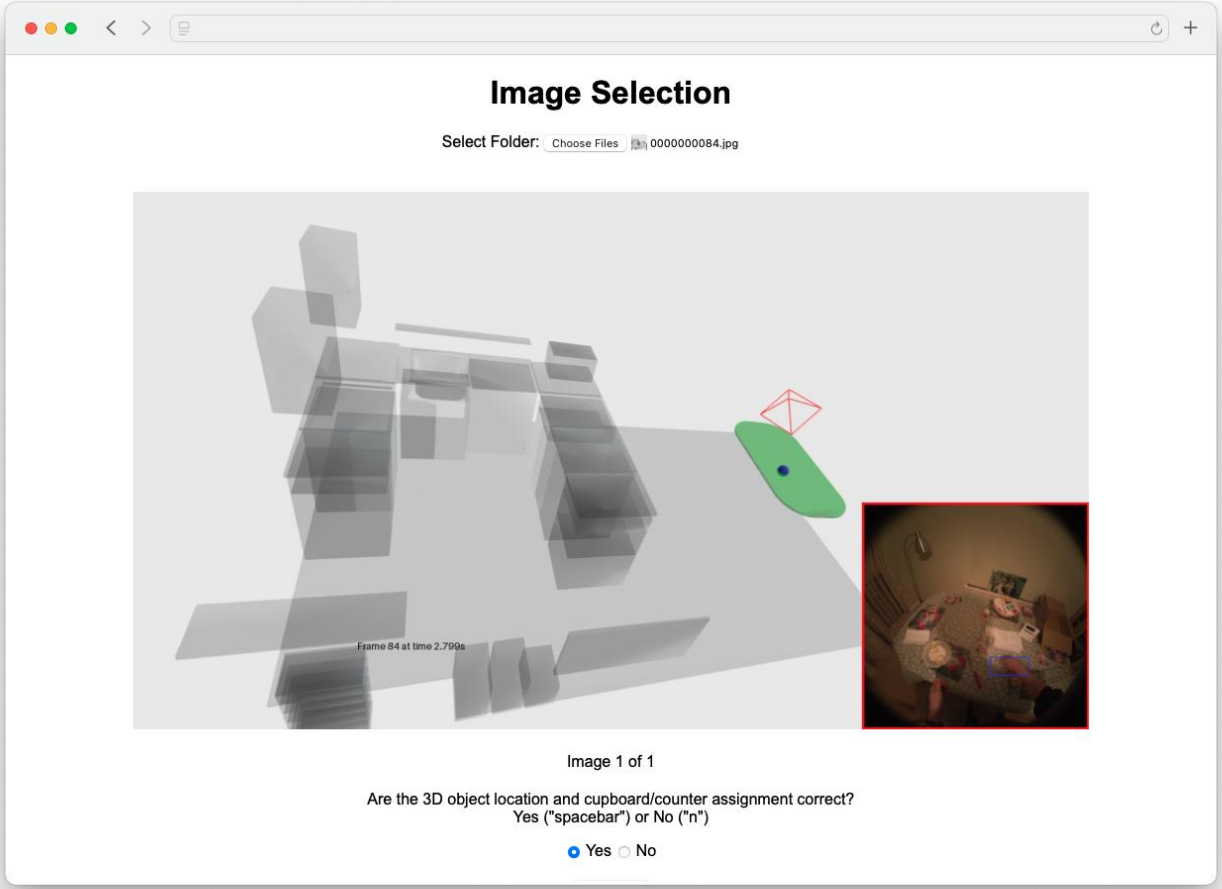
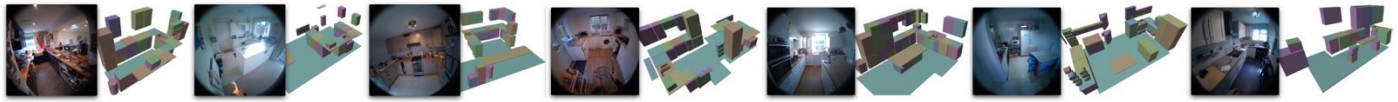
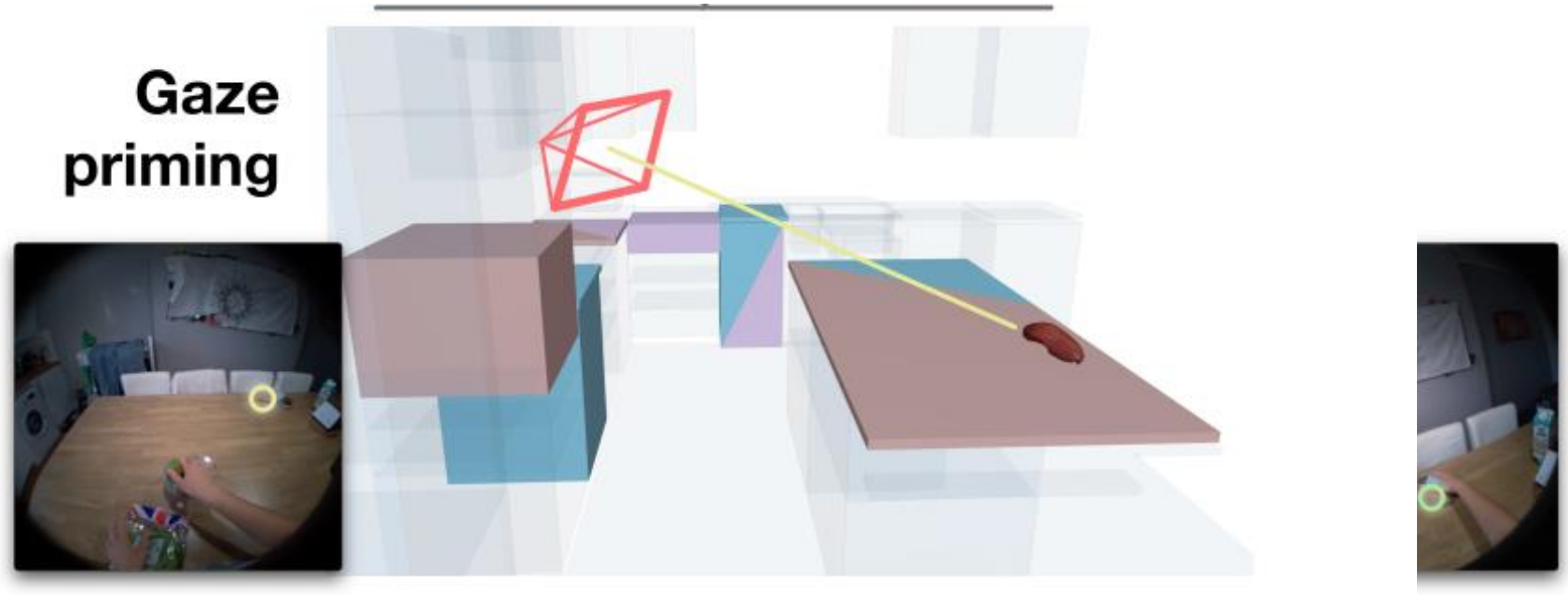


Image Selection
Select Folder:
Choose Files
0000000084.jpg
Frame 84 at time 2.799s
Image 1 of 1
Are the 3D object location and cupboard/counter assignment correct?
Yes ("spacebar") or No ("n")
Yes
No

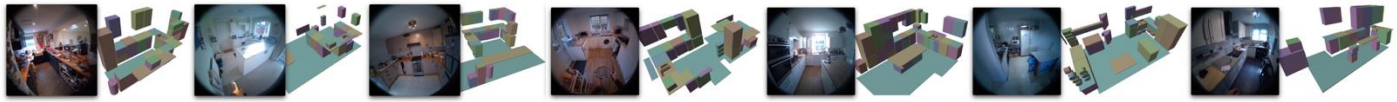
HD-EPIC

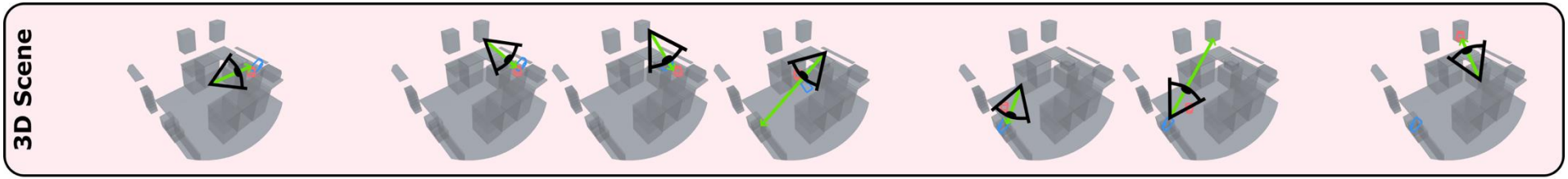


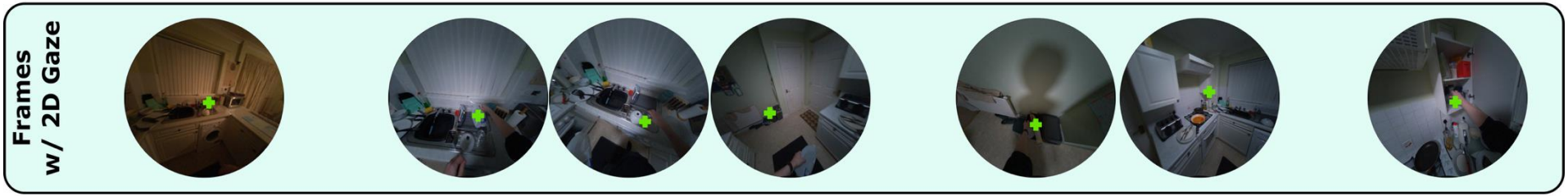


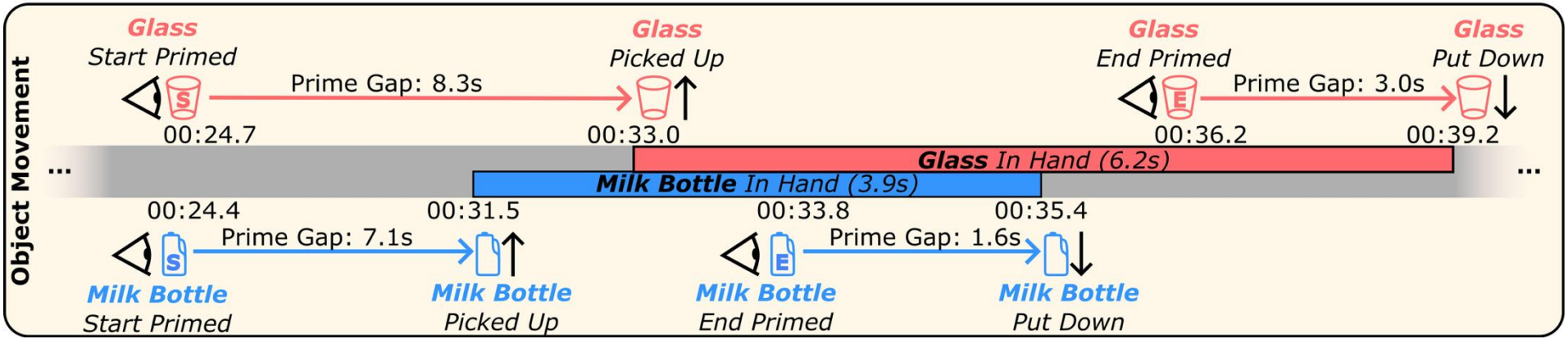
Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

HD-EPIC

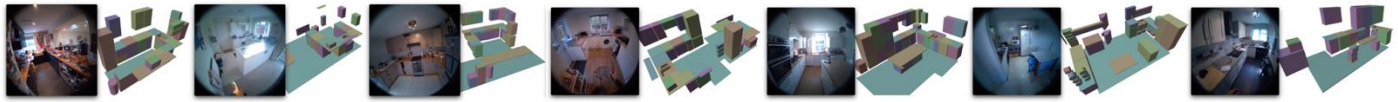


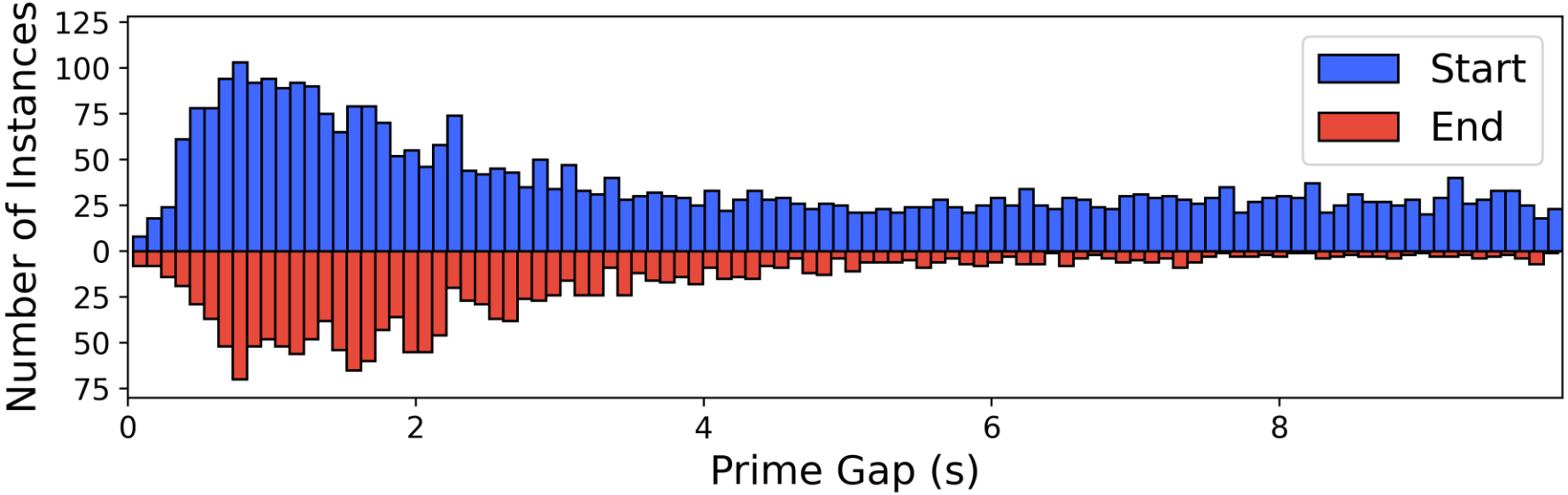

3D Scene


Frames w/ 2D Gaze


Object Movement
Glass
Start Primed
Prime Gap: 8.3s
Glass
Picked Up
Glass
End Primed
Prime Gap: 3.0s
Glass
Put Down
00:24.7
00:33.0
00:36.2
00:39.2
Glass In Hand (6.2s)
Milk Bottle In Hand (3.9s)
00:24.4
00:31.5
00:33.8
00:35.4
Prime Gap: 7.1s
Prime Gap: 1.6s
Milk Bottle
Start Primed
Milk Bottle
Picked Up
Milk Bottle
End Primed
Milk Bottle
Put Down

HD-EPIC



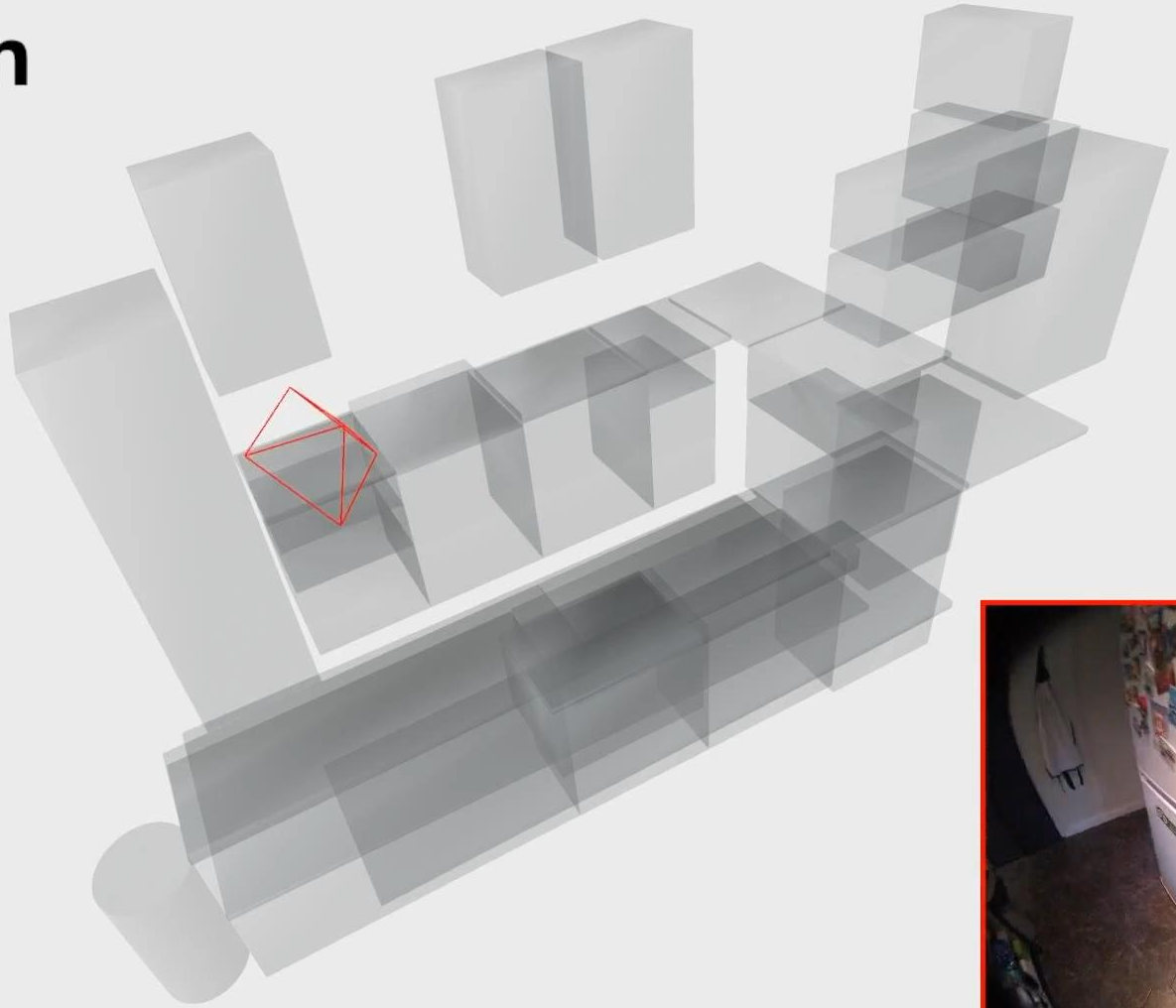


Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

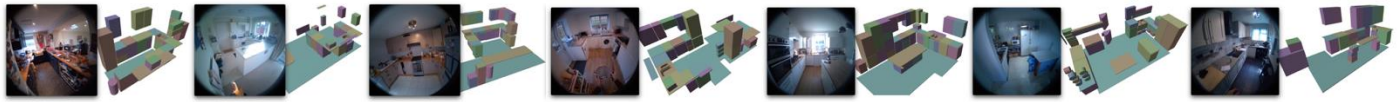
Digital Twin

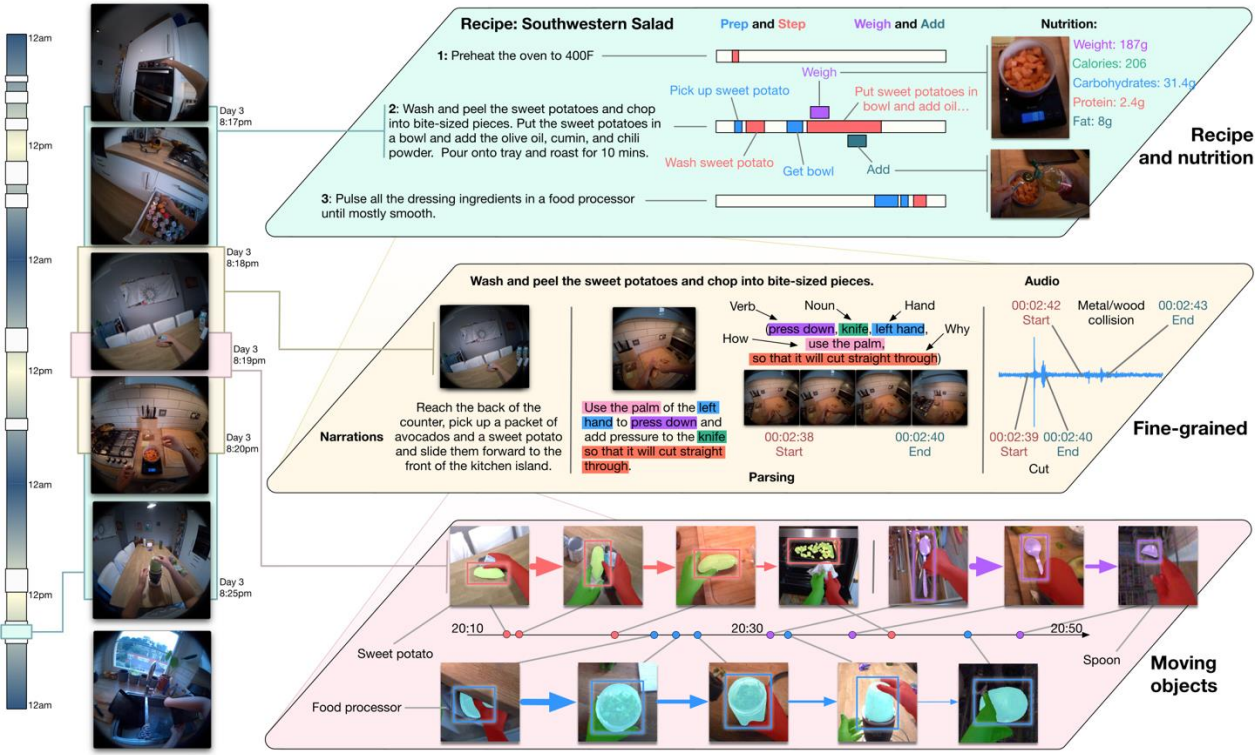
Fixtures

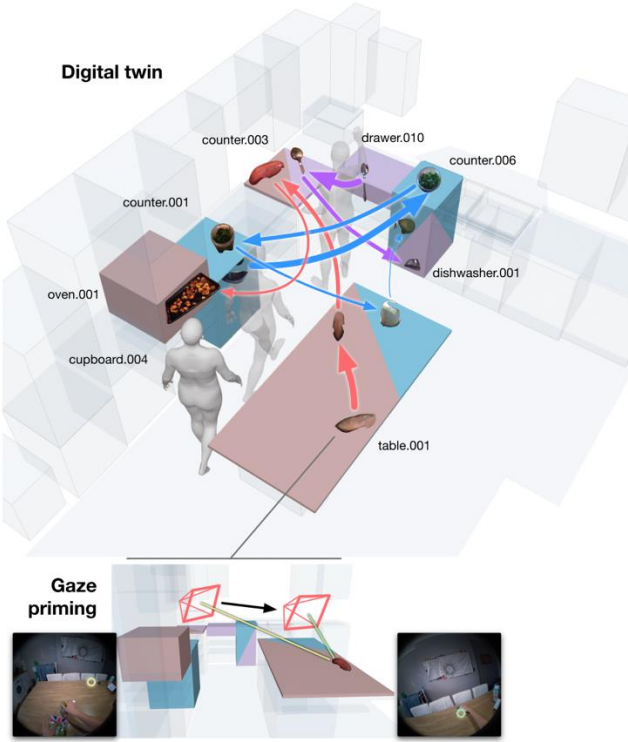
Open drawer



HD-EPIC



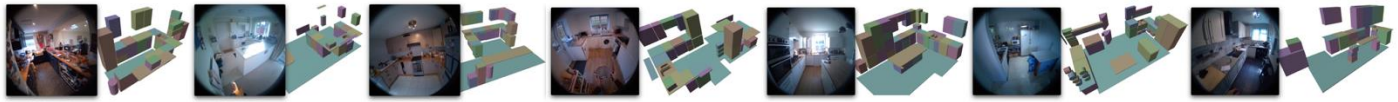




Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

Dima Damen
April 2025

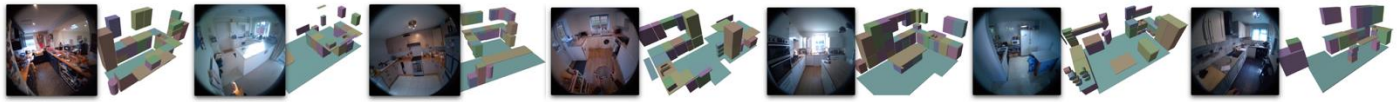
HD-EPIC

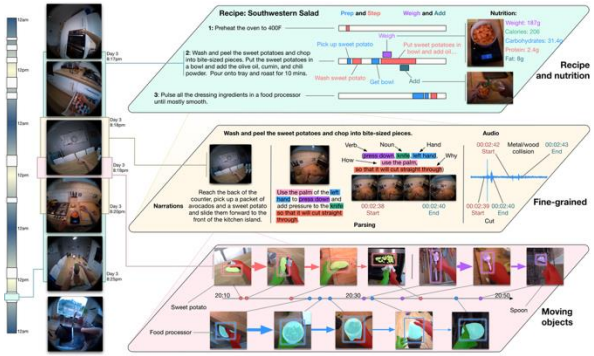


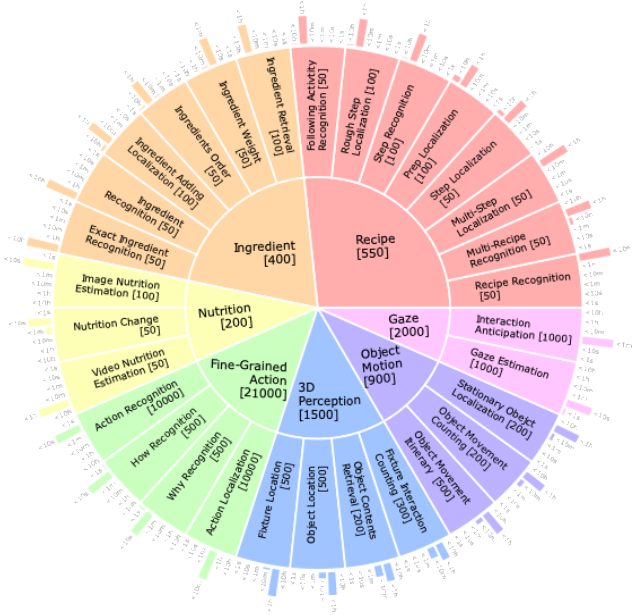
Annotation Type	Total annotations	Annotations/min
Narrations	59,454	24.0
Parsing (Verbs + Nouns + Hands + How + Why)	303,968	122.7
Recipes (Preps + Steps)	4,052	1.6
Sound	50,968	20.6
Action boundaries	59,454	24.0
Object Motion (Pick up + Put down + Fixtures + Bboxes + Masks)	153,480	62.0
Object Itinerary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

Table A3. HD-EPIC annotations per minute

HD-EPIC



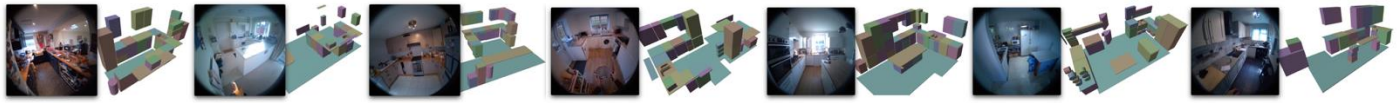




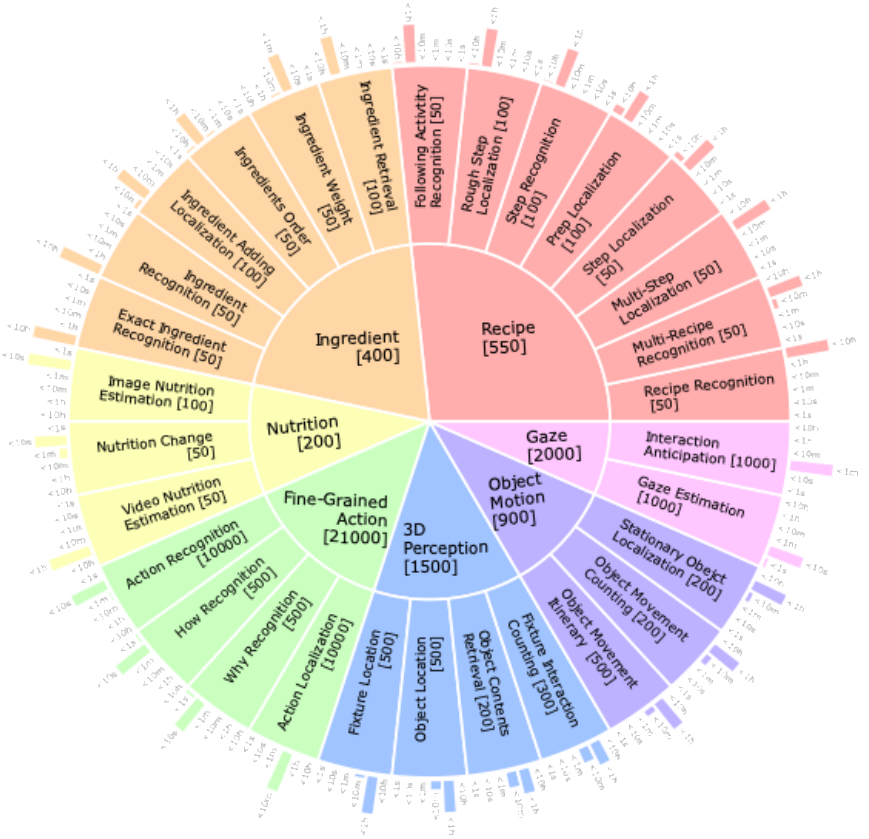
Sec 1: Highly-Detailed Dataset

Sec 2: HD-EPIC VQA Benchmark

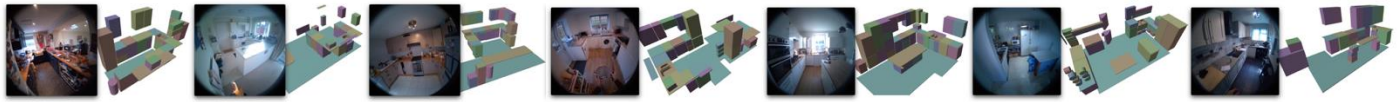
HD-EPIC



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.



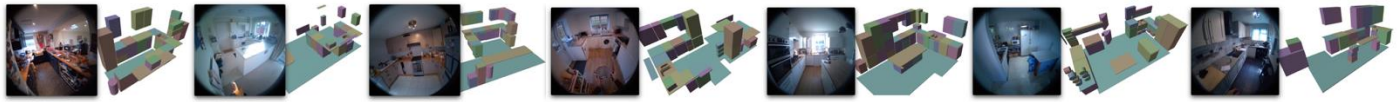
HD-EPIC



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	60.5	46.2	34.7	39.6	32.5	20.8	28.7	37.6
Sample Human Baseline	*96.7*	*96.7*	*85.0*	*92.5*	*93.8*	*92.7*	*75.0*	*90.3*

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

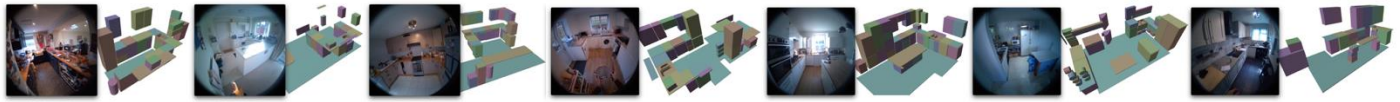
HD-EPIC



Input	Recipe	Ingredient	Nutrition	Action	3D	Motion	Gaze	Avg.
Llama 3.2								
A only	26.8	23.8	14.0	20.2	14.9	15.4	17.8	19.0
Q + A	33.5	25.0	36.7	23.3	22.3	25.3	19.5	26.5
GT Narrations + Q + A	70.8	46.3	34.0	62.5	42.9	28.7	29.4	45.0
Gemini Pro								
A only	29.6	21.0	17.7	19.2	18.9	16.3	18.0	20.1
Q + A	38.0	26.8	30.0	22.1	21.5	27.7	20.5	26.7
GT Actions + Q + A	79.0	54.8	36.3	31.3	42.5	32.8	25.5	43.2
GT Narrations + Q + A	82.6	57.5	36.7	63.6	47.6	38.5	29.0	50.8
Video + Q + A	60.5	46.2	34.7	39.6	32.5	20.8	28.7	37.6

Table A4. **VQA Input Ablation** Our benchmark cannot be solved by analysing Q+A pairs or external knowledge and is a challenge for state-of-the-art closed and open source video VLM models.

HD-EPIC



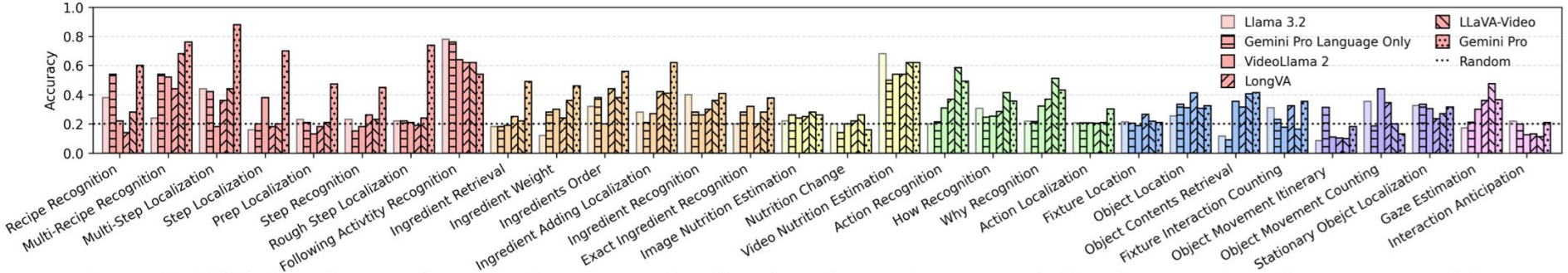
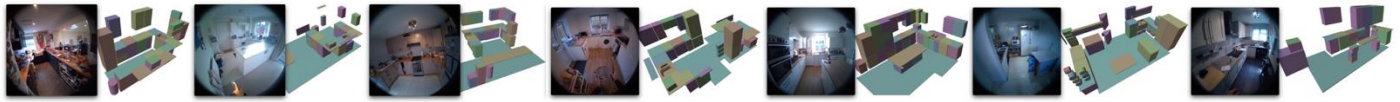


Figure 11. **VQA Results per Question Prototype**. Our benchmark contains many challenging questions for current models.

HD-EPIC



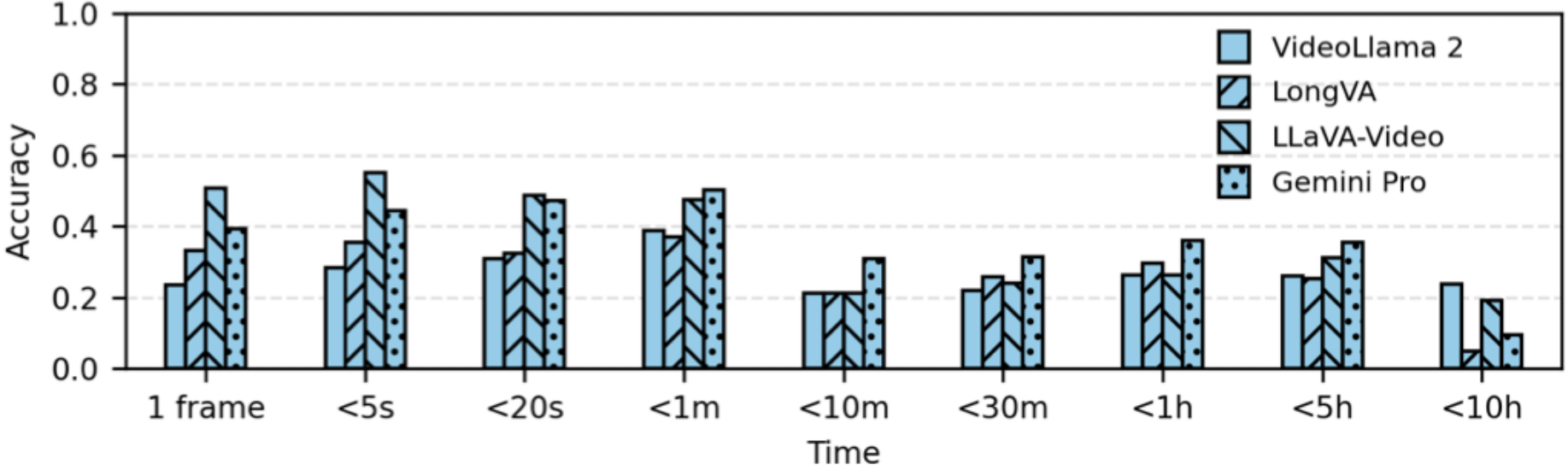
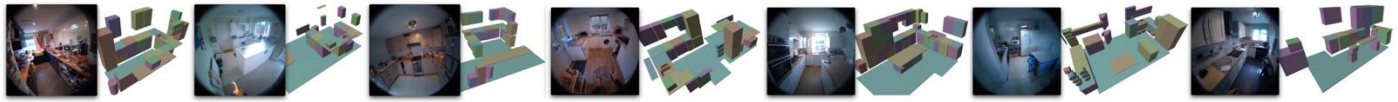


Figure 12. **Effect of Input Length**. Models struggle with questions of all video input lengths. s=second, m=minute, h=hour.

HD-EPIC



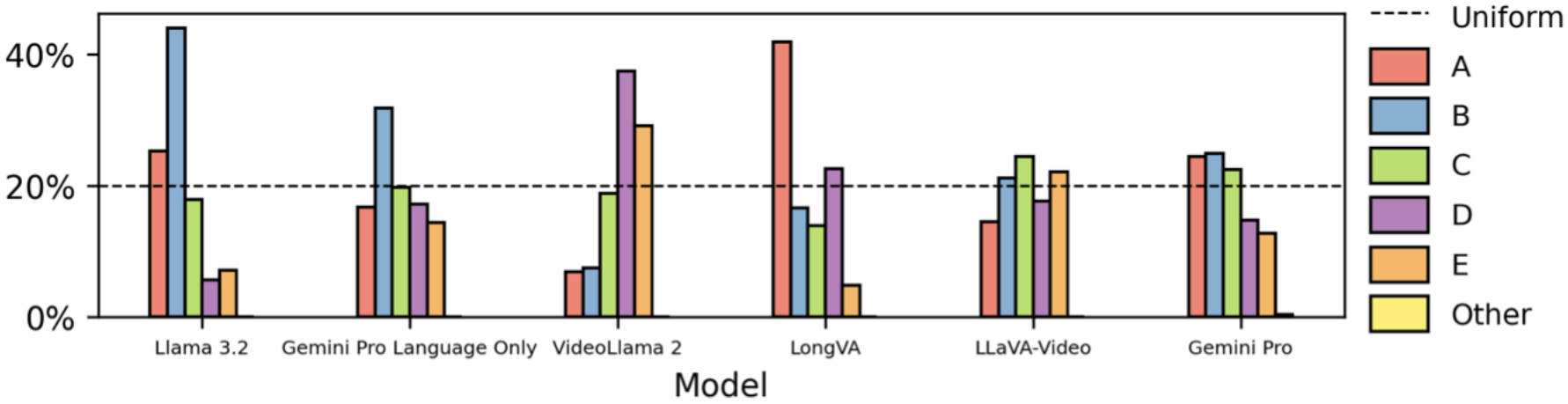
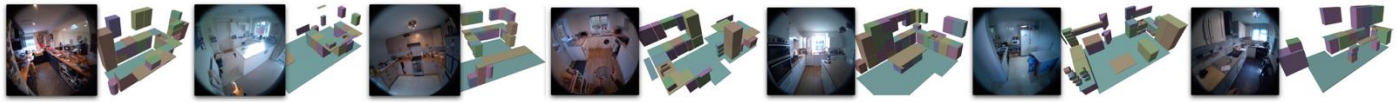
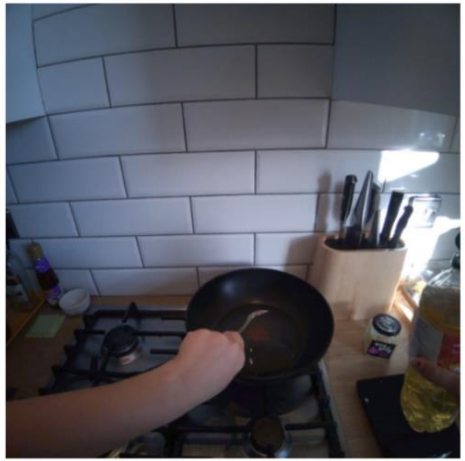
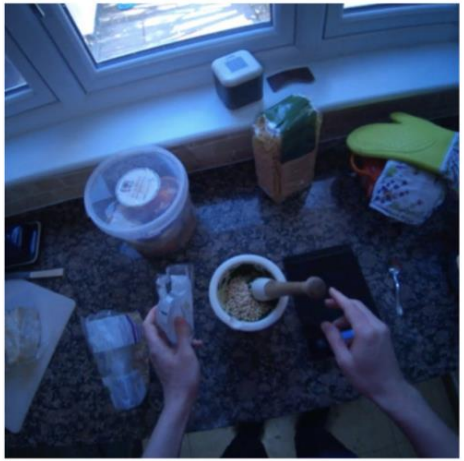
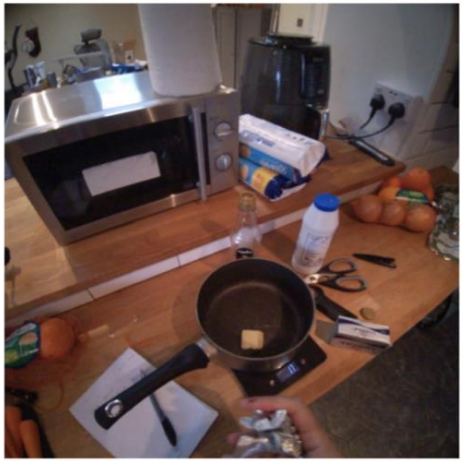


Figure A11. **Prediction Bias of Models**. Most models have a bias in answer, although it is different for each model.

HD-EPIC



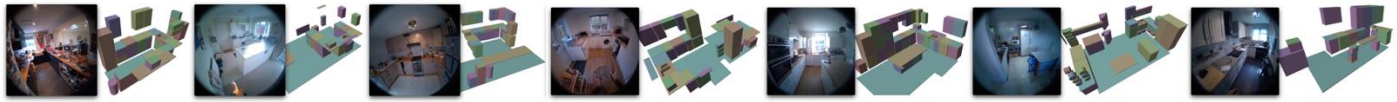

Which of the ingredients in these images showcase higher fat?

A. Oil
B. Pine nuts
C. Salted butter
D. Butter
E. Vegetable oil

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

Dima Damen
April 2025

HD-EPIC

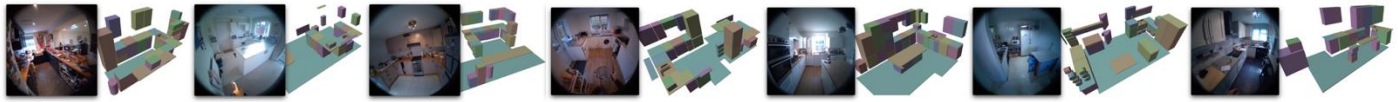


Which of these sentences best describe the action(s) in the video? [00:03:56 - 00:04:03]

- **A.** Wash the cutting board using the sponge in right hand, then, rotate the cutting board so that the back side can be washed
- **B.** With sponge in right hand, clean cutting board while holding board steady with left hand, then with left hand put cutting board under water to clean from soap
- **C.** With left hand, grab cutting board from dish rack, then, with both hands put cutting board down on kitchen counter
- **D.** With my left hand, pick up cutting board, then, with both hands, run the cutting board under water to clean
- **E.** Pick up cutting board from drying rack using right hand, then, dry the cutting board using tea towel in left hand whilst flipping and rotating the cutting board with right hand

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

HD-EPIC

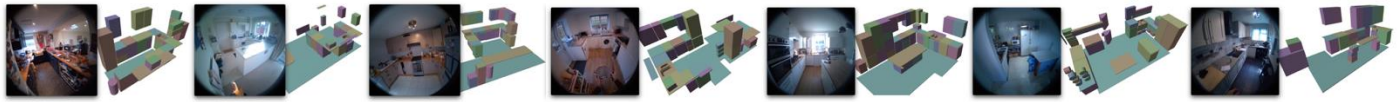


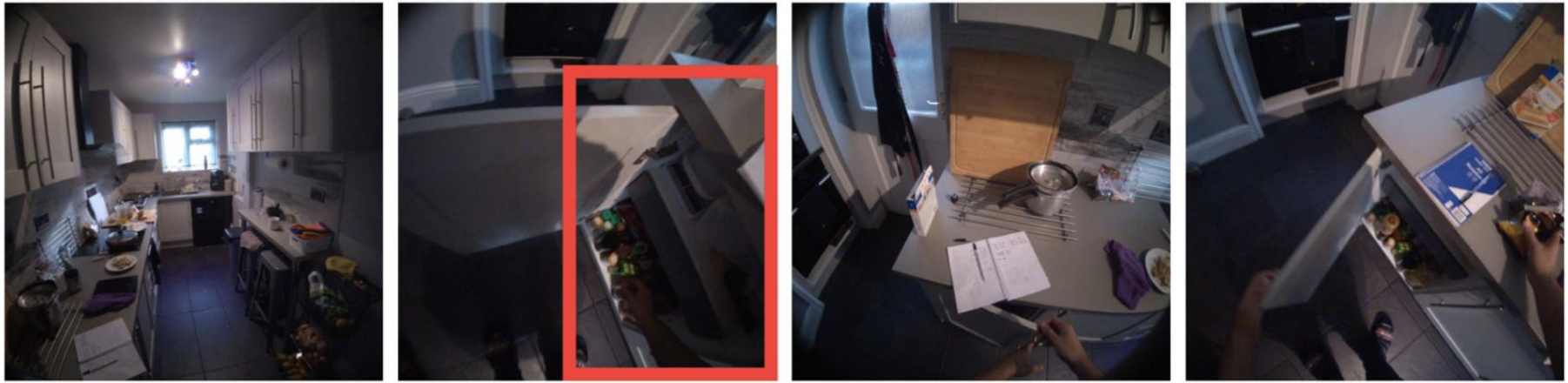
What is the best description for how the person carried out the action pick up bowl of coconut milk in this video segment? [00:18:44 - 00:18:46]

- **A.** Using both hands holding the bowl from bowl rim.
- **B.** By holding both sides using the oven gloves.
- **C.** using the right hand and lift the large white bowl up.
- **D.** using left hand and removing the fork used to stir it using right hand.
- **E.** using both hands from the kitchen top above the dishwasher.

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

HD-EPIC



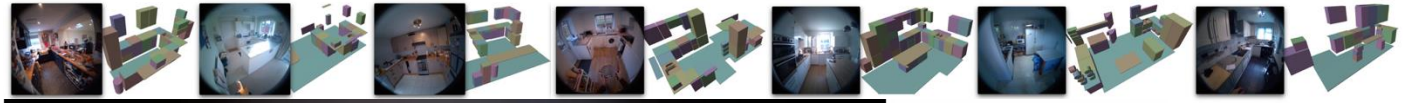


How many times did I open the item at bounding box (165, 452, 1408, 1408) in 00:00:57?

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

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

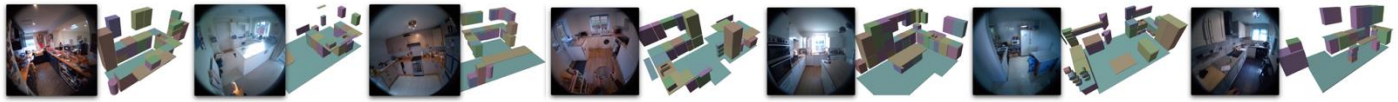
HD-EPIC

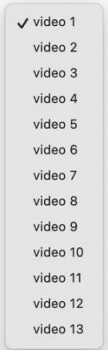


Perrett et al (2025). HD-EPIC: A Highly-Detailed Egocentric Video Dataset. CVPR.

Dima Damen
April 2025

HD-EPIC

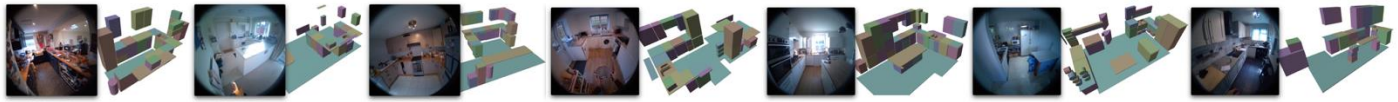


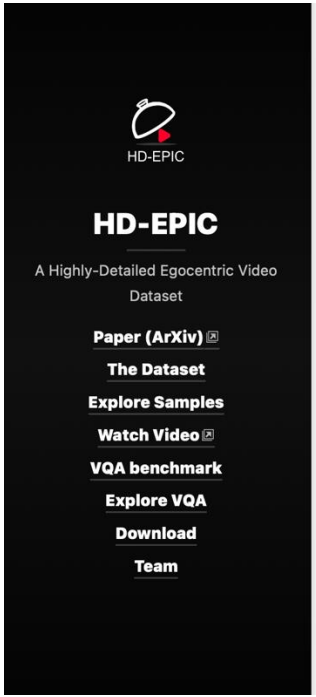


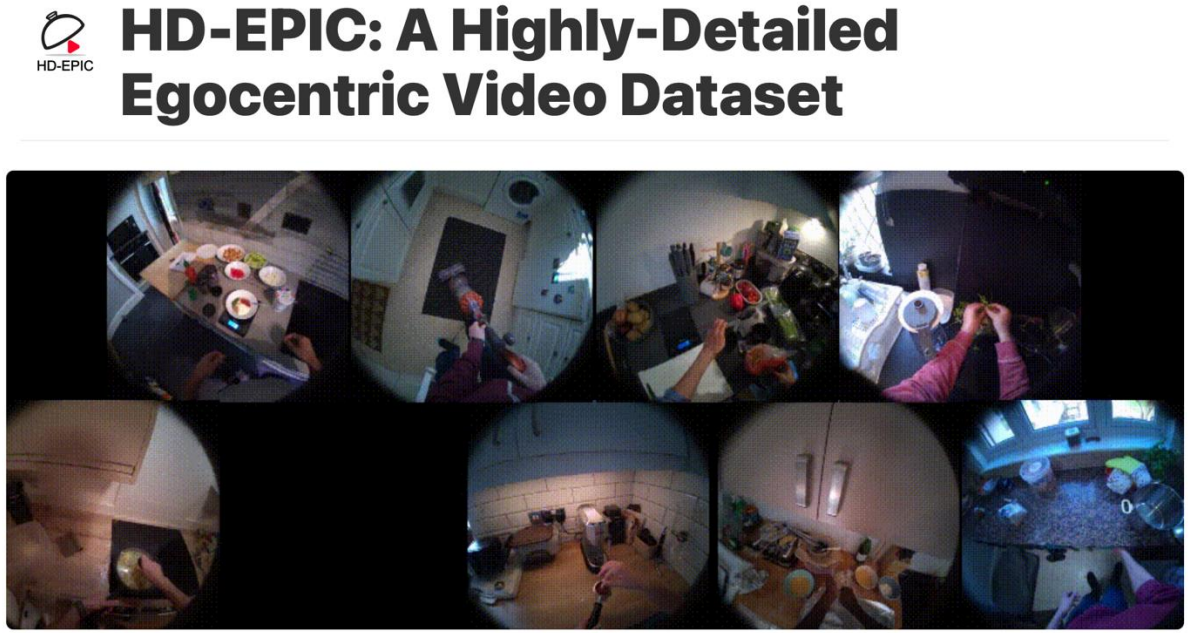
Question: Which of these recipes were carried out by the participant?

- ○ Chicken Korma
- ○ Mutton Vindaloo
- ○ Keema Matar
- ○ Mangsho Bhuna
- ○ Katsu Chicken

HD-EPIC







http://hd-epic.github.io

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

Dima Damen
April 2025