

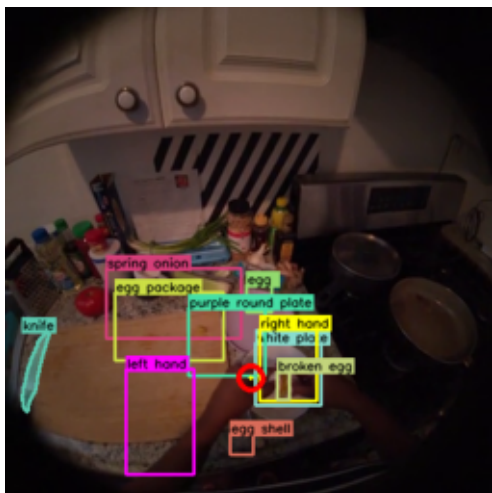
georgiatech_cooking_02_01_2

Ego-Exo4D | Cooking Tomato & Eggs (real egocentric)

24 annotated steps + annotation inventory + methodology



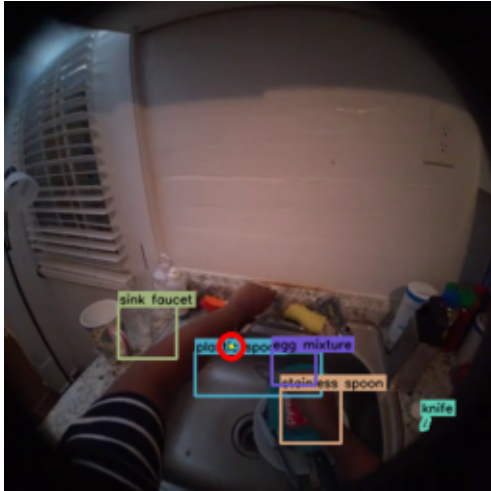
Step 1 [1-28s]: Check paper recipe - Read the paper with tomato and egg recipe (essential) | overlays: red=eye-gaze; masks=knife; boxes=kitchen countertop, recipe manual



Step 2 [28-52s]: Crack eggs into a mixing bowl - Crack an egg into a mixing bowl (essential) | overlays: red=eye-gaze; masks=knife; boxes=broken egg, egg, egg package, egg shell, left hand, purple round plate... | atomic@~40s: "C pours the egg into the bowl with her hands."



Step 3 [53-58s]: Check heat of the skillet - Heat the skillet over a medium heat (essential) | overlays: red=eye-gaze; boxes=frying pan, gas stove | atomic@~56s: "C turns the gas knob with her right hand."



Step 4 [58-72s]: Add water - Add water into a mixing bowl (essential) | overlays: red=eye-gaze; masks=knife; boxes=egg mixture, plastic spoon, sink faucet, stainless spoon | atomic@~65s: "C turns on the tap with her left hand."



Step 5 [72-94s]: Whisk until the egg whites and yolks are well integrated. - Whisk the egg mixture with a spoon (essential) | overlays: red=eye-gaze; masks=knife; boxes=stainless spoon | atomic@~83s: "C reads the recipe paper."



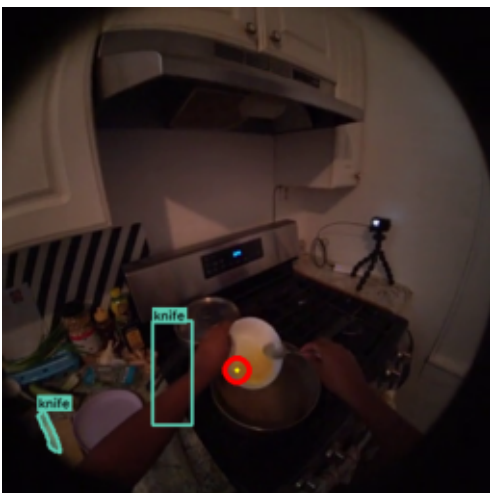
Step 6 [94-115s]: Add oil - Add the oil in a measuring spoon into the skillet (essential) | overlays: red=eye-gaze; boxes=kitchen sink, oil bottle , oil bottle cap | atomic@~104s: "C pours the oil in the measuring spoon into the saucepan with her left hand."



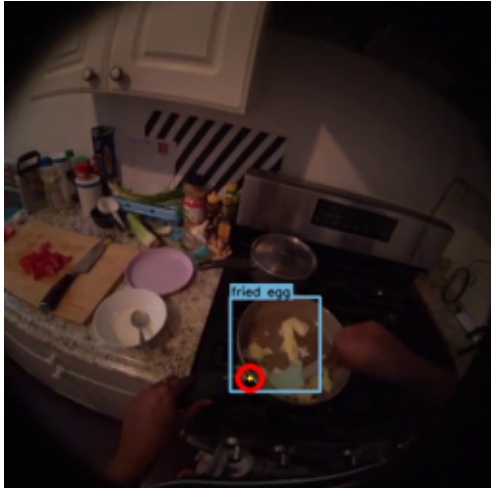
Step 7 [115-125s]: Tilt and rotate the skillet to allow oil or butter flow into empty space - Rotate the skillet to allow oil flow into an empty space (essential) | overlays: red=eye-gaze | atomic@~120s: "C places the sauce pan on the stove top with her right hand."



Step 8 [125-171s]: Cut tomato - Dice the tomato on the chopping board (essential) | overlays: red=eye-gaze; masks=knife; boxes=chopped tomato, cutting board, tomato | atomic@~148s: "C dices the cut tomato on the chopping board with the knife in her right hand."



Step 9 [171-194s]: Pour egg mixture into a skillet - Pour the egg mixture into the skillet (essential) | overlays: red=eye-gaze; masks=knife; boxes=knife | atomic@~182s: "C pours the egg mixture from the bowl into the skillet with her left hand."



Step 10 [194-240s]: As eggs begin to set, gently move spatula across bottom and side of skillet to form large, soft curds. - Stir the egg with a spatula to form soft curds (essential) | overlays: red=eye-gaze; boxes=fried egg | atomic@~217s: "C places the sauce pan on the stove top with her left hand."



Step 11 [240-285s]: Transfer scrambled egg - Transfer the scramble egg into a plate (essential) | overlays: red=eye-gaze; masks=knife | atomic@~263s: "C transfers the cooked egg to the plate with the spatula in her right hand."



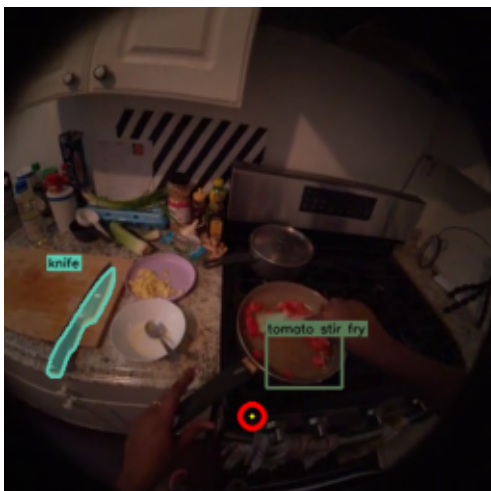
Step 12 [285-307s]: Add oil - Add oil to the skillet (essential) | overlays: red=eye-gaze; boxes=garlic clove, spatula | atomic@~296s: "C pours the oil in the bottle into the measuring spoon with her left hand."



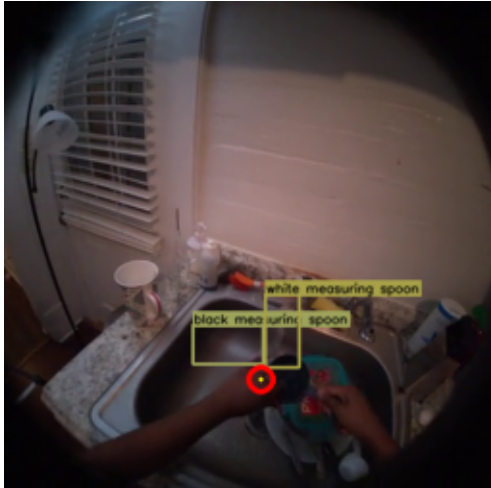
Step 13 [307-327s]: Add fresh ginger - Add fresh ginger into the skillet (essential) | overlays: red=eye-gaze; boxes=chopped garlic clove | atomic@~317s: "C dices the ginger on the chopping board with the knife in her right hand."



Step 14 [328-358s]: Stir the mixture - Stir fry the fresh ginger in the skillet (essential) | overlays: red=eye-gaze | atomic@~343s: "C reads the instruction manual on the countertop wall."



Step 15 [358-396s]: Stir the mixture - Stir fry the mix ingredients with a spatula (essential) | overlays: red=eye-gaze; masks=knife; boxes=tomato stir fry | atomic@~377s: "C regulates the stove heat with her left hand."



Step 16 [396-420s]: Add water - Add water into the skillet (essential) | overlays: red=eye-gaze; boxes=black measuring spoon, white measuring spoon | atomic@~408s: "C pours the water from the measuring cup into the sink with her right hand."



Step 17 [420-449s]: Stir the mixture - Stir the mixed ingredients with a spatula (essential) | overlays: red=eye-gaze; masks=knife



Step 18 [449-485s]: Cut celeries - Cut the celery on the chopping board (essential) | overlays: red=eye-gaze; masks=fried egg; boxes=chopped spring onion | atomic@~467s: "C picks the spring onions cut piece from the chopping board with her left hand."



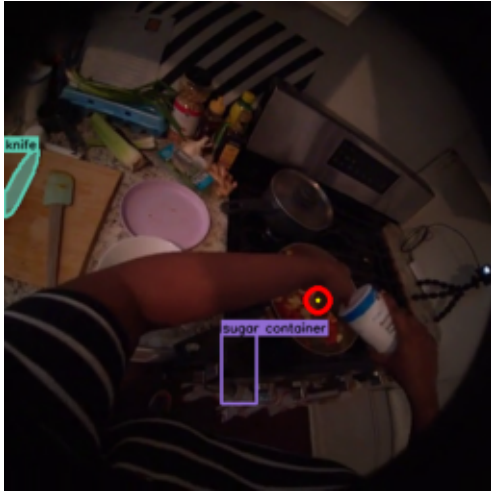
Step 19 [485-533s]: Stir the mixture - Stir the mixed ingredients with a spatula (essential) | overlays: red=eye-gaze; masks=fried egg



Step 20 [533-547s]: Stir the mixture - Stir the mixed ingredients with a spatula (essential) | overlays: red=eye-gaze; masks=knife | atomic@~540s: "C holds the handle of the saucepan with her left hand."



Step 21 [547-584s]: Add salt - Add sea salt into the skillet (essential) | overlays: red=eye-gaze; boxes=salt container, sugar container | atomic@~566s: "C picks a sugar container from the countertop with her left hand."



Step 22 [565-584s]: Add salt - Add sea salt into the skillet (essential) | overlays: red=eye-gaze; masks=knife; boxes=sugar container | atomic@~574s: "C adds sugar to the mix of ingredients in the sauce pan with her right hand."



Step 23 [592-622s]: Stir the mixture - Stir the mixed ingredients with a spatula (essential) | overlays: red=eye-gaze; masks=knife | atomic@~607s: "C stirs the mix of ingredients in the sauce pan with the spatula in her right hand. "



Step 24 [622-641s]: Transfer food to a plate - Transfer the tomato and egg into the plate (essential) | overlays: red=eye-gaze; masks=knife | atomic@~632s: "C places the sauce pan on the stove top with her left hand."

ANNOTATION INVENTORY -- georgiatech_cooking_02_01_2 (Cooking Tomato & Eggs)

take: 644s, 19312 frames @ 30fps ; ego RGB aria02_214-1.mp4 (448x448) ; scenario 'Cooking Tomato & Eggs'.

1. Eye-gaze 6437 per-frame 2D samples (also 3D + summary.json) -> red dot
[SPATIAL, sensor]
2. Keystep subgoals 24 segments (name+desc+essential) -> caption [TEMPORAL, human]
3. Relations - boxes 35 tracks (objects + L/R hands), 35 box annots -> labelled rectangles
[SPATIAL, human]
4. Relations - masks 2 objects masked (~1Hz COCO-RLE) -> filled+outlined [SPATIAL, human]
5. Atomic descriptions ... 492 timestamped action sentences -> caption / video subtitle
[TEMPORAL, human]
6. Expert commentary 2 spoken tip/error entries [TEMPORAL, expert]
7. Proficiency demo not annotated for this take [take-level, expert]

(Fields present in Ego-Exo4D but not spatial / not in the CoF focus scope:

procedure_understanding,

ego_pose, proficiency_demonstrator, physical_setting, visual_objects, participants -- listed for completeness.)

How these annotations were produced (Ego-Exo4D) -- ALL are real human/sensor labels on real video:

- Eye-gaze (red dot): Aria glasses' inward eye-tracking cameras; machine-perception computes per-frame 3D

gaze and projects it into the RGB camera (~10 Hz). The wearer's ACTUAL gaze, not a manual label.

- Keystep / subgoals: annotators segment the take into fine procedural steps from a per-scenario taxonomy;

each = start/end time + step name + description + 'essential' flag.

- Relations -- object/hand MASKS (filled + outlined) and BOXES (labelled rectangles): annotators use an

interactive point-click segmentation tool on SAMPLED frames (masks ~1 Hz) to mask/track task-relevant

objects and both hands; the same object is corresponded across ego<->exo views. Masks are stored as

LZString-compressed COCO-RLE; only a subset of objects is masked per take (varies 2-11 here).

- Atomic descriptions: narrators write timestamped fine-grained action sentences (subject + ego-visible flag).

- Expert commentary: a domain expert records spoken tips/error notes at timestamps.

- Proficiency demonstration (some takes): expert lists good executions + tips-for-improvement.